YRS 2016

YOUNG RESEARCHERS SYMPOSIUM 2016

14th – 15th April 2016 ▪ Fraunhofer-Zentrum Kaiserslautern

PROCEEDINGS

FRAUNHOFER VERLAG

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Young Researchers Symposium 2016 (YRS 2016)

Proceedings

April, 14th - 15th 2016
Fraunhofer-Zentrum Kaiserslautern

Organizers and Editors

TU-Nachwuchsring
https://nachwuchsring.uni-kl.de

Innovationszentrum Applied System Modeling for Computational Engineering (ASM4CE)
www.applied-system-modeling.de

FRAUNHOFER VERLAG

Imprint

Contact
Innovationszentrum Applied System Modeling for
Computational Engineering (ASM4CE)
c/o Fraunhofer-Zentrum
Fraunhofer-Platz 1
67663 Kaiserslautern
http://www.applied-system-modeling.de/

TU Kaiserslautern
TU-Nachwuchsring
Geschäftsführung Dr. Berenike Schröder
Gottlieb-Daimler-Straße, Gebäude 47
7663 Kaiserslautern
https://nachwuchsring.uni-kl.de

Bibliographic information published by Die Deutsche Nationalbibliothek

Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliografic data is available in the Internet at www.dnb.de.

Printing and Bindery:
Mediendienstleistungen des
Fraunhofer-Informationszentrum Raum und Bau IRB, Stuttgart

Printed on acid-free and chlorine-free bleached paper.

© by Fraunhofer Verlag, 2016, ISBN 978-3-8396-1010-7
Fraunhofer Information-Centre for Regional Planning and Building Construction IRB
P.O. Box 80 04 69, D-70504 Stuttgart
Nobelstrasse 12, D-70569 Stuttgart
Phone +49 (0) 7 11/9 70-25 00
Fax +49 (0) 7 11/9 70-25 07
E-Mail : verlag@fraunhofer.de
URL www.verlag.fraunhofer.de

Dear Colleagues,

we are delighted to welcome you to Young Researchers Symposium 2016!

Almost five years ago, the $(CM)^2$- Nachwuchsring invited young researchers at the University of Kaiserslautern to present their research projects to a broad scientific community. This is how the Young Researchers Symposium (YRS) came to be. The YRS is an interdisciplinary platform for informal exchange and networking which provides an overview of the research activities in Kaiserslautern.

Since then, the YRS has steadily grown. At its second edition it saw the addition of the Kaiserslautern Fraunhofer Institutes, including the Innovationszentrum Applied System Modeling. As the $(CM)^2$-Nachwuchsring has since been developed into the university-wide TU-Nachwuchsring, with the third edition we now open the YRS 2016 to all young researchers from Kaiserslautern for the first time!

Networking, exchange and interdisciplinarity – the YRS 2016 offers the opportunity to look beyond one's own scientific community and to meet other young researchers. The high number of submitted contributions clearly indicates how important the YRS is for networking and for discussion. As in previous years, the best contributions will be awarded according to the assessment of a jury and by votes of the attendees. For the first time posters are also presented - they will be part of the competition as well.

We would like to offer our gratitude to all the authors for the contributions they have submitted. We are also delighted by all the other attendees in the auditorium who, with their questions and discussions, make also valuable contributions to the success of the event.

We wish all participants a successful and interesting YRS 2016!

Prof. Dr. Arnd Poetzsch-Heffter,
Vice President for Research and Technology, University of Kaiserslautern and Patron of YRS 2016

Scientific Committee

Jun.-Prof. Dr.-Ing. Frank **Balle**
(Department of Mechanical and Process Engineering, University of Kaiserslautern)
Dr.-Ing. Christian **De Schryver**
(Department of Electrical Engineering and Computer Engineering, University of Kaiserslautern)
Sylvia **Gerwalin**
(Public Relations, Fraunhofer ITWM)
Prof. Dr.-Ing. Hans **Hasse**
(Department of Mechanical and Process Engineering, University of Kaiserslautern)
Dr. Joachim **Jonuscheit**
(Department of Materials Characterization and Testing, Fraunhofer IPM)
Dipl.-Ing. Matthias **Jung**
(Department of Electrical Engineering and Computer Engineering, University of Kaiserslautern)
Dr. habil. Thorsten **Kowalke**
(Office for Research Affairs and the Promotion of Young Researchers, University of Kaiserslautern)
Prof. Dr. Karl-Heinz **Küfer**
(Dept. of Optimization, Fraunhofer ITWM and Dept. of Mathematics, University of Kaiserslautern)
Dr. Thomas **Kuhn**
(Department of Embedded Software Engineering, Fraunhofer IESE)
Dr. Andreas **Morgenstern**
(Department of Embedded Software Engineering, Fraunhofer IESE)
Prof. Dr.-Ing. Ralf **Müller**
(Department of Mechanical and Process Engineering, University of Kaiserslautern)
Christian **Peper**
(Department of Embedded Software Engineering, Fraunhofer IESE)
Prof. Dr. Arnd **Poetzsch-Heffter**
(Vice President for Research and Technology, University of Kaiserslautern and Patron of YRS 2016)
Dr. Isabel **Sattler**
(State Research Center OPTIMAS)
Dr. Berenike **Schröder**
(Office for Research Affairs and the Promotion of Young Researchers, Manager of TU-Nachwuchsring, University of Kaiserslautern)
Dr. Konrad **Steiner**
(Department of Flow and Material Simulation, Fraunhofer ITWM)

Organizing Committee

Jun.-Prof. Dr.-Ing. Frank Balle
Dr.-Ing. Christian De Schryver
Sylvia Gerwalin
Dr. Isabel Sattler
Dr. Berenike Schröder

Table of contents

Table of contents

Tool-supported extraction of conceptual interoperability constraints of software units from UML Diagrams

Hadil Abukwaik
Department of Computer Science
University of Kaiserslautern
Kaiserslautern Germany
Email: abukwaik@cs.uni-kl.de

Dieter Rombach
Department of Computer Science
University of Kaiserslautern
Kaiserslautern Germany
Email: rombach@cs.uni-kl.de

***Abstract*—Successfully integrating an external software unit into a system requires software architects to check the conceptual constraints of this unit to ensure that it has no mismatches with the system. However, such constraints about software units are usually hidden within their architectural documents that are not publicly shared with clients. Hence, owners of the software units need to search for the constraints in the architecture document and provide them to clients. However, this manual task is not trivial and it is time consuming especially in the case of large software systems. In this paper, we demonstrate a tool-supported, systematic approach for extracting the interoperability-related constraints of software systems from their architecture and lower-level design documentation. Our proposed approach aims at helping architects in performing the conceptual interoperability analysis tasks in an effective and efficient manner. To bring the approach into practice, we developed an aiding tool that assists architects with easy-to-use interfaces. We plan to evaluate our approach empirically through a controlled experiment where we will test our hypotheses about its positive effect on architects effectiveness and efficiency in performing the interoperability analysis.**

Keywords- conceptual interoperability; interoperability analysis; information extraction; software architecture; tool support

I. INTRODUCTION

Interoperability is the ability of two or more separately-developed software units to communicate and exchange data in a seamless and meaningful way [1]. This ia an important property as it plays a vital role in today's large software systems. For example, a system-of-systems [2] needs interoperability to put together a number of information systems in order to achieve more functionality and better performance (e.g., a health care system-of-systems that combines application of doctor, patient, and pharmacy for better treatment). Similarly, cyber-physical systems [3] need interoperability to link computational software units with physical resources (e.g., an automated traffic control cyber-physical system that connects traffic lights and vehicles on the road). Having this being said, to successfully integrate software units, it is required to identify and resolve any technical or conceptual interoperability mismatches. While technical mismatches (e.g., different communication protocols, programming languages, data types, arguments orders, etc.) hinder the exchange of information and services among software units; conceptual mismatches (e.g., usage context, architectural constraints, terminologies, qualities, etc.) lead to meaningless or undesirable interoperation results. Such conceptual mismatches are expensive to resolve and should be found early in the integration project before spending effort on resolving the technical mismatches.

According to a scoping study we performed earlier [4], existing works focus mainly on achieving the technical level of interoperability of software units and lack the focus on the conceptual level. In the context of black-box interoperations, in which the source code of a software unit is not publicly shared, the typically available source of information about a software unit is the documentation of its Application Programming Interface (API). This API documentation describes the software unit in terms of its functionality, input, output, and data type. Hence, interested clients in a software unit investigate its API documentation to find the technical constraints [5].

However, API documentation of a software unit is insufficient source of information for a software architect who is assessing the conceptual interoperability mismatches. This is because API documentations are typically technical oriented and do not expose all conceptual and architectural constraints that are usually hidden in the unshared architectural documents. These architecture documents are usually written using the popular Unified Modeling Language (UML) [6] that enables describing the software structure (e.g., component diagram) and behavior (e.g., state chart diagram). On another hand, it is a tedious and time consuming task for the owners of a software unit to manually find the interoperability-relevant pieces of information about a specific software unit from across the whole system's UML diagrams, and then to add these information to the shared API documentation for interested clients.

In this paper, we present a tool-supported approach for extracting the **CO**nceptual **I**nteroperability co**N**straints (COINs) from UML diagrams. This work is an extension of our previously proposed framework for supporting interoperability analysis [7]. The goal of our presented COINs' extraction

approach is to alleviate the burden from software architects' shoulders in explicitly publishing the conceptual information about their interoperable software units, and accordingly to support interested clients in performing proper interoperability analysis in order to effectively detect the conceptual mismatches. This is achieved through the: (1) semi-automatic extraction of COINs from internal UML documentation, and (2) automatic documentation of extracted COINs in a standard ready-to-publish document that we call the "COINs Portfolio". Our work offers advantages for software companies that build interoperable systems. That is, on the short-term planning, we expect that the proposed extraction idea would help these companies increasing their effectiveness and efficiency in identifying and documenting the conceptual constraints of their interoperable software systems. Besides, on the long-term planning, these companies would grow their business value as a result of publishing sufficient interoperability information about their systems that would lead to more successful interoperations with less integration effort.

The rest of the paper is organized as follows: We start with a motivation scenario in Section II. Then, we describe our COINs extraction approach in Section III. In Section IV and V, we discuss the planned evaluation for our work and related work accordingly. We finally conclude and present future work in Section VI.

II. Motivating Example

This section outlines a brief example for the application of our tool-supported COINs extraction approach.

Imagine a company *Alpha* that has developed an *ATM (S1)*, which is a cash machine developed with the goal to make its software system interoperable. That is, *S1* is intended to exchange data and services in a meaningful way with other separately developed software systems (e.g., mobile apps of bank clients, bank management systems, etc.). Hence, the company has created the API documentation of *S1* and published it for interested third-party clients.

Afterwards, another company *Beta* gets interested in integrating an instance of *S1* within their *Bank system (S2)*. The responsible person for assessing the feasibility of building a successful interoperation in this scenario is the software architect of *S2*, Noah. He is responsible for analyzing the interoperability constraints for both *S1* and *S2* and he aims at detecting any conceptual mismatches between them. Hence, Noah starts with a manual investigation of the text in *S1*'s API documentation for its offered functionalities and related data (e.g., "Withdrawal" function and its related data element "Amount" that represents the requested money to be withdrawn). Based on this investigation, Noah could identify some technical mismatches, for example, he finds that *S1* uses the Secure Socket Layer (SSL [8]) protocol to ensure the security over communications, while *S2* uses another security protocol called the Transport Layer Security (TSL [9]). However, Noah could not detect some of the conceptual mismatches because their related constraints are hidden in the unshared UML diagrams of *S1*. Here are some explaining examples:

- A class diagram of *S1* shows that the "Withdrawal" transaction has a "history" file that contains up to 100 transactions, while *S2* assumes that the "history" contains all withdrawal transactions on the account within a year (which may be more than 100);
- A use case diagram of *S1* states that the "Withdrawal" transaction is permitted only for the owner of the account, while *S2* allows shared account for a husband and wife where some mechanism is used to handle simultaneous access);
- A sequence diagram of *S1* shows that the interaction type with the "Withdrawal" functionality is synchronous meaning that the user can make a second withdrawal transaction before receiving a confirmation on the first. While, *S2* uses asynchronous interactions that require a confirmation on a transaction before allowing the second one.

Obviously, having such information in the API document would enhance the results of Noah's analysis and would save the cost spent on handling the unexpected conceptual mismatches late in the integration project.

A possible approach to handle this issue is to contact company X asking for further information about *S1*. However, adopting this solution might be expensive for company X especially for repeated inquiries. Also, the waiting time might be inconvenient for clients. Alternatively, we propose that company X publishes the relevant conceptual information of *S1* in advance for interested clients. To make this practical and avoid the manual, time-consuming effort , especially for large systems, we support architects with a tool that semi-automate the extraction and documentation of the conceptual constraints of their interoperable systems!

III. Tool-supported Extraction of Conceptual Interoperability Constraints

This section describes our COIN model, the extraction process, tool support, and planned evaluation.

A. The Conceptual Interoperability Constraint (COIN)

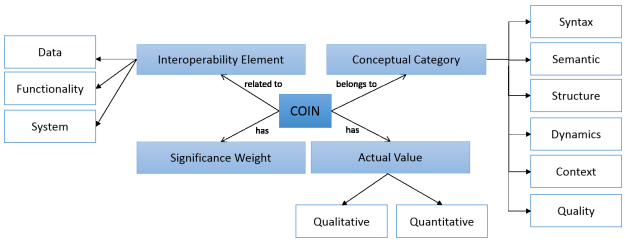


Fig. 1. A COIN meta-model.

The corner stone of our work is the **COINs** or the conceptual interoperability constraints, which we define as the conceptual characteristics of the software unit that if they are misassumed, they may lead to conceptually-wrong, meaningless, or improper interoperation results [7]. In Figure 1, we present our COIN meta-model and show its structure. Each

COIN in the interoperable software system is related to an element (i.e., data, function, component, or system), belongs to a conceptual category (i.e., syntax, semantic, structure, dynamic, context, and quality), has a value (i.e., qualitative and/or quantitative description), and has a significance weight (i.e., high, medium, or low).

The COINs we target in our extraction approach are the ones that represent a cause of conceptual interoperability mismatch between two software systems. Figure 2 presents the whole set of COINs and their categories with examples. Note that, the shaded rows outlines the so far supported COINs with the current version of our extraction tool.

Category	COIN name	Examples of value
Syntax	Lexical references	Dictionary, thesaurus, glossary, etc.
	Modeling lang.	XML, UML, ADL, WSDL, etc.
Semantic	Semantic references	Reference ontologies
	Semantic constraints	Data units and scale ratio
Structure	Data structural constraints	Invariants, inherited constraints, and multiplicity constraints
	Data storage	Relational database, flat files, etc.
	Distribution	Distributed, centralized
	Encapsulation	Encapsulated , not encapsulated
	Concurrency	Single-threaded, multi-threaded
	Layering	Layered, not layered
Dynamic	Data change	Periodic, irregular, continuous, etc.
	Service conditions	Pre, post, and time conditions
	Interaction properties	State(ful/less), (a)synchronous, etc.
	Interaction time constraints	Session timeline, acknowledgment timeline, response timeline, etc.
	Communication style	Messaging, procedure call, blackboard, streaming
Context	Usage context	device type, wired/wireless, access rate, time, location, etc.
	Intended users	Human/machine, gender, age, etc.
Quality	Data quality	Security, trust, accuracy, etc.
	Service quality	Safety, availability, efficiency, etc.

Fig. 2. Conceptual interoperability constraints (COINs).

B. Extracting COINs from UML Diagrams

As we mentioned in Section I and II, creating the COINs Portfolio manually is a tedious and expensive task. This is beacause it requires sifting through the UML documentation of the whole software system, and then extracting the only useful pieces of information for interoperability analysis. To address this issue, we have previously proposed an abstract idea about a Portfolio Generator [7]. In this subsection and the one after we describe this idea in details along with its supporting tool.

In our approach, we aim at gathering all related COINs of an interoperable system into a standard document called the "COINs Portfolio". The input to our COINs' extraction approach is a consistent, complete, and up-to-date UML document about the interoperable software unit. This UML document includes structural diagrams (e.g., component diagram, deployment diagram, class diagram, etc.) and behavioral diagrams (e.g., use case diagram, sequence diagram, etc.). This input goes through the following four activities in order to result in the desired COINs Portfolio (as seen in Figure 3):

Identification of interoperable elements. In this activity, the software architect identifies the UML elements (i.e., components, classes, use cases, or actors) of the system, which are involved in interoperations with other software systems. This identification happens in terms of assigning an "Interoperability Type" property for the element. For example, in the ATM example described in Section II, the "Withdrawal" use case in the use case diagram would have the interoperability property declared, and similarly this would be declared for the the "Amount" data element in the class diagram. Declaring this property for the elements directs the subsequent activities of the COINs extraction.

Automatic extraction of COINs. To enable the proposed automation in this activity, an Interoperability Knowledge Base (IKB) is created and charged with COIN extraction templates from UML diagrams. A COIN extraction template is a set of rules that if it is met by an interoperable element within a UML diagram, then a COIN candidate for the element is created and added to the list of COIN candidates. An example of a context COIN template is:

- **if** there is a use case within a use case diagram that is identified as interoperable element (e.g., the "Withdrawal" use case within the use case diagram of *S1*),
- **and** it is inherited from another use case (e.g., "Withdrawal" use case is one kind of the "Transaction" use case,
- **and** the inherited from use case has a constraint (e.g., the "Transaction" use case has a constraint that only the account owner can do it),
- **then**, a context COIN will be added to the list of COIN candidates for the use case element (e.g., the "Withdrawal" use case has a constraint that only the account owner can do it.

Note that the extraction activity starts with checking each UML diagram of the systems to find if it contains any element with the interoperability property declared. Then, only the interoperable elements are checked against the predefined COIN templates saved in the IKB.

Manual filtering for the extracted COIN candidates. This activity is performed manually by the system architects who have the final word to approve or disapprove the automatically extracted COINs within the final published COINs Portfolio. Furthermore, they can manually add more COINs to the Portfolio if they see it important and useful to share with clients.

Automatic generation for the COINs Portfolio. Finally, the approved and the manually added COINs are bundled together and categorized according to the elements they are related to. Then, they are documented in a standard form that is ready to share with clients who will use it for their interoperability analysis task if interested in interoperating with the software system.

C. The CoinsExtractor Tool

To make our proposed idea applicable in practice, we implemented the CoinsExtractor tool [10] that assists architects with easy-to-use interfaces in extracting and publishing the COINs of the interoperable units of their software systems. This section summarizes the main features, design, and limitations of the tool (for details please refer to [10]).

Main features. The CoinsExtractor has a number of features that supports architects through the aforementioned ac-

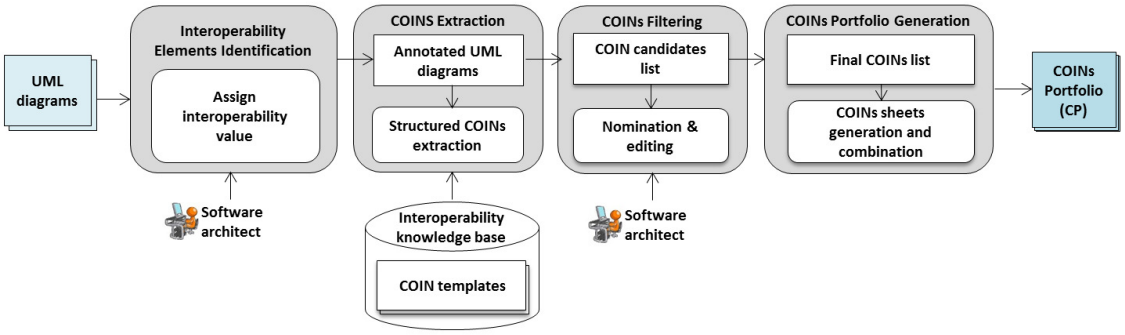


Fig. 3. COINs extraction process.

tivities in Subsection B, which takes some burden off the architects' shoulders as the following:

- *Interoperability Tags*. With these self-implemented tags for the UML elements, the architect can directly determine their systems' interoperable elements. Accordingly, if an element is tagged as interoperable, all its instances within all the diagrams get tagged as interoperable too. The tool maintains a table of interoperable elements, which represents the architect knowledge about the interoperable units of the system and it would be reused in all future integration projects.
- *COINs Extraction*. This feature implements the COIN templates that we described earlier in Subsection B. Hence, the tool saves the architect's time and effort by automatically looking for the relevant COINs about the interoperable elements only. In addition, this feature guarantees consistency in terms of what COINs are being extracted and how they are documented across projects of interoperable units or systems.
- *COINs Review*. The CoinsExtractor tool realize the COINs filtering by enabling the architects to efficiently review, update, approve, or delete the automatically extracted COINs. It offers two views: (1) *diagram-based view* where COINs can be navigated according to the source diagram that they are found in, and (2) *element-based view* where COINs can be navigated according to the the interoperable element they belong to.
- *COINs Portfolio Generation*. Finally, the tool creates a ready-to-share, web-based COINs Portfolio (see an example in Figure 4). The Portfolio arranges the COINs according to their categories to facilitate the conceptual interoperability analysis task that will be performed by interested clients.

Design and Implementation. The CoinsExtractor has a multi-layered architecture with a *presentation layer* that is responsible for tool-architect interactions, *business layer* that includes the logic units responsible for processing the UML diagrams and extracting their COINs, the *data access layer* that reads input from the UML database and writes results into the output file, and finally the IKB that stores all predefined COIN templates as we described in the previous subsection. These layers lead to the tool modularity that allows extending its COIN templates to cover more categories and accordingly enhancing the interoperability analysis results. The CoinsExtractor tool is implemented as an extension for the Enterprise Architect (EA) application, which is one of the powerful, widely-used architecture modeling tools [11].

Limitations. The current version of the tool accepts input (i.e., UML diagrams) created by the Enterprise Architect only. It also assumes that the UML input is created according to the UML standard notations specified by the Object Management Group (OMG) [12] and published by the International Organization for Standardization (ISO) [13]. To ensure that these assumptions are realistic enough to use the tool in industry, we have reviewed the tool with architect experts and we used their feedback to enhance the tool usability.

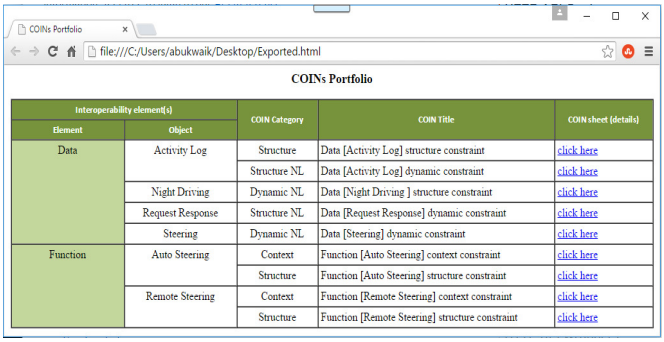


Fig. 4. COINs Portfolio generated by CoinsExtractor [10].

IV. Planned Empirical Evaluation

To empirically evaluate the ideas that we have presented in this paper, we plan for conducting a controlled experiment. The goal of this experiment, formulated by means of the Goal-Question-Metric template [14], is to *analyze* the tool-supported approach for extracting COINs from UML diagrams *for the purpose of* evaluation *with a focus on* effectiveness, efficiency, and acceptance *from the perspective of* software architects *in the context of* a controlled experiment with students. For more

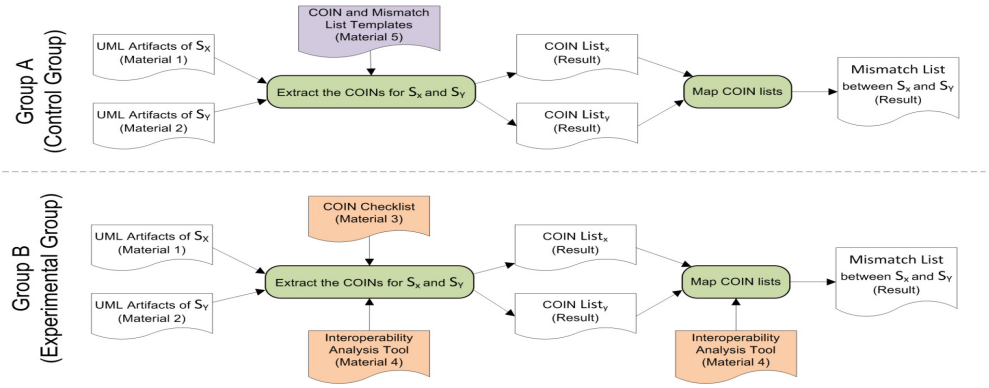


Fig. 5. Experimental design for evaluating the analysis approach.

details about the experiment (i.e., detailed information about the research questions, process, material, and data analysis plan) please refer to [15].

Figure 5 outlines the experimental design of the planned controlled experiment, in which we plan to have two groups of students who we hand them the same input (i.e., UML diagrams of two software systems) and we ask them to find the the COINs of each system in the first step and their mismatches in the second one. While, the control group applies ad-hoc approach in analyzing the input, the experimental group applies our proposed approach along with the support of our CoinsExtractor tool. Once we execute the experiment, we will analyze the collected data from both groups and compare between them with regards to correctness and completeness, and then we will publish the results.

V. Related Work

A. Detecting Mismatches in Black-box Context

Reuse analysis approaches have been proposed for finding component mismatches. Bhuta et al. [16] propose to manually create component definitions for technical and architectural assumptions, then to apply mismatch-detection rules on them. Our approach extends the architectural part of this work by further conceptual constraints to reduce the possibility of facing risks of unplanned conceptual mismatches. Besides, we try to reduce the cost by finding the already documented constraints in the UML with automation support, rather than writing definitions manually from scratch. **Testing-based techniques** such as Halle's [5] are useful in identifying mismatches, though, they require preparing complete, high-quality test suites and launching interoperation for each test case. Such intensive testing is expensive and impractical due to invocation costs. **Prototyping methods** for analyzing software components propose simulating its usage within other systems [17]; however, this may not always be feasible as it requires acquiring the component, learning, and evaluating it. **Conformance checking** has an active research community with design-by-contract enforcement techniques [18] where formal specification of methods pre-, post- conditions, and invariants are used in static [19] or dynamic [5] conformance analysis. Although this enables detecting technical conformance violations automatically, it does not address the conceptual ones. Therefore, our proposed idea stretches the conformance checking to the conceptual level.

B. Tools for Interoperation Analysis

Previous works have proposed tools to support the interoperability analysis task. For example, Bahuta et al. presented a tool called, the Integration Studio (iStudio) [16], which performs automatic assessment of architectures and proposes mismatch resolution. Compared to our CoinsExtractor tool, iStudio depends on a completely manual specification for the architectural interfaces, while our tool saves this cost through the semi-automatic extraction of such information from available UML documents. In addition, we extend the architectural attributes covered by iStudio with further conceptual features like context and quality. Another interoperability supporting tool was proposed by Ullberg et al. [20] for analyzing enterprise architecture models. This tool allows assessment theories to be specified using a formal modeling language (i.e., extension of the Object Constraint Language- OCL [21]), and then it utilizes it in the interoperability analysis. Relating this to our CoinsExtractor, our tool infers the conceptual features that would be the input for such assessment theories. On other hand, Buschle [22] developed a tool that analyzes properties including interoperability to provide decision-making support for information technology in enterprise architecture models. To the best of our knowledge, the CoinsExtractor is the first tool to semi-automate the extraction of conceptual interoperability constraints from UML diagrams, which aids architects and eliminates unexpected conceptual mismatches.

VI. Conclusion and Future Work

In this paper we have presented an approach for extracting the conceptual interoperability constraints of an interoperable software system from its UML diagrams. The approach aims at enhancing the effectiveness and efficiency of software architects in extracting and publishing their systems' COINs with third-part clients. We have also outlines the features of the supporting tool that we have implemented as an extension for the Enterprise Architect software application.

Based on our ongoing research, we aim to create further COIN templates to add them in our interoperability knowledge base, and to transform the COINs Portfolio into a formal notation to enable the automatic mapping of two interoperating systems to find their mismatches. We also plan to execute the designed controlled experiment and to evaluate the tool within industrial case studies too.

ACKNOWLEDGMENT

This work is part of the PhD research of Hadil Abukwaik under the supervision of Prof. Dieter Rombach and is funded by the PhD Program of Kaiserslautern University. The authors would like to thank Mohammed Abufouda, Shah Rukh Humayoun, and the anonymous reviewers for their constructive comments.

REFERENCES

[1] A. Geraci *et al.*, *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*, 1991.
[2] W. A. Crossley, "System of systems: An introduction of purdue university schools of engineering's signature area."
[3] R. Baheti and H. Gill, "Cyber-physical systems," *The impact of control technology*, vol. 12, pp. 161–166, 2011.
[4] H. Abukwaik, D. Taibi, and D. Rombach, "Interoperability-related architectural problems and solutions in information systems: A scoping study," in *Software Architecture*. Springer International Publishing, 2014, vol. 8627, pp. 308–323.
[5] S. Hallé, T. Bultan, G. Hughes, M. Alkhalaf, and R. Villemaire, "Runtime verification of web service interface contracts," *Computer*, vol. 43, no. 3, pp. 59–66, 2010.
[6] G. Booch, *The unified modeling language user guide*. Pearson Education India, 2005.
[7] H. Abukwaik, M. Naab, and D. Rombach, "A proactive support for conceptual interoperability analysis in software systems," in *Working Conference on Software Architecture. 2015. WICSA'15*, 2015.
[8] A. Freier, P. Karlton, and P. Kocher, "The secure sockets layer (ssl) protocol version 3.0," 2011.
[9] E. Rescorla, "The transport layer security (tls) protocol version 1.1," *Transport*, 2006.
[10] H. Abukwaik, M. Abujayyab, and D. Rombach, "Coinsextractor: The architects' buddy in identifying interoperability-relevant architectural constraints," in *European Conference on Software Architecture. 2015. ECSA'15*, 2015.
[11] "Enterprise Architect," http://sparxsystems.de/, Sparx System, accessed: 2015-05-08.
[12] O. CORBA and I. Specification, "Object management group," *Joint revised submission OMG document orbos/99-02*, 1999.
[13] J. Koppell, "International organization for standardization," *Handb. Transnatl. Gov. Inst. Innov*, vol. 41, no. 8, p. 289, 2011.
[14] V. R. Basili, "Software modeling and measurement: the goal/question/metric paradigm," 1992.
[15] H. Abukwaik, "Empirical evaluation for the conceptual interoperability analysis approach- controlled experiment design," Kaiserslautern University, CS Department, Tech. Rep., 2014. [Online]. Available: https://kluedo.ub.uni-kl.de/frontdoor/index/index/docId/3915
[16] J. Bhuta, "A framework for intelligent assessment and resolution of commercial-off-the-shelf product incompatibilites," Ph.D. dissertation, University of Southern California, 2007.
[17] R. Land, L. Blankers, M. Chaudron, and I. Crnković, "Cots selection best practices in literature and in industry," in *High Confidence Software Reuse in Large Systems*. Springer, 2008, pp. 100–111.
[18] B. Meyer, "Applying'design by contract'," *Computer*, vol. 25, no. 10, pp. 40–51, 1992.
[19] B. Rubinger and T. Bultan, "Contracting the facebook api," *arXiv preprint arXiv:1009.3715*, 2010.
[20] J. Ullberg, U. Franke, M. Buschle, and P. Johnson, "A tool for interoperability analysis of enterprise architecture models using Pi-OCL," in *Enterprise Interoperability IV*. Springer, 2010, pp. 81–90.
[21] O. A. Specification, "Object constraint language."
[22] M. Buschle, P. Johnson, and K. Shahzad, "The enterprise architecture analysis tool–support for the predictive, probabilistic architecture modeling framework," in *AMCIS 2013*, 2013, pp. 3350–3364.

Experimental investigation, modeling and simulation of vibration welded nanocomposites

Michael Albrecht[1]
[1]Chair of Composite Engineering
University of Kaiserslautern
Kaiserslautern, Germany
E-mail: michael.albrecht@mv.uni-kl.de

Leyu Lin[1]
[1]Chair of Composite Engineering
University of Kaiserslautern
Kaiserslautern, Germany
E-mail: leyu.lin@mv.uni-kl.de

Alois K.Schlarb[1,2,3]
[1]Chair of Composite Engineering
University of Kaiserslautern
Kaiserslautern, Germany
[2]Research Center OPTIMAS
University of Kaiserslautern
Kaiserslautern, Germany
[3]INM – Leibniz Institute for New Materials
Saarbrucken, Germany
E-mail: alois.schlarb@mv.uni-kl.de

I. INTRODUCTION

Composites are materials made from a matrix material, filled with another material in order to improve or add functional or mechanical properties, such as stiffness or Young's Modulus. Nanofillers in particular are a promising subject of research as they, compared to micro-scale filler materials, have geometries with at least one dimension of less than 100 nm and thus high surface to volume ratios, which affect the interaction between matrix and filler material and increase their capability of surface-dependent properties. While vibration welding, the most common joining technique for thermoplastics, has been shown to be capable of achieving weld strengths equivalent to the strength of injection molded parts for neat polymers at low welding pressures [1], nanocomposites [2] were shown to have lower weld strengths. This can be attributed to orientation effects during welding [1,2,3] and its unique morphological layer structure [1,2].
The aim of this work was to characterize, model and simulate the deformation behavior and morphology-dependent local properties of neat PP and a PP/TiO_2-nanocomposite, welded at two different pressures each, in microtensile test.

II. EXPERIMENT, MODEL AND SIMULATION

A. *Materials and processes*

Polypropylene (PP HD120 MO, Borealis GmbH, Vienna, Austria) was used as a matrix material with titatium dioxide (TiO_2) nanoparticles (Hombitec RM130 F, Huntsman, Duisburg, Germany) as inorganic nanofillers. They were compounded by a double-step extrusion of compounding and then diluting a masterbatch on a co-rotating twin-screw extruder (Theysohn, Theysohn Extrusionstechnik GmbH, Salzgitter, Germany) as described in literature [4] to optimize dispersion of nanoparticles. PP filled with 1 vol.-% of nano-TiO_2 (PP-T-V1) was then injection molded (Arburg Allrounder 420C, ARBURG GmbH + Co KG, Loßburg, Germany) to sheets of 50x50x4 mm³.

B. *Vibration welding*

Vibration welding was performed for both materials, neat PP and PP/TiO2 with 1 vol.-% of filler, at welding pressures of 0.4 MPa and 2.0 MPa each, a welding amplitude of 0.7 mm and a frequency of vibration of 240 Hz on a M-112H (Branson Ultraschall, Dietzenbach, Germany).

C. *Microtensile Testing*

Welded parts were then cut into thin films of 0.02x4x20 mm³ size (directions of amplitude, melt flow and meltdown) with a microtome (HYRAX M25, Carl Zeiss, Göttingen, Germany). Microtensile tests have then been perfomed on a Mikrozugmodul 5kN (Kammrath&Weiss, Dortmund, Germany) at a speed of 20 µm/s in meltdown direction and a starting length of 9 mm as shown in Figure 1. During microtensile test, sample morphology and deformation have been investigated and recorded by transmitting light microscopy (Nikon Eclipse LV100, Nikon GmbH, Düsseldorf, Germany). Using records of microscopy, local deformations of different layers (bulk material B, deformed spherulites D and inner area I / weld line W as described in previous works [1, 2]) of the weld have been calculated.

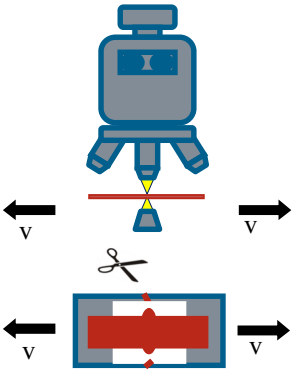


Figure 1: Setup of the micro-tensile test in transmitting light microscopy in front-view (top) and top-view of the sample (bottom).

D. *Modeling and simulation*

Based on the geometry of the morphology measured, a model of the weld and nearby bulk material has been created by using the element solid 186, consisting of 20 nodes on the corners and edges of a cuboid in ANSYS. By using an isotropic multilinear model and attaching local stress-strain curves measured during micro-tensile testing, a simulation of it was performed for small deformations in order to calculate the stress distribution inside the sample. In order to implement the effect of morphology and local properties, von Mises stress was compared.

III. RESULTS AND DISCUSSION

A. *Micro-tensile properties*

In comparison with tensile test results [2], the same effect of adding filler and increasing weld pressure on the reduction of tensile strength has been found for micro-tensile test as can be seen in Table 1. The results show, however, much smaller values for tensile strength and Young's modulus by about 30%, which is caused by the difference stress-distribution (plane stress in the film specimens and plain strain in the bulky tensile test specimens), the higher elongation rate at tested setup and the bigger impact of flaws in sample preparation due to the small cross-area.

Table 1: Ultimate tensile strength of tested non-welded and welded material in tensile and micro-tensile.

Weld pressure	No weld	0.4 MPa
	σ_u, MPa	
Tensile PP	39.5 ± 2.5	39.7 ± 0.5
Micro-tensile PP	29.1 ± 0.6	27.5 ± 2.2
Tensile PP-T-V1	37.9 ± 0.7	31.7 ± 1.7
Micro-tensile PP-T-V1	27.3 ± 1.4	21.3 ± 2.4

While neat PP welded at a low pressure of 0.4 MPa was breaking statistically in all layers, PP welded at 2.0 MPa and filled PP welded at both pressures break inside the weld only. Local deformations under tensile load have been found to be higher inside the weld than in the bulk material with their Young's modulus lower than in the bulk material, respectively as shown in Figure 2. This is due to the effect of highly oriented nanofiller in weld layers as described in [2].

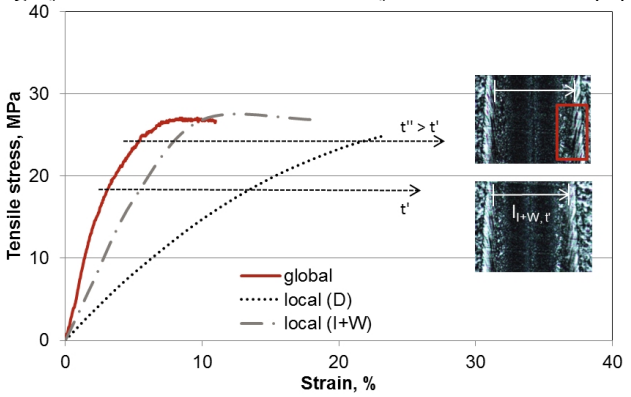


Figure 2: Micro-tensile curves for neat PP welded at 0.4 MPa measured and derived with local deformations.

B. *Modeling and simulation*

The simulation in Figure 3 shows the von Mises stress distribution inside and outside the weld with the colors representing the maximum stress reached for each layer. Although maximum stresses can be found at the edge between bulk material and weld bead and inside the bulk material, initial failure of welded nanocomposites takes place inside the welding areas first for all material-process-combinations tested except for neat PP welded at 0.4 MPa. Therefore, local tensile strength in deformed spherulites, inner area and weld line must be lower than in the bulk material. Different weld geometries achieved by changes in weld pressure lead to different stress distributions.

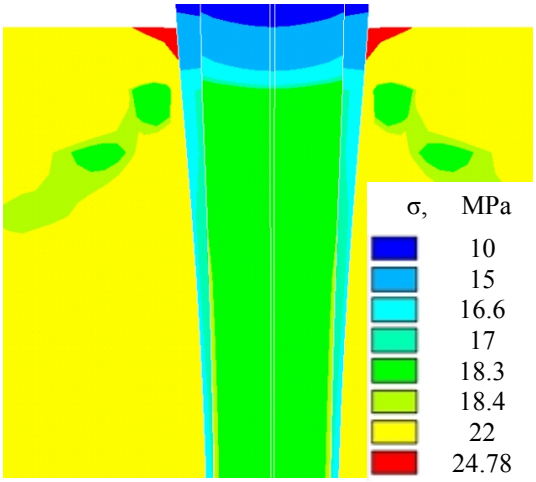


Figure 3: Von Mises stress-distribution for strain at ultimate stress of neat PP-T-V1 welded at 0.4 MPa in ANSYS.

IV. CONCLUSIONS

The investigation shows, that the weakest point of vibration welded nanocomposites is inside the weld area due to reduced tensile strength and geometry dependent stress maxima. Model and simulation can be used to find ways to optimize the weld geometry in order adjust the welding process to achieve higher weld strengths.

ACKNOWLEDGEMENT

The authors thank the German Research Foundation (DFG) for the financial support according to the DFG-project SCHL 280/19-1. The authors are grateful to Mr. V. Demchuk, Polymer Engineering Bayreuth, for the helpful cooperation. They also gratefully acknowledge Borealis Group and Huntsman for the kind donation of materials used.

REFERENCES

[1] Schlarb A.K., Ehrenstein G.W.; Polymer engineering and science, vol. 29, pp. 677-682, 1989

[2] Lin L.Y.; Dissertation, Technische Universität Kaiserslautern, 2013

[3] Bates J. et al.; Journal of Thermoplastic Composite Materials, vol. 20, pp. 5-16, 2007

[4] Schlarb A.K.; Masterbatch Asia, Singapore, 26th of March 2012

Detailed Modeling of RDC Liquid Extraction Column using the Sectional Quadrature Method of Moments (SQMOM)

Samer Alzyod[#1], Menwer Attarakih[*#2], Hans-Jörg Bart[#3]
Chair of Separation Science and Technology, TU Kaiserslautern, Postfach 3049 - 67653 Kaiserslautern, Germany
* Faculty of Engineering & Technology, Chemical Engineering Department., The University of Jordan, POBox 11942 Amman, *Jordan*
[1]samer.alzyod@mv.uni-kl.de
[2]attarakih@yahoo.com
[3]bart@mv.uni-kl.de

***Abstract*— The Sectional Quadrature Method Of Moments (SQMOM) is extended into the physical space to simulate the coupled hydrodynamics and mass transfer in a Rotating Disk Contactor (RDC) liquid-liquid extraction column. To facilitate the model implementation, a finite volume scheme with flux vector splitting technique was designed and coupled with the standard MATLAB ordinary differential equation solvers. The present numerical scheme is fifth order accurate in time and first order accurate in space. The quadrature nodes and weights are calculated analytically using the Two-Equal Weight Quadrature (TEqWQ) formula derived by Attarakih et al., (*Attarakih, M., Drumm, C., Bart, H-J., (2009), Solution of the population balance equation using the Sectional Quadrature Method Of Moments (SQMOM). Chem. Eng. Sci., 64, 742-752)*. As a numerical case study, the SQMOM is validated using PPBLab software which is based on the extended fixed pivot technique with multi-sectional grid w.r.t droplet diameter. In addition to this, the SQMOM prediction is validated against the published steady state experimental data for water-acetone-toluene chemical test system. The required number of sections to discretize the droplet diameter was found to be one, while the required number of spatial numerical cells was found to be 50 for accurate steady state pilot plant RDC column simulation. The SQMOM was found flexible to predict the column hydrodynamics as well as the mass transfer profiles.**

Keywords*— Population balances; SQMOM; TEqWQ; PPBLab; RDC extraction column

I. INTRODUCTION

Liquid-liquid extraction is considered as one of the most commonly used separation techniques after distillation in pharmaceutical and petroleum industries [1]. In such a technique, the droplet size distribution plays a decisive role in the determination of the physicochemical product properties of the dispersed phase. To date, the extraction equipment design and control procedures are still rely on the laboratory experiments and the lumped classical models. Indeed, these models predict neither the correct steady state nor the dynamic behavior of the dispersed phase. This is because they neglect the discreet nature of the dispersed phase by assuming a pseudo homogeneous phase. As a result of this, detailed mathematical modeling at the discrete level is an obvious need using the Detailed Population Balance Model (DPBM). Due to the growing computational power, the DPBM is accepted as an essential mathematical framework for modeling the steady state and dynamic behavior of multiphase flow problems. This behavior is characterized by the coupled interacting hydrodynamics and mass transfer phenomena such as: droplet breakage, coalescence, growth and interphase mass transfer [1]. The DPBM consists of nonlinear integral partial differential equations and hence it has no general known analytical solution except for a few simplified cases [2]. Therefore, there are many available numerical solution methods which include: stochastic Monte Carlo methods, moments based methods, and the recent maximum entropy methods [1]. Among these methods, the moments based methods were found to be efficient in terms of implementation and computational cost [1]. In this regard, Attarakih et al. [2] developed the Sectional Quadrature Method Of Moments (SQMOM) based on the primary and secondary particle concepts to solve the homogenous population balance equation. Unlike the classical moments based methods, the SQMOM is able to reconstruct the particle size distribution without any assumptions concerning the distribution shape supplied with two analytical solutions to calculate the quadrature nodes and weights (in case of two quadrature points). Attarakih et al., [2] showed that most of the moments based methods are special cases from the general SQMOM framework by varying the number of primary and secondary particles. For example, when only one primary particle is used, the Quadrature Method Of Moments (QMOM) is obtained, while if only one primary and one secondary particle are used, then the recent One Primary One Secondary Particle Method (OPOSPM) is obtained [3] which retains most of the dynamic features of the detailed continuous DPBM. Therefore, this method was successfully used as an alternative to the detailed DPBM for solving online inverse problems to find the optimal values of the coalescence model parameters [4]. In spite of its simplicity, OPOSPM found many CFD applications as a special case from the general SQMOM framework. Drumm et al. [5] implemented OPOSPM using

Fluent 6.3 to simulate the hydrodynamics behavior of a pilot plant RDC liquid extraction column. Attarakih et al. [6] and Alzyod [7] implemented OPOSPM as a reduced population balance model using MATLAB-SIMULINK environment to simulate the steady state and dynamic behavior of Kühni liquid -liquid extraction column. Recently, Kobert [8] solved the kinetic Bhatnagar Gross Krook (BGK) model of gas dynamics by coupling the SQMOM and FPM methods. In this work, the SQMOM is extended to solve the nonhomogeneous population balance model along the spatial domain to simulate the coupled hydrodynamics and mass transfer behavior of a pilot plant RDC DN80 liquid extraction column using a semi-implicit finite volume scheme. The required coalescence model parameters are obtained from Attarakih et al. [3], where they used OPOSPM as a reduced population balance model to simulate the same column geometry under the same operating conditions. Here, one section based on the Two-Equal Weight Quadrature (TEqWQ) was found enough to predict accurately the experimental data as compared to PPBLab extended fixed-pivot technique population balance equation solver. This is essentially a space-time multi-sectional population balance equation solver.

II. MATHEMARICAL MODEL

Following Attarakih et al. [9], the general population balance equation which couples the column hydrodynamics and mass transfer as formulated by the SQMOM can be written as follows:

$$\frac{\partial \mu_{r,m}^{\langle i\rangle}}{\partial t}+\frac{\partial}{\partial z}\left(\bar{F}_{r,m}^{\langle i\rangle}-D_y\frac{\partial \mu_{r,m}^{\langle i\rangle}}{\partial z}\right)=\hat{S}_{r,m}^{\langle i\rangle}\ \ r,m=0,..,2\times \mathrm{N}_q-1 \qquad (1)$$

where, the superscript i represents the i^{th} section number, N_q is the number of secondary particles, and D_y is the dispersed phase axial dispersion coefficient. The conserved quantity $\mu_{r,m}^{\langle i\rangle}$ is a two dimensional moment transformation function, while $\bar{F}_{r,m}^{\langle i\rangle}$ is the transformed numerical flux vector. These transformed terms are given by:

$$\mu_{r,m}^{\langle i\rangle}=\sum_{j=1}^{Nq}\sum_{n=1}^{Nq}d_j^{\,r}c_{y,n}^{\,m}w_{j,n} \qquad (2)$$

$$\bar{F}_{r,m}^{\langle i\rangle}=\sum_{j=1}^{Nq}\sum_{n=1}^{Nq}u_y\left(d_{j,n},\phi_y,P\right)d^{\,r}c_y^{\,m}w_{j,n} \qquad (3)$$

Here, d and w are the quadrature nodes and weights respectively, c_y is the dispersed phase solute concentration, and u_y is the dispersed phase mean droplet velocity. The general source term $\left(\hat{S}_{r,m}^{\langle i\rangle}\right)$ is a nonlinear term and it is given by:

$$\hat{S}_{r,m}^{\langle i\rangle}=u_y^{in}\mu_{in}^{\langle i\rangle}c_{yin}^{m}\delta\left(z-z_y\right)+S_{B,r,m}^{\langle i\rangle}+S_{C,r,m}^{\langle i\rangle}+S_{M,r,m}^{\langle i\rangle} \qquad (4)$$

In this equation, u_y^{in} is the dispersed phase inlet superficial velocity, $\mu_{in}^{\langle i\rangle}$ is inlet feed moments and c_{yin} is the dispersed phase inlet solute concentration. The first term on the right hand side represents the inlet feed distribution moments modeled as a point source term at $z = z_y$, while the second and third terms are the transformed droplets breakage and coalescence functions respectively. These transformed functions are written as follows [2]:

$$S_{B,r,m}^{\langle i\rangle}=-Dr^{\langle i\rangle}\left[\Gamma^{\langle i\rangle}\bullet w^{\langle i\rangle}\right]^T+\sum_{m=i}^{N_{pp}}C_r^{\langle i,m\rangle}\left[\Gamma^{\langle i\rangle}\bullet w^{\langle i\rangle}\right]^T \qquad (5)$$

$$S_{C,r,m}^{\langle i\rangle}=\sum_{k=1}^{i\times N_q}\left\{\sum_{j=k}^{i\times N_q}\psi_{k,j,r}^{\langle i\rangle}\omega_{j,k}w_j'w_k'-\eta_k\times\sum_{n=1}^{N_q\times N_{pp}}\left(d_k'^{\,r}\right)\omega_{k,n}w_k'w_n'\right\} \qquad (6)$$

where Γ and ω are the breakage and coalescence frequencies respectively. Theses functions will be discussed in detail in section 3. The last term in (4) takes into account the mass transfer rate from the continuous phase to the dispersed phase and vice versa. This transformed term is written as follows [7]:

$$S_{M,r,m}^{\langle i\rangle}=m\sum_{n=1}^{N_{sp}}c_y^{m-1}\dot{c}_y w_n \qquad (7)$$

The solute concentration in the continuous phase is given by the following solute concentration transport equation [3]:

$$\frac{\partial\left(\alpha_c c_c\right)}{\partial t}-\frac{\partial}{\partial z}\left(u_c\alpha_c c_c+D_c\frac{\partial\left(\alpha_c c_c\right)}{\partial z}\right)=u_c^{in}c_c^{in}\delta\left(z-z_c\right)-\int_0^{\infty}\int_0^{c_y^{\max}}\dot{c}_y\nu(d)n_{v,c_y}(d)\partial d\,\partial c_y \qquad (8)$$

In this equation, α_c is the continuous phase volume fraction, C_c is the solute concentration in the continuous phase, u_c is the continuous phase velocity and D_c is the continuous phase axial dispersion coefficient. In this work, the required quadrature nodes and weights are calculated analytically using the Two-Equal Weight Quadrature (TEqWQ) formula [2]:

$$d_{1,2}^{\langle i\rangle}=\hat{\mu}_1^{\langle i\rangle}\mp\frac{1}{\sqrt{3}}\sqrt{\frac{\hat{\mu}_3^{\langle i\rangle}}{\hat{\mu}_1^{\langle i\rangle}}-\left(\hat{\mu}_1^{\langle i\rangle}\right)^2} \qquad (9)$$

$$w_{1,2}^{\langle i\rangle}=\frac{1}{2}\hat{\mu}_0^{\langle i\rangle} \qquad (10)$$

Within this framework, the dispersed phase holdup $\left(\alpha_y\right)$ and the mean solute concentration in the dispersed phase $\left(\bar{c}_y\right)$ are given by the following equations:

$$\alpha_y=\frac{\pi}{6}.\sum_{i=1}^{N_{pp}}\mu_{3,0}^{\langle i\rangle} \qquad (11)$$

$$\bar{c}_y = \sum_{i=1}^{N_{pp}} \mu_{3,1}^{\langle i \rangle}(t,z) \Big/ \sum_{i=1}^{N_{pp}} \mu_{3,0}^{\langle i \rangle} \tag{12}$$

Unfortunately, the system of equations given by (1) can not be directly solved due to the coupled column hydrodynamics and mass transfer. To cure this problem, Attarakih et al. [9] proposed a solution methodology in which the pure column hydrodynamics transport equations are initially solved then the mass transfer transport equations are solved with subsequent update procedure if required. Indeed, this methodology is valid for liquid extraction equipment since the breakage and coalescence terms are weak functions of the solute concentration and the usual operating solute concentration is less than 10 percent [3].

III. Column Hydrodynamics

In this section, the required models and correlations for describing the column hydrodynamics are presented. These models and correlations include: single droplet terminal velocity, breakage and coalescence phenomena.

A. Droplet Terminal Velocity

The droplet terminal velocity (u_t) is considered as one of the important functions for liquid-liquid extraction equipment design and scale-up purposes [10]. Indeed, it helps to get the required information about the actual droplets swarm velocity inside the extraction equipment. In general, the droplet terminal velocity depends on the dispersed phase physical properties and the droplet diameter. In this work, the droplet velocity model of Vigness [11], which is recommended for high surface tension chemical systems, is used. Based on the droplet terminal velocity (u_t), the droplet rising velocity (u_y) along the column height as a function of droplet diameter (d) and dispersed phase holdup (α_y) is given by [3]:

$$u_y(\alpha_y, d) = k_s(d) \cdot (1-\alpha_y)^m \cdot u_t(\alpha_y, d) - \frac{u_c^{in}}{(1-\alpha_y)} \tag{13}$$

In this equation, k_s is the slowing factor due to column internals, u_c^{in} is the continuous phase inlet superficial velocity, and m is a swarm exponent. Here, the terminal velocity is modified to take into account the droplets swarm effect and the column internals. In this work, the slowing factor was taken as an average constant value (Kv = 0.51) based on the experimental data of Garthe [12].

B. Breakage and Coalescence Phenomena

The source term given by (4) is a nonlinear integral term which takes the breakage and coalescence events into account. The droplet breakage takes place in agitated columns due to the interaction occurring between the droplets swarm and the column internals. Garthe [12] reported that the breakage process occurs above a certain critical rotational speed ($N_{critical}$), which is a function of physical properties and the critical droplet diameter (d_{crit}). Based on his experimental work, Garthe developed the following breakage probability correlation for agitated liquid-liquid extraction columns [12]:

$$\frac{p_B}{1-p_B} = c_1 \cdot \left(\frac{W_{mod}}{1 + c_2 \cdot \eta_d \cdot \left[W_{mod} / (\gamma \cdot d_m \cdot \rho_d) \right]^{0.5}} \right)^{C_3} \tag{14}$$

where, W_{mod} is the modified weber number which depends on the critical rotational speed. In this work, the critical rotational speed is given by [12]:

$$n_{R,Crit} = c_4 \cdot \frac{d_A^{-2/3} \cdot \eta_d \cdot d_m^{-4/3}}{2 \cdot (\rho_c \cdot \rho_d)^{0.5}} + \left[\left(c_4 \cdot \frac{d_A^{-2/3} \cdot \eta_d \cdot d_m^{-4/3}}{2 \cdot (\rho_c \cdot \rho_d)^{0.5}} \right)^2 + c_5 \cdot \frac{\sigma}{\rho_c \cdot d_A^{4/3} \cdot d_m^{5/3}} \right]^{0.5} \tag{15}$$

Here, the fitting parameters (c_i) depend on the column type. These parameters are estimated by matching the breakage model prediction using the available breakage probability experimental data for each single droplet. In this work, the breakage probability model parameters are given by Garthe [12]. These parameters are shown in Table I. Unlike the breakage process, the coalescence mechanism is not well understood yet. In this regard, many theories were proposed to describe the coalescence mechanism. In this regard, Coulaloglou and Tavlarides [13] developed the following model based on the film drainage theory:

$$\omega(d, d', \phi_d) = \left[c_1 \frac{\varepsilon^{1/3}}{1+\phi} (d + d')^2 (d^{2/3} + d'^{2/3})^{1/2} \right] \times \left[\exp\left(- \frac{c_2 \eta_c \rho_c \varepsilon}{\sigma^2 (1+\phi)^3} \right) \left(\frac{d d'}{(d + d')} \right)^4 \right] \tag{16}$$

where, c1 and c2 are fitting parameters, and ε is the energy dissipation. The fitting parameters in Eq.(16) depend on the chemical system physical properties, rotational speed, and the total throughput. In this work, the coalescence model parameters are obtained from Attarakih et al. [3], where they solved an inverse optimization problem based on OPOPSPM as a reduced population balance model to simulate the same column geometry.

TABLE I. Breakage Model Fitting Parameters

C_1	C_2	C_3	C_4	C_5
1.29×10^{-6}	0.33	2.78	0.020	0.13

These parameters are given by c_1 = 0.03 and $c_2 = 1.83\times10^3$ (m^{-2}) respectively.

IV. Mass Transfer

A successful design and scale-up of liquid extraction equipment require a good understanding of the governed mass transfer phenomena. Indeed, the mass transfer mechanism is not well understood yet due to the experimental difficulties and the accompanied mathematical complexity [14]. Basically, the

mass transfer rate inside the extraction equipment depends on the overall mass transfer coefficient (k_{oy}) and is given by [3]:

$$\frac{\partial c_y}{\partial t} = \frac{6k_{oy}}{d}\left(mc_c - c_y\right) \tag{17}$$

where, k_{oy} is the overall mass transfer coefficient and m is the solute distribution coefficient. In this work, the solute distribution coefficient is correlated as a function of the solute mass fraction in the continuous phase (x) based on the experimental work of Garthe [12] and it is given by:

$$m = \exp\left(a \cdot x + b\right) \tag{18}$$

The fitting parameters a and b are estimated using PPBLab [15] activity coefficient optimization tool. These parameters are shown in Table II.

TABLE II. DISTRIBUTION COEFFICIENT CORRELATION PARAMETERS

Chemical system	*a*	*b*
Water-actone-toluene	3.66	-0.41040

The overall mass transfer coefficient is given by the two film theory of Whitman [16], where, the overall mass transfer coefficient is the reciprocal of the overall resistances and it is given as follows [16]:

$$k_{oy} = \left(\frac{m}{k_c} + \frac{1}{k_y}\right)^{-1} \tag{19}$$

where k_c and k_y are the individual mass transfer coefficients in the heavy and dispersed phases respectively. In this work, Kronig and Brink model [17], which is recommended for circulating droplets, is used to predict the individual mass transfer coefficient for both phases.

V. SQMOM SPATIAL DISCRETIZATION

In this section, the SQMOM is implemented and solved in the physical space using MATLAB software. Equation (1) represents a highly nonlinear system of equations and hence it has no general analytical solution. This system is dominant by convection [3] which complicates the numerical treatment. To overcome this problem, the solution procedure proposed by Attarakih et al., [9] is implemented here. First, the column pure hydrodynamics is solved and then the mass transfer equations are solved. Here, a semi-implicit finite volume scheme with flux vector splitting technique is implemented and coupled with the standard MATLAB ordinary differential equation solvers. The present numerical scheme is fifth order accurate in time and first order accurate in space. The column height (z) is discretized into n spatial numerical cells $I_j = \left[z_{j-1/2}, z_{j+1/2}\right]$ of size $\Delta z = z_{j+1/2} - z_{j-1/2}$, with $j = 1,2,3...n$. The control volume is given by $V = I_i \times \left[t^n, t^{n+1}\right]$ as shown in Fig. 1. Using the cell sliding average for the conserved quantity $\mu_{r,m}^{\langle k \rangle}$, equation (1) could be written in the following semi- discrete form:

$$\frac{d\,\mu_{r,j}^{\langle i \rangle}}{dt} + \frac{1}{\Delta z}\left[\bar{F}_{r,m}^{\langle i \rangle}\big|_{j+1/2} - \bar{F}_{r,m}^{\langle i \rangle}\big|_{j-1/2}\right]_j = \bar{S}\left(U\right)_j \;, j = 1,...,n \tag{20}$$

Where n, is the number of numerical cells used to discretize the physical spatial domain, and $\bar{F}_{r,m}^{\langle i \rangle}|_{j+1/2}$ and $\bar{F}_{r,m}^{\langle i \rangle}|_{j-1/2}$ are the right and left fluxes at the cell interface respectively. Finally, using Patankar and Spalding [21] compact notation, the numerical flux could be written as follows:

$$\bar{F}_{r,m}^{\langle i \rangle}|_{j+1/2} = \left\langle \bar{F}_{r,m}^{\langle i \rangle}|_j, 0 \right\rangle - \left\langle -\bar{F}_{r,m}^{\langle i \rangle}|_{j+1}, 0 \right\rangle \tag{21}$$

$$\bar{F}_{r,m}^{\langle i \rangle}|_{j-1/2} = \left\langle \bar{F}_{r,m}^{\langle i \rangle}|_{j-1}, 0 \right\rangle - \left\langle -\bar{F}_{r,m}^{\langle i \rangle}|_j, 0 \right\rangle \tag{22}$$

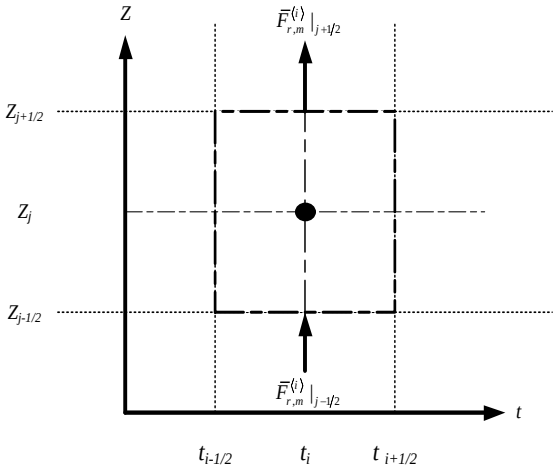


Fig. 1. Two-dimensional representation of the space-time grid.

VI. NUMERICAL ABD EXPERIMENTAL INVESTIGATIONS

In this section the SQMOM prediction is validated using the published steady state experimental data of Garthe [12]. In this case study, the simulation was conducted using a pilot plant RDC DN80 liquid-liquid extraction column where its detailed dimensions are shown in Table III.

TABLE III. RDC COLUMN GEOMETRY (ALL DIMENSIONS ARE IN M)

Column height	*Column diameter*	*Rotor diameter*	*Compartment height*
4.40	0.08	0.045	0.05

The chemical system used here is water-acetone-toluene chemical test system. The dispersed phase (toluene) flow rate is 48 liter/h, while the continuous phase flow rate is 40 liter/h. The rotational speed is fixed at 200 rpm. The initial conditions are zero for the droplets number and dispersed phase holdup

respectively, while the initial concentration is the same as the solute concentration in the continuous phase inlet feed. The inlet feed is taken as a normal distribution with mean and variance of 2.0 and 0.55 respectively for both chemical systems. The minimum and maximum droplet diameters are 0.025 and 8 mm respectively. In this work, the breakage probability is given by Garthe [12], while the coalescence frequency is given by Coulaloglou and Tavlarides [13]. These frequency functions are described in detail in section III. Fig. 2 shows a comparison between the simulated dispersed phase holdup profiles using the SQMOM and PPBLab [15] software and the published experimental data of Garthe [12]. It can be clearly observed that the SQMOM and PPBLab show a very good agreement with the experimental data using the same operating and numerical parameters.

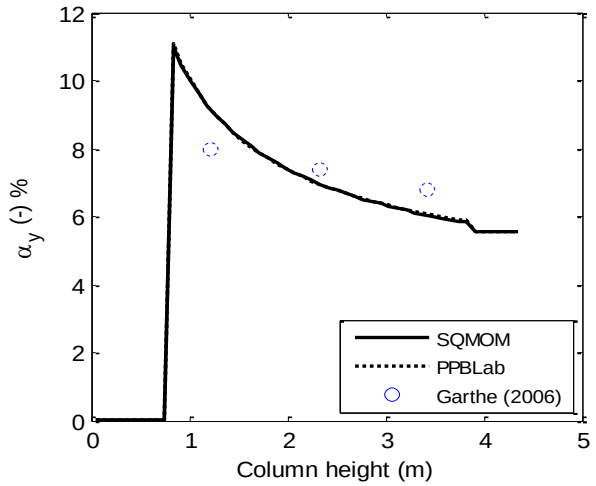


Fig. 2. Comparison between the simulated holdup profiles using the SQMOM and PPBLab [18], and the experimental data [14].

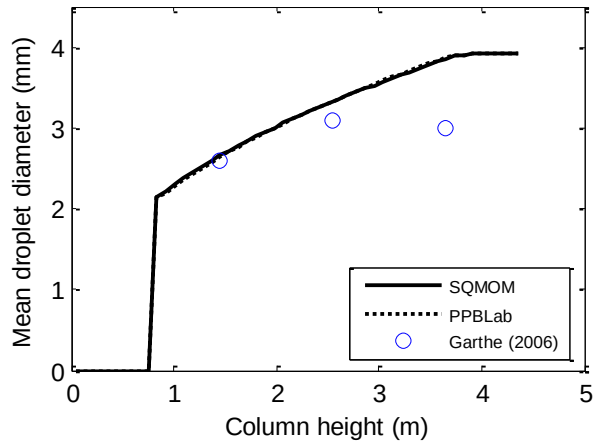


Fig. 3. Comparison between the simulated mean droplet diameter profiles using the SQMOM and PPBLab [18], and the experimental data [14].

Fig. 3 shows the simulated mean droplet diameter profiles along the column height using PPBLab and the SQMOM as compared to the experimental data. Unlike the holdup measurement, the mean droplet diameter measurement is less accurate and this explains the slight deviation between the simulation results and the experimental data [12]. As descried in section V, the column hydrodynamic equations are initially solved then the mass transfer profiles along the column height are estimated. Fig. 4 depicts a comparison between the simulated mass transfer profiles along the column height using the SQMOM and the experimental data, where a very good agreement was obtained using the recommended mass transfer correlations of Kronig and Brink [17] for circulating droplets.

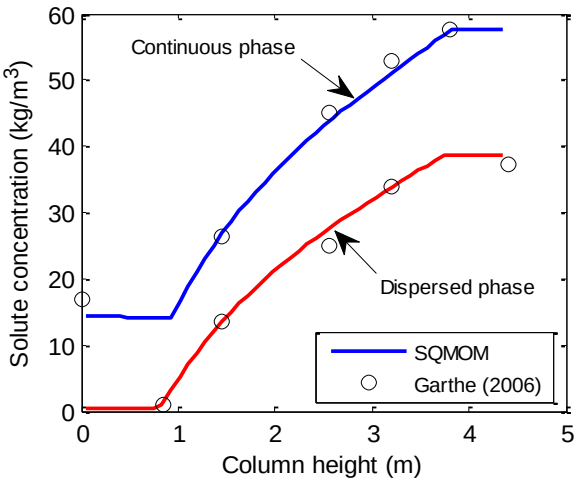


Fig. 4. Comparison between simulated the mass transfer profiles using the SQMOM and the experimental data [14].

VII. Conclusions

In this work, we extended the SQMOM to model the hydrodynamics and mass transfer behaviour of an RDC liquid liquid extraction column using a semi-impicit finite volume scheme. At the numerical level, the SQMOM showed a very good agreement with PPBLab software using the same operating conditions and the same coalescence model parameters. At the experimental level , the SQMOM was found flexible to predict the experimental data using water-acetone-toluene chemical test system. As a final conclusion, the SQMOM is considered as a promising framework for modelling the multiphase flow problems in general and liquid-liquid extraction equipment in particular.

Acknowledgment

The authors would like to acknowledge the financial support from TU Kaiserslautern and the Process Engineering and Technology Network of Competence (Pro3).

References

[1] M. Attarakih, H-J., Bart, “Solution of the Population Balance Equation using the Differential Maximum Entropy Method (DMaxEntM): An application to liquid extraction columns,” Chemical Engineering Science, vol. 108, pp. 123-133, 2014.

[2] M. Attarakih, C. Drumm, H-J. Bart, “Solution of the population balance equation using the Sectional Quadrature Method of Moments (SQMOM),” Chemical Engineering Science, vol. 64, pp. 42-752, 2009.

[3] M. Attarakih, M. Abu-Khader, H-J Bart, “Modelling and dynamic analysis of an RDC extraction column using OPOSPM”, Chemical Engineering Science, vol. 91, pp. 180-196, 2013.

[4] M. Attarakih, H.B. Jildeh,M. Mickler, H-J. Bart, “The OPOSPM as a Nonlinear Autocorrelation Population Balance Model for Dynamic Simulation of Liquid Extraction Columns,” Computer Aided Chemical Engineering, vol. 31, pp. 1216-1220, 2012.

[5] C. Drumm, M. Attarakih, M. Hlawitschka, H.-J Bart, “One group reduced population balance model for CFD simulation of a pilot-plant extraction column,” Industrial and Engineering Chemistry Research, vol. 49(7), pp. 3442-3451, 2010.

[6] M. Attarakih, S. Alzyod, M. Hlawitschke, H.-J Bart, “OPOSSIM: A population balance-SIMULINK module for modelling coupled hydrodynamics and mass transfer in liquid extraction equipment,” Computer Aided Chemical Engineering, vol. 37, pp. 257-262, 2015.

[7] S. Alzyod, “One primary one secondary particle method SIMULINK module for the steady state and dynamic modelling of Kühni extraction column,” master thesis, The University of Jordan, Amman, Jordan, 2013.

[8] M. Kobert, “Application of the finite pointset method to moving boundary problems for the BGK model of rarefied gas dynamics.,” Dissertation, TU Kaiserslautern, Germany, 2015.

[9] M. Attarakih, H-J. Bart, N. Faqir, “Numerical solution of the bivariate population balance equation for the interacting hydrodynamics and mass transfer in liquid–liquid extraction columns,” Chemical Engineering Science, vol. 61, pp. 113-123, 2006.

[10] C. Gourdon, G. Casamatta, G. Muratet, “Population Balance Based Modelling In Liquid-liquid extraction equipment,” Godfrey and Slater, Eds.; John Wiley and Sons: New York, 1994, pp.140-226.

[11] A. Vigness,“Hydrodynamique Des Dispersions - Mouvement D´un Globule Dans un Fluide Immobile et infini, ” Genie Chim, vol. 93, pp. 129-42, 1965.

[12] D. Garthe, “Fluid Dynamics and Mass Transfer of Single Particles and Swarms of Particles in Extraction Column,” Dissertation, Technischen Universität München, Germany, April 2006.

[13] C. Coulaloglou, L. Tavlarides, “Description of interaction processes in agitated liquid-liquid dispersions,” Chemical Engineering Science, vol. 32(11), pp. 1289-97, 1977.

[14] Slater, M. Rate coefficients in liquid-liquid extraction systems. In Liquid-liquid extraction equipment; Godfrey and Slater, Eds.; John Wiley and Sons: New York, 1994, pp.47-94.

[15] M. Attarakih, S. Al-Zyod, M. Abu-Khader, H.-J. Bart, “PPBLAB: A New Multivariate Population Balance Environment for Particulate System Modelling and Simulation,” Procedia Eng, vol. 42, pp. 1445-62, 2012.

[16] W.G. Whitman, “The film theory of absorption,” Chem. Met. Eng. vol. 29, pp. 147–151, 1923.

[17] R. Kronig, R., J.C, Brink, “On the theory of extraction from falling droplets,” Applied Scientific Research, vol. 2, pp. 142-154, 1951.

[18] M. Jaradat, M., Attarakih, H.-J. Bart, “ Effect of phase dispersion and mass transfer direction on steady state rdc performance using population balance modelling”, Chem. Eng. J.,vol. 165, pp. 379–387, 2010.

Automated Learning of Self-Similarity of Informative Structures in Architecture

Michael Arnold
Fraunhofer ITWM Kaiserslautern
Heidelberg Collaboratory for Image Processing
Email: michael.arnold@itwm.fraunhofer.de

Björn Ommer
Heidelberg Collaboratory for Image Processing
Email: ommer@uni-heidelberg.de

***Abstract*—Databases in cultural heritage are still not sufficiently annotated in respect to form and context in art and architecture. That makes automatic methods necessary, which avoid costly and tedious manual annotation. However computers still lack the ability to understand scenes and objects in images on a semantic level. Automated learning of meaningful objects in images without human supervision is challenging since there is no prior knowledge about the morphology and localization of important parts. In our method we used a Selective Search algorithm [10] to generate a set of object hypotheses. We encoded image regions with Fisher Vectors, enhanced by a spatial pyramid scheme. We were able to identify reoccurring facade elements with Exemplar SVM classifiers for all object hypotheses, by utilizing their self-similarity property. A final graphcut based reconstruction, explained the original facade, by the given automatically generated hypotheses. We evaluate the performance of our method on a challenging facade image of early modern architecture. In particular, we show that unsupervised learning methods can be used for facade segmentation.**

I. Introduction

Automated facade recognition is of interest for a broad range of applications like urban modeling, facade synthesis, facade detection, 3D scene parsing, facade decomposition and not least for applications in computational humanities. Facades as man-made structures also feature a particular set of properties, amongst which the occurence of strong repetitive elements, is the most prominent one. It is easy to see, that architecture has an underlying grammar, which arranges the facade elements in a functional and visually pleasing way. In the last years, the body of research adressing facade understanding has increased significantly.

In 2011 Nan et. al. [6], presented a work, that highlights underlying Gestalt principles in architecture. They apply these principles on architectural drawings in order to find good abstractions of facades. Their work is unique in adressing these architectural concepts, while remaining only applicable to synthetic data.

Nearly all methods in the field of facade understanding remain in the domain of supervised learning approaches. In 2013 Dai et. al. [3] presented a method for synthesizing complete, photo-realistic facade images, from a single example. In the field of facade detection Liu proposed in 2014 a novel regularity-driven framework for facade detection from aerial images of urban scenes [5], by leveraging the vertical and horizontal alignment of edge features in buildings. Other approaches like the work from Riemenschneider et. al. from 2012 [8] concentrate on the grammar of facades, whereas the method of Cohen et. al. from 2014 [2] concerns itself with efficient facade labeling.

There is only little work in the domain of facade understanding with a completely unsupervised setting. Wendel et. al. introduced a novel approach in 2010 [11] for finding repetitive patterns with the main focus on facade separation and segmentation, by applying a procedural modeling approach on real-world data.

In contrast to the work from Wendel et. al. [11], we do not take facade separation into account. Instead, we focus on finding and reconstructing facades by their constituent structural elements. This novel approach allows for applications in computational humanities and is aimed at closing the gap to a large body of challenging unlabeled real-world data. Typical drawings extracted from architectural literature provided by the University Library of Heidelberg and photos of early modern facades are illustrated in Figure 1.

In section 2 we explain our method, from constructing a strong representation for image regions, to finding good facade element hypotheses, training of hypothesis classifiers, followed by the description of our final graphcut based facade reconstruction. Section 3, evaluates and discusses our method based on the results, on the challenging early modern facade shown in Figure 2. Finally, we conclude this work in section 4.

II. Methods

In this work we aim to automatically identifying meaningful structures in facades, in a completely unsupervised setting. In this context we want to utilize the reoccurence behaviour of important facade elements. The Selective Search algorithm [10] is used to generate object hypotheses. Moreover self-similarites help identify reoccuring structures, based on strong HoG based representations feeded into an Exemplar Support Vector Machine. Graphcut based reconstruction, of facade elements with the hypothesis classifier response maps, finally gives us the automated semantic segmentation of the facade elements.

A. Representation

We are looking for reoccuring facade elements. To this end we have to describe image regions, in a way that is robust

Fig. 1. Diverse collection of drawings and photos, supported by the University Library of Heidelberg.

Fig. 2. Exemplary test image, showing a variety of facade elements. A range of different windows, pillars and roof elements with many repetitions exist.

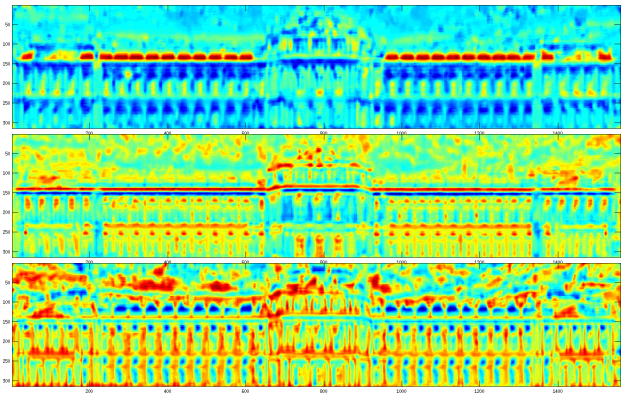


Fig. 3. Visualization of 3 out of 500 distinct feature dimensions in false colors, after final dimensionality reduction of the spatial pyramid enhanced Fisher Descriptor based encoding. The underlying region size for the shown representation is 9x22 pixels. Representations like these are computed for all region sizes occuring in the set of automatically generated set of hypothesis. Note that single feature dimensions, can be weakly associated with certain facade elements.

against typical challenges like lighting conditions, occlusions etc. In essence, we have to construct a compact representation with desired invariance properties.

We use the common Histogram of Oriented Gradient (HoG) [4] descriptor, which counts the occurences of gradient orientations in localized image regions. To further improve the quality of our image region descriptor, we tailor our representation to the underlying feature distribution by encoding them with Fisher Vectors [7]. To this end, we are learning the distribution of HoG features within a given image by fitting a Gaussian Mixture Model to them. Based on the learned Gaussian Mixture Kernels, we can encode image regions into high dimensional Fisher Vectors.

In addition to the descriptive power of Fisher Vectors, we are also interested in a good localization property of our descriptors. After all, we are interested in finding the position of specific facade elements. Correspondingly, we make use of a spatial pyramid, that describes a given image region not by a single, but several descriptors of smaller size. These descriptors are of partially different scales in a pyramid like assembly, describing subparts of the original image region. The descriptors of the individual pyramid cells are finally concatenated, forming an even higher dimensional descriptor. High dimensionality of features, whatsoever can lead to a problem called Curse of Dimensionality in the later learning stage. Principal Component Analysis is therefore used to reduce the dimensionality, to get to the final representation for our image region descriptors.

B. Selective Search

Searching the space of all possible image subregions to find meaningful structures is unfeasible. We therefore use the Selective Search algorithm from Uijlings [10], in order to determine a manageable number of object hypotheses. The algorithm combines the strength of exhaustive search and segmentation algorithms. Combining a range of segmentation criteria, like texture and color spaces with different invariance properties helps to improve the quality of the generated hypotheses.

Coupled with additional filters for size and aspect ratio, we also enforce the symmetry property of the generated hypotheses, to further reduce their number and to improve their quality. The remaining number of hypotheses are then used as elementary facade elements for the final facade reconstruction.

C. Exemplar-SVM Matching

Since we want to find all occurences for all generated hypotheses, we have to train a classifiers. In the light of the given unsupervised setting, there are no labels for object categories available. This is why we train for each hypothesis a model with an Exemplar Support Vector Machine with the given hypothesis as single positive sample (the exemplar). In order to find a good descision boundary, we need a carefully

Fig. 4. Automatically generated selective search hypotheses remaining after aspect ratio, size and symmetry based filtering.

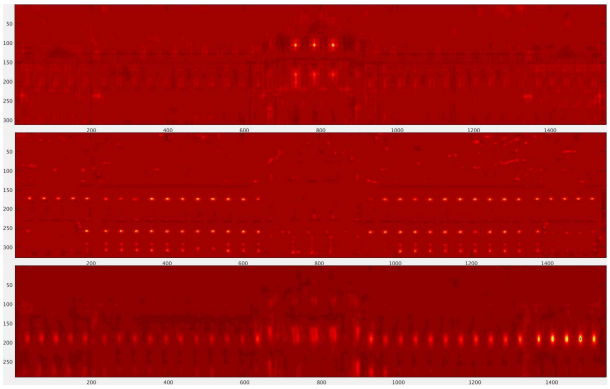


Fig. 5. Gibbs normalized Exempar SVM responses for three different window elements.

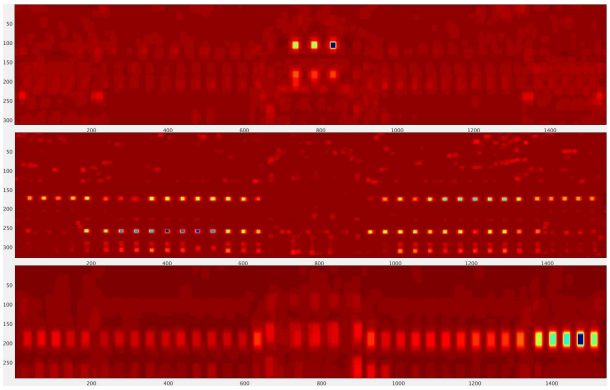


Fig. 6. Gibbs normalized Exempar SVM responses for three different window elements, filtered by a box shaped convolution filter with half the spatial dimensions of the corresponding hypothesis returning the maximal response value. It can be seen, that the strongly localized response values from Figure 5 have now the shape and size of their corresponding hypothesis. These responses are suited as costmaps for the graphcut based facade reconstruction.

Fig. 7. Facade reconstruction based on automatically generated hypotheses **without** hierarchical clustering based redundance filtering. The background label class was not visualized, to highlight the foreground background separation.

tailored set of negative samples. To this end, we sample around 500 negatives per object hypothesis. We keep the number of negatives rather small, in order to avoid image regions, that by chance are too similar to the exemplar.
Altogether, we do get a response map after Exemplar-SVM training for each hypothesis. Finally, a Gibbs normalization step enables the comparison of responses from different hypotheses with a probabilistic measure, in addition to improving the signal to noise ratio. These normalized response maps are the basis for the final facade reconstruction via graphcut segmentation (see Figure 5).

D. Graphcut Reconstruction

By explaining a given facade, based on the automatically generated hypotheses, we can learn something about the internal structure of the architecture. We aim for a consistent separation of facade elements from the background, as well as a robust separation of facade elements on the semantic level. To this end, we use a graphcut algorithm for a pixel-based reconstruction of the given facade. The graphcut algorithm can be understood, as an optimization problem, which aims to assign each pixel a label with minimal costs. In essence there is a cost for each pixel to assume a specific label, based on the responses of the different classifiers at that location. Furthermore, there is a smoothness term, that raises the cost for neighbouring pixels to have different labels. This helps with rooting out outliers, as well as improving the consistency of the reconstruction.

Since our response map signals at a certain pixel describe not only that pixel, but a region of the size of the corresponding hypothesis, we have to adjust our response maps, accordingly. We apply a box shaped convolution filter with half the spatial dimensions of the corresponding hypothesis and return the maximal response value (see Figure 6). This ultimately results in a filtered response map with hypothesis-region shaped responses. Note, that the spatial pyramid based representation, supports this property by improved response localization.
To enable robust background segmentation, we also add a separate background class. By assuming a uniform background label cost higher than the cost for good fits and a label cost cheaper than picking bad fits, we achieve a proper foreground background separation.
In the instance of multiple redundant hypotheses describing the same object label, confusion can occur. Given this point, we furthermore apply a hierarchical clustering based on the similarity of the overall response maps. Labels describing the same objects in image space will consequently be combined into a single label.

III. Experiments and Discussion

We show the performance of the used methods exemplarily on a challenging facade image (see Figure 2). We will also highlight the strength and weaknesses, as well as open challenges of our method. We will do so by evaluating the different stages of the algorithm separately.
Computing the described representation for an image region

Fig. 8. Facade reconstruction based on automatically generated hypotheses **with** hierarchical clustering based redundance filtering. The background label class was not visualized, to highlight the foreground background separation.

of a specific size at every pixel location is computationally expensive. We therefore use the Fast Local Area Invariant Representation (FLAIR-)scheme, described in [9], utilizing dynamic programming principles for an efficient computation. Still, the computation of the representation remains a bottleneck in our pipeline, because it has to be computed for all occuring hypotheses region sizes. In the past, we also considered representations, that are cheaper to compute, but lack in performance. Extensive testing has shown, that high quality responses with good localization are crucial. This is especially important, because due to the unsupervised setting, we have no labels to learn high quality object level subspaces. Our observations concerning the Selective Search algorithm [10], delivers on the whole good candidates for facade element hypotheses. This seems also to be supported by the fact, that man-made structures favor horizontally and vertically dominant elements in image space, that can easily be enclosed by rectangular image regions.

By filtering out outliers with extreme aspect ratios and region sizes, we were able to further inprove the overall quality of the hypotheses. Furthermore, enforcing the symmetry of the hypotheses, by thresholding the similarity to their mirrored counterpart, worked very well. A set of automatically generated object hypotheses can be seen in Figure 4. In any case, it should be noted, that a good set of hypotheses is important for a meaningful reconstruction during the later graphcut segmentation. In the case of hypotheses that overlap in image space, difficulties in foreground background separation can arise. In addition, multiple hypotheses of virtually the same facade element, can lead to label confusion during the graphcut segmentation, which is why we use the hierarchical clustering step described in the previous section. This highlights the importance of a good set of hypotheses. Therefore additional effort in improving the selection of good hypotheses helps to apply our method in a wider context.

For the Exemplar SVM, we apply the commonly used libsvm [1] with linear kernel. The results can be seen in Figure 5. In order to achieve a good performance in this step we carefully sampled negatives from the same image. One usually wants to sample negatives from other images, but due to the already high computational effort for the representation, we randomly sample from the given one. We limit the number of negatives to around 500. This keeps the chance of sampling false negatives very low. Extensive testing has also shown, that the estimation of a uniform Gibbs normalization parameter, enables good comparability between response maps of different hypothesis.

We have applied the graphcut reconstruction of the original facade on the max-response region filtered classifier response maps (see Figure 6). Since also redundant hypotheses exist in our facade element set, we test the performance of the reconstruction with, as well as without hierarchical clustering of the classifier response maps. Both variants can be seen in Figure 7 and 8 respectively. The number of resulting clusters were given as input to the clustering algorithm. One can clearly see, that the label refinement with hierarchical clustering improves the performance of the reconstruction. The drawback of providing the number of resulting clusters remains and is left for further optimization in the future.

IV. Conclusion

We have shown that unsupervised learning methods can be used for facade reconstruction. We provide a novel approach to facade segmentation that closes the gap to applications in computational humanities. In the future we will apply this method on large databases of images in order to support art historians with a tool for architectural analysis. We aim for the extention of hierarchical decomposition and the connecting of similar facade elements across images. In the long run, this will enable art historians to analyse architecture in a way, that otherwise would only be possible with supervised learning methods.

Acknowledgements

I thank the University Library of Heidelberg, for supporting a suitable facade image database.

References

[1] C.-c. Chang and C.-j. Lin. LIBSVM : A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2:1–39, 2011.

[2] A. Cohen, A. G. Schwing, and M. Pollefeys. Efficient Structured Parsing of Facades Using Dynamic Programming. pages 3206–3213, 2014.

[3] D. Dai, H. Riemenschneider, G. Schmitt, and L. Van. Example-Based Facade Texture Synthesis. *Computer Vision (ICCV), 2013 IEEE International Conference on*, (1):1065–1072, 2013.

[4] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, pages 886–893, 2005.

[5] J. Liu and Y. Liu. Local Regularity-driven City-scale Facade Detection from Aerial Images. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (1):3778–3785, 2014.

[6] L. Nan, A. Sharf, K. Xie, T.-T. Wong, O. Deussen, D. Cohen-Or, and B. Chen. Conjoining Gestalt rules for abstraction of architectural drawings. *ACM Transactions on Graphics*, 30:1, 2011.

[7] F. Perronnin, J. Sánchez, and T. Mensink. Improving the Fisher Kernel for Large-Scale Image Classification. *Computer Vision ECCV 2010*, 6314:143–156, 2010.

[8] H. Riemenschneider, U. Krispel, W. Thaller, M. Donoser, S. Havemann, D. Fellner, and H. Bischof. Irregular lattices for complex shape grammar facade parsing. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1640–1647, 2012.

[9] K. E. a. V. D. Sande, C. G. M. Snoek, and A. W. M. Smeulders. Fisher and VLAD with FLAIR. *Cvpr*, pages 2377–2384, 2014.

[10] J. R. R. Uijlings, K. E. a. van de Sande, T. Gevers, and a. W. M. Smeulders. Selective Search for Object Recognition. *International Journal of Computer Vision*, 104:154–171, 2013.

[11] A. Wendel, M. Donoser, and H. Bischof. Unsupervised Facade Segmentation Using Repetitive Patterns. *Pattern Recognition*, pages 51–60, 2010.

Fast 3D Terahertz Imaging Using a Sparse Array

Bessem Baccouche, Patrick Agostini and Fabian Friederich
Department of Materials Characterization and Testing
Fraunhofer Institute for Physical Measurement Techniques
Kaiserslautern, Germany

Wolfgang Sauer-Greff and Ralph Urbansky
Department of Electrical and Computer Engineering
University of Kaiserslautern
Kaiserslautern, Germany

***Abstract*—Terahertz imaging technology has proven to be highly suitable for nondestructive testing applications of dielectrics. Due to their capabilities to resolve depth information, frequency modulated continuous wave imaging systems operating in the lower terahertz regime have shown a great potential for volume inspection in industrial applications. We developed a 3D imaging system using a sparse multistatic line array in combination with a band-conveyor. The continuous-wave system operates in the frequency range from 75 GHz to 110 GHz with stepped frequency modulation. For 3D image generation we derivate a digital beam forming algorithm and show two possible implementations in order to accelerate image generation. The imaging performance of the system is experimentally verified.**

Keywords—Terahertz; multistatic; sparse; 3D imaging; real-time

I. INTRODUCTION

The spectrum of terahertz radiation ranges from around 100 GHz up to 10 THz, thus it closes the gap in the electromagnetic spectrum between millimeter wave radiation and infrared light. Since many dielectric materials such as plastics, ceramics and foams are transparent for terahertz waves especially in the lower terahertz regime (below 1 THz), terahertz imaging systems are valuable tools for nondestructive testing applications. Many interesting terahertz imaging systems targeting industrial applications have been developed in the last years. As reported in [1] focal plane array (FPA) imaging systems are suitable for this purpose and meanwhile operate at fast video rates. However, they still require a trade-off between field of view and image resolution. On the other side common concepts based on quasi-optics provide only a limited depth of focus. A very promising alternative which overcomes the mentioned limitations is the use of sparse arrays together with digital beam forming (DBF) techniques. These techniques reduce the amount of required sensor elements significantly and enable to benefit from powerful signal processing methods for 3D image generation such as being used in the fields of ultrasound and radar imaging. We developed a 3D terahertz imaging system using a sparse line array of transmitters and receivers in combination with a band-conveyor. In section II the system concept is presented. In section III a DBF algorithm is derived and two possible time accelerated implementations are shown. In section IV we experimentally verify the imaging quality of the system and in section V some conclusions are drawn.

II. SYSTEM CONCEPT

Fig. 1. Sparse line array based 3D terahertz imaging system.

The system operates in the frequency range from 75 GHz to 110 GHz using stepped frequency modulation with a bandwidth of 35 GHz in order to achieve high range resolution. The core component of the system is a sparse line array containing 12 transmitters and 12 receivers (Fig. 1). This is equivalent to an array with 144 virtual elements that sample the measurement scene along the array axis. A band-conveyor moves the object under test in the direction perpendicular to the array axis. This way a synthetic aperture is generated in this direction. While the object is moving at a constant velocity, the transmitters illuminate it sequentially and the receivers simultaneously record back-scattered radiation. Using DBF algorithms, which are suitable for general purpose computing on graphics processing units (GPGPUs), 3D terahertz images of the object are generated. More details on the system can be found in [2].

III. 3D DIGTAL IMAGE GENERATION

To illustrate the focusing method we consider the simple imaging problem depicted in Fig. 2. It shows a generic multistatic array containing N_{Tx} transmitters and N_{Rx} receivers and a volume V with an arbitrary shape. Assuming a linear scattering problem, the volume is considered to be composed of infinitesimal part volumes (voxels) located at $\vec{r}_V$ with a corresponding constant reflectivity a_V. Hence the volume is described by a so called reflectivity function $\sigma(\vec{r})$ which is given by,

$$\sigma(\vec{r}) = \int_V a_V \delta(\vec{r} - \vec{r}_V) d\vec{r}. \quad (1)$$

The coordinates of the receivers and the transmitters are given by $\vec{r}_{Rx}$ and $\vec{r}_{Tx}$, respectively. By ignoring the radiation patterns of the antennas and the free space propagation attenuation, the measured signal is modeled by the superposition of part waves reflected from the different voxels and is given by,

$$s(\vec{r}_{Tx}, \vec{r}_{Rx}, f) = \int_V \sigma(\vec{r}) \exp\left(j \frac{2\pi f d_B}{c}\right) d\vec{r}, \quad (2)$$

with,

$$d_B = |\vec{r}_V - \vec{r}_{Rx}| + |\vec{r}_V - \vec{r}_{Tx}|, \quad (3)$$

f is the signal temporal frequency and c is the speed of light in vacuum. An estimation of the reflectivity function at a desired voxel $\tilde{\sigma}(\vec{r}_V)$ can be obtained by correlating the measured signal with the complex conjugated of the response of this voxel,

$$\tilde{\sigma}(\vec{r}_V) = \sum_{N_{Tx}} \sum_{N_{Rx}} \sum_{N_f} s(\vec{r}_{Tx}, \vec{r}_{Rx}, f) s_V(\vec{r}_{Tx}, \vec{r}_{Rx}, f), \quad (4)$$

with

$$s_V(\vec{r}_{Tx}, \vec{r}_{Rx}, f) = \exp\left(-\frac{2\pi f d_B}{c}\right), \quad (5)$$

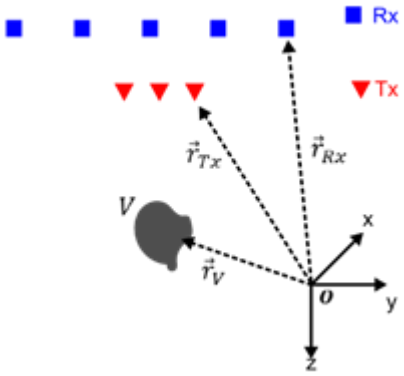


Fig. 2. Geometry of a generic imaging problem.

and N_f the number of frequency points. Equation (4) represents a matched filter. The direct implementation of (4) results in a computational load in the order of $O(N_{Tx}N_{Rx}N_fN_V)$ with N_V the number of pixels in the reconstructed image and makes the system unsuitable for real-time operation, hence we use algorithms that base on the Fast-Fourier transform (FFT) to accelerate the execution time of the algorithms. In the following two different FFT based implementations of (4) are explained. Using the inverse FFT $\mathcal{F}_f^{-1}\{.\}$ measured data can be focused in range and then interpolated to the desired voxel.

$$s(\vec{r}_{Tx}, \vec{r}_{Rx}, t) = \mathcal{F}_f^{-1}\{s(\vec{r}_{Tx}, \vec{r}_{Rx}, f)\} \quad (6)$$

$$\tilde{\sigma}(\vec{r}_V) = \sum_{N_{Tx}} \sum_{N_{Rx}} s(\vec{r}_{Tx}, \vec{r}_{Rx}, t = d_B/c) \quad (7)$$

Equation (7) is known as the global back projection (GBP) and its implementation has a computational load of $O_{GBP}(N_{Tx}\,N_{Rx}\,(N_V + N_f \log(N_f)))$. A further time efficient implementation of the GBP is achieved using the so called fast factorized back projection (FFBP) algorithm, in which the reconstruction scene is iteratively splitted into sub-images, and adjacent transmitter/receiver pairs are combined to focus to the respective sub-image centers. Thereby the final number of required back projections (7) for image reconstruction is reduced. The computational gain of this algorithm is mainly dependent on the factorization-depth, and computational loads down to $O_{FFBP}(N_V)$ can be attained [3].

IV. Measurements

To experimentally verify the imaging performance of the system a US Air Force resolution test pattern is used as a test object, which has been recorded in 60 cm distance from the center of the array. Fig. 3 shows a photograph of the sample and a reconstructed terahertz image. Features of 3 mm size can be resolved. Here we used a direct implementation of (4) to verify the imaging quality of the system.

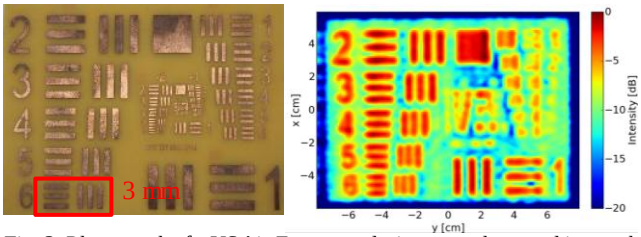


Fig. 3. Photograph of a US Air Force resolution test chart and its terahertz image.

V. Conclusion

A 3D terahertz imaging system based on a sparse line array has been developed. By using digital beam forming techniques we are able to generate 3D images with a spatial resolution of up to 3 mm. Ongoing works are focusing on accelerated image reconstruction and reduced measurement times.

Acknowledgment

This work was supported by FhG Internal Programs under Grant No. Attract 018-692 158, the European Fund for Regional Development (EFRE) under Grant No. 961-52 123/40(16) 81047970 and the innovation center for Applied System Modeling for Computational Engineering (ASM4CE) in Kaiserslautern, Germany.

References

[1] F. Friederich, et al., "THz active imaging systems with real-time capabilities," IEEE Trans. Terahertz Sci. Technol., vol. 1, no 1, pp. 183-200, Sept. 2011.

[2] B. Baccouche, et al., "A sparse array based sub-terahertz imaging system for volume inspection," in proc. Of the 45th European Microwave Conference., pp. 438-441, Sept. 2015.

[3] J. Moll, et al., "Towards three-dimensional millimeter-wave radar with the bistatic fast-factorized back-projection algorithm—Potential and limitations," IEEE Trans. Terahertz Sci. Technol., vol. 2, no 4, pp. 432-440, July 2012.

Multifunctionality by embedded steel fibers for improved aircraft composites

Sebastian Backe
Institute of Materials Science and Engineering (WKK)
Hybrid Materials Engineering Group
University of Kaiserslautern
Germany
sbacke@mv.uni-kl.de

Benedikt Hannemann
Institute for Composite Materials (IVW GmbH)
University of Kaiserslautern
Germany
benedikt.hannemann@ivw.uni-kl.de

***Abstract* - Composite airframe structures show beside an excellent strength to density ratio a poor electrical conductivity and ductility compared to metals. Against this background, the electrical and mechanical performance of our new composite reinforced with carbon and steel fibers, is tested and discussed in the paper.**

Keywords - Hybrid materials, Steel fibers, Multifuncitonality, Electrical conductivity, Damage tolerance

I. INTRODUTION AND MOTIVATION

The future civil aircraft production will meet different challenges. Pushed by airlines and politics the most important aims besides rising comfort requirement are saving operational aircraft costs and reducing pollution. One of the major contributors to operational costs is the primary structure. Reducing the weight of aircraft fuselage would not only cause less pollution but also saving fuel. Fig. 1 shows the total operational cost breakdown of an Airbus A320.

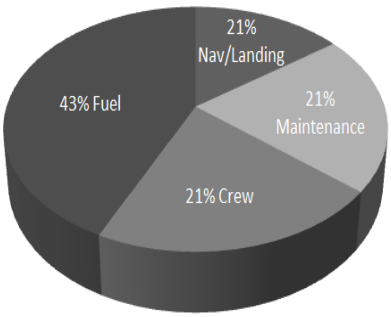


Fig. 1. *Cost breakdown for the Airbus aircraft A320 [1]*

Nearly half of the complete operational costs results from fuel consumption [1]. Reducing weight would directly lead to lower fuel consumption and therefore less operational costs. The past few years showed, one of the promising approaches to save weight is the implementation of Fiber Reinforced Polymers (FRP) instead of conventional metallic structures. The rise of parts made out of FRP began in the 1970´s with the Airbus A300. Several parts like leading and trailing edges on the tail fin were produced out of Carbon Fiber Reinforced Polymers (CFRP) [3]. The technical improvements over the last 40 years particularly in processing and production process afforded the constructors to realise a fast increase of the FRP wt. -% in primary structures, shown in Fig. 2.

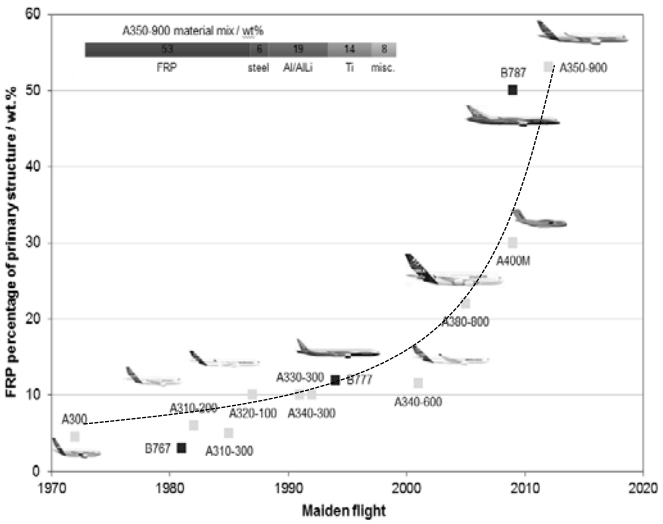


Fig. 2. *Use of FRP in primary aiframe structures [2]*

CFRP parts offer not only an outstanding structural performance to mass ratio, but also excellent fatigue behavior and corrosion resistance. However, the lightweight potential of CFRP is limited due to the poor electrical conductivity compared to metals, especially aluminum alloys. Many examples like metal mesh on outer skin for lightning strike protection, wires for electrical grounding or overbraiding of cables to provide electromagnetic shielding show the need of additional metallic masses to provide electrical functions in aircraft industry. Compared to metals further disadvantages are the brittle failure behavior and a low level of damage tolerance against probable impacts resulting in a minimum wall thickness, which exceeds strength and stability requirement, but also the lowest necessary structural weight. The challenge of modern CFRP airframe structure is a further mass reduction without losing important structural integrity but gaining

electrical conductivity. Smart material design is the key to a new level of multifunctionality. First researches concentrated on modifying the CFRP structures with carbon nanotubes to get an increased conductivity level [4]. However a level of conductivity providing the minimum necessary level of electrical properties for aircraft applications could not be reached. A promising new approach to multifunctional composites is the implementation of ductile steel fibers. First results in [4] show significant increases of energy absorption. Beside the improvements of mechanical properties, steel fibers have a much better electrical conductivity compared to carbon fibers. The idea to embed continuous steel fibers is used yet for reinforcement in tires for automotive industry [5]. The benefit of combining the positive properties from carbon and metal fibers are shown in Fig. 3.

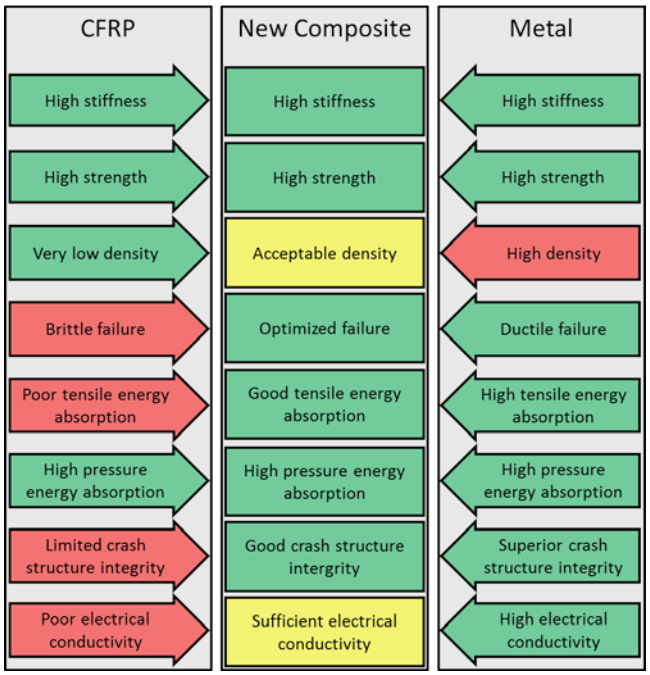


Fig. 3. *Benefits of Metal and Carbon Fiber Reinforced Polymers*

Especially for civil aircraft applications the benefit would be immense. New airliners like Boeing´s 787 or Airbus´ A350 have a fuselage made out of CFRP. The total amount of FRP-structures is over 50 wt. -% [7]. However, because of the poor electrical conductivity additional metallic masses are needed, to provide the necessary level of conductivity for lightning strike protection. But lightning strike protection is not the only function which cannot be adopted by CFRP airframe structures. Signal transfer, grounding, isolation or insulation are just a few examples for limits of CFRP`s multifunctionality. Additionally the low damage capability against probable impacts is the limiting driver for a minimum wall thickness criterion. So the lightweight potential for primary fuselage structures is not yet fully exploited. Fig. 4 shows a small overview of functions and their corresponding responsible materials. Obviously only a small range of functions is fulfilled by CFRP structures. Therefore further researches and developments are necessary to create a new composite material adapted to the needs of an airframe structure. The new developed composite specially focuses on the red marked demands.

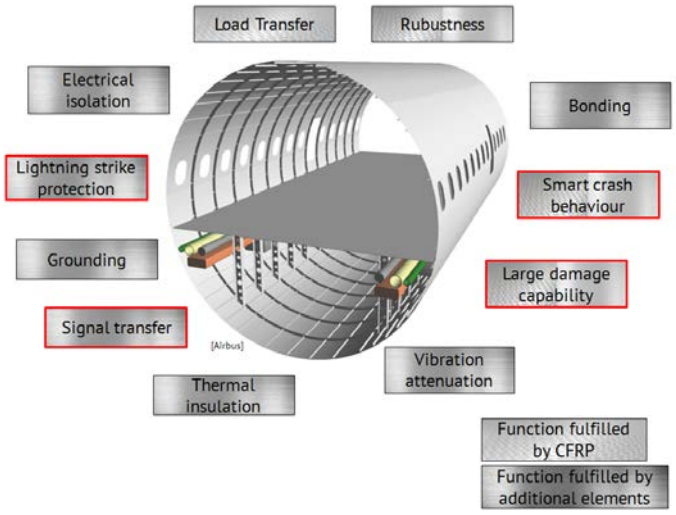


Fig. 4. *Function integration in fuselages*

II. Materials

For validation of mechanical and electrical properties six different types of unidirectional laminates were produced. The manufacturing process was a combination of tape deposition and filament winding. Preimpregnated carbon fiber layers (prepreg) were stacked on a flat steel sheet and wrapped in dry stainless steel fibers. The manufacturing process is schematically shown in Fig. 5.

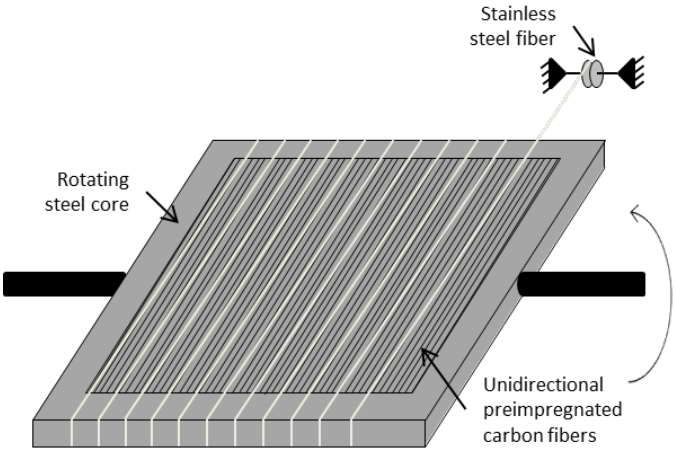


Fig. 5. *Manufacturing process of MCFRP*

Within this project two different types of fibers were used: A high tenacity carbon fiber from Toho Tenax and an annealed stainless steel fiber (1.4301). The stainless steel fiber occurs in bundles consisting of seven fibers, each with a diameter of 60 µm. The thickness of steel fibers is compared to carbon fibers ten times higher. The experimental identified specific values are listed in TABLE I Compared to the values in [9] the experimental show only a very small deviation.

TABLE I. EXPERIMENTAL VALIDATED PROPERTIES OF CARBON AND STAINLESS STEEL FIBER

	Carbon fiber	Steel fiber
Density / g/cm³	1.77	7.95 ± 0.01
Tensile modulus / GPa	240 ± 3	176 ± 7
Ultimate tensile strength / MPa	4806 ± 125	897 ± 2
Failure strain / %	1.85 ± 0.04	32.31 ± 2.01
Electrical resistivity / Ωm	1.6×10^{-5}	6.92×10^{-7}
Filament diameter / µm	5	60.0 ± 0.4
Roving Size	12k	7

The present study concentrates on six different types of composite explained afterwards.

A. *Carbon Fiber Reinforced Plastics (CFRP)*

One of the reference materials is conventional CFRP, consisting of 64 vol. -% carbon fibers and 36 vol. % epoxy resin. Eight prepreg layers were used. A cross section is shown in Fig. 6.

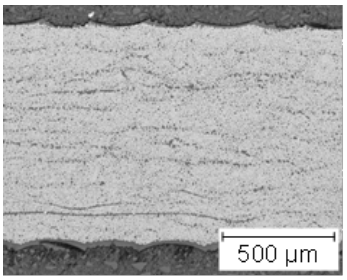


Fig. 6. *Lightmicroscopical image of a CFRP cross section*

B. *Metal Fiber Reinforced Polymers (MFRP)*

The second reference material is epoxy resin reinforced with stainless fibers. The whole composite consists of six layers steel fibers and 14 layers resin. All in all the fiber volume content is 59 %. Because of the bigger diameter the fiber volume content is a little lower compared to the CFRP. A cross section is shown below in Fig. 7.

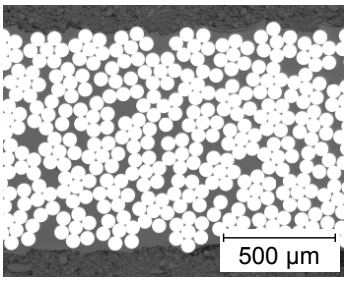


Fig. 7. *Lightmicroscopical image of a MFRP cross section*

C. *Metal and carbon fiber reinforced polymers (MCFRP) 10 / 20 / 20i / 20a*

The composites with steel and carbon fiber reinforcement were produced in four different variations. The first version shown in Fig. 8a has a steel fiber volume content of 10 % and 52 vol. -% carbon fibers. The layup is homogeneous, so it is an alternating process between stacking carbon fiber prepregs and winding steel fiber bundles. Fig. 8b shows a cross section of MCFRP 20. The layup is similar to MCFRP 10, but the winding feed is smaller and so the fiber volume content of steel is rising up to 18 vol. %, while the volume percentage of carbon fibers is decreasing to 47 %. In both concepts the layup consists of eight layers steel- and seven layers carbon fibers.

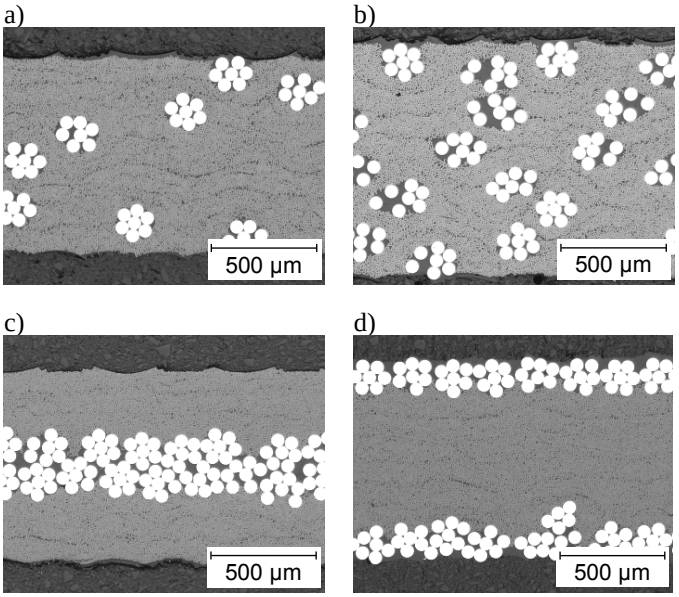


Fig. 8. *Lightmicroscopical cross section images of a) MCFRP 10, b) MCFRP 20, c) MCFRP 20i , d) MCFRP 20a*

The laminate layups from MCFRP 20i and MCFRP 20a shown in Fig. 8 c) and d) are different compared to homogeneous layups of MCFRP 10 and MCFRP 20. It is visible that steel- and carbon fibers are in separated layers. On the one hand Fig. 8 c) shows the steel fiber location in the core of laminate setup, on the other hand Fig. 8 d) shows the steel fibers located on the outside of the laminate. Both concepts include 21 vol. -% steel fibers and 46 vol. -% carbon fibers. The laminates have two layers of steel, winded with a much smaller feed than in the homogeneous concept and six prepreg layers of carbon fibers. All introduced unidirectional laminates have a thickness of roughly 1 mm. The density for the different materials is determined by the following equation (1) and listed in TABLE II.

$$\rho = \sum_i v_i * \rho_i \quad (1)$$

ρ: Density of the composite

v_i: Volume content of individual component

ρ_i: Density of individual component

The density of the hybrid composites rises with the steel fiber volume content. Compared to light metal alloys like aluminum the densities of the MCFRP´s are quite similar, however the mechanical properties should be much better.

TABLE II. DETERMINED DENSITIES OF LAMINATE SETUPS

	Density / g/cm³
CFRP	1.6
MFRP	5.4
MCFRP 10	2.25
MCFRP 20	2.79
MCFRP 20i	2.95
MCFRP 20a	2.95

III. ELECTRICAL IMPROVEMENTS

Electrical conductivity measurements for the laminate in-plane properties were examined by four wires Ohms measurements on a LabVIEW-based data logging system. For this purpose, direct current (DC) was applied to the short faces of rectangular specimens via copper electrodes with a defined pressure. Concerning the significant difference in electrical conductivity of carbon fibers and dielectric polymer matrix, the contact between composite sheet and metallic electrode is inherently discontinuous. To reduce the influence of contact resistance the following procedure was applied. The short faces of the specimen were prepared with a grit of 1200, cleaned by ethanol and brushed with a silver conductive paste. The surface temperature was monitored during the conductivity experiments by IR-thermography to exclude thermal-induced changes of the electrical conductivity. The experimental setup is presented in Fig. 9.

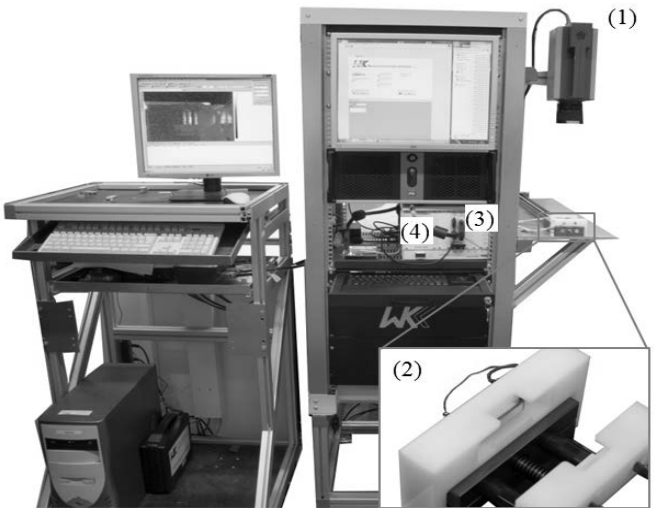


Fig. 9. *Experimental setup for electrical properties: (1) IR-Thermography, (2) specimen fixture with copper electrodes, (3) DC-supply, (4) voltage metering*

For a period of 10 seconds a various direct current of 200 mA in maximum was applied to the specimen. The reading was repeated for specimen length of 30 to 250 mm. The voltage was metered for DC stages of 10, 20, 50, 100 and 200 mA. Based on the specific geometry of each specimen the specific electrical conductivity was calculated by Ohm´s law. The results of the conductivity measurements for the different composite configurations are summarized in Fig. 10.

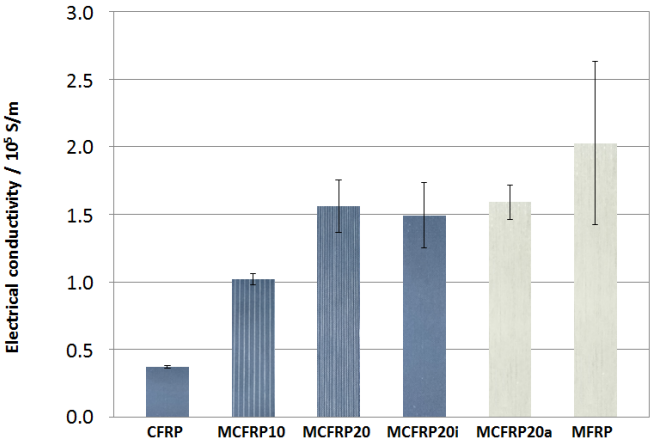
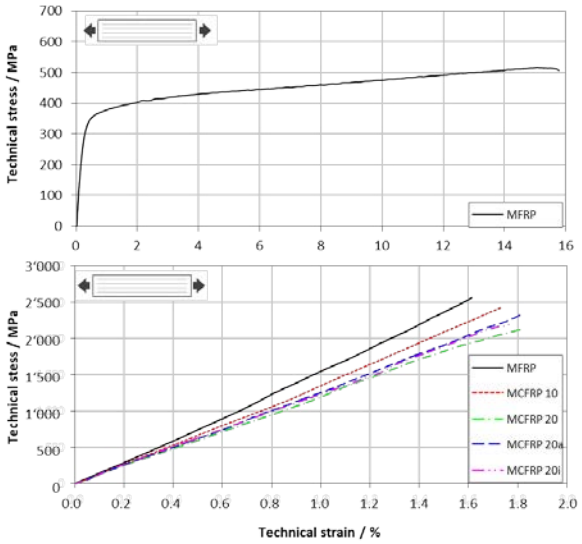


Fig. 10. *Electrical conductivity of unidirectional CFRP, MFRP and MCFRP*

The experimental results of the electrical measurements show a pronounced increase of electrical conductivity as a function of stainless steel fibers volume content. The hybrid composite reinforced with 20 vol. -% stainless steel fibers (MCFRP 20) are characterized by the highest electrical conductivity which is about four times higher than conventional CFRP. The level of electrical conductivity for MCFRP 20, MCFRP 20i and MCFRP 20a is comparable, so obviously the laminate layup has no influence on the in plane conductivity. The observed increased standard deviation of the MCFRP in comparison to CFRP is explainable by the complex contact conditions due to the three different phases of the SCFRP and the copper electrodes. However, the theoretical potential could not yet be verified by these experiments.

IV. MECHANICAL IMPROVEMENTS

Monotonic tensile tests longitudinal to the fiber orientation were performed with a hydraulic testing machine of type Zwick Roell HTM 50/20 in dependence on DIN EN ISO 527-5:2010-01. The rectangular specimens were sized 250 mm x 15 mm and provided with 3 mm thick chamfered aluminum end tabs. The specimens were clamped with a gauge length of 150 mm and loaded with a test speed of 3 mm/s. The deformation was recorded by a high-speed camera and interpreted by an optical 2D measuring system. Representative stress-strain curves are shown in Fig. 12.



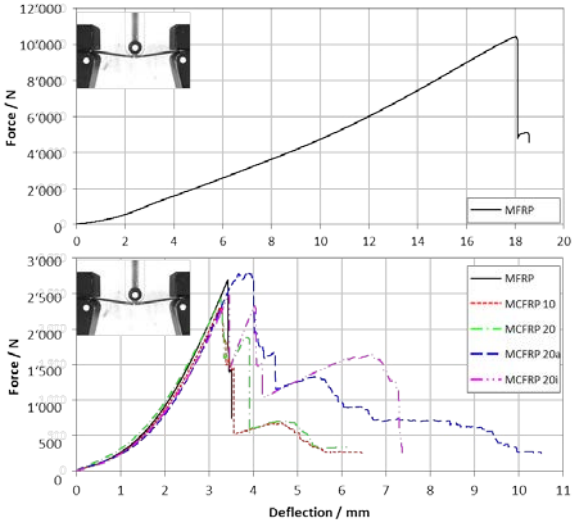
Material	E-Modulus / GPa	Strength / MPa	Elongation / %
CFRP	146 ± 5	2'491 ± 85	1.61 ± 0.06
MFRP	100 ± 3	514 ± 7	14.75 ± 0.96
MCFRP 10	128 ± 5	2'322 ± 73	1.74 ± 0.04
MCFRP 20	125 ± 7	2'093 ± 72	1.77 ± 0.08
MCFRP 20a	133 ± 3	2'259 ± 78	1.79 ± 0.04
MCFRP 20i	133 ± 4	2'215 ± 31	1.81 ± 0.04

Fig. 11. *Stress-strain relationship for unidirectional monotonic tension*

CFRP and the hybrid composites (MCFRP 10, MCFRP 20, MCFRP 20a, MCFRP 20i) show a linear elastic behavior with similar strains at failure between 1.61 % and 1.81 %. The mean ultimate tensile strength decreases from 2'491 MPa to 2'093 MPa, 2'322 MPa, 2'259 MPa and 2'215 MPa. Moreover, the stiffness declines from 146 GPa for pure CFRP to approximately 130 GPa for the hybrid composites. By contrast, pure SFRP shows a pronounced ductile material performance with a yield strength of 349 MPa, a UTS of 514 MPa and an ultimate strain at failure of 14.75 %.

Furthermore, flexure tension tests were carried out on a conventional testing machine of type Zwick 1474. For this purpose, rectangular coupons in the form of 125 mm length and 15 mm width flat strips were clamped at their short edges with a remaining span of 50 mm and centrally loaded by a rounded indenter with a cross head speed of 2 mm/min. The supports were chamfered with 5 mm radius. Fibers were orientated in parallel to the long edge of the specimen. Typical force-displacement curves are given in Fig. 12. The integrals of the curves yield the absorbed deformation energy.

Pure CFRP shows a brittle material performance. Failure occurs abruptly and singularly with a bearable deformation of 3.58 mm. By contrast, the incorporation of stainless steel fibers into CFRP causes to a noticeable post failure behavior. In case of the hybrid composites with homogenously distributed stainless steel fibers, first failure occurs at 3.36 mm (MCFRP 10) and 3.22 mm (MCFRP 20), respectively. Afterwards, the coupons can be further loaded at a reduced level of force. Total failure is reached at an increased deflection of 6.11 mm and 6.33 mm, respectively. An even better material performance can be realized by localizing the metal fibers in the core (MCFRP 20i) or top layers (MCFRP 20a) of the hybrid laminate. Again, similar to the other composites, first failure occurs at 3.48 mm and 3.90 mm, respectively. However, the maximum deflection can be further enhanced up to 7.32 mm and 10.09 mm, while the force remains on a higher level compared to the homogenized SCFRPs. Pure SFRP bears a deflection of 17.48 mm.



Material	Max. Force / N	Initial failure / mm	Max. deflection / mm	Absorbed energy / J
CFRP	2'715 ± 105	3.41 ± 0.04	3.58 ± 0.12	3.38 ± 0.09
MFRP	9'797 ± 742	17.14 ± 0.89	17.48 ± 1.20	75.96 ± 11.30
MCFRP 10	2'396 ± 93	3.36 ± 0.07	6.11 ± 0.28	4.24 ± 0.21
MCFRP 20	2'406 ± 57	3.22 ± 0.07	6.33 ± 0.36	5.16 ± 0.34
MCFRP 20a	2'710 ± 115	3.90 ± 0.08	10.09 ± 0.58	9.41 ± 0.48
MCFRP 20i	2'566 ± 58	3.48 ± 0.01	7.32 ± 0.29	8.47 ± 0.46

Fig. 12. Force-displacement diagrams for combined tension-bending load

In case of pure tensile load in parallel to the fiber orientation, both CFRP and the hybrid composites show brittle material performances with comparable ultimate strains to failure. Despite the incorporation of highly ductile stainless steel fibers, a pronounced post failure behavior could not yet be demonstrated with the analyzed material samples. Due to the homogenous stress state and material deformation, the carbon fibers generate elastic energy on the entire length of the specimen. At the time of their failure, this energy is abruptly released and transferred by the matrix to the ductile steel fibers. However, this energy transfer is limited very locally at the point of failure. By this means, the available ductility of the steel fibers can only be addressed on a small scale, causing merely a slight macroscopic elongation. The total elongation at break coincides with the strain at failure of the carbon fibers. Improvements are expected by higher steel fiber fractions (i.e., less elastic energy that must be absorbed by a bigger amount of plastifying steel fibers), but would cause higher material densities. This assumption is supported by the test results of pure SFRP, which demonstrate a pronounced plastic deformation. Furthermore, a reduced fiber-resin-adhesion could enable unhindered deformation of the metal fibers over more extensive distances, but would certainly impact other important properties.

The potential of incorporated steel fibers for an improved fracture and impact behavior of the hybrid composite becomes apparent in the flexure tension tests. As in many applications, load is very locally introduced. High stress peaks occur in the loaded regions, while stress states in the rest of the specimen are significantly lower. In case of the hybrid composites, a certain level of force can be obtained due to the incorporated steel fibers. Related to the very local loading, the carbon fibers store only a small amount of elastic energy. After sudden failure of the carbon fibers, the steel fibers continue to yield without failure and bear further deflections. The enhanced maximum deflection leads to an increase in energy absorption of 25 % for MCFRP 10 and 53 % for MCFRP 20, in comparison with pure CFRP. Positioning the steel fibers in the top or core layers of the laminate can further enhance the damage tolerance of the composite. In case of MCFRP 20i, the stiff and strong carbon fibers in the top layer enable a higher resistance against bending, while the steel fibers in the core layer can plasticize and maximize the structural integrity. The energy absorption can be increased by 151 %. In case of MCFRP 20a the steel fibers in the outer plies sustain the bearing cross section until higher deflections. By this means, the energy absorption can be risen by 179 %.

V. Conclusion and Outlook

The promising results in chapter III showed a pronounced increase of electrical conductivity with increased steel fiber volume content. With 20 vol. -% steel fiber reinforcement the electrical conductivity was improved by a factor of 4 compared to conventional CFRP. It was shown that the laminate layup has no influence on the electrical in-plane properties of the composite. In Chapter IV the results of the tensile tests showed no increase of strain to failure and a light decrease of stiffness. In case of bending-tension loading the achieved properties of MCFRP are more promising. A significant increase of absorbed energy is corresponding to an increase of steel fibers. Especially the laminate setup with separate layers of carbon- and steel-fibers offered an increased deformation energy by a factor of nearly 3compared to conventional CFRP. Transferring the results from unidirectional laminates into the production of the multidirectional setups, the separated layers seem to be the more promising. The first schematically cross sections are shown in Fig. 11.

a) b)

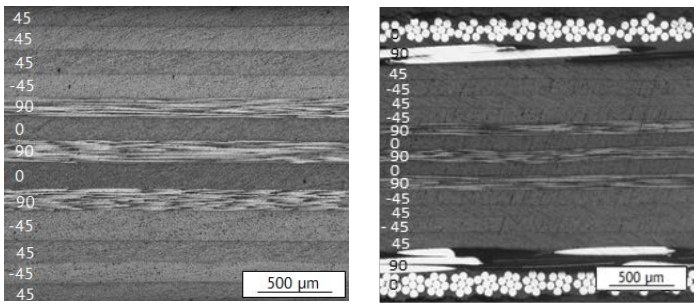


Fig. 13. *Lightoptical micrograph of a) CFRP and b) MCFRP multidriectional laminate layups*

The influence of embedded steel fibers to the fatigue properties of the composite are going to be evaluated in low- and high-cycle-fatigue tests. Additionally specific tests for aircraft applications like open- or filled-hole-tension are carried out in future.

In summary, the embedded steel fibers in conventional CFRP structures seem to be a promising new approach to a new level of multifunctionality, especially for the application in next generation aircraft structures.

VI. Acknowledgement

The financial support of the German Research Foundation (DFG) within the projects BR 4262/2-1 and BA 4073/6-1 is gratefully acknowledged. Prepreg and resin films were kindly supplied by Cytec Engineered Materials GmbH (Östringen, Germany).

VII. References

[1] Airbus, Flight Operations Support & Service, Issue 2, 2008

[2] U.P. Breuer, Civil Aircraft Composite Technology

[3] Airbus Group, URL: http://www.airbus.com/company/history/the-narrative/first-order-first-flight-1970-1972/, cited 27.10.2015

[4] Spitalsky, Z.; Tasis, D.; Papagelis, K.; Galiotis, C.: Carbon nanotube - Polymer composites: Chemistry, processing, mechanical and electrical properties. In: Progress in Polymer Science 35 (3), 2010, Pages 357–401.

[5] M.G. Callens, L. Gorbatikh, I. Verpoest, Ductile steel fibre composites with brittle and ductile matrices. In: Composites Part A, Applied Science and Manufacturing, 2014

[6] Rao S, Daniel IM, Gdoutos EE. Mechanical properties and failure behavior of cord/rubber composites, Applied Composite Materials 11: 353-375, 2004

[7] Airbus Group, URL: http://www.airbus.com/aircraftfamilies-/passengeraircraft/a350xwbfamily/technology-and-innovation/, cited 27.10.2015

[8] Boeing Group, URL: http://www.boing.com/ cited 27.10.2015

[9] Toho Tenax Europe GmbH: HTS Delivery program and characteristics. Version 05, 2014-09-30. http://www.tohotenax-eu.com. cited: 27.10.2015

In Situ Large Data Visualization using Layered Depth Images

T. Biedert and C. Garth
University of Kaiserslautern, Germany

***Abstract*—Contour trees and other topological visualization methods are promising tools for in situ analysis in large-scale, distributed scientific computing scenarios. We demonstrate in several experiments that such techniques provide effective data reduction in combination with layered image-based storage.**

I. Introduction

At the current scale of computational capability provided by large-scale parallel computer architectures such as commodity clusters and modern supercomputers, high-fidelity computational simulation models have assumed a significant role in scientific research and engineering applications. However, this increased amount of computation has incurred architectural trade-offs. While arithmetic capacity and in-core memory have grown at a tremendous rate, I/O subsystems have not been able to keep abreast in relative bandwidth [1]. As a consequence, numerical data produced during typical simulations cannot be persistently stored, e.g. to hard drives, in its entirety; a lack of available I/O bandwidth would make this prohibitively costly with respect to time.

Visualization of large-scale simulation output thus has to rely on a number of different strategies to facilitate meaningful analysis in reasonable time frames. Multi-resolution schemes represent data on varying scales of resolutions and have a long standing tradition in this setting. They enable an essential compromise between fidelity and accuracy of visualization results on one hand, and computation and I/O bandwidth expended on the other. Among the large set of available methods, topological techniques such as the *contour tree* stand out because they are able to provide meaningful simplification for scalar fields.

A further smart approach to reduce the volume of data generated from simulations while preserving a significant amount of analysis flexibility was recently introduced by Ahrens et al. [2]. The underlying idea is to generate in situ an image database that stores multiple layers of predefined visualization renderings. The layers can then be composited in post processing depending on specific demands by the scientist conducting analysis.

In this context, the intent of this paper is to study the combination of in situ topological analysis with the image-based approach of Ahrens et al. Based on in situ contour tree analysis and simplification, we obtain a segmented representation of scalar fields contained in the simulation data at every time step. A rendering of this segmentation is then generated describing all components visible in every pixel (similar to an A-buffer), and stored together with the simplified contour tree. These ingredients can then be used in post-analysis to flexibly select specific subsets of the segmentation, after further simplification if required.

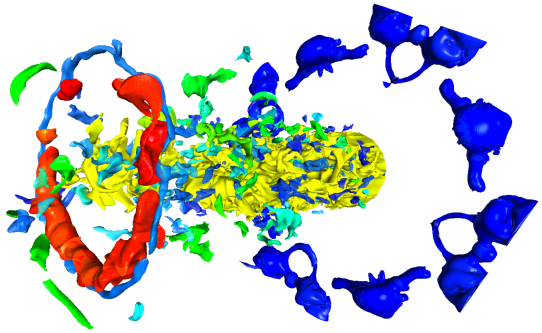


Fig. 1. Automatic persistence-guided simplification of a high-velocity fluid jet entering a medium at rest into 256 branches, which are visualized using an HSV color scale based on the average velocity value within each branch.

The overall intent of this paper is to investigate possible advantages of such an approach for the visualization of large-scale data. Specifically, after a brief review of relevant prior work (Section II), we make the following contributions:

- In Sections III through VII, we describe a system to combine in situ contour tree analysis, simplification, and image-based representation to facilitate reduced I/O requirements while preserving flexibility in visualization.
- We conduct several experiments to quantify the I/O savings possible from such an approach, and describe results and analysis in Section VIII.
- We anticipate that many enhancements and improvements are possible, and discuss a number of such opportunities in Section IX.

Our contribution is intended as a baseline demonstration of the feasibility and potential of the combination of topological analysis and image-based representation in large-scale in situ scenarios.

II. Related Work

A classic use case of topology in scalar field visualization is isosurface extraction, where typically several topological properties such as the number of connected components or the genus of the isosurface, i.e., the number of independent tunnels, are of central interest. Based on Morse theory, showing that topological changes in scalar fields defined on manifolds happen only at distinct critical points, Reeb graphs capture the

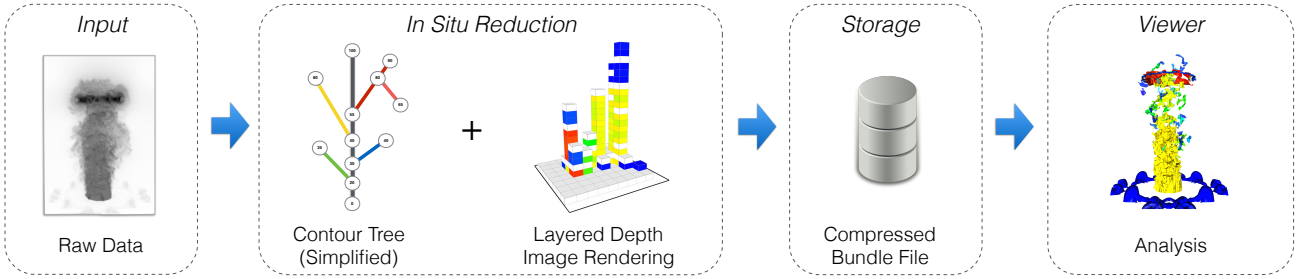


Fig. 2. Conceptual architectural overview of our contribution, providing a framework for the flexible exploration of in-situ generated compressed renderings. Volume data is segmented into topological regions based on its contour tree. After automatic filtration, the segments imposed by the simplified contour tree are intersected by a ray casting rendering approach. All resulting fragments form a layered depth image, which is compactly stored combined with the contour tree and of significantly smaller size than the original data.

topological evolution of individual contours using these critical points and their relationships. The efficient construction of Reeb graphs in general is still an active field of study [3]. However, for simply connected domains, the Reeb graph is always a tree structure [4], called *contour tree*, which is algorithmically computable for tetrahedral [5] and hexahedral [6] meshes.

Topological techniques have proven highly valuable for the analysis, visualization and exploration of scientific data. Typical applications include isosurface extraction and visualization [7], [8], [9], transfer function design [10], and scalar field exploration based on the detection of critical features and regions [11].

However, noise in data sets can lead to large numbers of irrelevant critical points, complicating feature-driven exploration based on topology. Topological simplification eliminates insignificant features by *cancellation*, i.e., removing irrelevant pairs of critical points. Pascucci et al. [12] introduced the *branch decomposition* of a contour tree, which can be interpreted as a hierarchy of contour tree simplifications based on leaf pruning and node collapses [13]. Since a branch is defined by a monotone path connecting a saddle and an extremum vertex, discarding a branch is equivalent to the topological cancellation of the respective two critical points. Our work is inspired by Weber et al. [14], who used the branch decomposition for the segmentation of a volume into regions of equivalent contour topology and applying separate transfer functions to each region.

In the context of large data, topological methods have been successfully applied to achieve compact representations of features [15], [16] and reduced data storage [17], while maintaining post-processing flexibilities [18]. More recently, also image-based approaches to large data storage and visualization emerged, enabling interactive exploration [19], [20] and reduction, trading off data size and complexity for scalar value uncertainty. Ahrens et al. [2] generated an in situ image database, storing multiple layers of predefined visualization renderings, which can then be composited in post processing. The central idea, which is also key to our work, is to achieve a massive data reduction when storing the simulation output of large-scale numerical simulations, while preserving visualization fidelity and flexibility for future post-processing.

III. Method Overview

We aim to achieve substantial reductions in data size by leveraging topological simplification and compact image-based storage. Our approach essentially consists of two components: an image *rendering library* and an interactive *viewer application*. A conceptual overview of the central steps in the processing workflow is depicted in Figure 2.

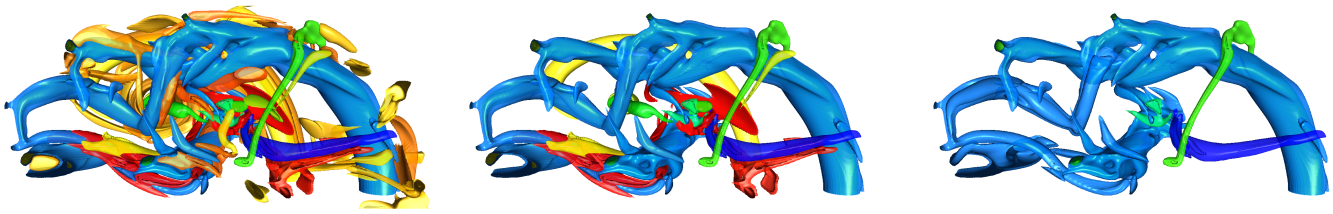
The rendering library is directly includable into simulation code and targets high-performance cluster environments or workstations. Simulation output data is segmented into topological regions based on its contour tree, which in turn are intersected by a ray casting rendering approach. However, instead of only determining the fragment closest to the camera, we store each intersection together with a set of local properties in a linked fragment list, similar to an A-buffer [21]. These properties include the element in the contour tree corresponding to the intersected segment and further additional attributes used for rendering such as the normal or ray depth at the intersection. The fragment lists combined with the contour tree are stored in a proprietary binary file format, which is considerably smaller in size than the original simulation data, yet provides enough flexibility for subsequent data exploration.

Once compressed layered depth image data has been generated, it represents a significant size reduction of the original input data and can be explored in the interactive viewer application running on the user's desktop computer. By interactively modifying visual properties of the regions imposed by the contour tree or applying further filtering schemes, the user can control which topological regions are displayed.

The following sections IV to VII outline the central concepts of our approach. For more theoretical and technical in-depth information please refer to [22].

IV. Segmentation and Filtering

Given a scalar field defined on a regular grid as input data, we construct the contour tree using the sweep and merge algorithm by [5], where the split and join sweep phases are executed in parallel. Afterwards, the so-called *branch decomposition* is computed using a variant of [12], which can be interpreted as a hierarchy of contour tree simplifications. The



(a) 90 branches (b) 30 branches (c) 10 branches

Fig. 3. Incremental simplification of the turbulent *plate* λ_2 data set by maximum branch persistence, enabling a flexible and topologically-guided exploration of vortex core structures. A standard HSV color scale is applied based on the branches' average λ_2 vortex detection values.

output of the above algorithm is a tree structure representing the branch decomposition of the contour tree and a mapping of vertices to their corresponding branch in the decomposition. Notably, the latter is used for rendering only, whereas the branch decomposition is stored in the final image.

Filtering is a central concept inherent to the hierarchical branch decomposition structure. In our context, if a branch is to be discarded, all vertices belonging to this branch are reassigned to the closest unfiltered parent branch. We employ this technique in two ways. First, the initial branch decomposition used for rendering can be controlled by a user-provided maximum number of branches. This is a crucial step to achieve significant data reduction at in situ time, as the initial unfiltered branch decomposition consists of numerous branches which are irrelevant for the later visualization, either due to negligible importance or being caused by noise. Second, in the interactive viewer application, the user can apply several consecutive filtering steps in order to simplify the layered depth image visualization. At any stage, branch selection for filtering can be performed manually by the user or automatically based on numerous criteria such as persistence or volume thresholds.

V. Depth Image Rendering

After the contour tree for a single simulation time step has been constructed, multiple layered depth image renderings can be computed based on its branch decomposition. Thus, given a scalar field defined on a regular grid, the vertex-to-branch mapping, camera position and orientation, resolution and an optional maximum number of depth layers, we perform a concurrent ray casting procedure for image generation.

The goal is to record all intersections of rays with the boundaries of the topological regions defined by the contour tree. Assuming a sequence of sample positions along the ray, we principally need to determine the branch index at each location and detect an intersection where it varies. To improve performance, our sampling algorithm traverses data voxels following a three-dimensional Bresenham approach. For each cell encountered, we check if all eight corner vertices belong to the same branch, in which case we can trivially use this respective branch index and return the entering intersection of the ray with the voxel boundary as intersection candidate. Otherwise, there are potential intersections with topological segment boundaries within the cell, which we approximate by a local uniform sampling scheme restricted to the inner cell volume. At each sample position, we need to determine the topological segment containing the given location, i.e., determine the respective branch index. We follow the approach presented in [14], which is based on the relation that monotone paths in the scalar field always map to monotone paths in the corresponding contour tree and vice versa. First, the unique monotone path going through the given sample position and two cell corners is determined by exploiting the linearity of the trilinear interpolant within the cell along axis-oriented lines. We then follow the monotone path in the contour tree until the branch containing the data value at the sample position is found.

Eventually, for each detected fragment, the associated branch index and additional optional parameters such as gradient of the scalar field or depth value are stored. Notably, the central idea of this proof-of-concept work is independent of the employed data sampling scheme, providing opportunities for performance optimization. Also, together with an appropriate intersection detection scheme, our approach is easily applicable to non-regular data.

We have implemented the above layered depth image rendering algorithm both as a CPU- and GPU-version. While the parallel CPU version of our library is based on the HPX framework [23] and targets rendering of large data sets in traditional distributed cluster environments without accelerator cards, we have additionally implemented GPU hardware acceleration as a proof of concept of our highly parallelizable algorithm. As described above, rendering by ray casting creates a list of fragments for each ray, representing the visual properties for all intersections with topological segments it encounters. Our implementation is conceptionally similar to per-pixel linked lists [24], a recent technique in computer graphics for hardware-accelerated order-independent transparency [25] based on the classic A-buffer approach [21]. For each pixel, a linked list of fragments keeps track of the intersections encountered during traversal. The respective fragment properties themselves are stored in a global shared fragment pool.

VI. Storage

Compact storage of results while maintaining flexibility for interactive exploration is central to our contribution. For a

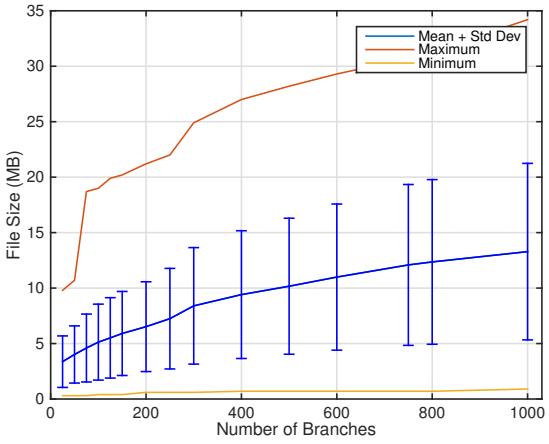


Fig. 4. Impact of limiting the maximum number of branches in the automatic simplification of the *jet5* λ_2 data set on the compressed layered depth image file size. Depth image file size can vary greatly depending on the complexity of the scene captured in the rendering, however in general stays significantly smaller than the input data size of 134 MB.

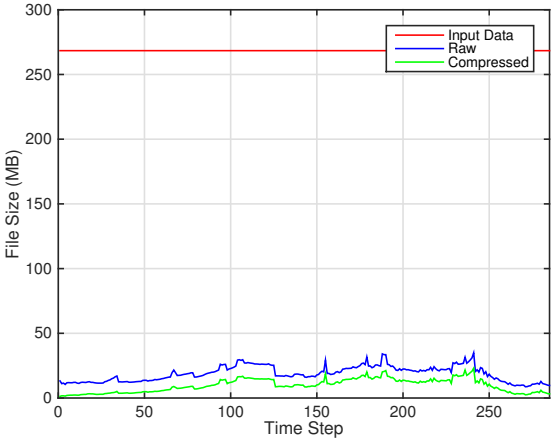


Fig. 5. Comparison of raw and compressed layered depth image sizes in relation to the input data size of the *plate* λ_2 data set across the full time range, using 256 branches. The reduction in file size is even more prominent for larger input data sets.

given (filtered) contour tree, multiple depth renderings from different perspectives and resolutions can be generated. These bundles are stored in a single file in HDF5 format with zlib deflate compression enabled at level 6, which we found to be a good compromise between size and speed.

The branch decomposition of the contour tree is serialized recursively in depth-first manner, where for each branch we store index, saddle value, extremum value, and volume. While during depth image rendering fragments are stored and manipulated as linked lists distributed across memory and potentially shared by rendering threads, this data can be compactly reorganized as contiguous blocks for permanent storage. For each pixel, we store the list of fragments, where a single fragment consists of its associated branch index and additional attributes such as normal.

Since the largest offset, the maximum number of fragments per pixel and the maximum branch index (i.e. total number of branches in the branch decomposition) are known, primitive data types used for storing the respective values can be intelligently chosen. As an example, if we can restrict the total number of branches to 256, this allows branch indices to be stored in single bytes, yet provides sufficient segmentation detail in many scenarios.

While storing the branch index of each fragment is mandatory, additional attributes are optional and depend on the intended visualization. In our studies, we additionally compactly store the normal at each intersection in two bytes using a spheremap transformation [26]. Also, we store the intrinsic and extrensic camera parameters used for image generation, serving as reference for lighting and shading in the viewer.

VII. Interactive Viewer

Compressed layered depth image bundle files can be loaded and interactively explored on the user's desktop machines in the interactive viewer application, the counterpart of the parallel rendering library in our framework. The interface is split in two parts: the visualization of the selected depth image and a tree widget used for the display and modification of the branch decomposition tree. Users can select multiple branches manually or automatically as described in section IV. Additionally, picking branches in the visualization using the mouse cursor is an efficient way of selecting regions of interest. Filters can be combined by applying them on top of previous filtering steps. Undo is supported by storing the history of filtering operations.

Depth images are interactively updated and displayed in real time using hardware-accelerated rendering similar to the techniques outlined in section V. For each pixel, the fragments are back-to-front composited with shading being computed based on the stored fragment normal and angle-based view-dependent transparency as presented in [27].

VIII. Results

We have tested our framework on two regular vector field data sets. The *jet5* data set (given on a 256×512×256 grid over 3000 timesteps) results from a direct numerical simulation of a jet of high-velocity fluid entering a medium at rest and exhibits progressively finer vortical structures in the velocity field. Similarly, the *plate* data set (with a resolution of 1024×256×256 over 285 timesteps) describes the mixing of fluid flowing past a plate at different speeds that undergo mixing due to viscous effects. See Figures 1 and 3 for example renderings of the above data sets.

All images depicted in this paper have been rendered at 1920x1080 resolution on a standard desktop workstation using an Intel Core i7-4770k quad-core CPU and a NVIDIA GeForce GTX 770 GPU. The complete topological segmentation of a single time step took less than one minute using the above hardware for both data sets. Rendering times varied

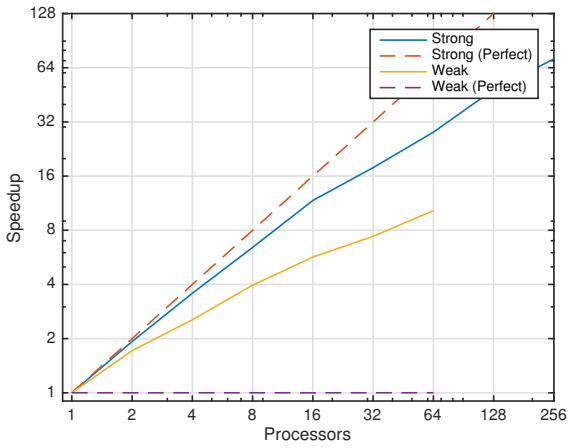


Fig. 6. Strong and weak scaling behavior of the distributed layered depth image rendering approach using HPX on the *Elwetritsch* cluster. Subblocks of the *jet5 velocity* data set are distributed across nodes and rendered in parallel. The resulting partial images are composited to a single layered depth image afterwards. We achieve overall good strong scaling and excellent weak scaling characteristics.

depending on the chosen perspective, but typically ranged from 20 to 60 seconds. The final compression and storage of the resulting depth images was performed in less than one second.

A. Compression

The key goal of our contribution is to provide a flexible trade-off between storage memory consumption and interactive data exploration, essentially controllable by the branch decomposition simplification level used for layered depth image rendering.

Figure 4 depicts the output bundle file sizes of our technique applied to the *jet5* data set in relation to several user-provided maximum branch numbers. For each number of branches, the graph shows the mean, standard deviation, minimum and maximum file sizes measured across the complete time range of the data set. Clearly, the mean file size is monotonously increasing with the number of branches. Notably, there is a small jump visible at 256 branches, when fragments are required to use shorts instead of bytes for branch index storage. However, in our studies, a maximum number of 256 branches has emerged as a very good compromise between file size and segmentation detail, providing a mean output size of approximately 7 MB, compared to the input data size of 134 MB for each scalar *jet5* time step, i.e., a size reduction of 95%.

The size reduction is even more prominent for full-view renderings of the larger *plate* data set. Figure 5 illustrates the raw output size and compressed output size of our algorithm applied to each time step in relation to the constant input data size. In contrast to the great variation in file size due to the continuously increasing complexity of the *jet5* data set over time, the graph reflects the rather uniform complexity of the *plate* data set. Also, the graphs clearly highlight the fundamental reduction in data size compared to the input data even when storing the layered depth images in raw format, i.e., without additional zlib deflate compression, which usually achieves further compression ratios of about 50-60%.

B. Analysis

By design, our approach features a powerful framework for hierachical simplification and automatic segmentation based on the branch decomposition of the contour tree, which is interactively controllable in the viewer application.

Figure 3a depicts a close-up on the central turbulences of the *plate* data set. After an initial simplification of the whole data set to 256 branches prior to layered depth image rendering, the surrounding branches have been manually pruned in the viewer application. The remaining 90 branches visible in the depicted scene have been colored using a HSV color scale based on the branches' average λ_2 values. In the following Figures 3b and 3c, the branch decomposition subsequently has been further simplified to 30 and 10 branches, respectively. Simplification was performed automatically based on sorted persistence as described in section IV, while maintaining the color scheme for non-discarded branches. One can clearly see the incremental reduction in complexity, which has been achieved with minimal user interaction and immediate visual feedback. Automatic persistence-based simplification to 256 branches of the *jet5 velocity* data set is shown in Figure 1.

In our studies, the topological segmentation based on the branch decomposition of the contour tree has proven itself useful as a flexible representation of the major structures of interest occurring in the data sets, which furthermore provides an intuitive approach to simplification and filtering in both pre- and postprocessing.

C. Scalability

We have studied the scaling characteristics of the task-parallel CPU-based layered depth image renderer on the *Elwetritsch* cluster at TU Kaiserslautern, which is depicted in Figure 6, using the example of the *jet5 velocity* data set.

Considering strong scaling, we achieve an overall good efficiency. As expected, there is a minor decrease when using more than 16 processors due to the required network communication between nodes for image composition after parallel rendering.

Moreover, the algorithm exposes excellent weak scaling behavior, which is of greater significance to practical considerations than strong scaling. To measure weak scaling, the *jet5 velocity* data set was resampled for each run proportionally to the respective increase in processors used. Interestingly, the results show that bigger volumes can be rendered using proportionally more processors in shorter time.

IX. Conclusion and Outlook

We have demonstrated the feasibility and potential of combining in situ topological analysis and compact image-based data representation. Our approach significantly reduces the amount of I/O bandwidth required to store the numerical

results of high-fidelity numerical simulations running on large-scale parallel computer systems, while preserving flexibility in visualization.

Based on in situ contour tree analysis and simplification, we obtain a segmented representation of scalar fields contained in the simulation data at every time step. Together with the simplified contour tree, we store a rendering of this segmentation that describes all components visible in every pixel. Rendering can leverage GPU-based hardware acceleration and exhibits good scaling behavior on distributed systems. The resulting compressed layered depth images can then be used in post-analysis to flexibly select specific subsets of the segmentation, and perform further topological simplification if required.

While our results already show substantial reductions in output file size, especially for larger data sets, our contribution is intended as a baseline demonstration investigating possible advantages of such an approach for the visualization of large-scale data. However, we anticipate that many enhancements and improvements of our approach are possible. Our technique could be easily extended to generate a complete in situ image data base from multiple perspectives, which can be combined in the viewer application to enable a flexible 3D data exploration, or even be used for reconstruction purposes. A critical shortcoming of our current implementation is frame-to-frame temporal consistency, since contour trees are computed and decomposed independently at each time step. Also, truely distributed parallel contour tree construction is still in its infancy [28]. Other optimizations might include multivariate topological segmentations, annotated contour trees or more sophisticated compression schemes. We will investigate these possibilities in future work.

ACKNOWLEDGEMENTS

This project was funded in part under EU Career Integration Grant #304099.

REFERENCES

[1] H. Childs, D. Pugmire, S. Ahern, B. Whitlock, M. Howison, Prabhat, G. Weber, and E. Bethel, "Extreme scaling of production visualization software on diverse architectures," *Computer Graphics and Applications, IEEE*, vol. 30, no. 3, pp. 22–31, May 2010.

[2] J. Ahrens, S. Jourdain, P. O'Leary, J. Patchett, D. H. Rogers, and M. Petersen, "An image-based approach to extreme scale in situ visualization and analysis," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '14. Piscataway, NJ, USA: IEEE Press, 2014, pp. 424–434.

[3] H. Doraiswamy and V. Natarajan, "Efficient algorithms for computing reeb graphs," *Comput. Geom. Theory Appl.*, vol. 42, no. 6-7, pp. 606–616, Aug. 2009.

[4] R. L. Boyell and H. Ruston, "Hybrid techniques for real-time radar simulation," in *Proceedings of the November 12-14, 1963, Fall Joint Computer Conference*, ser. AFIPS '63 (Fall). New York, NY, USA: ACM, 1963, pp. 445–458.

[5] H. Carr, J. Snoeyink, and U. Axen, "Computing contour trees in all dimensions," in *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '00, 2000, pp. 918–926.

[6] V. Pascucci and K. Cole-McLaughlin, "Parallel computation of the topology of level sets," *Algorithmica*, vol. 38, no. 1, pp. 249–268, 2004.

[7] M. van Kreveld, R. van Oostrum, C. Bajaj, V. Pascucci, and D. Schikore, "Contour trees and small seed sets for isosurface traversal," in *Proceedings of the Thirteenth Annual Symposium on Computational Geometry*, ser. SCG '97. New York, NY, USA: ACM, 1997, pp. 212–220.

[8] H. Carr and J. Snoeyink, "Path seeds and flexible isosurfaces using topology for exploratory visualization," in *Proceedings of the Symposium on Data Visualisation 2003*, ser. VISSYM '03. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2003, pp. 49–58.

[9] C. L. Bajaj, V. Pascucci, and D. R. Schikore, "The contour spectrum," in *Proceedings of the 8th Conference on Visualization '97*, ser. VIS '97. Los Alamitos, CA, USA: IEEE Computer Society Press, 1997.

[10] I. Fujishiro, Y. Takeshima, T. Azuma, and S. Takahashi, "Volume data mining using 3d field topology analysis," *Computer Graphics and Applications, IEEE*, vol. 20, no. 5, pp. 46–51, Sep 2000.

[11] G. H. Weber, G. Scheuermann, and B. Hamann, "Detecting critical regions in scalar fields," in *Proceedings of the Symposium on Data Visualisation 2003*, ser. VISSYM '03. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2003, pp. 85–94.

[12] V. Pascucci, K. Cole-McLaughlin, and G. Scorzelli, "Multi-resolution computation and presentation of contour trees," in *Proc. IASTED Conference on Visualization, Imaging, and Image Processing*, 2004.

[13] H. Carr, J. Snoeyink, and M. van de Panne, "Simplifying flexible isosurfaces using local geometric measures," in *Visualization, 2004. IEEE*, Oct 2004, pp. 497–504.

[14] G. H. Weber, S. E. Dillard, H. Carr, V. Pascucci, and B. Hamann, "Topology-controlled volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 2, Mar. 2007.

[15] P.-T. Bremer, G. Weber, V. Pascucci, M. Day, and J. Bell, "Analyzing and tracking burning structures in lean premixed hydrogen flames," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 2, pp. 1–1, 2010.

[16] P.-T. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. Bell, "Interactive exploration and analysis of large-scale simulations using topology-based data segmentation," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 9, pp. 1307–1324, Sept 2011.

[17] D. Thompson, J. Levine, J. Bennett, P.-T. Bremer, A. Gyulassy, V. Pascucci, and P. Pebay, "Analysis of large-scale scalar data using hixels," in *Large Data Analysis and Visualization (LDAV), 2011 IEEE Symposium on*, Oct 2011, pp. 23–30.

[18] A. G. Landge, V. Pascucci, A. Gyulassy, J. C. Bennett, H. Kolla, J. Chen, and P.-T. Bremer, "In-situ feature extraction of large scale combustion simulations using segmented merge trees," in *Proc. of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '14, Piscataway, NJ, USA, 2014, pp. 1020–1031.

[19] A. Tikhonova, C. Correa, and K.-L. Ma, "Visualization by proxy: A novel framework for deferred interaction with volume data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, no. 6, pp. 1551–1559, Nov 2010.

[20] S. Frey, F. Sadlo, K.-L. Ma, and T. Ertl, "Interactive progressive visualization with space-time error control," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 12, pp. 2397–2406, 2014.

[21] L. Carpenter, "The a -buffer, an antialiased hidden surface method," *SIGGRAPH Comput. Graph.*, vol. 18, no. 3, pp. 103–108, Jan. 1984.

[22] T. Biedert and C. Garth, "Contour Tree Depth Images For Large Data Visualization," in *Eurographics Symposium on Parallel Graphics and Visualization*. The Eurographics Association, 2015.

[23] H. Kaiser, T. Heller, B. Adelstein-Lelbach, A. Serio, and D. Fey, "Hpx: A task based programming model in a global address space," in *Proceedings of the 8th International Conference on Partitioned Global Address Space Programming Models*, ser. PGAS '14. New York, NY, USA: ACM, 2014, pp. 6:1–6:11.

[24] J. C. Yang, J. Hensley, H. Grün, and N. Thibieroz, "Real-time concurrent linked list construction on the gpu," in *Proceedings of the 21st Eurographics Conference on Rendering*, ser. EGSR'10. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2010.

[25] M. Maule, J. Comba, R. Torchelsen, and R. Bastos, "Memory-efficient order-independent transparency with dynamic fragment buffer," in *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*, Aug 2012, pp. 134–141.

[26] A. Pranckevičius. (2010, Mar.) Compact normal storage for small g-buffers. [Online]. Available: http://aras-p.info/texts/CompactNormalStorage.html

[27] M. Hummel, C. Garth, B. Hamann, H. Hagen, and K. I. Joy, "Iris: Illustrative rendering for integral surfaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1319–1328, Nov. 2010.

[28] D. Morozov and G. Weber, "Distributed contour trees," in *Topological Methods in Data Analysis and Visualization III*, 2014, pp. 89–102.

Usage Modeling of Commuters on Basis of Geographical Data for Vehicle Engineering

Christine Biedinger
Fraunhofer ITWM
Kaiserslautern, Germany

Sascha Feth
Fraunhofer ITWM
Kaiserslautern, Germany

***Abstract*—During their lifetimes vehicles suffer from various kinds of loads. These strongly depend on the quality and condition of traveled roads as well as the driving behavior of the motorist. Thus, the determination of representative measures for the usage of motorcars in predefined geographical areas is desired in vehicle development. In order to provide these, models for the typical use of private passenger cars were considered. Essentially basic patterns for commuters consisting of approached target points were described. It is shown how the tours during a week can be simulated including daily rides to and from work as well as less regular trips like going shopping or attending leisure time activities. Additionally, a possibility to handle missing statistics for a country is outlined. Finally, it is presented how the resulting journeys can be used to estimate region-specific operating conditions of vehicles by processing them with VMC® (virtual measurement campaign), a software developed at Fraunhofer ITWM. In some short case scenarios it is also demonstrated that the presented usage simulation based on geo-referenced data is not restricted to passenger cars but can also be applied to commercial vehicles.**

I. Introduction

In an early stage of vehicle development the design of particular components requires specific knowledge on the future operation conditions. During the construction of a drive train for an emerging market for instance, information on the hilliness of the expected site of operation is essential. The challenge here is to obtain data cost-efficiently in the considered area. Instead of conducting an expensive and time-consuming measurement campaign, we want to gain knowledge virtually. This additionally allows to easily transfer the applied methods to other regions later on. For our purpose only the regularly used parts of the road network are relevant. We extract these by simulating the journeys of several thousands typical customers. In the following we introduce a model for commuters and present mathematical methods to create representative routes. In that process, the simulated target points shall correctly reproduce available statistics for the considered region, for instance those on driven distances between home and work locations. We concentrate on private customers due to their homogeneity as a larger group, however, the models and algorithms described can be adapted to further types of vehicle usage. Additionally, the methods are not restricted to passenger cars and commuters.

In the simulation process we use the simplifying assumption that employees go to work five times a week. In a general situation, like the one in our example of constructing a drive train, we are not interested in the exact time or day in work week. In order to ease the notation we just number the routes from one to five. We also exclude extraordinary trips like going on vacation or weekend forays. Of course, it is possible to consider these, too. However, we want to keep the model simple and thus only add selected types of private matters in addition to the regular journeys to and from work. In the first step we randomly choose the people's home and target locations. The targets shall be given by their longitudinal and latitudinal coordinates. These are sufficient as an input for VMC® (virtual measurement campaign), a software developed at Fraunhofer ITWM which then applies a routing algorithm to determine the precise roads traveled and which is able to analyze the routes like described in section IV-B.

II. Related work

Analysis and simulation of commuters' daily travels has been subject of research for more than forty years and is still an active field. Thus, there exist plenty of publications dealing with this topic but as far as we know, non of them focuses on the determination of concrete locations with appropriate classification given by their coordinates. They are mainly interested in commuter data on household level as input for traffic flow simulations. In [1] for instance target points are selected in given zones for which population or employment distributions are given. Pairs of home and work places are estimated on basis of the number of people commuting in the same direction between these zones. In a similar way, households are assigned to districts in [2]. The aim here is the generation of synthetic data based on available aggregate data that can be used as an input for microsimulation. In general, these extensive results could enhance our procedure presented in the following, but again the lack of geo-coordinates makes them unserviceable.

The detailed considerations of Beckman, Baggerly and McKay given in [3] are not applicable in our case due to the fact that they do not only use disposable census data, but also a small representative original data sample. The simulated commuters are created by selecting subsamples from this pool. Though, our method shall be applicable generally and such input data is rarely available and impossible to be integrated for every region considered. An analogue model for travel patterns is shown in [4]. In contrast to the ones given in

the next sections, those are based on travel time restrictions and conditional probabilities for further trips whereas our approach concentrates on traveled distances directly and takes the probabilities for different types of trips as a required input. At last the work of Susilo and Kitamura [5] shall be mentioned. It deals with so called action spaces that represent places where people carry out activities. Like in our model, distances are considered here and interesting regions are also bounded by some ratio around the home location. However, all results are predicated on travel dairies of people living in Karlsruhe and Halle. These can not be easily extrapolated to other regions, like Germany as a whole for example, since they lack of representative data for especially rural areas.
Altogether it can be said that there exist lots of promising approaches for the simulation of commuters but most of them do not consider concrete coordinates that can be projected on the map. Additionally, multiple data sources are required that are not freely available. Our method presented in the following needs only a few travel statistics as an input that even do not have to be available for the considered region.

III. Usage modeling for commuters

Before starting with the detailed description about how to model commuters we have to introduce some notation. We will see that we need two types of data in simulation. On one hand side these are the usual ”points of interest”, POIs, directly given by their coordinates. On the other hand there are ”regions of interest”, ROIs, described by their boundaries. In both cases the positions are combined with a classification of the type of usage.

A. Basic modeling of trips

Typical routes of commuters basically consist of tours between home location and work place. Additionally, further different kinds of private trips towards destinations like supermarkets, pharmacies or gyms for example are performed more or less regularly. First of all, the home location as center of attention has to be chosen inside some residential neighborhood in the considered region. It stays the same for all five days and is used as starting and destination point of all trips. The work place is also chosen just once and approached every day. Its classification is more demanding. Places of employment are assumed to be settled at different kinds of sites. On the one hand they are expected to be located at zones labeled as commercial, industrial or retail. Of course, concrete places like supermarkets or pharmacies also make sense. What both have in common is that they need to have a reasonable distance to the home location. ”Reasonable” in this context depends on various factors. Usually, people have a limit in mileage they accept traveling every day. Thus, the distance between home and work place should not be too large. On the other hand, due to economic reasons, people might prefer walking or going by bike if the way is short enough. Additionally, some regional aspects like the size of the city or the employment situation have an influence. Results of polls regarding motor traffic statistics provide information on this and can be included in the simulation. Further details on this are given in section III-E. Supplementary to these daily trips, some journeys for private matters are considered. These include leisure-time pursuits, consultations and weekly shopping for instance. They are expected to take place after work. Though, it depends on the personal attitude if they are executed directly on the way home or if they form an additional tour.
Up to now these were only basic trips every commuter is expected to perform every week. Though, there are groups of people executing further regular tours. Two of them shall also be modeled. We start with parents taking their children to school or kindergarten. Thereby we distinguish between the ones just attending them on the way to work and the ones also collecting them after work. In that process, the restriction that the additional destination should not cause a large detour can be added. However, in order to reproduce the situation that parents do not always get a space at kindergarten that lies on their way we skip this in our general model. Since the cultural background also influences the attitude of parents, this has to be adopted depending on the considered region. In some countries for instance children usually go to school by bus whereas in other states parents prefer bringing them to school themselves. The second group of persons we want to include is the one using the break for private matters. We expect them to perform small tour going to cash points, pharmacies or post boxes for example. The case of people using the car to go having lunch a bit further away from their work place is also included.
For both types of additional trips we use POIs as destinations. Especially in the first case the specific locations of kindergarten and schools are essential.

B. Methods to select target points in POIs

We already shortly introduced the notion of POIs and ROIs in the beginning of this chapter. Points of interest are thereby used in the common sense that they mark important locations. According to their description they can be classified in different groups which are more or less relevant for the modeling of commuters. One category for instance combines different types of shops like supermarkets, malls or stores focusing on special goods like wine or cheese. An other one is used to summarize different kinds of leisure-time activities like gyms or cinemas. In the general simulation described here all POIs within one category are assumed to be of equal importance. Hence, if a new destination has to be selected, first all relevant groups are gathered and the linear distance between each point and the home or work location is computed. Which of the latter two is considered depends on the trip that is simulated in the specific situation and on the available statistics. In [6] for example the distances for private trips are measured from the home location whereas activities during break should be selected depending on the work location. Afterwards all remaining POIs fulfilling the limits for the distances are assumed to be equally important. One of them is selected by chance. Automatically, zones with a higher density

of POIs are chosen more frequently than less covered areas. This is expected to mirror well the situation that large towns are more attractive, containing more shops, work places etc. than small municipalities. In the case of commuters POIs are used as targets for all kinds of private trips no matter if they are performed before work like taking children to school, during break or after work. Additionally, they are can also be used as places of work for parts of the simulated commuters.

C. Methods to select target points in ROIs

The selection of a concrete point for ROIs requires more computational effort. Similar to the case of POIs, first all regions with suitable classification have to be taken. If the home locations has to be determined, this means we concentrate on residential areas. Since we want to chose the starting point of all trips, no distance calculation has to be performed. Instead, one of the districts is randomly selected. Currently, there are no population figures or densities available on basis of single residential neighborhoods. Thus, as a workaround until this problem is solved, the sizes of the areas are taken to estimate the proportion of inhabitants between all regions. Larger neighborhoods are assumed to have larger probabilities to contain home locations of commuters than smaller ones. After choosing a single area, a concrete place inside is picked. Therefor a bounding box is laid around the polygon and coordinates are chosen randomly. If the resulting point lies inside the polygon, it is accepted. Otherwise the process is repeated until a feasible place is found.
In the case that a work place from ROIs has to be determined, the procedure is similar. Again, the relevant regions have to be selected first. Naturally thereto belong commercial, industrial and retail areas. However, further districts are expected to be relevant, too. These include zones belonging to large buildings like hospitals or universities that are excluded from all other regions. They are marked as POIs and are thus already included in the simulation, though they get the same importance like small shops for instance. Since more work places are expected to be settled there, a new category of ROIs called ”work” is introduced including all larger buildings that should get more impact. Of course, their areas are relatively small compared to the ”real” regions but this mirrors the reality quite well. What kinds of buildings should be included is a matter of taste. Since we assume shopping malls or big supermarkets with their large parking lot to be completely included in commercial areas, we do not add them to the new category. The selection of a work place then once again starts with taking the desired categories of ROIs. Next, the ones in suitable distance to the home location are extracted. Therefor traffic statistics giving bounds on the accepted displacement are reapplied, see section III-E. Afterwards, coordinates in one of the resulting districts are selected. In doing so, their intersection planes with the circular ring around the domicile represented by the accepted distance limits are respected. Consequently, zones containing larger areas fulfilling the demanded requirements are chosen with higher probability than ROIs with smaller cut surfaces. Figure 1 sketches the locations for multiple draws. The dot in the middle of the draft marks a home in Kaiserslautern. All ROIs, except the residential ones, with a distance between 100m and 1km to the home location are taken to contain acceptable work places in this example. They are delineated as gray regions. A suitable distance in this context means that there exists at least one point inside or on the boundary of the polygon that fulfills the limits. It can be seen that the concept of randomly selecting positions, plotted as stars, works quite well. Larger areas inside the circular ring are chosen frequently whereas districts with small cut are selected rarely.

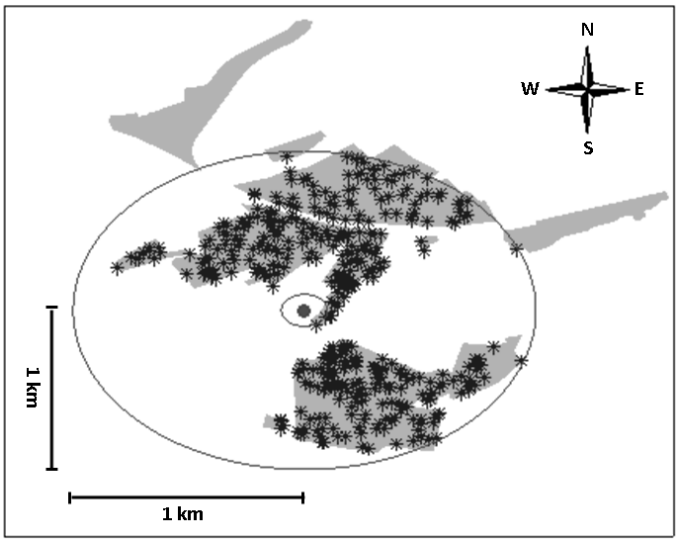


Fig. 1. Distribution of selected locations (*) around a home place (•) in Kaiserslautern. Drawn areas have points in distance of at least 100m and at most 1km from this center and are used as simulation base.

D. Using OpenStreetMap data

The usage model of commuters uses POIs and ROIs as basal data types given by coordinates. They allow the reasonable calculation of distances between locations during simulation. Additionally, the created trips, represented by single points, can be projected to maps. Thereby routing algorithms can be applied in order to determine the precise roads driven. Further evaluations enabled by this geo-reference are discussed in section IV-B. At this point we want to comment on the data that is used as an input. First of all it has to be stated that there are different sources of geo-referenced data available. Beneath various subject to charges. Additionally, their coverage with respect to single countries and specification of information differs strongly.
We only want to use data free of costs which can be processed easily. This reduces the variety drastically since a lot of gratuitous providers allow only the download of POIs of particular categories at a single blow which is not sufficient for our usage models and thus wastes of lot of time. More important is the fact that ROIs are rarely available. In most cases POIs are available in order to download as add-ons to navigation devices but land use information is assumed to be already existent in the ordinary map data. One provider of free of charge

POI and ROI data which can be handled comfortably is the OpenStreetMap (OSM) project [7]. It shows a good coverage for most regions of the world and is continuously extended and improved by over two million people contributing their collected data [8]. However, one always has to keep in mind that the quality of data strongly depends on their participation and is usually less good in isolated areas. Nevertheless, the simulations given in section IV are performed with this source. We extract the regions of interest from "ways" describing areas by their boundaries and a "landuse" specification. POIs can directly be derived from "nodes" concentrating on useful classifications.

E. Inclusion of poll data

In the selection of target points like the place of work or destinations for private matters suitable distances were needed. These can be be chosen reasonably on basis of available statistical data. Commuter surveys like [9] for example summarize the length of reported journeys in distance classes of different size. Similar statistics are also available for other countries or further purposes of trips like [10] or [6]. The given frequency distributions can be used to obtain limits on the acceptable distance for the currently simulated commuter. Though, two things have to be respected. First, the single distance classes have to be bounded. Otherwise the displacement between home and work for instance would have an open range which is not realistic. Hence, the respecting classes, which are commonly the last ones including all reported values larger than a predefined quantity, have to be cut wisely. Sometimes it might also be reasonable to remove the complete category and recompute all other frequencies for the reduced data base. The second and more involving problem is the one that usually polls ask respondents to report their driven distances. Though, the computation of target points is done with respect to linear distances. Since ways on roads are expected to be longer than the beeline between starting and arrival point, the routed distances after the usage simulation will be larger than requested. Therefore the classes in the summary statistics have to be recalculated. For that purpose circuity factors described in [11] can be used. These mirror the proportion between both considered lengths in elected regions. In order to obtain reasonable values also for countries that are not examined in detail, the also available standard deviations can be integrated in the transformation. The same procedure also allows to estimate the distances traveled for countries where no surveys are available. For them statistics from a different country, in which commuters are expected to behave similar, can be applied with adapted classes.
Surely, the distance limits simulated for a single commuter will not always be met exactly. This can result either from a rather general circuity factor which is not respecting regional special characteristics or from the routing to selected unreachable target points, like POIs inside buildings. However, in average the predefined distribution is expected to be reproduced.

TABLE I
DISTRIBUTION OF DIFFERENT TYPES OF COMMUTERS CHOSEN IN SIMULATION

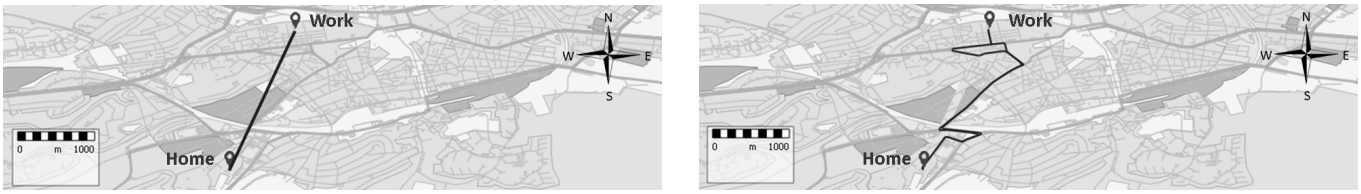
Type of commuter	Classification of additional stops	Fraction of simulated commuters
Only go to work and back home	No additional stopovers	70%
Attendance of children	School or kindergarten with distance smaller than or equal to 30km around home	20%, equal probability for just taking and additionally collecting children
Private matters in break	POIs of different kinds in radius of 10km around workplace	10%

IV. SIMULATION AND ANALYSIS OF ROUTES

After the description of the usage model we now simulate the trips of some commuters and use them as an input for VMC. Therein they get routed and different types of analysis are performed.

A. Choice of parameters and distances

Before starting the simulation various settings have to be made. Primarily, we use the already mentioned commuter survey [9] to estimate the distribution of distances between home and work locations. Additionally, we also use the results of [6] to determine reasonable target points for private activities after work. For both of these the domiciles are taken as reference locations. We do not have any statistics available when and how often private trips are performed during a workweek. Hence, we constitute that they occur equally distributed over the complete time slot and select the specific days randomly. Furthermore, we do not know if they are conducted rather directly after work or rather after an additional stop at home. There are only mean values regarding the total number of trips over a day available, but due to the simplification of our model we are not able to derive reliable conclusions from them. Thus, we assume that people slightly prefer a fewer number of trips and set the proportion between activities on the way home and leaving again after returning to 52.5% and 47.5%. Even if it is possible, we do not consider destinations in ROIs for private trips. Surely, we loose some visits at family members of friends by doing this but accept it as a negligible side effect of the model simplification. Indeed, the inclusion of these trips is possible yet raises the computational effort.
The last parameters we have to fix are related to extra tours. Once again we neither have any information regarding the accepted distance for taking children to kindergarten or school nor for activities in break available. So we select maximal values of 30km and 10km as educated guesses. Table I summarizes the additional assumptions on the distribution of the different types of commuters. Supplementary, we choose to select the workplaces as POIs and in ROIs in equal parts.



(a) After simulation (b) After routing

Fig. 2. Simulation results for a commuter only going to work.

TABLE II
SPLIT OF ALL ROUTES ON FIVE DAYS ACCORDING TO ROAD TYPE FOR A COMMUTER LIVING AND WORKING IN KAISERSLAUTERN

Motorway	Rural A	Rural B+C	Urban A	Urban B+C	Other
0%	3.1%	6.6%	21.6%	66.3%	2.3%

TABLE III
SPLIT OF ALL ROUTES ON FIVE DAYS ACCORDING TO SLOPE CLASSES FOR A COMMUTER LIVING AND WORKING IN KAISERSLAUTERN

Slope class	0-3%	3-6%	6-9%	9-12%	12-15%	>15%
Proportions	83.2%	14.5%	1.8%	0.2%	0.3%	0.0%

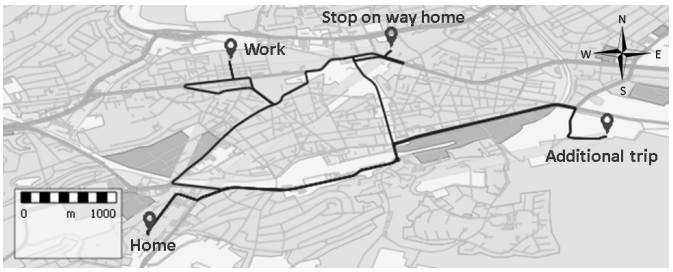


Fig. 3. Simulation of a commuter stopping after work and going out again.

B. Analyzing created routes with VMC

The result of simulating commuters consists of five files for each. They contain the route of one person per day represented by the coordinates of the starting and target points. These files can then be further processed with VMC. In a first step, the software projects the given locations to the nearest way and applies a routing algorithm to determine the traveled roads. Figure 2 shows the trips on one day for a commuter living and working in Kaiserslautern who only drives to work and back home. On the left hand side (a) the basic coordinates are given. The right hand side (b) displays the tour on the road network. Figure 3 illustrates how more complex days can look like. In this case the commuter stops on his way home and performs a further trip after having returned there.

When analyzing the routes, the resulting paths are subdivided in smaller parts afterwards. This procedure creates single segments split by different criteria. One of them is the classification of highway type. Here motorways, major and minor roads for instance are separated. A distinction between urban and rural roads is also possible. Topographical factors that can be evaluated include altitude or slope classes as well as hilliness or curviness investigations. Table II for instance summarizes the distribution of different road types traveled in work week. The allocation is chosen according to [12]. Obviously, the trips of this specific commuter only traveling inside the city when going to work and just leaving town for some leisure time activity are reflected well. Comparing table II and figure 3 it can be seen that for instance the lack of motorway contingents correctly mirrors the tours of this driver. They are shorter when passing through the city center instead of taking a ring road, an urban freeway is inexistent in Kaiserslautern. Of course, a relevant percentage of motorway kilometers is expected when simulating more days for the same person. Additionally to the already discussed factor, a possible classification due to slopes is given in table III. An example for the split regarding altitudes is skipped here since the considered commuter stays in Kaiserslautern only and thus the computed heights show few variations.

Supplementary to these evaluations on basis of topographical data only, it is also possible to start virtual measurements in which the driving style of the car owner and the specific vehicle model can be incorporated. Thereby speed profiles including acceleration and breaking events can be created. For that purpose traffic densities or average speeds for the roads traveled are required in addition to these two inputs. Furthermore, some random process deciding about the stopping at traffic lights and cross-walks as well as about giving way has to be adapted to the considered region just as to the habit of the motorist simulated. Figure 4 shows the speed profile for the trip from work back home shown above of the commuter in Kaiserslautern. The velocity in kilometers per hour is plotted against the distance from work given in kilometers. In order to create this we assumed the driver to behave ordinary while driving a standard passenger car. It can be seen that in some cases he drives a bit too quick but this does not exceed the speed limits too exorbitant. The profiles simulated over the whole week can then for instance be used to estimate damage values which can be compared to those obtained for different cars or different driving styles. VMC for example also allows the simulation of a rather aggressive behavior where speed limits are not respected that strictly. Details on the methodology of virtual measurements are given in [13].

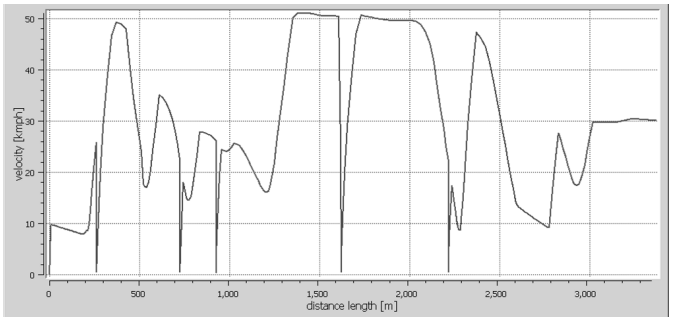


Fig. 4. Velocity in kilometers per hour simulated for the way home of the considered commuter.

V. Further usage models

In the usage model presented up to now we concentrated on regular trips of commuters driving passenger cars. Of course the introduced concepts are generally neither restricted to these motorists nor to that vehicle variant. An alternative field of application for passenger cars as well as smaller vans is the taxi business. In that case the daily mileage is larger and more trips with multifaceted destinations are performed. Typical situations include collecting people on train stations or airports as well we taking them there. A further common trip purpose is to see someone home after going clubbing. In this case the usage modeling is more complex and requires a grater extent of data or expert knowledge to adapt all parameters to the considered region. Additionally, a simulation of seven days per vehicle is necessary. Though, basically the methods described on how to select precise coordinates in POIs and ROIs can be employed. The same holds in the event that the use of light-duty commercial vehicles shall be modeled for instance. In that process two fundamental driving patterns can be determined which differ in the order of destinations. "Tours" typically occur in the case of delivery services starting from a distribution center, supplying all customers is a reasonable way and returning to the stock. In contrast to this, craftsmen for instance often visit only one client before returning to their office again. These patterns characteristically look like stars with different numbers of points with the business address in the midst. Again, journeys are given by single coordinates and are routed and evaluated with VMC.

VI. Summary and outlook

In this paper we described how the usage of vehicles can be modeled and how concrete routes in predefined regions can be simulated on basis of geo-referenced data. For that purpose we concentrated on commuters driving passenger cars. We extended the common concept of points of interest to regions of interest. For both we developed mathematical methods to select concrete destinations in suitable distances to the former visited locations. We also explained how available results of travel surveys can be included. A possibility to estimate reasonable inputs for regions where no data is obtainable from counts at hand was illustrated. A simulation was performed for exemplary parameter values. The resulting journeys represented by coordinates were routed and analyzed with VMC. In that process, evaluations on basis of topographical factors were performed. The possibility to create speed profiles in VMC was also illustrated. Therefor provided default driver and car models were used. It was explicated that the methods presented are not restricted to the modeling of commuters. Thereto models for taxis and salesmen were shortly introduced. However, the algorithms can also be adapted to simulate larger vehicles like trucks. A next step for instance is the creation of routes for the delivery traffic of supermarket chains. One possibility to extend the given methods for private vehicle usage is to include less regular and especially long journeys like going on vacation.

In order to estimate reliable customer specific damage values for desired vehicle parts, the results received from usage modeling can also be combined with real data obtained from measurements. How such campaigns have to be planned is explained in [12] for commercial light-duty vehicles. Of course it is also plausible to do something similar for commuters with passenger cars.

References

[1] R. Lovelace, D. Ballas, and M. Watson, "A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels," *Journal of Transport Geography*, vol. 34, pp. 282 – 296, 2014.

[2] M. Birkin and M. Clarke, "Synthesisa synthetic spatial information system for urban and regional analysis: Methods and examples," *Environment and Planning A*, vol. 20, no. 12, pp. 1645–1671, 1988.

[3] R. J. Beckman, K. A. Baggerly, and M. D. McKay, "Creating synthetic baseline populations," *Transportation Research Part A: Policy and Practice*, vol. 30, no. 6, pp. 415 – 429, 1996.

[4] C. Bhat, J. Guo, S. Srinivasan, and A. Sivakumar, "Comprehensive econometric microsimulator for daily activity-travel patterns," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1894, pp. 57–66, 2004.

[5] Y. Susilo and R. Kitamura, "Analysis of day-to-day variability in an individual's action space: Exploration of 6-week mobidrive travel diary data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1902, pp. 124–133, 2005.

[6] "Les déplacements locaux un jour de semaine selon les motifs -hors marche à pied, enquête nationale transports et déplacements (entd) 2008."

[7] "OpenStreetMap," see www.openstreetmap.org.

[8] "OpenStreetMap Wiki," see http://wiki.openstreetmap.org/wiki/Join_the_community, retrieved December 1st, 2015, last modified on September 3rd, 2015.

[9] "Les déplacements du domicile vers un lieu de travail fixe selon la zeat du domicile, enquête nationale transports et déplacements (entd) 2008," 2010.

[10] "Mobilität in deutschland 2008 (mid 2008), tabellenband," 2010.

[11] R. H. Ballou, H. Rahardja, and N. Sakai, "Selected country circuity factors for road travel distance estimation," *Transportation Research Part A*, vol. 36, pp. 843–848, 2002.

[12] T. Weyh, M. Speckert, A. Opalinski, and M. Wagner, "Planning a measurement campaign in eastern europe using fraunhofer vmc®("virtual measurement campaign")," in *13. International Conference Commercial Vehicles 2015 Truck, Bus, Van, Trailer*, 2015, pp. 207–217.

[13] M. Speckert, K. Dreßler, N. Ruf, T. Halfmann, and S. Polanski, "The virtual measurement campaign (vmc) – a methodology for geo-referenced description and evaluation of environmental conditions for vehicle loads and energy efficiency," in *3rd Commercial Vehicle Technology Symposium*, 2014, pp. 88–98.

Non-equilibrium electron-phonon coupling in dielectrics excited by ultrashort laser pulses

Nils Brouwer* and Bärbel Rethfeld
Department of Physics and OPTIMAS Research Center
TU Kaiserslautern
67663 Kaiserslautern
*Email: brouwer@physik.uni-kl.de

***Abstract*—Ultrashort laser pulses of high intensity are of increasing importance in material processing and fundamental research. To check the validity of equilibrium models, we calculate the non-equilibrium dynamics of electrons and phonons in laser-excited dielectrics using Boltzmann collision integrals.**

I. Physics of laser-excited dielectrics

With ever increasing availability of intense tabletop laser sources with pulse lengths of 10 - 100 femtoseconds (10^{-15} s), these lasers have become a standard tool in basic physics research. The high intensity of these very short pulses however also shows its potential in precise material processing, especially of transparent dielectrics, e. g. silicon dioxide [1], [2], where a high intensity is necessary to make the material absorbing at all. While the basic processes of the laser-excitation of dielectrics are known, their respective influence on material damage are still under debate, in particular on the femtosecond timescale. Since time-resolved measurements on this timescale are difficult and can access only integrated quantities, e. g. reflectivity, theory and numerical modeling are necessary for the correct interpretation of these experiments [3].

Since this work is targeted at a general audience, a few concepts of solid state physics shall be introduced. In solids, atoms are arranged in regular 3D-lattices. The lattice can be distorted leading to lattice vibrations, which is one, often the main, contribution to the temperature of the solid. These lattice vibrations are commonly described by the amplitude of normal modes. Due to quantum statistics these are quantized and the corresponding quasi-particles are called phonons. In equilibrium, these phonon modes are populated according to a Bose-Einstein distribution, because phonons are bosonic particles.

Electrons in solids are not restricted to discrete energies, like in atoms and molecules, but they occupy states in so-called energy bands. In contrast to a free electron, these energy bands have gaps, that is, there are energies that are forbidden. Unlike phonons, electrons are fermions and not bosons. This entails that they obey Pauli's principle, that means that each electron state can be occupied by only one electron, while there is no such limit for phonons. Therefore fermions also populate states according to a different distribution function, called Fermi distribution.

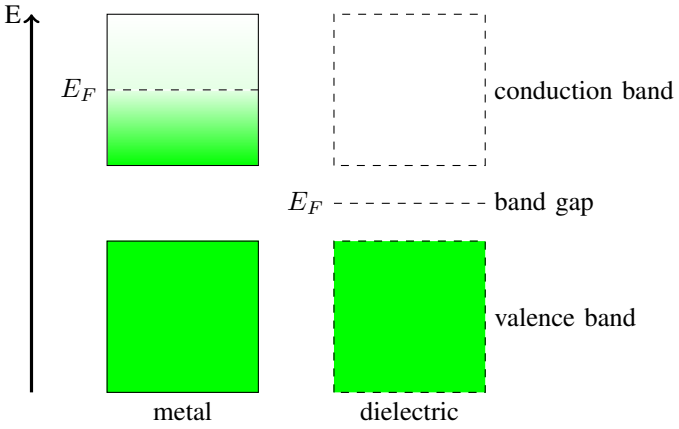


Fig. 1. Schematic: Electronic bands in metals and dielectrics. E_F denotes the Fermi energy, which is the highest energy occupied by electrons at 0 K. The shading indicates the electron distribution at a finite temperature ($T > 0$ K).

If such an equilibrium distribution is disturbed, e. g. by a laser pulse, it takes time for the system to thermalize, that is to reach a new equilibrium distribution. Typical thermalization times for electrons are on the femtosecond timescale, while phonon thermalization times are in the picosecond (10^{-12} s) regime.

The main difference between metals on the one hand and dielectrics and semiconductors on the other hand is that in metals, the highest energy band containing electrons at a temperature of 0 K is only half filled, allowing for conductivity and therefore is called conduction band. In dielectrics and semiconductors the highest occupied band is completely filled. Since, due to Pauli's principle, every quantum state can be occupied only by one electron, conduction is not possible. Energy bands that are completely filled at a temperature of 0 K are called valence bands (see figure 1). In semiconductors, the band gap between the highest occupied (valence) band and lowest unoccupied (conduction) band is small enough for a significant fraction of electrons to be excited to the higher energy band by thermal energy alone at room temperature. Therefore semiconductors also conduct current to some degree, while dielectrics, having a larger band gap, do not.

The absorption of visible light is also different in metals, semiconductors and dielectrics, respectively. Visible light is

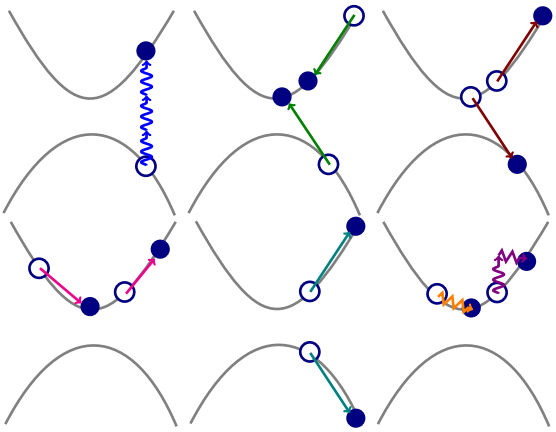


Fig. 2. Microscopic collision processes in laser excited dielectrics. From top right to bottom left: multi-photon ionization, impact ionization, Auger recombination, electron-electron collisions, interband thermalization, electron-phonon collisions and inverse Bremsstrahlung.

in general absorbed by electrons, and similar to conductivity, free states within the energy bands are necessary. Light is absorbed in the form of photons, which are particles with a fixed energy depending only on the wavelength. Metals, again, can very easily absorb light. For semiconductors, the band gap can be often bridged by the energy provided by one photon. In dielectrics the band gap is larger and therefore this energy is not sufficient and under usual conditions, many dielectrics are transparent (opaque dielectrics have other channels of light absorption, that are not considered in this work). However, if the light intensity is strong enough, several photons can be absorbed at once (a process called multi-photon ionization) and thus excite an electron to the conduction band. These electrons then can, as in a metal, absorb further photons gaining even more energy (by inverse Bremsstrahlung). Once sufficient energy is absorbed, they can collide with valence band electrons transferring enough energy to them to let them overcome the band gap as well (performing impact ionization). The newly excited electrons can then also absorb single photons, which can result in an avalanche of electron excitation [4], [5], [6]. Further processes include electron-electron inter- and intraband collisions that thermalize the system, Auger recombination that counteracts impact ionization and electron-phonon collisions that transfer energy to the lattice, increasing its temperature, which may finally lead to melting and therefore material damage. An overview over the microscopic collision processes in laser-excited dielectrics can be found in figure 2.

The generation of conduction band electrons (often called "free electrons") is often described by a simple rate equation [4], [5], [6]:

$$\frac{\mathrm{d}n_{\mathrm{free}}}{\mathrm{d}t} = \dot{n}_{\mathrm{mpi}}\left(E_L\right) + \alpha\left(E_L\right) n_{\mathrm{free}} \quad , \tag{1}$$

where n_{free} is the conduction band electron density, $\dot{n}_{\mathrm{mpi}}$ is multi-photon ionization rate, α is the avalanche coefficient and E_L is the laser field. To determine whether damage has occurred, a critical conduction band electron density criterion is introduced. It is based on the strong increase in light absorption when the electron density dependent plasma frequency matches the laser frequency. While this is an adequate model on the nanosecond timescale, its predictive powers on the femtosecond timescale are questionable at best [7], [8], [9].

To go beyond a simple critical electron density criterion, on can follow the evolution of electron and phonon (lattice) temperature using a two temperature model [10]. Here, damage can be accessed by observing when the phonon temperature reaches the melting point. The evolution of the electron and phonon temperatures in this model is described by the following system of equations:

$$\frac{\mathrm{d}U_{\mathrm{e}}}{\mathrm{d}t} = \nabla\cdot\left(\kappa_{\mathrm{e}}\nabla T_{\mathrm{e}}\right) - g\left(T_{\mathrm{e}} - T_{\mathrm{ph}}\right) + S_{\mathrm{Laser}} \quad , \tag{2}$$

$$\frac{\mathrm{d}U_{\mathrm{ph}}}{\mathrm{d}t} = \nabla\cdot\left(\kappa_{\mathrm{ph}}\nabla T_{\mathrm{ph}}\right) + g\left(T_{\mathrm{e}} - T_{\mathrm{ph}}\right) \quad , \tag{3}$$

where U_{e} and U_{ph} denote the internal energy, T_{e} and T_{ph} the temperature and κ_{e} and κ_{ph} the heat conduction coefficients of electrons and phonons, respectively. g is the electron-phonon coupling parameter and S_{Laser} is the laser energy absorption rate. This model describes the spatial energy transport as well as the coupling between electrons and phonons and is used down to the picosecond timescale. In order to work for semiconductors and dielectrics, this model was expanded to a density dependent two temperature model, using an equation similar to (1) [11], [12]. However, it assumes electrons and phonons to have a well defined temperature, that is their occupation of the states in the energy bands can be described by an equilibrium distribution function, a Fermi or Bose-Einstein distribution respectively. As the electron-electron interaction takes place on a femtosecond timescale, the applicability of (2) and (3) on an ultrashort timescale is also questionable.

In order to test the validity of the two temperature model on a femtosecond timescale, it is necessary to investigate the non-equilibrium dynamics and model the evolution of the electron and phonon distribution functions directly. One approach to achieve this, is by using Boltzmann collision integrals, presented in the next section.

II. Modeling

In order to model the non-equilibrium dynamics of electrons and phonons in laser excited dielectrics we calculate the distribution function using Boltzmann collision integrals. This allows us to trace the occupation of each energy, regardless whether the distribution resembles a Fermi distribution or not.

The evolution of the electron distribution function in the valence and conduction band is calculated with a homogeneous, isotropic approximation of the Boltzmann equation, which yields the following integro-differential equation for each band:

$$\begin{aligned}\frac{\partial f}{\partial t} = &\left.\frac{\partial f}{\partial t}\right|_{\text{multiph}} + \left.\frac{\partial f}{\partial t}\right|_{\text{imp. ion.}} + \left.\frac{\partial f}{\partial t}\right|_{\text{Auger}} \\ &+ \left.\frac{\partial f}{\partial t}\right|_{\text{e-e}} + \left.\frac{\partial f}{\partial t}\right|_{\text{e-e-intertherm}} + \left.\frac{\partial f}{\partial t}\right|_{\text{e-ph}} + \left.\frac{\partial f}{\partial t}\right|_{\text{e-ph-pt}}\end{aligned} \tag{4}$$

The $\frac{\partial f}{\partial t}$ terms denote collision integrals modeling the microscopic collision processes shown in figure 2. As an example the term for electron-phonon collisions reads:

$$\begin{aligned}\left.\frac{\partial f}{\partial t}\right|_{\text{e-ph}}(E) = \frac{1}{2\pi\hbar^3}\frac{m^*}{k(E)}\sum_\beta\Big\{ \\ \int_{q\in Q_\beta^+}\left[f(E+\hbar\omega_\beta(q))(1-f(E))(s_\beta(q)+1)\right. \\ \left.-f(E)(1-f(E+\hbar\omega_\beta(q)))s_\beta(q)\right]\ |\mathcal{K}(\beta,q)|^2\ q\ \mathrm{d}q \\ +\int_{q\in Q_\beta^-}|\mathcal{K}(\beta,q)|^2\ q\left[f(E-\hbar\omega_\beta(q))(1-f(E))s_\beta(q)\right. \\ \left.-f(E)(1-f(E-\hbar\omega_\beta(q)))(s_\beta(q)+1)\right]\ \mathrm{d}q\Big\}\ , \end{aligned} \quad (5)$$

where $f(E)$ denotes the electron distribution function at energy E and $s_\beta(q)$ denotes the phonon distribution function of mode β at wavenumber q. $\mathcal{K}(\beta,q)$ is electron-phonon matrix element, $\omega_\beta(q)$ is the phonon frequency, $\hbar$ is the reduced Planck's constant, m^* is the effective electron mass and $k(E)$ is the electron momentum. $Q_\beta^\pm$ are integration regions ensuring momentum and energy conservation. Equations that are similar to equation (4) represent the change of the energy distribution of the phonon modes, resulting in a total of five coupled integro-differential equations. Further details on this model can be found in [7], [13], [14]. To solve this equation numerically, the energy axis is discretized, which transforms the set of equations into a set of coupled non-linear ordinary differential equations, which are solved with an adaptive time-step method. The computationally expensive calculation of the collision integrals can be done for each discretization point in parallel, with data exchange between processes being necessary only once every time-step.

III. Model parameters

In the next section we will show results for a model-dielectric with parameters of silicon dioxide, taken from [7], [13], [14]. The following material parameters are used in the model:

- The (DOS-)effective valence and conduction band electron mass m^* ensures that the correct amount of electrons is modeled in our parabolic band approximation. It is set to the free electron mass ($m_e = 9.109\cdot10^{-31}$ kg) for conduction band electrons and to 3.52 times the free electron mass for valence band electrons.
- The band gap, which determines the energy separation of conduction and valence band and influences the speed of Auger recombination and impact ionization, is set to $\Delta = 9$ eV (1 eV $= 1.602\cdot10^{-19}$ J).
- The mass density,which enters the electron-phonon matrix element $\mathcal{K}(\beta,q)$, is set to $\rho = 2650$ kg/m^3.
- For the acoustic phonon mode, the sound velocity, which enters the phonon frequency ω_β, is set to 5935 m/s and the deformation potential, which enters the electron-phonon matrix element $\mathcal{K}(\beta,q)$, is set to $C_D = 6$ eV.

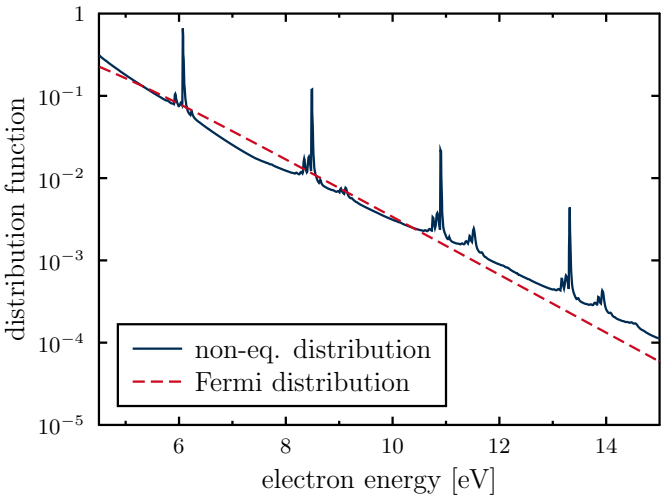


Fig. 3. Non-equilibrium distribution function of conduction band electrons at the peak of a 100 fs long, 400 nm wavelength laser pulse with a maximum intensity of $2\cdot10^{18}$ W/m^2, equivalent thermalized distribution added for comparison. Further details on the peak structure can be found in [14].

- For the two optical phonon modes, the phonon frequency is set to $\omega_\beta = 0.063$ eV/$\hbar$ and $\omega_\beta = 0.153$ eV/$\hbar$, respectively.
- The relative dielectric constant, which determines the optical parameters, is set to $\varepsilon_r = 2.161$ at 400 nm wavelength.

IV. Results

The main output of our calculations are the transient distribution functions of electrons and phonons.

As can be seen in figure 3, our calculations result in distribution functions, that are not Fermi-like. Instead, they show sharp peaks at certain electron energies [14]. Therefore, temperatures are not strictly well defined, as was to be expected on the ultrashort timescale considered here. This is of cause already an indicator that models based on equations (2) and (3) are not valid on this timescale, but the more interesting question is: How far off are they?

One of the main mechanism for material damage is melting, which requires the energy to be transferred from the electron system to the lattice. We can extract this energy transfer rate from the distribution functions of electrons and phonons with the integral:

$$\left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_{\text{e-ph}} = \int_{\mathrm{E}} E\ DOS(E)\ \left(\frac{\partial f(E)}{\partial t}\right)_{\text{e-ph}}\ \mathrm{d}E\ , \quad (6)$$

where $\left(\frac{\partial f(E)}{\partial t}\right)_{\text{e-ph}}$ is given by equation (5) and $DOS(E)$ is the density of states.

In figure 4, we compare the energy transfer rate from the conduction band electrons to the dominant, second optical phonon mode. The not-thermalized values were obtained from the distributions functions during a 100 fs long, 400 nm wavelength laser pulse with a peak intensity of $2\cdot10^{18}$ W/m^2

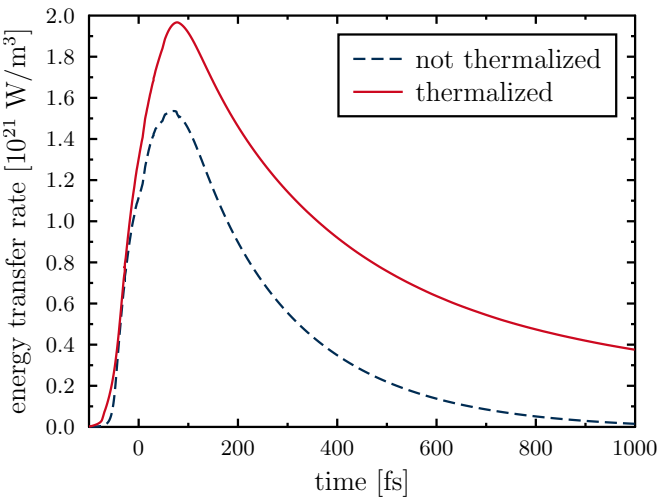


Fig. 4. Energy transfer from the conduction band to the dominant, second optical phonon mode during and after irradiation with 100 fs long, 400 nm wavelength laser pulse with a maximum intensity of $2 \cdot 10^{18}$ W/m^2 centered at 0 fs, shown for the real distribution and a thermalized equivalent (see text).

centered at 0 fs. They are compared with values of a thermalized system, where every distribution function at each instant of time was replaced with an equilibrium distribution, conserving the respective particle and energy densities. It is clearly visible that in the thermalized case, the energy is transferred to the lattice much faster since the plotted energy transfer rate is larger all the time.

Therefore, an electron-phonon coupling parameter based only on temperatures and electron densities typically used in the two temperature model is not applicable on this timescale.

V. Summary and Conclusions

We have given a short overview on the basics of laser-material interaction, including the different light absorption behavior of metals, semiconductors and dielectrics. Focusing on dielectrics, we presented the basic microscopic processes involved in the absorption of intense laser light by the electrons and how this energy is transferred to the phonon system leading to the heating of the lattice and possibly material damage. We have presented models frequently used to describe laser-excited solids.

Considering a dielectric irradiated by an intense, ultrashort laser pulse, we have traced the non-equilibrium distribution functions of electrons and phonons by calculating Boltzmann collision integrals. We have shown that the resulting distribution function differ significantly from equilibrium distributions.

We have analyzed the energy transfer from the electron system to the phonon system. Comparing results for thermalized and not-thermalized distributions, we found that using equilibrium distributions substantially overestimates the energy transfer rate, clearly showing that temperature-based models are not applicable on the femtosecond timescale.

Acknowledgment

The authors would like to thank the Allianz für Hochleistungsrechnen Rheinland-Pfalz for providing computing resources through project LAINEL on the Elwetritsch high performance computing cluster.

Financial support from the Deutsche Forschungsgemeinschaft through the Heisenberg program (grant No. RE 1141/15-1) is gratefully acknowledged.

References

[1] K. Mishchik, G. Cheng, G. Huo, I. M. Burakov, C. Mauclair, A. Mermillod-Blondin, A. Rosenfeld, Y. Ouerdane, A. Boukenter, O. Parriaux, and R. Stoian, "Nanosize structural modifications with polarization functions in ultrafast laser irradiated bulk fused silica." *Optics Express*, vol. 18, pp. 24 809–24 824, 2010.

[2] M. Rohloff, S. K. Das, S. Hohm, R. Grunwald, A. Rosenfeld, J. Kruger, and J. Bonse, "Formation of laser-induced periodic surface structures on fused silica upon multiple cross-polarized double-femtosecond-laser-pulse irradiation sequences," *Journal of Applied Physics*, vol. 110, p. 014910, 2011.

[3] B. Rethfeld, A. Rämer, N. Brouwer, N. Medvedev, and O. Osmani, "Electron dynamics and energy dissipation in highly excited dielectrics," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 327, pp. 78 – 88, 2014.

[4] N. Bloembergen, "Laser-induced electric breakdown in solids," *IEEE Journal of Quantum Electronics*, vol. 10, pp. 375–386, 1974.

[5] D. Arnold and E. Cartier, "Theory of laser-induced free-electron heating and impact ionization in wide-band-gap solids," *Physical Review B*, vol. 46, pp. 102–115, 1992.

[6] B. C. Stuart, M. D. Feit, S. Herman, A. M. Rubenchik, B. Shore, and M. Perry, "Nanosecond-to-femtosecond laser-induced breakdown in dielectrics," *Physical Review B*, vol. 53, pp. 1749–1761, 1996.

[7] A. Kaiser, B. Rethfeld, M. Vicanek, and G. Simon, "Microscopic processes in dielectrics under irradiation by subpicosecond laser pulses," *Physical Review B*, vol. 61, pp. 11 437–11 450, 2000.

[8] F. Quéré, S. Guizard, and P. Martin, "Time-resolved study of laser-induced breakdown in dielectrics," *Europhysics Letters*, vol. 56, pp. 138–144, 2001.

[9] B. Rethfeld, "Unified Model for the Free-Electron Avalanche in Laser-Irradiated Dielectrics," *Physical Review Letters*, vol. 92, no. 18, p. 187401, may 2004.

[10] S. I. Anisimov, B. L. Kapeliovich, and T. L. Perelman, "Electron emission from metal surfaces exposed to ultrashort laser pulses," *Zhurnal Eksperimentalnoi I Teoreticheskoi Fiziki*, vol. 66, pp. 776–781, 1974.

[11] H. M. van Driel, "Kinetics of high-density plasmas generated in Si by 1.06- and 0.53- μm picosecond laser pulses," *Physical Review B*, vol. 35, pp. 8166–8176, 1987.

[12] A. Rämer, O. Osmani, and B. Rethfeld, "Laser damage in silicon: Energy absorption, relaxation, and transport," *Journal of Applied Physics*, vol. 116, p. 53508, 2014.

[13] B. Rethfeld, H. Krutsch, and D. H. H. Hoffmann, "Tracing Laser-Induced Dielectric Breakdown in Solids," *Contributions to Plasma Physics*, vol. 50, pp. 16–20, 2010.

[14] N. Brouwer and B. Rethfeld, "Excitation and relaxation dynamics in dielectrics irradiated by an intense ultrashort laser pulse," *Journal of the Optical Society of America B*, vol. 31, p. C28, 2014.

Advantages and challenges of isogeometric structural analysis of thin-walled structures

Wolfgang Dornisch and Ralf Müller
Lehrstuhl für Technische Mechanik
Technische Universität Kaiserslautern
Gottlieb-Daimler-Str., 67663 Kaiserslautern, Germany
Email: {dornisch,ram}@rhrk.uni-kl.de

***Abstract*—Thin-walled structures occur frequently in nature and technology. Commonly, they are described by surfaces in CAD tools and analyzed by shell elements, using two separate models. Isogeometric analysis merges these two processes into an integrated approach. The use of Non-Uniform Rational B-Spline (NURBS) surfaces as a basis for finite shell elements is discussed in this contribution. The higher continuity of NURBS surfaces allows the computation of exact normal vectors and the automatic detection of kinks. However, these advantages can only be taken advantage of if the interpolation of vectors is adapted to the high continuity. An adapted interpolation method is compared to an interpolation method inherited from standard finite shell elements. A numerical example compares the two approaches.**

I. Introduction

Isogeometric analysis (IGA) is a new approach for Computer-Aided Engineering (CAE) which strives to integrate design and analysis, which have evolved as independent processes. IGA has been proposed in 2005 by Hughes et al. [1]. The design is commonly performed in a Computer-Aided Design (CAD) program, whereby objects are described with the help of Non-Uniform Rational B-Splines (NURBS) surfaces. Thick-walled structures are described by their surfaces and thin-walled structures are described by a reference surface in combination with an attributed thickness. The analysis of structures is usually performed with the help of the Finite-Element Method (FEM). The geometry is usually interpolated by linear Lagrange basis functions. Thus, curved domains cannot be described accurately. This yields an interpolation error of the geometry which propagates to errors in the simulation results. But the more important point is the need for a conversion from the design geometry model to the analysis geometry model, and vice-versa if the structure is to be optimized. The typical time for the creation of an analysis model from the design model is given by 21 % of the overall simulation time in [2].

The basic idea of isogeometric analysis is to use the geometry description of the CAD program also for the analysis using FEM. The major part of research in the field of isogeometric analysis uses NURBS as they are the prevailing standard in CAD software. This work focuses on the analysis of thin-walled structures using shell formulations, which describe the domain by a two-parametrical reference surface in space. If the geometry description of the design is used for the definition of the shell reference surface, then any model conversion can be avoided and a truly isogeometric formulation is obtained. Thus, the use of shell formulations for the isogeometric analysis of thin-walled structures is required. Several isogeometric shell formulations exist in the literature. The first isogeometric Reissner–Mindlin shell formulation proposed in 2010 in [3] carries over concepts known from standard FEM using Lagrange basis functions. This requires the definition of normal vectors at the control points. As is depicted in Fig. 2a, the control points of NURBS surfaces do not lie on the surface. Thus, the definition of normal vectors at control points is not trivial. Furthermore, the high continuity of the NURBS basis functions, which is depicted in Fig. 1, leads to an artificial thinning of the structure if the thickness is interpolated with the help of normal vectors at the control points. This effect is due to the triangle inequality. If standard concepts are used, then the advantages of NURBS described in Sec. III cannot be utilized. More recent isogeometric Reissner–Mindlin shell formulations [4], [5], [6], [7] are better adapted to the advantages and challenges of NURBS, but to the authors' opinion no fully satisfying formulation exists up to now. The properties of NURBS allow the formulation of Kirchhoff–Love shell elements [8], [9] but here additional measures at patch intersections are required [10], [11], [12].

The properties of NURBS surfaces include three peculiarities which constitute a huge obstacle for the industrial application of isogeometric shell elements in industrial applications. A NURBS surface patch is described with the help of a rectangular parameter space which is spanned by a tensor-product of two knot vectors, which is depicted in Fig. 2b. Thus, NURBS surface patches are always of rectangular topology. Furthermore, this precludes local refinement of elements as refinement propagates throughout the whole patch. Complex geometries are described by a multitude of NURBS patches in CAD programs. The individual patches are not connected in the geometry model, they are just placed adjoining. In general these patches are non-conforming to each other along the interface. Thus, methods for the coupling of non-conforming meshes are required for efficient computations. The coupling of deformations along the interface is ususally obtained by introducing additional constraints. Several formulations are given in the literature [13], [14], [15], [16], [17], [18], but to the authors' opinion no fully satisfying solution has been found up to now. The use of trimmed NURBS surface patches

in CAD programs also requires special attention. In this case both the parametric and the physical domain are trimmed by NURBS curves. This yields elements of arbitrary topology and requires special care for integration. Furthermore, the condition of the global system of equation can be problematic. Methods for the treatment of trimmed NURBS surface patches are proposed in [19], [20], [21], [22].

II. Geometry description by NURBS

Details about NURBS can be found in [23]. The most important points are given in the following. Each NURBS surface patch is created by a tensor-product combination of the two knot vectors $\Xi_1^k = \left\{\xi_1^1, \xi_2^1, \cdots, \xi_{n_1+p_1+1}^1\right\}$ and $\Xi_2^k = \left\{\xi_1^2, \xi_2^2, \cdots, \xi_{n_2+p_2+1}^2\right\}$, which yields a rectangular parameter space. The superscript k refers to the considered patch. The univariate B-spline basis functions $N_i^p(\xi)$ depend on the knot vector Ξ and the order p. They are defined recursively starting with

$$N_i^0(\xi) = \begin{cases} 1 & \text{if } \xi_i \leq \xi \leq \xi_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

for $p = 0$. Basis functions $N_i^p(\xi)$ with order $p > 0$ are computed with the Cox-de Boor formula

$$N_i^p(\xi) = \frac{\xi - \xi_i}{\xi_{i+p} - \xi_i} N_i^{p-1}(\xi) + \frac{\xi_{i+p+1} - \xi}{\xi_{i+p+1} - \xi_{i+1}} N_{i+1}^{p-1}(\xi) \,. \tag{2}$$

A set of sample basis functions N_i^p is given in Fig. 1 for an order $p = 3$. The univariate basis functions N_i^p are multiplied under consideration of the weight factor $w_{(ij)}^k$ by

$$N_I\left(\xi^1, \xi^2\right) := \frac{N_i^{p_1^k}\left(\xi^1\right) N_j^{p_2^k}\left(\xi^2\right) w_{(ij)}^k}{W^k\left(\xi^1, \xi^2\right)} \tag{3}$$

with $W^k\left(\xi^1, \xi^2\right) = \sum_{I=1}^{n_{en}^k} N_i^{p_1^k}\left(\xi^1\right) N_j^{p_2^k}\left(\xi^2\right) w_{(ij)}^k$ and a fixed assignment $I = f^e(i, j, k, e)$ between global and element indices in order to obtain the NURBS basis functions N_I. Details about the notation can be found in [24]. Besides the knot vector and the order p_α^k also the control point net is required for the definition of a patch. The physical coordinates of the control points

$$\boldsymbol{B}_{(ij)}^k = \left[x_{(ij)}^k, y_{(ij)}^k, z_{(ij)}^k, w_{(ij)}^k\right]^T = \begin{bmatrix} \boldsymbol{X}_I \\ w_{(ij)}^k \end{bmatrix} \tag{4}$$

are used for the interpolation of physical points

$$\boldsymbol{X}^k\left(\xi^1, \xi^2\right) = \sum_{I=1}^{n_{en}^k} N_I\left(\xi^1, \xi^2\right) \boldsymbol{X}_I \tag{5}$$

on the surface. An exemplary physical domain with control point net is given in Fig. 2a. The associated parametric domain and the basis functions are given in Fig. 2b.

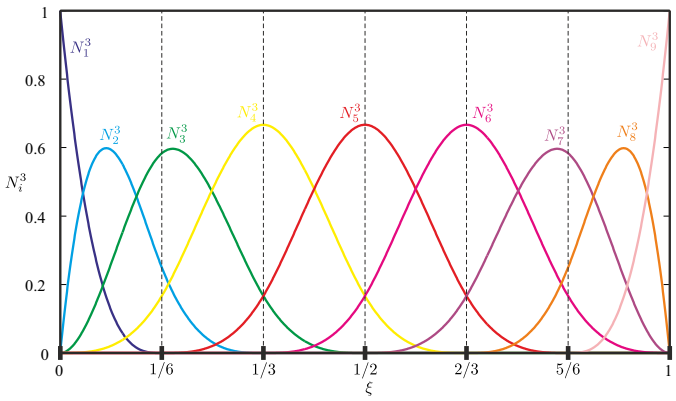


Fig. 1. B-spline basis functions for $p = 3$ and an exemplary knot vector $\Xi = \left\{0, 0, 0, 0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, 1, 1, 1, 1\right\}$.

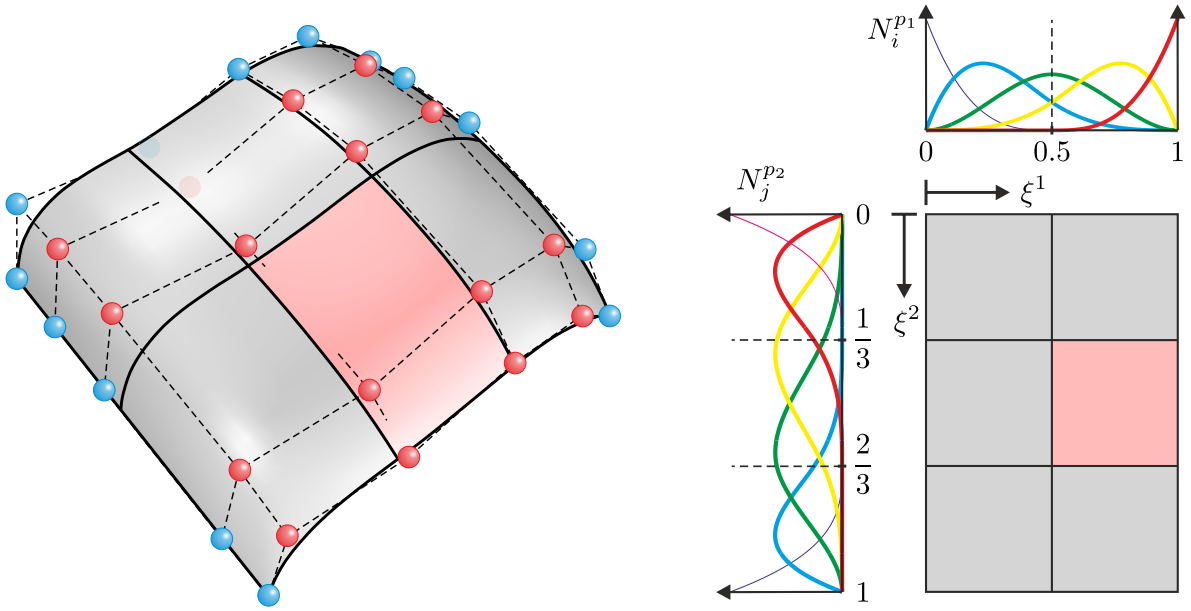
III. Fundamental properties of isogeometric analysis – Challenges and possibilities

The NURBS basis functions defined in Eq. (3) are used within the design geometry model and as isoparametric ansatz functions for the finite element analysis. The tensor product of two knot spans in the parametric space can directly be used as element; see the example given in Fig. 2b. In most cases the initial geometry has to be refined for the analysis. The initial geometry mesh can be refined both by elevating the order or subdividing elements, whereby both actions are trivial and preserve the exact geometry. Two choices for order elevation are possible. While p-refinement maintains the highest possible continuity between elements of the initial order, k-refinement allows the highest possible continuity between elements of the target order to be used. The former choice yields a significantly higher number of control points which are close to the elements they are supported on due to the low continuity. The latter choice yields a lower number of control points which can lie far away from the element due to the high continuity.

Due to the non-interpolatory property of NURBS the control point net and the physical mesh do not coincide in general, see Figs. 2 and 3. Thus, the nodes for the finite element analysis do not lie on the elements. It is not possible to assign one point on the physical mesh to each control point. Thus, the determination of nodal values from surface quantities defined within the elements is non-trivial. Here, commonly an L_2-projection as proposed in [4] is used.

The properties stated above should be taken advantage of in order to obtain an efficient, accurate and robust formulation. Thus, the main possibilities of isogeometric analysis are stated by:

- The geometry is defined exactly for every discretization. Furthermore, the derivatives, curvature and the normal vector are defined exactly in every point on the surface. This precludes the faceting effect, which occurs in Lagrange-based shell elements and may deteriorate the accuracy of stability analysis significantly.



(a) Physical space and control point net: The control point net is denoted by bullets connected by dashed lines. Red bullets denote control points which have influence on the red element. All other control points are plotted blue.

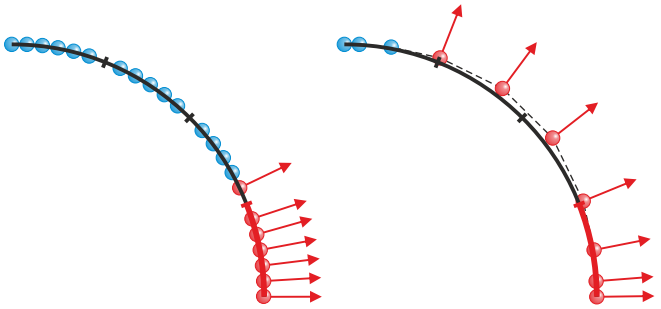
(b) Parametric space and univariate B-spline basis functions for the two parametric directions ξ^α: All basis functions which have influence on the red element are plotted bold.

Fig. 2. Physical space and associated parametric space of a free form surface: The red element in the parametric space is mapped to the physical space with the help of the NURBS basis functions. $\Xi_1 = [0, 0, 0, 0, 0.5, 1, 1, 1, 1]$, $\Xi_2 = [0, 0, 0, 0, \frac{1}{3}, \frac{2}{3}, 1, 1, 1, 1]$.

- Intentional kinks in the isogeometric geometry model can be identified uniquely. This is not possible in standard Lagrange-based shell elements, where intended kinks cannot be distinguished from kinks induced by faceting.
- The continuity of the basis functions rises in step with their order for k-refinement. The higher continuity allows a better approximation of solution fields in case they are smooth. In general, the higher smoothness attained by k-refinement yields considerably lower error levels [25] than standard Lagrange-based finite elements. The higher accuracy of isogeometric analysis allows model sizes to be reduced in comparison to standard Lagrange-based finite element analysis, while maintaining a desired error level.
- Solution spaces with differing desired orders and continuity between elements can be tailored for every solution variable while maintaining the exact geometry. This opens up new possibilities for mixed finite element methods.

The main challenges of NURBS-based isogeometric analysis are:

- Coupling of non-conforming patches without negative influence on the stability and accuracy of computations. A special focus has to be given to the condition number of the stiffness matrix to allow stable computations with large time steps for dynamic analysis.
- The computation of trimmed NURBS patches requires special efforts for the integration of the trimmed elements. Furthermore the condition of the stiffness matrix has to be controlled as trimmed elements may significantly deteriorate the condition number which results in deteriorated accuracy of the solution.
- A special focus has to be set on the interpolation of vectorial quantities. Due to the triangle inequality it is not possible to interpolate vectors with differing orientations exactly; see Fig. 3 for an illustration. Nodal values for



(a) $p = 6$, p-refinement, four non-zero elements, C^1-continuity between elements

(b) $p = 6$, k-refinement, four non-zero elements, C^5-continuity between elements

Fig. 3. Order elevation exemplified with a segment of a circle: The elements of the physical curve are separated by strokes. The control point net is denoted by bullets connected by a dashed line. Red bullets denote control points which have influence on the element marked red. All other control points are plotted blue. Red arrows denote the nodal director vectors which have influence on the red element.

the control points have to be computed using an L_2-projection. Interpolation of directed quantities should be avoided in element formulations. This is especially important in rotation-based shell formulations and requires new concepts for the interpolation of rotations.

IV. NURBS-BASED ISOGEOMETRIC SHELL ANALYSIS

Within this contribution the interpolation of vectorial quantities is chosen to depict the challenges and possibilities of NURBS-based isogeometric analysis. This is shown with the help of a rotation-based isogeometric Reissner–Mindlin shell formulation which is shortly sketched in the following.

A thin-walled body of thickness h in space is described by a reference surface in combination with a director vector $\boldsymbol{D}$. The position of a point is described by

$$\tilde{\boldsymbol{X}}(\theta^i) = \boldsymbol{X}(\theta^\alpha) + \theta^3 \boldsymbol{D} \quad \text{and} \quad \tilde{\boldsymbol{x}}(\theta^i) = \boldsymbol{x}(\theta^\alpha) + \theta^3 \boldsymbol{d} \tag{6}$$

in the reference and the current configuration, respectively. The current director vector is computed with the help of an orthogonal rotation

$$\boldsymbol{d} = \boldsymbol{R}(\boldsymbol{\omega})\, \boldsymbol{D} \qquad \boldsymbol{R}(\boldsymbol{\omega}) \in SO(3) \tag{7}$$

using the Rodrigues tensor, see [7]. The Green–Lagrange strains are decomposed into shell strains

$$\begin{aligned} \varepsilon_{\alpha\beta} &= \tfrac{1}{2}\left(\boldsymbol{x}_{,\alpha} \cdot \boldsymbol{x}_{,\beta} - \boldsymbol{X}_{,\alpha} \cdot \boldsymbol{X}_{,\beta}\right) \\ \kappa_{\alpha\beta} &= \tfrac{1}{2}\left(\boldsymbol{x}_{,\alpha} \cdot \boldsymbol{d}_{,\beta} + \boldsymbol{x}_{,\beta} \cdot \boldsymbol{d}_{,\alpha} - \boldsymbol{X}_{,\alpha} \cdot \boldsymbol{D}_{,\beta} - \boldsymbol{X}_{,\beta} \cdot \boldsymbol{D}_{,\alpha}\right) \\ \gamma_\alpha &= \boldsymbol{x}_{,\alpha} \cdot \boldsymbol{d} - \boldsymbol{X}_{,\alpha} \cdot \boldsymbol{D} \end{aligned} \tag{8}$$

and arranged in the column matrix

$$\boldsymbol{\varepsilon} = \left[\varepsilon_{11}, \varepsilon_{22}, 2\varepsilon_{12}, \kappa_{11}, \kappa_{22}, 2\kappa_{12}, \gamma_1, \gamma_2\right]^T \tag{9}$$

using Voigt notation. The second Piola–Kirchhoff stresses are pre-integrated in thickness direction neglecting higher order terms and assembled in the stress resultant column matrix

$$\boldsymbol{\sigma} = \left[n^{11}, n^{22}, n^{12}, m^{11}, m^{22}, m^{12}, q^1, q^2\right]^T . \tag{10}$$

The simplified analytic pre-integration of the constitutive relation is given by

$$\boldsymbol{\sigma} = \overline{\boldsymbol{D}}\boldsymbol{\varepsilon} \quad \text{with} \quad \overline{\boldsymbol{D}} = \begin{bmatrix} h\,\boldsymbol{C}_P & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{h^3}{12}\boldsymbol{C}_P & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & h\,\boldsymbol{C}_S \end{bmatrix} , \tag{11}$$

where $\boldsymbol{C}_P$ and $\boldsymbol{C}_S$ denote the linear elastic material matrices for in-plane and shear strains. The purely displacement-based variational formulation is based on the weak form

$$\int_{\Omega_0} \delta\boldsymbol{\varepsilon}^T \boldsymbol{\sigma}\, \mathrm{d}A - \int_{\Omega_0} \delta\boldsymbol{v}^T \bar{\boldsymbol{p}}_0\, \mathrm{d}A - \int_{\Gamma_0^N} \delta\boldsymbol{v}^T \bar{\boldsymbol{t}}_0\, \mathrm{d}s = 0 \tag{12}$$

of the equilibrium, where boundary tractions and surface loads are denoted by $\bar{\boldsymbol{t}}_0$ and $\bar{\boldsymbol{p}}_0$, respectively.

The isogeometric discretization of the problem (12) uses the NURBS basis functions N_I for the interpolation of the reference geometry

$$\boldsymbol{X} = \sum_{I=1}^{n_{en}} N_I \boldsymbol{X}_I \qquad \boldsymbol{D}^h = \sum_{I=1}^{n_{en}} N_I \boldsymbol{D}_I \tag{13}$$

as well as for the current position vector and its variations

$$\boldsymbol{x}^h = \sum_{I=1}^{n_{en}} N_I \boldsymbol{x}_I \qquad \delta\boldsymbol{x}^h = \sum_{I=1}^{n_{en}} N_I \delta\boldsymbol{u}_I . \tag{14}$$

The interpolation of the current director vector $\boldsymbol{d}$ and its variations $\delta\boldsymbol{d}$ can be performed in different ways. Two possible choices are given in the following and the resulting numerical results are compared in Sec. V:

- **discrete concept:** The current director vector is interpolated from rotated nodal director vectors

$$\boldsymbol{d}^h = \sum_{I=1}^{n_{en}} N_I \boldsymbol{R}_I^i \boldsymbol{D}_I \qquad \Rightarrow |\boldsymbol{d}^h| \neq |\boldsymbol{D}^h| , \tag{15}$$

  which yields artificial thinning of the geometry if the structure undergoes bending deformations. The variation is computed by

$$\delta\boldsymbol{d}^h = \sum_{I=1}^{n_{en}} \boldsymbol{T}_I \delta\boldsymbol{\beta}_I \qquad \boldsymbol{T}_I = N_I \boldsymbol{W}_I^T \boldsymbol{H}_I \boldsymbol{T}_{3I} . \tag{16}$$

  The update of the rotational tensor is performed in an additive manner by

$$\boldsymbol{R}_I^i = \boldsymbol{R}(\boldsymbol{\omega}_I^i) \qquad \boldsymbol{\omega}_I^i = \boldsymbol{\omega}_I^{i-1} + \boldsymbol{T}_{3I} \Delta\boldsymbol{\beta}_I . \tag{17}$$

  This rotational concept is the standard concept for Lagrange-based shell formulations. Its accuracy is sufficient for linear Lagrange basis functions with C^0-continuity between elements, but as it is shown in [24] this is not the case for higher order NURBS basis functions.

- **continuous concept:** The current director vector is rotated from the interpolated reference director vector $\boldsymbol{D}^h$ by

$$\boldsymbol{d}^h = \boldsymbol{R}^i \sum_{I=1}^{n_{en}} N_I \boldsymbol{D}_I \quad \Rightarrow |\boldsymbol{d}^h| = |\boldsymbol{D}^h| \tag{18}$$

  in every integration point. The variation of Eq. (18) is expressed in terms of the variation of the axial vector of the rotation $\boldsymbol{\omega}$. It results in

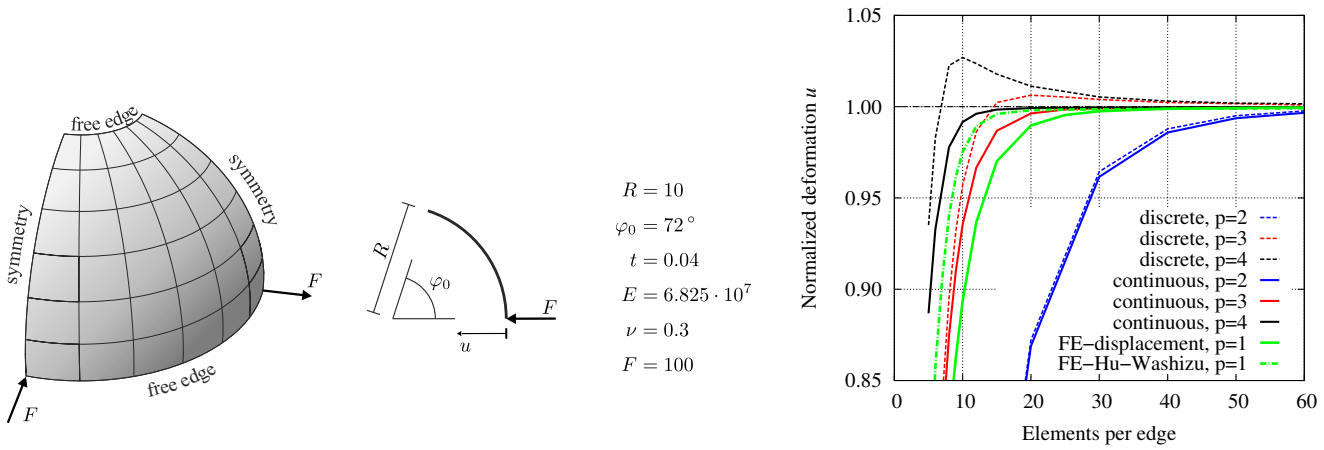
$$\delta\boldsymbol{d}^h = \sum_{I=1}^{n_{en}} \boldsymbol{T}_I \delta\boldsymbol{\beta}_I \qquad \boldsymbol{T}_I = \boldsymbol{W}^{hT} N_I \boldsymbol{T}_{3I} . \tag{19}$$

  A multiplicative update formulation of the rotational state is applied

$$\boldsymbol{R}^i = \boldsymbol{R}(\Delta\boldsymbol{\omega}^h)\boldsymbol{R}^{i-1} \qquad \boldsymbol{R}_I^i = \boldsymbol{R}(\Delta\boldsymbol{\omega}_I)\boldsymbol{R}_I^{i-1} . \tag{20}$$

V. NUMERICAL EXAMPLE

The hemispherical shell with hole proposed in [26] is used as numerical example to study the impact of the rotational concept on the deformation results. A system sketch, the Young's modulus E, the Poisson ratio ν and the wall thickness t are given in Fig. 4a. The deformation u of the inward loaded point is used to compare the deformation convergence behavior for mesh refinement for different orders p of the NURBS basis functions. For all orders the two presented rotational concepts



(a) System sketch, boundary conditions and material constants

(b) Effect of different rotational concepts.

Fig. 4. Hemispherical shell with hole.

are compared. Full Gauss integration is used. The results are given in Fig. 4b. Computations with the standard discrete rotational concept converge from above and the accuracy of results deteriorates with rising order of the basis functions. The deformations are overestimated by the discrete rotational concept. This is due to an artificial thinning of the structure. The discrete interpolation of the director vector given in Eq. (15) adds up vectors with differing direction. Due to the triangle inequality, the interpolated current director vectors are shorter than the interpolated reference director vectors. This effect is especially visible if large changes of curvature occur, which is the case in this example with large deformations and finite rotations. The continuity and thus the area of influence of each control point grows with rising order of the basis functions, see Fig. 3. The influence of the artificial thinning yielded by the triangle inequality rises with the continuity. This explains the deterioration of the accuracy of the discrete concept for rising orders of the basis functions.

In contrast to that, computations with the continuous concept yield convergence from below and the error level decreases with rising order of the basis functions. Artificial thinning is precluded by the use of the interpolation given in Eq. (18). This ensures convergence from below and improved accuracy for rising orders of the basis functions.

The discrete concept yields good results for an order $p = 2$, which is close to its standard field of application in Lagrange-based shell formulations with linear basis functions. If higher orders are used, then the discrete concept fails to produce higher accuracy for higher orders which naturally require higher computational effort. The situation is completely different if the continuous concept, which is adapted to the needs of NURBS-based isogeometric analysis, is used. Here the use of basis functions with higher orders yield results with higher accuracy. This shows that the presented concept of continuous rotations is able to make use of the advantaged offered by NURBS and at the same time solves one challenge introduced by the use of NURBS. The comparison to standard finite shell elements (FE-displacement and FE-Hu–Washizu in Fig. 4b) shows the need for methods against locking effects, which have to be adapted to isogeometric shell elements.

VI. Conclusion

The challenges and possibilities of isogeometric analysis have been illustrated. The sample challenge of interpolating vectorial quantities has been shown with the help of an isogeometric Reissner–Mindlin shell formulation with rotational formulation for the determination of the director vector. The numerical example has shown that the issue of interpolation of vectors has been solved in a way that at the same time uses the advantage of the higher continuity which is used to obtain higher accuracy of the numerical results.

References

[1] T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs, "Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement," *Comput. Meth. Appl. Mech. Engrg.*, vol. 194, pp. 4135–4195, 2005.

[2] J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs, *Isogeometric analysis: Toward integration of CAD and FEA*. Chichester: Wiley, 2009.

[3] D. J. Benson, Y. Bazilevs, M.-C. Hsu, and T. J. R. Hughes, "Isogeometric shell analysis: The Reissner–Mindlin shell," *Comput. Meth. Appl. Mech. Engrg.*, vol. 199, pp. 276–289, 2010.

[4] W. Dornisch, S. Klinkel, and B. Simeon, "Isogeometric Reissner–Mindlin shell analysis with exactly calculated director vectors," *Comput. Meth. Appl. Mech. Engrg.*, vol. 253, pp. 491–504, 2013.

[5] R. Echter, B. Oesterle, and M. Bischoff, "A hierarchic family of isogeometric shell finite elements," *Comput. Meth. Appl. Mech. Engrg.*, vol. 254, pp. 170–180, 2013.

[6] W. Dornisch and S. Klinkel, "Treatment of Reissner–Mindlin shells with kinks without the need for drilling rotation stabilization in an isogeometric framework," *Comput. Meth. Appl. Mech. Engrg.*, vol. 276, pp. 35–66, 2014.

[7] W. Dornisch, R. Müller, and S. Klinkel, "An efficient and robust rotational formulation for isogeometric Reissner–Mindlin shell elements," *Comput. Meth. Appl. Mech. Engrg.*, vol. 303, pp. 1–34, 2016.

[8] J. Kiendl, K.-U. Bletzinger, J. Linhard, and R. Wüchner, "Isogeometric shell analysis with Kirchhoff-Love elements," *Comput. Meth. Appl. Mech. Engrg.*, vol. 198, pp. 3902–3914, 2009.

[9] D. J. Benson, Y. Bazilevs, M.-C. Hsu, and T. J. R. Hughes, "A large deformation, rotation-free, isogeometric shell," *Comput. Meth. Appl. Mech. Engrg.*, vol. 200, pp. 1367–1378, 2011.

[10] J. Kiendl, Y. Bazilevs, M.-C. Hsu, R. Wüchner, and K.-U. Bletzinger, "The bending strip method for isogeometric analysis of Kirchhoff–Love shell structures comprised of multiple patches," *Comput. Meth. Appl. Mech. Engrg.*, vol. 199, pp. 2403–2416, 2010.

[11] D. J. Benson, S. Hartmann, Y. Bazilevs, M.-C. Hsu, and T. J. R. Hughes, "Blended isogeometric shells," *Comput. Meth. Appl. Mech. Engrg.*, vol. 255, pp. 133–146, 2013.

[12] M. Breitenberger, A. Apostolatos, B. Philipp, R. Wüchner, and K.-U. Bletzinger, "Analysis in computer aided design: Nonlinear isogeometric B-Rep analysis of shell structures," *Comput. Meth. Appl. Mech. Engrg.*, vol. 284, pp. 401–457, 2015.

[13] W. Dornisch and S. Klinkel, "Boundary Conditions and Multi-Patch Connections in Isogeometric Analysis," *Proc. Appl. Math. Mech.*, vol. 11, no. 1, pp. 207–208, 2011.

[14] A. Apostolatos, R. Schmidt, R. Wüchner, and K.-U. Bletzinger, "A Nitsche-type formulation and comparison of the most common domain decomposition methods in isogeometric analysis," *Int. J. Numer. Meth. Engng.*, vol. 97, no. 7, pp. 473–504, 2014.

[15] M. Ruess, D. Schillinger, A. I. Özcan, and E. Rank, "Weak coupling for isogeometric analysis of non-matching and trimmed multi-patch geometries," *Comput. Meth. Appl. Mech. Engrg.*, vol. 269, pp. 46–71, 2014.

[16] V. P. Nguyen, P. Kerfriden, M. Brino, S. P. A. Bordas, and E. Bonisoli, "Nitsche's method for two and three dimensional NURBS patch coupling," *Comput. Mech.*, vol. 53, no. 6, pp. 1163–1182, 2014.

[17] W. Dornisch, G. Vitucci, and S. Klinkel, "The weak substitution method – An application of the mortar method for patch coupling in NURBS-based isogeometric analysis," *Int. J. Numer. Meth. Engng.*, vol. 103, no. 3, pp. 205–234, 2015.

[18] W. Dornisch and R. Müller, "Dual and approximate dual basis functions for B-splines – Comparison at the example of mortar patch coupling in isogeometric analysis," *sub. to Comput. Meth. Appl. Mech. Engrg.*, 2015.

[19] H.-J. Kim, Y.-D. Seo, and S.-K. Youn, "Isogeometric analysis for trimmed CAD surfaces," *Comput. Meth. Appl. Mech. Engrg.*, vol. 198, pp. 2982–2995, 2009.

[20] R. Schmidt, R. Wüchner, and K.-U. Bletzinger, "Isogeometric analysis of trimmed NURBS geometries," *Comput. Meth. Appl. Mech. Engrg.*, vol. 241–244, pp. 93–111, 2012.

[21] D. Schillinger, L. Dedè, M. A. Scott, J. A. Evans, M. J. Borden, E. Rank, and T. J. R. Hughes, "An isogeometric design-through-analysis methodology based on adaptive hierarchical refinement of NURBS, immersed boundary methods, and T-spline CAD surfaces," *Comput. Meth. Appl. Mech. Engrg.*, vol. 249-252, pp. 116–150, 2012.

[22] A. P. Nagy and D. J. Benson, "On the numerical integration of trimmed isogeometric elements," *Comput. Meth. Appl. Mech. Engrg.*, vol. 284, pp. 165–185, 2015.

[23] L. Piegl and W. Tiller, *The NURBS book*, 2nd ed., ser. Monographs in visual communications. Berlin: Springer, 1997.

[24] W. Dornisch, *Interpolation of Rotations and Coupling of Patches in Isogeometric Reissner–Mindlin Shell Analysis*. Ph.D. thesis. Lehrstuhl für Baustatik und Baudynamik, RWTH Aachen, 2015.

[25] D. Schillinger, J. A. Evans, A. Reali, M. A. Scott, and T. J. R. Hughes, "Isogeometric collocation: Cost comparison with Galerkin methods and extension to adaptive hierarchical NURBS discretizations," *Comput. Meth. Appl. Mech. Engrg.*, vol. 267, pp. 170–232, 2013.

[26] R. H. Macneal and R. L. Harder, "A proposed standard set of problems to test finite element accuracy," *Finite Elem. Anal. Des.*, vol. 1, no. 1, pp. 3–20, 1985.

Linking Bioprocess Engineering and Electrochemistry for Sustainable Biofuel Production

Mareike Engel*, Julia Thieringer* and Nils Tippkötter*
*Junior Research Group BioSats, Bioprocess Engineering, University of Kaiserslautern, Kaiserslautern, Germany

***Abstract*—Electromicrobial engineering is an emerging, highly interdisciplinary research area linking bioprocesses with electrochemistry. In this work, microbial electrosynthesis (MES) of biobutanol is carried out during acetone-butanol-ethanol (ABE) fermentations with *Clostridium acetobutylicum*. A constant electric potential of −600 mV (vs. Ag/AgCl) with simultaneous addition of the soluble redox mediator neutral red is used in order to study the electron transfer between the working electrode and the bacterial cells. The results show an earlier initiation of solvent production for all fermentations with applied potential compared to the conventional ABE fermentation. The final butanol concentration can be more than doubled by the application of a negative potential combined with addition of neutral red. Moreover a higher biofilm formation on the working electrode compared to control cultivations has been observed. In contrast to previous studies, our results also indicate that direct electron transfer (DET) might be possible with *C. acetobutylicum*. The presented results make microbial butanol production economically attractive and therefore support the development of sustainable production processes in the chemical industry aspired by the "Centre for resource-efficient chemistry and raw material change" as well as the the project "NanoKat" working on nanostructured catalysts in Kaiserslautern..**

I. Introduction

Fermentative butanol production was first discovered by Louis Pasteur in 1862 [1]. Since then, microbiologically produced butanol (biobutanol) has gained much attention because of its use as bulk chemical and solvent. More recently, biobutanol has been considered as a sustainable alternative to fossil fuels because it is very similar to gasoline [2], [3].

Since its isolation in 1912, the anaerobic, gram-positive strain *Clostridium acetobutylicum* has been used as a model organism for other solvent producing clostridial species [4]. The strain's metabolism is divided into two different phases: the acidogenic and the solventogenic phase. During the acidogenic phase, the bacterium produces acetate and butyrate from glucose. In the later solventogenic phase, acetone, butanol and ethanol (ABE) are secreted [5], [6].

Although *Clostridium acetobutylicum* has been studied intensively, the shift between the acidogenic and solventogenic phase is not yet completely understood and butanol productivities are not yet satisfying. For the moment, petrochemically produced butanol is still much cheaper than the biotechnological product. Jiang et al. [7] calculated that petrochemically produced n-butanol costs 1.52 $/kg, whereas butanol from a corn based ABE fermentation costs 1.87 $/kg. As Kumar and Gayen [8] pointed out, using lingocellulosic feedstock or industrial wastes could be a comparatively cheap substrate for ABE fermentation. But even in this case, product yield has to be further improved. The stimulation of the clostridial metabolism by application of an electric potential during the fermentation could be a new approach to increase butanol production. This methods links conventional ABE fermentations with electrochemistry and will be described in the following.

II. Microbial electrosynthesis of biobutanol

A. Bioelectrochemical systems

For combining a biocatalyst with an electrochemical reaction step, bioelectrochemical systems (BES) are used. A typical BES contains a two- or three-electrode-setup with a working (WE) and a counter electrode (CE) and if necessary an additional reference electrode (RE). Working and counter electrode can be either in one compartment or separated, e.g. by a proton exchange membrane. At the WE the interaction between the microorganism (which generally forms a biofilm) and the electrode takes place. At the counter electrode a counter reaction occurs. The reference electrode is used in a three-electrode-setup in order to apply or to measure a precise potential at the WE.

BES have been known for over a century in the form of microbial fuel cells (MFCs) [9]. In this case, the WE acts as electron acceptor (anode) during the fermentation, thus producing an electric current. Microbial electrosynthesis (MES) has been discovered more recently and describes the inverse process. A separated bioelectrochemical system with a three-electrode-setup for MES is illustrated in Figure 1. During MES, the microbes take up electrons from the WE (cathode) in order to use these additional electrons for product formation. The mechanisms of extracellular electron transfer between a microorganism and an electrode is still a key aspect of ongoing research [10].

B. Extracellular electron transfer

Currently, different mechanisms of extracellular electron transfer are discussed [11]. For direct electron transfer (DET) the microorganisms attach to a solid-state electrode which is used as electron donor or acceptor [12]. Heme containing c-type cytochromes, as well as other metal containing proteins like hydrogenase are discussed as possible transmembrane transport chain for electrons [5], [11], [13], [14]. For a longer range DET conductive appendages, like pili, also called nanowires, are discussed [14], [15], [16]. Electron transfer can also occur with soluble compounds as electron donors/acceptors. During indirect electron transfer (IET), electrochemically synthesized compounds like hydrogen or formic

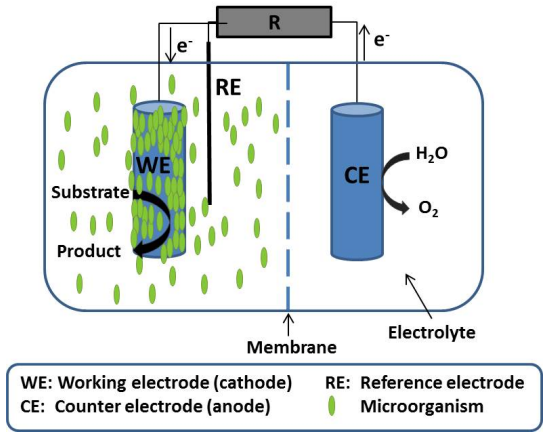


Fig. 1. Separated BES for microbial electrosynthesis. Electron transfer takes place from the working electrode (cathode) to the cells. Water is oxidized as a counter reaction at the CE (anode).

acid are metabolized by the microorganism. Alternatively, a redox mediator substance can be used as shuttle between the electrode and the microbe (mediator based electron transfer, MET). This redox mediator can be secreted by the microorganism (e.g. flavines or phenazines) or added abiotically into the system. Typical artifical redox mediators are viologens or redox dyes like neutral red [14].

DET and MET have been investigated intensively for the electroactive, gram-negative species *Shewanella oneidensis* and *Geobacter sulfurreducens* for MFC use [11]. For gram-positive bacteria and the MES mode, results are still scarce. This work aims to further elucidate the interaction of gram-positive bacteria with solid-state electrodes and its impact on product formation using *Clostridium acetobutylicum* and the MES of biobutanol as model system.

C. Metabolism of Clostridium acetobutylicum

In anaerobic ABE fermentations with *C. acetobutylicum* using glucose as sole carbon source and electron donor, product formation takes place in two different phases [17]. During the growth-associated, acidogenic phase, acetate and butyrate are mainly produced. When growth starts to stagnate, the bacterium enters the solventogenic phase and begins to metabolize glucose and the earlier secreted acids simultaneously. During the solventogenic phase, the main products are acetone, butanol and ethanol (ABE) with a typical mass ratio of 3:6:1 [5], [17], [18]. Different intra- and extracellular parameters affect the shift between the acidogenic and the solventogenic phase. An extracellular pH < 4.5, intracellular acid accumulation, especially high butyrate concentrations, nutrient limitations or high glucose concentrations are reported to trigger the shift [19], [20], [21], [22]. Different authors also report that external stress, e.g. addition of reagents can lead to the initiation of solventogenesis [23], [24]. The effect of electrochemically induced stress has not yet been studied in detail. Some authors investigated the influence of abiotically added electron mediator (viologen dyes or neutral red) with or without application of an electric potential [19], [25], [26]. They found an altered electron flow leading to higher butanol production with these substances during ABE fermentation. Peguin and Soucaille [27] found a 51 % increase in butanol yield when 1 mM methyl viologen was added into the fermentation medium and a potential of −800 mV (vs. Ag/AgCl) was applied. When only a potential of −800 mV (vs. Ag/AgCl) was applied, Peguin et al. [26], [27] did not observe any metabolic changes. This leads to the conclusion that DET is not possible. In this work, we confirm that electron transfer for *C. acetobutylicum* can be successfully mediated using neutral red and also show that DET might be possible.

III. Materials and Methods

A. Reactor design

In this study the gram-positive, strictly anaerobic bacterium *Clostridium acetobutylicum* was cultivated in H-shaped BES. These H-cells consist of two compartments made of two 120 mL glass bottles separated through a flange containing a proton selective Nafion membrane (QuinTech). WE and CE are made of 8.75 cm^2 pieces of a carbon fabric purchased at Kynol® Europa. The carbon fabric was dried at 100 °C overnight, put into 2-propanol for 30 min and then washed three times in deionized water. A silver/silver chloride electrode (Sensortechnik Meinsberg) was used as a reference. The electrodes were connected to a multi-potentiostat MultiEMStat with six measurement channels from PalmSens using platinum wires for the WE and stainless steel wires for the CE. Six miniature magnetic stirrers placed in an isothermal incubator hood allowed stirring of the working electrode compartments of six H-cell bioreactors in parallel while maintaining a constant temperature of 37 °C. For pH-control, pH-probes were inserted into the WE compartments and connected to a Profilux aquaristic computer with expansion box, enabling the addition of acid or base using a self-made feed station.

B. Cultivation conditions

The used biocatalyst *C. acetobutylicum* (DSM 792) was obtained from DSMZ - German Collection of Microorganisms and Cell Cultures and maintained under anaerobic conditions throughout the experiments and storage. The strain was stored at −80 °C in 35 % glycerol. Precultures were cultivated in 200 mL serum bottles containing 100 mL medium 104b, recommended by the DSMZ with an initial pH value of 6.8. The medium was boiled in a microwave, closed with a butyl rubber stopper and then cooled down under N_2 flow before autoclaving for 20 min at 121 °C. Fermentations were carried out in modified P2-medium (MP2-medium) containing 45 $g \cdot L^{-1}$ glucose, 1 $g \cdot L^{-1}$ ammonium acetate, 0.5 $g \cdot L^{-1}$ KH_2PO_4, 0.5 $g \cdot L^{-1}$ K_2HPO_4, 0.2 $g \cdot L^{-1}$ $MgSO_4 \cdot 7\ H_2O$, 0.01 $g \cdot L^{-1}$ $FeSO_4 \cdot 7\ H_2O$, 0.01 $g \cdot L^{-1}$ $MsSO_4 \cdot H_2O$, 1 $mg \cdot L^{-1}$ para-aminobenzoic acid, 0.01 $g \cdot L^{-1}$ biotin with an initial pH of 6.8. The medium was autoclaved separately from the H-cells for 20 min at 121 °C and filled into the working electrode compartments under aseptic conditions. The counter electrode compartment was filled with a sterile 0.1 M potassium phosphate buffer with

the same pH as the medium. Heat-labile vitamins, as well as neutral red for the study of mediated electron transfer, were sterile filtered and were added into the medium before inoculation. During the fermentations, pH was adjusted to 5.0 using 1 M NaOH in the working electrode compartment. Inoculum was 10 % (v/v) of a preculture grown at 37 °C for about 24 h. For electrochemical cultivations a constant electric potential of −600 mV was applied at the WE and the electrodes were polarized overnight.

C. Substrate and product analytics

Bacterial growth was monitored by measurement of the optical density at 700 nm with a spectrophotometer Viktor X3 from Perkin Elmer. Glucose concentration in the culture broth was measured by HPLC using a Ca^{2+} ion exchange column from Dr. Maisch GmbH (length 300 mm, diameter 8 mm). Acetone, butanol, ethanol, butyrate and acetate were analyzed with a gas chromatography system Clarus 600 with a Rtx-5 amine column (length 30 m, internal diameter 0.32 mm, film thickness 1.5 µm) and a refractive index detector. The initial temperature of 50 °C was held for 1 min before increasing the temperature up to 120 °C with a heating rate of 30 °C·min^{-1}. Helium was used as carrier gas with a flow of 250 mL·min^{-1}.

IV. Results

A. Biofilm formation during MES

In order to understand the influence of an applied potential during ABE-fermentation with *C. acetobutylicum* on the product formation, MES is carried out in BESs with the working electrode poised at −600 mV and medium supplementation of 125 µM neutral red in comparison to control fermentations in the same system without applied potential. During all fermentation experiments a visible biofilm formation occurred on the WE. For cultivations with an applied potential this biofilm formation was higher than for control fermentations. The cultivations based on mediated electron transfer show a 26 % increase in total biofilm mass on the WE (see Figure 2) relating to the control. The higher biofilm formation in the electrochemical fermentations is probably due to the interaction of the cells with the electrode for electron transfer. It has been reported that reduced neutral red (colorless) binds to the cell membrane where it is chemically oxidized by NAD^+ [28]. During the fermentation experiments with neutral red, it was indeed visible that the oxidized (red) neutral red was mainly attached to the cells in the biofilm and not dissolved in suspension.

The higher biofilm mass on the electrodes during fermentations with applied potential is also reflected in the lower optical density of the culture broth. In the control fermentation (Figure 3), the optical density is higher throughout the fermentation with an additional increase at the end of the fermentation. This increase might be due to biofilm detachment. When a potential is applied (Figure 4), the optical density is lower and no biofilm detachment takes place. This leads to the hypothesis that the cells are attached differently to the electrode when a potential is applied in comparison to a conventional biofilm.

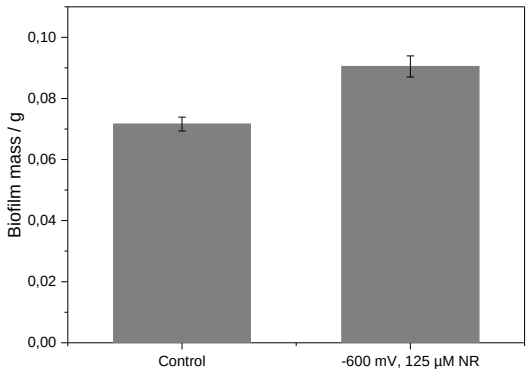


Fig. 2. Biofilm masses on working electrodes after 48 h fermentation for control cultivation and cultivations with an applied potential of −600 mV and 125 µM neutral red (NR). Mean values and standard deviations are from two independent experiments.

First fermentation experiments with the exclusive application of a potential of −600 mV without neutral red addition also showed higher biofilm formation than the control fermentation. Choi et al. [15] found that for *Clostridium pasteurianum* cells in the biofilm on the cathode, the electronegativity as well as the extracellular structure changed when a potential was applied. The authors measured a less negative zeta potential of the cells cultivated with a negative potential in comparison to a control cultivation [15]. Similar changes in the zeta potential might take place with *C. acetobutylicum* and might allow a higher attraction of the cells and the electrode by electrostatic forces. Furthermore, the same authors observed the formation of extracellular appendages which probably allow a better attachment of the cells among each other and on the electrode surface. They suggest that these appendages are conductive nanowires allowing a direct electron transfer [11], [15]. To our knowledge, nanowire formation has not been proven for *C. acetobutylicum*, but since the strain forms peritrichious flagella [17] it might be possible that those flagella act as nanowires in microbial electrosynthesis. Biofilm formation as well as the existence of nanowires has to be studied in more detail in order to understand the possible mechanism of direct electron transfer for *C. acetobutylicum*.

B. Changes in product formation during MES

Figures 3 and 4 also show the development of the acetate, butyrate and butanol concentrations during the fermentation experiments. Examining the product formation and the development of the optical density, differences in the acid-solvent shifts of the fermentation experiments can be distinguished. Growth levels off after 12 h for both cultures, but butanol production begins after 12 h only for the fermentation with applied potential. Final butanol concentration is highest with 2.3 g·L^{-1} when a potential of −600 mV is applied and neutral red is added. For the control fermentation, butanol production only starts at 32 h and reaches about 1 g·L^{-1} after 48 h. The application of the potential combined with the addition of the redox mediator substance advances the initiation of the

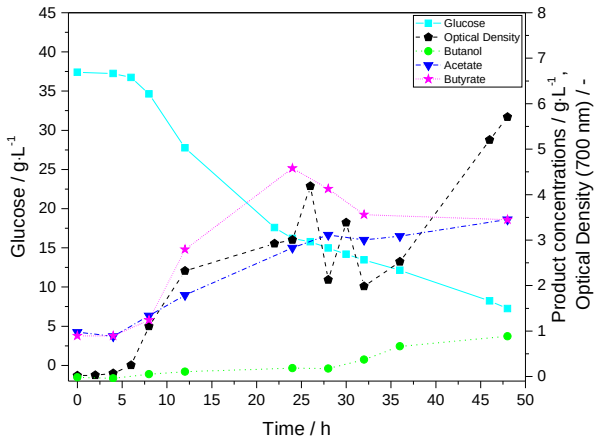


Fig. 3. Optical density, substrate and product concentrations during control fermentation

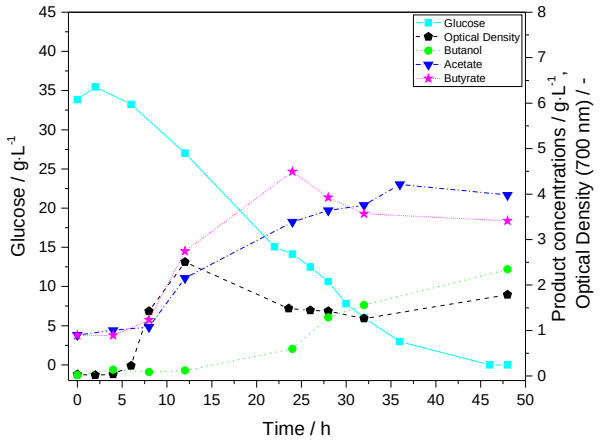


Fig. 4. Optical density, substrate and product concentrations during ABE fermentations with an applied potential of −600 mV and 125 µM neutral red.

solventogenesis by 24 h and leads to a 130 % increase in butanol concentration. Interestingly, during the electrochemical cultivation the butyrate concentration still increases when butanol is produced (Figure 4). There is not a distinct shift between acidogenesis and solventogenesis but an intermediate phase where acid and solvent production takes place. The control cultivation in contrast shows a clear decrease in butyrate concentration (Figure 3) before butanol is measured. Therefore the described reassimilation of butyrate for the conversion to butanol takes place whereas during microbial electrosynthesis the butanol seems to be exclusively produced from glucose in the beginning of the solventogenic phase.

Several authors have reported an increase in reduced product formation by the use of redox mediator substances like methyl viologen or neutral red in ABE fermentation [5], [25], [26]. Kim and Kim [25] found a 26 % increase in butanol concentration using 2 mM methyl viologen. Peguin and Soucaille [27], [26] studied the influence of methyl viologen with and without application of a negative potential of −800 mV (vs. Ag/AgCl). When only a potential of −800 mV (vs. Ag/AgCl) was applied, they did not observe any metabolic changes. But once 1 mM methyl viologen was added, solvent production started almost immediately leading to a 51 % increase in final butanol yield. Girbal et al. [5] studied the influence of the addition of 1 mM neutral red without the application of a potential. They found a higher specific butanol production rate of 2.51 $mmol \cdot h^{-1} \cdot g^{-1}$ compared to 0.01 $mmol \cdot h^{-1} \cdot g^{-1}$. It has been suggested that both methyl viologen and neutral red can replace ferredoxin in reactions involving ferredoxin oxidoreductase [26], [27]. This would lead to a higher NAD(P)H/NA(P)D^+ ratio during MET which is essential for the production of the reduced product butanol. But since methyl viologen is unable to cross the cytoplasmic membrane, it was also suggested that it acts directly on periplasmatic hydrogenases for gram-negative bacteria [14]. It is not yet clear how methyl viologen acts on gram-positive bacteria. Neutral red in contrast is able to bind to the cell membrane and can be oxidized by intracellular NAD^+ [29]. Therefore the altered product formation could be due to different effects of the mediators on the bacterial cells.

The 130 % increase in butanol concentration obtained in this work confirms the previously found increase in reduced products by the addition of neutral red. However, the exact mechanism of influence on the metabolism could not yet be satisfyingly explained. So far, no DET and therefore no influence of the exclusive application of a negative potential on the metabolism of *C. acetobutylicum* during MES has been observed [26]. Preliminary studies for the application of a potential without addition of neutral red revealed that DET might be possible. This has to be further studied in the future.

V. Conclusion

Linking bioprocess engineering and electrochemical techniques is a new approach to alter the electron flow in whole cell biocatalysts. This study shows that by the application of a negative electric potential in combination with the addition of a soluble redox mediator, the metabolism of the gram-positive bacterium *C. acetobutylicum* can be influenced. The already studied positive effect of MET by neutral red on the butanol production has been confirmed in this study and has shown an even higher increase in butanol production compared with results for methyl viologen. Additionally, the results support the idea of a DET mechanism for *C. acetobutylicum*. It now has to be elucidated, how the bacterial cells interact with the electrode in the case of MET and DET. Especially the role of hydrogenases and nanowires have to be examined in detail. Overall, the results are promising for a more economic, high yield ABE fermentation process based on the application of electrochemical methods. In the surroundings of the “Centre for resource-efficient chemistry and raw material change” as well as the the project “NanoKat” working on nanostructured catalysts in Kaiserslautern, ongoing studies are supposed to clarify how solvent production with clostridia as biocatalyst can be controlled electrochemically during ABE fermentation.

ACKNOWLEDGMENT

This work was supported by the network of promotion of young scientists at the University of Kaiserslautern "Nachwuchsring TU Kaiserslautern". The authors would like to thank the workgroup of biochemical engineering at DECHEMA Research Institute for their support.

REFERENCES

[1] L. Pasteur, "Quelques résultats nouveaux relatifs auf fermentations acétiques & butyrique: Some new results concerning the acetate and butyrate fermentations," 09.05.1862.

[2] J. Zheng, Y. Tashiro, Q. Wang, and K. Sonomoto, "Recent advances to improve fermentative butanol production: genetic engineering and fermentation technology," *Journal of bioscience and bioengineering*, vol. 119, no. 1, pp. 1–9, 2015.

[3] M. Köpke, S. Noack, and P. Dürre, "The Past, Present, and Future of Biofuels – Biobutanol as Promising Alternative," in *Biofuel Production-Recent Developments and Prospects*, Dos Santos Bernardes,Marco Aurelio, Ed. InTech, 2011.

[4] D. Hönicke, T. Lütke-Eversloh, Z. Liu, D. Lehmann, W. Liebl, and A. Ehrenreich, "Chemostat cultivation and transcriptional analyses of Clostridium acetobutylicum mutants with defects in the acid and acetone biosynthetic pathways," *Applied Microbiology and Biotechnology*, vol. 98, no. 23, pp. 9777–9794, 2014.

[5] L. Girbal, I. Vasconceles, Saint-Amans, and Silvie, Soucaille, Philippe, "How neutral red modified carbon and electron flow in Clostridium acetobutylicum grown in chemostat culture at neutral pH," *FEMS Microbiology Reviews*, vol. 16, pp. 151–162, 1995.

[6] N. P. Minton and D. J. Clarke, Eds., *Clostridia*. New York: Plenum Publishing Corporation, 1989.

[7] Y. Jiang, J. Liu, W. Jiang, Y. Yang, and S. Yang, "Current status and prospects of industrial bio-production of n-butanol in China," *Biotechnology advances*, 2014.

[8] M. Kumar and K. Gayen, "Developments in biobutanol production: New insights," *Applied Energy*, vol. 88, no. 6, pp. 1999–2012, 2011.

[9] M. C. Potter, "Electrical Effects Accompanying the Decomposition of Organic Compounds," *Proceedings of the Royal Society of London*, vol. 84, no. 571, pp. 260–276, 1911.

[10] B. Korth, Rosa, Luis F M, F. Harnisch, and C. Picioreanu, "A framework for modeling electroactive microbial biofilms performing direct electron transfer," *Bioelectrochemistry (Amsterdam, Netherlands)*, vol. 106, no. Pt A, pp. 194–206, 2015.

[11] A. Sydow, T. Krieg, F. Mayer, J. Schrader, and D. Holtmann, "Electroactive bacteria–molecular mechanisms and genetic tools," *Applied Microbiology and Biotechnology*, vol. 98, no. 20, pp. 8481–8495, 2014.

[12] B. Erable, N. M. Duţeanu, M. M. Ghangrekar, C. Dumas, and K. Scott, "Application of electro-active biofilms," *Biofouling*, vol. 26, no. 1, pp. 57–71, 2010.

[13] S. Peguin and P. Soucaille, "Modulation of Carbon and Electron Flow in Clostridium acetobutylicum by Iron Limitation and Methyl Viologen," *Applied and environmental microbiology*, vol. 61, no. 1, pp. 403–405, 1995.

[14] M. Rosenbaum, F. Aulenta, M. Villano, and L. T. Angenent, "Cathodes as electron donors for microbial metabolism: which extracellular electron transfer mechanisms are involved?" *Bioresource technology*, vol. 102, no. 1, pp. 324–333, 2011.

[15] O. Choi, T. Kim, H. M. Woo, and Y. Um, "Electricity-driven metabolic shift through direct electron uptake by electroactive heterotroph Clostridium pasteurianum," *Scientific reports*, vol. 4, no. 6961, pp. 1–10, 2014.

[16] S. Kato, "Biotechnological Aspects of Microbial Extracellular Electron Transfer," *Microbes and environments / JSME*, vol. 30, no. 2, pp. 133–139, 2015.

[17] P. Dürre, *Handbook on clostridia*, 10th ed. CRC Press, Tailor & Francis Group, 2005.

[18] F. Harnisch, Rosa, Luis F M, F. Kracke, B. Virdis, and J. O. Krömer, "Electrifying white biotechnology: engineering and economic potential of electricity-driven bio-production," *ChemSusChem*, vol. 8, no. 5, pp. 758–766, 2015.

[19] G. Rao and R. Mutharasan, "Altered Electron Flow in Continuous Cultures of Clostridium acetobutylicum Induced by Viologen Dyes," *Applied and environmental microbiology*, vol. 53, no. 6, pp. 1232–1235, 1987.

[20] I. S. Maddox, E. Steiner, S. Hirsch, S. Wessner, N. A. Gutierrez, J. R. Gapes, and K. C. Schuster, "The Cause of Acid-Crash and Acidogenic Fermentations During the Batch Acetone-Butanol-Ethanol (ABE-) Fermentation Process," *Journal of Molecular Microbiology & Biotechnology*, vol. 2, no. 1, pp. 95–100, 2000.

[21] P. Dürre, "Fermentative butanol production: bulk chemical and biofuel," *Annals of the New York Academy of Sciences*, vol. 1125, pp. 353–362, 2008.

[22] G. Rao and R. Mutharasan, "NADH levels and solventogenesis in Clostridium acetobutylicum: new insights through culture fluorescence," *Applied Microbiology and Biotechnology*, vol. 30, pp. 59–66, 1989.

[23] M. Kumar, S. Saini, and K. Gayen, "Elementary mode analysis reveals that Clostridium acetobutylicum modulates its metabolic strategy under external stress," *Molecular bioSystems*, vol. 10, no. 8, pp. 2090–2105, 2014.

[24] T. Millat, H. Janssen, G. J. Thorn, J. R. King, H. Bahl, R.-J. Fischer, and O. Wolkenhauer, "A shift in the dominant phenotype governs the pH-induced metabolic switch of Clostridium acetobutylicumin phosphate-limited continuous cultures," *Applied Microbiology and Biotechnology*, vol. 97, no. 14, pp. 6451–6466, 2013.

[25] Kim, Sung T. and B. H. Kim, "Electron flow shift in Clostridium acetobutylicum Fermentation by Electrochemically Introduced Reducing Equivalent," *Biotechnology Letters*, vol. 10, no. 2, pp. 123–128, 1988.

[26] S. Peguin and P. Soucaille, "Modulation of metabolism of Clostridium acetobutylicum grown in chemostat culture in a three-electrode potentiostatic system with methyl viologen as electron carrier," *Biotechnology & Bioengineering*, vol. 51, pp. 342–348, 1996.

[27] S. Peguin, P. Delorme, G. Goma, and P. Soucaille, "Enhanced Alcohol Yields in Batch Cultures of Clostridium acetobutylicum using a Three-Electrode Potentiometric System with Methyl Viologen as Electron Carrier," *Biotechnology Letters*, vol. 16, no. 3, pp. 269–274, 1994.

[28] D. H. Park, M. Laivenieks, M. V. Guettler, M. K. Jain, and J. G. Zeikus, "Microbial Utilization of Electrically Reduced Neutral Red as Sole Electron Donor for Growth and Metabolite Production," *Applied and environmental microbiology*, vol. 65, no. 7, pp. 2912–2917, 1999.

[29] D. H. Park and J. G. Zeikus, "Utilization of Electrically Reduced Neutral Red by Actinobacillus succinogenes: Physiological Function of Neutral Red in Membrane-Driven Fumarate Reduction and Energy Conversion," *Journal of Bacteriology*, vol. 181, no. 8, pp. 2403–2410, 1999.

On segregation in dry granular material flows in mixing processes

Dominik Gilberg
Fraunhofer ITWM, Germany
Email: gilberg@itwm.fhg.de

***Abstract*—Today, several models exist to describe segregation in granular materials for specific processes. In this work a mixture theory based segregation model and a continuum model of Navier-Stokes-Type for granular flows have been coupled to describe the segregation in general industrial processes. The focus lies on dry granular material flows of particles of different size. To test the model the rotational mixing of small and large particles in a tumbler is simulated.**

***Index Terms*—granular media, granular segregation, mixing**

I. Introduction

A. Motivation and industrial relevance

GRANULAR materials consist of macroscopic particles of different size and kind. They regularly appear in most people daily lives *(e. g. morning cereals)* as well as in the fields of geophysics *(dunes and avalanches)*, agriculture *(corn silos)* and in many different industrial processes. The production of grains and aggregates reaches 10 billion metric tons every year and their processing consumes 10% of all energy produced on earth. It is the most manufactured material in industry after water [1]. For these reasons granular materials are the subject of intensive engineering research. A well-known phenomenon of granular materials is the segregation of particles due to different size, density or shape. In this paper the focus lies on size segregation as it is the most dominant reason for segregation. Segregation describes the effect of polydisperse granular matter to demix under motion. While the granular system is in motion small voids appear, where small particles are more likely to fill these voids than the large ones. In the mining industry it is a helpful process, but it leads to problems in most other industrial areas. In the food and pharmaceutical industry mixing processes are important to guarantee the quality of products. The phenomenon of segregation counteracts the mixing processes, which can worsen the products. For all these reasons it is helpful to simulate the flow behaviour of granular materials including the effect of segregation.

B. Existing models

Simulating the flow behaviour and the occurring phenomena of granular media can be done quite exact using the discrete element method (DEM), that describes the motion of each particle. As in industrial processes the number of particles is very high, the simulation time is very long and costly. To overcome the long computation times, several continuum models have been found to describe the flow behaviour of granular media. Unfortunately, these models need to be fitted to experimental measurements and mostly can only describe specific industrial processes or natural phenomena. For all that, there are promising models like the Bouchaud-Cates-Ravi Prakash-Edwards (BCRE) model. The model combines the flowing and static behaviour of granular materials during the formation of sand piles [2], [3]. Latz and Schmidt modified a system of hydrodynamic equations of Navier-Stokes-Type to cover the regime of fast dilute flow as well as slow dense flow, where the density of the granular material is close to the maximum packing density [4]. Further, there are several models focusing on the phenomenon of segregation, where the most models are made to describe granular avalanches. One of the first models to describe segregation was proposed by Bridgwater, Foo and Stevens in 1985 [5]. It is an advection equation for one of two particle phases. The segregation flux in this equation shuts off when only one particle phase is present. Models of similar structure were derived by Savage and Lun [6] in 1988 and Gray and Thornton [7] in 2005.
All these segregation models only describe the segregation process for a given flow field that is typical for avalanches. The previously mentioned models for the granular flow behaviour try to capture the different flowing regimes. In general they assume identical particles and they are not able to capture segregation. In the following a granular flow model and a model for the segregation have been combined to derive a system of equations that can describe the segregation in a general industrial mixing process.

C. Structure of the work

After introducing the topic and the state of the art the next section shows the mathematical models. First, the Latz-Schmidt model is presented in the used version. Second, a modified version of the Gray-Thornton model is introduced, which will be combined with the Latz-Schmidt model. Section 3 shows simulation results of the final coupled model. The framework for these simulations is a rotating tumbler half-filled with small and large particles. Finally, the quality of the model is evaluated.

II. Mathematical model

The mathematical model consists of a segregation equation, based on mixture theory, and a continuum model of Navier-Stokes-Type describing the behaviour of granular flows. It is assumed that the granular mixture consists of small and large

particles with equal material densities $\rho^{s*} = \rho^{l*} = const$. In the standard derivation processes the equations are formulated in terms of mass- or partial densities. They are given by a linear volume fraction scaling

$$\rho^\nu = \phi^\nu \rho^{\nu*}, \tag{1}$$

where ϕ^ν is the volume fraction of phase ν with

$$\phi^\nu \in [0, \Phi_{max}], \text{ for } \nu \in \{s, l\} \tag{2}$$

and Φ_{max} is the maximal allowed space occupied by granular material. From mixture theory it is known that the mass density for the mixture is given by

$$\rho = \rho^s + \rho^l \tag{3}$$

and

$$\rho = \Phi \rho^*, \tag{4}$$

where Φ is the volume occupied by all particles. It holds

$$\phi^s + \phi^l = \Phi \tag{5}$$

and consequently $\rho^* = \rho^{s*} = \rho^{l*} = const$. As all upcoming terms are linear in the density, all equations are directly given in terms of the respective volume fractions to avoid further variable changes.

A. The Latz-Schmidt model

The origin of the Latz-Schmidt model is a hydrodynamic model derived from kinetic theory. The derivation process was done by using Chapman-Enskog theory. As kinetic theory is just able to reproduce dilute flow regimes they modified the model using a theory from soil mechanics. The resulting system of equations is a hybrid model of kinetic theory and the soil mechanical approach. This extends the applicability of the hydrodynamic model to arresting granular flows. In this work a version of the Latz-Schmidt model similar to that derived in [4] is used. To capture the main granular aspects the model is given by mass and momentum balances

$$\partial_t \Phi + \nabla \cdot (\Phi \mathbf{u}) = 0 \tag{6}$$

$$\partial_t (\Phi \mathbf{u}) + \nabla \cdot (\Phi \mathbf{u} \otimes \mathbf{u}) = \nabla \cdot \boldsymbol{\sigma} + \Phi \mathbf{g}, \tag{7}$$

where the stress tensor is given by

$$\boldsymbol{\sigma} = \eta \nabla \mathbf{u} - p\mathbf{I}, \tag{8}$$

and additionally an equation for the granular temperature T, which is the mean fluctuating kinetic energy

$$\partial_t (\Phi T) + \nabla \cdot (\Phi T \mathbf{u}) = \frac{3}{2}(\dot{\gamma}^2 - \nabla \cdot \mathbf{q}) - \varepsilon \Phi T. \tag{9}$$

The equations (6), (7) and (9) are solved for the granular volume fraction Φ, the velocity $\mathbf{u}$ and the granular temperature T. Further, η defines the granular viscosity, $\mathbf{g}$ the gravitational acceleration and p the pressure. The temperature equation (9) has the form of a heat equation. The temperature grows with the shear rate $\dot{\gamma}$, which is a product of viscous forces. The second term of the right hand side is the heat flux

$$\mathbf{q} = -\lambda \nabla T, \tag{10}$$

where λ denotes the thermal conductivity. Special is the last term, which accounts for the loss of energy due to inelastic collisions, where ε is the energy dissipation rate. Due to this term a moving system will rapidly come to rest, if there are no forces acting on the granular material.
To close the system, equations of state, based on kinetic theory, are used as described in [8]. In the following the kinetic expressions are denoted by the subscript K. In the final hybrid model these expressions are modified by introducing the so-called yield pressure p_Y, that is active in the dense flow regime $\Phi > \Phi_0$. The pressure p is given by

$$p = p_K + p_Y \tag{11}$$

$$p_K = \Phi T g(\Phi) \tag{12}$$

$$p_Y = \Theta(\Phi - \Phi_0) T_0 (\Phi - \Phi_0) g(\Phi), \tag{13}$$

where $\Theta(\cdot)$ is the Heaviside step function and

$$g(\Phi) := \left(1 - \frac{\Phi}{\Phi_{max}}\right)^{-1} \tag{14}$$

the radial distribution function at contact. The transport coefficients are given by

$$\eta = \eta_K (1 + \frac{p_Y}{p_K}), \qquad \eta_K = \eta_0 \sqrt{T} g(\Phi) \tag{15}$$

$$\varepsilon = \varepsilon_K (1 + \frac{p_Y}{p_K}), \qquad \varepsilon_K = \varepsilon_0 \sqrt{T} \Phi g(\Phi) \tag{16}$$

$$\lambda = \lambda_K (1 + \frac{p_Y}{p_K}), \qquad \lambda_K = \lambda_0 \sqrt{T} \Phi g(\Phi). \tag{17}$$

All constants involved, subscribed with 0, either have to be looked up in literature or validated by experiments. In the dilute regime the yield pressure p_Y is zero and the model is equal to a pure kinetic model, as all variables and coefficients reduce to their kinetic expressions. In the dense regime p_Y is non zero and hence, the model shows different flow behaviour. A more detailed derivation of the model is given in [4].

B. Segregation equation adapted by Gray and Thornton

For the derivation of the segregation equation the modelling steps were performed like in [7]. Further, the model has been adapted to be in the preferred form. First, it has been assumed that each material involved can be described as a continuum. Let there be two particle phases of small s and large l particles of equal density $\rho^{s*} = \rho^{l*} = const$. Each phase fulfils conservation of mass and momentum

$$\partial_t \phi^\nu + \nabla \cdot (\phi^\nu \mathbf{u}^\nu) = 0 \tag{18}$$

$$\partial_t (\phi^\nu \mathbf{u}^\nu) + \nabla \cdot (\phi^\nu \mathbf{u}^\nu \otimes \mathbf{u}^\nu) = \nabla \sigma^\nu + \phi^\nu \mathbf{g} + \beta^\nu \tag{19}$$

for $\nu \in \{s, l\}$, where β^ν is the interaction force on phase ν. Introducing the relative velocity between phase ν and the mixture

$$\mathbf{u}^{\nu m} = \mathbf{u}^\nu - \mathbf{u} \tag{20}$$

the mass balance can be rewritten in the form

$$\partial_t \phi^\nu + \nabla \cdot (\phi^\nu \mathbf{u}) = -\nabla \cdot (\phi^\nu \mathbf{u}^{\nu m}). \tag{21}$$

In equation (21) the left hand side describes the flow with the bulk velocity of the mixture, whereas, the term on the right

hand side represents the segregation as it is a motion relative to the bulk flow. Hence, one needs to find an expression for the relative velocity $\mathbf{u}^{\nu m}$. Similar to Gray and Thornton it can be assumed that the segregation only takes place in the direction of the gravitational force. In the used convention this is the negative $z-$axis. Further, one can assume that the phase velocity in $x-$ and $y-$direction is equal to the bulk velocity

$$u_1^\nu = u_1 \quad \text{and} \quad u_2^\nu = u_2 \quad \forall \nu. \tag{22}$$

The mass balance then simplifies to

$$\partial_t \phi^\nu + \nabla \cdot (\phi^\nu \mathbf{u}) = -\partial_z(\phi^\nu u_3^{\nu m}). \tag{23}$$

The relative velocity $u_3^{\nu m}$ can be gained from the momentum balance by introducing a new pressure scaling

$$p^\nu = f^\nu p \quad \forall \nu, \tag{24}$$

where the factor f^ν determines the proportion of the load carried by phase ν. Consequently, it holds

$$f^s + f^l = 1. \tag{25}$$

Gray and Thornton formulated the condition in such a way that the small particles carry less of the overburden pressure than the large particles. So the small particles can percolate through the granular matrix to the ground. They found that

$$\begin{aligned} f^l &= \phi^l + B\phi^s\phi^l, \\ f^s &= \phi^s - B\phi^s\phi^l, \end{aligned} \tag{26}$$

where B is a non-dimensional magnitude. Under the assumptions done so far it is necessary to scale (26) by $\frac{1}{\Phi}$ to fulfil (25) as well as the saturation condition (5). Hence, the load functions are given by

$$\begin{aligned} f^l &= \frac{1}{\Phi}(\phi^l + B\phi^s\phi^l), \\ f^s &= \frac{1}{\Phi}(\phi^s - B\phi^s\phi^l). \end{aligned} \tag{27}$$

In contrary to Gray and Thornton it has been assumed in this model that B depends on the granular temperature T. If the system is in motion the granular temperature is high and the system is less densely packed than in the static regime. This means the load factor especially for the small particles shrinks. Hence, the magnitude B must grow with the granular temperature T. Here, the temperature stands representative for the ability of the system to segregate. The main cause for segregation is shear. Due to shear small voids appear and the small particles can percolate through the granular matrix. Considering the granular temperature equation (9) in a steady state the granular temperature is given by

$$T = \frac{2}{3}\frac{\dot\gamma^2}{\varepsilon\Phi}. \tag{28}$$

Hence, the shear rate $\dot\gamma$ is proportional to the square root of the granular temperature

$$\dot\gamma \propto \sqrt{T}. \tag{29}$$

Consequently, the following form for $B(T)$ has been assumed:

$$B(T) = \begin{cases} 0 & \text{if } T < T_c \\ b\sqrt{T} & \text{if } T \geq T_c, \end{cases} \tag{30}$$

where the critical temperature T_c defines the point where the system changes from a static to a dense flow regime.
Further, Gray and Thornton argue that the interaction force should take the form

$$\boldsymbol{\beta}^\nu = p\nabla f^\nu - \phi^\nu c(\underbrace{\mathbf{u}^\nu - \mathbf{u}}_{\mathbf{u}^{\nu m}}) \quad \forall \nu. \tag{31}$$

The first term ensures that the percolation process is driven by intrinsic rather than partial pressure gradients [7]. The second term is a Darcy like drag term as proposed by Morland [9], where c is the coefficient of inter particle drag.
Neglecting the acceleration terms in the momentum balance and combining the previous assumptions as shown in [7], the relative velocities are given by

$$u_3^{sm} = +S_{sl}\sqrt{T}\phi^s \tag{32}$$

$$u_3^{lm} = -S_{sl}\sqrt{T}\phi^l, \tag{33}$$

respectively, where

$$S_{sl} = -b\frac{g}{c} \tag{34}$$

is the mean segregation velocity that needs to be fitted to experimental measurements. If the system is in a static state the granular temperature is small. As a consequence, $B(T)$ becomes zero and the relative velocity as well. The particles cannot segregate any more. Finally, the segregation equations are given by

$$\partial_t \phi^s + \nabla \cdot (\phi^s \mathbf{u}) = \partial_z(S_{sl}\sqrt{T}\phi^s\phi^l), \tag{35}$$

$$\partial_t \phi^l + \nabla \cdot (\phi^l \mathbf{u}) = \partial_z(S_{ls}\sqrt{T}\phi^s\phi^l). \tag{36}$$

Due to the structure of the equations the segregation process stops, if solely one phase is present. Then the phase moves with the bulk velocity of the mixture.

C. Combining the equations

The previously given segregation equations are a tool to compute the demixing of small and large particles. To do this, the flow field of the granular mixture is needed as additional information. This can be defined for simple flow fields, but for general industrial processes either it is not known or it cannot be described in an easy way. Therefore, the flow field can be computed by solving the Latz-Schmidt model. This leads to a one way coupling of the models. Therefore, the models have been introduced in the preferred form for the coupling.
With the help of the mixture mass balance (6) and the saturation condition (5) the volume fraction of the large particle phase is known from the volume fraction of the small particle phase and vice versa. Combining equation (35) with the saturation condition the Latz-Schmidt model needs to be extended by only one additional equation. The final system of equations can be summarized like in the following way:

$$\begin{aligned} \partial_t \Phi + \nabla \cdot (\Phi \mathbf{u}) &= 0, \\ \partial_t (\Phi \mathbf{u}) + \nabla \cdot (\Phi \mathbf{u} \otimes \mathbf{u}) &= \nabla \cdot (\eta \nabla \mathbf{u}) - \nabla p + \Phi \mathbf{g}, \\ \partial_t (\Phi T) + \nabla \cdot (\Phi T \mathbf{u}) &= \frac{3}{2}(\dot\gamma^2 - \nabla \cdot \lambda \nabla T) - \varepsilon \Phi T, \\ \partial_t \phi^s + \nabla \cdot (\phi^s \mathbf{u}) &= \partial_z(S_{sl}\sqrt{T}\phi^s(\Phi - \phi^s)), \end{aligned}$$

with the closures and relations given by equations (10)-(17).

Figure 1. Asymptotic size segregation result in a tumbler mixer. The small particles segregate to the centre. Whereas, the large particles concentrate in the periphery region. Experiment from [10]

III. Simulation of a mixing process

A. The rotating tumbler

Mixing is very important to guarantee the quality of many products produced in industry. A very common type of mixer is the rotating tumbler. There can be additional baffles in it of different shapes to improve the mixing process. Today, it is well known that in a plain cylinder the rotation of mixtures of particles differing in size leads to a radial segregation. For a half-filled tumbler the result is a mainly unmixed system where the small particles rapidly concentrate in the middle region and the large particles in the periphery. Only at the surface of the mixture the particles are in motion. All other particles are in a static state where no segregation can happen, which ends in the mentioned pattern. In the following this framework is used to test the previously presented model, as it needs to capture flowing regimes with segregation as well as static regimes without segregation. In [10] the authors analyse the segregation process in rotating tumblers and the influence of different baffles to mixing quality. Their experimental results serve as comparison. One of their results of a tumbler without baffles, half-filled with small and large particles, can be seen in figure 1.

B. Simulation results

For the numerical simulations the finite volume software CoRheoS has been used, developed at Fraunhofer ITWM, where the Latz-Schmidt model is already implemented. The existing model has been extended by the segregation equations given in the previous section. Details of the implementation of the Latz-Schmidt model can be found in [4] and [11]. The framework shows the cross section of a cylindrical geometry, that is half-filled with some sandy granular material with a density of 2650 $\frac{kg}{m^3}$ and a mean particle size of 0.2 mm. The material initially is in a perfectly mixed state, where half of the mixture volume is occupied by small particles and the other half by large particles. The geometry then rotates in a clockwise direction with a speed of 3 rounds per minute (rpm). Figure 2 shows the volume fraction of small

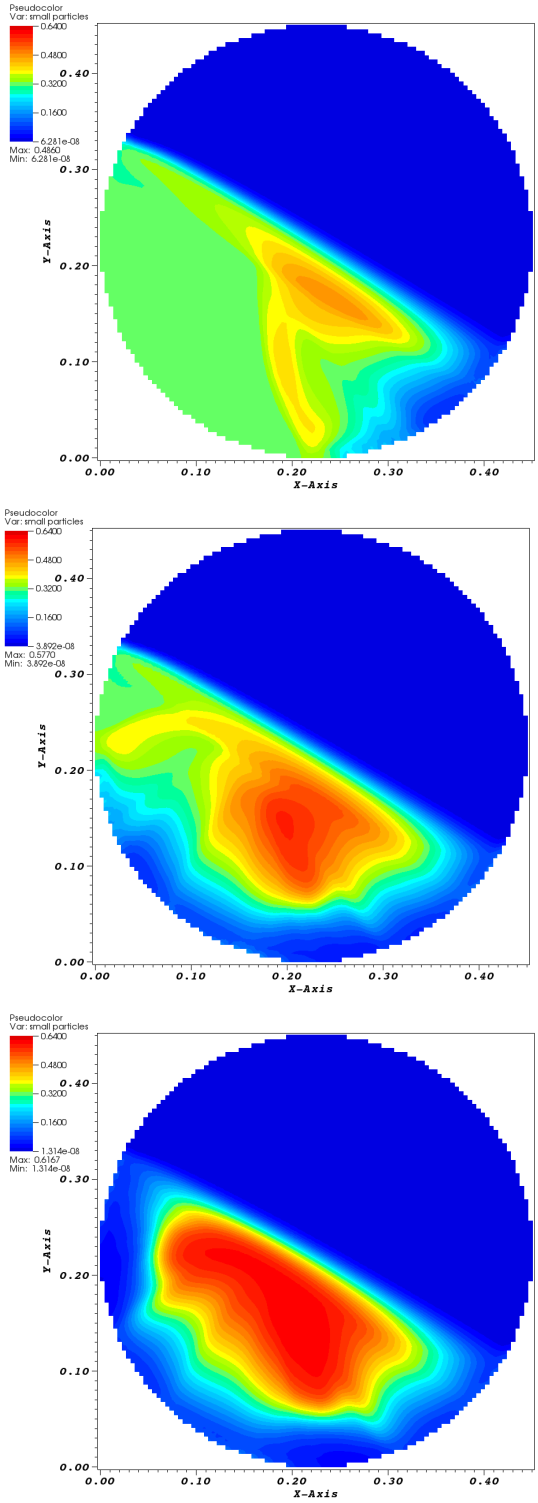


Figure 2. Volume fraction of the small particles after 0.25, 0.5 and 3 rounds in a half-filled rotating tumbler. The rotational velocity is 3 rpm. The colour table is fixed between 0% *(blue)* and 64% *(red)*.

particles in the granular system after 0.25, 0.5 and 3 rounds. One can see that a region of small particles forms rapidly in the centre of the mixture, as expected. It is experimentally proven that the maximum packing fraction lies between 63 % and 64 % for equal sized spherical particles. Consequently,

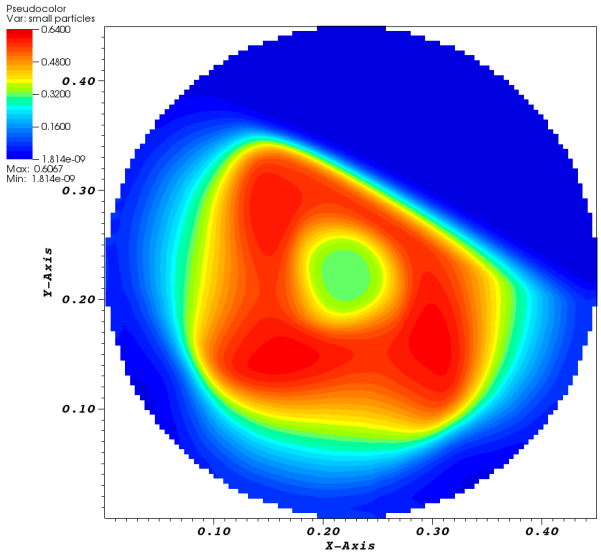


Figure 3. Volume fraction of the small particles after 3 rounds in an rotating tumbler. The tumbler is filled to a height of 70 % of the tumbler's diameter. The rotational velocity is 3 rpm. The colour table is fixed between 0% *(blue)* and 64% *(red)*.

this is the maximum value the small particle phase can reach in the centre region, as the left space is occupied by air. The maximal value of small particles reached in figure 2 is 48.6 %, 57.7 % and 61.7 % in the respective plots. For comparison figure 1 shows the asymptotic mixing result of the same process derived experimentally in [10]. For the parameters in the granular model the default values in the software CoRheoS where used. For the segregation equation $S_{sl} = 0.005$ has been chosen, which just indicates a segregation velocity between particles of different size. For a real application the value needs to be validated by experiments or DEM simulations for a set of particles of two specific sizes.

To test the model further, the same framework has been used, but the filling height of the tumbler has been changed. Figure 3 shows the simulation result after three rounds of rotation. The tumbler is filled up to 75 % with granular material. All parameters are chosen like in the first simulation. One can see, that after three rotations a ring of small particles has been formed. The large particles again concentrate in the periphery. Due to the bigger height of granular material, there is a region in the middle of the tumbler that stays untouched of the segregation process. This region stays in the initially mixed state during the whole simulation. This pattern formed in the chosen framework, matches to those given in the literature. A review about pattern formation in granular media is given by Aranson and Tsimring [12].

IV. Conclusion

In this work two models were presented, modified and combined. A set of parameters has been found for this model to reproduce the radial segregation effect occurring in a rotating tumbler filled with a randomly chosen material. Making the segregation equation depend on the granular temperature the segregation process can be regulated in the different regimes. The presented simulation results show that the segregation only happens in the avalanche layer at the surface of the granular mixture. Whereas, in the static regime the particles rotate with the tumbler without further segregation, as it is well known from experiments. Good simulation results have been achieved for the classical experiment of the half-filled tumbler as shown in figure 2. Recently, good results have been gained for simulating tumblers with different filling heights. One example is given in figure 3. Even if the simulations lead to good results, one should keep in mind, that the simulation of specific materials and the applicability of the model to other processes still need to be tested and validated by DEM simulations or experiments.

References

[1] J. Duran, *Sands, powders, and grains: An introduction to the physics of granular materials*, ser. Partially ordered systems. New York and London: Springer, 2000.
[2] J.-P. Bouchaud, M. E. Cates, J. R. Prakash, and S. F. Edwards, "A model for the dynamics of sandpile surfaces," *Journal de Physique I*, vol. 4, no. 10, pp. 1383–1410, 1994.
[3] ——, "Hysteresis and metastability in a continuum sandpile model," *Physical Review Letters*, vol. 74, no. 11, pp. 1982–1986, 1995.
[4] A. Latz and S. Schmidt, "Hydrodynamic modeling of dilute and dense granular flow," *Granular Matter*, vol. 12, no. 4, pp. 387–397, 2010.
[5] J. Bridgwater, W. S. Foo, and D. J. Stephens, "Particle mixing and segregation in failure zones—theory and experiment," *Powder Technology*, vol. 41, no. 2, pp. 147–158, 1985.
[6] S. B. Savage and C. K. K. Lun, "Particle size segregation in inclined chute flow of dry cohesionless granular solids," *Journal of Fluid Mechanics*, vol. 189, pp. 311–335, 1988.
[7] Gray, J. M. N. T. and A. R. Thornton, "A theory for particle size segregation in shallow granular free-surface flows," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 461, no. 2057, pp. 1447–1473, 2005.
[8] N. Brilliantov and T. Pöschel, *Kinetic theory of granular gases*, ser. Oxford graduate texts. Oxford: Oxford University Press, 2004.
[9] L. W. Morland, "Flow of viscous fluids through a porous deformable matrix," *Surv. Geophys. (Surveys in Geophysics)*, vol. 13, pp. 209–268, 1992.
[10] W. L. Vargas, S. K. Hajra, D. Shi, and J. J. McCarthy, "Suppressing the segregation of granular mixtures in rotating tumblers," *AIChE Journal*, vol. 54, no. 12, pp. 3124–3132, 2008.
[11] S. Schmidt, "On numerical simulation of granular flow." PhD-Thesis, University of Kaiserslautern, 2009.
[12] I. S. Aranson and L. S. Tsimring, "Patterns and collective behavior in granular media: Theoretical concepts," *Reviews of Modern Physics*, vol. 78, no. 2, pp. 641–692, 2006.

Holistic Surgery Monitoring of Angioplasty

Christina Gillmann, Thomas Wischgoll, Hans Hagen

Abstract—Cardiovascular diseases, mainly caused by lipids accumulating within the vessel wall creating plaque, are one of the leading causes of death in modern society. In advanced cases, surgery is required to avoid strokes or heart attacks. Therefore the aim of the presented approach is to assist in the entire surgery process, identify risks, discuss options, and examine the success. Unfortunately, the state of the art slice-by-slice reviewing method lacks in the ability to investigate CT scans in their entirety, as it only shows a single slice of the raw CT scan at a time. Advanced visualization approaches bring the potential to solve these problems, but are not designed to allow a retracing of the image generation process which is important to make trustworthy decisions. Therefore, this paper presents a holistic tool to insert visual exploration in all stages of the surgery process. An interactive linked view system process is introduced that contains a standardized tree view of the coronary arteries and additional supporting and analyzing views to orient oneself in space and analyze the recorded CT scan. As the visualization is derived as part of the medical workflow it addresses the user's needs and opens the black box of vessel visualization to obtain user approval.

Index Terms—Medical Visualization, Visual Exploration, Linked Views

1 Introduction

As studies show, cardiovascular diseases are one of the leading causes of death in modern society [25].

Most of them are caused by an accumulation of plaque in the coronary arterial system. This can result in reduced blood flow or blockage, especially if the afflicted area ruptures. The process is continuous (see Figure 1) and the more advanced it is the more dangerous it becomes. At a later stage of these diseases, the only viable option is to surgically open narrowed parts of the vascular system, called angioplasty [9].

As open heart surgery comes with a high risk, the overall aim is to visually monitor the entire process. To achieve this, a survey showed, that in medical doctors' point of view important aspects to improve on are the *Assessment of surgery*, *Teaching for future doctors*, *Safety during surgery*, and *Patient information* [19]. To achieve these aims a holistic visualization of patient records is required.

CT (Computed Tomography) scans are a common imaging method in medicine resulting in a stack of gray scaled images, each showing a recorded slice of the patient. State of the art of CT scan reviewing is a slice-by-slice approach where doctors scroll through the CT-scanned slices and build a three dimensional image in their head, which is a long process to learn. Although this method is very accurate as it displays the raw CT scan, it is very slow to use and not as effective in case of surgery monitoring as it does not allow measurements, annotations or comparative visualizations.

Advanced visualization approaches can be adapted to the surgery monitoring workflow, but most of them lack in the possibility to let the user retrace the image generation process. As medical decisions can have a huge influence on patients health, a suitable visualization needs to provide doctors with the possibility to judge on its correctness. If this is not the case, medical users often reject new visualization methods and stay with the slice-by-slice technique.

In this paper a holistic, open-box approach for planning angioplasty for coronary vessels based on CT-scanned data is introduced. Therefore, the coronary arterial tree is projected to a plane to provide the user with an overview of the recorded vascular system. In order to provide the user the to relate the visualization to the state of the art slice-by-slice reviewing method and to provide a spatial context supporting views are contained in the presented system. This allows a robust, easy to understand and trustworthy visualization for the surgery monitoring process. To monitor different stages of an angioplasty, analysis methods are given, where important features can be highlighted and the condition of vessels in surgery can be reviewed and compared in different CT scans.

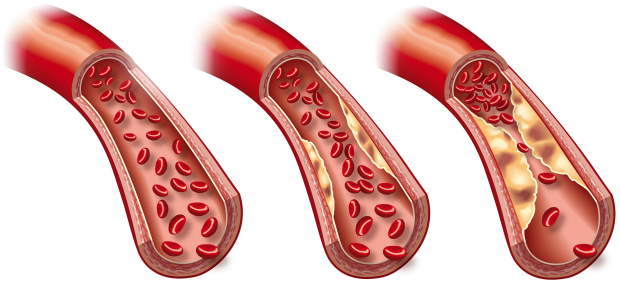


Fig. 1: Stages of atherosclerosis. Left: No plaque at the vessels wall. The blood flow is not affected. Middle: Starting plaque at the vessels wall. The blood flow starts to be compressed. Right: Massive plaque at the arteries wall. Blood flow is highly compressed and the risk of ruptures increases [23].

The paper is structured as follows: First, the state of the art in vessel visualization is reviewed in Section 2. After that, an overview over the application and the available data is given in Section 3. Based on the resulting requirements, the used visualization methods are shown in 4. To show the applicability of the presented system, results are given in Section 5 and discussed in Section 6. At last a conclusion and future directions will be presented in Section 7.

2 Related Work

The following section will discuss the state of the art in advanced vessel visualization and their application to surgery monitoring.

CPR (Curved Planar Reformation) relates to the flattening of a vessel along its centerline by sampling the vessel cross-sections [1]. The quality of the resulting planar gray-scale images are highly dependent on the chosen sampling strategies. In addition, the visualization of several vessels raise the question of their arrangement in a plane [11], [14]. The application of CPR is a basic tool in surgery monitoring [27] as medical users are trained to understand gray-scale images. Unfortunately, this method is not able to provide spatial context or allow visual exploration during the surgery.

In contrast to planar visualization approaches volume rendering can be used to create a three-dimensional visualization of the monitored heart. The main problem for this class of visualization is the applied transfer function. Especially in medical datasets they can be hard to find. Transfer functions assign a color and transparency to each voxel of the volume. Several approaches use additional knowledge obtained from cardiovascular modeling or simulations to create suitable and more accurate transfer functions [6], [8], [13]. Volume rendering can be applied to surgery planning and monitoring [2], [15], [21], [24] by rendering the raw CT scan or segmented structures as isosurfaces and provide interaction methods to navigate through the volume. Besides the induced occlusion problems of these techniques volume rendering approaches do not directly display the original dataset. This often results in rejection of new visualization techniques in medicine as the user cannot be entirely sure if the visualization shows the recorded dataset correctly.

To combine the benefits of volume rendering and CPR hybrid methods, such as focus & context [18], [26] or linked views [12], are available in vessel visualization. They are also available as commercial software [7], [10], [16], [20]. The suitability of hybrid methods highly depend on the methods used and the fashion of their connectivity. Examples show, that the combination of 3D and 2D techniques have a high potential, but for the scenario of surgery monitoring is not provided so far.

Therefore the aim is to provide a linked view system combining the benefits of 2D and 3D vessel visualizations and therefore provide the user with a holistic tool to visually explore CT scanned cardiac vessels during the complete process of surgery.

3 Application and Data

The following section will describe the surgery phases, their tasks for visual exploration, the available data in this scenario and the consequences for suitable visualizations.

3.1 Surgery Monitoring of Angioplasty

As mentioned before, a massive calcification can only be resolved via an angioplasty. The three main stages of a surgery are **(1)** Preparation, **(2)** Processing, and **(3)** Recovery [3].

Preparation includes the medication of the patient and the decision on how the surgery will be performed. In this phase it is important to locate and quantify the plaque in the cardiovascular system. Additionally, the size of the used tool as well as possible risks needs to be discussed.

Processing refers to the execution of the angioplasty. This means an expansion of the narrowed part in the vascular system. This is accomplished by the insertion of a catheter implanting a stent (small pipe), that is stretched by an inflatable balloon to keep the vessel open wider. During this phase, it is important to monitor the insertion of the tool to the vessel and the inflation of the balloon.

Recovery is the phase after the surgery where intricacies can occur and further treatments need to be planned. For this phase, it is important to compare a record of the vessel after the surgery with one from before to see if the aim of enlarging the diameter of the vessel was accomplished or not. Additionally, the successive treatment is planned.

For all stages the standard slice-by-slice reviewing method is not optimal as there are no mechanisms for the user to compare datarecords, investigate images from different points of views or a possibility to visually explore the CT scanned data. Therefore, suitable visualization are required to allow a holistic visual monitoring of angioplasty.

3.2 Available Data

The following section will list the properties of CT-scans and the geometric features that can be generated feature based on them.

3.2.1 Computed Tomography Scans

CT datasets are acquired by X-ray based scanners resulting in a volume containing a scalar value for each voxel. The intensity values are typically calibrated using the Hounsfield Unit (HU). Although CT scans are widely available in the medical domain, diagnosis of cardiovascular issues can be difficult as the images can contain several artifacts and inaccuracies due to, e.g., low resolution, motion, or image artifacts [22].

In this paper the CT scans have a resolution of 512x512 pixel per slice and around 200 slices in summary. Each resulting voxel cpatures a cube of 0.6x0.6x1.0mm.

The images are captured from a pig with previously surgically extracted hearts. A detailed description of the data acquisition can be found in the essay of Huo et al. [29].

3.2.2 Centerline

After recording the raw CT scan, features embedded within the 3D image can be extracted. This paper will focus on centerline data. The data available in this case was extracted with an algorithm presented by [28].

The centerline is a topological line-representation of the vascular system that is located in the center of the vessel [4]. This means that the centerline covers all branches and connects continuous components. For centerline extractions there are multiple classes of algorithms available that are summarized by Cornea et al. [5].

The algorithm used in the presented case is geometry-based and utilizes boundary information of the vessels as an input to calculate a Delaunay tetrahedrization of the vessel interior. Normal vectors of the boundary are used to generate a vector field in the vessel interior. For the resulting vector field extreme points, i.e. singularities, are identified. For each of these points, the extraction algorithm follows the vector field direction and checks if it can be connected with other points to a line.

This paper refers to an arbitrary centerline point as c_i with its values $c_i(p)$, the location of the point in space, and $c_i(r)$ the vessel radius at the examined point.

4 Visualization Techniques

According to the special needs of the medical user and the defined tasks in surgery monitoring, the resulting framework was designed as an interactive linked view system to capture the benefits of planar as well as 3D approaches. Therefore, three classes of views are provided: An overview containing a planar tree representation of the examined vascular system. Supporting views containing methodologies to orient in space and relate to the state of the art slice-by-clice reviewing technique and an analysis views to generate a timeline of recorded CT scans and examine and a stacked view to estimate the amount of plaque in a depicted part of the vascular system.

4.1 Overview

The planar tree visualization of the vascular system was designed to give the user an overview over the examined vascular system and identify spots where further analysis is required. Figure 2 gives an overview over the tree generation steps, that will be described in the following.

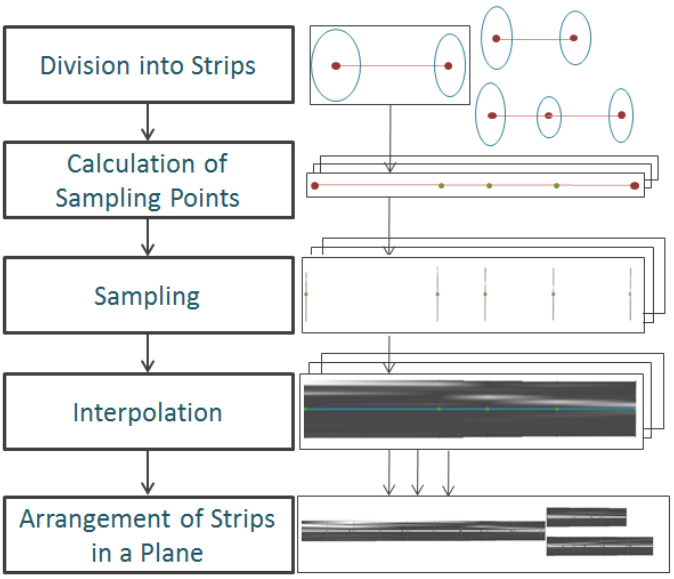


Fig. 2: The five computational steps to create the presented tree-view of the presented visualization.

4.1.1 Division of Centerline into Strips

The input dataset can contain several subtrees of the vascular system as well as unconnected components resulting from the centerline extraction algorithm.

After the root node is selected by the user, the algorithm follows its neighborhood information to capture all points that are connected in this subtree. While doing this, the points are clustered into *Strips*. Two centerline points c_i, c_j belong to one *Strip* if there is a path from c_i to c_j and the number of neighbors for each point on this path is not higher than two. With this separation it is possible to perform the following calculation steps without involving special cases for the graph.

Each *Strip* with n points consists of $n-1$ Segments *Seg*. A Segment describes the two neighbored centerline points of a strip and holds information about the applied sampling points, length ($Strip_{i,j}(l)$) and the segment direction ($Seg_i(d)$). The calculation of this information will be shown in the following.

4.1.2 Calculation of Sampling Points

In many cases the extracted centerline is not dense enough to be used as sampling points solely. Consequently, there are additional sampling points required. Although a constant sampling rate assignment would be straight forward it lacks in the ability to adjust the number of points to the length of the segment what would result in unequally sampled strips. To solve this problem a distance-based rate assignment was used in this case. This means that the rate of a segment is higher if the segment is longer. More formal, this can be formulated as:

$$Seg_i(r) = f * Seg_{i,j}(l) * \frac{1}{\Sigma} \tag{1}$$

with $Seg_{i,j}(l)$ as the current length of the segment and Σ as the sum of all segment length. The calculated rate can be controlled by a global scaling factor f. As a result, the longer a segment the more sampling points will be assigned. Each segment will be equally divided by its number of rate.

4.1.3 Sampling

For each centerline point and sampling point the cross-sectional area can be defined. This area is the plane in space that intersects the sampling point and has the direction $Seg_i(d)$ as its normal vector. For each cross-section the algorithm performs a CPR that samples the CT volume.

To create a direction independent sampling, a circular approach can be used [17]. In this case circles with increasing size surrounding the center point are sampled and the maximum calculated. To avoid different sampling rates on circles with different sizes the points are set with a user-defined distance on the circles.

As this approach reduces each circle to a point (the maximum), sampling information is lost. To solve this problem, the Sampling is mirrored at the middle line and displayed again with the mean value of the sampled circle. This covers a wider range of values and gives a better impression of the cross-sectional area (see Figure 3).

4.1.4 Interpolation

After each sampling point is evaluated the resulting images are arranged in a line, depending on their position in the segment. The gaps between the sampled lines are interpolated according to their position in the trapezoid formed by the previous and following sampled values.

4.1.5 Arrangement of Strips in a Plane

Once all single *Strips* are calculated they need to be arranged in a two-dimensional plane. In contrast to other cases, this visualization requires the preservation of the edge length. To obtain a better impression of the vessel length a parallel view is chosen.

For parallel alignment the algorithm starts by assigning a box ($Strip_i(b)$) to each strip, starting with the root. For each leaf the height of the corresponding image will be defined as its box. After that, the box of each node will be obtained from the sum of widths from all of its children. After all strips have an assigned box they can be arranged in the plane starting from the root. Each strip will be placed in the center of its assigned box followed by recursively drawing its children by dividing its size to the child strips.

4.2 Supporting Views

During the process of tree-view generation, spatial information is lost as well as not all sampled points can be displayed. Therefore, two additional views are added to restore this information.

4.2.1 Volume View

The volume view aims to add the spatial context that is lost in the tree view during the sampling process. In this view, the CT scan is rendered with a one-dimensional transfer function. This transfer function ranges from dark gray (small values) over light gray (user defined threshold) to red (light values), what can be seen in Figure 3. As light values indicate plaque in the vessel, this results in a highlighting of plaque in the volume view. To follow the surgery path in space, the selected start and end point as well as their connecting path can be displayed.

The volume view is able to point to the corresponding location in space selected by the user in the tree view. This helps the user to get a feel for the geometric arrangement of the vasculature within the tree view.

4.2.2 Slice View

The slice view is an additional option to help the user following the workflow of the algorithm. It shows the cross-sectional area of a depicted centerline point in the CT volume. As doctors are used to review the gray scale slices this provides a correlation of the tree view to their previous used reviewing method.

Additionally, for the selected cross-section the centerline point, sampling points, and radius will be added to show how the cross-section is sampled. This allows the user to control the CPR settings and see if the vessel is covered properly. In the case of over or undersampled vessels or shifted centerline points, the user can correct the sampling settings and therefore enhance the resulting visualization.

As the retracing of image generation is an important point in the medical domain, this view helps doctors to understand the visualization process and therefore increase the acceptance of the presented approach.

4.3 Analysis Views

The analysis view aims to provide the user with the possibility to create data records for selected parts of the vascular tree and allow him to examine and compare them in different stages of the angioplasty.

The view contains the selected part of the vascular tree embedded in a timeline view to monitor its changes over time. Several interaction methods as well as a stacked view for better vessel estimation are available.

4.3.1 Event View

The event view contains a timeline, where the user can mark different events. For an event, the user has to select a time step and an arbitrary part of the vascular system. In the event view, the depicted part will be displayed according to its position on the timeline and the user can define which kind of event is displayed (Preparation, Processing and Recovery). The system can be linked to the patient's history showing the users additional information as medication or prior surgeries.

To highlight interesting features in surgery planning several features can be added to the CPR. First, it is possible to add a scale to each strip starting from the root and marking the distance to it. The diameter is shown on this scale as it can be seen in Figure 3 in blue lines. The calculated sampling points can be visualized with green dots on the middle line.

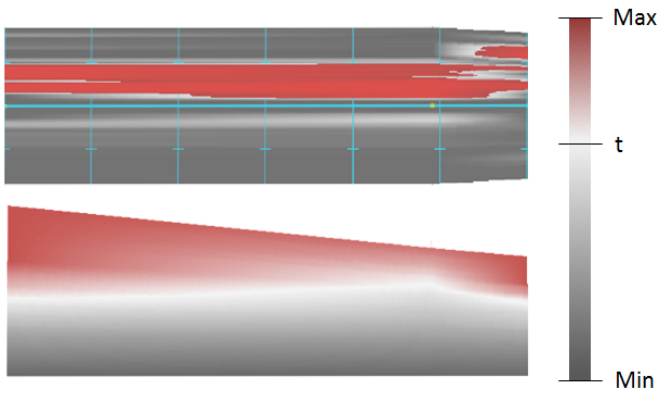


Fig. 3: Example of an analysis view. Top: Selected part of tree view with a highlighting of all values greater then 190, the added scale and the sampling points. Bottom: Stacked view of the selected CPR.

Additionally, to highlight calcification, the user is able to choose an isovalue. All gray scale values on the resulting visualization that are higher then the depicted isovalue will be colored in red with an opacity value depending on the gray-scale value. This results in darker red colored voxels the higher the gray-scale value what can be seen in Figure 3.

At last, the system allows the user to add annotations, thus the doctor can mark parts of interest and note some important findings from the planning.

4.3.2 Stacked View

As the generation of a CPR causes a loss of information due to the sampling and mapping process, it is possible that the resulting visualization does not show the exact amount of calcification. For a trustworthy visualization that can be used to plan an angioplasty this information is required. Therefore, the stacked view is introduced to provide the user with a control mechanism over the sampling process. The view corresponds to the selected part of the overview.

In this view, all sampled values of a cross-section are visualized as a stacked graph. Therefore all sampled values are sorted by their gray scale value (starting with the smallest value on the bottom). To make this visualization comparable to the chosen sampling strategy in the event view, the stacked view shows the selected event. To allow a better comparability, the same color scheme as in the event view is used.

With this view it is possible to compare the exact amount of sampled points in a specific range and compare the result to the event view. As Figure 3 shows, the CPR displays a nearly equal amount of plaque over the whole vessel. In comparison to that, the stacked view detects that there are different amounts of plaque along this vessel. This can help the user to decide exactly where a surgery needs to be performed.

With a combination of all presented views, a holistic visual tool can be build to allow the monitoring of a surgery for coronary arteries in all stages.

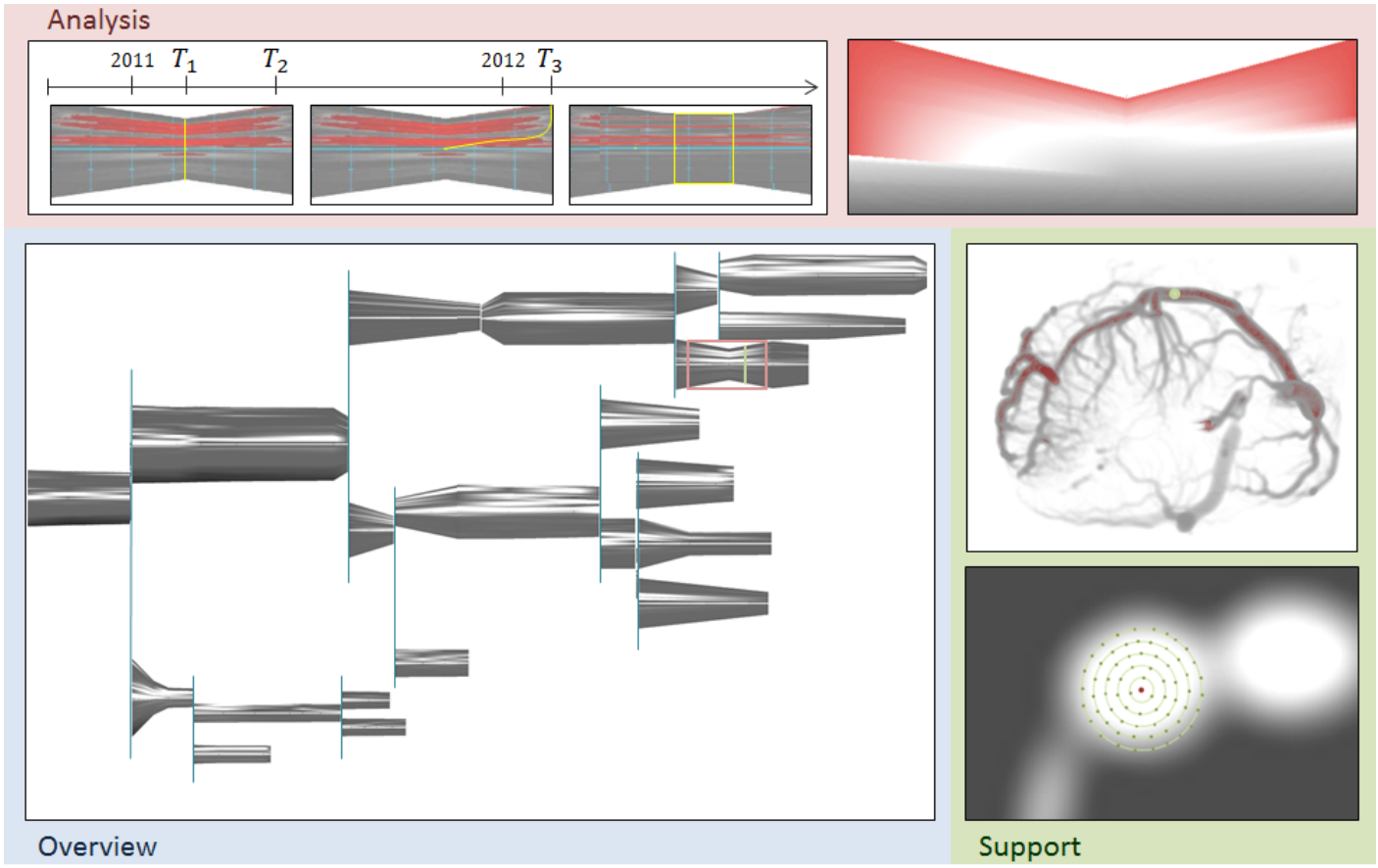


Fig. 4: A possible real world example visualized with the presented system. Blue: The entire right coronary arteries tree of a pig. The pink box indicates the selected vessel part for a surgery. Red: Visual exploration for the three stages of surgery on the timeline (preparation, process and recovery) applied to the selected vessel part. The stacked view shows all sampling points of the dataset visualized in the preparation step. Green: Supporting views for for orientation and review of image generation process. The volume view holds the depicted point in the tree-view, whereas the slice view shows the cross-section of the depicted point.

5 Results

In the following, a potential real world example will be discussed to prove the applicability of the presented system. Figure 4 shows the presented visualization applied to the right coronary arteries of a pig that is highly affected by calcification.

The tree view is shown with branching vessels connected by a blue bar. The pink frame indicates the selected calcification that needs to be treated. Although there are further affected areas this part is narrowed in addition to the plaque. The supporting views (green frame) show the volume rendered heart and the depicted slice. The orientation in space is very intuitive with the volume view and the corresponding cross-section can be reviewed with the slice view.

The analysis view (red box) holds three events (T_1, T_2 and T_3) where T_1 is the preparation, T_2 is the processing and T_3 is the recovery process.

The three boxes show the corresponding sampling, based on the user's choice. Gray scale values higher then 190 are highlighted with the red colorscheme to identify calcification.

With the chosen red color scheme and the parallel alignment of the strips, this combination of features is easy to detect.

In the preparation process (T_1), the current diameter of the vessel is examined (1.62 cm), which can be seen via the yellow bar. With this information, the medical doctor is able to choose a tool with a proper size and decide how to perform the surgery the best way. The additional stacked view shows, that the selected area needs to be treated slightly shifted to the left. This was not able to be detected by the timeline view alone, which makes the stacked view an important feature to analyze the vessel of interest.

During the actual surgery process (T_2), the planned surgery path can be monitored, as the yellow line shows. As the visualization is capable of showing the entire vascular system, surgical risks and other affected vessels can be discussed.

At last, the recovery process (T_3) can be monitored by obtaining a new CT scan of the vascular system and the calculation of the corresponding centerline. In this example it can be seen that the diameter was increased by the angioplasty to a diameter of 1.8. Although this is a small change the direct comparison of the vessel before and after the surgery shows, that the obstruction was reduced thereby increasing blood flow.

6 Discussion

As mentioned before, surgery monitoring aims at enhancing several aspects. This section discusses how these can be achieved.

The *assessment of surgery* and *number of complications* can be improved upon with the presented visualization. As the example showed it is possible to visually explore the surgery process in all stages, which is not possible with other techniques. With the

filtering tool, all parts of the visualization that exceed a certain diameter can be skipped. This allows the medical doctor to try several scenarios of different surgical tools and find calcified vessels. The technique is well suited for medical use as it links to the state of the art technique which is very accurate. Thus, the image generation process can be controlled and therefore the black box issue of visualization is avoided. With the additional 3D view, the spatial context of the selected points is ensured. To amplify this results and find potential of improvement a user study is required. In contrast to the slice-by-slice reviewing method alone this visualization gives the user more possibilities in planning surgeries.

The *Teaching of future doctors* and *Patient Information* have a different target group. Here, the addressed users are not strongly related to the state of the art visualization technique but need a quick and easy to understand overview over the planned process. The stacked view is a good method to show the flaws of CPR and therefore help students in medicine better understand it. For future doctors, the visualization can be used to setup a database with performed and recorded surgeries. With this database medical students can be taught frequently occurring problems during surgery or the general process of planning. With the presented visualization it is possible to show the patient the planned procedure and inform him or her about the risks and planned stages in the surgery.

7 Conclusion and Future Work

In this paper we presented an interactive, standardized linked view visualization to visually monitor all stages of an angioplasty. As the design of the visualization follows the workflow of the medical user and related to the state of the art technique, this work presents an open-box approach. The result is a holistic visualization framework that is capable of guiding the medical user through all stages of the surgery.

This work provides new insights into CT scans and their features. It enhances the surgery monitoring workflow with existing knowledge. The presented example shows how important the study of the addressed user is to create suitable visualization and obtain user acceptance.

As future work it is planned to extend the presented visualization to further tasks in vascular diagnosis and provide a general framework that is able to go beyond current datasets, such as simulated pressure values of the coronary arterial tree. Additionally, the clinical applicability will be tested in a formal user study.

Acknowledgments

We would like to thank Ghassan Kassab's research team for providing the datasets used in this study, including Jenny S. Choy, Yunlong Huo, and Tong Luo.

References

[1] G.-P. Bonneau, T. Ertl, and G. M. Nielson, editors. *Scientific Visualization: The Visual Extraction of Knowledge from Data*. Springer, 1 edition, 2006.

[2] A. M. Chiu, D. Dey, M. Drangova, W. D. Boyd, and T. M. Peters. 3-D Image Guidance for Minimally Invasive Robotic Coronary Artery Bypass (MIRCAB).

[3] L. W. Cohen, Frances Poon. Coping with surgery: Information, psychological preparation, and recovery. pages pp. 375–382, 1980.

[4] R. L. Cook. Shade Trees. In *Computer Graphics*, volume 18, page pp. 210226, 1984.

[5] N. D. Cornea and D. Silver. Curve-Skeleton Properties, Applications, and Algorithms. In *IEEE Transactions on Visualization and Computer Graphics*, volume 13, pages 530–548, 2007.

[6] C. Correa and K.-L. Ma. Size-based Transfer Functions: A New Volume Exploration Technique. volume 14, pages 1380–1387, 2008.

[7] Fujifilm. Synapse 3d, Aug. 2015.

[8] C. Gillmann and P. Salz. Improving Visual Communication for EIT-Based Lung Research. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE*, volume 10, pages 291–295, 2014.

[9] A. R. Grntzig, k. Senning, and W. E. Siegenthaler. Nonoperative dilatation of coronary-artery stenosis. *New England Journal of Medicine*, 301(2):61–68, 1979.

[10] G. Healthcare. Autobone and vesselqexpress, Aug. 2015.

[11] A. Kanitsar, R. Wegenkitti, D. Fleischmann, and M. E. Gröller. Advanced Curved Planar Reformation: Flattening of Vascular Structures. In *IEEE Transactions on Visualization and Computer Graphics*, volume 1, pages 99–108, 2012.

[12] C. Kubisch, S. Glaer, M. Neugebauer, and B. Preim. Vessel Visualization using Curved Surface Reformation. In *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE Scientific Visualization 2013)*, volume 19, pages 2858–2867, 2013.

[13] G. Läthén, S. Lindholm, R. Lenz, and M. Borga. Automatic Tuning of Spatially Varying Transfer Functions for Blood Vessel Visualization. In *IEEE TVCG (SciVis Proceedings)*, volume 18, pages 2345–2354, 2012.

[14] N. Lee and M. Rasch. Tangential Curved Planar Reformation for topological and orientation invariant Visualization of Vascular Trees. In *EMBS Annual International Conference*, volume 28, 2006.

[15] N. J. B. McFarlane, X. Lin, Y. Zhao, G. J. Clapworthy, F. Dong, A. Redaelli, O. Parodi, and D. Testi. Visualization and simulated surgery of the left ventricle in the virtual pathological heart of the virtual physiological human. 1(3):374–383, 2011.

[16] Millensys. Vision tools 3d views, Aug. 2015.

[17] G. Mistelbauer, A. Morar, A. Varchola, R. Schernthaner, I. Baclija, A. Kchl, A. Kanitsar, and E. G. S. Bruckner. Vessel Visualization using Curvicircular Feature Aggregation. volume 32, pages 88–93, 2013.

[18] S. G. Nikolov, D. Agraotis, D. R. Bull, C. N. Canagarajah, and M. G. Jones. Focus+Context Visualisation for Fusion of Volumetric Medical Images. In *Proceedings of the 4th International Conference on Information Fusion*, volume 1, 2001.

[19] K. J. Oldhafer, B. Preim, C. Drge, H. O. Peitgen, and C. E. Broelsch. Acceptance of computer-assisted sur gery planning in viscreal (abdominal) surgery - result s of a nationwide survey. 127(2):128–33, 2015.

[20] Orobix. Vmtk- the vascular visualization toolkit, Aug. 2015.

[21] L. D. Paolis, M. Pulimeno, and G. Aloisio. Visualization and interaction systems for surgical planning . In *Information Technology Interfaces (ITI)*, volume 32, pages 269 – 274, 2010.

[22] B. Preim and D. Bartz. *Visualization in Medicine*. Elsevier, 2007.

[23] O. Soehnlein, M. Drechsler, Y. Dring, D. Lievens, H. Hartwig, K. Kemmerich, A. Ortega-Gmez, M. Mandl, S. Vijayan, D. Projahn, C. D. Garlichs, R. R. Koenen, M. Hristov, E. Lutgens, A. Zernecke, and C. Weber. Distinct functions of chemokine receptor axes in the atherogenic mobilization and recruitment of classical monocytes. *EMBO Molecular Medicine*, 5(3):471–481, 2013.

[24] L. Soler, H. Delingette, G. Malandain, N. Ayache, C. Koehl, J. M. Clement, O. Dourthe, and J. Marescaux. An automatic virtual patient reconstruction from ct-scans for hepatic surgical planning. *Stud Health Technol Inform*, 70:316–22, 2000.

[25] P. Stang, A. Carson, K. Rose, J. Mo, S. Ephross, E. Shahar, and M. Szklo. Headache, cerebrovascular symptoms, and stroke: the arteriosclerosis Risk in Communities Study. In *Stroke*, volume 11, pages p. 2718–2721, 2005.

[26] M. Straka and M. Cerven. The VesselGlyph: Focus and Context Visualization in CT-Angiography. In *Eurographics Conference on Visualization*, volume 32, pages 88–93, 2013.

[27] R. H. Taylor, S. Lavealle, G. C. Burdea, and R. Mosges, editors. *Computer-Integrated Surgery: Technology and Clinical Applications*. MIT Press, Cambridge, MA, USA, 1st edition, 1995.

[28] T. Wischgoll. Modeling and Visualization of Cardiovascular Systems. In *Dagstuhl Publishing*, page pp. 210226, 1998.

[29] T. W. Yunlong Huo, Jenny Susana Choy, T. Luo, S. D. Teague, D. L. Bhatt, and G. S. Gebab. Computed tomography-based diagnosis of diffuse compensatory enlargment of coronary arteries using scaling power laws Interface. In *Royal Society*, volume 10, pages H514–H523, 2002.

A stochastic model featuring acid induced gaps during tumor

Sandesh Athni Hiremath
Felix Klein Center of Mathematics
Technical University of Kaiserslautern
Paul-Ehrlich-Str. 31, 67663 Kaiserslautern, Germany
Email: hiremath@mathematik.uni-kl.de

Christina Surulescu
Felix Klein Center of Mathematics
Technical University of Kaiserslautern
Paul-Ehrlich-Str. 31, 67663 Kaiserslautern, Germany
Email: surulescu@mathematik.uni-kl.de

***Abstract*—Here we take an ecological and evolutionary stance and investigate a particular behavior of cancer. The behavior in mind is the formation of gaps and infiltrative patterns mediated by acidic extracellular environment. We setup a phenomenological model for the above behavior by building upon the previously existing models. In this model acidity plays the role of natural selector giving rise to more resistant and invasive cancer progeny. This new model not only predicts the occurrence of gaps between the tumor and the stroma but also elucidates several features of local invasion like formation of spikes, buds, islands, and cavities. We use semigroups method to prove the wellposedness of the model and perform 1D and 2D numerical simulations to illustrate the model predictions and draw conclusions based on the observed behavior.**

I. The link between evolutionary biology and cancer

A. Prologue

According to the theory of evolution, multicellular organisms have evolved from unicellular ones, over a period of millions of years, through the random processes of genetic alterations and natural selection. The tree of evolution clearly indicates a bottom up process where the unicellular organisms are at the root while multicellular organisms occupy their place in the trunk and leaves of the tree. This is the case because genetic/chemical alterations and environmental forces propelled unicellular organisms to aggregate and start to collaborate which then gave rise to a functionally more sophisticated life units like tissues, organs and in turn organisms. There are also cases where complex multicelluar organisms have evolved (or de-evolved) to simpler forms of life like viruses for e.g. certain species of Cnidaria have de-evolved from multicellular free-living autonomous animals into parasitic organisms comprising of only a handful of cells. Such parasites constitute to the class of organisms called Myxozoans [19]. This unique reverse evolution is again due to genetic alterations and environmental selection.

Cancer is a disease in which some of the cells within a multicelluar life unit (e.g. tissue or organ) have functionally turned into a more primitive life form. What this means is that tumor cells are no more adhering to the rules of their pre-destined functionality but instead they become rouge cells acting in a primitive and selfish way which is often deleterious to the multicellular life unit that it is still a part of. A simple e.g. is the case of melanoma (a type of skin cancer) the pigment producing cells called melanocytes undergo transformation (via mutations induced mainly through exposure to UV radiations) and are no more just responsible for melanine production but they start exhibiting the features of an independent unicellular organism. This means that melanocytes after being carcinogenized start replicating profusely, start to boost its nutrition supply, may even start to migrate, etc. All these are features of an autonomous living organism (say like protozoans). Thus carcinogenesis is, in a very vague sense, similar to the de-evolution of Cnidarians to Myxozoans, with the main difference being that carcinogenesis induces a switch from a stable, cooperative (altruistic) and functionally pre-determined behavior to an instable, selfish and functionally free behavior while the de-evolution process of Cnidarians to Myxozoans involves a rather drastic physical transformation. This novel analogy can be backed up by the biological finding that many protozoan life cycles features have been observed in mammalian cancers undergoing polyploidy [20].

Another aspect of cancer is that it not a modern disease, instead its an ancient one, whose roots go back to the time of the beginning of the cellular life itself. In the early stages of unicellular life, natural selection favored organisms that were fast growing, space invading and had low mortality rate etc. However, for the multicellular to evolve there had to be a paradigm shift, i.e. natural selection had to favor cooperative cells over selfish ones. According to [2], [20] this can be a reason for the rarity of cancers in multicellular organisms and also a reason

why the occurrence of cancer represents a reversal from altruistic behavior multicellular cells to selfish behavior unicellular cells.

A Conventional approach for understanding cancer is to look for cellular features that are specific to neoplastic cells. Hanahan and Weinberg [11] provide a fairly comprehensive review on the cellular capabilities that characterize a cancerous cell. These are namely: 1. Excessive proliferation, 2. Evasion of apoptosis, 3. Angiogenesis, 4. Metastasis, 5. Replicative immortality, 6. Genetic instability, 7. Reprogramming of energy metabolism, etc. Though crucial, just these cellular mechanisms are, however, not sufficient for understanding the disease. To complete the picture one would also need to understand the role of microenvironment and look at cancer from an ecological and evolutionary perspective. This is a currently growing trend in cancer research and cancer therapy. The reason for this being that: current approach in cancer therapy (chemotherapy) often leads to the selection of more resistant and malignant strains of cancer. This is analogous to the modern ecological problem where spraying of pesticides and insecticides leads to the emergence of a more resistant strain of pests. Another common example is the emergence of drug resistant viruses due to genetic variations. Thus, from the ecological perspective, a successful treatment of cancer means extinction of certain class or family or species of neoplastic cells. Therefore, the study of cancer as an ecosystem and understanding the role of the microenvironment of a heterogeneous neoplastic cells is very important.

Having emphasized the importance of evolutionary and ecological viewpoint of cancer, we now come to main topic of this paper, i.e. to study the role of acidity in the migratory and malignancy property of cancer.

B. Microenvironmental role of acidity

Protons (H^+) are one of the most reactive molecule in a living cell. The amount of active protons (and proton equivalents) present in the solution determines the acidity of that solution and is expressed in terms of pH. Intracellular pH ($\mathrm{pH_i}$) is a crucial regulator/inhibitor/activator of cellular functions. It influences many cellular processes like protonation and de-protonation events, structural configuration of proteins, movement of ions and solutes across membrane, exitability of membranes, cell contractivity, etc. [4], [26], [23]. Normal cells usually maintain their $\mathrm{pH_i}$ well within the range of 7.3 to 7.4. This $\mathrm{pH_i}$ homeostasis is possible due to several regulatory mechanisms viz- 1. intrinsic and mobile buffers, 2. membrane transporters like NHE, AE, NDCBE, NBC, MCT, etc. , 3. proton sequestration by cell organelles in order to maintain their own pH levels. Apart from the intracellular $\mathrm{pH_i}$ it is also highly important that the extracellular pH ($\mathrm{pH_e}$) is regulated (via buffers, vasculature, diffusion etc) and is maintained at a higher level compared to $\mathrm{pH_i}$, i.e. in a normal adult differentiated cell $\mathrm{pH_i}$ is lower than $\mathrm{pH_e}$.

In early 20th century Warburg observed that tumor cells consistently use glycolysis for their energy requirements and that their extracellular pH is relatively smaller [29]. There are many theories as to why exactly this behavior has emerged. Among them, a strongly viable hypothesis is that (see for e.g. [10], [21]) due to high density of a neoplastic tissue some of the cells experience ischemia due lack of profuse vasculature which consequently results in insufficient oxygen supply. Due to this the inhabitant cells utilize the secondary pathway for energy production namely glycolysis. Now because of genetic variation within the neoplastic tissue some of the cells with enhanced acid extruders are able to expel the intracellular protons to the extracellular side thereby making the extracellular acidic. The ones which are unable to expel the acid succumb to intracellular acid accumulation, while the surviving ones are subject to high acidic extracellular environment. This inturn selects for cells that are capable of surviving in this harsh acidic environment. Altogether, the microenvironmental factors like overcrowding of cells, low oxygen supply, high extracellular acidity, etc. selects for cells capable of developing aerobic glycolytic phenotype i.e. cells that mainly use glycolysis for their energy requirements even under normoxic conditions. Thus over a period of time and after some cycles of replication, the progenitor cells of a neoplastic tissue posses a glycolytic phenotype and it is characteristic of such cells that they consistently maintain a reverse pH gradient i.e. $\mathrm{pH_i}$ is higher that $\mathrm{pH_e}$ as opposed to a normal tissue [30]. With this in mind we next briefly introduce the cancer dynamics and build a phenomenological model for the formation of gaps and infiltrative patterns mainly driven by micorenvironmental acidity.

II. Introduction

Cancer in nothing but an ecosystem of neoplastic and stromal cells inhabiting the tissue containing resources for cell sustenance. Using this ideology we model (see III) the evolution of cancer cell and stromal cell population mediated by the cellular acidity and tissue. Malignancy of cancer can be qualitatively categorized based on its strength (or severity) of tissue invasion (or destruction) which in turn can be judged based on the histological patterns generated by a progressing

tumor. Infiltrative growth pattern (INF) classification is one such classification scheme, as defined by the *Japanese Gastric Cancer Association* [1], [14]. Based on this, invasion can be assigned to three categories, namely (a) INFa : An expanding tumor core with a clear separation between its boundary and the stromal cells, (b) INFb : An intermediate stage of tumor expansion with or without clear separation from the stromal cells and (c) INFc : An expanding tumor core whose boundary cells have infiltrated the stromal cells and there is mixing of tumor and stromal cells. The INFa marks the appearance of a gap between cancer cells and stromal cells. This gap may be composed of the cellular byproducts secreted mostly by the cells on the progressing boundary of the tumor core. This progression can be either due to cell proliferation or due to spatial movement induced by random motion or taxis [25], [6]. In either case the cells rely on glycolysis for energy production [28], [27], [3], as a result of which the gap consists of acidic metabolites. The acidic contents enhances in turn cell motility [22] by making room for movement through degradation of ECM fibres [24], [9] and inducing apoptosis in normal cells [16]. Motivated by this we rely on the *go-or-grow* hypothesis (see [5], [7]) to propose a highly nonlinear stochastic model for the gap formation between cancer and stromal cells. Here we use acidity as the key ingredient regulating the appearance and disappearance of a gap. Also the nonlinear coupling between the proton dynamics and cell-population dynamics leads to interesting infiltrative patterns of tumor cells.

III. Modeling

In this section we briefly explain the modeling procedure (refer [13] for a more detailed explanation of the model setup and illustration of the involved coefficient functions). The four variables in which we are interested are as follows: (a) C denotes the population density of cancer cells; it is measured in $cells/vol$, (b) N is the population density of stromal (normal) cells; it is measured in $cells/vol$, (c) H_i represents the concentration of protons in the intracellular region of a cancer cell; it is measured in Mol/vol and (d) H_e is the concentration of protons in the extracellular region due to cancer cells; it is measured in Mol/vol. The system of equations describing the dynamics of the proton concentrations and cell population densities is as follows:

$$\begin{aligned}\frac{d}{dt}H_i &= J(C/K_C)[-T_1 - T_2 + T_3 - Q + q_1] \\ &\quad + \gamma_\xi J(C/K_C)\frac{H_i}{K_w}\xi_t \qquad (1)\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial t}H_e &= J(C/K_C)[T_1 + T_2 - T_3] - q_2 H_e + \gamma_{_D}\Delta H_e \\ &\quad + \nabla\cdot\Big(\gamma_g g(\frac{H_i}{K_w},\frac{H_e}{K_w},\frac{C}{K_C})\nabla C\Big) \\ &\quad + \nabla\cdot\Big(\gamma_h h(\frac{H_e}{K_w},\frac{N}{K_C})\nabla N\Big) \qquad (2)\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial t}C &= C(1-\frac{C}{K_C})(\Lambda_1(\overline{H}_i,\overline{H}_e)+\Lambda_2(\overline{H}_i,\overline{H}_e)) \\ &\quad + \nabla\cdot\Big(\gamma_a a(\overline{H}_i,\overline{H}_e,\overline{C},\overline{N})\nabla C\Big) \\ &\quad - \frac{\gamma_b}{K_w}b(\overline{H}_i,\overline{H}_e)\nabla H_e\cdot\nabla C \qquad (3)\end{aligned}$$

$$\begin{aligned}\frac{d}{dt}N &= \Big(\gamma_{\Lambda_4}\Lambda_4(\overline{H}_e,\gamma_{\Lambda_{4,1}}) - \gamma_{\Lambda_3}\Lambda_3(\overline{H}_e)\Big)\Big(N-\frac{N^2}{K_N}\Big) \\ &\quad - \gamma_{_N}\frac{C}{K_C}N, \qquad (4)\end{aligned}$$

where $\overline{H}_i$, $\overline{H}_e$, $\bar{C}$ and $\bar{N}$ are the nondimensional analogues of H_i, H_e, C and N respectively.

A. Intracellular proton dynamics (IPD) described by (1):

The terms in (1) describe the following: (a) T_1 and T_2 represent the efflux of intracellular protons across the cell membrane due to NHE and NDBCE transporters, respectively. (b) T_3 represent the influx protons (i.e. cytosol acidification) due to AE (Cl^--HCO_3^- or *a*nion *e*xchanger). (c) The production rate of acid due to metabolic activity (mainly aerobic glycolysis) is represented by q_1. (d) The sequestration of acid by intracellular organelles is represented by H_i dependent function Q. (e) ξ_t is a stochastic process which phenomenologically accounts for intracellular fluctuations affecting the proton dynamics. Such fluctuations can occur due to: (i) uncertainty of the effects of various other biochemical process, (ii) random biochemical process like gene expression, random behavior of membrane transporters, etc. (f) The function J that depends on cancer density is a flux modulation function, which ensures that the proton dynamics is dominant on the proliferating edge of the tumor.

B. Extracellular proton dynamics (EPD) described by (2):

The terms in (2), accounting for extracellular proton dynamics (EPD), are as follows: (a) The functions T_1, T_2 and T_3 are the same as described above. Here they appear with an opposite sign since they represent the efflux of intracellular protons which inturn means influx of extracellular protons. (b) q_2 represents the dissipation of protons through the blood vessels.(c) The Laplace operator models the random movement of extracellular protons. (d) The nonlinear operators

$\nabla \cdot (g \nabla C)$ and $\nabla \cdot (h \nabla N)$ for non-negative functions g and h characterize the effect of proton repulsion from highly dense regions of cancer and normal cells towards less dense regions.

C. Cancer cell population dynamics (CPD) described by (3):

The growth of the cancer cell population is characterized by the intra-species competition term $C(1 - C/K_C)$, where K_C is the carrying capacity of the cancer cell population. This logistic-type growth is modulated by a proliferation function Λ_1 and a recession function Λ_2. The non-linear diffusion and the advection terms model the movement of cancer cells due to acid induced cyto-skeleton remodelling and availability of free space due to ECM degradation. One novel feature is that the functions Λ_1, Λ_2 and b (shortly called as GGR functions) have mutually exclusive supports. Thus in this way we are able to model the go or grow or recede phenomenon [5], [7].

D. Normal cell population dynamics (NPD) described by (4):

The dynamics of normal cells is essentially governed by growth and decay terms. The logistic term $N(1 - N/K_N)$ is modulated by proliferation and recession function Λ_4 and Λ_3. Apart from this there is linear decay term with a decay rate proportional to cancer cell density.

IV. Numerical simulations

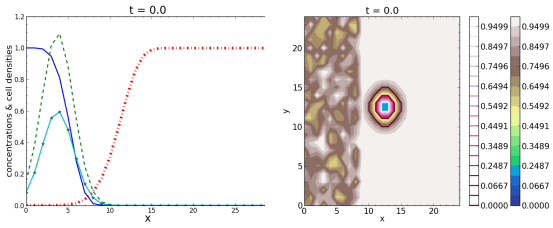
In this section we perform numerical simulations for the proposed model (to be precise the simulatios are carried out for the non-dimensionalized equations). The simulations are done both in one-dimensional (1D) and two-dimensional (2D) spatial domains: The former allow, to better visualize the dynamics of the propagating wave fronts, while the latter are better suited to visualize the infiltrative growth patterns. In both cases we use the Ornstein-Uhlenbeck process $\xi_t := (O_t)_{t \geq 0}$, with

$$O_t = e^{-\nu t} O_0 + \mu(1 - e^{-\nu t}) + \sigma \int_0^t e^{-\nu(t-s)} dW_s$$

as the noise process ξ_t. For the details concerning the parameters chosen and the numerical scheme used for 1D and 2D simulations we refer the reader to [13]. The chosen initial conditions are as shown in Figure 1. For this and all subsequent pictures we will use the following legends: (a) 1D simulations: Solid curve (——): cancer cell density C; dashed curve (− − −): extracellular proton concentration H_e; barred curve (|||): normal cell density N, and solid line with asterisks (–⋆–⋆–): intracellular proton concentration H_i, (b) 2D simulations: The solid curves (—) indicate the level sets of cancer cell density C, while filled regions indicate level sets of normal cell density N. The values corresponding to these level sets are indicated by the colorbars adjacent to the plot and on the right side to the figure. In order to see the effects of spatial heterogeneity, we have added some random perturbation to the initial value of the normal cell density only on the left side of the xy-plane. Thus, these perturbations are seen as patches on the left half of the plot (see Figure 1b).

Fig. 1: Initial conditions in the 1D and 2D case

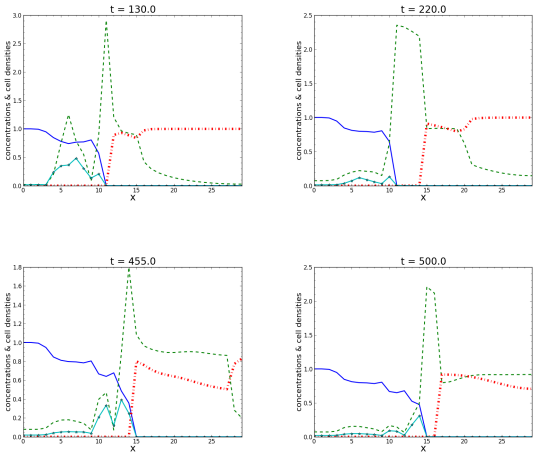
(a) Initial condition in the 1D case (b) Initial condition in the 2D case



A. 1D simulations

Figures 2 depict a temporal sequence of a single sample path of the solution. Several features can be inferred from these plots: (a) In all sample solutions the dynamics of H_i is dominant at the falling edge of C, i.e. near the tumor-stroma interface, exceeding the values it takes in the bulk region of the tumor. Phenomenologically, this captures the high metabolic rates of the cells on the tumor edge to perform cytoskeleton remodeling and taxis, which results in acidic byproducts. (b) In all sample solutions the concentration of H_e exceeds that of H_i on the support on cancer cell density. This is in accordance to the reverse pH gradient observed in the tumor microenvironment. However, due to fluctuations in H_i (induced by the noise), it may happen that near the tumor-stroma interface, the concentration of H_i exceeds that of H_e for a brief period of time. (c) In all sample solutions the concentration of H_e is high at the interface between cancer and normal cells, which captures phenomenologically the accumulation of acid due to high metabolic rates. This accumulation influences in turn the dynamics at the population level in the following way:
(c1) Depending on the H_i and H_e concentrations, the cancer dynamics may be in either one of the go, grow, or recede modes: If the H_i, H_e values are in the support of b, then the cancer cells are in the go-mode, so the cells move in the direction of higher H_e. If the H_i, H_e values are in the support of Λ_1, then the cancer cells are in the grow-mode, hence the cells proliferate. Lastly, if the H_i, H_e values are in the support of Λ_2,

Fig. 2: Time snapshots of the sample solution 90, in the case of a 1D domain.



the cancer cells are in the recede-mode, meaning a decay of tumor cell density. (c2) Depending on γ_N, γ_{Λ_3} and γ_{Λ_4}, the depletion of normal cells varies based on the concentration of H_e and on the density of C. Thus according to the interplay between the go-grow-recede functions and the decay and remodelling rates of normal cells, we can expect appearance (opening) and disappearance (closing) of gaps between cancer and stromal cells.

For a more detailed explanation of the results we refer the reader to the original paper [13]. To summarize the 1D plots show the occurrence of gaps (empty regions, with zero population and relatively high H_e concentration). The gap is a transient phenomenon which closes and opens at different points in time and space, and for some sample paths no gaps appear. Due to this the expectation solution may not show any gaps. However one may recover them by increasing the normal cell decay rate γ_3 and decreasing the normal cell remodelling rate γ_4.

B. 2D simulations

Figure 3 represent the contour plots of a sample path of the solution. In there the first row depicts the time snapshots of cancer cell and normal cell evolution while the second row depicts the time snapshots of the evolution of proton concentrations. The solid curves (—) indicate the level sets of cancer cell density C, while filled regions indicate level sets of normal cell density N. The white regions appearing in the center are the gaps $\Big($regions of nearly zero cell density $(< 10^{-7})\Big)$. To test the influence of spatial heterogeneity we perturb the initial condition of the normal cell density in such a way that it is smooth on the right three-quarters of the spatial plane and it is rough and uneven on the left quarter of the spatial plane. This produces a bias in the direction of invasion of cancer cells. The preferred direction is towards the left side of the 2D plane i.e. toward the rough and uneven side of the tissue. For a more detailed description of the 2D simulation results we refer the reader to the original paper [13]. In short, apart form the appearance of gaps (like in the 1D case) the 2D simulations also show the emergence of different types transient of spatial patterns.

V. Discussion

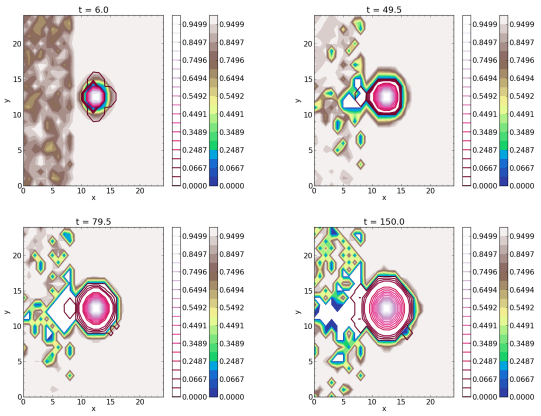
Through numerical simulations we have established that acidity plays a paramount role as a modulator of the tissue environment and as a selector of resistive and malignant cancer progeny. The impact of acidity on the invasiveness of cancer is not a direct one but through a cascade of multiple events which altogether has a negative influence on the tissue architecture and tissue homeostasis. This is generally the case in an complex ecosystem where a reversal of a certain functionality of a key ecological player results in major drift in the evolution of certain populations of the ecosystem. In the current context the establishment of a reverse pH gradient brings about a drift in the behavior of the tumor.

The main highlight of the model is the incorporation of a stochastic term in a realistic way (i.e. by considering a almost surely (in a probabilistic sense) continuous process) and the tight coupling (via GGR functions) of the proton dynamics with the population dynamics. As a result, we observe that: starting form a deterministic initial values the evolution of protons and cancer population, qualitatively diverges into various different patterns. This is clearly evident for the kind of qualitatively different gaps that occur at different point in space over a course of time. Even more surprising, but natural, observation is the lack of gaps in the expected solution. This, in general, highlights a crucial drawback of deterministic modeling, which in a vague sense can be viewed as an averaged result of various possible outcomes. Thus a deterministic model is unable to capture rare events which, from an evolutionary viewpoint, are very significant.

Numerical experiments also highlighted the role of spatial heterogeneity and spatial randomness on the invasive capability of neoplastic cells. This again reiterates the pivotal role of random interactions of passive entities of the ecosystem.

In summary, for just a few choices of Ξ_M and Ξ_R (migration parameters and reaction parameters), the model phenomenologically captured various aspects of the tumor advancement like gap formation, buds and spikes formation, island and cavity formation in the

Fig. 3: Time snapshots of the sample solution number 5, in the case of a 2D domain.



stromal region, and up to some extent even INFb and INFc infiltrative growth patterns were observed. Such vast coverage of invasive features was possible, on the one hand, due to the highly nonlinear coupling via GGR functions and the flux modulation function J and on the other hand the noisy perturbations could bring about different variations in the patterns. Moreover, the model also highlighted that spatial perturbations/unevenness/heterogeneity in structural density of the stroma have a strong influence on the invasion patterns. Hence, it is expected that the range of possible INF patters would be enlarged if one were to incorporate spatial noise both in proton and normal cell dynamics.

REFERENCES

[1] Japanese classification of gastric carcinoma: 3rd english edition. *Gastric Cancer*, 14(2):101–112, 2011.

[2] Audrey Arnal, Beata Ujvari, Bernard Crespi, Robert A. Gatenby, Tazzio Tissot, Marion Vittecoq, Paul W. Ewald, Andreu Casali, Hugo Ducasse, Camille Jacqueline, Dorothe Miss, Franois Renaud, Benjamin Roche, and Frdric Thomas. Evolutionary perspective of cancer: myth, metaphors, and reality. *Evolutionary Applications*, 8(6):541–544, 2015.

[3] M.E. Beckner, M.L. Stracke, L.A. Liotta, and E. Schiffmann. Glycolysis as primary energy source in tumor cell chemotaxis. *Journal of the National Cancer Institute*, 82(23):1836–1840, 1990.

[4] J.R. Casey, S. Grinstein, and J. Orlowski. Sensors and regulators of intracellular pH. *Nature Review Molecular Cell Biology*, 11(1):50 – 61, 2010.

[5] A. Corcoran and R.F. Del Maestro. Testing the go or grow hypothesis in human medulloblastoma cell lines in two and three dimensions. *Neurosurgery*, 53, 2003.

[6] P. Friedl and K. Wolf. Tumour-cell invasion and migration: diversity and escape mechanisms. *Nature Review Cancer*, 3(5):362–374, 2003.

[7] T. Garay, E. Juhász, E. Molnàr, M. Eisenbauer, A. Czirók, B. Dekan, V. László, M.A. Hoda, B. Döme, J. Tímár, W. Klepetko, W. Berger, and B. Hegedűs. Cell migration or cytokinesis and proliferation? Revisiting the go or grow hypothesis in cancer cells in vitro. *Experimental Cell Research*, 319(20):3094 – 3103, 2013.

[8] R.A. Gatenby and E.T. Gawlinski. A reaction-diffusion model of cancer invasion. *Cancer Research*, 56(24):5745–5753, 1996.

[9] R.A. Gatenby, E.T. Gawlinski, A.F. Gmitro, B. Kaylor, and R.J. Gillies. Acid-mediated tumor invasion: a multidisciplinary study. *Cancer Research*, 66(10):5216–5223, 2006.

[10] R.A. Gatenby and R.J. Gillies. Glycolysis in cancer: A potential target for therapy. *The International Journal of Biochemistry and Cell Biology*, 39(78):1358 – 1366, 2007.

[11] D. Hanahan and R.A. Weinberg. Hallmarks of cancer: The next generation. *Cell*, 144(5):646 – 674, 2011.

[12] S. Hiremath and C. Surulescu. A stochastic multiscale model for acid mediated cancer invasion. *Nonlinear Analysis: Real World Applications*, 22(0):176 – 205, 2015.

[13] Sandesh Athni Hiremath and Christina Surulescu. A stochastic model featuring acid induced gaps during tumor progression. 2015.

[14] E. Ito, S. Ozawa, H. Kijima, A. Kazuno, T. Nishi, O. Chino, H. Shimada, M. Tanaka, S. Inoue, S. Inokuchi, and H. Makuuchi. New invasive patterns as a prognostic factor for superficial esophageal cancer. *J. Gastroenterol.*, 47:1279–1289, 2012.

[15] A. Jentzen and P.E. Kloeden. *Taylor Approximations for Stochastic Partial Differential Equations*. SIAM, 2011.

[16] Y. Kato, S. Ozawa, C. Miyamoto, Y. Maehata, A. Suzuki, T. Maeda, and Y. Baba. Acidic extracellular microenvironment and cancer. *Cancer Cell International*, 13(24):89, 2013.

[17] Peter E. Kloeden, Stefanie Sonner, and Christina Surulescu. A nonlocal sample dependence sde-pde system modeling proton dynamics in a tumor. 2015.

[18] G. Meral and C. Surulescu. Mathematical modeling, analysis and numerical simulations for the influence of heat shock proteins on tumor invasion. *Journal of Math. Analysis and Applications*, 408:597–614, 2013.

[19] B. Okamura, A. Gruhl, and J.L. Bartholomew. *Myxozoan Evolution, Ecology and Development*. Springer International Publishing, 2015.

[20] P. Pontarotti. *Evolutionary Biology from Concept to Application*. Springer Berlin Heidelberg, 2008.

[21] E. Racker. Why do tumor cells have a high aerobic glycolysis? *Journal of Cellular Physiology*, 89(4):697–700, 1976.

[22] E.K. Rofstad, B. Mathiesen, K. Kindem, and K. Galappathi. Acidic extracellular pH promotes experimental metastasis of human melanoma cells in athymic nude mice. *Cancer Research*, 66(13):6699–6707, 2006.

[23] A. Roos and W.F. Boron. Intracellular pH. *Physiological Reviews*, 61(2):296–434, 1981.

[24] J. Rozhin, M. Sameni, G. Ziegler, and B.F. Sloane. Pericellular pH affects distribution and secretion of cathepsin b in malignant cells. *Cancer Research*, 54(24):6517–6525, 1994.

[25] E. Ruoslahti. How cancer spreads. *Scientific American*, 275(3):72–77, 1996.

[26] P. Swietach, M. Zaniboni, A.K. Stewart, A. Rossini, K.W. Spitzer, and R.D. Vaughan-Jones. Modelling intracellular H^+ ion diffusion. *Progress in Biophysics and Molecular Biology*, 83(2):69 – 100, 2003.

[27] S. Takumi, J. Verdone, J. Huang, U. Kahlert, J. Hernandez, G. Torga, J. Zarif, T. Epstein, R. Gatenby, A. McCartney, J. Elisseeff, S. Mooney, S. An, and K. Pienta. Glycolysis is the primary bioenergetic pathway for cell motility and cytoskeletal remodeling in human prostate and breast cancer cells. *Oncotarget*, 5(0), 2014.

[28] M.G. Van der Heiden, L.C. Cantley, and C.B. Thompson. Understanding the warburg effect: The metabolic requirements of cell proliferation. *Science*, 324(5930):1029–1033, 2009.

[29] O. Warburg. The metabolism of tumours. 1930.

[30] B.A. Webb, M. Chimenti, M.P. Jacobson, and D.L. Barber. Dysregulated pH: a perfect storm for cancer progression. *Nature Review Cancer*, 11(9):671–7, 2011.

[31] S.D. Webb, J.A. Sherratt, and R.G. Fish. Alterations in proteolytic activity at low pH and its association with invasion: a theoretical model. *Clin Exp Metastasis*, 17(5):397–407, 1999.

[32] S.D. Webb, J.A. Sherratt, and R.G. Fish. Mathematical modelling of tumor acidity: Regulation of intracellular pH. *Journal of Theoretical Biology*, 196(2):237 – 250, 1999.

Numerical simulation of phase separation in cathode materials of lithium ion batteries

Tobias Hofmann
Fraunhofer ITWM, Kaiserslautern
Email: tobias.hofmann@itwm.fhg.de

Ralf Müller
Technical Mechanics, University Kaiserslautern
Email: ram@rhrk.uni-kl.de

Heiko Andrä
Fraunhofer ITWM, Kaiserslautern
Email: heiko.andrae@itwm.fhg.de

***Abstract*—Phase separation during the intercalation of lithium ions leads to degradation effects in some cathode materials, shortens life-time and decreases capacity. A nonlinear initial boundary value problem for the lithium ion concentration on the microscale is introduced. The Cahn-Hilliard equation is used to describe the phase transition from solid-solution diffusion to two-phase dynamics. The resulting phase-field model for the lithium ion concentration is then discretized on a regular voxel mesh. A first-order finite-volume scheme with adaptive time stepping is applied. The parameters and their effects in the non-convex Helmholtz energy are investigated and explained. Furthermore, the numerical convergence of the scheme is examined. In order to illustrate the method, the effect of the parameters interface width and interface energy density is discussed on a random initial configuration.**

***Index Terms*—lithium batteries, finite volume methods, electrochemical processes**

I. Introduction

Today lithium ion batteries form an indispensable component for electronic devices or electric vehicles. Even though lithium ion electrodes are very versatile in battery production due to their high energy density, the diverse fields of application require the prediction of life-time and capacity fade and the modeling of aging mechanisms. A lot of electrode materials show degradation during usage. If a large current is applied at the poles of the battery during discharge, the diffusion of lithium ions inside the battery from anode to cathode is not fast enough and concentration gradients arise. In some materials the restructuring of the lithium ions inside the crystal structure of the electrode material gives rise to large strains [2]. From experiments it is known that the stresses related to these strains can cause mechanical damage effects in materials including lithium tin oxide [4], lithium manganese oxide [9], lithium titanate oxide [5] and lithium iron phospate [6]. The mechanical stresses arising in a porous electrode made of lithium manganese oxide have been numerically simulated with a model based on diffusive dilute solution theory [14].

But especially in lithium iron phospate, the diffusion of the lithium ions from electrolyte into the active material cannot be modeled by a regular diffusion equation. While in a lot of materials the diffusion leads to an even ion distribution inside the material, for lithium iron phospate a separation into areas with a maximum concentration of lithium ions and areas where no lithium ions are present [7], [17]. Lithium-rich areas emerge during charging and persist. The distribution of the lithium ions inside the material can then be described by two different phases, one phase enriched with lithium ions and one phase devoid of lithium ions. The process of separation into different phases is called spinodal decomposition. The problem of describing the movement of the boundaries between both phases is a Stefan problem and may be approached by adaptive meshes and front-tracking methods [11].

Another approach called phase-field method was introduced by Cahn-Hilliard [3] and is based on a thermodynamical approach involving a non-convex Helmholtz energy functional. In a general phase-field method, the boundary between two phases is discretized and a fine regular spatial mesh is used. There have been applications to electrode material for different electrode structures [16], [9], [1]. The Cahn-Hilliard equation involved in the application of this method is a fourth-order non-linear partial differential equation.

In Section II the chemical model for a dilute solution battery on the microscale [10] and a phase-field model for phase separation are introduced [16]. An alternative, more accessible representation of the phase-field potential is derived and the phase dynamics are discussed depending on the material parameters of the chemical potential. In Section III a finite-volume discretization and an adaptive time integration algorithm are presented. The numerical convergence of the discretization for mesh width and time step size is shown. In Section IV the effect of interface width L and interface energy density G on the time evolution of a random initial configuration is examined. In Section V the presented work is summarized and an outlook to future work is given.

II. Chemical model

In this section the equations for the lithium ion distribution in an electrode material of a lithium ion battery are presented on the microscale. After introducing the spatial domain involved the transport equations [10] for the lithium ions are proposed. A phase-field model is applied to describe the phase separation process. Boundary and initial conditions complete the model.

A rectangular domain $\Omega = (0, L_x) \times (0, L_y) \subset \mathbb{R}^2$ of solid electrode material, the time interval $T = (0, t_0]$ and their Cartesian product $\Omega_T = \Omega \times T$ are considered. The function spaces $V = \{f : \Omega \to \mathbb{R}\}$ and $V_T = \{f : \Omega_T \to \mathbb{R}\}$ are introduced to describe the lithium ion concentration.

TABLE I
MATERIAL PARAMETERS

Name	Character	Value	Unit
Concentration	$c(x,y;t)$	-	$\frac{\text{mol}}{\text{l}}$
Universal gas constant	R	8.314	$\frac{\text{J}}{\text{mol K}}$
Temperature	T	300	K
Faraday constant	F	96485	$\frac{\text{A s}}{\text{mol}}$
Cathode diffusion coefficient	D_0	10^{-10}	$\frac{\text{cm}^2}{\text{s}}$
Maximum lithium ion concentration	c_m	20	$\frac{\text{mol}}{\text{l}}$
Enthalpy of mixing per site	θ	$1.110 \cdot 10^4$	$\frac{\text{J}}{\text{mol}}$
Phase-field interface energy	κ	$3.020 \cdot 10^{12}$	$\frac{\text{J}}{\text{mol cm}}$
Reference length	L_0	10^{-7}	cm

The following equations involve the function $c(x,y;t) \in V_T$ describing the lithium ion concentration in $\frac{\text{mol}}{\text{l}}$ and the chemical potential $\mu(x,y;t) \in V_T$ in $\frac{\text{J}}{\text{mol}}$.

A. Governing equations

An electrochemical model for the transport of lithium ions is given in [10]. In the continuity equation

$$\partial_t c + \nabla \mathbf{f} = 0, \quad (x,y;t) \in \Omega_T. \tag{1}$$

the divergence of a diffusion flux $\mathbf{f}$ gives rise to local concentration changes. This flux $\mathbf{f}$ is proportional to the gradient of a chemical potential $\mu \in V_T$

$$\mathbf{f} = D_s(c)\nabla\mu(c), \quad (x,y;t) \in \Omega_T. \tag{2}$$

For thermodynamic consistency, the diffusion coefficient $D_s(c)$ is chosen depending on the local lithium ion concentration c as

$$D_s(c) = \frac{D_0}{RT} c \left(1 - \frac{c}{c_m}\right). \tag{3}$$

In dilute solution theory, the chemical potential μ in Eq. (2) may be chosen as a classical diffusion potential

$$\frac{\mu}{RT} = \log \frac{c}{c_m}, \tag{4}$$

resulting in a parabolic equation [8]. However, in a phase-field method the chemical potential μ is defined as the variational derivative of a non-convex free energy $F(p)$ [3],

$$\frac{\mu}{RT} = \frac{\delta F(p)}{\delta p}. \tag{5}$$

The phase-field order parameter p is the normalized lithium ion concentration, $p = \frac{c}{c_m}$. In [9] the free energy is assumed as

$$\begin{aligned} F_0(p) &= H_1 + H_2, \quad \text{where} \\ H_1 &= p\log p + (1-p)\log(1-p), \\ H_2 &= \frac{\theta}{RT} p(1-p). \end{aligned} \tag{6}$$

The first term H_1 is related to a diffusion potential based on one-body terms in a Hamiltonian of the crystalline structure in active material. The second term H_2 results from a mean-field approximation of two-body interaction terms in the Hamiltonian.

The interface between lithium-rich phase and lithium-poor phase is related to misfits in the crystal structure. Therefore in phase-field methods a penalty term involving a norm of the gradient of the concentration is added to the free energy to receive

$$F(p, \nabla p) = \alpha L_0^2 \frac{G}{L} F_0(p) + \frac{\beta G L}{2} (\nabla p)^2. \tag{7}$$

The parameters G and L are introduced as an alternative description of the parameters phase-field model. They represent energy density and the width of the interfacial region, respectively. Moreover, α and β are dimensionless scalar parameters. All of these parameters are derived analytically from the representation of the free energy with the Euler-Lagrange equations in Remark 1.

From Eq. (7), the chemical potential $\mu \in V_{c,T}$ is now given as the variational derivative by

$$\frac{\mu}{RT} = \alpha L_0^2 \frac{G}{L} F_0'(\frac{c}{c_m}) - \beta G L \Delta \left(\frac{c}{c_m}\right) \quad (x,y;t) \in \Omega_T. \tag{8}$$

Combining Eq. (1), Eq. (2) and Eq. (8), the equations governing the concentration distribution c and the chemical potential μ are therefore

$$\begin{aligned} 0 &= \partial_t c + \nabla(D_s(c)\nabla\mu), \\ \mu &= RT\alpha L_0^2 \frac{G}{L} F_0'(\frac{c_c}{c_m}) - RT\beta G L \Delta \left(\frac{c_c}{c_m}\right), \\ &\quad (x,y;t) \in \Omega_T. \end{aligned} \tag{9}$$

The chemical potential is here introduced as an additional unknown. The resulting equation system is made of two second-order equations.

Remark 1 (Derivation of the alternative phase-field representation)**.** *In this remark an alternative representation of the chemical phase-field potential μ is derived in terms of the interface width L and the interface energy density G as presented in [13].*

With the free energy F_0 in Eq. (6) *and the mixing enthalpy θ given in Table I, the minima of the free energy can be derived from the equation $F_0'(p) = 0$ as $p_1 \approx 0.013$ and $p_2 \approx 0.987$. The corresponding concentrations are called the equilibrium concentrations $c_1 \approx 0.013\, c_m$ and $c_2 = c_m - c_1 \approx 0.987\, c_m$.*

The shifted free energy $\Delta F_0(p) = F_0(p) - F_0(p_1)$ allows to define the activation energy needed during a phase transformation from concentration p_1 to p_2 as an integral.

The coefficients in Eq. (7) *can be identified with corresponding values given in [15] as*

$$\alpha L_0^2 \frac{G}{L} = 1, \quad \beta G L = \frac{\kappa}{L_0^2 c_m N_A RT}. \tag{10}$$

With this and the use of the Euler-Lagrange equations, G and L are calculated as

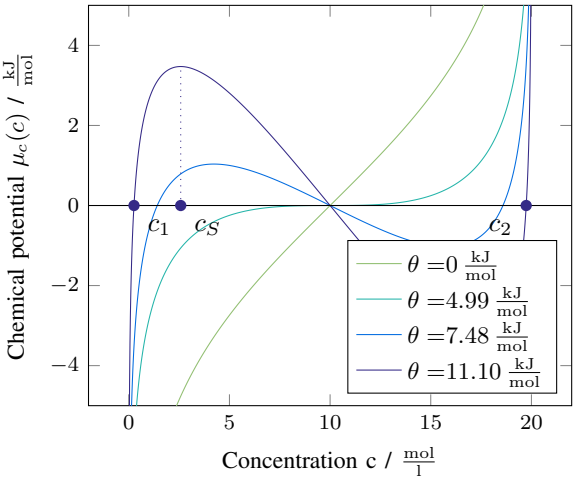


Fig. 1. The bulk chemical potential $\mu_c(c)$ for different values of the phase-field enthalpy θ. Equilibrium concentrations c_1 and c_2 and the local maximum c_S are depicted for the value θ =11.1 $\frac{\text{kJ}}{\text{mol}}$.

$$G = \frac{1}{L_0}\sqrt{2\frac{\kappa}{L_0^2 c_m N_A RT}}\int_{p_1}^{p_2}\sqrt{\Delta F_0(p)}dp \approx 2.09\cdot 10^{-7}\frac{1}{\text{cm}},$$
$$L = L_0(p_2-p_1)\sqrt{\frac{\kappa}{2L_0^2 c_m N_A RT\Delta F_0(0.5)}} \approx 3.33\cdot 10^{-7}\text{cm}. \tag{11}$$

The scalar parameters α and β can now be deduced as

$$\alpha \approx 1.593\cdot 10^{-14}, \quad \beta \approx 1.444\cdot 10^{10} \tag{12}$$

The application of a phase-field method requires a fine discretization of the interface width. From the interface width being L = 3.3 nm it can be concluded that the spatial discretization size should not be larger than $h \approx 1$ nm. This way the phase interface between different phases is resolved by at least three or four voxels.

Remark 2 (Equilibrium concentration)**.** *Figure 1 shows the influence of the mixing enthalpy parameter θ on the bulk chemical potential $F_0'(c)$*

$$F_0'(c) = RT\log\frac{c}{c_m-c} + \theta\left(1-2\frac{c}{c_m}\right) \tag{13}$$

in Eq. (8)*. For a value of $\theta < \theta_{crit} = 0.052\,\frac{\text{kJ}}{\text{mol}}$, this chemical potential is monotonous. There are no extrema and no phase separation will occur. $\theta > \theta_{crit}$ allows phase coexistence, where the equilibrium concentration values are given by the root marks in Figure 1. An over-saturated state of charge between the equilibrium concentrations $c_1 \approx 0.013\,c_m$ and the concentration $c_s \approx 0.13\,c_m$ at the maximum S of the chemical potential gives rise to a spinodal decomposition. In this domain, small perturbations lead to phase separation into a lithium-poor phase with concentration c_1 and a lithium-rich phase with concentration $c_2 = c_m - c_1$.*

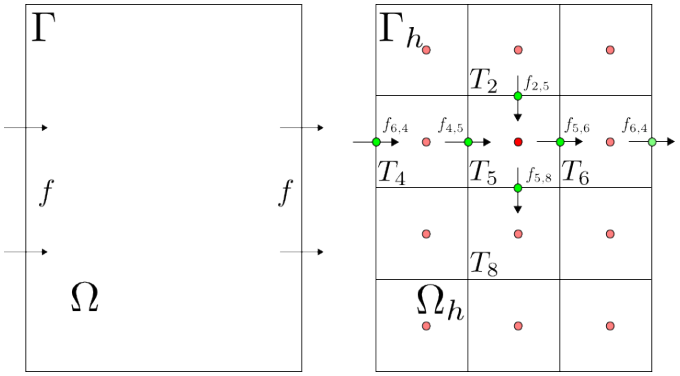


Fig. 2. Discretization of a box domain into cells. Arrows denote the periodic boundary conditions as well as the cell fluxes in the finite-volume method.

TABLE II
NON-DIMENSIONALIZATION OF UNKNOWNS

$\tilde{\mu}(\tilde{c}) = \frac{1}{RT}\mu(c_m\tilde{c})$	$\tilde{c} = \frac{c}{c_m}$	$\tilde{\phi} = \frac{F}{RT}\phi$
$\tilde{t} = \frac{D_e}{L_x^2}t$	$\tilde{\nu} = \frac{\nu}{i_{\text{in}}}$	$\tilde{x} = \frac{x}{L_x}$

B. Boundary and initial conditions

On the boundary Γ periodic boundary conditions are prescribed for the lithium ion concentration c,

$$\begin{aligned} c(x,0;t) &= c(x,L_y;t), \quad x\in(0,L_x)\,t\in T,\\ c(0,y;t) &= c(L_x,y;t), \quad y\in(0,L_y)\,t\in T, \end{aligned} \tag{14}$$

and also for the concentration flux $\mathbf{f}$

$$\begin{aligned} D_s(c)\nabla\mu(x,0;t) &= D_s(c)\nabla\mu(x,L_y;t), \quad x\in(0,L_x)\,t\in T,\\ D_s(c)\nabla\mu(0,y;t) &= D_s(c)\nabla\mu(L_x,y;t), \quad y\in(0,L_y)\,t\in T. \end{aligned} \tag{15}$$

As an initial concentration distribution $c_0 \in V$ is assumed and $c(x,y;0) = c_0, (x,y)\in\Omega$ is defined. It is important that c_0 has only values in $(0,c_m)$. Large initial gradients in c_0, for example stemming from a random initialization with values in $(0,c_m)$ lead to very small time steps in the beginning.

III. Numerical method

In this section, a general first-order, finite-volume, cell-centered scheme for discretizing a parabolic differential equation with non-linear conductivity coefficients is presented. Grid function spaces for the concentration are introduced.

A. Non-dimensionalization

Table II introduces dimensionless unknowns by a tilde-notation. The size of the domain in the first dimension L_x is used as a reference length. This enables us to describe both spatial and time derivatives in a non-dimensional way as

$$\frac{\partial}{\partial\tilde{x}_i} = L_x\frac{\partial}{\partial x_i}, \quad \tilde{\nabla} = L_x\nabla, \quad \frac{\partial}{\partial\tilde{t}} = \frac{D_e}{L_x^2}\frac{\partial}{\partial t}. \tag{16}$$

For the subsequent derivation dimensionless equations are utilized without explicit usage of the tilde notation.

B. Discretization

Suppose that N_x and N_y are positive integers such that $\frac{L_x}{N_x} = \frac{L_y}{N_y}$. Let $h \coloneqq \frac{L_x}{N_x}$ and consider a uniform voxel mesh Ω_h on Ω, defined by

$$\begin{aligned}\Omega_h \coloneqq \Big\{ (x,y) \in \Omega | x = ih - \frac{1}{2}, y = jh - \frac{1}{2}, \\ 1 \leq i \leq N_x, 1 \leq j \leq N_y \Big\}.\end{aligned} \tag{17}$$

The number of grid cells is $N = N_x N_y$. Let V_h denote the linear space of real-valued functions defined on Ω_h. Figure 2 shows how to introduce a uniform staggered mesh $y_{ij} \in \Sigma_h$ and establishes a notation relating the staggered mesh to the original mesh Ω_h. The elements in Σ_h are

$$\begin{aligned}\Sigma_h \coloneqq \{(x_i, y_j) \in \bar{\Omega} | x_i = ih, y_j = jh, \\ 0 \leq i \leq N_x, 0 \leq j \leq N_y\}.\end{aligned} \tag{18}$$

This definition includes mesh points on the boundary of Ω. Points on opposite sides of Ω are identified with each other as

$$\begin{aligned}(0, jh) \cong (L_x, jh) \quad 0 \leq j \leq N_y, \\ (ih, 0) \cong (ih, L_y) \quad 0 \leq i \leq N_x.\end{aligned} \tag{19}$$

Let U_h denote the linear space of real-valued functions defined on Σ_h.

Given a solution $c_T \in V_T$, $c = c_T(\cdot, t) \in V$ is called the current time step and $\check{c} = c_T(\cdot, t - \tau) \in V$ is called the previous time step. Integration of the diffusion equation in Eq. (1) over $[t - \tau, t]$ gives

$$\begin{aligned}0 = &\int_{t-\tau}^{t} \partial_t c_T dt - \int_{t-\tau}^{t} \nabla(D_s(c_T)\nabla\mu_c(c_T))dt = \\ &= c - \check{c} - \int_{t-\tau}^{t} \nabla(D_s(c_T)\nabla\mu_c(c_T)dt = \\ &= c - \check{c} - \tau\nabla(D_s(c)\nabla\mu_c(c) + O(\tau)\end{aligned} \tag{20}$$

Now take $c_h, \mu_h \in V_h$ as the discretized lithium ion distribution and discretized chemical potential, respectively. The components of c_h are called $c_i, i \in \{1 \dots N\}$. Eq. (20) is now integrated over a cell T_i with volume $|T_i| = h^2$. The surface areas are $|\partial T_i| = h$ corresponding to line segments. The surface integrals are calculated by integration of the constant flux f_{ij} on Σ_h.

$$\begin{aligned}0 &= \int_{T_i} (c - \check{c})dV - \tau \int_{T_i} \nabla(D_s(c)\nabla\mu(c))dV + O(\tau) \\ &= \int_{T_i} (c - \check{c})dV - \tau \int_{\partial T_i} (D_s(c)\nabla\mu(c))dS + O(\tau) \\ &= h^2(c_i - \check{c}_i) - h\tau \sum_{(i,j)\in\Sigma_h} f_{ij}(c) + O(h + \tau), \\ &(x,y) \in \Omega_h\end{aligned} \tag{21}$$

and a finite-difference approximation for the gradient of μ,

$$\begin{aligned}f_{ij} &= D_s(c)\nabla\mu(c) = \\ &= D_s\left(\frac{c_i + c_j}{2}\right)\frac{\mu(c_j) - \mu(c_i)}{h} + O(h) \text{ if } f_{ij} \in \Sigma_h.\end{aligned} \tag{22}$$

Eq. (8) on the chemical potential $\mu \in V_T$ is discretized similar in space. This concludes the discretization of the governing equations in time and space.

The periodic boundary conditions are realized by the discretization with a periodic identification of the interface fluxes $f_{i,j}$ in Eq. (22). The discretized initial condition $c_{0,h} \in V_h$ is defined from the initial condition $c_0 \in V$ as

$$c_{0,h} = c_0(x), \quad (x,y) \in \Omega_h \tag{23}$$

After restructuring, the discretization includes these two equations

$$\begin{aligned}\frac{c_i - \check{c}_i}{\tau} =& \frac{1}{h^2} \sum_{y_{ij}\in\Sigma_h} D_s\left(\frac{c_i + c_j}{2}\right)(\mu_j - \mu_i) \\ \mu_i =& \alpha F_0'(c_i) + \frac{\beta}{h^2} \sum_{y_{ij}\in\Sigma_h} (c_j - c_i) \\ &\forall x_i \in \Omega_h.\end{aligned} \tag{24}$$

These equations can be applied to each other. By this a single equation is received that only involves the unknowns for the current discretized time step c_i and the discretized previous time step $\check{c}_i$. For the next section this equation is assumed to be given in an abstract formulation as

$$f(c_i, \check{c}_i) = 0 \quad (x,y) \in \Omega_h \tag{25}$$

C. Linearization and adaptive algorithm

Numerical simulations of phase transitions inside active materials require either very small constant time steps or an adaptive scheme. The non-linear system in N unknowns is solved with a damped Newton-Raphson algorithm with an adaptive time step algorithm.

In the last subsection a discretization of the electrochemical model equations is presented. Now the equations are assumed to be given as an ordered nonlinear equation system $f(c_h, \check{c}_h) = 0$, where c_h and $\check{c}_h$ denote the discretized solution and the discretized from the previous time step, respectively. Now in terms of a general nonlinear equation solver, $f(c_h)$ is denoted as the residuum and $J(c_h) = \dfrac{\partial f(w_h, \check{c}_h)}{\partial w_h}(c_h)$ is denoted as the Jacobian. The Newton direction vector d is the solution to the linear equation system given by $Jd = f$.

An adaptive damped Newton-Raphson scheme is applied for the time integration. The index k denotes the current Newton iteration. The index l denotes the current dampening line search iteration. Depending on the iteration number of the last time step, the time step size is either increased, retained or decreased. In each iteration a damped line search with a dampening parameter is performed with dampening

TABLE III
PARAMETERS FOR THE TIME INTEGRATION SCHEME

Name	Symbol	Value
Error criterion	ϵ	10^{-10}
Maximum number of Newton iterations	$k_{\max}$	20
No. of undamped steps until time step trust	N_{trust}	10
Maximum state of charge	$\overline{SOC}$	0.99
Reduction factor in the line search algorithm	σ	$\frac{1}{2}$
Initial line search step size	ω_0	1
Minimum line search step size	$\omega_{\min}$	10^{-3}
Minimum time step size	$\tau_{\min}$	10^{-6}
Maximum time step size	$\tau_{\max}$	$\frac{t_0}{100}$

parameter ω_l. Succeedingly smaller parameters are tried, until an error criterion for the new residual $f(c_{k,l})$ is met. The line search algorithm ensures global linear convergence and local quadratic convergence [12]. The error e in the algorithm and the line search corresponds to a discrete L^1-norm for a grid function c_h and is calculated as

$$||f(c_h)||_1 = \max_{i\in\{1,N\}} |c_i| \tag{26}$$

Table III gives the parameters used in the schemes. The symbols $c_{k,l}$, f and d for vectors in $\mathbb{R}^N$ denote a solution vector, a residuum vector and a Newton direction, respectively.

D. Numerical test

In this section the numerical convergence of the method is examined. An initial configuration is chosen as

$$\begin{gathered} c_0(x,y) = 0.4\sin\left(\frac{2\pi x}{L_x}\right)\sin\left(\frac{2\pi y}{L_y}\right) + 0.5, \\ (x,y) \in (0,L_x)\times(0,L_y). \end{gathered} \tag{27}$$

The domain size $L_x = L_y = 50$ nm is used. The simulation is run until for $t_0 = 10^{-3}$ s. The time interval $[0, t_0]$ is divided into N_t steps of size τ_{Ref}. Then a discretized norm $||c_h||_h$ for a discretized solution for the concentration c_h in $\Omega_h \times T_\tau$ is given as

$$||c_h||_h = \frac{1}{N_x N_y N_t}\sqrt{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_t} c_h(ih, jh; k\tau)^2} \tag{28}$$

A relative error measure with respect to a reference solution c_{ref} is defined as

$$e = \frac{||c_h - c_{\text{Ref}}||_h}{||c_{\text{Ref}}||_h}; \tag{29}$$

The error in a similar discretized parabolic equation is $O(h+\tau)$ [8], where h is the mesh size and τ is the time step size. The numerical simulation is conducted for pairings of grid size and time step size (h, τ) given in Table IV. In the logarithmic plot in Figure 3 the logarithmic error ξ plotted against the discretization fineness $\frac{\tau}{\tau_{\text{Ref}}} = \frac{h}{h_{\text{Ref}}}$ can be seen. The convergence order is linear with $O((h+\tau)^{0.96})$.

TABLE IV
NUMERICAL CONVERGENCE DATA, THE SOLUTION ON THE FINEST GRID IS USED AS A REFERENCE SOLUTION.

h / nm	1	0.5	0.25	0.125	0.0625	0.03125
τ / µs	51200	25600	12800	6400	3200	1600

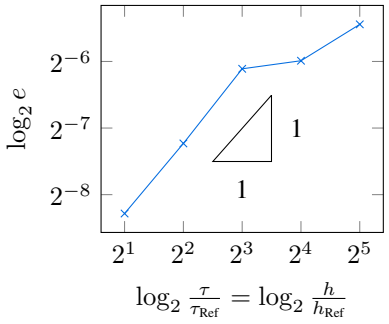


Fig. 3. Results of the numerical convergence test in a binary logarithmic plot. Depicted is the error in the lithium ion concentration (blue) and a linear slope for comparison.

IV. SPINODAL DECOMPOSITION

A simulation is run for $t_0 = 0.01$ s on a quadratic domain with $L_x = L_y = 50$ nm. For the initial concentration, a random uniform distribution in the range $[0.1\,c_m; 0.9\,c_m]$ is used as seen in Figure 4. Figure 5 shows the concentration distribution in the material at the end of the simulation time for different values of the phase-field interface width L and phase-field interface energy density G, using the same color bar as Figure 4. The top left picture shows the result of the simulation with the values for interface width and interface energy density given in [15]. The spinodal decomposition and minimization of interface energy density leads to several distributed lithium rich and lithium poor phases inside the material.

The other simulations show results for the interface width in the range of 3.3 nm to 6.6 nm and for the interface energy density in a range from $0.209\frac{1}{\text{cm}}$ to $2.09\frac{1}{\text{cm}}$. Even though the simulation results are numerically different, small variations in both parameters do not affect the general evolution of the shapes.

With larger interface width, the contours of the phase separation interface into lithium rich and lithium poor get more blurred. In the case of a small interface energy density and a large interface width they disappear nearly completely and the future concentration distribution is dominated by diffusion instead of phase separation.

Figure 5 also shows the effect of a larger interface energy density during the process of Ostwald ripening. Ostwald ripening denotes the disappearing of small phase separated regions and the growth of already larger phase separated regions. A large interface energy density and large interface width also enable faster diffusion inside the particle, resulting in a nearly laminar structure of lithium rich and poor phases. This structure minimized curvature and is a stable equilibrium for periodic phase-field problems.

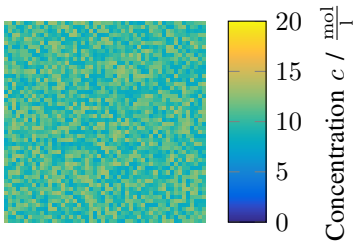


Fig. 4. Random initialization in the range $[0.1\,c_m; 0.9\,c_m]$

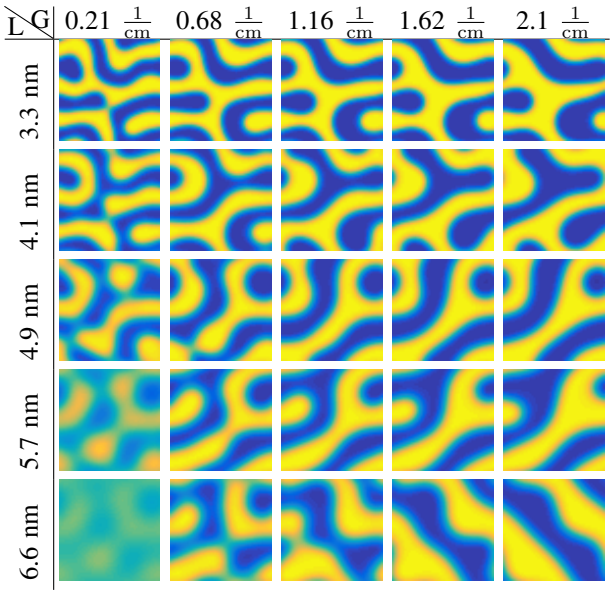


Fig. 5. Spinodal decomposition depending on interface width and interface energy density

V. Summary and Conclusions

An electrochemical model implementing the transport of lithium ions inside electrode material is introduced in Section II. While previous theory is limited to a diffusion model, the model is extended to include arbitrary electrochemical potentials. A phase-field potential is introduced and the interface length is derived as 3.3 nm in Remark 1.

The finite volume method presented in Section III is suitable for regular voxel meshes and enables the use of accelerated numerical methods such as Fourier methods for parabolic equations. The method shows linear numerical convergence for a finer discretization. The adaptive time integration method is useful for finding the maximum time step size during the numerical simulation. In the application of phase-field methods the emergence of phases requires small time steps while during most of the simulation time larger time steps are possible.

In Section IV the effects of interface length and interface energy density G are shown on a random initial configuration. The shorter the interface length L, the more defined and sharper are the interfaces between lithium-rich and lithium-poor phases. For larger interface energy density, the lithium-rich phases emerge faster and the interface areas are more flattened.

Future work includes an electrochemical model for a battery cell including anode and electrolyte domain. Here Butler-Volmer transmission conditions are required to define the interface currents between electrode and electrolyte. The extension of the model onto elasticity depending on the concentration gradients is crucial for the explanation of aging processes during cycling. The computational costs can be significantly reduced by the application of fast Fourier methods for the spatial differentiation on regular voxel meshes instead of finite differences.

References

[1] M. Z. Bazant. Theory of Chemical Kinetics and Charge Transfer based on Nonequilibrium Thermodynamics. *Accounts of chemical research*, 46(5):1144–1160, 2013.

[2] M. Broussely, P. Biensan, F. Bonhomme, P. Blanchard, S. Herreyre, K. Nechev, and R. Staniewicz. Main aging mechanisms in Li ion batteries. *Journal of Power Sources*, 146(1-2):90–96, 2005.

[3] J. Cahn and J. Hilliard. Free Energy of a Nonuniform System. I. Interfacial Free Energy. *The Journal of Chemical Physics*, 28(1958):258–267, 1958.

[4] D. Chen, S. Indris, M. Schulz, B. Gamer, and R. Mönig. In situ scanning electron microscopy on lithium-ion battery electrodes using an ionic liquid. *Journal of Power Sources*, 196(15):6382–6387, 2011.

[5] Z. Choi, D. Kramer, and R. Mönig. Correlation of stress and structural evolution in Li4Ti5O12-based electrodes for lithium ion batteries. *Journal of Power Sources*, 240:245–251, 2013.

[6] D. a. Cogswell and M. Z. Bazant. Coherency strain and the kinetics of phase separation in LiFePO4 nanoparticles. *ACS nano*, 6(3):2215–25, 2012.

[7] M. Ebner, F. Marone, M. Stampanoni, and V. Wood. Visualization and Quantification of Electrochemical and Mechanical Degradation in Li Ion Batteries. *Science*, 342(6159):716–720, 2013.

[8] L. Evans. *Partial Differential Equations*. American Mathematical Society, 1996.

[9] M. Huttin. Phase-field modeling of the infuence of mechanical stresses on charging and discharging processes in lithium ion batteries. 2014.

[10] A. Latz and J. Zausch. Thermodynamic consistent transport theory of Li-ion batteries. *Journal of Power Sources*, 196(6):3296–3302, 2011.

[11] S. Mitchell and M. Vynnycky. Finite-difference methods with increased accuracy and correct initialization for one-dimensional Stefan problems. *Applied Mathematics and Computation*, 215(4):1609–1621, 2009.

[12] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag New York, 1999.

[13] D. Schrade. *Microstructural modeling of ferroelectric material behavior*. PhD thesis, TU Kaiserslautern, 2010.

[14] M. Taralov. *Simulation of Degradation Processes in Lithium-Ion Batteries*. PhD thesis, TU Kaiserslautern, 2015.

[15] Y. Zeng, P. Albertus, R. Klein, N. Chaturvedi, A. Kojic, M. Z. Bazant, and J. Christensen. Efficient Conservative Numerical Schemes for 1D Nonlinear Spherical Diffusion Equations with Applications in Battery Modeling. *Journal of the Electrochemical Society*, 160(9):A1565–A1571, 2013.

[16] Y. Zeng and M. Z. Bazant. Phase separation dynamics in isotropic Ion-intercalation particles. 74(4):980–1004, 2014.

[17] Y. Zhu, J. W. Wang, Y. Liu, X. Liu, A. Kushima, Y. Liu, Y. Xu, S. X. Mao, J. Li, C. Wang, and J. Y. Huang. In Situ Atomic-Scale Imaging of Phase Boundary Migration in FePO4 Microparticles During Electrochemical Lithiation. *Advanced Materials*, 25(c):5461–5466, 2013.

Multiscale Modeling of Glioma Invasion

Alexander Hunt*, Christina Surulescu
Fachbereich Mathematik
Technische Universität Kaiserslautern
Email: *hunt@mathematik.uni-kl.de

***Abstract*—Gliomas are a common brain tumor type, characterized by fast invasion and highly anisotropic spread. We consider a model which describes this dynamics upon relying on kinetic equations. It accounts for the brain tissue anisotropy by using DTI medical data. The cell proliferation is modeled in the kinetic equations context by using a cell-tissue interaction term as in [9]. An appropriate parabolic scaling leads to an effective equation for the macroscopic behavior of the tumor cell density, thus allowing for predictions about the tumor extent in the particular brain tissue provided by the data.**

I. Introduction

Gliomas are primary brain tumors originating from glia cells. Especially glioblastoma multiforme, one of the most frequent brain tumors, has a poor prognosis, as it is very diffusive and it is characterized by a highly anisotropic spread, leading to fractal tumor margins which cannot be assessed by medical imaging. Hence it is very difficult (if at all) possible to resect and/or irradiate the tumor in an exhaustive way. Therefore, more information is needed for the therapy of this type of brain tumor. Mathematical modeling, especially multiscale modeling based on medical data, might overcome this issue, as the simulations of these models may reveal the microlocal tumor structure in higher detail. This paper is a review of [9]. For more details about diffusion tensor imaging (DTI) and its use in glioma modeling we refer to section III below and to [9], [8], [19].

II. Modeling

In recent years a new class of models and modeling techniques evolved to enhance the description accuracy of such biological phenomena. They aim at taking into account the intrinsic multiscale structure of different biological processes like the migration of tumor cells, on which we focus here. As in [9], [8], [19] we start from a kinetic description of the (mesoscopic) tumor cell density modeled via a velocity-jump-process. Thereby, the density function p depending on the variables $(t,x,v) \in [0,T] \times \mathbb{R}^d \times s\mathbb{S}^{d-1}$ satisfies a kinetic transport equation of the form

$$\partial_t p + div_x(vp) = \lambda \int_V (T(v,v')p(t,x,v') - T(v',v)p(t,x,v))\ dv',$$

where $V = s\mathbb{S}^{d-1}$ and $T(v,v')$ denotes the turning kernel describing changes from a previous velocity regime v' into a new one, v. The velocity also depends on the cell speed s, however (as we are mainly interested in the direction of cells and since it is very difficult to assess the individual cell speeds in an evolving tumor), we will assume s to be a constant. The function λ characterizes the cell turning rate at the reorientation sites and is assumed to be constant in most models (see e.g., [19]). In our model we make a more realistic choice and allow it to depend on the subcellular dynamics of receptor binding to insoluble signals from the tumor microenvironment (as described below).

In the following we assume that we have a description of the brain tissue fiber density[1] (denoted by $q(x,v)$) which satisfies

$$\int_V q(x,v)\ dv = 1 \quad \text{and} \quad \int_V vq(x,v)\ dv = 0 \quad \forall x \in \mathbb{R}^d,$$

and that the reorientation kernel is given by the fiber orientation: $T(v,v') = q(x,v)$. The first condition is only the explicit reformulation of q being an orientation density, while the second one relates to the symmetry of this distribution, which means that the single fibers are assumed not to be oriented (i.e. a cell can move up or down a fiber with equal efficiency).

While this model only involves so far a single scale, tumor invasion is a multiscale process involving a plethora of events taking place on several levels, from the subcellular one over the individual cell behavior and up to the macroscopic level of the tumor itself. The latter is the scale on which the evolution of the neoplastic tissue is observed by medical imaging, however, that evolution is controlled by and influences the processes on the lower modeling levels, see e.g., [11]. Here we focus only on certain aspects of the subcellular dynamics, as including more biological detail would complicate too much the setting and require more model parameters to be assessed. In order to survive, proliferate, and migrate all cells (including glioma) need to bind to the surrounding tissue via surface receptors (e.g., integrins, tyrosine kinase receptors, etc.). This leads us to adding a variable for the density of bound cell surface receptors (denoted by $y \in Y = (0,1)$). Here we only consider the receptor binding dynamics (and ignore the subsequent intracellular signaling cascade leading to cell polarization, formation of protrusions, advancing through the cell cycle etc.):

$$\frac{d}{dt}y = G(Q,y) = -(k^+Q + k^-)y + k^+Q, \qquad (1)$$

where k^+ and k^- are rates characterizing the reversible binding of receptors to the tissue. The corresponding mass

[1] more on that in the section regarding medical data

action kinetics can be schematically represented by

$$Q + (1-y) \underset{k^-}{\overset{k^+}{\rightleftharpoons}} y,$$

where the free receptors $(1-y)$ bind to the extracellular matrix Q and so form bound receptors y. As mentioned above, this binding is a reversible process.

Thereby, k^+ is a measure of how many cell surface receptors are bound to the tissue per unit time, while k^- denotes the detachment rate. The quantity $Q(x)$ is the macroscopic volume fraction of tissue components in a spatial point x. This model is very simplifying, nevertheless still capable to capture essential features in the qualitative behavior of a cell interacting with its surroundings in the absence of soluble ligands in the peritumoral region.

If we now follow the typical procedure of deriving mean-field equations (see e.g., [15] for an application of that procedure in the tumor invasion context), we arrive at the equation

$$\partial_t p + div_x(vp) + div_y(G(x,y)p) = \lambda(y)\left(q(x,v)\bar{p} - p\right), \quad (2)$$

where we introduced the notation $\bar{p}(t,x) = \int_V p(t,x,v,y)\ dv$. The turning rate λ now depends on the amount y of receptors bound to the tissue hence contains subcellular level indormation. The latter is also contained in the supplementary transport term $div_y(G(x,y)p)$. To make the analysis feasible, we transform equation (2) by $y \to z$, where the variable $z = y^\star - y$ measures the deviation of the variable y from the only steady state $y^\star = f(Q)$ of (1). After this transformation (2) becomes

$$\begin{aligned} &\partial_t p(t,x,v,z) \\ &\quad + div_x(vp) + \partial_z\left(\left(-(k^+Q + k^-)z + f'(q)v\cdot\nabla Q\right)p\right) \\ &\qquad = \lambda(z)(q\bar{p} - p). \end{aligned} \quad (3)$$

In the following, as done in [9], [10] in a related, but different context, we select the turning rate to be linear: $\lambda(z) = \lambda_0 - z\lambda_1$. To include proliferation into this multiscale model, we rely on the biological fact that receptor binding is the onset of a network of intracellular events culminating in cell division [17]. Hence we use the term

$$\mathcal{P}(p) = \mu(x,u,v)\int_Z \chi(x,z,z')p(t,x,v,z')Q(x)\ dz' \quad (4)$$

for modeling proliferation, where μ is the growth rate dependent on the macroscopic cell density $u = \int_Z \bar{p}\ dz$, and $Z = (y^\star - 1, y^\star)$. This form is also motivated by the so-called proliferative interaction (between p and Q) in the kinetic theory of active particles (KTAP) framework introduced by [3]. Thus we consider the complete multiscale model for tumor proliferation and migration

$$\begin{aligned} \partial_t p + div_x(vp) + \partial_z\left(\left(-\left(k^+Q + k^-\right)z + f'(Q)v\cdot\nabla Q\right)p\right) \\ = \lambda(z)(q\bar{p} - p) + \mathcal{P}(p), \end{aligned} \quad (5)$$

together with (1). The model is globally well-posed, as can be proven by using a result established in a more general framework [18].

III. Medical Data

The evolution of a tumor in a particular brain tissue is of course patient specific, thus medical data should be included into the multiscale model (5) to properly characterize the underlying brain structure. This procedure has different constraints, so we cannot assume high-resolution data with high angular precision to be available. Nevertheless, it should remain possible to include exactly this high-precision data without modifying any assumptions. The data we want to use should also be a medical standard in glioma therapy. Moreover, we must not tie ourselves to a special type of data. The available data types include diffusion tensor imaging (DTI), Q-Ball imaging and the so-called HARDI (high angular resolution diffusion imaging) data. All these types of data are obtained by medical imaging and offer different diffusion images per brain voxel. The diffusion tensor can be reconstructed by using at least 7 diffusion images: 1 weighted diffusion image and 6 gradient images. This will not result in high accuracy data, but is cheaper and therefore more often done. The Q-Ball and HARDI data take up to 48 gradient directions with one ore more weighted diffusion images.

What we utilize here is the concept of an orientation distribution function [1], which is available for all of these data types. This seems to be the most flexible way to get the fiber orientation density q for our model. The macroscopic tissue density Q has to be estimated in some way. We did a heuristic assessment involving the diffusion tensor (it can be reconstructed for all the mentioned data) and some characteristics of the Brownian motion associated to the diffusion of water molecules over a previous setting [8] using the fractional anisotropy as cell density.

The fractional anisotropy is defined for a diffusion tensor $\mathbb{D}$ as $FA = \frac{\sqrt{3}}{\sqrt{2}}\frac{\sqrt{(\nu_1-\bar{\nu})^2+(\nu_2-\bar{\nu})^2+(\nu_3-\bar{\nu})^2}}{\sqrt{\nu_1^2+\nu_2^2+\nu_3^2}}$. Here ν_i $(i=1,2,3)$ denote the eigenvalues of $\mathbb{D}$ and $\bar{\nu}$ is the mean $\frac{\nu_1+\nu_2+\nu_3}{3}$. These eigenvalues exist and are real numbers, as the diffusion tensor is reconstructed in a symmetric and positive definite way.

Recently a computational method has become available for the reconstruction of tissue density per voxel involving inverse problems [20]. Unfortunately this method requires data of higher resolution than is available to us.

The data we use is provided by Dr. Carsten Wolters (WWU Münster) and stems from a healthy subject. For the computations we superimpose in the initial configuration a hypothetical tumor on the brain structure obtained from this data.

IV. Limit Equations

Using parabolic scaling we extract in the macroscopic limit the effective equation for the overall cell density u

$$\begin{aligned} \partial_t u - \nabla\nabla : (\mathbb{D}_T(x)u) + \nabla\cdot(g(Q(x))\mathbb{D}_T(x)\nabla Qu) \\ = Q(x)\mu(x,u)u. \end{aligned} \quad (6)$$

Thereby the (tumor) diffusion tensor $\mathbb{D}_T$ can be explicitly computed and takes the form

$$\mathbb{D}_T = \frac{1}{\lambda_0}\int_V v\otimes vq(x,v)\ dv,$$

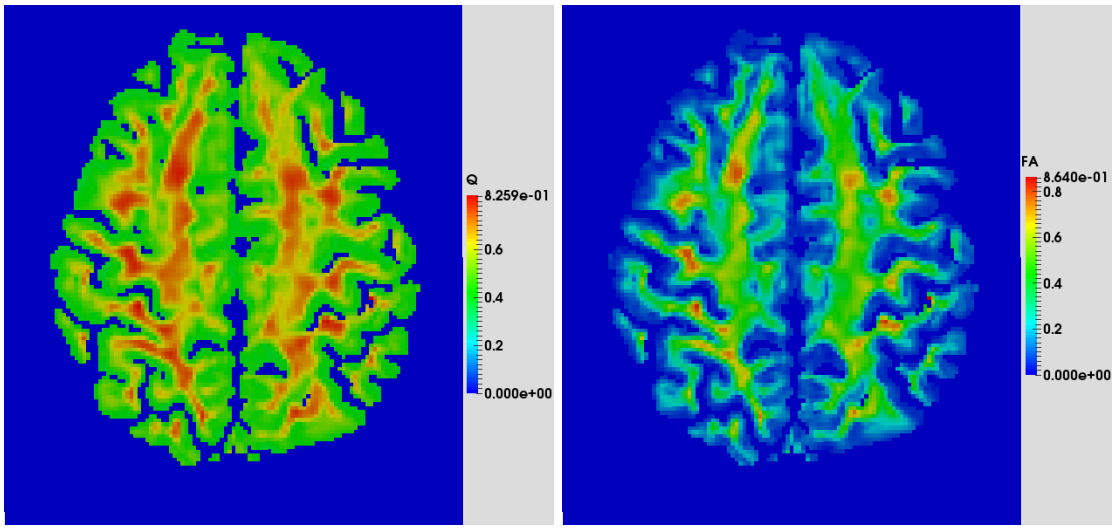


Figure 1: Difference between estimated Q and fractional anisotropy FA

while the function g is given via

$$g(Q(x)) = \lambda_1 \left(\lambda_0 + k^+ Q(x) + k^-\right)^{-1} f'(Q(x)).$$

We selected the growth functional coefficient μ to be $c_g(1-u)$, where c_g is the tumor growth rate[2], however other (more general) choices are possible as well. The constants c_g, λ_0, cell speed s and detachment rate k^- can be measured [2], [5], [16], [12]. The rate k^+ is assumed to be of the same order of k^- and the only free parameter λ_1 is selected as in [8].

The equation (6) supplemented with no-flux boundary conditions is well posed and the solution is nonnegative. This can be proven by standard arguments.

V. Numerical Results

We solve equation (6) with respect to homogeneous Neumann boundary conditions using a star like initial condition by way of a standard finite volume method. For the drift term we use an upwind procedure to guarantee stability. This is necessary to get a non-negative solution. Different methods like HDG-methods or discontinuous Galerkin methods are attractive, but do not conserve positivity in general. However, the latter is clearly required, so we cannot use them directly.

VI. Discussion

Due to the highly anisotropic spread and the fast invasion, gliomas have a very poor prognosis. The classical diagnostic methods (computer tomography or magnetic resonance imaging) cannot assess the complete extent of the tumor, but only the bulk of the cells. Hence mathematical modeling can be used to enhance the quality of a therapy relevant tumor margin.

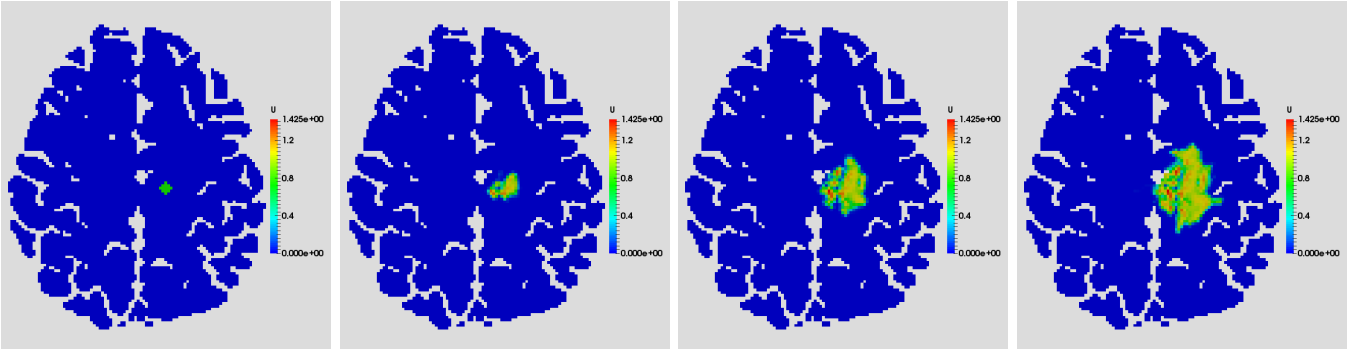
In characterizing proliferation, the presence of the rather general functions μ and χ ensure a higher model flexibility. The most adequate choice of these functions can be approached by using cell growth data, if this data becomes available.

[2]this corresponds to the usual logistic growth

The simulations we did (based on the DTI data) are in accordance to the clinically observed finger-like structures formed by glioma [6], [7]. On the other hand, existing (monoscale) models [14] can only predict a smoother, more compact shape of the tumor mass, hence they are not able to reveal the actual extent and shape of the neoplastic lesion.

There are still several interesting issues (both from mathematical and the biomedical point of view) concerning the multiscale model presented. On the mathematical side we cannot rigorously prove the convergence of our parabolic scaling approach to the limit equation. This is a highly nontrivial issue and only done in a monoscale setting, for a simple problem [4] not involving cell-tissue interactions. From the biological point of view, the most interesting issue is to look at groups of cells moving collectively. This will lead to adhesion terms. Unfortunately, these terms require finer space description than available in the data used. From the medical point of view, the model may be extended by considering therapy effects. We have proposed a way to do this in [13].

Another demanding issue is the verification of the model. It is known ([16], [21]) that cell surface receptors play a crucial role in cell motility. Hence it is necessary to include at least one internal variable term into a tumor model which aims at describing migration or tumor invasion. The mass action kinetics we used here is simplifying. There are different types of receptors playing different roles in the migration behavior and triggering multiple intracellular signalling cascades [21]. Hence the use of one variable for all types of intergrins is but a first step in the direction to a multiscale model describing tumor spread. The main problem is that exactly these subcellular signalling cascades are not known in detail and hence cannot be modeled in detail. Our simplified model represents a possibility to overcome these problems by using the most basic features of receptor dynamics (binding and unbinding, for which data for the corresponding rates is still available [16]). Hence it is questionable whether a validation on merely this part of the model would be reliable. Moreover, it is not our point to have the most exact model on the microscopic scale;



(a) Day 0 (b) Day 200 (c) Day 400 (d) Day 600

Figure 2: Simulation results for (6) with the estimated Q.

rather, the implications of subcellular (microscopic) dynamics on the macroscopic scale are the relevant ones from a practical (i.e., medical) point of view. For this aim, we have to check our model against real patient data - a problem which we are working on at the moment. The crucial part is the data. The verification of the simulated spread of a tumor would require at least one diffusion measurement (DTI / HARDI / QBall) as well as a segmentation at the initial time and a segmentation on the same grid at a later time point. In the time span between these times therapy effects should be excluded to get the pure migration and proliferation. Some of these measurements are available to us, but only with a time difference of around one month, which is exactly the time span between the tumor detection and start of a therapy. This is often a problem, as in this relatively short time span there are few alterations in the size and shape of the tumor, and this applies also to the numerical simulations.

References

[1] I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel. Reconstruction of the orientation distribution function in single and multiple shell q-ball imagaing with constant solid angle. *Magnetic Resonance in Medicine*, 64(2):554–566, 2010.

[2] A. Belkin, G. Tsurupa, E. Zemskov, Y. Veklich, J. Weisel, and L. Medved. Transglutaminase-mediated oligomerization of the fibrin(ogen) αC domains promotes integrin-dependent cell adhesion and signaling. *Blood*, 105(9):3561–3568, 2005.

[3] N. Bellomo. *Modeling Complex Living Systems. A Kinetic Theory and Stochastic Game Approach*. Birkhäuser, 2008.

[4] F. Chalub, P. Markovich, B. Perthame, and C. Schmeiser. Kinetic models for chemotaxis and their drift-diffusion limits. *Monatshefte für Mathematik*, 142(1-2):123–141, 2004.

[5] M. R. Chicoine and D. L. Silbergeld. Assessment of brain tumor cell motility in vivo and in vitro. *Journal of Neurosurgery*, 82(4):615–622, 1995.

[6] S. Coons. Anatomy and growth patterns of diffuse gliomas. pages 210–225, 1999.

[7] C. Daumas-Duport, P. Varlet, M. Tucker, F. Beuvon, P. Cervera, and J. Chodkiewicz. Oligodendrogliomas. part i: Patterns of growth, histological diagnosis, clinical and imaging correlations: A study of 153 cases. *Journal of Neuro-Oncology*, 34:37–59, 1997.

[8] C. Engwer, T. Hillen, M. Knappitsch, and C. Surulescu. Glioma follow white matter tracts; a multiscale DTI-based model. *J. of Math. Biol.*, 2014.

[9] C. Engwer, A. Hunt, and C. Surulescu. Effective equations for anisotropic glioma spread with proliferation: a multiscale approach and comparisons with previous settings. *Math.Med.Biol*, 2015.

[10] R. Erban and H. Othmer. From signal transduction to spatial pattern formation in *e. coli*: a paradigm for multiscale modeling in biology. *Multiscale Modeling and Simulation*, 3(2):362–394, 2005.

[11] D. Hanahan and R. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, 2011.

[12] T. Hoshino, C. Wilson, M. Rosenblum, and M. Barker. Chemotherapeutic implications of growth fraction and cell cycle time in glioblastomas. *J. Neurosurg.*, 43(2):127–35, 1975.

[13] A. Hunt and C. Surulescu. A multiscale modeling approach to glioma invasion with therapy. *preprint, TU Kaiserslautern, submitted*, pages –, 2015.

[14] A. Jbabdi, E. Mandonnet, H. Duffau, L. Capelle, K. Swanson, M. Pelegrini-Issac, R. Guillevin, and H. Benali. Simulation of anisotropic growth of low-grade gliomas using diffusion tensor imaging. *Mang. Res. Med.*, 54:616–624, 2005.

[15] J. Kelkel and C. Surulescu. A multiscale approach to cell migration in tissue networks. *Mathematical Models and Methods in Applied Sciences*, 23(3), 2012.

[16] D. Lauffenburger and J. Lindermann. *Receptors. Models for binding, trafficing and signaling*. Oxford University Press, 1993.

[17] K. Legate, S. Wickström, and R. Fässler. Genetic and cell biological analysis of integrin outside-in signaling. *Genes Dev.*, 23:397–418, 2009.

[18] T. Lorenz and C. Surulescu. On a class of multiscale cancer cell migration models: Well-posedness in less regular function spaces. *Math. Models Meth. Appl. Sci.*, 24, 2014. 54 pages.

[19] K. Painter and T. Hillen. Mathematical modelling of glioma growth: the use of diffusion tensor imaging (DTI) data to predict the anisotropic pathways of cancer invasion. *J. Theoretical Biol.*, 323:25–39, 2013.

[20] M. Reisert, I. Mader, R. Umarova, S. Maier, L. Tebartz van Elst, and V. G. Kiselev. Fiber density estimation from single q-shell diffusion imaging by tensor divergence. *NeuroImage*, 77:166–176, 2013.

[21] J. Uhm, C. Gladson, and J. Rao. The role of integrins in the malignant phenotype of gliomas. *Frontiers in Bioscience*, 4:d188–199, 1999.

Parameter Estimation for the Modelling and Simulation of Expanding Polyurethane Foams

Ikenna E. Ireka
[a]Fraunhofer ITWM, Kaiserslautern, Germany.
[b]Department of Mathematics Obafemi Awolowo University, Ile-Ife, Nigeria.
ireka@itwm.fhg.de

Dariusz Niedziela
Fraunhofer ITWM,
Kaiserslautern, Germany.
niedziel@itwm.fhg.de

Jürgen Tröltzsch
Department of Mech. Engineering,
TU Chemnitz, Germany
juergen.troeltzsch@mb.tu-chemnitz.de

***Abstract*—This paper presents a numerical study and approach for estimating associated parameters resulting from the mathematical modelling of self expanding polyurethane (PU) foams. In reaction injection molding of PU foams, the reactant polymer mixture undergoes phase transition resulting from exothermic chemical reaction, evolution of gas and network formation. Thereby, exhibiting the chemorheological property of the mixture viscosity which strongly influence the final structure of the expanded foam. More so, measuring this viscosity experimentally becomes extremely difficult, hence, the resolve to numerical investigations. However, to account for this behaviour mathematically, the foam viscosity is assumed to depend on the rate of conversion of the reactants mixture (degree of cure), temperature of the system and the volume fraction of evolving gas attributed to the chemical reaction. With this form of coupling in the state variables it becomes very challenging to estimate the associated model parameters analytically. We therefore explore a purely numerical approach in estimating some of these parameters. The coupled system of time dependent PDEs governing the foam expansion process is solved on a numerical simulation platform (CoRheoS) based on finite volume method. With graphical illustrations, we discuss the influence of these parameters on the foam viscosity and then validate our results with available experimental data.**

I. Introduction

Advances in reaction injection molding (RIM) process of Polyurethane (PU) foam have greatly impacted the global economy in a positive sense. Due to their interesting physical properties several industries ranging from automotive, aerospace, acoustic and structural industries (to mention a few) have engaged intensely in exploring the vast applications of PU foam products. Howbeit, there are still certain level of challenges encountered during the production phase of these foam materials. These difficulties are often attributed to the chemorheological properties (i.e. dependence of flow material variables on chemical reaction) of the mixture viscosity. Furthermore, measuring this viscosity experimentally is also very complicated, hence, the recourse to computational modelling approach.

PU foam production process involves chemical reaction between polyol and isocyanate functional groups in the presence of blowing agents and suitable catalyst. Depending on the blowing agent used (physical or chemical blowing) the reaction can become highly exothermic, thereby leading to evolution of gas and heat creation [1]. In addition, the expansion process and final structure of the foam are often characterized by spatial and temporal changes that occur in the reactant mixture [2]–[4]. Therefore, creating appropriate mathematical models capable of describing the physics as well as the interplay between relevant state variables in the system becomes quite challenging. Nonetheless, adequate understanding of the complex behaviour exhibited by these self expanding PU foams will aid process and design engineers in optimizing the production processes of these foams [3].

Due to the vast applications of PU foams, several theoretical and experimental studies, [4]–[12], aimed at understanding and/or predicting some attributes of expanding foam systems have been conducted. In the work of Lefebvre and Keunings [7], a two dimensional mathematical model was used to describe the flow of a reacting polymer system under certain time and geometric conditions. Results from this study predicted the propagation of the macroscopic temperature, velocity, material stresses as well as species concentration in the flow. In a related study aimed at predicting the flow field, distribution of foam density, thermal conductivity and the progression of the flow front in mold filling processes of PU foam [10], a theoretical three dimensional model, which extends the work of Baser [8] was derived and solved numerically. The setup assumed empirical models for the densities and viscosity of bubble suspension in the homogeneous phase of the reactant mixture. Their results gave good prediction for the state variables of interest.

Furthermore, a mathematical framework capable of describing the self-expanding process of physically blown foams from a reactive polymer system contained in a closed complex geometry was presented in [13]. Here, an empirical density model which depends on time and temperature was assumed. Their numerical results were validated with available experimental data and they showed good qualitative agreement, although some physical attributes were lost to the modelling. On the other hand, Bikard et al [11], [12] adopted a mesoscale modelling approach for the foam growth phenomena which was based on the expansion of gas bubbles within the rheologically transforming polymer matrix.

One important feature in the study of PU foams is the degree at which the foaming mixture cures [3], [14]. This phenomenon plays an crucial role in the spatio-temporal

evolution of the mixture viscosity. Hence, as a critical gel point (ζ_c) is approached the foaming mixture evolves in its material property such that the mixture viscosity grows infinitly. Therefore, to describe the overall foam expansion process mathematically one ought to account for intricacies such as the mixture viscosity, degree of cure/polymerization, temperature and gas/liquid volume fraction. The resulting model equations are often transient and highly coupled with parameters that are difficult to estimate analytically.

In our previous work [4], we proposed and implemented an experimentally motivated mathematical model capable of predicting the foam expansion process. In addition, we obtained both qualitative and quantitative agreements of the spatio-temporal distribution of temperature when compared with available experimental data. Our results gave some insights on the spatial distribution of viscosity in the expanding foam system. However, to quantify the viscosity of most fluids experimentally, the torque ($\mathbf{M}$) response obtained from rheometric measurement is often a vital resource [5], [15]. In our recent investigation, [16], we estimated the viscosity of expanding PU foams (experimentally) using the information from viscometric torque measurements and a simplified Newtonian fluid assumption for the reacting mixture as in [5]. In this current study however, we will adopt the modelling approach of [4] to numerically investigate the influence of some relevant model parameters on the expanding foam system in the viscometeric setup described in [16].

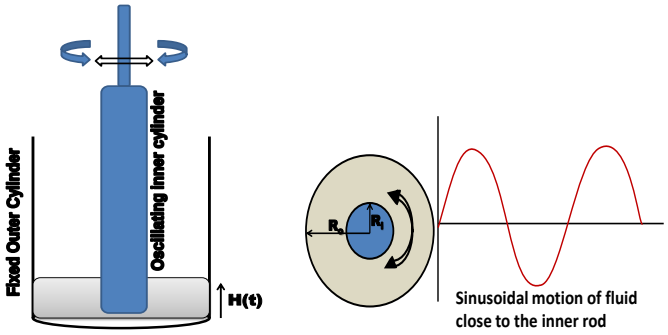
Our preliminary analysis will center on addressing the following questions. " How fast does the foaming mixture reach it's gelling point (ζ_c)? Secondly, what effect does the time to gel (influenced by the degree of polymerization) have on the mixture viscosity considering the torque measurements from experiments? "

The resulting transient coupled nonlinear system of partial differential equations (PDEs) governing the physical process are solved via finite volume method. In particular, we carry out numerical simulations of the relevant torque measurement setup presented in [16]. Furthermore, we present and discuss our results using graphical illustrations and then validate them with the available experimental data.

The remaining part of the paper is structured as follows: In section II we present a description of the mathematical framework for the setup under investigation and in III we present and discuss our numerical results. Finally in section IV we give some concluding remarks for the paper.

II. Mathematical Formulation

To model the torque measurement experiment setup described in [16] we assume a concentric cylinder of inner radius R_i and outer radius R_o with both cylinders closed at one end. Hence, the exothermic reacting pseudo-homogeneous mixture of polyol and isocyanate group is contained in the gap between them, see setup in Fig.1. In the experiment, the outer cylinder is fixed while the inner cylinder oscillates in a periodic manner through an angle θ in rads with angular speed ω rads/s. As in [4], we assume that the foaming mixture



(a) (b)

Fig. 1. Geometric description of the viscometeric device used to obtain the torque measurements: (a) Front view of the setup, (b) Aerial view of the setup.

expands uniformly in the domain over time, so that at any given time the height ($H(t)$) is obtained. In addition, we obtain the relevant expansion source term necessary for the conservation of mass (details of the modelling can be found in [4], [17].

In line with [4], the system of equations governing the expansion process are given by:

Conservation of mass:

$$\nabla . \mathbf{v} = S_p, \tag{1}$$

Conservation of momentum

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) = \nabla \cdot \eta_m \mathbf{D} - \nabla P + \rho \mathbf{g} , \tag{2}$$

Conservation of energy

$$\rho C_p \left(\frac{\partial T}{\partial t} + \mathbf{v} . \nabla T \right) = \nabla \cdot (k \nabla T) + \frac{1}{2} (\eta_m \mathbf{D} : \mathbf{D}) + \rho H_R \frac{d\zeta}{dt}, \tag{3}$$

Degree of cure

$$\frac{d\zeta}{dt} \equiv \frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla \zeta = (k_1 + k_2 \zeta^m)(1-\zeta)^n, \tag{4}$$

Volume fraction of liquid

$$\frac{\partial \varphi}{\partial t} + \mathbf{v} . \nabla \varphi = -\varphi S_p, \tag{5}$$

Viscosity model

$$\eta_m = \begin{cases} \eta_F, & \text{for} \quad \varphi \geq \varphi_s, \\ \eta_F \frac{\varphi}{\varphi_s} + \eta_g \left(1 - \frac{\varphi}{\varphi_s}\right), & \text{for} \quad 0 \leq \varphi < \varphi_s \ , \end{cases} \tag{6}$$

and

$$\eta_F = \eta_{00} \exp\left(\frac{E_\eta}{RT}\right) \cdot \left(\frac{\zeta_c}{\zeta_c - \zeta}\right)^{1.5+\zeta} (0.8 - 1.2\varphi_g + 0.5\varphi_g^2). \tag{7}$$

$\mathbf{v}$ is the flow velocity, S_p is a time dependent expansion source term which drives the foam expansion process, ρ is the mixture density, $\mathbf{D} = \nabla \mathbf{v} + (\nabla \mathbf{v})^t$ is the rate of strain

tensor and C_p is the heat capacity of the reacting mixture. k is the thermal conductivity, P and T are respectively the pressure and temperature in the system, H_R is the heat of reaction and $\mathbf{g}$ is acceleration due to gravity. Furthermore, the degree at which the reacting mixture cures/polymerizes is represented by ζ and ζ_c is the gelling point. At the gelling point, the material properties of the mixture transforms from a purely viscous liquid to a viscoelastic foam matrix. k_1 and k_2 are Arrhenius rate laws for the primary (isocyanate - polyol) reaction and secondary (water-isocyanate) reactions and the respective reaction order n and m are constants such that $m + n = 2$ and $n < m$. In general, k_1 and k_2 are defined by

$$k_j = A_j \exp(-E_j/RT), \tag{8}$$

where A_j and $E_j, j \in \{1, 2\}$ are the Arrhenius prefactor and activation energy respectively. For simplicity however, k_1 and k_2 (to be obtained in this study) are assumed to be constant. In addition, φ and φ_g are the volume fraction of the liquid mixture and gas respectively. η_m and η_F are the mixture and foam viscosity respectively. φ_s is a switch parameter see [4] for description. The activation energy of the expanding foam is given by E_η, R is the rate constant and η_{00} is a prefactor for the foam viscosity.

Heat transfer process in reaction molding can be described as a one dimensional static system [5]. Hence, to obtain appropriate input parameter for our 3D simulations we first study (in 1D) the influence of the associated state variables and parameters on the foam viscosity as described by equation (7). In this regard, our preliminary analysis will hinge on numerical solutions to the 1D heat equation (i.e. equation (3) without the convective and viscous heating terms) and the equation for the degree of cure equation (4) without the convection term. Furthermore, we adopt the prediction from these 1D results and obtain appropriate input parameters k_1, k_2 and η_{00} . Hence, the emphasis of this study is on the choice of parameters for the degree of cure adequate for our 3D simulation of the torque measurements.

In our 3D simulations the equations of state (equations 1-5) are solved with the same approach as in [4] under the same boundary conditions except for the temperature conditions and velocity at the rotating walls. On this boundary we prescribe a Robin boundary condition for the temperature in the form

$$\alpha T + (1 - \alpha)\nabla T|^{<} \cdot \mathbf{n} = \alpha T_o, \tag{9}$$

where $0 < \alpha < 1$, $T_o = 25^oC$ is the room temperature and $|^{<}$ is indicates heat flux into the inner cylinder. In addition, the velocity on the wall corresponding to this inner cylinder is also prescribed as

$$\mathbf{v} = A \sin(\omega t), \tag{10}$$

where A is the amplitude of oscillation and ω is the angular speed. t is time in seconds. All our simulations start at the onset of expansion which is usually about 40 seconds after injecting the reacting mixture into the mold.

In the next section we present some numerical results describing the heat transfer process, degree of cure and viscosity in the domain and discuss how the associated viscosity parameter affects the time to cure which in retrospect affects the torque. We also present some results from our 3D simulations.

III. Numerical Results and Discussion

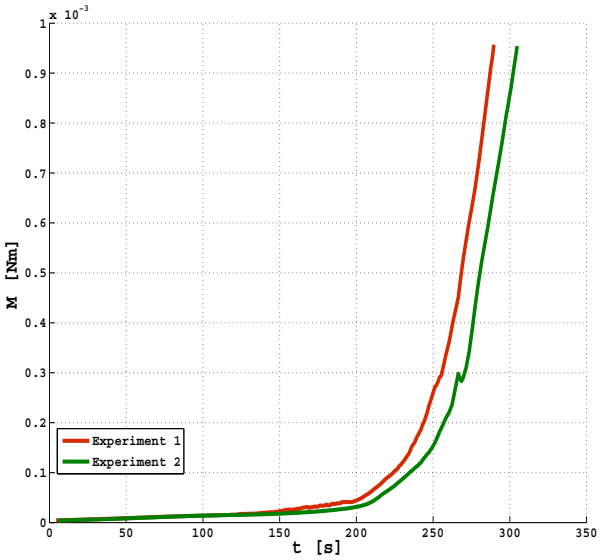


Fig. 2. Torque measurement from experiments.

As earlier mentioned, the focus of this paper is on estimating appropriate parameters for our 3D simulations of the torque experiment described in section II as well as understanding the interplay between flow viscosity, degree of cure and temperature in the expanding foam. We note that the spatial changes of viscosity (in the system) is driven by local variations in temperature and the degree of cure.

Preliminary investigation of experimental data (from the torque measurements) available to us, Fig. 2, suggests possibility of a slowly gelling process (in time) before the gel point ζ_c is reached. The torque values is seen to increase gradually in the first 200 seconds followed by a rapid growth over a shorter period of time. This rapid growth characterized by the gellation of the foam mixture signifies a transformation from a viscous liquid to a viscoelastic material. This information is therefore vital for our choice of the cure parameters (k_1 and k_2) in equation (4). Generally, the coupling between the degree of cure ζ (for varying k_1, and k_2 values) and temperature is such that at higher temperature the foam mixture gels faster, see for example [14]. Therefore, under adiabatic conditions, one expects the mixture to polymerize/gel relatively faster. From Figs. 3 and 4 the effect of the parameters k_1 and k_2 on the gelling process is such that as these values decrease we notice a relative time delay in the gelling process before ζ_c is reached. The consequence of this time delay on the viscosity is shown in Figs. 5 and 6. Therefore, as each of k_1 and k_2 decreases the viscosity slowly approach infinity in time. This observation

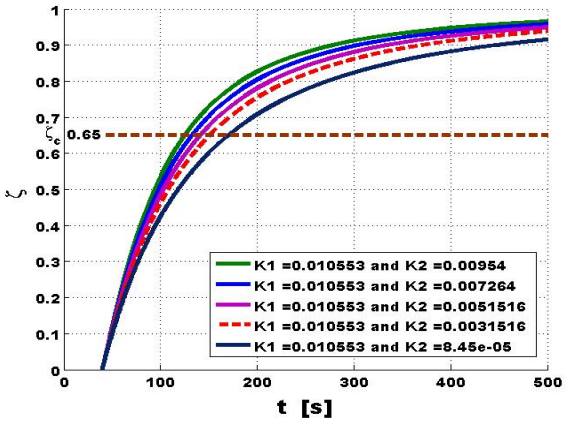


Fig. 3. Influence of the cure parameter k_1 and k_2 on the degree of cure for fixed k_1 and different values for k_2: *The brown dashed line indicates the critical gel point ζ_c chosen from literature to be 0.65 for the given material.*

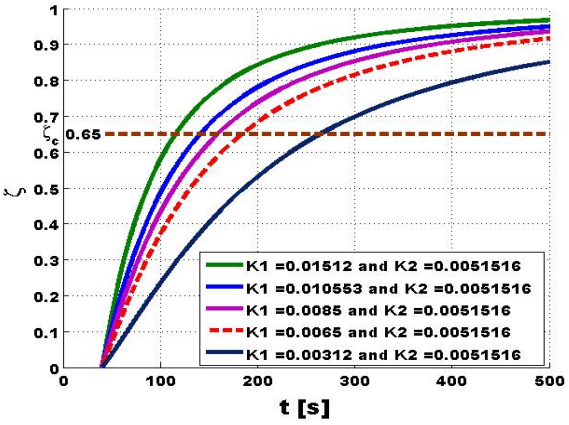


Fig. 4. The effect of the cure parameter k_1 and k_2 on the degree of cure for fixed k_1 and different values for k_2: *The brown dashed line indicates the critical gel point ζ_c chosen from literature to be 0.65 for the given material.*

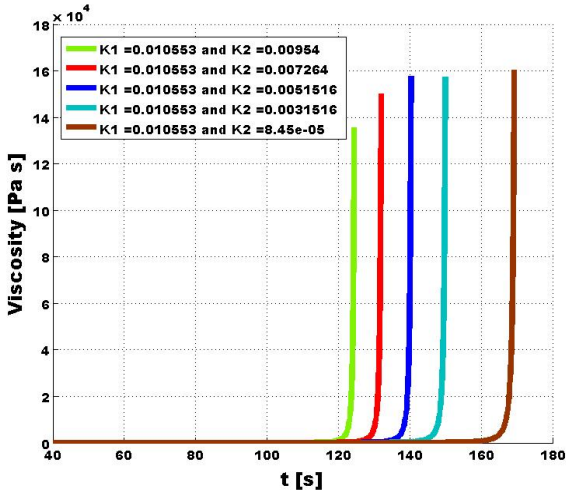


Fig. 5. Sensistivity of foam viscosity η_F to the cure parameters k_1 and k_2 for fixed k_2 and various k_1 values.

therefore inform us that a slowly gelling process will result to

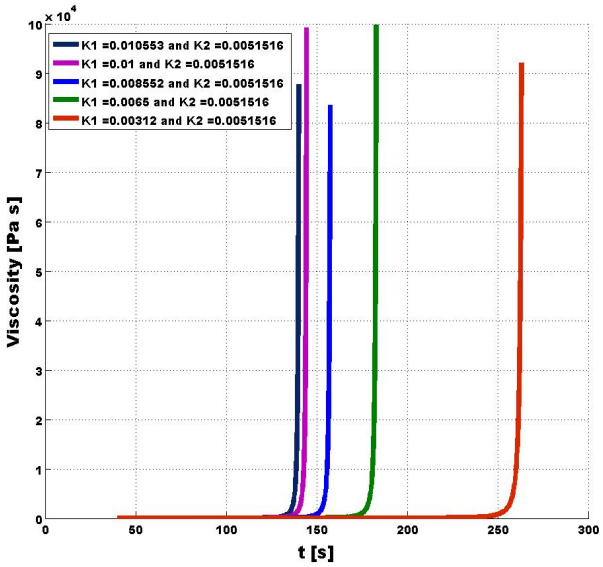


Fig. 6. Graphical illustration of the influence of k_1 and k_2 on the foam viscosity η_F for fixed k_1 and different k_2 values.

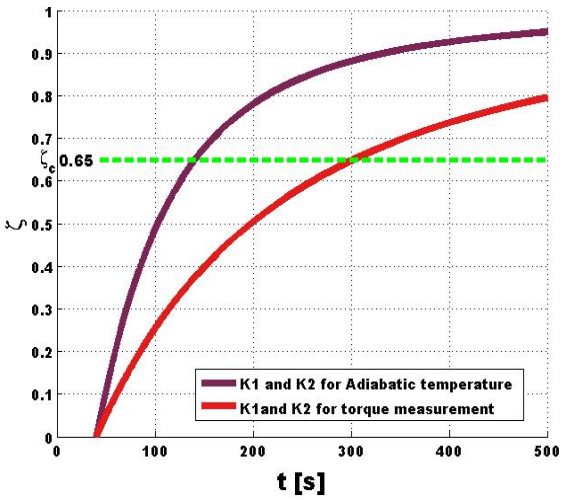


Fig. 7. Comparison of the degree of cure for values of k_1 and k_2 chosen for adiabatic temperature measurement and torque measurement: *The green line indicates the critical gel point ζ_c chosen to be 0.65 for the given material. Observe the significant variation in the time to reach $\zeta_c = 0.65$ in both cases.*

a corresponding but gradual increase in the torque values.

In [4], we obtained k_1 and k_2 values for the degree of polymerization in an expanding foam setup under adiabatic conditions Fig. 7. The viscosity for this choice of k_1 and k_2 values is observed to increase to infinity (rapidly) in a short time Fig. 8. However, since there are some heat loss to the environment during the torque measurement experiment our strategy is therefore to estimate k_1 and k_2 values which will induce an appropriate time delay in the gellation process before ζ_c is reached, Fig. 7, thereby, affecting the viscosity of the foaming mixture in a similar manner Fig. 9. This therefore provides us some insight on what to expect in our 3D simulations of the torque experiment. In addition, these

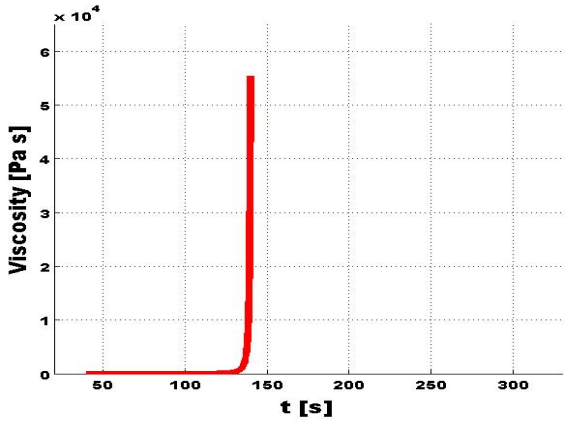


Fig. 8. Rapid temporal changes in the foam viscosity η_F for adiabatic values of k_1 and k_2.

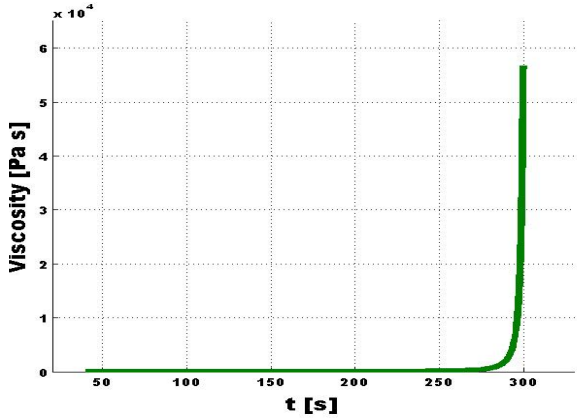


Fig. 9. Time delay of gellation induced by low values of k_1 and k_2 reflected in the foam viscosity η_F.

k_1 and k_2 serves as as input to our 3D simulations of the experiment.To validate the results from our 3D simulation, we compare the torque measurements from experiments with those from simulation Fig. 10. Our results show qualitative agreement with those from experiment. Though our emphasis in this study is on viscosity, however, to further validate our 3D numerical solutions, the measured time change of temperature at a given position in the domain of the experiment setup is compared with those from our simulation (obtained at the same point). Our result also show very good agreement with the experimental temperature data, see Fig. 11.

Although our results show good agreement with those from experiment, we note that optimally, the mathematical description for the cure parameters k_1 and k_2 show a dependence of these parameters on temperature. Accounting for temperature dependence in these cure parameters makes the problem highly coupled and more complicated. In this case we have to estimate the unknown Arrhenius prefactors A_1 and A_2 and the activation energies E_1 and E_2 in both k_1 and k_2. However, we are currently investigating approaches to achieve Arrhenius dependence of the cure parameters on temperature.

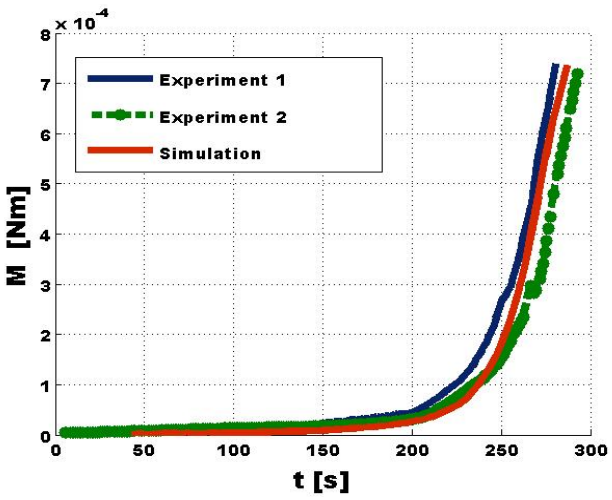


Fig. 10. Comparison of torque measurement from experiments with result from our simulation.

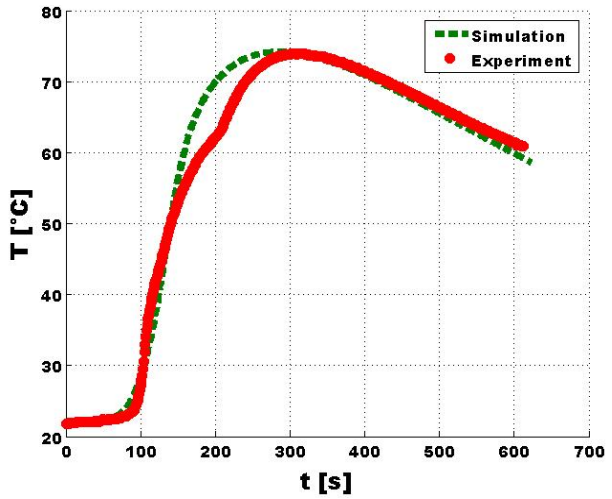


Fig. 11. Plots comparing temperature measurement from (the torque) experiment with that from our simulation.

IV. Conclusion

We have numerically studied the influence of the relevant cure parameters necessary for predicting the time change of viscosity in the expanding PU foam system. This analysis was purely for the purpose of obtaining adequate input parameters for our 3D simulation of the torque measurement experiment discussed in our earlier work [16].

Available torque measurement data provided insight on the behaviour of the material viscosity over time. Hence, from our 1D simulations and analysis we were able to estimate the values for the cure parameters so that the corresponding viscosity behaviour agrees with the information we could gather from the torque measurement data. As earlier mentioned, it will be adequate to make the choice of the cure parameter (k_1 and k_2) independent of any particular experiment. To achieve this, it is necessary to account for temperature dependence of these cure parameters. In this regard, it is worth mentioning that

work is currently underway to estimate the four parameters that would arise from making k_1 and k_2 temperature dependent in a general way.

REFERENCES

[1] Kaneyoshi Ashida (2007), Polyurethane and Related Foams: Chemistry and Technology, CRC Press Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 334872742, ISBN 1587161591.

[2] Lindt J. T. and W. Kostrzewski, (1981) Flow Mechanics of Polyurethan Foam Formation, (Chapter 13) Urethane Chemistry and Applications, *Editor Edwards N. Kenneth*. American Chemical Society Symposium 172: Washington DC.

[3] A. Yousefi and P. G. Lafleur (1997), Kinetic Studies of Thermoset Cure Reactions: A Review. Polymer Composites, 18(2):157-168.

[4] I.E. Ireka, D. Niedziela, K. Schäfer, J. Tröltzsch, K. Steiner, F. Helbig, T. Chinyoka, and L. Kroll, (2015) Computational Modelling of the Complex Dynamics of Chemically Blown Polyurethane Foam, Physics of Fluids, 27, 113102, doi: 10. 1063/1.4935788.

[5] E. Broyer, C. W. Macosko, F. E. Critchfield an L. F. Lawler (1978), Curing and Heat Transfer in polyurethane Reaction Molding. Polymer Engineering and Science, 18 (5):382 - 387

[6] J. D. Domine and C. G. Gogos (1980), Simulations of reactive Injection molding, Polymer Engineering and Science, 20 (13):847 - 858

[7] L. Lefebvre and R. Keunings (1995), Finite Element Modelling of The Flow of Chemically Reactive Polymeric Liquids, Int. Journal for Numerical Methods in Fluids, 20:319-334.

[8] S. A. BASER and D. V. Khakhar (1994), Modelling of the Dynamics of Water and R-11 Blown Polyurethane Foam Formation Polymer Engineering and Science, 34 (8):643 - 649

[9] D. Seo, J. R. Youn and C. L. Tucker III (2003), Numerical simulation of mold filling in foam reaction injection molding, International Journal of Numerical Methods in Fluids, 42:1105 -1134,

[10] D. Seo and J. R. Youn (2005), Numerical Analysis on Reaction Injection molding of Polyurethane Foam by Using Finite Volume Method. Polymer, 46:6482 - 6493,

[11] J. Bikard, J. Bruchon, T. Coupez and B. Vergnes (2005), Numerical prediction of the foam structure of polymeric materials by 3D simulation of their expansion by chemical reaction based on multidomain method. Mechanical Behaviour of Cellular Solids: Journal of Materials Science, 40: 5875 - 5881.

[12] J. Bikard, J. Bruchon, T. Coupez and L. Silva (2007), Numerical Simulation of 3D Polyurethane expansion during manufacturing process. Colloids and Surfaces A: Physicochem Eng. Aspect, 309:49-63.

[13] R. R. Rao, L. A. Mondy, D. R. Noble, H. K. Moffat, D. B. Adolf and P. K. Notz (2011), A level set method study of foam processing: a validation study, International journal for Numerical Methods in Fluids, 68:1362-1392.

[14] M. C. Saha, B. Barua, S. Mohan (2011), Study on the Cure Kinetic Behavior of Thermosetting Polyurethane Solids and Foams: Effect of Temperature, Density, and Carbon Nanofiber. Journal of Engineering Materials and Technology, 133: 0110153 -6.

[15] C. Macosko, (1994), Rheology: Principles, Measurements and Applications. Wiley-VCH ISBN 0-471-18575-2.

[16] D. Niedziela, I. Ireka, K. Shäffer, J. Tröltzsch, Fr. Helbig, K. Steiner, L. Kroll. (2015), Computational Modelling of Polyurethane expansions for Lightweight Sandwich structures 2nd International MERGE Technologies conference IMTC 2015 Lightweight Structures. MERGE conference proceedings. Oct 1-2 2015, TU Chemnitz. ISBN 978-3-95t35-025-1

[17] I. E. Ireka, (2015), Computational Analysis of Non-Isothermal Flow of Non-Newtonian Fluids, Phd Thesis, University of Cape-Town, Cape-Town, South Africa.

Code coverage metrics to support testing of concurrent software

Jasmin Jahić
Fraunhofer IESE
Kaiserslautern, Germany
jasmin.jahic@iese.fraunhofer.de

***Abstract*—Multicore processors execute programs concurrently, but program interactions can lead to concurrency bugs. Testing all interaction combinations is expensive. We introduce coverage metrics for concurrent software to quantify testing and reduce the number of test cases. Interactions are mapped to existing coverage metrics of sequential programs.**

Keywords— Multicore, testing, code coverage, LLVM

I. INTRODUCTION

According to Moore's Law, the number of transistors on a microchip doubles every year. In the years following the formulation of this law, the increase in the number of transistors on a microchip was followed by an increase in the working frequency of transistors. Consequently, processors became more complex, with ever increasing clock speed. While the increase in performance was mainly driven by raising the clock speed, the complex structure of the processor supported elements for efficient power consumption, enhancement of computing technologies implemented directly in hardware, etc. However, from 2004 onwards, certain aspects of semiconductor technology have been stalling (Figure 1). Further frequency increases are unlikely due to the nonlinear increase in heating and power consumption. Today, the only valid aspect of Moore's Law is the doubled number of transistors on a microchip every year. The law will be valid as long as we do not reach the physical limits of silicon semiconductor technology, but the clock speed of the processors will no longer rise significantly. Actually, it might even be reduced to mitigate negative effects of shrinking transistor's surface.

Software continues to require more computation power. Instead of trying to achieve higher computing power by increasing the frequency of processors, the industry has turned to the idea of processing several tasks concurrently. Concurrent processing considers the concurrent execution of program parts on several computing units (cores) and merging the results of the execution. The performance of multicore processors is driven by the number of cores and does not necessarily require an increase in frequency. Higher numbers of cores on a microchip are still supported by Moore's law, as we continue to shrink the surface of transistors. However, the new multicore technology does have its challenges. Existing software must be parallelized - divided into parts that can be executed concurrently. The software parts executed concurrently (threads) can access the same memory locations. Access to the shared memory must be synchronized, otherwise the access results in invalid or inconsistent data (Example 1.a).

// thread t1	*// thread t2*
…	*…*
account=account+v;	*account=account-v;*
…	*…*
Sequence 1: - t1 and t2 read the account value. - t1: write back new account value. - t2: write back new account value. Operation performed by t1 is overwritten. Sequence 2: - t1 and t2 read the account value. - t2: write back new account value. - t1: write back new account value. Operation performed by t2 is overwritten.	

a)

// thread t1	*// thread t2*
acquire (accLock);	*acquire (accLock);*
account=account+v;	*account=account-v;*
release (accLock);	*release (accLock);*
Sequence: - t1: acuire lock on the account. - t2: try to acuire lock on the account. - t2: wait until the lock is released. - t1: read the account value. - t1: write back new account value. - t1: release the lock on the account. - t2: acuire lock on the account. - t2: read the account value. - t2: write back new account value. - t1: release the lock on the account.	

b)

Example 1: Concurrent access to the shared variable account without any synchronization mechanism (a) and with a synchronization mechanism (b)

The invalid data is a consequence of a special unsynchronized access order performed by two concurrent threads. In (Example 1.a), the bug occurs only if one thread reads the value of the shared variable and does not write back the results of its processing before the second thread reads the value of the same variable. In this case, the second thread operates on the old value of the variable and is unaware of the processing done by the first thread. Whichever thread is the last to write the value of the shared variable will overwrite operations performed by the other thread as both are processing the same old, not updated value and hence will ignore all the processing done by the other thread. In order to cope with these challenges, several synchronization mechanisms have been developed. One of the most frequently used synchronization mechanisms in software development today is based on locking the shared resources. Properly synchronized using a lock synchronization mechanism, the previous example is shown in (Example 1.b).

The synchronization mechanisms are the solution for the shared memory challenge. However, the correctness of concurrent software depends on the correct implementation of the synchronization mechanisms. Concurrent software is developed by humans and is error-prone due to various factors. For example, new developers often do not have an overview of the system and are unaware of shared variables between threads. A bug is created when a developer forgets to use or partially uses synchronization, as shown in (Example 2).

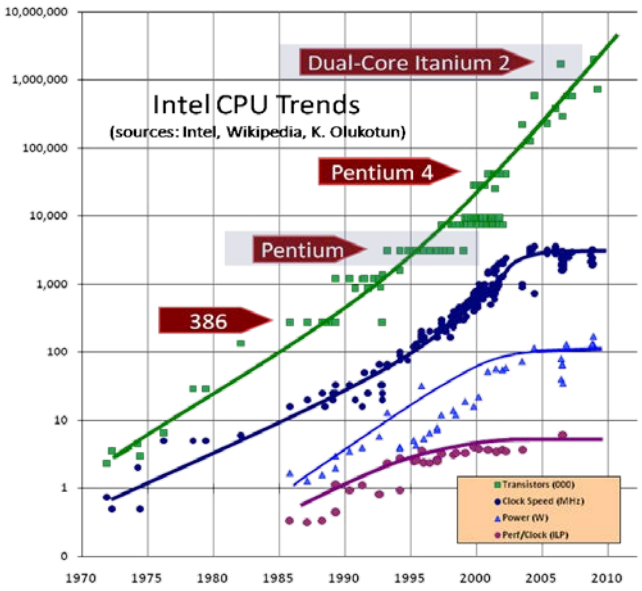


Figure 1: Intel CPU trends [18], different aspects of Moor's law

Often, quick fixes that include access to shared variables introduce hidden concurrency bugs. So the real challenge in concurrent software development is to ensure that a developer has correctly synchronized access to shared variables, which is a difficult task as the developer needs to think in several concurrent dimensions. The introduced bug will remain hidden and will be present in the delivered software if the specific order of shared-memory accesses leading to the bug is not observed during testing. There exist various techniques for testing concurrent software (cf. Related Work section), but there also exists a gap in quantifying the tested coverage of concurrent software. Coverage of concurrent software must include coverage of the individual software threads and their interaction. Coverage metrics for sequential software can be applied to the individual software threads, but their interaction is a multicore-specific challenge. Due to the lack of proper coverage metrics, it is hard to claim that the software is free from concurrency bugs and to have confidence in the correctness of concurrent software, especially if it is to be deployed in safety-critical systems. Concurrency bugs have caused severe damages in the real world. One of the most infamous examples is Therac-25, a radiation therapy machine, which due to a concurrency bug sometimes gave its patients radiation doses that were hundreds of times greater than normal [19]. Therefore, new approaches for quantifying the coverage of concurrent software are needed. In this paper, we are introducing coverage metrics for quantifying concurrent software testing. Our approach maps the interactions of mutually concurrent threads to the coverage metrics of sequential programs. Our proposed coverage metrics is complementary with existing algorithms for concurrency bugs detection. It is based on the code coverage metrics used for sequential code and it reduces a number of needed test cases for full coverage. Section III describes the conceptual solution, while in the section IV we demonstrate an implementation of the concept. Section V is reserved for a discussion about the constraints of our approach and ideas for future work.

II. Related work

Software testing involves the execution of a software unit to evaluate one or more properties of interest. If the result of the evaluation is different than expected, the software contains a bug. The software unit is a self-contained group of code statements, with input and output that together perform a task. In sequential software, a software unit is tested across different inputs. As the input data is not a limited set, the number of test cases is practically infinite. In order to solve this challenge, coverage metrics have been designed to quantify code testing. When achieved, coverage metrics guarantee a certain quality level of the code.

However, the behavior of concurrent software units depends on the input data and on the interaction among concurrent threads. A simple extension of sequential testing techniques for multicore software is to set up a testing environment that enables testing of all concurrent shared-memory access combinations. This is a challenging task, as the setup of such a testing environment requires a lot of effort [10]. Also, the number of test cases increases exponentially with every memory access, leading to an explosion of the test space. A common approach in concurrent software testing is to complement the test cases with an analysis component. Test cases are used to gather the execution traces through the execution of software units. An execution trace is a set of executed instructions, such as memory access, branching instructions, function calls, etc. (*Example 2*). The analysis component is responsible for extracting the data from the execution trace and exposing concurrency bugs by applying different algorithms.

A. Code coverage metrics

Software is a collection of software units (*Figure 2.b*). If a software unit contains control flow structures, a single execution of the software unit does not guarantee execution of all statements. However, even the statement coverage – the execution of every statement – does not guarantee the absence of bugs. Control structures (such as *if* statements) create execution branches. The evaluation of the control structure decision determines which software branch is executed. The decision is defined as a Boolean expression composed of conditions and zero or more Boolean operators. Individually correct statements can lead to a bug when combined in a specific branch. Branch coverage is achieved when every decision is evaluated at least once to true and false.

Complex program structures include nested decisions leading to branches and sub-branches. Besides the evaluation of a decision to true or false, it is important to reason about which conditions influence the decision, as they might consequently influence the execution of sub-branches. Condition coverage measures which condition has been evaluated both to true and false. The condition is a Boolean expression that cannot be broken down into simpler Boolean expressions. Path coverage measures which paths (routes)

through the software unit have been executed. Full path coverage is usually impractical or impossible to achieve as a software unit with a succession of n decisions can have up to 2^n paths within it. Basis path coverage is based on achieving complete branch coverage without achieving complete path coverage [2]. The basis path technique covers a set of linearly independent paths of execution. Modified Condition/Decision Coverage (MC/DC) considers that every point of entry and exit in the program has been invoked at least once, every condition in a decision in the program has taken all possible outcomes at least once, and each condition has been shown to affect that decision outcome independently [3]. Coverage analysis of concurrent software has to date been mostly researched via reachability analysis [4, 5] or generation of thread schedules to fulfill coverage requirements [6]. In the context of concurrently executed code, notable work has been done in quantifying coverage based on the level of Concurrent Function Pairs (CFP) [7].

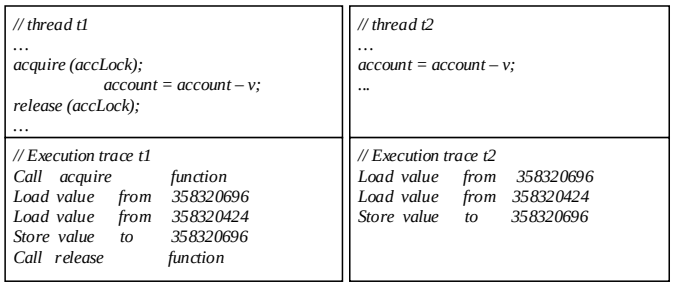
B. *Concurrent software testing*

There exist many testing approaches that base the detection of concurrency bugs on the analysis of execution traces [11-14]. They mainly differ w.r.t. the approach used for the collection of the execution traces and mostly use variations of the same algorithms for analyzing the collected trace. Lockset and happens-before are algorithms commonly used in concurrent software testing approaches. The lockset algorithm [8] is based on two sets of locks. Candidate locksets C(v) contains all locks that may be protecting the variable, while locks_held(t) contains the locks currently held by a thread. If at any access to v, $C(v) \cap locks\_held(t)$ is an empty set, there is a potential concurrency bug. The happens-before algorithm is based on a relation between the results of two events. If one event should happen before another event, the result must reflect that, even if those events are in reality executed out of order [9].

Approaches based on static analysis and model checking [15-17] are used for the detection of concurrency bugs, but are beyond the scope of this paper.

III. CONCEPT OF CONCURRENT SOFTWARE CODE COVERAGE

A software program is composed from a set of instructions. In sequential software, testing ensures the absence of bugs by executing all instructions and relevant combinations of instructions in a software unit under test. In industry, MC/DC coverage is used for compliance with the highest safety levels (avionics software development DO-178B and DO-178C Level A, Automotive Safety Integrity Level D). However, concurrent software units are mutually dependent. Achieving MC/DC coverage of a software unit individually does not guarantee the absence of bugs, as the concurrent bugs are related to the interaction between threads over the shared memory. Execution of all shared-memory access combinations is not practical. Observation of the dynamic behavior is performed through the collection of an execution trace (Example 2). Commonly used, synchronization locking mechanisms are implemented as function calls (e.g., acquire() for locking and release() for unlocking). From the execution trace, the analysis component calculates concurrency-related data of the software, such as shared memory access and use of synchronization mechanisms. There exist different algorithms and different approaches for analyzing concurrency data. Every concurrency analysis approach has a set of rules that are applied to the calculated concurrency data to reason about whether an access to a shared resource is a concurrency bug.



Example 2: Access to shared resources and the resulting execution trace

Software is composed from software units, where each software unit has a specific structure defined by its source code (Figure 2.b). Software under test can be represented on different levels of abstraction (e.g., source code, intermediate representation, binary executable, etc.). The prerequisite of our approach is that a software structure – a list of software units - can be derived from the software representation in a structure that enables the calculation of execution paths. Consequently, it must be possible to derive the structure of every software unit, with condition structures. The structure of a software unit can be seen as a flow graph with conditional and unconditional blocks (Figure 2.a). A block represents a set of instructions, which, if entered, is fully executed.

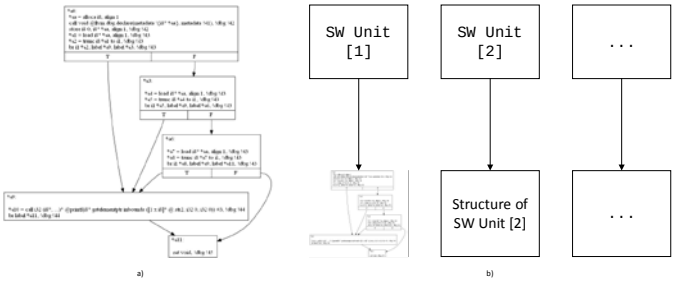


Figure 2: SW unit structure (a), SW composed from distinct SW units (b)

On the logical level, concurrent software consists of runnables which are equivalent to a specific tasks (e.g. Control Throttle runnable). A thread executes a runnable composed of at least one software unit. The initial software unit of every runnable is known in advance. However, determining other software units composing a runnable is an undecidable task (Example 3) and can only be determined during execution. The mapping of software units to runnables is unknown a priori. Therefore, the first challenge in quantifying code coverage is to determine the scope of every runnable (Figure 4). The scope of a runnable is determined by software units composing a runnable and their execution paths. Our approach incrementally builds a

scope of a runnable by analyzing the execution trace. For reasons of simplicity, let us assume that the software units are equivalent to functions. If the execution trace of a runnable contains a function call, then the called function belongs to the scope of the runnable. Software execution is characterized by an execution path, where every execution path does not necessarily include the execution of all code statements belonging to a runnable. Whenever an execution path that includes an access to some new function is executed, the scope of the runnable is incrementally expanded for the new function. The structure of the new function is analyzed and its execution paths are calculated. The execution path of a software unit can be reconstructed from the software unit structure and the branching instructions present in the execution trace. The scope of a runnable is complete when path coverage of all associated software units has been achieved. If all execution paths of a software unit have been executed, then the complete execution trace for the software unit is collected.

An access to a shared variable is a part of an execution path. Synchronization mechanisms invoked on the execution path define the context of the access to the shared variable. The complete execution trace of a software unit contains access to every variable, in every possible context. Common causes of concurrency bugs are situations where an access to a shared variable is synchronized only for some execution paths, or when accesses to the shared resource are synchronized with different locks. For developers, it is very hard to notice these types of errors. The worst attribute of the concurrency software is that it will function flawlessly, until the specific combination of memory accesses is observed. In some cases, it means that software will function several years without problems, until specific timing circumstances lead to an erroneous combination of memory access. In the example in (*Figure 3*), t_1 accesses the *sharedVal* using $lock_1$, and t_2 using $lock_1$ and $lock_2$. If t_1 is using $lock_1$ and concurrently t_2 is using $lock_2$, a concurrency bug will occur as the shared resources is protected but with unrelated locks. Therefore, the context of access to shared resources is the key in exposing concurrency bugs. From the branching instructions of the execution trace it is possible to extract individual execution paths. On the level of each execution path, it is possible to identify memory access instructions and function call instructions representing locking synchronization mechanisms. If an access to a memory was performed between, for example, the acquire() and release() functions, then the access is potentially synchronized. From the execution trace, it is also possible to extract function parameters; therefore, it is possible to detect specific, used lock. Our reasoning is that the analysis component is capable of and responsible for finding all concurrency bugs when the complete scope of every runnable associated with the execution trace is provided. This reasoning reduces the problem of concurrent software coverage to the sequential coverage of software units. In software coverage terms, the challenge of concurrent software coverage is reduced to the path coverage of the individual software units.

Code coverage analysis extracts function calls and branching instructions from the execution trace, deriving executed paths. Code coverage analysis must be able to derive the software structure from the software representation (e.g., code, intermediate representation, etc.) in order to model all execution paths. The complement of the *All Execution paths* set and the *Executed paths* set are *Unexecuted paths* = All Execution paths ∩ Executed paths

$$\textbf{Code coverage}(\%) = \frac{100 * \sum_{i=1}^{n} \frac{(\textbf{Number of executed paths})_i}{(\textbf{Number of all execution paths})_i}}{n}$$

- Where *n* is the number of threads and *i* is the thread ID.
- Relation $(Number\ of\ executed\ paths)_i/(Number\ of\ all\ execution\ paths)_i$ represents path code coverage per thread *i*.

For *Code coverage(%) = 100*, the analysis has enough data to fully reconstruct the context of access to memory locations.

```
// thread [1]
void t1() {
  funcOverPointer (func1);
}

// thread [2]
void t2() {
  funcOverPointer (func2);
}
```

```
void funcOverPointer ( void (* fpr) (void)) {
  (*fpr) ( ) ;
}
void func1 ( void ) {
  ...;
}
 void func2 ( void ) {
  ...;
}
```

Example 3: Undecidable problem - function call over a pointer

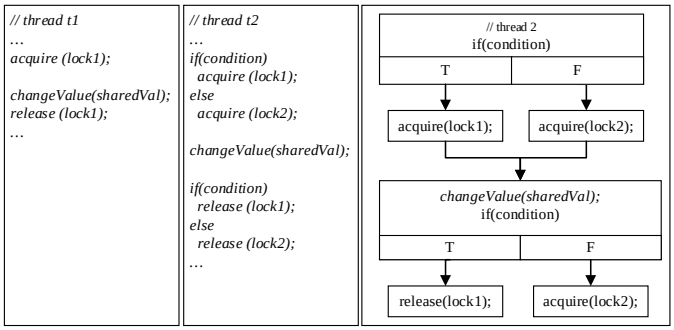


Figure 3: Context of access to the shared variable

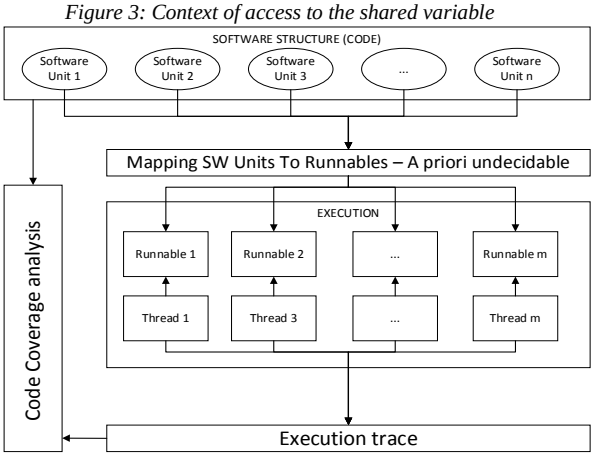


Figure 4: Mapping of SW units to runnables, code coverage using the execution trace

In summary, our conceptual solution of code coverage requires software structure which enables derivation of the execution paths, a priori known initial software unit of every thread and their complete execution traces including branching instructions and function calls.

IV. CONCURRENT SOFTWARE CODE COVERAGE USING LLVM

Our implementation of the concept solution is based on a virtual environment simulator – FERAL [22] – and an instruction set simulator – LLVM. LLVM includes compiler components (front end, optimization, back end) and components for software execution (JIT, interpreter) [21]. We have developed an interface between FERAL and LLVM that enables triggering of the desired software units. The LLVM interpreter executes the software, under the control of FERAL. Our modification of the LLVM interpreter allows us intercepting of every executed instruction. Modified LLVM interpreter is performing a task of observing the software execution, collecting the execution trace, and reporting the collected execution trace back to FERAL [20]. The analysis of the execution trace is performed by code coverage analysis component is implemented in FERAL.

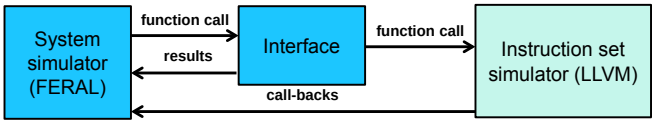


Figure 5: Concept of software virtual testing

A. *Determining a thread's scope and coverage through the incremental analysis*

In FERAL, we define testing scenarios by specifying the software units to be run, their corresponding input data sets, and the default values of shared variables. The LLVM execution environment is set up over the interface and the LLVM interpreter is triggered by the name of the software unit and the corresponding parameters. We have modified the interpreter so that it reports the execution of interesting instructions (e.g., access to memory) back to FERAL, building an execution trace during the execution. Monitoring the execution in this manner, we do not have to change the source code of the software. In order to execute software with LLVM, it is necessary to compile the source code into the LLVM Intermediate Representation (LLVM IR). The LLVM IR is an assembly-like, structured representation of a source code. Every software unit (e.g., a function) is represented as an execution flow diagram (*Figure 2.a*), composed of basic blocks. A basic block is an uninterrupted stream of instructions. If the execution enters the basic block, all its instructions are executed. Basic blocks are structures with a single entry and a maximum of two exit points as well as a unique ID number. Conditional statements (e.g., *if*) determine the number and the connections of basic blocks. A basic block with two exit points is a conditional block and reflects a condition from the source code. If there are no conditional statements in a software unit, it will be composed of only one basic block. Relations between basic blocks, defined by LLVM IR, are successor or predecessor, and True or False exit points of a basic block in case of a conditional basic block. Using the interface between LLVM and FERAL, FERAL obtains the structure of all software units present in the software, which is a first prerequisite for our conceptual solution. The scope determining of runnables/threads starts with the execution of the initial software units associated with each thread. During the execution, FERAL collects instruction traces. Branching instructions are executed on entrance to a basic block and contain the ID of the basic block that is executed. Our analysis reconstructs the execution path by correlating the basic block IDs from the branching instructions with basic block IDs from the structure of software units.

If the software is compiled with debug information, the code coverage component lists the unexecuted paths of the software unit with the associated source code line numbers. Based on this information, a tester can modify a testing scenario in FERAL and adjust the input data of the software unit to achieve better coverage. The code coverage component provides online results of the execution process in the form of the structure shown on (Figure 6). Every thread/runnable is associated with a list of executed software units, with every software unit being associated with a list of executed and unexecuted paths. The structure of every software unit is known to FERAL. The code coverage component performs an online analysis of the execution trace associated with the execution path of every thread invocation. It updates the list of the executed basic blocks of each software unit belonging to a thread. The scope of a thread is expanded based on the instructions in the execution trace if an execution path includes a new function call. For the current scope of a thread, FERAL can provide information about the path coverage, such as the percentage of executed paths, unexecuted paths, as well as the unexecuted evaluations of block conditions. For example, FERAL can provide a note to the tester: "*If the basic block [75813] is evaluated to false, then a remaining unexecuted path would be executed, leading to the higher coverage*".

B. *Identification of shared memory locations*

In the real system, threads will execute runnables concurrently. Concurrent execution of runnables ensures that they share the same memory space. If runnables are executed sequentially and independently, the analysis of concurrency bugs based on the execution trace is not possible, as every independent runnable execution will have a different address for the shared memory. Therefore, it is impossible to identify which variables are shared between which runnables using the common address because they simply have different addresses in every run. As the concurrent running of threads is not always the most practical solution, the technical challenge consists of ensuring the sequential triggering of threads in a shared memory space. In our approach, the LLVM interpreter loads the LLVM IR of the complete software, including all runnables, and stores the global, shared variables in a common address space. FERAL triggers the execution of any software unit (function) an unlimited number of times. By triggering the initial functions of threads, we simulate the execution of a

thread. Functions are executed sequentially, in the common memory space.

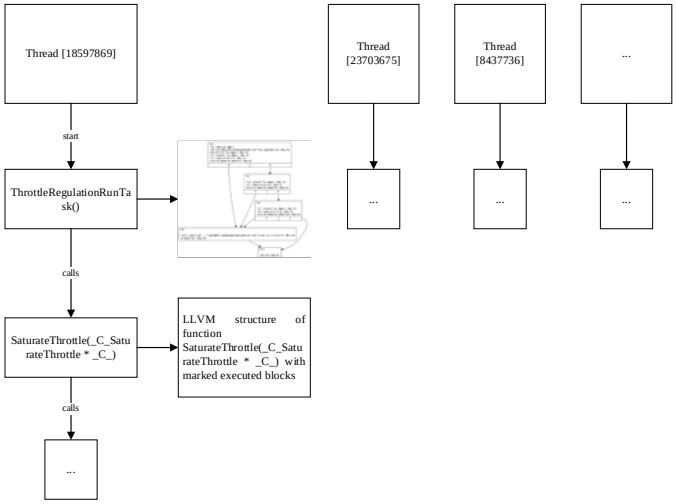


Figure 6: Incremental analysis, scope of threads

V. CONCLUSION, THREATS TO VALIDITY AND FUTURE WORK

Our approach analyzes an execution trace to quantify the coverage of the tested software and identify contexts in which a shared variable can be accessed. Any concurrency bug analysis algorithm operating on a standard execution trace can be complemented with our approach, to gain credibility of its analysis results. When achieved, our code coverage guarantees that all shared variables are accessed in all possible contexts.

However, there are certain aspects of our approach that still require improvement. Achieving path coverage is a hard task. When a code contains loops, there exist a huge number of execution paths. Also, the scope of our analysis is relative. For example, let us assume that there exist n linearly independent paths in a runnable. If n-1 paths are executed, the metrics will show that the coverage is relatively high, under the assumption that n is much greater than 1. However, if the remaining path includes calls to other software units with a large number of execution paths, the metrics are misleading. Our current approach considers testing of software that uses only locking synchronization mechanisms. If a shared variable is protected by other synchronization type, our approach will report false positive. In practice, non-locking synchronization mechanisms (e.g., ring structures) are becoming more common. In embedded systems, many developers use the guaranteed properties of the platform as a synchronization mechanism. Finally, certain paths are impossible to execute concurrently as they are mutually exclusive.

In this paper, we have demonstrated a concept solution, using existing coverage metrics in context of concurrent software. We are aware of challenges related to the path coverage and therefore we are not quantifying reduction in number of test cases. Following this concept, in the future we will experiment with other, more practical, sequential metrics, such as MC/DC to achieve the same result. We are also planning extensions w.r.t user-defined and platform-guaranteed synchronizations to reduce the number of false positives. We aim to keep our coverage metrics compatible with existing testing approaches.

REFERENCES

[1] Glenford J. Myers, “The Art of Software Testing, 2nd edition. Wiley”, 2004, ISBN 0-471-46912-2

[2] Linda Westfall, “The Certified Software Quality Engineer Handbook”, 2008, ASQ Quality Press. pp. 436–437. ISBN 978-0-87389-730-3.

[3] J.J. Chilenski, S.P. Miller, "Applicability of modified condition/decision coverage to software testing", in Software Engineering Journal, 1994

[4] Gwan-Hwan Hwang, Heng-Yi Lin, Shao-Yan Lin, Che-Sheng Lin, "Statement-Coverage Testing for Nondeterministic Concurrent Programs", in Theoretical Aspects of Software Engineering (TASE), July 2012

[5] S. R. S. Souza, P. S. L. Souza, M. A. S. Brito, A. S. Simao, E. J. Zaluska, "Empirical evaluation of a new composite approach to the coverage criteria and reachability testing of concurrent programs", February 2015

[6] Shin Hong, Jaemin Ahn, Sangmin Park, Moonzoo Kim, Mary Jean Harrold, "Testing concurrent programs to achieve high synchronization coverage", in International Symposium on Software Testing and Analysis (ISSTA), 2012

[7] Dongdong Deng, Wei Zhang, Shan Lu, "Efficient concurrency-bug detection across inputs", in OOPSLA 2013: 785-802, 2013

[8] S. Stefan , B. Michael , N. Greg , "Eraser: a dynamic data race detector for multithreaded programs", in ACM Transactions on Computer Systems (TOCS), v.15 n.4, p.391-411, Nov. 1997

[9] L. Leslie (1978). "Time, Clocks and the Ordering of Events in a Distributed System", Communications of the ACM, 21(7), 558-565.

[10] Lu, S.; Soyeon Park, Yuanyuan Zhou, "Finding Atomicity-Violation Bugs through Unserializable Interleaving Testing", 2012

[11] Chao Wang, M. Said, A. Gupta, "Coverage guided systematic concurrency testing", on 33rd International Conference Software Engineering (ICSE), 2011

[12] K. Baris, Z. Cristian, C. George, “Data races vs. data race bugs: telling the difference with portend”, 2012

[13] D. Dongdong, Z. Wei, Shan Lu, "Efficient concurrency-bug detection across inputs", 2013

[14] N. Honarmand, J. Torrellas, "Replay debugging: Leveraging record and replay for program debugging", 2014

[15] A. Miné, D. Delmas, "Towards an industrial use of sound static analysis for the verification of concurrent embedded avionics software", on International Conference on Embedded Software (EMSOFT), 2015

[16] F. Ling; T. Kitamura, "Formal Model-Based Test for AUTOSAR Multicore RTOS", on Software Testing, Verification and Validation (ICST), 2012

[17] S. Keul, "Tuning Static Data Race Analysis for Automotive Control Software", 11th IEEE International Working Conference on Source Code Analysis and Manipulation (SCAM), 2011

[18] H. Sutter, "The free lunch is over: A fundamental turn toward concurrency in software.", in Dr. Dobb’s Journal 30(3), March 2005

[19] N.G. Leveson, C.S. Turner, "An investigation of the Therac-25 accidents", 1993

[20] J. Jasmin, T. Purusothaman, Markus D., Thomas Kuhn, Peter Liggesmeyer and Christoph Grimm, "Automatic Test Coverage Measurements to support Design Space Exploration", in IFIP First International Workshop on Design Space Exploration of Cyber-Physical Systems (IDEAL), 2014

[21] C. Lattner and V. Adve, “LLVM: A compilation framework for lifelong program analysis & transformation. In Code Generation and Optimization”, 2004

[22] T. Kuhn, T. Forster, T. Braun, and R. Gotzhein. FERAL - Framework for Simulator Coupling on Requirements and Architecture Level. In Eleventh ACM-IEEE International Conference on Formal Methods and Models for Codesign, October 2013

Playing Checkers on a Donut: Visualizing the Only Vertex-Transitive Bipartite Quartic Integral Graph on 32 Vertices

Marsha Kleinbauer
TU Kaiserslautern
Kaiserslautern, Germany
Email: kleinbauer@cs.uni-kl.de

Brendan McKay
Australian National University
Canberra, Australia
Email: brendan.mckay@anu.edu.au

***Abstract*—The visualization of research outcomes can make complex content more readily accessible as well as inspire new ways of thinking about a problem. As a case in point, we present a recent result in spectral graph theory together with an intuitive visualization through which it can easily be grasped. An integral graph is a graph with only integers as eigenvalues, where the eigenvalues are calculated from the matrix representation of the graph. The spectrum of a graph is the set of distinct eigenvalues with their multiplicities. A list of candidates for the spectrum of a quartic integral graph that is both bipartite and vertex-transitive was previously determined and the question still remained: which graphs have a spectrum from this list? We are able to take the candidates from this list for graphs with 32 vertices, the smallest unknown case, and show that only one spectrum is realizable and that the graph that exists with this spectrum is unique. Despite substantial technical details required to derive the proof, we are also able to give a simple way of visualizing the unique graph that illustrates this result.**

I. Introduction

We prove, with extensive visual proof techniques, that there is a unique quartic vertex-transitive integral graph on 32 vertices.

A *quartic graph* is a graph in which every vertex is connected to exactly 4 others. The *automorphism group* $\mathrm{Aut}(G)$ of the graph G is the set of edge preserving mappings of the vertex set onto itself. A *vertex-transitive graph* is a graph for which for every pair of vertices u and v there is a mapping in the automorphism group that maps u to v. An *integral graph* is a graph with integer eigenvalues with respect to the adjacency matrix of the graph.

Previous results give a list of possible spectra for graph candidates that are quartic, vertex-transitive, and integral [1]. This list contains a unique entry for graphs with 32 vertices: $\{4, 2^{12}, 0^6, -2^{12}, -4\}$ where x^y denotes that the eigenvalue x occurs y times. Given these eigenvalues, it can be shown that the graph G we seek has no 4-cycles.

As shown in Section VII, the only vertex-transitive graph with these eigenvalues is the graph displayed in Figure 1.

This paper is organized as follows: Section II outlines the techniques that we apply to reduce our search space for finding the graph G by looking at a related set of smaller graphs called *quotient graphs*. Section III outlines our methods for determining a unique candidate out of the set of quotient graphs given some of the properties that G should exhibit. Sections IV, V, and VI investigate the path from the unique quotient graph candidate to the graph G and explain how the correct graph G is determined. A more detailed theoretical proof of our result is given in Section VII. This is followed by a section of concluding remarks.

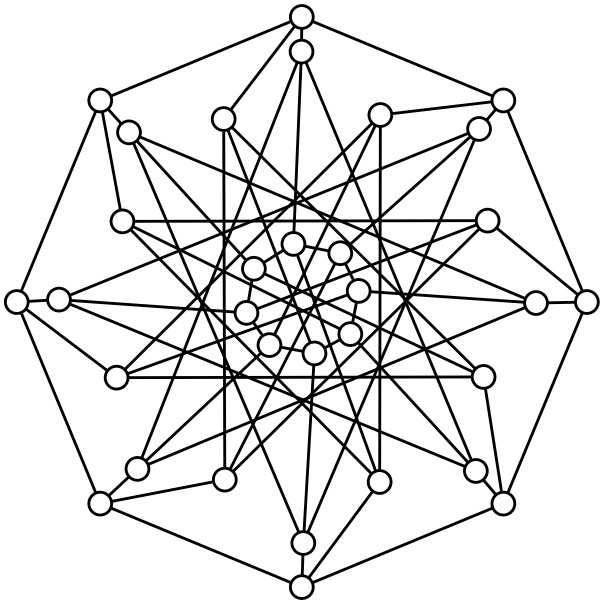


Fig. 1. The only quartic vertex-transitive integral graph on 32 vertices.

II. Building Quotient Graphs

We found a mapping σ in the automorphism group of G (please see Section VII for details) that can be used to partition the vertices into sets such that

1) there are two vertices in each set, and

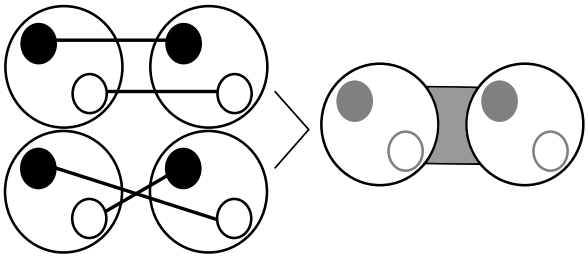


Fig. 2. The vertices of the quotient graph H consist of pairs of vertices of the original graph. Pairs of vertices from H must have exactly 0 or 2 edges from the original graph between them, corresponding to no edge or one edge in H. The edges of H can be thought of as parallel or crossed.

2) given any two sets, each vertex in the first set has the same number of edges to the second.

Let the 16 sets be the vertices of a new graph H, called *the quotient graph of* G.

Either both vertices in the same set have an edge to a different set or neither do by 2). Thus, we represent a pair of edges between sets by a single edge in H.

Using σ, step 1), and step 2), more than one quotient graph H can be built.

III. Finding the Correct Quotient Graph

Now, since the vertices of G were grouped into pairs using α, either the pair of vertices in every set has an edge between them or none of the sets contain an edge, i.e. the possibilities:

- every vertex of H is connected to 3 others, or
- the graph H is quartic.

Let *checked** denote the following process:

- Generate all graphs that are satisfying the properties of a candidate graph H so far, using the tool genbg of the program nauty [2].
- Test the eigenvalues of each graph for integrality.

Since G was vertex-transitive and integral, H must also be vertex-transitive and integral.

The 16 vertex graphs with property 1) were checked* and none were integral.

The only checked* graph that satisfied property 2) and was integral has the form given in Figure 3.

IV. From Quotient Graph to Quartic Graph

We find the possibilities for going from the only lasting candidate H (in Figure 3) to G.

- Replace each edge in H by a pair of edges.
- Configure each pair as *parallel* or *crossed* (see Figure 2).
- Keep the graph you find as a candidate.

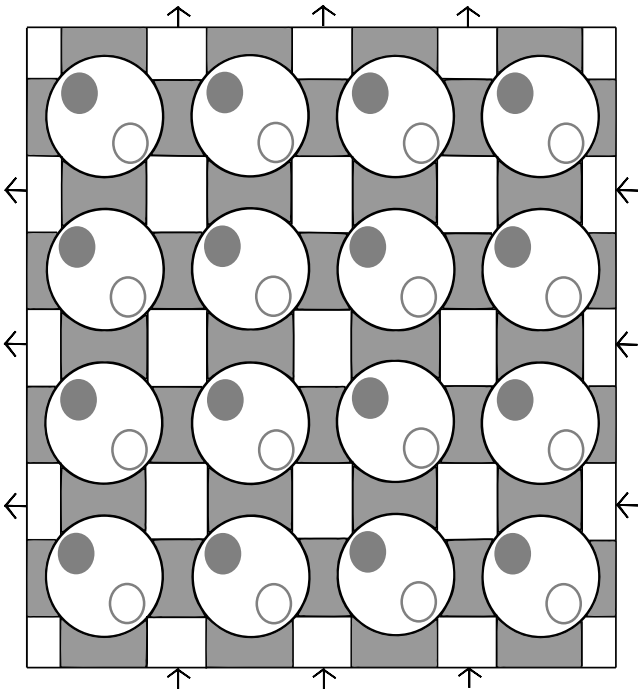


Fig. 3. The only quotient graph H that satisfied all the desired properties described in the first three sections. The arrows on the top are textitattached to the arrows on the bottom joining all corresponding edges of H. Similarly the left arrows are textitattached to the arrows on the right.

Fig. 4. By replacing every edge of Figure 3 by a parallel edge, the graph G is equivalent to a checkerboard on a donut with squares the vertex set and with edges connecting squares of the same color that meet at a corner.

V. What if all the edges are parallel?

This would be the moves that a single checker's piece could get to from some square if your 4×4 checkerboard was shaped like a donut.

This is not a good choice for our graph G:

- In a regular move in checkers, you can only get from the starting dark square to all the other dark squares. This means that the graph G would consist of two disconnected components: light squares/vertices and dark squares/vertices. The graph G is connected so this is not a good candidate for G.

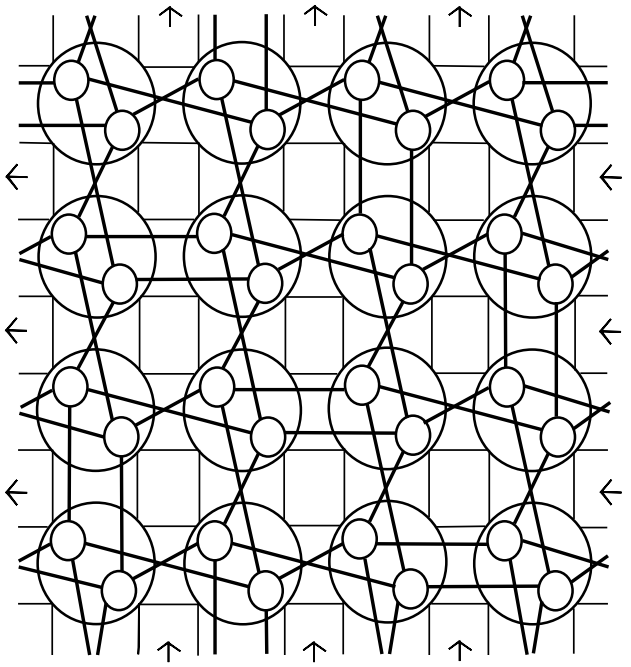


Fig. 5. The unique way (up to isomorphism) to configure the edges of H as parallel or crossed such that the resulting graph G does not have any 4-cycles is shown here.

- The second problem is that you can move in a 4-cycle and this is contrary to the information given by the eigenvalues (as discussed in the introduction). This is another reason why all parallel edges does not produce a good candidate for G.

VI. The Correct Quartic Graph

G cannot contain 4-cycles, thus every 4-cycle in H must have an odd number of parallel edges. Therefore, Figure 5 shows the only way to configure the edges appropriately.

This graph can be redrawn to give the graph in Figure 1.

VII. The Detailed Proof

The basis of our visualization is a result by Minchenko [3], which we reiterate here. While this proof requires substantial minutiae, the actual results can be made understandable by visualization as was demonstsrated by the previous sections.

As above, we consider a vertex-transitive graph G with 32 vertices, girth 6, and spectrum $\{4, 2^{12}, 0^6, -2^{12}, -4\}$. Since G is vertex-transitive, the orbit-stabilizer theorem (see [4] for example) gives us that $|G| = 2^5$ divides $|\operatorname{Aut}(G)|$.

A *p-subgroup* P of a group Γ is a subgroup of Γ where every element $h \in P$ has order a power of the prime p. A *Sylow p-subgroup* P of a group Γ is a p-subgroup that is contained in no larger p-subgroup of Γ. Let A be a Sylow 2-subgroup of $\operatorname{Aut}(G)$ [5]. The *center* $Z(\Gamma)$ is defined as $Z(\Gamma) = \{z \in \Gamma \mid z\alpha = \alpha z \text{ for all } \alpha \in \Gamma\}$. If $Z(\Gamma) = \{e\}$ where e is the identity element of Γ then we say that the center of Γ is *trivial*.

The following theorem appears in [6, Thm 3.4].

Theorem 1. *Let p be a prime number and let A be a Sylow p-subgroup of $\operatorname{Aut}(G)$. Suppose p^s divides $|\operatorname{Aut}(G)v|$ for a vertex $v \in V(G)$. Then p^s also divides $|Av|$.*

Now since G is transitive, $|\operatorname{Aut}(G)v| = 2^5$ for all $v \in V(G)$. By Theorem 1, 2^5 divides $|Av|$ for all $v \in V(G)$ and thus $Av = V(G)$. Therefore, A is transitive on the vertices of G.

Since A is finite and non-trivial, it has a non-trivial center [5]. Let σ be a central element in A of order 2. Since σ is a central element, the group $\langle\sigma\rangle$ is a normal subgroup of A.

Given a graph G_1, a *partition* Π of $V(G_1)$ is a set of disjoint non-empty subsets of $V(G_1)$ whose union is $V(G_1)$. We will refer to these subsets as cells. A partition, $\Pi = (C_1, \ldots, C_k)$ is *equitable* if for every choice of i and j, each vertex in C_i has the same number of neighbours in C_j. Given an equitable partition Π of a graph G_1, the *quotient graph* G_1/Π of G_1 with respect to Π is the graph with the cells of Π as its vertices and with edge (C_i, C_j) for every edge $(x, y) \in E(G_1)$ where $x \in C_i$ and $y \in C_j$. Thus, a quotient graph may have multiple edges and loops.

The partition, $\Pi = (C_1, \ldots, C_k)$, of $V(G)$ into orbits of σ is an equitable partition of the graph G [4]. Since the group $\langle\sigma\rangle$ is normal, each cell contains two vertices [6, Prop 7.1]. Now since A is transitive on $V(G)$, A acts transitively on these cells [7, p. 173]. Thus, the quotient graph $H = G/\Pi$ of G with respect to Π is a transitive multigraph with 16 vertices.

Every eigenvalue of G/Π is an eigenvalue of the graph G [8]. For this reason, we only consider graphs with integer eigenvalues. For any integral graph that satisfies the known conditions of G/Π, we check that it lifts to G.

The pair of vertices in one cell is mapped to the pair of vertices in another cell and so these pairs of edges in G/Π can be thought of as a single edge in the 'frame' of G/Π. Each pair of vertices in a cell is either adjacent or non-adjacent. This gives two cases:

- The frame of G/Π is a cubic transitive graph. It can be checked that there is no such cubic graph that is integral.
- The frame of G/Π is a quartic transitive graph. It can be checked that a unique quartic graph candidate exists that satisfies integrality: $C_4 \times C_4$ (where $C_4 \times C_4$ is the direct product of the cyclic group on 4 elements with itself).

In lifting from $C_4 \times C_4$ to G each edge between cells is replaced by two edges. These edges can be thought of as 'parallel' or 'crossed', depending on whether they join vertices in the same copy of $C_4 \times C_4$ or not. Since G has no 4-cycles, each cycle must have an odd number of 'parallel' edges to be a suitable candidate for G. Using this observation, it can be checked by exhaustion that there is a unique way to lift G/Π

to G. This unique completion is isomorphic to the graph in Figure 1.

Let $G_1 \times G_2$ denote the graph with vertex set the Cartesian product $V(G_1) \times V(G_2)$ with (x, a) adjacent to (y, b) if and only if x is adjacent to y in G_1 and a is adjacent to b in G_2. Let K_2 denote the graph that consists of two vertices and the single edge between them. The following result can be found in a paper by Stevanović [9]:

Theorem 2. *If G_1 is a quartic integral graph, then the bipartite graph $G_1 \times K_2$ is a bipartite quartic integral graph.*

There is no bipartite quartic integral graphs on 64 vertices. Therefore, by Theorem 2, there are no non-bipartite quartic integral graphs with 32 vertices. Thus, this G is the unique quartic integral graph on 32 vertices that is vertex-transitive.

VIII. Conclusions

Even if it were possible to play checkers on a donut, the unique quartic vertex-transitive integral **graph on** 32 **vertices would be merely related to the checkerboard** but not the moves that can be taken in the game.

There are still many more cases where we have a set of eigenvalues but we do not know whether a unique graph, many graphs, or no graphs exist for the set.

Some of the small unsolved cases include

- $40 = 5 * 8$ vertices and
- $42 = 7 * 6$ vertices.

Using the techniques presented here, we can partition the vertices into

- 8 sets of size 5 and
- 6 sets of size 7 respectively.

We checked* the graphs that satisfy these properties as before. We find 5 and 4 respective candidates for the quotient graph. Unfortunately, having the possibility of 4-cycles in these two cases has made things difficult enough that we have not yet come up with the solution set of graphs.

The graphical presentation of this result brings insight and understanding of a mathematical finding and technique to a wider audience. Additionally, the working ideas and methods of the authors are made readily available. Drawing on familiar metaphors can help build a mental model of the problem before requiring exposure to the technical more abstract mathematical notions.

References

[1] M. Minchenko and I. M. Wanless, "Quartic integral Cayley graphs," *Ars Math. Contemp.*, vol. 8, pp. 381–408, 2015.

[2] B. D. McKay and A. Piperno, "Practical graph isomorphism, II," *Journal of Symbolic Computation*, vol. 60, no. 0, pp. 94 – 112, 2014.

[3] M. E. Minchenko, "Counting subgraphs of regular graphs using spectral moments," Ph.D. dissertation, Monash University, 2014.

[4] C. D. Godsil and G. Royle, *Algebraic Graph Theory*. New York: Springer-Verlag, 2001.

[5] J. B. Fraleigh, *A first course in abstract algebra*, 7th ed. Pearson, 2003.

[6] H. Wielandt, *Finite Permutation Groups*. New York: Academic Press, 1964.

[7] N. Biggs, *Algebraic Graph Theory*. Cambridge: Cambridge University Press, 1974.

[8] W. H. Haemers, "Regularity and the spectra of graphs," in *Surveys in Combinatorics 2009*, ser. London Mathematical Society Lecture Notes, S. Huczynska, J. D. Mitchell, and C. V. Roney-Dougal, Eds. Cambridge University Press, 2009, vol. 365, pp. 75 – 90.

[9] D. Stevanović, "4-regular integral graphs avoiding ± 3 in the spectrum," *Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat.*, pp. 99–110, 2003.

Constant-Input Observability of DAEs with Application to Power Networks

Ferdinand Küsters
System Analysis, Prognosis and Control
Fraunhofer ITWM
Germany

Andreas Wirsen
System Analysis, Prognosis and Control
Fraunhofer ITWM
Germany

***Abstract*—We consider an observability notion suitable for a linearized power network model. *Constant-input observability* is a variant of observability for constant, but unknown inputs. This property is characterized for linear DAEs and then applied to a power network model. It is shown that constant-input observability of this system depends solely on the network topology and can be expressed by a simple matrix criterion.**

I. Introduction

The recent interest in the analysis of power networks is caused by the integration of renewable energy sources. Renewable energy sources increase the network's complexity and make control of the system much more involved. Former power networks had a top-to-bottom structure with conventional energy sources only on the top level. This does not hold true any more as renewable energy sources are integrated at every level.

For stabilization and control it it necessary to analyze sophisticated models of power grids. An important question for model-based control is state-estimation, i.e. the estimation of the system's state by measurements. A property required for this is observability, i.e. the theoretical ability to recover the state from input and output (for the undisturbed system).

This work deals with the observability of a power network model. The model considered here is the classical swing equation [1], [10]. It describes a transmission grid, i.e. the highest level of a power grid.

State estimation of power grids is classically done using static equations, see [11]. We chose a different approach and consider the observability of a dynamical system.

As the classical observability concept is not suitable for the model, we introduce the notion of constant-input observability. For ODEs such an approach is given in [8] and more generally in [7].

In Section II the swing equation will be discussed. Section III deals with observability. Constant-input observability will be introduced and characterized. The new notion will then be applied to the power network model in Section IV. Some conclusions will be drawn and several examples will be given to illustrate the result.

The following notion will be needed to state the main result: For a matrix $M \in \mathbb{R}^{n\times n}$ and two index sets $V, W \subseteq \{1,\dots,n\}$ denote by $M_{V,W}$ the $|V| \times |W|$ submatrix consisting of the rows corresponding to V and the columns corresponding to W. Let $M_{V,\odot}$ ($M_{\odot,W}$) describe the matrix for which only the number of rows (columns) is reduced.

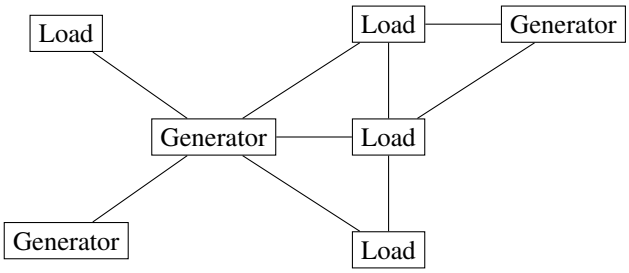


Fig. 1. Example of a power grid

II. Power network model

We consider a classical model called *swing equation*. It describes a transmission grid and consists of synchronous generators, loads and transmission lines (see Figure 1). The generators are described by ODEs, the load nodes impose algebraic constraints. Hence the overall system is a differential-algebraic equation (DAE).

A. Swing equation

A synchronous generator i is described by the ODE

$$M_i\ddot{\theta}_i(t) + D_i\dot{\theta}_i(t) = P_i(t) - f_i(\theta(t)),$$

where

- $\theta_i(\cdot)$ is the voltage phase angle at the generator node,
- M_i is the moment of inertia,
- D_i is the damping constant,
- $P_i(\cdot)$ is the (external) power infeed or extraction, called *load*
- $f_i(\theta)$ is the power coming from or flowing into the network. It depends on the voltage angles θ.

A load node i is modeled by the algebraic equation

$$0 = P_i(t) - f_i(\theta(t)).$$

Additionally, we allow for a second kind of load node which we call *dynamic load node* in contrast to *constant load node*

for the former. The dynamic load node is governed by the equation

$$D_i\dot{\theta}_i(t) = P_i(t) - f_i(\theta(t)).$$

This models a frequency-dependent load and occurs e.g. in [4].

The power flow $f_i(\theta)$ is given by the *load flow equation*:

$$f_i(\theta(t)) = \sum_{j\neq i} Y_{i,j}\sin\left(\theta_i(t) - \theta_j(t)\right).$$

$Y_{i,j}$ is the admittance between node i and node j. It is nonzero whenever there is a direct connection between node i and node j. The load flow equations can be obtained from the power flow equations by a common, yet restrictive decoupling assumption. This is reasonable for lossless transmission lines and thus the model is only useful for the transmission grid.

The overall model with n_g generators, n_{dl} dynamic loads and n_{cl} constant loads (with $n = n_g + n_{dl} + n_{cl}$) is then given by

$$M_{i_g}\ddot{\theta}_{i_g} + D_{i_g}\dot{\theta}_{i_g} = P_{i_g} - \sum_{j\neq i_g} Y_{i_g,j}\sin\left(\theta_{i_g} - \theta_j\right), \tag{1a}$$

$$i_g = 1,\dots,n_g, \tag{1b}$$

$$D_{i_{dl}}\dot{\theta}_{i_{dl}} = P_{i_{dl}} - \sum_{j\neq i_{dl}} Y_{i_{dl},j}\sin\left(\theta_{i_{dl}} - \theta_j\right), \tag{1c}$$

$$i_{dl} = n_g + 1,\dots,n_g + n_{dl}, \tag{1d}$$

$$0 = P_{i_{cl}} - \sum_{j\neq i_{cl}} Y_{i_{cl},j}\sin\left(\theta_{i_{cl}} - \theta_j\right), \tag{1e}$$

$$i_{cl} = n_g + n_{dl} + 1,\dots,n, \tag{1f}$$

where we omitted the time-dependency.

B. Linearization

The model (1) is a nonlinear DAE. To analyze its observability properties we will linearize it.

Define θ_g, θ_{dl} and θ_{cl} to be the vectors of voltage angles corresponding to generator nodes, dynamic load nodes and constant load nodes, respectively. Analogously, P_g, P_{dl} and P_{cl} are defined. Let $\omega = \dot{\theta}_g$ be the voltage angular velocity at the generators. By M_g, D_g and D_{dl} denote the diagonal matrices with the corresponding parameters. Let L be defined by

$$L_{i,j} = \begin{cases} -Y_{i,j}, & i \neq j, \\ \sum_{k\neq i} Y_{i,k}, & i = j. \end{cases}$$

L is the weighted laplacian of the underlying graph. Partitioning for the different node types yields

$$L = \begin{bmatrix} L_{g,g} & L_{g,dl} & L_{g,cl} \\ L_{dl,g} & L_{dl,dl} & L_{dl,cl} \\ L_{cl,g} & L_{cl,dl} & L_{cl,cl} \end{bmatrix}.$$

Linearization of (1) around an equilibrium for $P = 0$ gives

$$\mathfrak{E}\dot{x}(t) = \mathfrak{A}x(t) + \mathfrak{B}u(t) \tag{2}$$

with

$$\mathfrak{E} := \begin{bmatrix} 0 & I & 0 & 0 \\ M_g & 0 & 0 & 0 \\ 0 & 0 & D_{dl} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathfrak{A} := \begin{bmatrix} I & 0 & 0 & 0 \\ -D_g & -L_{g,g} & -L_{g,dl} & -L_{g,cl} \\ 0 & -L_{dl,g} & -L_{dl,dl} & -L_{dl,cl} \\ 0 & -L_{cl,g} & -L_{cl,dl} & -L_{cl,cl} \end{bmatrix},$$

$$\mathfrak{B} := \begin{bmatrix} 0 & 0 & 0 \\ I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}$$

and

$$x = \begin{pmatrix} \omega \\ \theta_g \\ \theta_{dl} \\ \theta_{cl} \end{pmatrix}, \quad u = \begin{pmatrix} P_g \\ P_{dl} \\ P_{cl} \end{pmatrix}.$$

Linearization around an equilibrium of (1) for $P \neq 0$ also gives a system of the same structure, but with a slightly different meaning of L and u. However, L would still be a laplacian.

Solvability of a linearized power DAE similar to (2) has been studied by [5]. Analogous to [5] one can show that (2) is regular and of index one if and only if there is no connected component consisting only of constant load nodes. Equivalently:

$$\text{(2) is regular} \quad \Leftrightarrow \quad \det\left(L_{cl,cl}\right) \neq 0.$$

Remark 1 (Regularity of linear DAEs, [12]). A linear DAE $E\dot{x}(t) = Ax(t) + f(t)$ is called *regular* if and only if $\det(sE - A) \in \mathbb{R}[s] \setminus \{0\}$, i.e. $\det(sE - A)$ is not the zero polynomial. Regularity is equivalent to existence and uniqueness of solutions (for arbitrary inhomogenity $f(\cdot)$ and suitable initial values $x(0)$).

Remark 2 (Consensus). The linear DAE (2) includes also the consensus problem of static networks, see [9]. If we assume each node to be a dynamic load node with damping constant $d_i = 1$ we arrive at the system

$$\dot{x}(t) = -Lx(t) + P(t).$$

This is the consensus equation (with a driving force for each node). The forthcoming characterization of constant-input observability is thus not only interesting for power networks but possibly also for the agreement problem of multi-agent networks.

III. Observability

Observability is a classical property considered in system theory, which basically says that the system's state can be reconstructed from input and output. As stated in the introduction, observability is relevant for the estimation of the system's state. For an observable system one can construct an observer, for example the Luenberger-Observer. The following ideas can easily be adapted for determinability.

A. Classical observability

Linear systems are of the form

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t) \tag{3}$$

with input $u(\cdot)$, state $x(\cdot)$ and output $y(\cdot)$. A triple (u, x, y) is called solution of (3) if it solves the equation for some initial value $x(0) = x_0$. The corresponding DAE problem is

$$E\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t). \tag{4}$$

There are several different observability concepts for DAEs, see the survey article [3]. The one used here is based on the behavioral description and equivalent to observability of the underlying ODE.

Definition 1. The regular DAE (4) is called *observable* if and only if it holds for any two solutions (u, x, y), $(\tilde{u}, \tilde{x}, \tilde{y})$:

$$(u \equiv \tilde{u} \ \wedge \ y \equiv \tilde{y}) \quad \Rightarrow \quad x \equiv \tilde{x}.$$

Observability can be characterized by the Hautus-Lemma:

Lemma 1 ([13, Theorem 7])**.** *The regular DAE* (4) *is observable if and only if it holds*

$$\ker \begin{bmatrix} sE - A \\ C \end{bmatrix} = \{0\} \quad \forall s \in \mathbb{C}.$$

B. Constant-input observability

For classical observability the input is assumed to be known. The inputs of the power network DAE model (2) are the loads at the different nodes. As it is not reasonable to assume all the loads to be known, we need a stronger observability notion taking this into account.

Such a concept is *strong observability*, see [6]. Strong observability means that the state can be reconstructed from the output for any unknown input. An ODE is strongly observable if and only if it holds

$$\operatorname{rank} \begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} = n + \operatorname{rank} \begin{bmatrix} B \\ D \end{bmatrix} \quad \forall s \in \mathbb{C}.$$

If we want to recover also the input u we have to require that the matrix $\left[\begin{smallmatrix} B \\ D \end{smallmatrix}\right]$ has full rank.

[8] and [7] considered the unknown input as governed by a dynamical system (with known dynamics). Assuming the input to be constant is a special case of this approach that has gained some attention in [8]. We will give a statement analogous to [14, Theorem 2] for DAEs. Assuming the input to be constant is reasonable if it changes slowly compared to the dynamics of the system.

If the input is assumed to be constant, it can be interpreted as an additional state governed by the equation $\dot{u} = 0$. This gives the *augmented system*

$$\begin{aligned} \begin{bmatrix} E & 0 \\ 0 & I \end{bmatrix} \frac{d}{dt} \begin{pmatrix} x(t) \\ u(t) \end{pmatrix} &= \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} \begin{pmatrix} x(t) \\ u(t) \end{pmatrix}, \\ y(t) &= \begin{bmatrix} C & D \end{bmatrix} \begin{pmatrix} x(t) \\ u(t) \end{pmatrix}. \end{aligned} \tag{5}$$

It can be easily seen that (5) is regular if and only if (4) is regular and that both systems have the same index.

Definition 2. The system (4) is called *constant-input observable* if and only if the augmented system (5) is observable.

Using Lemma 1, i.e. the Hautus-Test, constant-input-observability can be characterized as follows:

Theorem 1. *A regular DAE* (4) *is constant-input observable if and only if it is observable and*

$$\ker \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \{0\}.$$

In particular, the number of outputs has to be at least as large as the number of inputs. Constant-input observability is indeed weaker than strong observability (with $\left[\begin{smallmatrix} B \\ D \end{smallmatrix}\right]$ having full rank), as the following example illustrates:

Example 1. The ODE (3) with

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \ B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \ D = 0$$

is observable and constant-input observable, but not strongly observable. Indeed, $x(0) = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and $u(t) = \mathrm{e}^t$ give $y = 0$.

Remark 3 (Observer). If the system is constant-input observable, one can use a Luenberger observer for the augmented system to estimate the system's state and input. This observer works sufficiently well for slowly time-varying input. Under a mild additional assumption, this works not only for ODE systems but also for DAE systems. (For observers of DAEs we refer to [2].)

IV. Application to the power network model

In the remainder we will characterize constant-input observability for the power network (2).

We assume that for certain set $S \subseteq \{1, \dots, n\}$ of nodes both the inputs P_i and the states θ_i can be measured while for the remaining nodes $U := \{1, \dots, n\} \setminus S$ no information is available. This means the output of (2) is of the form

$$y = \begin{bmatrix} 0 & I_{S,\odot} \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ I_{S,\odot} \end{bmatrix} u. \tag{6}$$

Classical observability of a power network DAE depends on the parameters of the individual nodes and the node types as the following example shows.

Example 2. Consider the power DAE network in Figure 2 with one generator, two dynamic loads and one constant load.

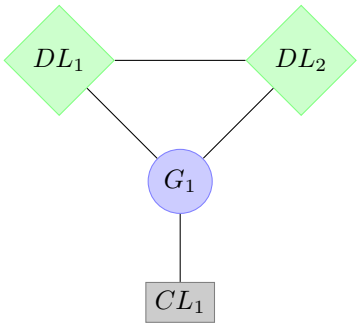


Fig. 2. Network of Example 2

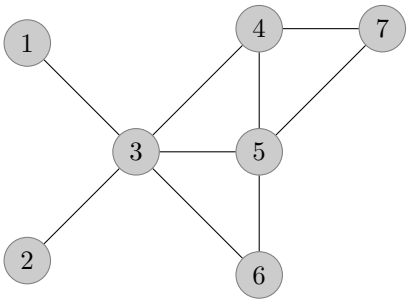


Fig. 3. Graph of the network used in Example 3.

We assume that all transmission lines have the same parameter and $m_g = 1$, $d_g = 1$. The system is then given by

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & d_1 & 0 & 0 \\ 0 & 0 & 0 & d_2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \frac{d}{dt} \begin{pmatrix} \omega \\ \theta_g \\ \theta_{dl1} \\ \theta_{dl2} \\ \theta_{cl} \end{pmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & -3 & 1 & 1 & 1 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{pmatrix} \omega \\ \theta_g \\ \theta_{dl1} \\ \theta_{dl2} \\ \theta_{cl} \end{pmatrix} + \begin{pmatrix} 0 \\ P_g \\ P_{dl1} \\ P_{dl2} \\ P_{cl} \end{pmatrix},$$

$$y = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} \omega \\ \theta_g \\ \theta_{dl1} \\ \theta_{dl2} \\ \theta_{cl} \end{pmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} P_g \\ P_{dl1} \\ P_{dl2} \\ P_{cl} \end{pmatrix}.$$

If $d_1 = d_2$, the system is not observable. However, for $d_1 \neq d_2$ the system is observable. Also, if the node DL_2 is changed to a constant load node, the system becomes observable.

A. Characterization

In contrast to classical observability, constant-input observability of the power network DAE depends solely on the network topology:

Theorem 2. *Constant-input observability of the power network* (2)*,* (6) *is equivalent to*

$$\ker L_{S,U} = \{0\}. \tag{7}$$

As a direct consequence of the theorem, the following are necessary conditions for constant-input observability of (2):

- $|S| \geq \frac{n}{2}$, i.e. at least every second node has to be measured;
- any node in U has to be connected directly with a node in S. (Otherwise $L_{S,U}$ would have a zero column.)

On the other hand, the following does not influence constant-input observability:

- the node types (except for the regularity of the system);
- the node parameters such as moment of inertia and damping;
- the connections between nodes in S and the connections between nodes in U.

The proof of Theorem 2 (given in the appendix) shows that the condition for constant-input determinability is the same as for constant-input observability. More interesting, the seemingly stronger notion of strong observability is also satisfied for the power network model if and only if (7) holds.

Remark 4 (Line switching). It might be of interest not only to check that the system is constant-input observable, but also that this property is preserved if a transmission line is switched of. (Compare the concept of "$N-1$ stability".) Theorem 2 gives a simple criterion to check for each network configuration. In particular, a change of a line that does not connect a node in S with a node in U does not affect constant-input observability.

B. Examples

The following examples show that (7) is a rather restrictive assumption.

Example 3. The grid described in Figure 1 gives a regular DAE (2). The corresponding graph is given in Figure 3. The network has 7 nodes, hence at least 4 sensor nodes are necessary. $S = \{2, 3, 5, 7\}$ gives

$$L_{S,U} = \begin{bmatrix} 0 & 0 & 0 \\ L_{3,1} & L_{3,4} & L_{3,6} \\ 0 & L_{5,4} & L_{5,6} \\ 0 & 0 & L_{7,6} \end{bmatrix},$$

hence the system is for any edge-weights constant-input observable.

Next we give two worst case examples:

Example 4 (Star graph)**.** Consider a "star" with n nodes: a central node and $n-1$ nodes that are only connected with this central node. S has to contain $n-1$ nodes, i.e. the maximal number. The particular choice of the sensor nodes is not important.

Example 5 (Complete graph)**.** The other extreme is a system where all nodes are connected (and all edges have the same weight), i.e.

$$L = nI - \mathbb{1},$$

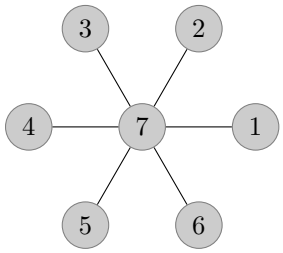


Fig. 4. Graph of the network used in Example 4

s

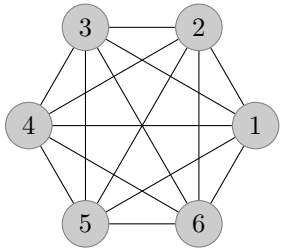


Fig. 5. Graph of the network used in Example 5

where $\mathbb{1}$ denotes a matrix having ones in each entry. Again, the maximal number of nodes $(n-1)$ is required. The particular choice of sensor nodes is not important. In contrast to the last example, the weights of the edges are crucial here.

C. Sensor placement

A question closely related to state estimation is that of sensor placement: How many sensors are needed and where shall they be placed to get measurements that allow to estimate the state.

We are interested in finding the minimal number and a corresponding placement of sensors to make the system (2) constant-input observable. Theorem 2 showed that at least every second node has to be measured and any node has to be directly connected to a measured node.

A very simple, yet not optimal approach to this problem is to choose the sensor nodes iteratively and add them to an (insufficient) sensor node set until constant-input observability is achieved. Hence successively choose $\tilde{s} \in U$ with

$$\operatorname{rank} L_{S,U} > \operatorname{rank} L_{S\cup\{\tilde{s}\},U\setminus\{\tilde{s}\}}. \tag{8}$$

The node $\tilde{s}$ is then added to the sensor nodes, i.e.

$$S \leftarrow S \cup \{\tilde{s}\}, \quad U \leftarrow U \setminus \{\tilde{s}\}.$$

The algorithm stops with a sensor set S that gives $\operatorname{rank} L_{S,U} = |U|$, i.e. with a sensor set that makes the system constant-input observable.

To improve this algorithm, note that any $\tilde{s} \in U$ with $\operatorname{rank} L_{S,U} = \operatorname{rank} L_{S\cup\{\tilde{s}\},U\setminus\{\tilde{s}\}}$ does not satisfy (8) for any $\hat{S} \supset S$, $\hat{U} = \hat{S}^c$ either. Hence one can store the information about all nodes that do not reduce the unobservable subspace in a set IM ("indirectly measured" nodes).

Data: S_0
Result: S
$S \leftarrow S_0$, $U \leftarrow S_0^c$, $IM \leftarrow \emptyset$;
while $\ker L_{S,U} \neq \{0\}$ **do**
  Choose $\tilde{s} \in U$;
  if $\operatorname{rank} L_{S,U} = \operatorname{rank} L_{S\cup\{\tilde{s}\},U\setminus\{\tilde{s}\}}$ **then**
    $IM \leftarrow IM \cup \{\tilde{s}\}$, $U \leftarrow U \setminus \{\tilde{s}\}$;
  else
    $S \leftarrow S \cup \{\tilde{s}\}$, $U \leftarrow U \setminus \{\tilde{s}\}$;
  end
end

Algorithm 1: Sensor placement.

V. Conclusion

Constant-input observability has been introduced and characterized for linear DAEs. It has been shown that for the power network DAE this observability notion breaks down to a simple algebraic criterion. Several small examples have been considered.

Appendix

Proof of Theorem 2:

First of all we show that $\ker L_{S,U} = \{0\}$ is equivalent to $\ker \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] = \{0\}$ for (2), (6):

$$\begin{aligned} & \ker \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \{0\} \\ \Leftrightarrow \quad & \ker \begin{bmatrix} L & I \\ I_{S,\odot} & 0 \\ 0 & I_{S,\odot} \end{bmatrix} = \{0\} \\ \Leftrightarrow \quad & \ker \begin{bmatrix} L_{\odot,U} & I_{\odot,U} \end{bmatrix} = \{0\} \\ \Leftrightarrow \quad & \ker L_{S,U} = \{0\}. \end{aligned}$$

If we can show that $\ker L_{S,U} = \{0\}$ implies classical observability of (2), (6), the proof is finished. Hence it remains to show that $\ker L_{S,U} = \{0\}$ implies $\ker \left[\begin{smallmatrix} sE-A \\ C \end{smallmatrix}\right] = \{0\}$ for all $s \neq 0$. Let $s \neq 0$ and denote

$$E = \left[\begin{array}{c|ccc} 0 & I & 0 & 0 \\ \hline M_g & 0 & 0 & 0 \\ 0 & 0 & D_{dl} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} 0 & E_1 \\ \hline E_2 & E_3 \end{array}\right],$$

$$A = \left[\begin{array}{c|c} I & 0 \\ \hline -D_g & \\ 0 & -L \\ 0 & \end{array}\right] = \left[\begin{array}{c|c} I & 0 \\ \hline A_2 & -L \end{array}\right].$$

This gives:

$$\begin{aligned} & \ker \begin{bmatrix} sE-A \\ C \end{bmatrix} = \{0\} \\ \Leftrightarrow \quad & \ker \begin{bmatrix} -I & sE_1 \\ sE_2 - A_2 & sE_3 + L \\ 0 & I_{S,\odot} \end{bmatrix} = \{0\} \\ \Leftrightarrow \quad & \ker \begin{bmatrix} -I & (sE_1)_{\odot,U} \\ sE_2 - A_2 & (sE_3 + L)_{\odot,U} \end{bmatrix} = \{0\}. \end{aligned}$$

The number of rows of E_1 corresponds to the number of generators. By $\hat{S}$ and $\hat{U}$ denote the subsets fo S and U that correspond to generator nodes. Then we get by reordering of rows and columns that $\left[\begin{smallmatrix} sE-A \\ C \end{smallmatrix}\right]$ has a trivial kernel if and only if the kernel of

$$\begin{bmatrix} -I & 0 & (sE_1)_{\hat{S},U} \\ 0 & -I & (sE_1)_{\hat{U},U} \\ (sE_2-A_2)_{S,\hat{S}} & (sE_2-A_2)_{S,\hat{U}} & (sE_3+L)_{S,U} \\ (sE_2-A_2)_{U,\hat{S}} & (sE_2-A_2)_{U,\hat{U}} & (sE_3+L)_{U,U} \end{bmatrix} \tag{9}$$

is trivial. For the next step, note that E_1, E_2, E_3 and A_2 are block matrices with entries only on (block) diagonals. Hence a restriction to the rows corresponding to $\hat{U}$ and columns corresponding to S gives the zero matrix for E_1. Thus the matrix (9) is equal to

$$\begin{bmatrix} -I & 0 & 0 \\ 0 & -I & (sE_1)_{\hat{U},U} \\ (sE_2-A_2)_{S,\hat{S}} & 0 & L_{S,U} \\ 0 & (sE_2-A_2)_{U,\hat{U}} & (sE_3+L)_{U,U} \end{bmatrix}.$$

If $L_{S,U}$ has a trivial kernel, so does the matrix written above. ■

REFERENCES

[1] A. Arapostathis, S.S. Sastry, and P. Varaiya. Global analysis of swing dynamics. *Circuits and Systems, IEEE Transactions on*, 29(10):673–679, Oct 1982.

[2] Thomas Berger and Timo Reis. Observers and dynamic controllers for linear differential-algebraic systems. preprint.

[3] Thomas Berger, Timo Reis, and Stephan Trenn. Observability of linear differential-algebraic systems. In Achim Ilchmann and Timo Reis, editors, *Surveys in Differential-Algebraic Equations III*, Differential-Algebraic Equations Forum. Springer-Verlag, Berlin-Heidelberg, 2015. submitted for publication.

[4] F. Dorfler. *Dynamics and Control in Power Grids and Complex Oscillator Networks*. PhD thesis, Mechanical Engineering Department, University of California at Santa Barbara, September 2013.

[5] Tjorben B Gross, Stephan Trenn, and Andreas Wirsen. Topological solvability and index characterizations for a common DAE power system model. In *Control Applications (CCA), 2014 IEEE Conference on*, pages 9–14. IEEE, 2014.

[6] Malo L. J. Hautus. Strong detectability and strong observers. *Lin. Alg. Appl.*, 50:353–368, 1983.

[7] CD Johnson. On observers for systems with unknown and inaccessible inputs. *International Journal of Control*, 21(5):825–831, 1975.

[8] JS Meditch and GH Hostetter. Observers for systems with unknown and inaccessible inputs. *International Journal of Control*, 19(3):473–480, 1974.

[9] Mehran Mesbahi and Magnus Egerstedt. *Graph theoretic methods in multiagent networks*. Princeton University Press, 2010.

[10] S. Sastry and P. Varaiya. Hierarchical stability and alert state steering control of interconnected power systems. *Circuits and Systems, IEEE Transactions on*, 27(11):1102–1112, Nov 1980.

[11] F.C. Schweppe and J. Wildes. Power system static-state estimation, part i: Exact model. *Power Apparatus and Systems, IEEE Transactions on*, PAS-89(1):120–125, Jan 1970.

[12] Stephan Trenn. Switched differential algebraic equations. In Francesco Vasca and Luigi Iannelli, editors, *Dynamics and Control of Switched Electronic Systems - Advanced Perspectives for Modeling, Simulation and Control of Power Converters*, chapter 6, pages 189–216. Springer-Verlag, London, 2012.

[13] Elizabeth L. Yip and Richard F. Sincovec. Solvability, controllability and observability of continuous descriptor systems. *IEEE Trans. Autom. Control*, AC-26:702–707, 1981.

[14] P. C. Young and J. C. Willems. An approach to the linear multivariable servomechanism problem. *International Journal of Control*, 15(5):961–979, 1972.

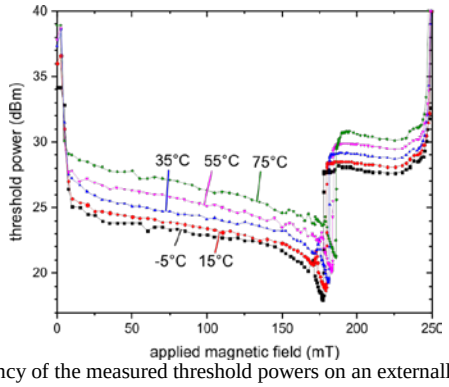
Influence of a thermal gradient on parametrically excited magnons in YIG-Pt bilayers

T. Langner, A.A. Serga, B. Hillebrands, V.I. Vasyuchka
Fachbereich Physik and Landesforschungszentrum OPTIMAS, TU Kaiserslautern
67663 Kaiserslautern, Germany
tlangner@rhrk.uni-kl.de

Our life is dominated by sophisticated electronic logic circuits. There is a need for advanced data processing schemes, but we face the problem that electronics approaches its ultimate physical limits. Integrated circuits have reached the minimal possible size and, as additional technological, economic and ecological problem, generate a lot of waste heat. Therefore, alternative data transfer and processing methods are needed. In our research group investigate magnetic excitations as information carriers, so called spin waves or magnons. These offer the vision of a "green" and fast technology. Due to their wave nature, new logic concepts can be realized [1].

Basic concept of a spin wave. The red arrows are the magnetic moments

Basic concept of a spin wave: In a ferromagnetic material all magnetic moments, or elementary magnets, are oriented parallel to an applied external magnetic field. If one magnetic moment is tilted with respect to this equilibrium position the magnetic field induced by this tilted magnetic moment forces neighbored magnetic moments to react. If this perturbation of the system is excited by an oscillating magnetic field we, thus, observe a propagation of this tilting through the material, the spin wave. In quantum mechanics, an elementary excitation of the system is described by a quasiparticle, here the magnon.
There are two main mechanisms for the interaction between two magnetic moments: The dipolar interaction, that means the attraction of different magnetic poles and the repulsion by magnetic poles with same sign on one hand, and the exchange interaction on the other hand. This quantum mechanical interaction forces neighbored magnetic moments to be aligned parallel. Due to their very short wavelength (less than micrometers) these exchange-dominated waves are of special interest for logic devices in microstructured systems. The well-established mechanism to excite those spin waves is the parametric pumping process. Hereby, an oscillating magnetic field with twice the frequency of the generated spin wave is applied parallel to the direction of the magnetization. Using this technique, two antiparallel spin waves with half the excitation frequency will be generated due to energy conservation. It is important to remark that this technique is characterized by a threshold process. The oscillating magnetic field needs to have a field strength that is large enough to overcome intrinsic magnetic damping processes. This threshold field can be determined by detecting the detuning of a microwave resonator applied to the investigated system [2].
Recent experimental results have shown that intrinsic magnetic damping can be influenced by the so called spin Seebeck effect [3]. It means that a temperature gradient along the interface between a magnetic layer (in our case the material Yttrium Iron Garnet, YIG) and a nonmagnetic metal (Platinum) generates thermally excited spin waves that contribute to the intrinsic damping. Due to very precise measurements this modification of the damping should be clearly detectable in a change of the applied microwave power necessary to reach the threshold condition. Our results show that a temperature gradient will not change the threshold and, thus, the magnetic damping. Furthermore our results show, that a homogeneous temperature change leads to a pronounced difference in the threshold field. Our results raise the question if the reported influence of the spin Seebeck effect on the damping really exists or if there is just an effect caused by homogeneous temperature change. Nevertheless, the effect is by far not as pronounced as reported. In my presentation the physical concept and measurement technique as well as the experimental results will be discussed.



Dependency of the measured threshold powers on an externally applied magnetic field for different constant temperatures of the sample. It is shown, that the threshold changes with temperature.

References:

[1] A.V. Chumak, V.I. Vasyuchka, A.A. Serga, B. Hillebrands, Nature Phys. **11**, 453-461 (2015).
[2] T. Neumann, A.A. Serga, V.I. Vasyuchka, B. Hillebrands, Appl. Phys. Lett. **94**, 192502 (2009).
[3] M.B. Jungfleisch, T. An, K. Ando, Y. Kajiwara, K. Uchida, V.I. Vasyuchka, A.V. Chumak, A.A. Serga, E. Saitoh, B. Hillebrands, Appl. Phys. Lett. **102**, 062417 (2013).

Simulation of spin-dependent hot-electron transport

Dennis Michael Nenno, Hans Christian Schneider
Physics Department and Research Center OPTIMAS
University of Kaiserslautern
67663 Kaiserslautern
Email: nenno@physik.uni-kl.de

***Abstract*—Current magnetic storage devices based on electronic currents are in need of faster data throughput. The technology in use is physically limited to read and write processes in the nanosecond regime. Experiments with ultrashort laser pulses have shown that it is possible to change the magnetization on a femtosecond time scale ("femtomagnetism"). We investigate the transport of electrons after excitation by ultrashort pulses in order to obtain a microscopic understanding of processes which may lead to progress in data storage.**

I. Introduction

Ever increasing hard drive capacities require the read and write techniques to keep up with the pace. Future storage densities will exceed $1\,\mathrm{Tbit/in^2}$ and current data transfer rates are on the order of a few Gigabit per second. Evidently, the quest for faster throughput techniques is on.

In magnetic hard drives, which is still the predominant form of storage technology, a bit is encoded in a very small region with a distinct magnetization direction. In a simple picture, the information can be read or modified by applying a magnetic field by the read head. The underlying technology of present-day read heads is based on the effect of giant magneto-resistance (GMR), for which Albert Fert and Peter Grünberg shared the 2007 Nobel Price in Physics. This effect was discovered in 1988 [1], [2], and found its first commercial application less than 10 years later in 1997.

If one expresses current data rates as related to one single event, i. e. switching a bit, they correspond to a switching time of less than a nanosecond. Both strong and short magnetic pulses are used in order to accomplish such comparatively fast switching times. However, it can be shown that this technique is limited to the picosecond timescale [3] and can not be improved below this.

II. Ultrafast demagnetization

In a ground-breaking experiment in 1996, Beaurepaire *et al.* [4] found that when irradiating a ferromagnetic nickel probe with an ultrashort, intense laser-pulse, the magnetization of the sample can be dramatically reduced within a few hundred some femtoseconds. This behavior is still surprising some 20 years after its discovery, despite a variety of possible explanations. In fact, before this experiment, it was thought that the magnetization dynamics were limited to a timescale of a few hundred picoseconds because the interaction of the laser-pulse with the electrons, whose angular momentum, or *spin*, is responsible for magnetism, acts in this regime. However, this experiment seemed to indicate that faster switching times can in principle be achieved, and a whole new field of "femtomagnetism" was started.

In search of the microscopic origin of this behavior, quite a few effects have been identified and it is very likely that all of them contribute to some extent. Most of them somehow relate the change in magnetization to the interaction of electrons with crystal deformations, magnetic excitations or directly with the laser field. Another promising suggestion are so called *superdiffusive* spin currents [5]. We will try to explain the idea behind this concept.

In the simple picture of an electron as a tiny bar magnet, a magnetic material consists of an unequal number of magnetic bars aligned facing north and south, which are supposed to represent the up- and down-electrons. We call the dominant contribution the majority electrons, as opposed to minority electrons. The laser now excites electrons from their equilibrium to states with higher energy in which they can — at first — freely propagate through the material. The optical excitation itself does not change the imbalance of majority and minority carriers, but just supplies energy. The argument, why the magnetization drops is then as follows. Due to the different energies, the electrons have different velocities and a varying probability to interact (or get scattered), mostly with other electrons. Because the transport out of the region where the electrons were excited by field thus becomes different for minority and majority electrons, the magnetization in this region changes. Thus, the difference in the transport properties of the carriers needs to be understood.

Much of the current understanding of the optically excited transport behavior comes from recent measurements [6] of

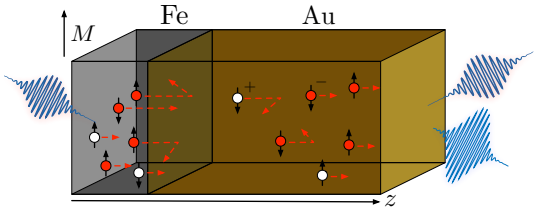


Fig. 1. Schematic setup for a pump-probe experimental on an iron/gold-bilayer. The magnetization of the iron (Fe) layer is indicated by M. The red laser pulse excites carriers in iron which subsequently propagate into the gold (Au) slab. The dynamics of the propagating carriers is measured on the right side the gold layer. Spin-up and spin-down electrons are indicated by arrows.

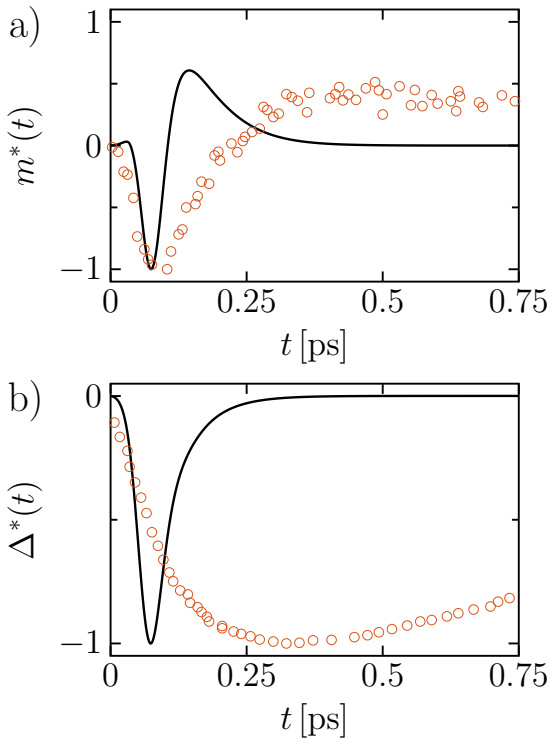


Fig. 2. Experimental data [6] (circles) and simulation results (solid line) for the spin (a) and charge dynamics (b) at the interface of the gold layer (see text).

the magnetization and charge dynamics in an non-magnetic gold slab adjacent to a thin magnetic iron layer. The setup is depicted in Figure 1 and consists of a thin (15 nm) iron layer and a larger 100 nm gold layer. After the optical excitation in the magnetic iron slab, the spin-polarized carriers propagate into the gold slab and are scattered along their path. The resulting dynamics is probed with an optical technique at the vacuum interface in gold. The detection method can distinguish between the charge and the spin of the incoming carriers. The experimentally obtained signal is shown in Figure 2. The signal corresponding to the spin imbalance at the interface shows a drop and a subsequent recovery, which can be attributed to carriers of different spin. In the charge signal, for which both spin-up and down electrons are counted, this feature is absent: only a fast rise and slow decay are observed. The authors of reference [6] offered a simple explanation in terms of the mean free path, that is the product of velocity and scattering time of the carriers, which was not based on a detailed microscopic analysis.

III. Simulating ultrafast transport

The existing model [5] for the transport of spin-polarized, hot (that is, excited) carriers is built on macroscopic concepts which are not derived from microscopic evolution equations. Starting with the basic Schrödinger equation of quantum mechanics and the model of free electrons in a crystal, one can derive the transport equation for an electron population. It tracks the distribution of electrons at different points in space and different velocities and is therefore able to capture the whole, complex dynamics after the excitation. After reducing the many-particle equations and interaction to a single particle picture, the resulting equation reads [7]

$$\begin{aligned}\frac{\partial \rho(\boldsymbol{R},\boldsymbol{k},t)}{\partial t} &- \frac{1}{2}\left\{\nabla_{\boldsymbol{R}}\,\bar{\epsilon}(\boldsymbol{R},\boldsymbol{k},t)\,,\nabla_{\boldsymbol{k}}\,\rho(\boldsymbol{R},\boldsymbol{k},t)\right\}\\ &+\frac{1}{2}\left\{\nabla_{\boldsymbol{k}}\,\bar{\epsilon}(\boldsymbol{R},\boldsymbol{k},t)\,,\nabla_{\boldsymbol{R}}\,\rho(\boldsymbol{R},\boldsymbol{k},t)\right\} \qquad (1)\\ &= \left.\frac{\partial \rho(\boldsymbol{R},\boldsymbol{k},t)}{\partial t}\right|_{\text{scat.}}.\end{aligned}$$

It models the evolution of the electronic distribution function taking into account their behavior as fermionic particles (two particles with the same quantum numbers can not be in the same state). The distribution function ρ contains the information of the probability to find a carrier at a certain point $\boldsymbol{R}$ in real space with wave-vector $\boldsymbol{k}$ or, equivalently, velocity $\boldsymbol{v}$. The different terms have a relatively clear physical interpretation, see reference [7].

We want to focus here on the last term in equation 1 which accounts for the scattering effects with electrons and their surrounding, that is the other electrons and the crystal lattice. It induces complex, non-linear dynamics and is numerically extremely demanding. Therefore, we approximate this term using the so called relaxation time approximation. The scattering term then reads

$$\left.\frac{\partial \rho(\boldsymbol{R},\boldsymbol{k},t)}{\partial t}\right|_{\text{scat}} = -\frac{\rho-\rho_{\text{eq}}}{\tau}\,. \qquad (2)$$

All single collisions between particles, which could be incorporated in detail, are summarized into the asymptotic behavior of the distribution function. From a non-equilibrium state induced by the optical excitation, it relaxes towards an equilibrium state ρ_{eq} with a particular relaxation time, depending on the incorporated scattering process.

To solve equation 1 the real space $\boldsymbol{R}$ variable and the wave-vectors $\boldsymbol{k}$ are numerically discretized. It is not possible to implement the full distribution function as function of three space and three wave-vector dimensions due to memory restrictions, and we therefore project the distribution function to an effectively two dimensional space, thereby drastically reducing the required memory. Technically speaking, in order to solve the hyperbolic, non-linear evolution equation, we use an operator splitting technique together with semi-Lagrangian solvers to deal with the remaining single transport steps [8].

Using this model with the experimental conditions from reference [6], we model the temporal evolution of magnetization and charge at the right gold interface. Typical results are shown in Figure 2. Although our model does not capture the exact temporal evolution as it is observed in the experiment, the main characteristics of the curves are well reproduced. In particular, we can explain the change of sign in the magnetization signal corresponding to the different contributions due to majority and minority carriers. In addition, in our microscopic model,

we can study the influence of model parameters on the measured dynamics, which are not easy to study experimentally. In cooperation with experiment, we can then obtain information about the physical mechanisms at work in ultrafast spin-dependent transport.

IV. Conclusion

Using a microscopic model for spin-dependent electronic transport, we simulated the complex dynamics after ultrashort laser excitation for a generic experimental setup. This constitutes an important step towards the understanding of microscopic spin-polarized electronic transport, which is one possible explanation for the enigmatic ultrafast demagnetization. In the bigger picture, future investigations can concentrate on the design of new techniques for magnetic recording, energy-efficient storage and sample optimization.

References

[1] M. N. Baibich, J. M. Broto, A. Fert, F. N. Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, "Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices," *Phys. Rev. Lett.*, vol. 61, p. 2472, 1988.

[2] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, "Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange," *Phys. Rev. B*, vol. 39, p. 4828, 1989.

[3] T. Gerrits, H. Van Den Berg, J. Hohlfeld, L. Bär, and T. Rasing, "Ultrafast precessional magnetization reversal by picosecond magnetic field pulse shaping," *Nature*, vol. 418, no. 6897, p. 509, 2002.

[4] E. Beaurepaire, J.-C. Merle, A. Daunois, and J.-Y. Bigot, "Ultrafast spin dynamics in ferromagnetic nickel," *Phys. Rev. Lett.*, vol. 76, p. 4250, 1996.

[5] M. Battiato, K. Carva, and P. M. Oppeneer, "Superdiffusive spin transport as a mechanism of ultrafast demagnetization," *Phys. Rev. Lett.*, vol. 105, p. 027203, 2010.

[6] A. Melnikov, I. Razdolski, T. O. Wehling, E. T. Papaioannou, V. Roddatis, P. Fumagalli, O. Aktsipetrov, A. I. Lichtenstein, and U. Bovensiepen, "Ultrafast transport of laser-excited spin-polarized carriers in Au/Fe/MgO(001)," *Phys. Rev. Lett.*, vol. 107, p. 076601, 2011.

[7] J. L. Cheng and M. W. Wu, "Spin diffusion/transport in n-type gaas quantum wells," *Journal of Applied Physics*, vol. 101, no. 7, p. 073702, 2007.

[8] G. Dimarco, Q. Li, L. Pareschi, and B. Yan, "Numerical methods for plasma physics in collisional regimes," *Journal of Plasma Physics*, vol. 81, p. 305810106, 2015.

Interpolated cut cell method for simulating behavior of granular materials

David Neusius[1,2], Sebastian Schmidt[1], Axel Klar[2]

[1]Fraunhofer Institute for Industrial Mathematics (ITWM), Department Flow and Material Simulation, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany

[2]University of Kaiserslautern, Department of Mathematics, Technomathematics Group, Erwin Schrödinger Straße, 67663 Kaiserslautern, Germany

***Abstract*—We develop methods for well-approximating boundaries of complicated domains in numerical partial differential schemes. The main interest is methods to solve non-Newtonian Navier-Stokes based systems of equations representing the behavior of granular materials. We present the developed interpolated cut-cell method (ICCM), which includes different interpolation schemes.**

I. Introduction

One might think at first glance that the importance of granular flow is insignificant compared to larger fields like fluid dynamics and the physics of solids. Yet, at a closer look one will discover many examples in everyday life, ranging from sand, soil and snow to powders, coffee beans, cereals and much further. Richard [1] states that "Granular materials are ubiquitous in nature and are the second-most manipulated material in industry (the first one is water)". As such the ability to reliably predict granular flow would be of great importance in industry as well as in the prevention or attenuation of disasters like avalanches or landslides.

The first preserved research of granular material was done by Coulomb in his laws of friction [2]. Yet, despite extensive ongoing investigations into granular flow, the prediction of granular behavior is not nearly as far progressed as it is for fluids. One main issue is the wide range of different regimes that is included in this topic.

It has been observed, that granular material can exhibit behavior similar to three states of matter: Solid, fluid and gas like behavior. Which characteristics are present is mostly dependent on the kinetic energy of the individual particles. A possible measure of this magnitude in the so called granular temperature, see subsection II-A.

The granular model that is used in this work was developed by Latz and Schmidt [3], [4] as a simplification of a model proposed by Savage [5], while still retaining all essential characteristics. It is based on a continuum approach, using modified non-Newtonian Navier-Stokes-Equations (NSE). In a non-Newtonian fluid the viscosity is dependent on the shear rate. In this granular flow model this also holds for other parameters and the dependence is non-linear. The modelling is based on spherical particles but by means of adjusting parameters via physical experiments, it is able to simulate more complicated particles.

So far, this model has only been successfully numerically solved on a rectilinear grid with a first order Finite Volume method and the possibility to locally refine the grid. This restriction to a rectilinear grid is due to the fact that the non-linearities result in a very fragile numerical stability. An extremely small time step is required to successfully solve the equations. When such a grid is used on complicated domains, a smooth boundary is discretized as a "stair"-like structure, if the boundary is not parallel to the Cartesian grid. For a compressible fluid this results in an additional error in density. Since the fluid properties of the macroscopic granular flow (see [3]) are dependent on the density, this density error tends to result in a large global error.

Our aim is to modify and apply an immersed boundary method that is well-suited to the three dimensional macroscopic granular model, in particular for numerical applications of this model that are often required. An immersed boundary in general is any object or numerical boundary that interacts with the fluid that is simulated, but is not included in the original rectilinear mesh formulation of the numerical method. It ranges from a solid wall to a highly elastic moving membrane. Depending on the method used, it could also represent an open boundary condition, like an in flow to or out-flow out of the simulation domain. Usually, such an immersed boundary does not coincide with the surfaces of the rectilinear mesh, often not even being parallel to any.

In this work we choose the cut-cell method among the known branches of immersed boundary methods. There have been many applications of the cut-cell method in the last decade. It has mostly been used for compressible non-viscous flows, e.g. [6], [7], [8]. Many papers have also shown its applicability to various incompressible viscous flows, see [9], [10], [11]. Only very recently people have started using it on the compressible viscous NSE, see [12], [13].

As one large focus in the modification of the method is the choice of interpolation method, we call the developed method interpolated cut-cell method (ICCM). The development of the method has been done within the framework of the PhD-thesis of David Neusius. Details about the method will be published there in 2016.

II. The continuum granular flow model

Interest in simulating granular materials is twofold. Firstly it is scientifically challenging due to the complex models involved. Despite the equations similarity to the NSE which have already been investigated in much detail in particular in a computational fluid dynamics context, many challenging aspects are specific to the granular flow model. Secondly they are economically interesting due to its widespread use in industrial processes. When an inside view into a production process is not possible, a simulation is important for better understanding and optimization.

Particle-based simulations of granular material are limited by particle numbers because of computation time and storage. This limit may be too small to simulate a complete production process. Furthermore, complex non-spherical and non-uniform particles pose difficult modeling challenges.

A continuum model can be derived via the Boltzmann equation using Chapman-Enskog theory. The derivation is very complex and can not be presented in this work (see [14]). We will only display the resulting macroscopic equations. Rather than scaling in run-time

with the number of particles as particle models do, the continuum model depends on the number of grid cells. Numerically this enables a comparison between coarse and fine grid simulations and thus, different precisions and run-times for identical particle sizes and domains. Since the cells are fixed in space, only interaction between neighboring cells has to be considered. Moreover, the granular properties required for the simulation, e.g. shear stresses, can be obtained from macroscopic laboratory experiments.

The model by Latz and Schmidt [3] which is shortly presented here has shown to be valid in two regimes of behavior of granular material: In the dilute regime binary collisions are dominant and kinetic gas theory can be applied, see [15], [14]. Even for higher densities Bocquet et al. [16] show agreement of the numerical simulation with experimental results, yet only under the assumption of a permanent source of granular temperature, in which case the granular flow shows liquid-like behavior. In the second regime, the static regime, the granular material is allowed to come to rest. This happens for instance in a sandheap. Since the kinetic model is not able to reproduce this, Savage [5] developed an extension to include this regime. The model by Latz and Schmidt [3] is a simplification of the one developed by Savage. The aim of the simplification is to be able to calibrate the model with as few parameters as possible. The model has been further reviewed and improved by Zemerli [17]. The complete set of equations is the following:

$$\partial_t c + \nabla \cdot (cu) = 0 \tag{1}$$

$$\partial_t (cu) + \nabla \cdot (cuu^T) - \nabla \cdot \sigma + \nabla p - F = 0 \tag{2}$$

$$\partial_t (cT) + \nabla \cdot (cTu) - \frac{3}{2}(\eta\kappa : \kappa - \nabla \cdot q) + \varepsilon cT = 0. \tag{3}$$

Equations (1) and (2) are the conservation equations for mass c and momentum cu. The mass is written as a volume fraction c , which is a density, scaled into the regime $[0, 1]$. Equation (3) is the temperature equation, see subsection II-A.

In these equations the following relations hold:

$$\sigma = \eta\kappa \qquad \kappa_{ij} = \frac{\partial u_i}{\partial x_j} \tag{4}$$

$$g(c) = \left(1 - \frac{c}{c_{\max}}\right)^{-1} \tag{5}$$

$$p = p_k + p_y \tag{6}$$

$$p_k = Tg(c)c \qquad p_y = \Theta(c - c_0)T_0(c - c_0)g(c) \tag{7}$$

$$q = -\lambda\nabla T \tag{8}$$

$$\lambda = \lambda_k(1 + \frac{p_y}{p_k}) \qquad \lambda_k = \lambda_0\sqrt{T}cg(c) \tag{9}$$

$$\eta = \eta_k(1 + \frac{p_y}{p_k}) \qquad \eta_k = \eta_0\sqrt{T}cg(c) \tag{10}$$

$$\varepsilon = \varepsilon_k(1 + \frac{p_y}{p_k}) \qquad \varepsilon_k = \varepsilon_0\sqrt{T}g(c) \tag{11}$$

The equations (4) denote the stress strain relation. It is necessary to compute forces that occur, if different parts of granular material move at different velocities while interacting with each other. The compressibility factor (5) is explained in subsection II-B. In the equations (6)-(11) we differ between kinetic terms with index k and yield terms with index y. Kinetic terms are analog to simulation of fluids, while yield terms are shortly explained in subsection II-C. Equations (6) and (7) denote the pressure, (8) and (9) the temperature diffusion and conductivity, (10) the viscosity and (11) the temperature dissipation (see subsection II-A).

A. Granular Temperature

The granular temperature T is essential for the granular model and is thus used in some way in most publications in this topic, see e.g. [3], [5], [14], [16], [18]. The granular temperature resembles the energy of "random movement" of particles, similar to the thermodynamic temperature being a measure of the random movement of molecules. This similarity indicates that the granular temperature dominates the dilute regimes, where granulate behaves in many respects like a gas. Assuming constant volume fraction, the higher the granular temperature the more often particles, just as molecules would, will interact.

Interactions between granular particles are non-elastic collisions. This leads to a constant loss of energy and is modelled as a dissipation term εcT (see (3) and (11)). At the same time shearing effects lead to an increase in granular temperature, which results in the viscous heating term $\eta\kappa : \kappa$ (see (3), (4) and (10)).

B. Maximum volume fraction

A further quantity required is that of a maximum volume fraction $c_{\max}$, i.e. the material dependent smallest number for which $c < c_{\max}$ always holds. The definition of the volume fraction $c = \frac{\rho_G}{\rho_M}$ combined with the incompressibility of the granular material, which implies $\rho_G \leq \rho_M$, lead to the trivial bound of $c_{\max} \leq 1$. Yet, since one has to expect remaining empty spaces between particles even for ideal ordering, the value is usually much lower.

The compressibility factor $g(\rho)$ (see (5)) is used in the model to ensure that given valid initial conditions any solution of the equation satisfies this bound on the volume fraction.

C. Yield expressions and transition density

With increasing density as well as decreasing granular temperature, the finite radius of particles requires additional forces that are not present in standard fluid equations. The assumption of instantaneous binary collisions of the kinetic model is now not valid anymore as particles may even rest permanently touching multiple neighbors.

A comparison of molecular temperature with granular temperature shows clear differences in this regime: Any material whose molecular temperature approaches zero Kelvin will contract strongly. This does not apply for a granulate which comes to a rest, which is equivalent to the granular temperature converging to zero. As stated in the previous section, the upper volume fraction limit $c_{\max}$ exist. Yet, one can further show in experiments with arbitrary particles that in addition to this random close packing $c_{\max}$ so called random loose packings exist. These are granular states that are mechanically stable and at rest but have a lower volume fraction. Such packings are possible due to friction and are furthermore very dependent on the shape of the particles. For spheres a random loose package of 0.55 can be observed, see [19], while the random close package $c_{\max}$ is about 0.64. For other types of particles it may be much lower in value as well as in relation to $c_{\max}$.

III. The interpolated cut-cell method

On one hand, the first order finite volume methods (FVM) on a Cartesian or rectilinear grid are very well suited methods for solving the granular flow model. In general these methods require little memory due to their simple geometric structure, consisting only of cuboids with edges parallel to the Cartesian axes, and are relatively fast and stable. Both of the latter properties are of outmost importance in numerically solving the granular model. The model shows inherent instabilities due to its nonlinear nature, see [3], [4]. This also implies that very many small time steps are required, making the duration of the simulation a major issue.

On the other hand, the granular flow model also requires well approximated boundary for numerically reproducing physical solutions. Yet, rectilinear grids can only approximate complicated boundaries step-wise, see Figure 1. Even using a high resolution, this is not sufficient in some examples, see e.g. subsection IV-A. Initially small errors in volume fraction may have a strong impact on the solution due to the non-linear dependency on volume fraction of the terms in the model.

Cut-cell methods improve rectilinear grids on complicated boundaries, while retaining as many of the positive properties of the rectilinear Finite Volume method as possible. This is done by introducing non-cuboid cells near the boundary, mostly without changing cells that are completely in the interior of the domain. Due to a strongly varying convection or diffusion dominance in the granular model, cut-cell methods found in literature cannot sufficiently improve the numerical solver on complicated domains.

To resolve these issues we develop the interpolated cut-cell method (ICCM). The ICCM consists of a grid generation and an interpolation. The grid generation ensures well approximating the geometrical domain boundary. The interpolation takes care of well discretizing the granular flow model on the approximated domain. Some interpolation is necessary to ensure first order convergence near the boundary, since the introduction of the cut-cells results in shifted center points of cells and faces and the creation of non-orthogonal boundary faces.

Depending on the choice of grid and interpolation method we name the method following the scheme

$$\mathrm{ICCM}^{\text{grid generation method}}_{\text{interpolation method}}.$$

For instance one specific ICCM method could be named $\mathrm{ICCM}^{\mathrm{V}}_{\mathrm{MS}}$. This denotes the combination of the Voronoi based grid generation step (V) with the modified Shepard interpolation step (MS). All possible choices as well as references are given in the following method overview.

The general outline of creating the final discrete equations with the ICCM is the following. We assume that a set of continuous equations and an arbitrary domain in two or three dimensions is given.

ICCM

1) Find a fitting underlying rectilinear mesh, considering the following:
   (i) Alignment.
   (ii) Resolution.
   (iii) Type of rectilinear mesh and aspect ratio.
2) Create a cut-cell grid.
   a) $\mathrm{ICCM}^{\mathrm{MC}}$: Cut-cell grid via Marching Cubes algorithm.
   b) $\mathrm{ICCM}^{\mathrm{BL}}$: Cut-cell grid via Marching Cubes algorithm with added boundary layers. Requires convex domain.
   c) $\mathrm{ICCM}^{\mathrm{V}}$: Voronoi grid creation. Allows boundary layers. Requires analytical description of domain.
3) Merge small cells. This is not necessary for $\mathrm{ICCM}^{\mathrm{V}}$.
4) Calculate cell properties: Cell volume, cell centroids, face areas, face normals.
5) Choose interpolation method.
   a) $\mathrm{ICCM}_{\mathrm{ML}}$: Multi-linear interpolation.
   b) $\mathrm{ICCM}_{\mathrm{MMLB}}$: Mixed multi-linear barycentric interpolation.
   c) $\mathrm{ICCM}_{\mathrm{MS}}$: Modified Shepard method, also called inverse distance interpolation.
6) Discretize on cut-cell grid including the chosen interpolation method.

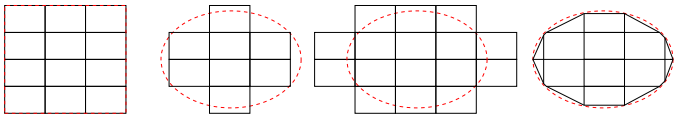


Figure 1. Two dimensional grid examples. Left: geometry fitting rectilinear grid. Center left: non-fitting rectilinear grid. Center right: Extending the rectilinear grid such that the desired domain is a subset of the union of cuboids. Right: cut-cell grid.

A. Grid generation

The automation of the generation of a general domain fitting grid with good numerical properties is subject to ongoing research. The cut-cell grid generation is one of the simpler variants, as it is based on a rectilinear mesh where only relatively few changes are necessary. The interior parts of the rectilinear mesh remain unchanged. This is an advantage, because it allows the usage of simpler and computationally faster Finite Volume implementations which only work on rectilinear meshes. On the other hand, the quality of the resulting grid near the boundary is usually lower than with more complicated domain fitting grid creation procedures, such as grid generation through differential systems or manually separating the domain into regions that are separately meshed. This is compensated by post processing the generated grid and adapting the numerical method.

A straight forward choice for the grid generation is the so called cutting procedure. One may regard it as first creating a rectilinear grid, overlapping the desired domain. Afterwards any parts of the cuboids that are not inside the domain are cut away, leaving polyhedron cells at the boundary, see Figure 1. This can be done in three dimensions by using the Marching Cubes [20] algorithm.

In an extension to this approach we create a layer of cells parallel to the boundary. The aim of this is to increase the regularity of the grid in the boundary layer regime of the flow. We will call these boundary-parallel cells “boundary layer”. This method is restricted to convex domains.

For these two approaches a merging algorithm is necessary to avoid too small cells. The criteria on the merging will be the average normal of all boundary faces of a small cell.

An alternative approach is using the Voronoi partition. This method can also include boundary layers but requires an analytical description of the domain. The principle of this grid creation is the following.

We choose points on the surface of our domain and add an interior Voronoi point in inner normal direction and an exterior Voronoi point in outer normal direction, both with the same distance. Doing a Voronoi diagram of these points, the interfaces between the interior and the exterior points will be an approximation of the domain boundary. The more points we add, the better the approximation becomes. Further adding Voronoi points on a rectilinear mesh in the interior of the domain, the set of all interior Voronoi cells can be used as the cells of our grid for the numerical computations.

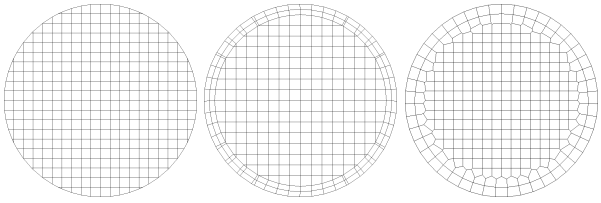


Figure 2. Comparing grid generation results for a circle using the three methods

B. Interpolation

Here, we will first give a list of requirements on an interpolation method to be used for the granular flow model. We then explain the motivation and check these requirements for various methods. Afterwards, we will shortly explain two of the methods which satisfy most requirements. Details on the other methods can be found in literature.

1) Requirements:

a) Continuity of the interpolation function $\hat{\xi}$ for arbitrary values on the source points and arbitrary positioning of source points. The latter is only demanded within the requirements of the rectilinear mesh based cut-cell method, i.e. we can assume one source point in every interior cell of the underlying rectilinear mesh.

b) Correct value at the interpolation sources. Together with (a) this is equivalent to $\lim_{x \to x_s} \hat{\xi}(x) = \xi_s = \hat{\xi}(x_s) \ \forall s \in S$.

c) Exact reproduction of a constant function. If weight functions exist (compare (e)) this is equivalent to a partition of unity of the set of weight functions for all source points $s \in S$.

d) Applicability to mostly unstructured as well as structured interpolation sources.

e) Possibility of precomputation of weights for each source point. Only one weight per source is necessary. This is equivalent to the interpolation process itself being fast and storage efficient during simulation.

f) Local interpolation: Includes as many points as necessary for first order but as few as possible. Due to the nature of numerical solver for linear systems, this is equivalent to minimizing the increase of computation time of implicit solver components.

g) Cheap precomputation, compare (e). Being a one-time step this has limited relevance.

2) Motivation for requirements: Requiring continuity of the interpolation function (a) is of special need in our specific case since we have to numerically approximate gradients within the interpolation function using two arbitrarily close points. Should there be a discontinuity in the interpolation function in between two such points, this could easily lead to a decrease in numerical precision or even stability issues. This is strongly connected with (b) since the discrete values on the source points are used without interpolation function within rectilinear parts of the domain. If our interpolation function does not reproduce these values we will have a discontinuity in between the interpolation regime near the cut-cells and the rectilinear interior.

The importance of the correct reproduction of a constant function (c) should be trivial. Most of the suggested interpolation methods satisfy this by normalizing the weights. Requiring almost arbitrary distribution if source points (d) follows directly from the observation of having a rectilinear as well as a mostly unstructured regime in the cut-cell method. The last three items are all concerned with computational efficiency. The method is applied to complex equations, which results in the numerical method being stiff for many configurations, despite being partly implicit. Thus, many time steps are needed. Not satisfying (e) means including a lengthy computation in between each one of those. At the same time, the implicit parts of the numerical method cause the requirement (f), as a larger stencil leads to a less sparse matrix, strongly increasing the computation time of any linear system solver.

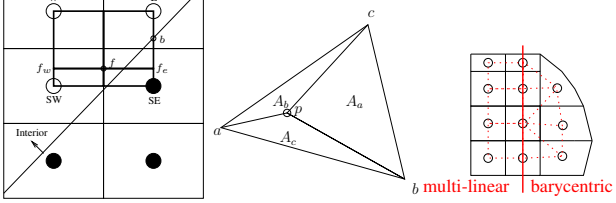
3) Fulfillment of the requirements: Table I gives an overview over which requirements are satisfied by which method.

4) Multi-linear and barycentric: Multi-linear interpolation only works with rectilinearly aligned sources. At the same time barycentric interpolation is not unique for Cartesian aligned sources, since even Delaunay triangulation is not unique in this case. Thus we suggest the combination of both, applying multi-linear interpolation in regions

Interpolation method	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Multi-linear	✓	(✓)	✓	✗	✓	✓	✓
Polynomial	✗	✓	✓	✓	✗	✓	-
Barycentric	✓	✓	✓	✗	✓	✓	✗
Mixed multi-linear/barycentric	✓	✓	✓	✓	✓	✓	✗
Least squares approximation	-	-	✓	✓	✗	✓	-
Natural neighbor	✓	✓	✓	✓	✓	(✗)	✗
Shepard's	✓	✓	✓	✓	✓	✗	✓
Modified Shepard's	✓	✓	✓	✓	✓	✓	✓

Table I
(a): CONTINUITY. (b): CORRECT SOURCE VALUE. (c): EXACT FOR CONSTANT FUNCTION. (d): STRUCTURED AND UNSTRUCTURED SOURCES. (e): PRECOMPUTATION. (f) LOCAL INTERPOLATION. (g): CHEAP PRECOMPUTATION.

where the source points are aligned and barycentric interpolation in other regions.



Bilinear Barycentric Mixed

5) Shepard's Method: Inverse distance weighting approach to interpolation was first published by Shepard [21]. A function ξ can be interpolated at any point p from a discrete set of known points S as

$$\hat{\xi}(x_p) = \begin{cases} \xi_s & , \text{ if } d(x_s, x_p) = 0 \text{ for some } s \in S \\ \frac{\sum_{s \in S} w_s(x_p)\xi_s}{\sum_{s \in S} w_s(x_p)} & , \text{ else} \end{cases} . \tag{12}$$

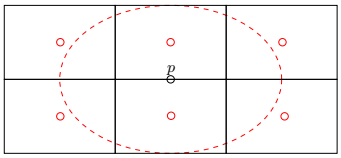
In Shepard's method the weighting w with regard to distance function $d(\cdot, \cdot)$ given as

$$w_s(x) = \frac{1}{d(x_s, x)^\beta} \tag{13}$$

This approach allows arbitrarily placed source points and ensures smoothness properties depending on $d(\cdot, \cdot)$ and $\beta > 0$. Yet, the original method is a global interpolation while we require a local interpolation. This is satisfied by the modified Shepard's method, first suggested by Franke [22]. It introduces a maximal distance R of a source point $s \in S$ to the interpolation point p:

$$w_s(x) = \max\left(0, \frac{1}{d(x_s, x)} - \frac{1}{R}\right)^\beta , \tag{14}$$

We use an elliptic norm as distance function to make sure that the structure of the underlying rectilinear grid is considered in the interpolation method. The elliptic radii will be the sizes of the rectilinear cells. This ensures that almost the complete two adjacent cells of any face are can be included in the interpolation area while avoiding unnecessary interpolation sources



IV. Numerical Results

We show some mostly qualitative numerical results of granular flow simulations with the ICCM. A detailed analysis and precise description and explanation of all parameters would go beyond the scope of this paper. The choice of examples covers some interesting and difficult aspects of the granular flow model.

A. Rotating cylinder

In many applications of the granular model it suffices to add average normal information of the curved domain to a standard Cartesian Finite Volume solver in order to achieve satisfying results. Yet, there are some examples where this approach fails. One such example is shown in Figure 3. To further simplify the example, we have omitted the temperature equation in this case and set a constant granular temperature $T = 10^{-5}$. We can assume that in the analytical solution the density and velocity norm should only depend on the radial distance. Yet, we have unrealistic piling of material in the Cartesian solver while the ICCM gives a plausible result. In Table II we further see that the convergence rate of the ICCM in L_1 as well as in L_∞ norm is close to one, which is the best value to be expected. Comparing this to the convergence rate of the Cartesian solver, we see that the ICCM not only gives much better results for the mesh plotted in Figure 3 but also improves much faster on finer meshes.

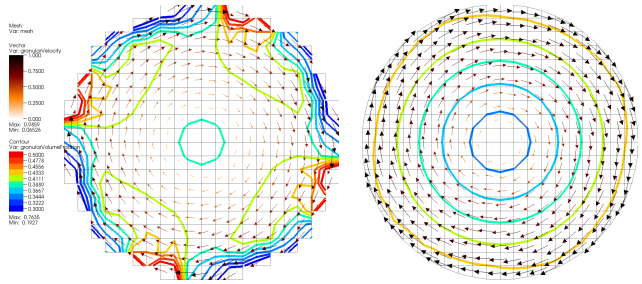


Figure 3. Rotation of granular material induced by a uniform tangential velocity on a circular no-slip boundary. Uniform initial condition $c = 0.4, u = 0$. Solid wall with tangential boundary condition $u = 0.5$ and no-slip condition.(left): Cartesian grid. (right): $\text{ICCM}^{\text{MC}}_{\text{MS}}$

h	Cartesian		$\text{ICCM}^{\text{MC}}_{\text{MS}}$	
h	$\frac{1}{10\pi}L_1$	$\frac{1}{10}L_\infty$	$\frac{10^{-3}}{\pi}L_1$	$\frac{1}{100}L_\infty$
Density Errors				
1/50	2.09	2.25	9.03	1.00
1/100	1.54	3.19	5.82	0.852
1/200	1.06	3.54	2.81	0.513
1/400	0.760	3.76	1.60	0.249
Density Convergence Rate				
1/100	0.445	-0.500	0.634	0.235
1/200	0.539	-0.150	1.05	0.731
1/400	0.476	-0.0873	0.810	1.04

Table II
Errors and convergence rates for this example using different uniform grid sizes h in both coordinate directions. An $\text{ICCM}^{\text{MC}}_{\text{MS}}$ solution with $h = 10^{-3}$ has been used as reference.

B. Heap flow

In contrast to the previous example, a more complicated domain and granular flow is presented here. This shows an interesting property of granular material, the capability of forming stable piles. This numerical experiment was suggested by group Midi [18].

The domain is composed of a truncated cone that is filled with a given volume fraction of granular material. This cone is connected to a mostly empty box, where below the cone a closed boxed is placed inside, leading to a step-structure. During the simulation, the granular material flows due to gravity out of the cone, into the larger box and onto the smaller box. This leads to pile of granular material first on the smaller box and, once these become too large and thus unstable in the lower parts of the larger box.

A reference solution is not available but the two as well as the three dimensional results (Figure 4 and 5) cover the behavior one would expect well. It could be compared to the behavior of sand in such a case. Furthermore, we see that the granular flow at the walls of the cone is very smooth, which would not be the case if a purely Cartesian solver would be used instead of the cut-cell method.

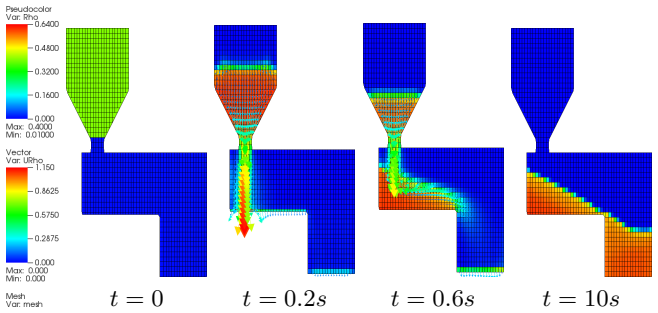


Figure 4. $\text{ICCM}^{\text{MC}}_{\text{MS}}$ results: Plot of the volume fraction and momentum of the flow from the truncated cone at different points in time for the two dimensional simulation.

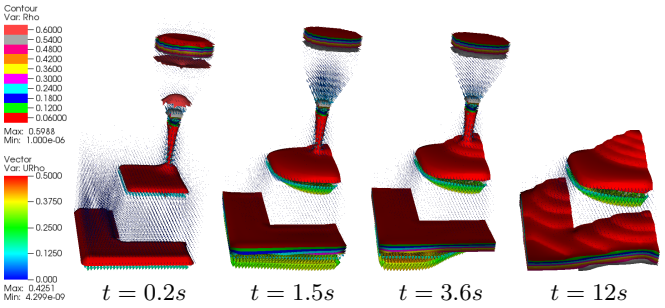


Figure 5. $\text{ICCM}^{\text{MC}}_{\text{MS}}$ results: Plot of the volume fraction and momentum of the flow from the truncated cone at different points in time for the three dimensional simulation.

C. Filling of a rotating cylinder

Another more complex example without a reference solution in the filling of a rotating cylinder. A further closed rotated cylinder in placed in the center of the cylinder that is being filled. This example is only solver three dimensionally but with two different grids, the cut-cell and the Voronoi grid, see Figure 6.

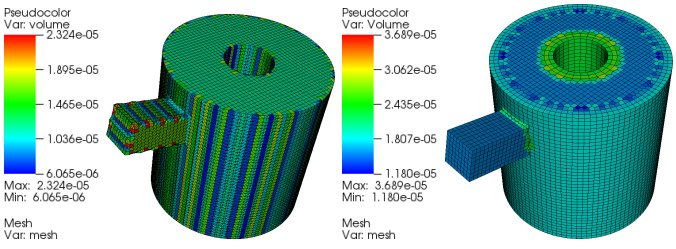


Figure 6. ICCM^{MC} and ICCM^{V} grid. Cartesian Grid: 60x44x44. Spatial step-size around 0.023m

As one can see in the results Figure 7 both methods seem to capture a realistic solution. One can see mostly in the later points of time that the result using a Voronoi grid is slightly more smooth which we think indicates that it is close to the correct solution.

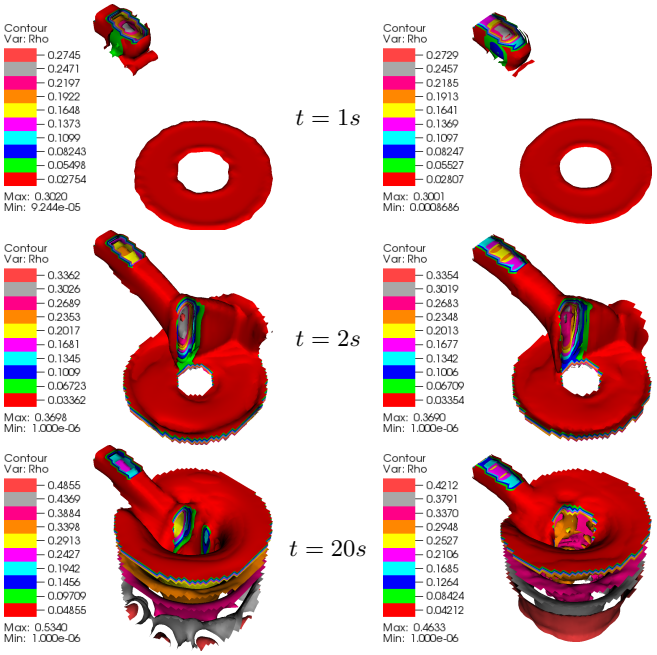


Figure 7. Left: $\text{ICCM}^{\text{MC}}_{\text{MS}}$. Right: $\text{ICCM}^{\text{V}}_{\text{MS}}$. Simulation time given row-wise.

V. Conclusion

We have presented the ICCM, which was developed at "Fraunhofer ITWM" in cooperation with "TU Kaiserslautern" in the PhD-thesis of David Neusius. It can be seen as a cut-cell extension in two and three dimensions to numerically solving the hydrodynamic granular flow model. In that model, granular flow has been characterized by a Navier-Stokes like set of equations with non-Newtonian non-linear rheology. We were able to include the complete model in our numerical results.

The granular flow model could previously only be numerically solved on a rectilinear grid which did not allow to correctly simulate granular flow at boundaries of complex domains. With the ICCM we have now resolved this issue.

To achieve this, we have presented three approaches of constructing a cut-cell grid that is to a varying degree based on an underlying Cartesian or rectilinear mesh. We have researched interpolation methods suitable for our specific task. We have presented and compared multiple possible interpolation approaches and have given reasons specific to the granular model as well as overall numerical scheme for our choice of method.

In an example we have shown first order convergence in different norms, including a maximum norm that is sensitive to artifacts that are commonly created directly at a non-grid fitting boundary. Second order convergence was neither in rectilinear grids, nor in cut-cell grids within the scope of this work. We furthermore have presented some qualitative numerical examples of granular flow with the different grid approaches in two and three dimensions.

Possible future work could address the inclusion of moving boundaries in the ICCM. This has been done before for other models. An open question is how large the additional computational effort of such a step would be. We expect that it would require a parallelization of the grid creation process.

There are a number of further numerical experiments that could be done in the future. First a numerical comparison of the different suggested interpolation methods, with respect to computation time as well as quality of results would be possible. Furthermore, three dimensional convergence studies have not yet been conducted.

Lastly, it would be interesting to try a higher order solver for the granular equations. We do not know if this would be possible even for a rectilinear grid, as stability issues would certainly be much stronger in such a scheme. Yet, if this could be achieved, we are confident that it would also be possible to extend the cut-cell method to an overall second order.

References

[1] P. Richard, M. Nicodemi, R. Delannay, P. Ribière, and D. Bideau. Slow relaxation and compaction of granular systems. *Nature Materials*, 4:121–128, 2005.

[2] C. A. Coulomb. *Essai sur une application des régles de Maximis et Minimis à quelques Problèmes de Statique, relatifs à l'Architecture.* Mem. Academ. Roy. 1776.

[3] Arnulf Latz and Sebastian Schmidt. Hydrodynamic modeling of dilute and dense granular flow. *Granular Matter*, 12(4):387–397, apr 2010.

[4] Sebastian Schmidt. *On Numerical Simulation of Granular Flow*. PhD thesis, 2009.

[5] S.B. Savage. Analysis of slow high-concentration flows of granular materials. *J. Fluid Mech.*, 377:1–26, 1998.

[6] G. Yang, D. M. Causon, and D. M. Ingram. Calculation of compressible flows about complex moving geometries using a three-dimensional Cartesian cut cell method. *Int. J. Numer. Meth. Fluids*, 33:1121–1151, 2000.

[7] R. Klein, K. R. Bates, and N. Nikiforakis. Well-balanced compressible cut-cell simulation of atmospheric flow. *Phil. Trans. R. Soc. A*, 367:4559–4575, 2009.

[8] Z. Z. Hu, D. M. Causon, C. G. Mingham, and L. Qian. Numerical simulation of floating bodies in extreme free surface waves. *Nat. Hazards Earth Syst. Sci.*, 11(2):519–527, feb 2011.

[9] T. Ye, Rajal Mittal, H. S. Udaykumar, and W. Shyy. An Accurate Cartesian Grid Method for Viscous Incompressible Flows with Complex Immersed Boundaries. *Journal of Computational Physics*, 156:209–240, 1999.

[10] Meng-Hsuan Chung. Cartesian cut cell approach for simulating incompressible flows with rigid bodies of arbitrary shape. *Computer & Fluids*, 35:607–623, 2006.

[11] Mark W. Johnson. A novel Cartesian CFD cut cell approach. *Computers & Fluids*, 79:105–119, apr 2013.

[12] Daniel Hartmann, Matthias Meinke, and Wolfgang Schröder. A strictly conservative Cartesian cut-cell method for compressible viscous flows on adaptive grids. *Computer Methods in Applied Mechanics and Engineering*, 200(9-12):1038–1052, feb 2011.

[13] Marsha Berger, Mercer St, Michael J Aftosmis, Nasa Ames, Moffett Field, and Steven R Allmaras. Progress Towards a Cartesian Cut-Cell Method for Viscous Compressible Flow. Technical report, 2012.

[14] Nikolai V. Brilliantov and Thorsten Pöschel. *Kinetic theory of granular gases*. Oxford University Press, 2003.

[15] James W. Dufty. Kinetic Theory and Hydrodynamics for a Low Density Gas. page 10, 2001.

[16] L. Bocquet, W. Losert, D. Schalk, T. C. Lubensky, and J. P. Gollub. Granular shear flow dynamics and forces: Experiment and continuum theory. *Physical Review E*, 65(1):011307, dec 2001.

[17] Clement Zemerli. *Continuum Mechanical Modeling of Dry Granular Systems : From Dilute Flow to Solid-like behavior*. PhD thesis, 2013.

[18] GDR Midi. On dense granular flows. *The European physical journal. E, Soft matter*, 14(4):341–365, 2004.

[19] Melissa Jerkins, Matthias Schröter, Harry L. Swinney, Tim J. Senden, Mohammad Saadatfar, and Tomaso Aste. Onset of Mechanical Stability in Random Packings of Frictional Spheres. *Phys. Rev. Lett.*, 101(1):18301, jul 2008.

[20] William E. Lorensen and Harvey E. Cline. Marching Cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4), 1987.

[21] Donald Shepard. A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM national conference*, pages 517–524, 1968.

[22] Richard Franke. Scattered data interpolation: tests of some methods. *Mathematics of Computation*, 38(157):181–181, 1982.

Improving Environmental Stress Cracking Resistance of glassy polymers by the incorporation of nanofillers

Jiraporn Nomai[1]
[1]Chair of Composite Engineering
University of Kaiserslautern
Kaiserslautern, Germany
E-mail: jiraporn.nomai@mv.uni-kl.de

Alois K.Schlarb[1,2,3]
[1]Chair of Composite Engineering
University of Kaiserslautern
Kaiserslautern, Germany
[2]Research Center OPTIMAS
University of Kaiserslautern
Kaiserslautern, Germany
[3]INM – Leibniz Institute for New Materials
Saarbrucken, Germany
E-mail: alois.schlarb@mv.uni-kl.de

I. Introduction

Environmental stress cracking (ESC) is defined as the premature initiation of cracking and embrittlement of plastic due to the simultaneous action of mechanical stress and aggressive fluids (known as stress cracking agents). In this case, the stress to failure of the polymeric material is significantly lower than the normal failure limit. ESC is one of the most common causes of unexpected brittle failure of polymers. It is responsible for approximately 25% of all serious premature failures of polymer-based components and particular risk when the product is used in the medical care [1-3]. This may result in the danger of medicine leakages in the therapy system which is serious problems for the safety of patients and users [3]. Amorphous polymers are extensively used in various applications including medical care products, whereas they are particularly susceptible to ESC. This is due to the lack of ordered structure in the amorphous polymers. Stress cracking agents are better able to diffuse into their molecular network, which promotes premature failure via molecular disentanglement. Thus, molecular weight plays an important role in the ESC behavior of amorphous polymers [4, 5].

Reinforcing polymers with small amounts of nano-sized fillers is one of the most effective methods for improving polymer properties and creep stability [6]. However, only a few studies have reported the behavior of reinforced polymers in terms of ESC resistance. Our previous research found that the addition of nano-SiO_2 particles into polycarbonate (PC) leads to a significant improvement in ESC resistance, which increases the service lifetime of the PC by about 72% [2].

To increase the service lifetime of amorphous polymer-based components and ultimately contribute to an improvement in the user safety, the effects of nanoparticle contents and the molecular weight of PC on the ESC resistance of PC nanocomposites in isopropanol were investigated in this study.

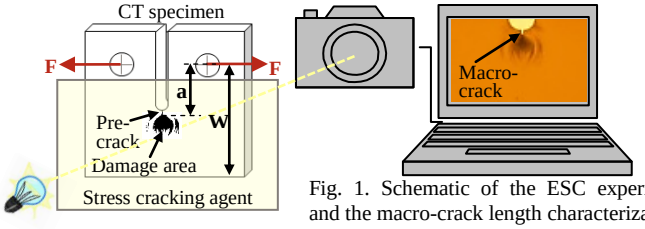


Fig. 1. Schematic of the ESC experiment and the macro-crack length characterization

II. Experimental and Characterization

A. Materials

Two different molecular weights of Bisphenol A polycarbonate (Bayer MaterialScience AG (Germany), Makrolon 2405 and Makrolon RX 1805) were used as polymer matrices. The average molecular weight (Mw) of the low-molecular weight PC (PC-L) and the high-molecular weight PC (PC-H) are 24,000 g/mol and 31,000 g/mol, respectively. A type of hydrophobic nano-SiO_2 (Evonik Industries (Germany), Aerosil R8200) was used as the nano-sized filler.

B. *PC/SiO_2 nanocomposites preparation*

PC nanocomposites with different nano-SiO_2 contents (0, 1, and 4 vol.%) were prepared using an optimized co-rotating twin-screw extruder (Theysohn, Theysohn Extrusionstechnik GmbH, Korneuburg) followed by injection molding (Arburg Allround 420C, ARBURG GmbH + Co KG, Loßburg). The procedure is described in the literature [2] in detail.

C. *Specimen preparation*

The compact tension (CT) specimens (36×36×4 mm^3) were machined according to ASTM D5045-99 and ASTM E399-90 standards. A pre-crack was made in the specimens by lightly tapping a fresh razor blade into the bottom of the saw slot. The pre-crack length (a_0) ranged from $0.45< a_0/W<0.55$.

D. *Experimental and Characterization of ESC*

A constant mechanical load in mode I (tension) was applied for the ESC testing specimens at a constant temperature of 23±1 °C. Isopropanol (CAS-No.: 108-32-7) was used as the stress cracking agent. The development of the macro-crack length was monitored and documented using a CCD-camera (EOS 1200D, Canon). Then, the macro-crack length (a) can be directly measured from the captured images, as shown in Fig. 1. The stress intensity factor (K_I) of a CT specimen in mode I was calculated using linear fracture mechanics which is described in detail in [2].

E. *Morphology of fractured surfaces*

The fractured surfaces of failed specimens were examined by using a Polarized Optical Microscope (POM) (ECLIPSE LV100, Nikon GmbH, Düsseldorf) and a Scanning Electron Microscope (SEM) (SU8000, Hitachi).

III. Results and Discussion

The macro-crack propagation rate as a function of the K_I factor of the polymer and its important parameters are shown in Fig. 2. The entire process of crack propagation of the polymer can be divided into three stages: the 1st stage - crack initiation (if $K_I<K_{Ith}$, the crack does not propagate and the material can withstand the applied load for an unlimited time), the 2nd stage - crack propagation (if $K_I>K_{Ith}$, the crack starts to propagate and the crack propagation rate increases linearly with the range of K_I), and the 3rd stage - final failure of material (if $K_I>K_{Ic}$, the unstable crack propagation begins and the material fractures in a right short time) [4, 7]. Table I shows K_{Ith}, K_{Ic}, and slope of curve as a function of nanoparticle contents with different molecular weights of PC. It is noticed that the slope of curve $((\Delta da/dt)/\Delta K)$ in the stable crack growth region decrease with increasing nano-SiO_2 contents. Simultaneously, K_{Ith} and K_{Ic} values increase with increasing nano-SiO_2 contents, especially in the case of high molecular weight PC. These results demonstrate that the addition of nano-SiO_2 particles to PC leads to slower crack growth and thereby increases the ESC resistance and service lifetime of PC-based nanocomposites. A comparison at any given nano-SiO_2 contents reveals an improvement in ESC resistance with increasing molecular weight of PC. This result is reflected in a decreased crack propagation rate while K_{Ith} and K_{Ic} are shifted toward higher values. This is attributed to the higher molecular weight provides more chain entanglement of the polymer, which lead to greater resistance to the stress cracking agent transport [1, 4].

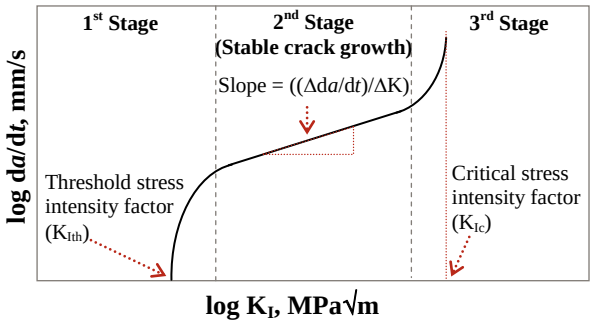


Fig. 2. Macro-crack propagation rate as a function of K_I factor of the polymer

TABLE I. K_{Ith}, K_{Ic}, AND SLOPE OF CURVE AS A FUNCTION OF NANOPARTICLE CONTENTS WITH DIFFERENT MOLECULAR WEIGHTS OF PC

Samples	SiO_2 (vol.%)	K_{Ith} (MPa√m)	K_{Ic} (MPa√m)	Slope m (-)
PC-L-S0	0	3.32±0.04	4.23±0.05	3.51±0.33
PC-L-S1	1	3.39±0.05	4.58±0.05	3.06±0.07
PC-L-S4	4	3.44±0.03	4.79±0.12	3.03±0.29
PC-H-S0	0	3.92±0.07	5.09±0.17	2.97±0.20
PC-H-S1	1	4.15±0.03	5.78±0.15	2.19±0.20
PC-H-S4	4	4.32±0.10	6.39±0.12	1.65±0.08

Optical micrographs of the fractured surfaces of failed specimens are shown in Fig. 3 (left). The onset of catastrophic failure increases with the increased of nano-SiO_2 contents and the increased of molecular weight of PC. This result correlates with the higher lifetime of PC nanocomposites. SEM micrographs of the fractured surfaces (Fig. 3, right) reveal cavities, particle/matrix debonding and localized micro-deformations of the nanocomposites matrix which are the major energy dissipation mechanism. This characteristic relates

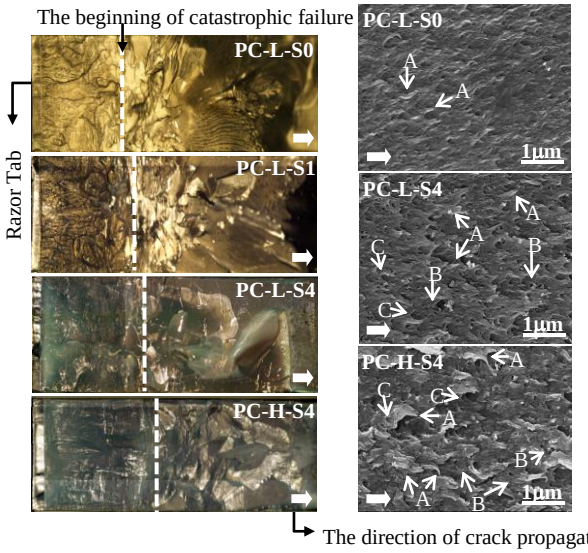


Fig. 3. Optical micrographs (left) and SEM micrographs (right) of fractured surfaces of PC and its nanocomposites. Features A, B, and C in SEM micrographs indicate plastic deformation, formation of cavities/nanoparticle debonding, and structures in parabolic shape, respectively

to the higher fracture toughness of PC nanocomposites. With increasing the molecular weight of PC, the SEM micrograph shows more dissipative energy structures of the matrix.

IV. Conclusions

It can be stated that the incorporation of nano-fillers as well as the higher molecular weight of the polymer shift the crack growth threshold to higher values, leads to slower crack growth, and higher critical stress intensity factor, and accordingly increases the service lifetime of polymer-based components significantly. The improvement due to the incorporation of 4 vol.% nano-filler seems to be on the same level as the higher molecular weight. However, the addition of nano-fillers is more effective with the higher molecular weight of the PC. Fractography analysis suggests that the formation of cavities, nanoparticle/matrix debonding, and localized micro-deformation of the matrix induced by nanoparticles as a result of energy dissipation are responsible for the improvement of the ESC resistance. These results ultimately contribute to an improvement in the user safety of polymer-based components in various applications.

Acknowledgment

The authors thank the German Research Foundation for the financial support. We are grateful to Nano Structuring Center, TU Kaiserslautern for the cooperation of experiments. We also gratefully acknowledge the Bayer MaterialScience and the Evonik Industries, Germany, for the supply of the materials.

References

[1] N.B. Sanches, M.L. Dias, and E.B.A.V. Pacheco, Polym. Eng. Sci., vol. 48, pp. 1953-1962, 2008.
[2] L. Lin, and A.K. Schlarb, J. Mater. Sci., vol. 47, pp. 6614-6620, 2012.
[3] A.K. Schlarb, Med. Device. Technol., vol. 13, pp. 30-34, 2002.
[4] V. Altstädt, Adv. Polym. Sci., vol. 188, pp. 105-152, 2005.
[5] V. Altstädt, S. Keiter, M. Renner, and A.K. Schlarb, Macromol. Symp., vol. 214, pp. 31-46, 2004.
[6] D. Pedrazzoli, and A. Pegoretti, Polym. Bull., vol. 71, pp. 2247-2268, 2014.
[7] J. Mencík, Mechanics of Components with Treated or Coated Surfaces, vol. 42. Springer Netherlands, 1996.

A combined sensitivity analysis and model reduction workflow for the simulation of cake filtration

Sebastian Osterroth[1,2], Oleg Iliev[1,2], René Pinnau[2]

[1]Fraunhofer Institute for Industrial Mathematics (ITWM), Department Flow and Material Simulation,
Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
e-mail: {osterroth,iliev}@itwm.fraunhofer.de

[2]University of Kaiserslautern, Department of Mathematics, Technomathematics Group,
Erwin Schrödinger Straße, 67663 Kaiserslautern, Germany
e-mail: pinnau@mathematik.uni-kl.de

***Abstract*—The macroscopic modeling of cake filtration involves a lot of parameters. Therefore it is important to know, how the solution depends on the parameters. Having sensitivity information available, we can use the information in combination with model reduction for the simulation of cake filtration. Two techniques, which use the sensitivity information in the POD snapshot matrix are considered. They differ in the choice of the weighting of the different entries of the snapshot matrix. In one technique the weights are changed for every new parameter configuration. This leads to higher computational costs, since for every new parameter configuration the POD modes are computed. To overcome this drawback, we combine the technique with interpolation. This reduces the computational costs up to 5 times.**

I. Introduction

A. Motivation

The separation of small particles from a suspension is an important task in many industrial applications. For example the clarification of water or the filtration of oil in an automotive engine can be named. Thus, mainly a porous medium, also called filtering medium, is used in such processes. Two important key factors of those processes are the particle capturing efficiency and the flow rate - pressure drop ratio. The first factor is connected to the purity of the fluid whereas the second one is mainly measuring the energy efficiency [12]. A third important criterion is the capability of a filtering medium to store dirt. This is mostly connected to the lifetime of a filter, but for our considerations we relax this criterion, since we want to study a phenomenon called cake filtration (e.g. [21]). Here we explicitly force the filtering medium to be completely clogged, such that particles are deposited at the surface of the filtering medium. The build-up of dirt at the surface is called filter cake.

From a mathematical point of view different challenges can be identified. The most obvious might be the description of the flow of a fluid, both in the free region and through the porous medium (e.g. [12]). But we do not address this topic in more details, but take care of the capturing of the particles inside and on the surface of the medium, which is related to the growth of the filter cake. This means that we are considering two filtration mechanisms in parallel. The deposition of particles inside a filtering medium (and also inside a filter cake) is called *depth filtration*, the deposition on the surface cake filtration. Thus we are considering a combined cake- and depth filtration model. Similar studies can be found in [15], [22]. The concentration of suspended particles is described by an advection-diffusion-reaction equation, the concentration of deposited particles by a kinetic expression. Those expressions are well-known from depth filtration theory, e.g. [6], [10], [11], [24]. The equations are accomplished with an equation for the growth of the cake, which is an evolution equation depending on fluid, particle and cake properties. This leads to a moving boundary problem.

B. Sensitivity analysis and model reduction

Since different process conditions affect the result, the influence of system parameters on the solution is investigated. Therefore a sensitivity analysis is performed. Since from a numerical point of view for every new set of parameters the whole system has to be computed, we try to reduce this afford by using model reduction. The basic idea of model reduction is to express a system by means of a few basic functions (e.g. [16], [18], [23]). Therefore data have to be available either in form of experiments or precomputed solution trajectories.

In this article we use different model reduction techniques, which are based on proper orthogonal decomposition (POD). The idea is to combine information from the solution and the sensitivities to obtain an improved reduced basis. We will just consider the sensitivities of the solution of our system as in [4], [5], [19], but other authors are also considering the sensitivities of the reduced basis functions [9], [14], [20], [25]. We are using the techniques first as proposed by the authors and then combine them with interpolation.

The article is organized as follows. In the next section we discuss the mathematical model. In section III we show how to compute sensitivities via a partial differential equation (PDE). This is followed in section IV by the presentation of different model reduction techniques. After some comparison results in section V the article closes with some conclusions.

II. Description of the mathematical model for cake filtration

A. Model assumptions

For the formulation of the mathematical model several assumptions are needed. First we assume that we have relatively dilute suspensions, i.e. the particle concentration is low. Further we assume a constant inflow flux C_{in}, which will lead to an increase of pressure during the filtration process. A constant amount of particles is deposited at the surface of the medium and is building a filter cake with a given cake porosity ϕ_{cake}. The filter cake itself is considered as incompressible. The remaining amount of particles is entering the filter cake and the filtering medium in a depth filtration manner.

We model the process on a macroscopic scale based on a phenomenological approach [6], [10], [11]. Let $\Omega = \Omega_p(t) \cup \Omega_f(t)$ denote the domain, splitted in $\Omega_p(t)$ the porous domain and $\Omega_f(t)$ the fluid domain. The domains depend on time, since we have a moving

boundary problem and the interface between the domains moves with time. The transport of dirt (particles) of concentration C with a fluid of velocity u is modeled by an advection-diffusion-reaction equation

$$\frac{\partial C}{\partial t} + u \cdot \nabla C - D\Delta C = -\frac{\partial M}{\partial t} \quad \text{in } \Omega_p(t), \tag{1}$$

where the reactive part $\frac{\partial M}{\partial t}$ (so called adsorption rate) is modeling the deposition of particles. The adsorption rate is depending on the mass flux [6] and can in the simplest case be modeled by a kinetic expression as

$$\frac{\partial M}{\partial t} = \alpha \|u\| C \quad \text{in } \Omega_p(t), \tag{2}$$

where α is the so called adsorption rate. For more involved models we refer to the literature [10], [24]. As initial conditions we have $C(0, L_0) = (1-\lambda)C_{in}$, $M(0, L_0) = 0$ and $L(0) = L_0$. For the thickness of the cake L the above assumptions are used. Since a given amount of particle concentration λC_{in} is building the cake with a given porosity, the following formula can be derived [13]

$$\frac{\partial L}{\partial t} = \lambda \frac{C_{in}\|u\|}{\rho_s(1-\phi_{cake})}, \tag{3}$$

where ρ_s is the density of the particles.

B. Numerical solution

For our study we are just considering the one-dimensional case. Thus the above model can be simplified. For the numerical solution the well known Landau transformation (see e.g. [1]) is used

$$y = \frac{x}{L(t)}. \tag{4}$$

This method is classified in the framework of front-fixing methods [1], also called boundary immobilization (BIM) technique [17]. The equations (1) and (2) (in 1d) are transformed to

$$\begin{aligned} \frac{\partial C}{\partial t} + \left(\frac{u}{L(t)} - y\frac{L'(t)}{L(t)}\right)\frac{\partial C}{\partial y} - \frac{D}{L(t)^2}\frac{\partial^2 C}{\partial y^2} \\ = -\frac{\partial M}{\partial t} + y\frac{L'(t)}{L(t)}\frac{\partial M}{\partial y} \end{aligned} \tag{5}$$

$$\frac{\partial M}{\partial t} - y\frac{L'(t)}{L(t)}\frac{\partial M}{\partial y} = \alpha\|u\|C. \tag{6}$$

Note that we skipped changing the notation of C and M for the transformed variables.

III. Sensitivity analysis

As in many realistic applications the above mathematical model is depending on a bunch of parameters. For design or optimization of the process it is often necessary to investigate the influence of the parameters on the process output. Also the understanding of the effect of a small change in the input parameters on the output parameters is often of interest. Sometimes the input parameters depend on measurements with some uncertainty, thus the input parameters lie in a specific range and it is interesting to know, if also the output stays in a bounded range.

For our considerations we take a look at the following reference problem for some variable z: Given a set of parameters $\mu \in \mathcal{D} \subset \mathbb{R}^p$, $\mu = [\mu_1, ..., \mu_p]^T$ and the system

$$\frac{\partial z}{\partial t} + \vartheta_1(t)A_1(y;\mu)\frac{\partial z}{\partial y} - \vartheta_2(t)A_2(\mu)\frac{\partial^2 z}{\partial y^2} = 0 \tag{7}$$

with parameters varying in a box $[\mu_{low}, \mu_{up}]$, we want to know how the solution $z(t, y; \mu)$ behaves under parameter change. Therefore we define the sensitivity as

$$s_i(t, y) = \frac{\partial z}{\partial \mu_i}(t, y; \mu), \ i = 1, ..., p. \tag{8}$$

Then the equation for computing the sensitivity can be deduced from (7) by differentiating the equation wrt μ_i, $i = 1, ..., p$. This yields

$$\begin{aligned} \frac{\partial s_i}{\partial t} + \vartheta_1(t)A_1(y,\mu)\frac{\partial s_i}{\partial y} - \vartheta_2(t)A_2(\mu)\frac{\partial^2 s_i}{\partial y^2} = \\ - \vartheta_1(t)\frac{\partial}{\partial \mu_i}A_1(y,\mu)\frac{\partial z}{\partial y} + \vartheta_2(t)\frac{\partial}{\partial \mu_i}A_2(\mu)\frac{\partial^2 z}{\partial y^2} \end{aligned} \tag{9}$$

for $i = 1, ..., p$. Also the boundary conditions have to be differentiated. Note that the sensitivity equations is always linear, also if the original equation is nonlinear. Therefore computing the sensitivities is in the most cases easier than solving the original system. Due to this reason we prefer the above method to computing the sensitivities by finite differences.

IV. Model reduction

A. Proper orthogonal decomposition (POD)

In this subsection we briefly introduce the model reduction technique proper orthogonal decomposition. For more details we refer to [16], [18], [23]. Assume that the solution of a discretized version of equation (7) (e.g. by finite volume or finite element method) is available at a time grid $0 = t_0 < t_1 < ... < t_{N_T} = T_{end}$. Denote $z_j = z(t_j, \cdot), j = 0, ..., N_T$. Those information can either be obtained by experiments or numerical simulations. Then the following minimization problem can be formulated

$$\begin{aligned} \min_{\phi_1,...,\phi_\ell} \mathcal{J}(\phi_1, ..., \phi_\ell) = \sum_{j=0}^{N_T} \left\| z_j - \sum_{i=1}^{\ell} \langle z_j, \phi_i \rangle \phi_i \right\|^2 \\ \text{s.t. } \langle \phi_i, \phi_j \rangle = \delta_{ij}, \quad i, j = 1, ..., \ell \end{aligned} \tag{10}$$

in some given norm induced by the inner product $\langle \cdot, \cdot \rangle$. The orthonormal basis solving this problem is called POD basis or POD modes. The solution is given by an eigenproblem, namely

$$ZZ^T\phi_i = \lambda_i \phi_i, \tag{11}$$

where $Z = [z_1|...|z_{N_T}]$ is the so-called snapshot matrix. The problem can alternatively be solved by singular value decomposition or the method of snapshots (see e.g. [18], [23]). The error (or residual) can be formulated in terms of the eigenvalues

$$\sum_{j=0}^{N_T} \left\| z_j - \sum_{i=1}^{\ell} \langle z_j, \phi_i \rangle \phi_i \right\|^2 = \sum_{i=\ell+1}^{d(n)} \lambda_i,$$

where $d(n)$ denotes the number of nonzero eigenvalues of Z. This error also motivates the choice of ℓ. It should be chosen in such a way, that the error is sufficiently small. A measure for ℓ based on the total energy can be found in the literature.

B. Reduced-order modeling

Assume that we have computed a POD basis $\Phi = [\phi_1, ..., \phi_\ell]$ and consider the discretized version of (7). Then a reduced-order system can be defined by using Galerkin projection and $z \approx \tilde{z} = \Phi z_\ell$ to obtain

$$\frac{\partial z_\ell}{\partial t} + \vartheta_1(t)A_{1\ell}(y,\mu)\frac{\partial z_\ell}{\partial y} + \vartheta_2(t)A_{2\ell}(\mu)\frac{\partial^2 z_\ell}{\partial y^2} = 0. \tag{12}$$

The reduced matrices are defined as $A_{1\ell} = \Phi^T A_1 \Phi$ and $A_{2\ell} = \Phi^T A_2 \Phi$. Note that also the initial condition needs to be projected. (12) is now a low dimensional problem of size ℓ for determining the coefficients z_ℓ. In the sequel we investigate the relative error

$$e_{rel} = \frac{\|z(t;\mu) - \tilde{z}(t;\mu)\|}{\|z(t;\mu)\|}. \tag{13}$$

The formulation in (12) allows for an offline/online decomposition (e.g. [8]). Since the reduced matrices does not depend on time t, they can be precomputed in an offline step and in each online step, just an update with the scalar functions $\vartheta_1(t)$ and $\vartheta_2(t)$ is necessary.

C. Compact POD (CPOD)

This technique was proposed in [4], [5] for parametrized discrete static systems. As mentioned in their conclusions the technique can be extended to transient problems and we will use it in this framework. Therefore we just present the basic idea, for details we refer to [5]. The idea is to include the sensitivities into the considerations for determining a reduced basis. This allows a meaningful truncation of the basis. Assume that z is the solution of the parameter dependent equation (7) for a given parameter μ^0. Then for a new parameter $\bar{\mu}$ Taylor expansion can be applied to get

$$z(\bar{\mu}) \approx z(\mu^0) + \sum_{i=1}^{p} \frac{\partial z}{\partial \mu_i}(\mu^0) \cdot (\bar{\mu}_i - \mu_i^0). \tag{14}$$

We restrict ourselves to a first order expansion, this can be generalized to any order as long as the sensitivities are available. Then we can define a new snapshot matrix $X(\mu^0, \bar{\mu})$ as

$$X(\mu^0, \bar{\mu}) = W(\mu^0)\Gamma(\mu^0, \bar{\mu}), \tag{15}$$

where

$$W(\mu^0) = \left[z(\mu^0) \ \frac{\partial z}{\partial \mu_1}(\mu^0) \ \ldots \ \frac{\partial z}{\partial \mu_p}(\mu^0) \right] \tag{16}$$

$$\Gamma(\mu^0, \bar{\mu}) = diag\left(1, (\bar{\mu}_1 - \mu_1^0), ..., (\bar{\mu}_p - \mu_p^0)\right). \tag{17}$$

D. Derivative extended POD (DEPOD)

Schmidt and coworkers [19] proposed this technique in the framework of parameter estimation problems with (parabolic) PDE constraints. As in the case of CPOD, parametric derivatives are included in the POD snapshot matrix. Here the snapshot matrix is defined as

$$X(\mu^0) = \left[\zeta_0 z(\mu^0) \ \zeta_1 \frac{\partial z}{\partial \mu_1}(\mu^0) \ \ldots \ \zeta_p \frac{\partial z}{\partial \mu_p}(\mu^0) \right], \tag{18}$$

where ζ_0 and ζ_k, $k = 1, ..., p$ are weighting factors. The authors propose a weighting with the norm, i.e.

$$\zeta_0 = 1/\|z\|, \quad \zeta_k = \theta / \left\| \frac{\partial z}{\partial \mu_k} \right\|, k = 1, .., p \tag{19}$$

with an additional scaling $\theta \in (0, 1]$.
Note that in comparison with CPOD the snapshot matrix is not depending on $\bar{\mu}$.

E. Interpolation of POD modes

Since for the case of CPOD the snapshot matrix is reformulated for every new parameter value $\bar{\mu}$, the computation can be very expensive. Therefore one can think on computing the CPOD basis just in specific points and then use it in a specific neighborhood[1]. This can be interpreted as a localization technique in parameter space (contrary to a global basis for the whole parameter space). But it can also be viewed as a kind of nearest neighbor interpolation. Thus one can also think on using other interpolation techniques between multiple reduced bases to improve the approximation result. Therefore we briefly mention an interpolation technique for reduced bases. One difficulty is to ensure one essential property of the basis, namely the orthogonality. Therefore the bases are considered as elements on the Grassmann manifold of subspaces. Then an interpolation is taking place in the tangent space to this manifold. For details we refer to [2], [3], [7], [25].

[1]Private communication with Kevin Carlberg

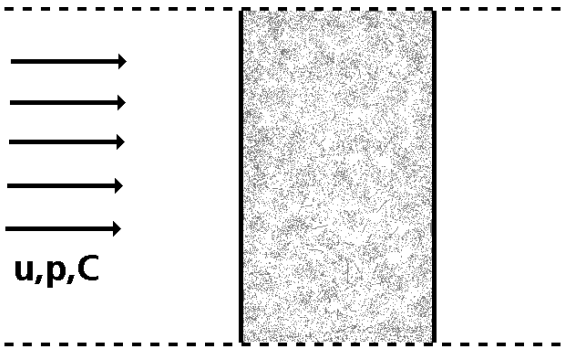


Fig. 1. A schematic plot of the flat filtering medium in 1d with one-sided flow

V. Results

We test the above methods for a flat filtering medium with constant and uni-directional flow (see figure 1). This allows us to reduce the model to 1d, such that we can use equations (5) and (6). These equations are repeated 117 times, since we have that number of different particle sizes (with different parameter values). As parameters we are considering the velocity u (now 1d) and the initial concentration C_{in}. The equations are discretized with a finite volume scheme and then solved in MATLAB. Since two parameters are considered the parameter domain $\mathcal{D} \subset \mathbb{R}^2$ is a rectangular. It is depicted in figure 2. To test the above methods we are just considering data in the midpoint of the domain. Assume that at this point solution and sensitivity information are available. We are comparing our results to two bounding cases: (i) The optimal POD approximation, which is obtained by first computing the whole solution trajectory for a new parameter configuration and then build the reduced basis from this information. This can be seen as the best achievable approximation, but computationally most costly. (ii) The old POD approximation (also called baseline in [9] or fixed in [14]), obtained from using the basis from the reference point in the middle in every other point. This can be seen as best achievable approximation with data from just one sampling point, but is computationally the cheapest alternative. We

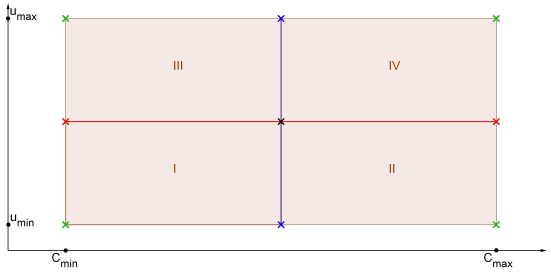


Fig. 2. 2d parameter domain with different sampling points.

are considering three different methods of using the sensitivities as additional information. The CPOD snapshot matrix is defined as in (15) and the DEPOD snapshot matrix (18) is defined in two ways: With the weights as in (19) and with all weights equal to 1. To the latter we refer as SenPOD. We investigate two cross-sections in the parameter domain: The blue line with constant concentration and varying velocity and the red line with vice versa roles (see fig. 2). Discretizing with eight points in space and 60 time steps, we have for 117 particle sizes 112320 unknowns. We use $\ell = 4$ POD modes for approximation. The results are shown in figure 3 for the blue line and in 4 for the red line. The midpoint of the figure corresponds in

both cases to the sampling point. One can see that in this point the optimal, old and CPOD method coincide, since they are considering the same snapshot matrix and thus yield the same reduced basis. Let us first take a closer look at CPOD. In the neighborhood of the sampling point it follows the curve of the old approximation. This means that near the sampling point the sensitivity information seems less important or due to the scaling too small to be included in the basis. Then from a point on sensitivity information is included and the approximation is improved in comparison to the old approximation. Looking at the DEPOD and SenPOD approximation we have to differ for the considered parameter. For the variation of velocity (fig. 3) the two models provide an approximation, which has an larger error than compared to the old and the optimal. Due to construction it also does not coincide in any point. The variation of concentration is different (fig. 4): Here the relative error is in the same order of magnitude, but the old approximation has an larger error. Therefore DEPOD and SenPOD offer a good approximation over the whole parameter range. In the neighborhood of the sampling point they can clearly not capture the specific solution. Comparing these two methods the behaviour is similar. This may mean that the influence of the weighting is not that large, if is taken constant over the whole parameter range. Changing the weights as in CPOD can be seen to give better results.

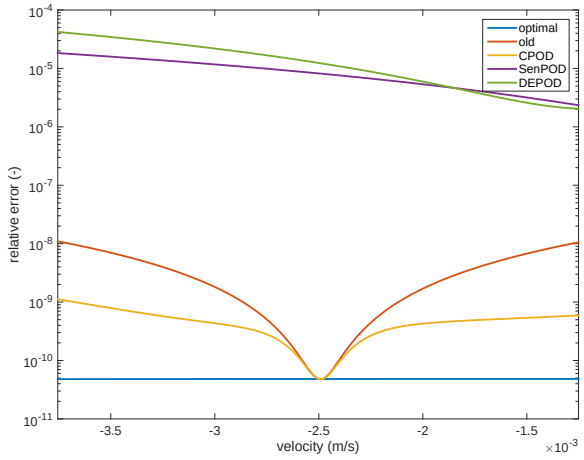


Fig. 3. Relative error for the variation of velocity for $T_{end} = 60s$ and $\ell = 4$

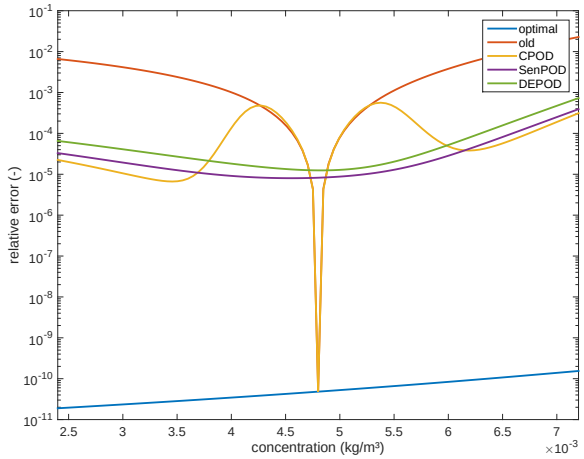


Fig. 4. Relative error for the variation of initial concentration for $T_{end} = 60s$ and $\ell = 4$

For an increased time horizon of 3600 time steps, we have 6739200 unknowns. We use $\ell = 7$ POD modes to approximate the solution.

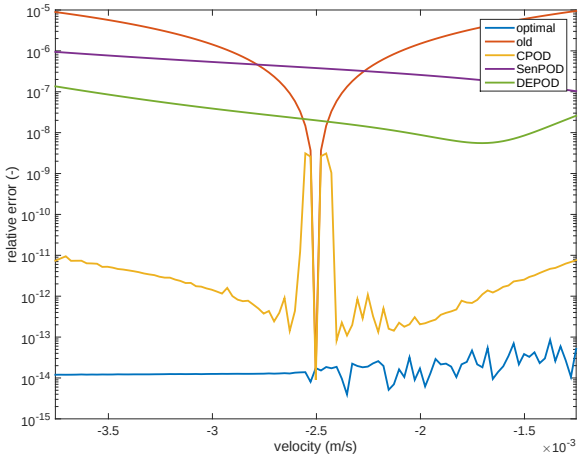


Fig. 5. Relative error for the variation of velocity for $T_{end} = 3600s$ and $\ell = 7$

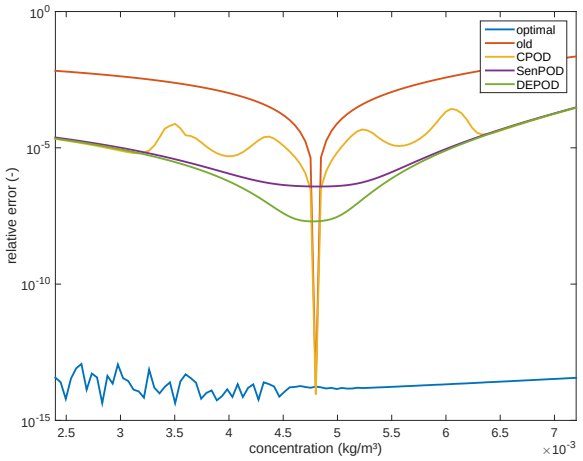


Fig. 6. Relative error for the variation of initial concentration for $T_{end} = 3600s$ and $\ell = 7$

The results can be found in the figures 5 and 6. Looking first to the figure of varying velocity with fixed concentration (fig. 5) one can see again the behaviour that the CPOD approximation follows the old approximation in a neighborhood of the sampling point. Then from a certain point one, the sensitivity information comes more into play and the approximation is directly getting better. Then the error slightly increasing in the directions of the interval boundaries. Taking a look at DEPOD and SenPOD the difference between the two methods is much larger. In this case, weighting seems to be much more important. Both now yield an overall better approximation than in the case before. If we look at the variation of concentration (fig. 6) one can see that outside a neighborhood around the sampling point the three methods yield the same error. In a neighborhood around the sampling point again CPOD follows the old approximation. This shows that in this case for a larger number of modes, the sensitivity does not contribute very much to the basis. This can be seen by the fact that three different weightings yield the same result over a larger parameter range.

In table I the relative computation times in comparison to the time for solving the full-dimensional system are shown. One can see that the offline cost are up to 6.45 times higher, but therefore in the online phase they are decreased up to ten times. Looking at the result for CPOD one can see that for the short time horizon the online computational cost are higher than the full-dimensional

TABLE I
COMPUTATION TIMES RELATIVE TO FULL-DIMENSIONAL SOLVE

Method	short		long	
	Offline	Online	Offline	Online
old	2.25	0.35	1.00	0.13
CPOD	4.00	1.60	2.83	0.54
DEPOD	4.77	0.33	3.56	0.13
SenPOD	4.73	0.33	3.05	0.13
CPOD nearest	6.45	0.35	3.63	0.13
CPOD bilinear	6.45	0.35	3.63	0.13

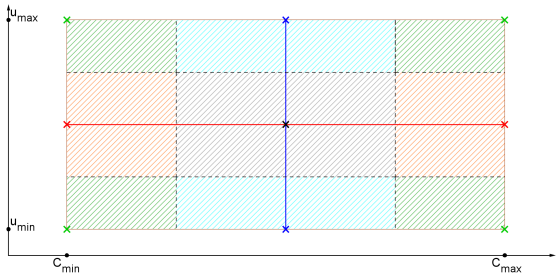


Fig. 7. Nearest neighbor region for each of the sampling points

solve, therefore it is not worthwhile to use the technique. For the long time horizon it is better and just needs half of the time of a full-dimensional solve. Note that we used CPOD in a standard way and the computations are not optimized (e.g. using QR factorization). DEPOD and SenPOD behave similar. For the short time horizon the online cost is just one third, for the long time horizon just 13%. Therefore the online computational costs are directly comparable with the cost of the old approximation.

A. Combination with interpolation

Since CPOD has shown to be a powerful tool for the approximation we want to overcome the drawback that for every parameter configuration a new snapshot matrix has to be build up with recomputation of the modes, which increases the computational cost a lot. Therefore we compute the CPOD basis just in a few sample points (see crosses in fig. 2) and then use interpolation between this states. The most natural technique might be nearest neighbor interpolation, where the corner point with the closest Euclidian distance is chosen (see fig. 7). A further idea to improve this is to use a more involved interpolation technique. In our case we choose for the 2d domain bilinear interpolation. Therefore we track, in which of the four sub-rectangles the actual point is located and use the four corner points as interpolation nodes (see fig. 2).

Looking first at the computational costs in table I one can see that using interpolation increases the offline time (since the mapping to the Grassmann manifold is necessary), but decrease the online cost a lot. They are now comparable with the cost of the other methods and improve around 5 times. Comparing to the solution of CPOD, nearest neighbor interpolation (see figures 8 to 11), it follows around the sampling the old approximation curve, since in this region both methods coincide. On the other part of the domain it follows the curve of CPOD and is sometimes even better. For bilinear interpolation, the results differ a lot more. For varying velocity (fig. 8 and 10) the results stay below the bound of the old approximation and yield in the neighborhood of the sampling point an improved result in comparison to nearest neighbor interpolation. On the rest of the domain it is slightly worse. For varying concentration the results are in contrary

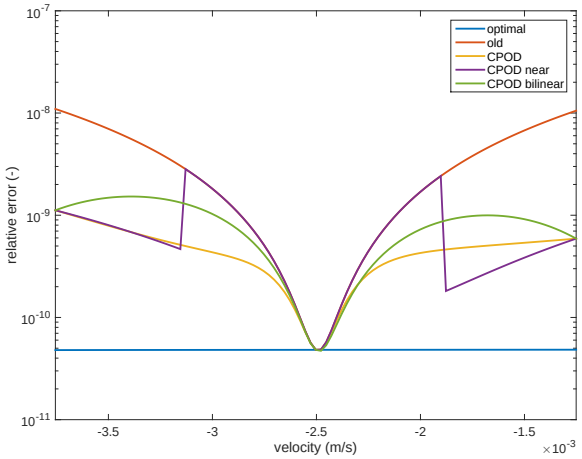


Fig. 8. Relative error for the variation of velocity using interpolation for $T_{end} = 60s$ and $\ell = 4$

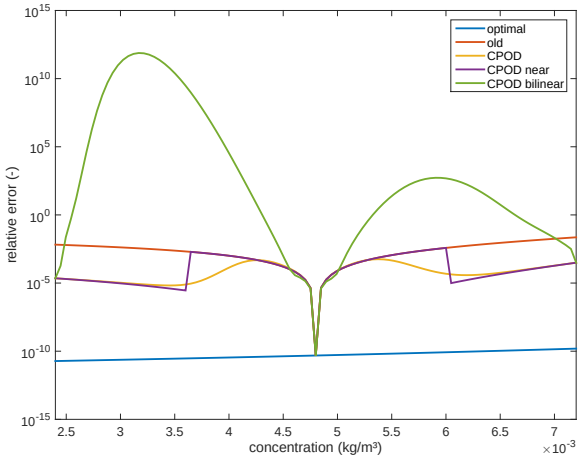


Fig. 9. Relative error for the variation of initial concentration using interpolation for $T_{end} = 60s$ and $\ell = 4$

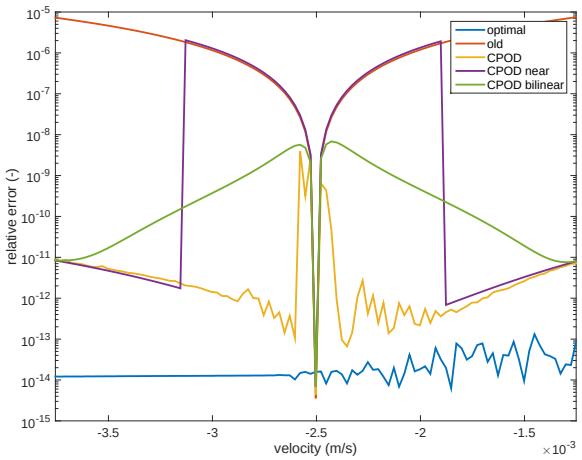


Fig. 10. Relative error for the variation of velocity using interpolation for $T_{end} = 3600s$ and $\ell = 7$

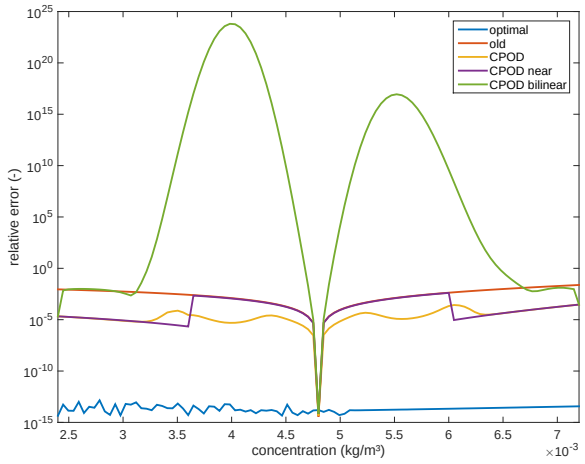


Fig. 11. Relative error for the variation of initial concentration using interpolation for $T_{end} = 3600s$ and $\ell = 7$

very bad (fig. 9 and 11). There the interpolation completely fails and yields an large error. The reason might be that with extrapolated data from just one point the interpolation is not meaningful. Thus bilinear interpolation can in our case not ensure to work well.

VI. Conclusions

We considered a model for combined cake and depth filtration, which is subject to parameter variation. To solve this problem over a whole range of parameters efficiently, we applied sensitivity analysis and model reduction. To get the sensitivities wrt a given parameter always an additional PDE solve is necessary, but the resulting PDE is always linear. For the use of proper orthogonal decomposition some precomputed solution trajectories are necessary. Thus there is a primary amount of computational cost necessary to compute the reduced basis. This amount is measured in the so-called offline phase. In the online phase the computed basis can be reused or is subject to slight adaptions. Therefore a speed-up is achieved, but an additional approximation error is introduced.

We used two techniques, where the sensitivity information is used in the snapshot matrix. Other techniques are using the sensitivity information on the reduced level (see e.g. [9]). CPOD uses sensitivity information with a Taylor expansion approach. Therefore the snapshots are weighted for every parameter configuration in a different way. For a 1d test case, this leads to good approximation results, but the computation costs are large (with standard implementation). Using DEPOD and its variant SenPOD the online computational cost are on the same level as for the old approximation, but for a good approximation a higher number of POD modes is needed. To overcome the drawback of CPOD we introduced interpolation. We tested two techniques, which reduce the online cost to the level of DEPOD and SenPOD, namely nearest neighbor and bilinear interpolation. Nearest neighbor interpolation leads to good approximations, whereas bilinear interpolation fails in some cases. Therefore nearest neighbor interpolation helps to improve CPOD for our case and the model equations can be solved for a multi-query context efficiently. The order of improvement is also depending on the simulation time horizon.

References

[1] *Free and moving boundary problems*. Oxford university press inc., 1984.

[2] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80(2):199–220, 2004.

[3] David Amsallem and Charbel Farhat. Interpolation method for adapting reduced-order models and application to aeroelasticity. *AIAA journal*, 46(7):1803–1813, 2008.

[4] Kevin Carlberg and Charbel Farhat. A compact proper orthogonal decomposition basis for optimization-oriented reduced-order models. *AIAA Paper*, 5964:10–12, 2008.

[5] Kevin Carlberg and Charbel Farhat. A low-cost, goal-oriented 'compact proper orthogonal decomposition' basis for model reduction of static systems. *International Journal for Numerical Methods in Engineering*, 86(3):381–402, 2011.

[6] Faruk Civan and Maurice L Rasmussen. Analytical models for porous media impairment by particles in rectilinear and radial flows. *Handbook of Porous Media, second edition, K. Vafai Ed., Taylor and Francis, New York*, pages 485–542, 2005.

[7] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[8] Bernard Haasdonk and Mario Ohlberger. Efficient reduced models and a posteriori error estimation for parametrized dynamical systems by offline/online decomposition. *Mathematical and Computer Modelling of Dynamical Systems*, 17(2):145–161, 2011.

[9] Alexander Hay, Jeffrey T Borggaard, and Dominique Pelletier. Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition. *Journal of Fluid Mechanics*, 629:41–72, 2009.

[10] JP Herzig, DM Leclerc, and P Le Goff. Flow of suspensions through porous media – application to deep filtration. *Industrial & Engineering Chemistry*, 62(5):8–35, 1970.

[11] Oleg Iliev, Ralf Kirsch, and Zahra Lakdawala. Exact solutions and parameter identification for a class of depth filtration processes.

[12] Oleg Iliev, Ralf Kirsch, Zahra Lakdawala, Stefan Rief, and Konrad Steiner. Modellierung und Simulation von Filtrationsprozessen. In *Mathematik im Fraunhofer-Institut*, pages 167–238. Springer, 2015.

[13] Oleg Iliev, Ralf Kirsch, and Sebastian Osterroth. Cake filtration simulation for poly-dispersed spherical particles. In *Proceedings of the FILTECH 2015 Conference*, 2015.

[14] Christopher Jarvis. Sensitivity based proper orthogonal decomposition for nonlinear parameter dependent systems. In *American Control Conference (ACC), 2014*, pages 135–140. IEEE, 2014.

[15] Michael Kuhn and Heiko Briesen. Dynamic modeling of filter-aid filtration including surface- and depth-filtration effects. *Chemical Engineering & Technology*, pages n/a–n/a, 2015.

[16] Karl Kunisch and Stefan Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische mathematik*, 90(1):117–148, 2001.

[17] S Kutluay, AR Bahadir, and A Özdeş. The numerical solution of one-phase classical Stefan problem. *Journal of computational and applied mathematics*, 81(1):135–144, 1997.

[18] René Pinnau. Model reduction via proper orthogonal decomposition. In *Model Order Reduction: Theory, Research Aspects and Applications*, pages 95–109. Springer, 2008.

[19] A Schmidt, Andreas Potschka, S Korkel, and Hans Georg Bock. Derivative-extended pod reduced-order modeling for parameter estimation. *SIAM Journal on Scientific Computing*, 35(6):A2696–A2717, 2013.

[20] PMA Slaats, J De Jongh, and AAHJ Sauren. Model reduction tools for nonlinear structural dynamics. *Computers & structures*, 54(6):1155–1171, 1995.

[21] Chi Tien. *Introduction to cake filtration: analyses, experiments and applications*. Elsevier, 2006.

[22] Chi Tien, Renbi Bai, and BV Ramarao. Analysis of cake growth in cake filtration: Effect of fine particle retention. *AIChE journal*, 43(1):33–44, 1997.

[23] Stefan Volkwein. Model reduction using proper orthogonal decomposition. *Lecture Notes, Institute of Mathematics and Scientific Computing, University of Graz. see http://www. uni-graz. at/imawww/volkwein/POD. pdf*, 2011.

[24] Amir Zamani and Brij Maini. Flow of dispersed particles through porous media – deep bed filtration. *Journal of Petroleum Science and Engineering*, 69(1):71–88, 2009.

[25] R Zimmermann. A locally parametrized reduced-order model for the linear frequency domain approach to time-accurate computational fluid dynamics. *SIAM Journal on Scientific Computing*, 36(3):B508–B537, 2014.

Energy dissipation in laser-excited semiconductors

Anika Rämer* and Bärbel Rethfeld
Department of Physics and Research Center OPTIMAS, TU Kaiserslautern, Germany
*araemer@physik.uni-kl.de

***Abstract*—The excitation of semiconductors with an ultrashort laser pulse induces an electron-hole plasma. We study the excitation dynamics of the electron and hole system on ultrashort timescales as well as the subsequent energy and particle transport and relaxation processes. We report damage threshold calculations that are in good agreement with experimentally determined values over a wide range of pulse durations.**

I. Introduction

For many years now, the interaction of laser pulses with solid matter has been subject of both experimental as well as theoretical research. Particularly ultrashort laser pulses causing minimal collateral damage to the surrounding material are of high interest for applications in medical surgery [1], micromachining [2] and nanostructuring [3].

For the theoretical description of laser-excited solids, models range from nonequilibrium descriptions over temperature-based models to atomistic calculations. Nonequilibrium models, like the Boltzmann equation [4] and kinetic Monte Carlo simulations [5] are able to describe the strong nonequilibrium conditions during the excitation process itself in detail but are computationally highly demanding. To address the latter point, there are some simplified nonequilibrium models like the multiple rate equation [6] but still, these models can only be solved with high computational effort. Temperature-based models like the density-dependent two-temperature model [7], [8], on the other hand, are computationally far less demanding than nonequilibrium descriptions and thus able to describe the relaxation on longer timescales as well as transport effects. They are based on the assumption that the strong nonequilibrium in the laser-excited solid will last only on a very short timescale of approximately one hundred femtoseconds and that its effect on longer-lasting processes, like e. g. lattice heating is small. Finally, atomistic calculations, like molecular dynamics simulations [9], can describe the actual structural changes of the material during the heating and melting process but are however again computationally far more demanding than the temperature-based modeling described here.

In the following we will first introduce the density-dependent two-temperature model in more detail. Then we will demonstrate that it is quite able to make predictions on damage thresholds in silicon that are in very good agreement with experimental data.

II. Modeling Excitation and Relaxation

At room temperature, semiconductors possess a valence band that is almost filled up with electrons, while only few free electrons are present in the conduction band. Thus, they have a low electrical conductivity that may, however, increase with increasing temperature.

When a semiconductor is irradiated with an ultrashort laser pulse, the light is at first mainly absorbed by valence band electrons which are thereby excited across the band gap into the conduction band. Each excited electron leaves behind a hole in the valence band. Thus, the laser pulse induces an electron-hole plasma within the material. The free carriers contained in this plasma may then absorb further photons gaining a higher kinetic energy, a process called free carrier absorption (FCA). Once carriers have gained a sufficiently high kinetic energy, they may excite additional valence band electrons to the conduction band via a process called collisional or impact ionization. In addition, electrons in the conduction band may recombine with holes in the valence band transferring their kinetic energy to the remaining free carriers. Moreover, carriers loose energy to the atoms of the solid, i. e. the lattice system.

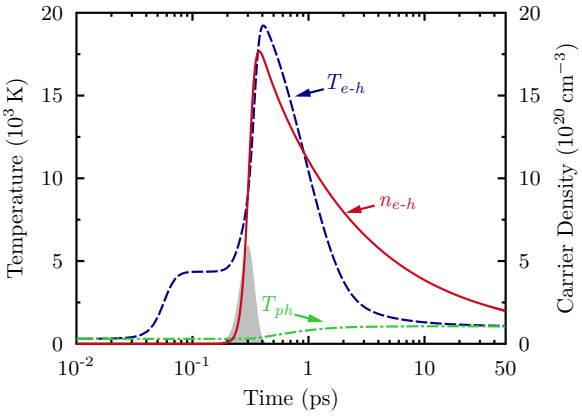


Figure 1. Temperature and density evolution during and after excitation calculated with the nTTM. The laser intensity is sketched in arbitrary units as a gray area. Figure adapted from Ref. [8].

All these processes as well as carrier and heat diffusion within the semiconductor can be modeled using the density-dependent two-temperature model (nTTM) [7], [8]. It is governed by three equations,

$$\frac{\partial u_{e-h}}{\partial t} = -\vec{\nabla}\cdot\vec{w}_{e-h} - g(T_{e-h} - T_{ph}) + S(\vec{r},t)\,, \quad (1)$$

$$\frac{\partial u_{ph}}{\partial t} = -\vec{\nabla}\cdot\vec{w}_{ph} + g(T_{e-h} - T_{ph})\,, \quad (2)$$

$$\frac{\partial n_{e-h}}{\partial t} = -\vec{\nabla}\cdot\vec{j} + G_{e-h} + R_{e-h}\,, \quad (3)$$

describing the evolution of the energy densities of the electron-hole pairs u_{e-h} and the lattice system u_{ph} as well as the free carrier density n_{e-h}. The first term on the right-hand side of each equation describes transport. Here, carrier transport happens via ambipolar diffusion, thus, electrons and holes move together. Additionally, we have to take into account that the carrier heat current density $\vec{w}_{e-h}$ is connected with the particle current density $\vec{j}$ via Peltier and Seebeck effect. The second term in the equations for the energy densities describes the energy exchange between the carriers and the lattice system. Note that the coupling parameter g is not a constant but may depend on carrier density as well as on temperature. The energy increase in the carrier system due to laser absorption is given by $S(\vec{r},t)$. Finally, G_{e-h} describes the

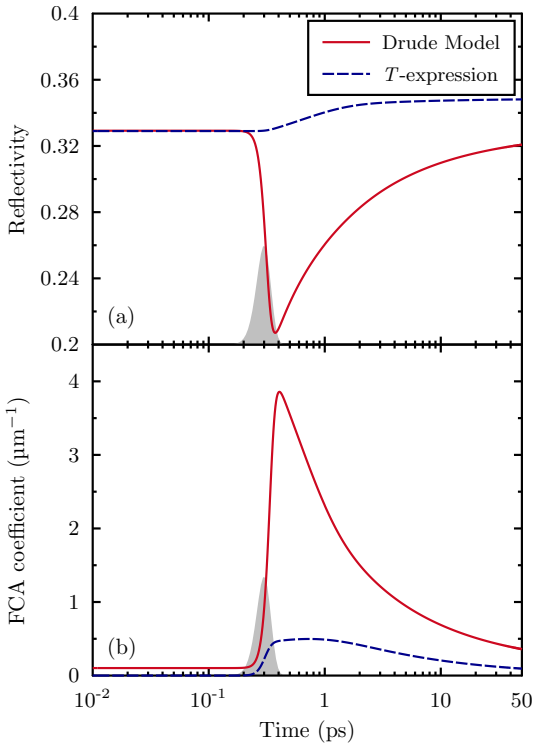


Figure 2. Reflectivity (a) and absorption coefficient (b) for calculations using either the Drude model or the T-expression to describe the optical parameters. Figure adapted from Ref. [8].

generation of free carriers due to laser absorption and impact ionization, while R_{e-h} is the recombination of electron-hole pairs.

The given set of equations is solved numerically in one dimension, the depth of the material, using a finite difference scheme. All model parameters are taken from Ref. [8]. The results of such an nTTM calculation at the surface of silicon irradiated with a 100 fs (full width half maximum) Gaussian laser pulse with a fluence of 130 $\mathrm{mJ/cm^2}$ at 800 nm wavelength is shown in Fig. 1. Here, it is visible that the carrier density strongly increases during pulse duration. Afterwards, recombination starts to dominate causing the carrier density to decrease. Similarly, the carrier temperature at the surface increases during the duration of the laser pulse and later decreases due to lattice coupling and transport. The lattice temperature rises due to the energy transfer from the carriers until both systems share a common temperature, which may take several tens of picoseconds.

III. Optical Parameters

In the last section, we have shown that during the excitation of semiconductors, the free carrier density may increase by several orders of magnitude. This will lead to severe changes in the optical parameters, namely the reflectivity and the free carrier absorption (FCA) coefficient [8]. However, when modeling semiconductors, the reflectivity is often assumed to depend solely on lattice temperature (T-expression) [7], [10]. Instead, we apply a Drude model with a transient Drude frequency accounting for the highly transient free carrier density.

Figure 2 (a) shows the reflectivity evolution during the irradiation with the same pulse as described above. We immediately notice major differences: while the T-expression causes an increase in reflectivity, the reflectivity calculated with the Drude model decreases, a behavior also observed in experiments [11]. In addition, the FCA coefficient calculated with the Drude model is much larger than the one calculated with the T-expression as is shown in Fig. 2 (b). As can be seen in Fig. 3, such strong differences in optical parameters also yield huge differences in the calculated density and temperature curves and may strongly influence any predictions made with the model.

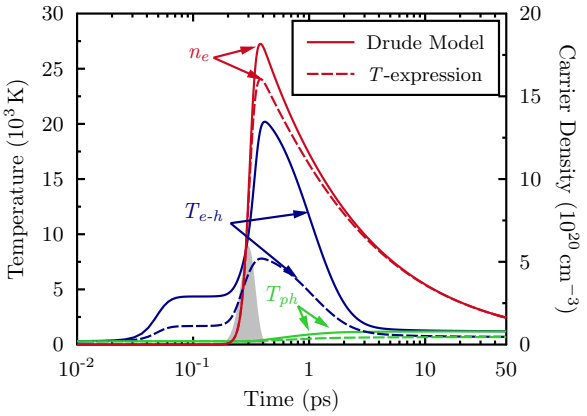


Figure 3. Temperature and density evolution calculated using either the Drude model or the T-expression to describe the optical parameters. Figure adapted from Ref. [8].

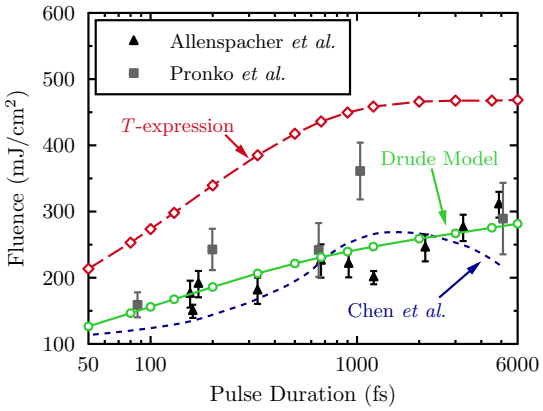


Figure 4. Damage thresholds estimated using either the T-expression or the Drude model to calculate the transient optical parameters in comparison to experimental data obtained by Allenspacher *et al.* [12] and Pronko *et al.* [13] as well as thresholds simulated by Chen *et al.* [10]. Figure adapted from Ref. [8].

IV. Estimation of Damage Thresholds

One major application of the nTTM is the prediction of damage thresholds. Here, we will demonstrate that these predictions strongly depend on the quality of the description of the optical parameters investigated above.

We estimate damage thresholds in silicon assuming that a material is damaged once its melting temperature is exceeded. The results are shown in Fig. 4 in comparison to experimentally determined values as well as a calculation by Chen *et al.* [10] that involved a fitting procedure. It is quite obvious that the calculations using the T-expression highly overestimate the damage thresholds, while the thresholds estimated using a Drude model are in very good agreement with experimental

data even over a wide range of pulse durations. The agreement is even better than the one achieved by Chen *et al.* [10] as it reproduces the trend in the experimentally determined values also for long pulse durations.

V. Conclusion

We have demonstrated that when modeling the laser excitation and relaxation of semiconductors, the description of the optical parameters is crucial. Once a Drude model with a transient Drude frequency accounting for the highly transient carrier density is implemented for the description of the optical parameters, the nTTM is able to predict damage thresholds that are in very good agreement with experimental data even over a wide range of pulse durations. Thus, the nTTM has been proven to be a flexible and easy to handle model that nevertheless yields highly satisfying results when accounting for density-depending optical parameters.

Acknowledgment

The authors thank O. Osmani, K. Sokolowski-Tinten and V. P. Lipp for helpful suggestions. Financial support by the Deutsche Forschungsgemeinschaft through the Heisenberg program (Grant No. RE 1141/15) and the State of Rhineland-Palatinate through the State Research Center Optics and Material Sciences in Kaiserslautern is gratefully acknowledged.

References

[1] A. Vogel and V. Venugopalan, *Chemical Reviews*, vol. 103, no. 2, pp. 577–644, 2003.

[2] M. Li, K. Mori, M. Ishizuka, X. Liu, Y. Sugimoto, N. Ikeda, and K. Asakawa, *Applied Physics Letters*, vol. 83, no. 2, pp. 216–218, 2003.

[3] A. Rodriguez, M. Echeverría, M. Ellman, N. Perez, Y. K. Verevkin, C. S. Peng, T. Berthou, Z. Wang, I. Ayerdi, J. Savall, and S. M. Olaizola, *Microelectronic Engineering*, vol. 86, no. 4—6, pp. 937 – 940, 2009.

[4] N. Brouwer and B. Rethfeld, *J. Opt. Soc. Am. B*, vol. 31, no. 11, pp. C28–C35, 2014.

[5] N. Medvedev, U. Zastrau, E. Förster, D. O. Gericke, and B. Rethfeld, *Phys. Rev. Lett.*, vol. 107, no. 16, p. 165003, 2011.

[6] N. Medvedev and B. Rethfeld, *Journal of Applied Physics*, vol. 108, no. 10, p. 103112, 2010.

[7] H. M. van Driel, *Phys. Rev. B*, vol. 35, no. 15, pp. 8166–8176, 1987.

[8] A. Rämer, O. Osmani, and B. Rethfeld, *Journal of Applied Physics*, vol. 116, no. 5, p. 053508, 2014.

[9] V. P. Lipp, B. Rethfeld, M. E. Garcia, and D. S. Ivanov, *Phys. Rev. B*, vol. 90, no. 24, p. 245306, 2014. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevB.90.245306

[10] J. K. Chen, D. Y. Tzou, and J. E. Beraun, *International Journal of Heat and Mass Transfer*, vol. 48, no. 3–4, pp. 501 – 509, 2005.

[11] K. Sokolowski-Tinten and D. von der Linde, *Phys. Rev. B*, vol. 61, no. 4, pp. 2643–2650, 2000.

[12] P. Allenspacher, B. Hüttner, and W. Riede, *Proc. SPIE*, vol. 4932, pp. 358–364, 2003.

[13] P. P. Pronko, P. A. VanRompay, C. Horvath, F. Loesel, T. Juhasz, X. Liu, and G. Mourou, *Phys. Rev. B*, vol. 58, no. 5, pp. 2387–2390, 1998.

Fatigue behavior of martensitic steels for low-pressure steam turbine blades in the very high cycle fatigue (VHCF) regime

Fabian Ritz
TU Kaiserslautern
Institute of Materials Science and Engineering (WKK)
Kaiserslautern, Germany
ritz@mv.uni-kl.de

Tilmann Beck
TU Kaiserslautern
Institute of Materials Science and Engineering (WKK)
Kaiserslautern, Germany
beck@mv.uni-kl.de

***Abstract*— In the present study, the fatigue behavior of X10CrNiMoV12-2-2 is investigated in the VHCF regime for two different notched sample geometries. The tests were performed up to 2 x 10^9 cycles using an ultrasonic fatigue testing machine developed at WKK at room temperature. For the notched samples with a stress concentration factor of 1.09 a transition from surface to subsurface cracks is observed at about 1 x 10^7 cycles. Crack initiation starts at oxide inclusions of the type (CaO)x-Al_2O_3. The maximum cycle number at which fatigue fracture occurred for these samples is about 4 x 10^7 cycles. For the second sample geometry with a notch factor of 2.42 no VHCF-failure could be observed until now and the maximum number of cycles where sample failure occurs is 1 x 10^6. These cracks initiate at surface defects resulting from sample preparation. The calculated fatigue strength until 2 x 10^9 load cycles is about 530 MPa (notch factor: 1.09) and approximately 245 MPa (notch factor: 2.42). The ratios between the notch factors and the fatigue limits until 2 x 10^9 are similar.**

Keywords*— *VHCF, martensitic steel, steam turbine

I. INTRODUCTION

Quantitative understanding of fatigue behavior of components in nearly all fields of mechanical engineering is a very important topic because fatigue failure under cyclic loading occurs at load amplitudes well below static yield strength. Fatigue tests are categorized into LCF- (low cycle fatigue, $N_f < 10^3$), HCF- (high cycle fatigue, $10^3 < N_f < 10^7$) and VHCF (very high cycle fatigue, $N_f > 10^7$) which are the focus of the present study. In the past, fatigue investigations of materials usually where performed until 10^7 load cycles. Meanwhile it is accepted that for many multi-phase materials like the X10CrNiMoV12-2-2 steel investigated in the present work, failure occurs even at stress amplitudes below the classic fatigue limit at cycle numbers far beyond 10^7. To avoid failure of safety critical components like the wheel axis failure of a ICE 3 in Cologne in 2008 (see Fig. 1), e.g, knowledge of fatigue behavior in the VHCF-regime is extremely important. For such investigations appropriate ultrasonic test equipment is necessary to realize effective test frequencies > 1000 Hz because using classic servo-hydraulic setups (5 Hz) one single fatigue test would take several years to reach 2 x 10^9 load cycles. With a total of 4 ultrasonics fatigue test rigs, the authors' institute is one of Germany's largest laboratories in this field.

Fig. 1. ICE wheel axis failure in Cologne 2008 (source: DDP)

For multi-phase materials as investigated in the present work, a transition of the crack initiation and –propagation from the surface (LCF- and HCF-regime) to the volume (VHCF) takes place at a certain number of cycles. Because many components reach the VHCF regime during their service life, the fatigue behavior at cycle numbers beyond 10^7 is of great relevance for reliable design.

The material investigated in this work is state of the art for last stages in the low pressure steam turbines of power plants. During their lifetime they undergo cyclic bending and torsional loads due to inhomogenities within the flow field with a frequency of about 2 kHz [1]. The typical service life of approx. 30 years clarifies that knowledge of the VHCF behavior of this material is necessary for a reliable component design with reasonably low safety coefficients. The fatigue load is superimposed by high mean stresses caused by centrifugal forces due to rotation at a frequency of 50 Hz. To analyze the fatigue behavior with a realistic load profile, it is necessary to use ultrasonic fatigue testing equipment with the possibility to superimpose high mean stresses. In the present

study the fatigue behavior of X10CrNiMoV12-2-2 steel is investigated for very high number of cycles under the influence of two different notch geometries. The samples are manufactured form the original components by wire spark eroding and subsequent turning.

II. Material and Testing Equipment

The chemical composition of the investigated martensitic steel X10CrNiMoV12-2-2 is summarized in Table I.

TABLE I. CHEMICAL COMPOSITION OF X10CRNIMOV12-2-2

element	C	Cr	Ni	Mo	Mn	V	Si	Cu
ma.-%	0.10	12.02	2.64	1.59	0.75	0.34	0.18	0.11

The most important alloying elements are chromium, nickel, molybdenum and carbon. The three-step heat treatment includes at first martensitic hardening (60 min at 1040 °C / cooling by compressed air) followed by a three hour tempering at 660 °C and stress relief annealing at 640 °C for four hours. The resulting tempered martensitic microstructure contains finely distributed carbides at former austenite grain boundaries as well as between the lamellar martensite laths. The resulting mechanical properties are a good compromise of high strength and ductility (Table II).

TABLE II. MECHANICAL PROPERTIES OF X10CRNIMOV12-2-2

mech. properties	yield strength	tensile strength	elongation at fracture	Young`s modulus	hardness
value	843 MPa	1001 MPa	17.7 %	213 GPa	321 HV

Opposed to earlier investigations [2] on the same material notched samples were used in the present study as shown in Fig. 2 and Fig. 3. The first sample geometry with $\alpha_k = 1.09$ results from the ratio of Murakami`s weighting factors for surface- and subsurface cracks [3].

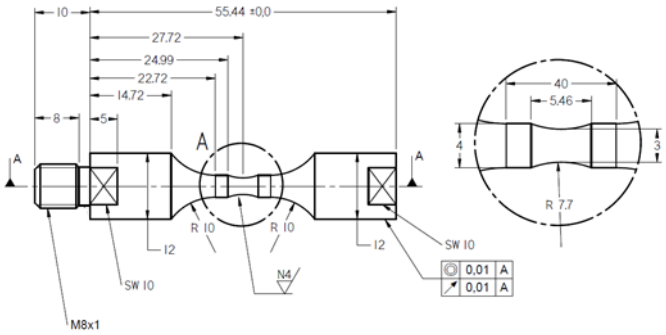


Fig. 2. Notched sample geometry ($\alpha_k = 1.09$ / R = -1)

The second sample geometry ($\alpha_k = 2.42$) represents a component relevant notch factor. Both sample types are manufactured for testing at a load ratio of R = -1 and for tests at a constant mean load (R ≠ -1).

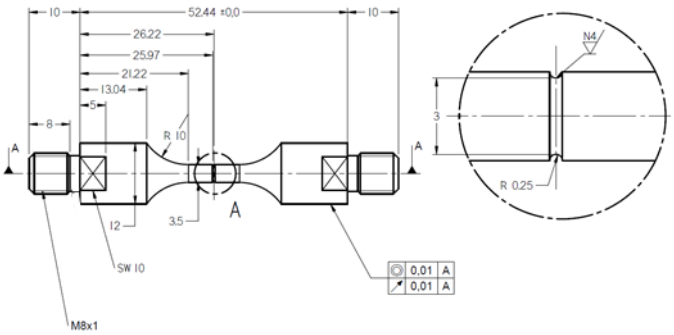


Fig. 3. Notched sample geometry ($\alpha_k = 2.42$ / R ≠ -1)

The investigations were performed using an ultrasonic fatigue testing system developed and built at the authors' institute (Fig. 4) with an operation frequency of 20 kHz. To limit the maximum sample temperature due to self-heating by microplastic deformation, it is necessary to run the tests in pulse-pause mode. In this case the duration of an ultrasonic pulse is 500 ms followed by a cooling time of 3500 ms. The effective testing frequency is about 2.5 kHz. This testing frequency further decreases in case of tests with measurement of electrical resistance because this data is collected in extended pauses.

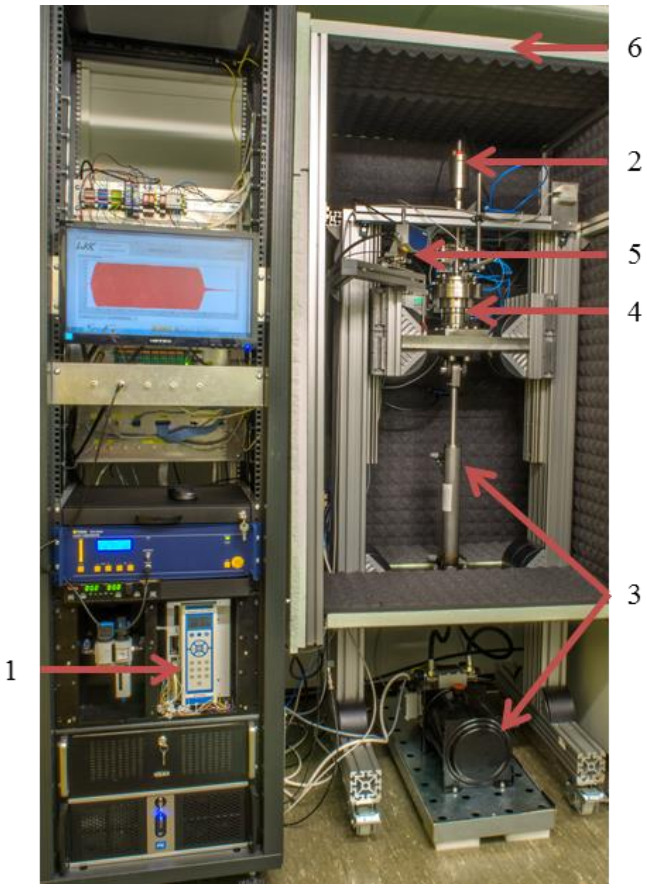


Fig. 4. Ultrasonic fatigue testing system (20 kHz)

The ultrasonic generator (1) transfers the supply voltage to a high-frequent (approximately 20 kHz) AC voltage. In the converter (2) this AC voltage is converted into a mechanical oscillation of the same frequency by using the inverse piezo electrical effect. A booster is attached to the converter which stabilizes and increases the oscillation. The sample itself is an active part of the system and vibrates in resonance. Therefore each of the described parts is designed with a length of λ/2 to realize the necessary stationary wave. In case of tests with a load ratio of $R \neq -1$ the mean load is applied by a servo-hydraulic system (3) placed below the load frame. To control the mean stress, a load cell (4) is attached below the second booster. In addition to the data collected from the ultrasonic generator, i.e. power, amplitude and frequency, the displacement of the sample can be measured by a laser vibrometer (5), cyclic creep is observed by a separate distance measuring sensor and temperature in the gauge length is obtained by an infrared system. All data is collected and processed continuously while the test is running. Opposed to that, electrical resistivity is measured in certain intervals during the pauses between two ultrasonic pulses. The whole testing facility is integrated in an insulating box to reduce thermal influences and to minimize noise in the laboratory (6).

III. Results and Discussion

The knowledge of the exact oscillation is very important to avoid unintended additional loads like bending or torsional stresses. By measuring the oscillation by 3D-laser-scanning-vibrometery, the actual sample deformation can be observed in a wide range of frequencies. The results show torsional and bending modes at 18.2 kHz and 21.4 kHz for tests with R = -1. Both frequencies are not in the range of the ultrasonic generator. For 20.1 kHz a uniaxial load in longitudinal direction without other stress components was found. In Fig. 5 the displacement in four different states of the oscillation is plotted. The first picture shows the state of minimum displacement.

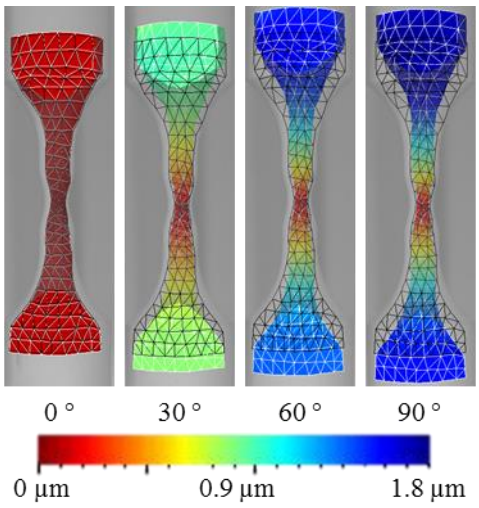


Fig. 5. 3D-Laser-vibrometer measurement (R = -1)

Increasing displacements and therefore increasing mechanical stresses are documented in the following three pictures. For tests with constant mean load ($R \neq -1$) other oscillation forms than uniaxial loading are also out of the generator`s frequency range For 20.2 kHz in Fig. 6 the oscillation at $R \neq -1$ is shown. No bending or torsional load occurs, proving good quality of load train alignment. The correlation between sample displacement and stress amplitude was determined by elastic FEM simulations using Abaqus CAE. A direct relationship between the measured displacement at the bottom of the sample and the control voltage of the ultrasonic generator was obtained in calibration tests and used to control the fatigue tests.

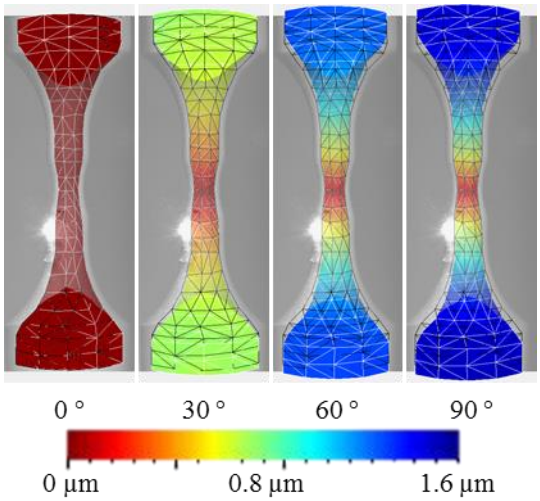


Fig. 6. 3D-Laser-vibrometer measurement under constant mean load

Fig. 7 shows the S-N-curves for the fatigue tests on both notched geometries at R = -1. For the lower notch factor a transition from surface to volume initiated cracks can be seen at about 1×10^7 cycles. In comparison to that the second sample geometry ($\alpha_k = 2.42$) has a distinctly higher stress concentration at the notch surface that subsurface cracks are completely prevented in the experiments ran so far. The maximum number of cycles at which failure occurred is about 4×10^7 for $\alpha_k = 1.09$ and approx. 1×10^6 for $\alpha_k = 2.42$ which means that up to now no real VHCF cracks were observed at the higher notched samples for R = -1. Initiation of both, surface and subsurface cracks take place at oxide inclusions of the type $(CaO)_x$-Al_2O_3 for $\alpha_k = 1.09$. The cracks generally initiate at surface defects in case of the sharper notch. The calculated fatigue strength for $\alpha_k = 1.09$ is about 530 MPa and approx. 245 MPa for $\alpha_k = 2.42$ for. The ratio between these two calculated fatigue strengths is close to the ratio of the respective notch factors. For the cracks initiated by inclusions Murakami`s suggestion that non-metallic inclusions can be treated as small cracks with the same cross-section area (√area-concept) [4]are confirmed by the present results. Additional to the results obtained at notched samples Fig. 7 show earlier results for samples with cylindrical gauge length of the same material [2].

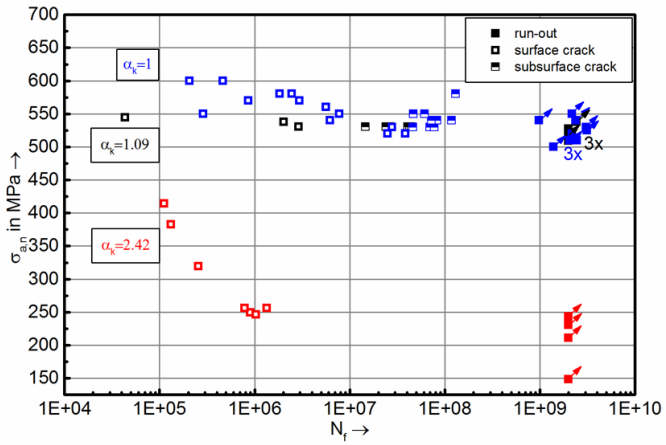


Fig. 7. S-N Curves of X10CrNIMoV12-2-2 for three different sample geometries

At very high cycle numbers (> 10^7) there is no significant difference between the notched and cylindrical samples, i.e. the stress concentration in the notch has no influence on the crack initiation. However, as soon as cracks are initiated at the surface, which is the case for $N_f < 10^7$, a slight difference in nominal stress amplitude of approx. 10 % is observed which correlates well with the notch factor and the associated surface stress increase.

With that, the present study confirms Mughrabi`s widely accepted idea of crack initiation transition from surface to volume as soon as the VHCF regime is reached for multi-phase materials [5]. For application in the design of low pressure steam turbine blades it should be noted that even though there is a change in fracture mechanism from surface to volume initiated cracks at a certain number of load cycles the slope of the S-N curve (Fig. 7) remains extremely flat throughout the VHCF-regime, resulting in the possibility to exclude VHCF failures by choosing appropriate safety factors even to data in the cycle range up to 10^7.

Current and future investigations focus on the influence of mean stress on the fatigue behavior of the same material with identical notch geometries as used for push-pull loading. In a first step, tests are performed at sample with notch factor 1.09 with a load ratio of R = 0.5. Cyclic creep which is observed at cylindrical samples in case of high load ratios is also expected to occur for α_k = 1.09, however – due to inhomogenous stress distribution across the sample – to a lower extent.

IV. Summary

3D-laser-scanning-vibrometry confirmed that perfectly uniaxial stress state without any significant superimposed loads could be reached with the applied ultrasonic fatigue testing system both for R = -1 and R ≠ -1. Other oscillation modes of the sample are sufficiently far out of the frequency range of the ultrasonic generator used in the fatigue testing system.

The fatigue tests for R = -1 show for the slightly notched sample a change from surface to subsurface crack initiation at about 1 x 10^7 cycles. For both, surface and the subsurface cracks oxide inclusions of the type $(CaO)_x$-Al_2O_3 initiate the cracks. In case of volume initiation the resulting S-N curve exactly resembles the values obtained at smooth samples in earlier studies. In case of surface crack initiation, the sustained stress amplitude is reduced by the notch factor. For α_k = 2.42 the maximum number of cycles at which fracture occurred is approx. 1 x 10^6 which means that no real VHCF-fracture could be observed so far. The crack in this case generally initiates at surface defects resulting from manufacturing of the samples. The calculated fatigue limit for α_k = 2.42 is 245 MPa and 530 MPa for α_k = 1.09, respectively. As for cyclindrical samples (α_k = 1.0) a very flat slope of the S-N curve in VHCF was found.

Acknowledgment

The authors gratefully acknowledge funding by the German Research Foundation (DFG) in the framework of the priority program 1466.

References

[1] Richter, C.H., Structural design of modern steam turbine blades using ADINA (TM). Computers & Structures, 2003. 81(8-11): p. 919-927.

[2] Kovacs, S., T. Beck, and L. Singheiser, Influence of mean stresses on fatigue life and damage of a turbine blade steel in the VHCF-regime. International Journal of Fatigue, 2012

[3] Murakami, Y., S. Kodama, and S. Konuma, Quantitative evaluation of effects of non-metallic inclusions on fatigue strength of high strength steels. I: Basic fatigue mechanism and evaluation of correlation between the fatigue fracture stress and the size and location of non-metallic inclusions. International Journal of Fatigue, 1989. 11(5): p. 291-298.

[4] Murakami, Y., Effects of Nonmetallic Inclusions on Fatigue Strength, in Metal Fatigue2002, Elsevier Science Ltd: Oxford. p. 75-127.

[5] Mughrabi, H., *Fatigue, an everlasting materials problem - still en vogue.* Procedia Engineering, 2010. 2(1): p. 3-26.

Terahertz Compressive Imaging Spectroscopy

Klemens M. Schmitt, Zinching Dang and Marco Rahm
University of Kaiserslautern, Department of Electrical and Computer Engineering and Research Center OPTIMAS,
Erwin-Schroedinger-Strasse, 67663 Kaiserslautern, Germany

***Abstract*—We present a fast terahertz (THz) imaging spectroscope using compressive sensing methods. In our measurement scheme, random THz projections of a scene are recorded by a single-pixel detector. We implemented the projection masks by spatially modulating the probing THz beam with patterned copper masks on PCB boards. We analyzed several image reconstruction methods for retrieving the image from the spatial projections. In the spectral range from $0.14\,\mathrm{THz}$ to $0.35\,\mathrm{THz}$ we reconstructed the image with an ℓ_1-minimization and a total-variation-minimization with different compression ratios.**

I. Introduction

In the electromagnetic spectrum, terahertz (THz) waves are located between the microwave and infrared spectrum at frequencies between about 0.1 THz and 10 THz. Like microwaves, THz radiation can penetrate non-metallic materials, yet offers a higher spatial resolution, which makes it suitable for many non-invasive imaging applications [1]. Since, unlike X-Rays, THz radiation does not have ionizing effects, it is potentially useful as a screening source in biological or medical imaging applications. Nowadays, THz technology has advanced to the maturity level of commercial applications as e.g. in non-destructive material inspection or security scanners at airports. In this context, the THz technology faces manifold challenges. In terms of spectroscopic applications, THz waves with a wide spectral bandwidth are required to resolve the spectral fingerprints of various substances. For imaging applications, the scene must be recorded with high spatial resolution while the data acquisition time must be held at a minimum.

As THz detectors are expensive, THz imaging mostly relies on raster scan methods using single-pixel detectors at the cost of two major drawbacks. First, the data acquisition rate of raster scan methods is inherently low and second, either the THz beam or the object must be constantly moved in the focal plane which is undesirable for sensitive setups and fragile samples.

In this respect, compressive imaging methods offer a potential solution for eliminating these drawbacks [2–5]. While it still reverts to the use of a single-pixel detector, the object does not need to be scanned. Furthermore, the compressive approach can reduce the number of required measurements by built-in compression and thus the data acquisition time.

II. THz Time-Domain Spectroscopy

Terahertz time domain spectroscopy (THz-TDS) is a technique to measure the time trace of the amplitude and phase of the complex THz electric field. The frequency spectrum of the THz field can be readily obtained by Fourier transform of the time trace of the electric field.

In our experimental setup (see Fig. 2), THz radiation is generated by focusing femtosecond laser pulses between the electrodes of a photoconductive switch (PCS). A PCS consists of two opposed, voltage biased electrodes on a semiconductor substrate. The laser pulses excite free carriers in the substrate, which are accelerated in the electric field between the electrodes and therefore radiate in a wide spectrum in the range from 0.1 THz to about 2.0 THz. The time duration of the emitted THz pulses lies in the range of 0.5 ps.

The THz pulses are measured in the time domain by an electro-optic sampling (EOS) process. In this method, the THz pulses travel through an electro-optic crystal and induce a birefringence by the Pockels effect. The birefringence is linearly dependent on the instantaneous electric THz field amplitude and can be probed by an optical sampling pulse. For this purpose, a certain fraction of the THz generating femtosecond laser pulses is coherently deviated to the electro-optic crystal. Due to the THz field induced birefringence, the polarization of the optical probe pulse is rotated. Hereby, the angle of the polarization rotation is proportional to the instantaneous THz field and the rotation direction gives the phase. Because the femtosecond optical sampling pulse is much shorter than the THz pulse, the latter one can be sampled by temporally delaying the probe pulse with respect to the THz pulse.

There are different ways to obtain the time delay. One possibility is to use a mechanical delay stage, which shifts the detection pulse relatively to the generation pulse. Another possibility is asynchronous optical sampling (ASOPS), where two femtosecond lasers with slightly different repetition rates are used. At some time point, the two lasers emit a pulse at the same time. The next pulse has a time delay according to the difference in repetition rate. The pulse after the next pulse has twice the time delay. After many cycles, the time delay equals the repetition period, i. e. both pulses are emitted synchronously again, and the whole terahertz pulse is sampled.

III. Compressive Imaging

As seen in Fig. 1, THz single-pixel imaging uses a spatial terahertz modulator (STM) to modulate a beam that illuminates the object to be imaged. After scattering from the object, the THz beam is focused onto the single pixel detector. Using different modulation patterns and recording one single-pixel data point for each pattern, it is possible to reconstruct the object. We can describe the N pixels of the one– or two-dimensional object as a vector x. Each measurement corresponds to a scalar product of the illumination pattern Φ_i

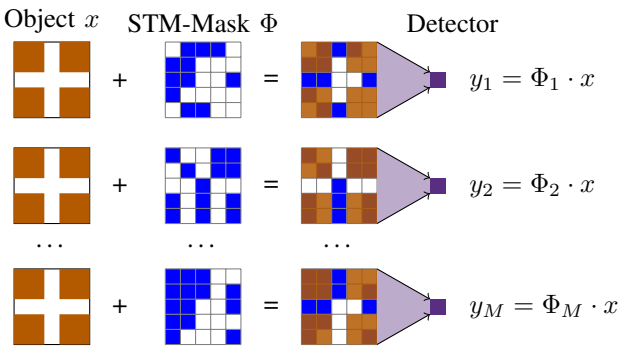


Fig. 1. Schematic drawing of the compressive imaging process. The spatial terahertz modulator (STM) modulates the THz beam which illuminates the object. The scattered THz beam is focused onto a single-pixel detector. This example uses an inverse cross as an object and a 5x5 pixel random mask which is superposed. The measurement process can be described by a scalar multiplication of the object vector x and the mask vector Φ_i.

and the object x. Combining the M patterns Φ_i into a single $M \times N$ measurement matrix Φ the imaging process can be described as:

$$y = \Phi x \tag{1}$$

where y is a vector composed of the measured values y_i. Knowing the measurement matrix Φ and the measurement vector y, the original image can be reconstructed by simple matrix inversion.

Compressive imaging samples the image irregularly [6–8], thus the Nyquist-Shannon sampling theory does not apply in this case. Most images contain less information than their number of pixels. By transforming the image, it is possible to obtain a sparse image where many coefficients are zero or at least negligibly small. These transforms are also used in algorithms for image compression. For example, the JPEG standard uses a discrete cosine transform (DCT). The transform can be described by matrix multiplication $x_s = \Psi x$ where x_s is the sparse representation of the image. For a good image approximation, only the largest coefficients are important. Consequently, acquiring $M < N$ measurements is sufficient in order to obtain the most important information, increasing the number of measurements add more details, which are not always necessarily needed.

The mentioned compression algorithms are applied after conventional image acquisition. Compressive imaging integrates the transform matrix Ψ into the measurement matrix Φ. By this means, the compression takes place during the measurement process and the number of individual measurements can be reduced.

As a measure for compression, it is convenient to define the compression ratio as the number of pixels of the (uncompressed) image divided by the number of measurements.

As it is not clear which transform yields an optimal compression, a binary random measurement matrix Φ is used. Most random matrices automatically obey the condition that the individual masks are linearly independent. Reconstructing the sparse image with compressive imaging is not a simple matrix inversion because the system is under-determined. For this reason, we need to define an additional sparsity condition. Sparsity means that the image has few non-zero pixels. This can be described mathematically by the ℓ_0 norm which counts the number of non-zero elements of a vector. Thus, one additional condition can be the ℓ_0-minimization

$$\tilde{x} = \min_x \|x\|_0 \quad \text{subject to} \quad y = \Phi x \tag{2}$$

This problem cannot be computed in a trivial manner, yet it is mathematically equivalent to the well-known ℓ_1-minimization [9]:

$$\tilde{x} = \min_x \|x\|_1 \quad \text{subject to} \quad y = \Phi x \tag{3}$$

where $\tilde{x}$ is an approximate reconstruction of the image. This is a well known problem that can be solved by linear programming.

As experimental data is subject to noise, the condition $y = \Phi x$ cannot be exactly fulfilled. Introducing a noise level ε results in:

$$\tilde{x} = \min_x \|x\|_1 \quad \text{subject to} \quad \|\Phi x - y\|_2 \leq \varepsilon \tag{4}$$

ℓ_1-minimization can be used to directly reconstruct a sparse image. If the image is not sparse in the position space, it can also be combined with a 2D-DCT or some other transform in order to reconstruct an image which is sparse in its corresponding spatial frequency space.

Another possible optimization condition is the total-variation-minimization (TV-minimization) [10]:

$$\tilde{x} = \min_x TV(x) \quad \text{subject to} \quad \|\Phi x - y\|_2 \leq \varepsilon \tag{5}$$

with the total variation $TV(x) = \sum_{ij} \|\nabla_{ij}\|_2$, where ∇_{ij} is the discrete gradient of the image.

Depending on the image, different conditions yield different results. ℓ_1-minimization is good for images which are mostly black (i.e no intensity) and have few regions with information. TV-minimization yields reconstructions with large regions with the same amplitude. That way, it is suitable for reconstructing images with areas of the same amplitude which is not necessarily black.

The additional conditions do not only allow the compression of the images, but make it also possible to fully sample or even oversample the image. The additional condition can then be used to reduce the noise of the image.

With compressive imaging, it is also possible to start with a small number of measurements, which yield a poor reconstruction quality and then to iteratively add more measurements until the image quality is satisfactory.

The reconstruction of an image takes some computational time, but on recent computers, it is negligible compared to the time required for THz imaging. Thus, reducing the number of required THz measurements greatly decreases the time needed for acquiring a complete THz image.

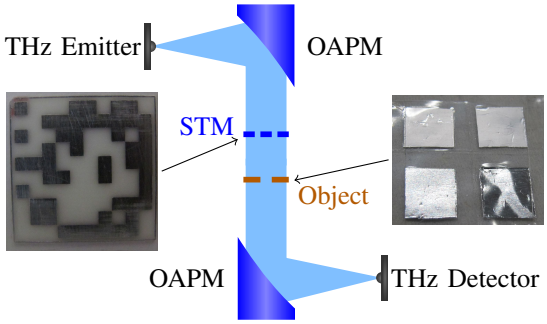


Fig. 2. Simplified schematic of the experimental setup. The emitted THz pulses are collimated and focused onto the THz detector by two off-axis parabolic mirrors (OAPM). A photography of the spatial terahertz modulator (STM), which is made of standard $0.5\,\mathrm{mm}$ FR-4 PCB, is shown on the left. The object made of aluminum foil is depicted on the right. Both are $23\,\mathrm{mm} \times 23\,\mathrm{mm}$ in size and are placed between the two OAPMs in the parallel THz beam.

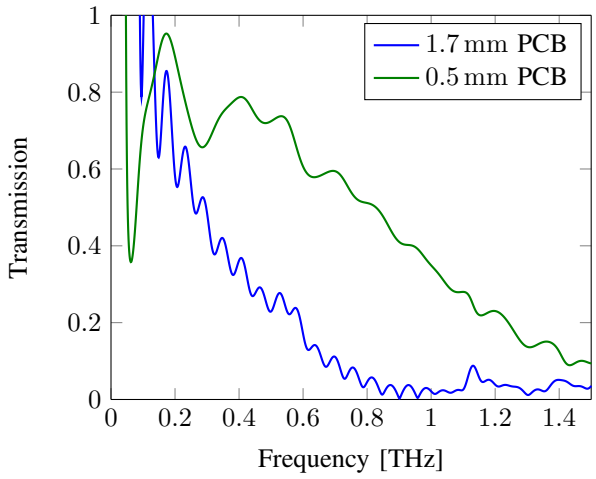


Fig. 3. Transmission spectra of FR-4 PCBs with thicknesses of $1.7\,\mathrm{mm}$ (blue) and $0.5\,\mathrm{mm}$ (green). Both PCBs show low pass characteristics. The 1/e-bandwidth of the $0.5\,\mathrm{mm}$ PCB is higher than the bandwidth of the $1.7\,\mathrm{mm}$ thick board.

IV. Experimental Setup

Our experimental setup consists of a THz-TDS system with EOS detection and ASOPS. A simplified schematic setup is depicted in Fig. 2. The THz pulses are generated by a PCS and collimated by an off-axis parabolic mirror (OAPM). After traveling through the spatial THz modulator and being scattered from the object, the THz radiation is collected by a second OAPM and focused onto the THz detector.

The parallel beam has a Gaussian beam profile which means that its intensity at the center is higher than at the edges. This inhomogeneous illumination of the object has to be taken into account in the proper reconstruction of the image.

The object is shown on the right in Fig. 2 and displays a $23\,\mathrm{mm} \times 23\,\mathrm{mm}$ inverted metallic cross. The metallic layer consists of aluminum foil and reflects the THz beam. As a result, the THz radiation images the shape of the cross.

A. Printed Circuit Board Spatial Terahertz Modulator

For the THz beam modulation, we use $23\,\mathrm{mm} \times 23\,\mathrm{mm}$ and $0.5\,\mathrm{mm}$ thick standard FR-4 printed circuit boards (PCB). The galvanized copper layer on the PCB-STM blocks the THz beam at defined spatial locations and represents the modulation patterns.

An ideal STM would block all THz radiation at the non-transparent pixels and be fully transparent at the location of the transparent pixels. In other words, the modulation depth of an ideal STM would be 100%, which would be true if the metal structure was in air. Yet, in our experiment, the STM consists of individual printed circuit boards, one for each spatial pattern. This means that the substrate, in our case FR-4, adds to absorption and thus diminishes the transmission of the transparent pixels. In order to optimize the transmission properties, we hence tested the THz transmission through commercially available FR-4 PCBs of different thicknesses. Figure 3 shows the spectral transmission through PCBs with thicknesses of $0.5\,\mathrm{mm}$ and $1.7\,\mathrm{mm}$, respectively. Both PCBs show low pass characteristics, whereby the 1/e transmission bandwidth of the $0.5\,\mathrm{mm}$ thick PCB is $0.98\,\mathrm{THz}$, whereas the 1/e bandwidth of the $1.7\,\mathrm{mm}$ thick PCB is significantly lower at a value of $0.36\,\mathrm{THz}$. Therefore, it is obvious that $0.5\,\mathrm{mm}$ PCB is a more suitable substrate thickness for the implementation of an STM.

The size of the pixels is an important factor for the imaging process. On the one hand, it is desirable to have a high number of small pixels to have a good spatial resolution of the image. On the other hand, the spatial diffraction of THz radiation at the transmitting STM pixels is a limiting factor with respect to spatial resolution. Due to the long wavelength of $1\,\mathrm{mm}$ at $0.3\,\mathrm{THz}$, we minimized the effect of diffraction by using pixel sizes of $2.3\,\mathrm{mm} \times 2.3\,\mathrm{mm}$.

V. Results

We perform the compressive imaging reconstruction by means of the algorithms introduced in Section III, i.e. by ℓ_1-minimization and TV-minimization. We use the Matlab library "l1magic" [6] to solve the minimization problems according to Equations 4 and 5. This library provides various algorithms for solving minimization problems, including ℓ_1- and TV-minimization. We investigate and compare the dependence of the reconstruction quality of both algorithms on the noise level ε of the recorded images. Additionally, the THz beam profile has to be considered in the reconstruction, since it has a Gaussian intensity shape.

We acquire the scene projection of each measurement, i. e. of each STM mask superposed with the object, in the time domain by use of the single-pixel detector. In order to gain spectroscopic information from this data, we Fourier-transform the time signal. The spectral transmission maximum of the $0.5\,\mathrm{mm}$ thick PCB, shown in Fig. 3 in the previous section, lies in the frequency range between $0.14\,\mathrm{THz}$ and $0.35\,\mathrm{THz}$. For this reason, it is instructive to use THz STMs

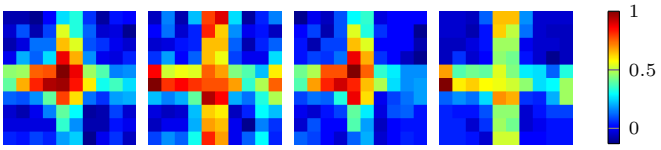


Fig. 4. Reconstructed image of the inverted cross without compression. All four images are based on the same data and use 100 measurements, therefore no compression was intended. (a) and (b) show the reconstructed image using ℓ_1-minimization. Since in (a) the THz beam profile was not considered, the intensity decreases at the edges, whereas in (b) the entire cross has a similar intensity distribution. In analogy to that, (c) and (d) show the reconstructed images using TV-minimization. In (c), the THz beam intensity profile was not taken into account, thus the reconstructed image reveals an intensity maximum in the center of the image as an artifact, while in (d) the cross has a homogeneous intensity distribution due to consideration of the THz beam intensity profile.

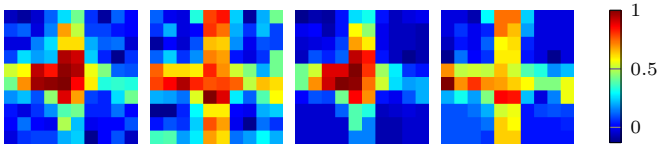


Fig. 5. Reconstructed image of the inverted cross with compression. The images are based on the same data as in Fig. 4, but use only 80 measurements. Thus, the images have a compression ratio of $100 : 80$. The images are noisier, however they show the same basic features as in Fig. 5. (a) and (b) show the reconstructed images using ℓ_1-minimization, while in (c) and (d) TV-minimization is used. In (a) and (c), the intensity maximum is in the center of the images because THz beam intensity profile remained unconsidered. In (b) and (d), the shape of the cross is more prominent and the image shows a higher contrast towards the borders due to consideration of the Gaussian beam intensity profile.

in this frequency band for investigating the aforementioned image reconstruction algorithms with respect to reconstruction quality of the reconstructed images.

In Eq. 1, each component y_i corresponds to the measured amplitude or intensity of the scene projection through mask number i. Since we obtain spectral amplitude and phase information by use of our time-domain spectroscope, we only consider the spectral amplitude and average it over a range from $0.14\,\mathrm{THz}$ to $0.35\,\mathrm{THz}$ to obtain the average y_i component. By repeating this procedure for each individual mask, we eventually determine the full vector y in Eq. 1.

For ℓ_1-minimization, we solve Eq. 4 with the acquired data vector y, the known STM matrix Φ, which includes the 2D DCT transform matrix, and a user-defined noise ε. Because the 2D DCT transform matrix is already included, the solution vector $\tilde{x}$ needs to be re-transformed.

Figure 4(a) shows a reconstructed image of the cross object. The ℓ_1 reconstruction uses 100 measurement points and thus does not imply compression. The cross can be recognized, but at the edges of the image the intensity of the reconstruction decreases, which is due to the Gaussian beam profile. If we take the beam intensity profile into account in Fig. 4(b), the center pixels have the same intensity as the arms of the cross.

Figures 4(c) and 4(d) show a reconstruction from the same measurement data, but this time by use of a TV-minimization algorithm. It can be seen that the dark areas are more homogeneous, as this algorithm tries to reconstruct large areas with the same intensity. Also in this case, a reconstruction, that takes the THz beam intensity profile into account, has a better reconstruction accuracy at the edges of the arms of the cross, as seen in Fig. 4(d).

In Fig. 5, the number of measurements is reduced to 80, whereas the reconstruction algorithms are the same as in Fig. 4. In this case, the reconstruction algorithms provide the additional conditions to resolve the equation system and the imaging process can be considered as being compressive. Figures 5(a) and 5(b) display the reconstructed images using ℓ_1-minimization, while the reconstruction of the images in Figs. 5(c) and 5(d) was based on a TV-minimization

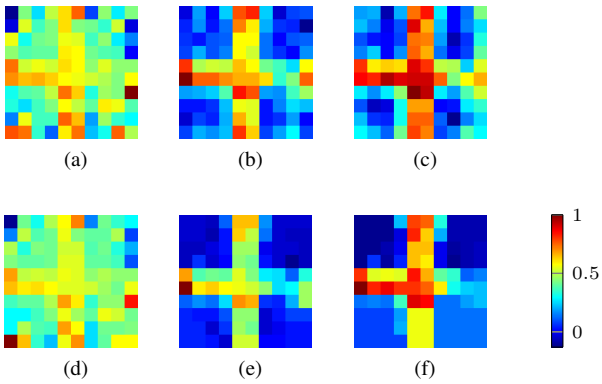


Fig. 6. Impact of different noise levels ε on the image reconstruction quality. Figures (a) to (c) show the reconstruction using the ℓ_1-minimization with noise levels of $\varepsilon = 0.2$, $\varepsilon = 0.5$, and $\varepsilon = 0.8$, respectively. In Fig. (a), the noise level is too low, resulting in a failed reconstruction. A higher noise level, shown in Figs. (b) and (c), implies better reconstructions. The reconstructed images by use of TV-minimization with the same noise levels are shown in Figs. (d) to (f). An insufficient noise level also results in a failed reconstruction, as shown in Fig. (d), whereas an exceeding noise level can cause artifacts, as shown in Fig. (f).

algorithm. Furthermore, the reconstruction processes for the images in Figs. 5(a) and 5(c) did not take the Gaussian beam intensity profile into account, whereas it was considered in the reconstruction of the images in Figs. 5(b) and 5(d). In comparison with the reconstructed images in Fig. 4, the compressed images in Fig. 5 are more noisy, yet the cross object can still be identified. Further reduction of the number of measurements yields additional noise until the object disappears in the noisy background. For real-world applications, a trade-off between acquisition time and noise in the image must be found.

For the presented measurement, the achieved compression ratio is $100 : 80 = 1.25$. This is due to the fact that the image has only 10×10 pixels, which is a very small amount compared to the several megapixels of an average-quality image taken with nowadays cameras in the optical frequency range. With a larger number of pixels, a higher compression ratio can be obtained because higher resolution usually means more pixels for only slightly more information.

One critical factor for the reconstruction algorithm is the noise level parameter ε that limits the reconstruction accuracy in Eqs. 4 and 5. In order to reconstruct the image correctly, the ε has to be adjusted to the actual noise level of the experimental measurement. In this context, Fig. 6 shows the influence of the noise level ε on the reconstructed image. All images in Fig. 6 are reconstructed using 100 measurements and taking the THz beam intensity profile into account. For the ℓ_1-minimization shown in Figs. 6 (a) to (c), the reconstruction basically fails for $\varepsilon = 0.2$ [see Fig. 6(a)]. In contrast, as can be seen from Figs. 6(b) and 6(c), the images, that are reconstructed with a noise level of $\varepsilon = 0.5$ and $\varepsilon = 0.8$, yield sufficiently high quality for the cross to be clearly visible.

As we consider the TV-minimization, the reconstruction also fails for a noise level parameter of $\varepsilon = 0.2$, as shown in Fig. 6(d). In Fig. 6(e), the reconstruction for $\varepsilon = 0.5$ provides a good image of the cross. At a noise level of $\varepsilon = 0.8$, some artifacts in the dark areas get stronger and the recognizability of the cross is reduced, as can be seen in Fig. 6(f). The study above clearly demonstrates that an optimal adaption of the noise level ε to the actual noise level of the measurement is of crucial importance for the reconstruction quality of the image.

VI. Conclusion

We presented a proof-of-principle for the applicability of compressive imaging techniques in an imaging terahertz time-domain spectroscope. Using a THz beam modulated by a spatial terahertz modulator and a single-pixel detector, it is possible to image an object with a series of measurements using different modulation masks. Both, ℓ_1-minimization and total-variation-minimization are suitable for reconstructing the images. Depending on the measurement scenario, the reconstruction algorithms can be used for noise reduction in the images or to reduce the number of performed measurements that are necessary to reconstruct the image. For real-world applications, an optimal trade-off between image quality and compression-ratio/measurement-time has to be found.

In future work, it is desirable to increase the number of image pixels to achieve a better compression ratio. It is also a goal to implement dynamically reconfigurable spatial terahertz modulators to speed up the measurement process and eventually install a feedback loop to selectively enhance the image quality in specific regions of interest.

References

[1] D.M. Mittleman, M. Gupta, R. Neelamani, R.G. Baraniuk, J.V. Rudd, and M. Koch. Recent advances in terahertz imaging. *Applied Physics B*, 68(6):1085–1094, 1999.

[2] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman. A single-pixel terahertz imaging system based on compressed sensing. *Applied Physics Letters*, 93(12):121105, 2008.

[3] D. Shrekenhamer, C. M. Watts, and W. J. Padilla. Terahertz single pixel imaging with an optically controlled dynamic spatial light modulator. *Opt. Express*, 21(10):12507–12518, May 2013.

[4] C. M. Watts, D. Shrekenhamer, J. Montoya, G. Lipworth, J. Hunt, T. Sleasman, S. Krishna, D. R. Smith, and W. J. Padilla. Terahertz compressive imaging with metamaterial spatial light modulators. *Nat Photon*, 8:605–609, 2014.

[5] W. L. Chan, M. L. Moravec, R. G. Baraniuk, and Daniel M. Mittleman. Terahertz imaging with compressed sensing and phase retrieval. *Optics Letters*, 33(9):974–976, 2008.

[6] E. Candes and J. Romberg. ℓ_1-MAGIC : Recovery of Sparse Signals via Convex Programming. 2005.

[7] Candes, E. J. Compressive sampling. In *Proceedings oh the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, pages 1433–1452, 2006.

[8] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[9] D. L. Donoho. For most large underdetermined systems of equations, the minimal 1-norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*, 59, 2006.

[10] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

Influence of the size effect on the flexural tensile strength of filigree UHPC components

Milan Schultz-Cornelius
Institute of Concrete Structures and Structural Engineering
Technische Universität Kaiserslautern
Kaiserslautern, Germany
milan.schultz-cornelius@bauing.uni-kl.de

Matthias Pahn
Institute of Concrete Structures and Structural Engineering
Technische Universität Kaiserslautern
Kaiserslautern, Germany
matthias.pahn@bauing.uni-kl.de

***Abstract* - The flexural tensile strength of structural elements made of concrete depends on the tensile strength and on the depth of the structural element. An increasing depth leads to a decreasing flexural tensile strength approaching to the tensile strength. This characteristic is called size effect. If the flexural tensile strength is tested on standard specimens with a defined depth this effect cannot be considered. Although size effect is already theoretically investigated and mentioned within the norms there is no knowledge on the behaviour of thin structural elements with a depth of a few centimetres.**

Within this paper flexural tensile strength specimens with a depth of 15 – 50 mm are presented and evaluated. The main ambition of these experiments is to gain more knowledge on the influences of size effect on flexural tensile strength to enable the calculation of the real flexural tensile strength of a structural element with a low depth on the basis of the results of a standard test of flexural tensile strength.

Keywords - (ultra) high-performance concrete, size effect, flexural tensile strength

I. INTRODUCTION

A. General

New materials such as ultra high-performance concrete allow to reduce the thickness of structural elements. One field of application is as façade elements. With UHPC they can be built thinner without reducing the size of the elements.

They can be used as curtain-wall facings or as facings for sandwich walls. To enable a wise and material-saving calculation of such filigree façades it is necessary to check present approaches. Façades are mostly punctually supported slabs with a multiaxial load bearing behaviour.

They are not load bearing elements because the main function is to protect the thermal insulation from external influences. The main design loads are wind suction, wind pressure and temperature.

Especially for unreinforced artificial stone façades the most important mechanical parameter is the flexural tensile strength which is the base of the calculation. This material parameter is tested according to different regulations. The most common one is DIN 196-1 [8] because every Testing Institute for Building Materials has standardized formwork and testing machines for this regulation. This prevents an influence of a varying geometry of the specimens or the experimental setup. The measurements of the specimens allow an easy handling and a low material consumption. After the flexural tensile strength tests the compressive strength of the both resulting halves can be tested.

Because of the steady measurements of the specimens (40 x 40 x 160 mm) the size effect of structural elements with varying depths is not taken into account. The below presented investigations enable the calculation of the real flexural tensile strength of a structural element with a certain depth on the basis of test results according to [8].

B. Size effect

The influence of the depth of a structural element on the mechanical characteristics is called size effect. The geometry affecting the flexural tensile strength is an acquainted phenomenon in material testing.

In 1939 Weibull [1] developed a statistical concept on the basis of the idea of the weakest chain link. This concept assumes statistically dispersed imperfections inside of the material volume. An increasing material volume leads to an increasing amount of imperfections and consequently to a decreasing strength.

A further concept was developed by Carpinteri and Chiaia [2]. They assume a maximum inhomogeneity of the concrete which is affected by the maximum grain size. If the maximum grain size is small with respect to the depth of the structural element the concrete approximately behaves like a homogeneous material. But if the depth of the structural element is low the specimen conducts inhomogeneously and the influence of the maximum grain size is high.

The fictitious crack model [3] from DAfStb Heft 444 explains the size effect of flexural tensile strength tests. As soon as the fracture strain is reached during a flexural tensile strength test a fictitious crack opens. During increasing deformation further stresses are transferred through the crack as shown in Figure 1.

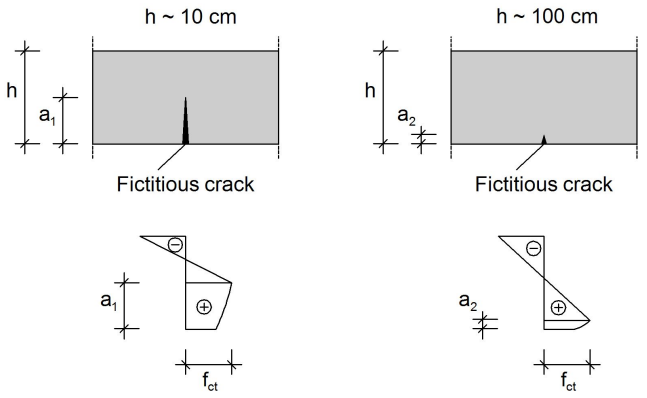


Fig. 1. Stress distribution of bended concrete specimens with various depths at breaking load [3]

C. *Aim of the investigations*

The results of previous investigations are adopted in Model Code 1990 [4] and in Eurocode 1992-1-1 [5]. The listed equations enable to calculate the flexural tensile strength on the basis of the tensile strength. Those equations apply to structural elements with $h \geq 50$ mm.

This paper presents experiments with specimens made of UHPC with depths of 15 mm $\leq h \leq$ 50 mm which are not covered by the stated concepts. This is particularly interesting to filigree façades which can be realized very thinly using UHPC and which are, as exposed concrete elements, measured for the absence of cracks according to serviceability limit state. They are used as curtain walls or facing panels. Further applications are stairways, lost formwork or roof cladding. The consideration of the size effect enables to capture the initial crack formation in a realistic way.

II. MATERIAL

A. *Concrete composition*

The appropriated UHPC was produced with the binder premix Dyckerhoff NANODUR® Compound 5941. This premix consists of 59% cement NANODUR® CEM II/B-S 52,5 R and 41% quartz powder. The aggregates are dried sand 0/2 mm and basalt 2/5 mm. The precise composition is displayed in Table 1.

TABLE I.

Concrete composition		
Material	***Raw density***	***Weight***
NANODUR® Compound 5941	2.860 kg/m³	1.066,5 kg
Sand 0/2 mm	2.650 kg/m³	436,7 kg
Basalt 2/5 mm	3.060 kg/m³	893,8 kg
Plasticizer Glenium ACE 430	1.060 kg/m³	14,1 kg
Water	1.000 kg/m³	156,8 kg

B. *Fresh concrete properties*

The fresh concrete properties of the self-compacting concrete are tested according to [9], [10] and [11]. The results are shown in the following Table 2.

TABLE II.

Fresh concrete properties		
Property	***Unit***	***Value***
Slump flow measure after 30 seconds	Cm	55 / 55
Slump flow measure after 60 seconds	Cm	62 / 64
Air void content	%	1,20
Fresh concrete temperature	°C	23,50
Room temperature	°C	22,40

C. *Hardened concrete properties*

The following Table 3 illustrates the mean values of the compressive strength and the modulus of elasticity tested on three specimens in each case. At the beginning the specimens are stored in water for 28 days and afterwards dried for three days. The tests are performed on cubes according to DIN EN 12390-3 [6] and on cylinders according to DIN 1048-5 [7] after 31 days.

TABLE III.

Hardened concrete properties		
Property	***Unit***	***Value***
Compressive strength $f_{ck,cube}$	N/mm²	155,43
Modulus of elasticity E_m	N/mm²	42.992,22

III. EXPERIMENTAL INVESTIGATION

A. *Geometry of the specimens*

The specimens are realized in accordance to DIN EN 196-1 [8] with regard to the tests of the flexural tensile strength. Their measurements are 160 mm x 40 mm x h as shown in Figure 2.

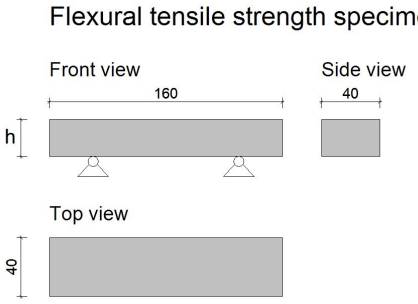


Fig. 2. Measurements of the flexural tensile strength specimens

B. Production of the specimens

The specimens are produced in the concrete laboratory of the University of Kaiserslautern. The formwork for the specimens consists of wood.

After the concrete is casted into the formwork the specimens are covered with foil. The formwork is stripped after one day and the specimens are stored in water at a temperature of 20°C for 28 days. Afterwards they are dried at a temperature of 20°C for 3 days up to the date of the tests.

C. Experimental programme

In total 70 flexural tensile strength tests are performed with varying depths. The experimental programme is presented in Table 4.

TABLE IV.

Experimental programme of flexural tensile strength tests		
Concrete	***Depth [mm]***	***Amount***
NANODUR®	50	10 each time
	45	
	40	
	35	
	30	
	25	
	20	
	15	

D. Experimental setup and procedure

The flexural tensile strength tests are performed with a special testing machine of the type ToniNorm 2060. The load speed is 50 N/s according to [8]. The support distance is 100 mm on standard specimens with the measurements of 160 mm x 40 mm x 40 mm. Consequently the ratio between the depth of the specimen and the support distance is 0.4. While varying the depth of the specimens this ratio is kept constant (Figure 3).

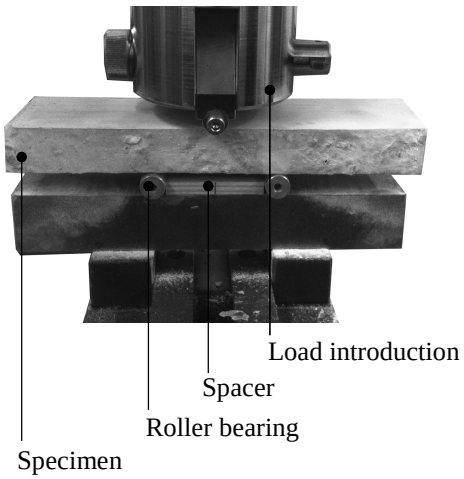


Fig. 3. Experimental setup

IV. Results and Discussion

A. Results

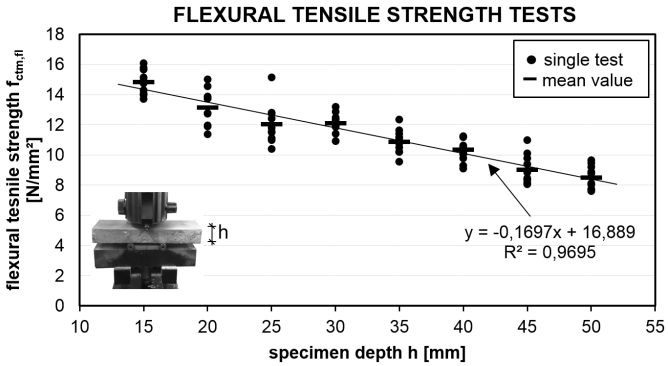


Fig. 4. Breaking stresses and mean values of the flexural tensile strength tests

Figure 4 shows the breaking stresses of the specimens. Every point represents a single test and every bar represents the mean value a test series. The diagram displays an increase of the breaking stresses with a decreasing depth clarified by the ascending trend line. The increase can be assumed as linear. It is noticeable that the dispersion of the test series is collectively small. The test series with h = 25 mm shows comparatively lower breaking stresses.

B. Discussion

The increase of the flexural tensile strength can be explained with the help of the fictitious crack model [3]. The dispersion of the test results can be caused by irregularities in the geometry, the modified experimental setup or the distribution of the aggregation in the cross section.

The low breaking stresses of the test series h = 25 mm indicate a systematic error. A plausible explanation is a slightly oversized spacer between the two roller bearings. This could be the result of an inaccurately produced spacer or a stuck pollution.

V. Conclusion

A. Calculation approach for thin structural elements with regard to the determination of the real flexural tensile strength on the basis of the results of standard tests

Currently no calculation approach exists that exclusively considers the influence of the depth on the flexural tensile strength. The increase of the flexural tensile strength, illustrated in Figure 4, shows that the real flexural tensile strength is up to 43 % higher than the results of the standard tests with a stated depth of h = 40 mm. In order to calculate the real flexural tensile strength for thin structural elements a size effect factor $k_{f,ct,fl}$ is established. This factor calculates the real flexural tensile strength on the basis of the test results of standard specimens with h = 40 mm according to DIN EN 196-1 [8].

The equation for the size effect factor is derived from the linear trend line in Figure 5. The following equation is proposed:

$$k_{f,ct,fl} = -0{,}017 \cdot h + 1{,}68 \quad (1)$$

The real flexural tensile strength can be calculated on the basis of the mean value of the test results of standard specimens according to DIN EN 196-1 [8] as follows:

$$f_{ct,fl,h} = k_{f,ct,fl} \cdot f_{ctm,fl,h=40mm} \quad (2)$$

The following Figure 5 shows the recalculation of the test results with the help of the recommended calculation approach.

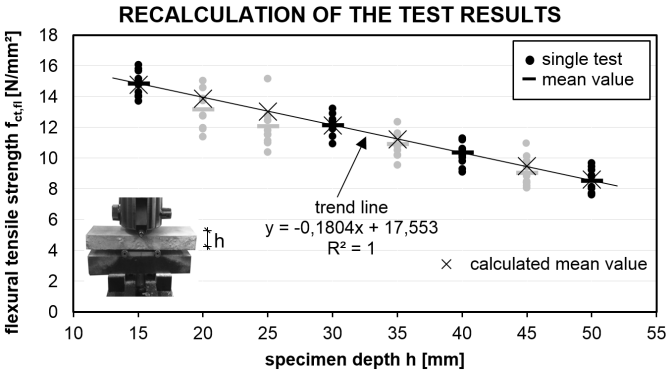


Fig. 5. Recalculation in order to examine the plausibility of the recommended calculation approach

The diagram illustrates that the trend line does not intersect every mean value. The grey test series were not involved in the establishment of the trend line. It is noticeable that the less dispersing black test series lie on a perfectly straight line. The coefficient of determination $R^2 = 1$ is optimal. The values of the flexural tensile strength which were calculated according to [3] and [4] on the basis of the mean values of h = 40 mm almost exactly lie on the trend line.

VI. Summary

A. General

This paper shows that the size effect also exists with respect to specimens with a depth of h < 50 mm. The test results expose approximately linearly increasing breaking stresses with decreasing depths.

A size effect factor is proposed in order to calculate the real flexural tensile strength depending on the depth of the structural element. With the help of this size effect factor the flexural tensile strength of a structural element with a certain depth can be determined on the basis of test results of standard specimens with a depth of h = 40 mm according to DIN EN 196-1 [8]. This enables a more economical calculation of concrete façades.

B. Outlook and next steps

Further investigations focus on the influence of size effect on the tensile strength f_{ctm} of UHPC. This material property is important to the calculation in case of centric tension inside of the façade caused by temperature effects. It is difficult to experimentally test the relative low tensile strength of concrete. Therefore the presented results are supposed to enable a conversion from flexural tensile strength to tensile strength.

References

[1] Weibull, W. (1939) A statistical theory of the strength of materials. The royal Swedish Institute for Engineering Research 151, p. 5-45, 1939

[2] Carpinteri, A, Chiaia, B. Multifractal scaling law for the fracture energy variation of concrete structures. Fracture Mechanics of Concrete Structures, ed. F.H. Wittmann, Aedificatio Publishers, p. 581-596, 1995

[3] Deutscher Ausschuss für Stahlbeton: Zum Zug- und Schubtragverhalten von Bauteilen aus hochfestem Beton – Heft 444, Berlin, 1994

[4] CEB-FIP Model Code 1990: Design Code, 1993

[5] Deutsches Institut für Normung: Bemessung und Konstruktion von Stahlbeton- und Spannbetontragwerken – Teil 1-1: Allgemeine Bemessungsregeln und Regeln für den Hochbau – DIN EN 1992-1-1, 2011

[6] Deutsches Institut für Normung: Prüfung von Festbeton – Teil 3: Druckfestigkeit von Probekörpern – EN 12390-3, 2009

[7] Deutsches Institut für Normung: Prüfverfahren für Beton; Festbeton, gesondert hergestellte Probekörper – DIN 1048-5, 1991

[8] Deutsches Institut für Normung: Prüfverfahren für Zement – Teil 1: Bestimmung der Festigkeit – DIN EN 196-1, 2005

[9] Deutsches Institut für Normung: Prüfung von Frischbeton – Teil 6: Frischbetonrohdichte – DIN EN 12350-6, 2011

[10] Deutsches Institut für Normung: Prüfung von Frischbeton – Teil 7: Luftgehalt - Druckverfahren – DIN EN 12350-7, 2009

[11] Deutsches Institut für Normung: Prüfung von Frischbeton – Teil 8: Selbstverdichtender Beton - Setzfließversuch – DIN EN 12350-8, 2010

Stochastic Micro-Macro Models for Acid Mediated Tumor Invasion

Stefanie Sonner [a]
[a] Felix-Klein-Zentrum für Mathematik
TU Kaiserslautern
67663 Kaiserslautern, Germany
Email: sonner@mathematik.uni-kl.de

Christina Surulescu [a]
Email: surulescu@mathematik.uni-kl.de

Peter Kloeden [a,b]
[b] School of Mathematics and Statistics
Huazhong University of Science & Technology
Wuhan 430074, China
Email: kloeden@mathematik.uni-kl.de

***Abstract*—Most tumors exhibit regions of increased acidity. The acidic environment promotes apoptosis of normal cells, while cancer cells can survive and proliferate, which facilitates invasion. We propose and analyze multiscale models for acid-mediated tumor invasion accounting for stochastic effects on the subcellular level.**

I. Introduction

Tumor spreading and migration through the surrounding tissue is a crucial stage in cancer development and a highly complex phenomenon, involving processes on several different space and time scales. Mathematical modeling can provide a framework for the interpretation of experimentally assessed knowledge about the various biochemical events influencing the evolution of the neoplastic tissue. During the last decade *multiscale approaches* linking two or more biological scales have proved to be well suited for integrating subcellular and single cell information in order to allow predictions on the tumor level and to uncover cross-scale mechanisms. Multiscale settings raise challenges with respect to their mathematical and numerical analysis, as they typically involve differential equations of several types that are strongly coupled and highly nonlinear.

A. Multiscale modeling approaches

A crucial stage in cancer development is the invasion of the surrounding normal tissue. This highly complex phenomenon involves many interrelated processes at several biological space and time scales ranging from the molecular level up to the scale of tissues and organs. Central to the invasion is the alternation of the local microenvironment, which can promote or suppress apoptosis, tumor malignancy and growth, and influence its response to therapies [10], [20], [24]. Moreover, subcellular events regulate cell behavior on the macroscale and vice versa, macroscopic effects have a significant impact on the subcellular dynamics [29]. A large variety of cancer migration models has been proposed, but most focus on biological processes at a single scale. Interconnecting two or more space and/or time scales leads to multiscale models, a recent modeling approach that is attracting increased attention. Linking multiple biological scales significantly increases the complexity, comprises novel features, and raises new challenges. However, it also proposes more realistic and predictive models that provide new insights and a better understanding of the different processes involved in cancer migration and their inter-correlations [4].

S. Sonner and C. Surulescu are supported by DFG grant SU807/1-1.

B. Acid-mediated cancer cell invasion

Most tumors are characterized by regions of acidity and hypoxia. Malignant cells typically exhibit altered metabolic patterns marked by increased glucose uptake and elevated glycolysis [10], [20], [24]. Compared to the aerobic metabolism this inefficient energy production forces tumor cells to upregulate their glycolytic pathway, which increases the production and release of H^+-ions. It results in an acidification of the local microenvironment, which promotes apoptosis of normal cells, while cancer cells can survive and proliferate [17], [21], [24]. Hence, cancer cells get an evolutionary advantage against normal cells, which facilitates invasion. Based on these facts, Gatenby & Gawlinsky [9] proposed a mathematical model for acid-mediated tumor invasion involving reaction-diffusion equations for the dynamics of extracellular protons and the tumor cell density that are coupled to an ordinary differential equation modeling the evolution of the normal cell density. Over the years, the model has been widely studied and several extensions have been proposed and analyzed, see e.g., [11], [16], [21].

C. Regulation of extra-and intracellular pH

The acid-mediated cancer invasion model [9] and its extensions are formulated at the macrolevel. However, the subcellular dynamics regulates and is essentially influenced by the cell behavior on the macroscale. In spite of the increased acidity in the environment the intracellular pH level of cancer cells is at the alkaline side of neutrality since tumor cells are capable to maintain their intracellular pH level through several membrane based ion transport systems. Focusing on the proton dynamics, Webb et al. [29] proposed mathematical models for the influence of microenvironmental changes on the activity of these ion transport mechanisms. An extended model [28] involves even more biological details, e.g. it takes intracellular proton buffering, effects on the expression/activation of matrix degrading enzymes and proton removal by vasculature into

account. It can be seen as a first step towards a multiscale approach to tumor invasion. Very few multiscale models integrating the intracellular proton dynamics have been developed so far. Surulescu and co-authors [6], [18], [23], recently proposed two-scale models for acid-mediated tumor invasion, connecting the subcellular dynamics with the cell behavior on the macroscale. The models couple the dynamics of intra- and extracellular protons with the evolution of tumor cells and normal tissue on the macroscale. Other multiscale approaches to cancer invasion focusing on different mechanisms were proposed and analyzed, e.g., by Chaplain [2] and Trucu et al. [27].

D. Stochasticity in the subcellular dynamics

While all mentioned models, except for [6], are deterministic, it is essential to take stochasticity into account, as it is a relevant feature inherent to many biological processes and there are variations and uncertainties in the behavior of every single cell. In particular, it seems to greatly influence the subcellular dynamics and individual cell behavior, see e.g. [25]. Furthermore, the distribution of intracellular pH at any value of extracellular pH was found to be broader than what was predicted by theoretical models based on machine noise and stochastic variations in the activity of membrane-based mechanisms regulating intracellular pH [15]. Moreover, excess current fluctuations have been observed in the gating of the ion channels [7]. The recent model by Hiremath & Surulescu [6] takes stochastic fluctuations in the intracellular proton dynamics into account. The intracellular proton concentration is modeled by a random ordinary differential equation featuring a stochastic term and this equation is coupled to an ordinary differential equation for the normal tissue and reaction-diffusion equations for the extracellular proton concentration and cancer cell density.

E. Stochastic models with nonlocal sample dependence

We aim to develop a new class of stochastic micro-macro-models for acid-mediated cancer invasion where the subcellular dynamics is governed by stochastic differential equations. Additional novel features arise if reaction functions do not only depend locally on the solutions, but also on their expectation values. This setting is motivated by the fact that stochastic fluctuations are highly relevant in subcellular processes while the effect of averaged quantities is observed on the macroscale. Stochastic micro-macro-models with nonlocal sample dependence coupling reaction-diffusion-equations, ordinary differential equations and stochastic differential equations have not been considered so far.

We first focus on a partial system and consider the interplay between intra- and extracellular protons. The evolution of the intracellular proton concentration is described by a stochastic differential equation, which is coupled to a reaction-diffusion equation governing the dynamics of the extracellular protons. Hence, due to the coupling, this setting opens the possibility of carrying stochasticity from the microscopic, subcellular level to the macroscopic level of protons diffusing in the tumor microenvironment. The governing equation for the extracellular protons hence becomes a random partial differential equation. We also introduce a model with nonlocal sample dependence where the reaction-diffusion equation for the extracellular proton concentration features only the averaged random fluctuations from the subcellular level. In this novel setup the partial differential equation is genuinely deterministic.

Scalar stochastic differential equations with nonlocal sample dependence relating to, but extending mean-field SDEs like those in [8], [26] have been proposed and analyzed in [12]. A solution theory for general local and nonlocal systems coupling stochastic differential equations and reaction-diffusion equations has been developed in [14]. Sufficient conditions for the existence and uniqueness of solutions have been formulated, as well as for their non-negativity and boundedness. These are essential properties in our modeling context since the solutions represent densities or concentrations. Whether a model ensures non-negativity and boundedness of solutions it is not a trivial question within the framework of stochastic differential equations [3].

The paper is organized as follows: In Section II we suggest stochastic models for the intra- and extracellular proton dynamics in a tumor, where local as well as nonlocal models are considered. In Section III we present numerical simulations illustrating the qualitative behavior of solutions.

II. Stochastic models for extra- and intracellular pH in a tumor

Crucial for cancer cell invasion is the acidification of the local microenvironment. While it promotes the death of normal cells, cancer cells can survive and proliferate since they are able to regulate their intracellular pH level through several membrane based transport mechanisms (e.g., Na^+/H^+ exchanger NHE1, monocarcoxylate transporters MCT1/MCT4). We propose mathematical models describing the extra- and intracellular proton dynamics in a tumor where stochastic fluctuations are taken into account in the subcellular dynamics.

The model is formulated as system of differential equations in a bounded spatial domain $D \subset \mathbb{R}^n, n = 1, 2, 3$, with smooth boundary ∂D, $x \in D$ denotes the position and $t \geq 0$ the time. The dependent model variables are the concentrations of extra- and intracellular protons, denoted by X and Y, respectively. While X satisfies a classical reaction diffusion equation, the subcellular dynamics is modeled by an Itô stochastic differential equation,

$$\begin{aligned}
&\partial_t X = \delta \Delta X + r(X, Y) - \alpha X \\
&\partial_\nu X|_{\partial D} = 0, \quad X(0) = \zeta, \\
&dY = (-r(X, Y) - \beta Y + \varphi(Y))dt + \gamma Y(1 - Y)dW \\
&Y(0) = \eta,
\end{aligned}$$

where W is a standard scalar Wiener process and dW the corresponding Itô differential. The constants α, β, γ and δ are

positive, Δ denotes the Laplace operator and ∂_ν the outward unit normal derivative at the boundary ∂D. The initial data is given by the non-negative and bounded functions ζ and η.

Homogeneous Neumann boundary conditions are specified for the extracellular proton concentration mimicking that no protons can leave the system. The model variables are normalized w.r.t. the maximum proton concentrations, i.e., X and Y take values within the unit interval $[0, 1]$.

Assuming that the averaged effects of the stochastic fluctuations on the intracellular level are observed in the macroscopic dynamics leads to the nonlocal model

$$\begin{aligned}
&\partial_t X = \delta \Delta X + r(X, \mathbb{E}(Y)) - \alpha X \\
&\partial_\nu X|_{\partial D} = 0, \quad X(0) = \zeta, \\
&dY = (-r(X, Y) - \beta Y + \varphi(Y))dt + \gamma Y(1 - Y)dW \\
&Y(0) = \eta,
\end{aligned}$$

where $\mathbb{E}$ denotes the expectation value. The dependence of r on $\mathbb{E}(Y)$ instead of Y highlights that the subcellular effects are averaged when considering the proton concentration in the extracellular space. In the intracellular space, however, Y is seen as a genuine stochastic process influencing as such the proton extrusion. Thus, X becomes a deterministic quantity satisfying a reaction-diffusion equation, while Y is a stochastic process evolving according to an Itô stochastic differential equation.

The following processes are included in the model:

- Extracellular protons are transported by diffusion which is modeled according to Fick's law. The diffusion coefficient δ is constant.
- The proton transport mechanisms through the cell membrane are described by the function r, which is of the form
$$r(x, y) = a_1 \frac{y}{1 + y^2 + a_2 x^2} - a_3 \frac{x}{1 + a_4 y^2},$$
according to [23], where $a_1, \ldots, a_4$, are positive constants. The choice of r was motivated by the functions chosen in [6] and [29] that in turn were based on quantitative information in [1], where the rates of proton flux due to NDCBE, NHE, and AE transporters were measured [1].
- The constant decay rate α characterizes the loss of extracellular protons by processes other than membrane based transport into the tumor cells (e.g., uptake by vasculature, normal cells, buffering etc.).
- In the intracellular space, the function φ models production of Y by glycolysis (which in cancer cells is much amplified when compared to normal cells and hence non-negligible),
$$\varphi(y) = \varphi_0 \frac{y}{1 + a_2 y},$$
where φ_0 is a positive constant.
- The decay term βY describes the loss of protons by intracellular buffering (e.g., by organelles).
- The stochastic perturbation $\gamma Y(1 - Y)$ quantifies the (stochastic) variability in the production/decay of intracellular protons. It accounts for values of Y ranging between 0 (complete alkalinization) and 1 (maximum acidification). Both bounds are lethal for the cell and hence there is no variability at these threshold values. Also, a maximum is achieved in the middle of this interval, suggesting that larger spreads of the Y distribution are not allowed.

A solution theory for coupled systems of local and nonlocal stochastic differential equations and reaction diffusion equations such as the proton dynamics models was developed in [14]. In particular, the results imply the existence and uniqueness of solutions. If, moreover, the parameters satisfy the following relation

$$\begin{aligned}
\frac{a_1}{2 + a_2} &\leq \frac{a_3}{1 + a_4} + \alpha, \\
\frac{\varphi_0}{1 + a_2} + \frac{a_3}{1 + a_4} &\leq \frac{a_1}{2 + a_2} + \beta,
\end{aligned} \tag{1}$$

then the solutions remain non-negative and bounded by 1 as long as they exist.

III. Numerical simulations

In this section we illustrate the qualitative behavior of solutions of the proton dynamics models by numerical simulations. We consider and compare the local version

$$\begin{aligned}
\partial_t X &= \delta \Delta X + r(X, Y) - \alpha X \\
dY &= (-r(X, Y) - \beta Y + \varphi(Y))dt + \gamma Y(1 - Y)dW,
\end{aligned}$$

with its nonlocal counterpart

$$\begin{aligned}
\partial_t X &= \delta \Delta X + r(X, \mathbb{E}(Y)) - \alpha X \\
dY &= (-r(X, Y) - \beta Y + \varphi(Y))dt + \gamma Y(1 - Y)dW,
\end{aligned}$$

in a spatially one-dimensional domain $D = (0, 1)$. The extracellular and intracellular proton concentrations are normalized w.r.t. their maximum values and necessarily take values within the unit interval $[0, 1]$.

For the simulations we choose the initial and boundary values

$$\begin{aligned}
\partial_x X_t|_{x=0} &= 0, & \partial_x X_t|_{x=1} &= \mu, \\
X_0 &= \zeta, & Y_0 &= \eta,
\end{aligned}$$

where μ is a positive constant and

$$\zeta(x) = 0.6e^{-x^2}, \quad \eta(x) = 0.3e^{-2x^2}, \qquad x \in D.$$

The parameters used in the simulations are summarized in Table I. We remark that the analytical results mentioned in the previous section extend to this setting allowing for nonhomogeneous Neumann-type boundary conditions. This ensures the existence and uniqueness of solutions. Moreover, for the chosen parameters the conditions (1) are certainly

[1] NDCBE (Na^+ dependent Cl^--HCO_3^- exchanger), NHE (Na^+ and H^+ exchanger) and AE (Cl^--HCO_3^- or *a*nion *e*xchanger) are specific ion transporters on the cell membrane.

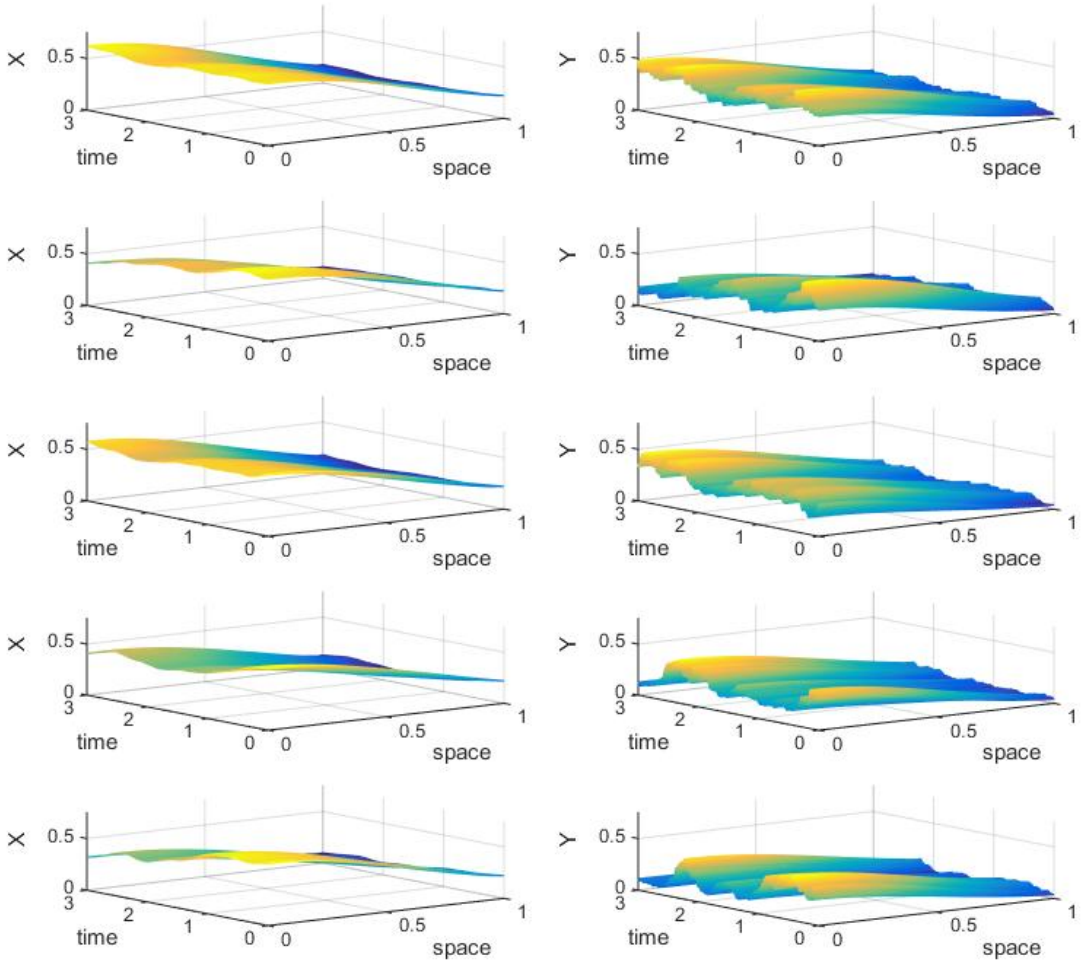


Fig. 1: Five sample paths for the solutions of the local proton dynamics model.
Left column: concentration X of extracellular protons. Right column: concentration Y of intracellular protons.

satisfied and the solutions remain non-negative and bounded from above by 1.

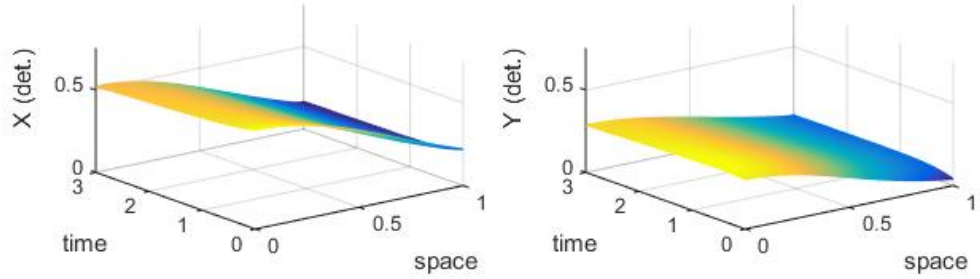
To solve the reaction diffusion equation for X we use finite differences and an explicit Euler scheme, and the Euler-Mayurama method [13] to evaluate the solution of the stochastic differential equation.

The shape of the initial data is motivated by the assumption that the core of the tumor is located at the left end of the spatial interval. Cancer cells emit protons rendering the tumor microenvironment acidic. The acidity is supposed to decrease towards the tumor edge, i.e. advancing in space towards the right end of the interval. The level of the concentration of intracellular protons should always be lower than the value of its extracellular counterparts, in order to allow the survival and proliferation of cancer cells.

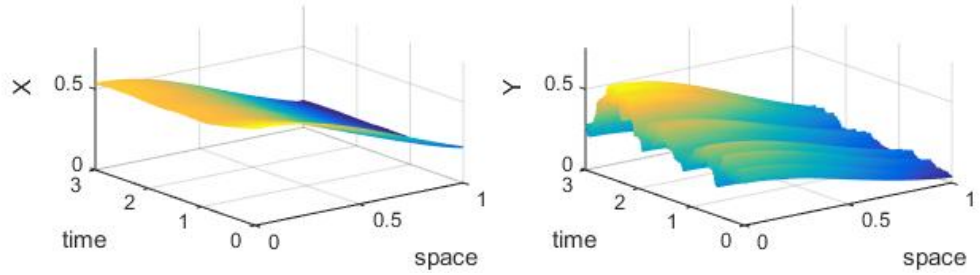
TABLE I: Parameter values

parameter	symbol	value
diffusion coefficient	d	0.0001
decay rate for X	α	1
decay rate for Y	β	2
activity of transporter 1	a_1	2
normalization constant	a_2	0.001
activity of transporter 2	a_3	0.5
normalization constant	a_4	0.1
production rate of Y	φ_0	2
intensity of the stochastic perturbation	γ	1
Neumann boundary conditions at $x = 1$	μ	0.5

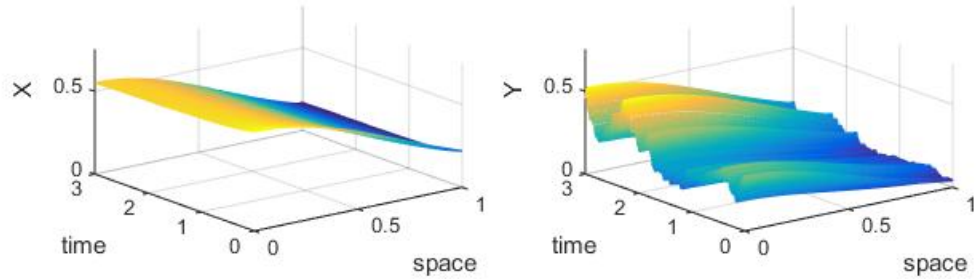
Figure 1 illustrates the time evolution of extra- and intracellular proton concentrations as predicted by the local stochastic model. We plotted five sample paths for the solutions. Remarkable are the inter-path differences in both proton concentra-



(a) deterministic proton dynamics model



(b) nonlocal model: X evaluated using the average over 20 sample paths (left), one sample path for Y (right)



(c) nonlocal model: X evaluated using the average over 1000 sample paths (left), one sample path for Y (right)

Fig. 2: Simulations for the deterministic and nonlocal proton dynamics model.
Left column: concentration X of extracellular protons. Right column: concentration Y of intracellular protons.

tions. This suggests a tumor-to-tumor variability w.r.t. acidity, although the same type of cancer cells is considered.

In Figure 2 we show the qualitative behavior of solutions of the deterministic model and of the nonlocal proton dynamics model. The expectation values of the solution of the stochastic differential equation involved in the reaction-diffusion equation for extracellular protons were evaluated by averaging over 20 (Figure 2b) and 1000 sample paths (Figure 2c). The solutions of the deterministic model, i.e., the proton dynamics model with $\gamma = 0$, are plotted in Figure 2a. In each of the Figures 2b and 2c the solution X of the nonlocal stochastic model and one sample path for the intracellular protons Y are shown.

As expected, when averaging over a large number of tumors the differences on the extracellular, macroscopic level between the nonlocal model and the purely deterministic one are small. However, if a significantly smaller number of tumors is considered (corresponding to a less frequent cancer type and a low number of sample paths) randomness, i.e. inter-tumor variations, seems to play an important role in the acidification of the extracellular microenvironment. This observation can be relevant for the sensitivity of tumors against therapies and for tumor aggressiveness, as it is well known that low values of extracellular pH and the gradients between intracellular and extracellular pH significantly influence the response to treatments like radiotherapy and chemotherapy [5], [19], [22]. Our modeling approach and numerical simulations seem to endorse the necessity of paying particular attention to individualized treatment, especially in case of rare tumors, where it is difficult to rely on clinical experience acquired by a comparatively small number of patients.

IV. Concluding remarks

Extending the model by including macroscopic equations for the dynamics of cancer cells and normal tissue is part of ongoing research. We aim to assess the behavior of the tumor in interaction with its surroundings and under the influence of the stochastic proton dynamics studied here. Thereby, the

challenge comes from the nonlinear diffusion of tumor cells and the performed taxis, along with the lack of diffusivity of normal tissue.

REFERENCES

[1] M. J. Boyer and I. F. Tannock, Regulation of intracellular pH in tumor cell lines: influence of microenvironmental conditions, *Cancer Res.*, **52** (1992), 4441–4447.

[2] M. A. J. Chaplain and G. Lolas, Mathematical modelling of cancer invasion of tissue: dynamic heterogeneity, *Netw. Heterog. Media*, **1** (2006), 399–439.

[3] J. Cresson, B. Puig and S. Sonner, Stochastic models in biology and the invariance problem, *Discrete Cont. Dyn. Syst. Ser. B*, to appear.

[4] T. S. Deisboeck, Z. Wang, P. Macklin and V. Cristini, Multiscale cancer modeling, *Annu. Rev. Biomed. Eng.*, **13** (2011), 127–155.

[5] M. L. Freeman and E. Sierra, An acidic extracellular environment reduces the fixation of radiation damage. *Radiat. Resist.*, **97** (1984), 154–161.

[6] S. Hiremath and C. Surulescu, A stochastic multiscale model for acid mediated cancer invasion, *Nonlinear Anal. Real World Appl.*, **22** (2015), 176–205.

[7] E. Jakobsson and S. W. Chiu, Stochastic theory of ion movement in channels with single-ion occupancy. Application to sodium permeation of gramicidin channel,, *Biophys. J.*, **52** (1987), 33–45.

[8] Y. Jiongmin, Linear-quadratic optimal control problems for mean-field stochastic differential equations, *SIAM J. Control Optim.*, **51** (2013), 2809–2838.

[9] R. A. Gatenby and E. T. Gawlinski, A reaction-diffusion model of cancer invasion, *Cancer Res.*, **56** (1996), 5745–5753.

[10] R. A. Gatenby, E. T. Gawlinski, A. F. Gmitro, B. Kaylor and R. J. Gillies, Acid-mediated tumor invasion: a multidisciplinary study, *Cancer Res.*, **66** (2006), 5216–5223.

[11] J. B. McGillen, E. A. Gaffney, N. K. Martin and P. K. Maini, A general reaction-diffusion model of acid-mediated cancer invasion, *J. Math. Biol.*, **68** (2014), 1199–1224.

[12] P. E. Kloeden and T. Lorenz, Stochastic differential equations with nonlocal sample dependence, *Stochastic Anal. Appl.*, **28** (2010), 937–945.

[13] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin-Heidelberg, 1992.

[14] P. E. Kloeden, S. Sonner and C. Surulescu, A nonlocal sample dependence SDE-PDE system modeling proton dynamics in a tumor, *Discrete Cont. Dyn. Syst. Ser. B*, to appear.

[15] A. H. Lee and I. F. Tannock, Heterogeneity of intracellular pH and of mechanisms that regulate intracellular pH in populations of cultured cells, *Cancer Res.*, **58** (1998), 1901–1908.

[16] N. K. Martin, R. A. Gatenby, E. T. Gawlinski and P. K. Maini, Tumor-stromal interactions in acid-mediated invasion: a mathematical model, *J. Theor. Biol.*, **267** (2010), 461–470.

[17] R. Martínez-Zaguilán, E. A. Seftor, R. E. B. Seftor, Y. W. Chu, R. J. Gillies and M. J. C. Hendrix, Acidic pH enhances the invasive behavior of human melanoma cells, *Clin. Exp. Metastasis*, **14** (1996), 176–186.

[18] G. Meral, C. Stinner and C. Surulescu, A multiscale model for acid-mediated tumor invasion: therapy approaches, *J. Coupled Syst. Multiscale Dyn.*, 3 (2015), 135–142.

[19] A. De Milito and S. Fais, Tumor acidity, chemoresistance and proton pump inhibitors. *Future Oncol.*, **1** (2005), 779–786.

[20] I. F. Robey, B. K. Baggett, N. D. Kirkpatrick, D. J. Roe, J. D. Bonnie, F. Sloane, A. I. Hashim, D. L. Morse, N. Raghunand, R. A. Gatenby and R. J. Gillies, Bicarbonate increases tumor pH and inhibits spontaneous metastases, *Cancer Res.*, **69** (2009), 2260–2268.

[21] K. Smallbone, D. J. Gavaghan, R. A. Gatenby and P. K. Maini, The role of acidity in solid tumor growth and invasion, *J. Theor. Biol.*, **235** (2005), 476–484.

[22] C. W. Song, R. Griffin, and H. J. Park, Influence of Tumor pH on Therapeutic Response, in *Cancer Drug Resistance*, Edited B. Teicher, Humana Press Inc., Totowa, NJ (2006).

[23] C. Stinner, C. Surulescu and G. Meral, A multiscale model for pH-tactic invasion with time-varying carrying capacities, *IMA J. Appl. Math.*, **80** (2015), 1300–1321.

[24] C. Stock and A. Schwab, Protons make tumor cells move like clockwork, *Pflugers Arch. - European J. Physiology*, **458** (2009), 981–992.

[25] C. L. Stokes, D. A. Lauffenburger and S. K. Williams, Migration of individual microvessel endothelial cells: stochastic model and parameter measurement, *J. Cell Science*, **99** (1991), 419–430.

[26] J. Touboul, G. Herrmann and O. Faugeras, Noise-induced behaviors in neural mean field dynamics, *SIAM J. Appl. Dynamical Syst.*, **11** (2012), 49–81.

[27] D. Trucu, P. Lin, M. A. J. Chaplain, and Y. Wang, A multiscale moving boundary model arising in cancer invasion, *Multiscale Model. Simul.*, **11** (2013), 309–335.

[28] S. D. Webb, J. A. Sherratt and R. G. Fish, Alternations in proteolytic activity at low pH and its association with invasion: a theoretical model, *Clin. Exp. Metastasis*, **17** (1999), 397–407.

[29] S. D. Webb, J. A. Sherratt and R. G. Fish, Mathematical modelling of tumour acidity: regulation of intracellular pH, *J. Theor. Biol.*, **196** (1999), 237–250.

Influence of nanofillers on crystallization of polypropylene at rapid cooling

Buncha Suksut[1]
[1]Chair of Composite Engineering
University of Kaiserslautern
Kaiserslautern, Germany
E-mail: buncha.suksut@mv.uni-kl.de

Alois K.Schlarb[1,2,3]
[1]Chair of Composite Engineering
University of Kaiserslautern
Kaiserslautern, Germany
[2]Research Center OPTIMAS
University of Kaiserslautern
Kaiserslautern, Germany
[3]INM – Leibniz Institute for New Materials
Saarbrucken, Germany
E-mail: alois.schlarb@mv.uni-kl.de

I. INTRODUCTION

In general, it is known that the mechanical properties of semi-crystalline polymers are highly dependent on the morphological structure, which is governed by the respective crystallization process during polymer processing. Previous studies [1-2], demonstrated that the presence of fillers initiated faster crystallization and changed the final morphology of injection-molded PP parts.

The aim of this work was to investigate the crystallization of isotactic polypropylene (iPP) containing different types of nanoparticles by using conventional differential scanning calorimetry (DSC), and flash DSC. While the conventional DSC is accepted for characterization of certain materials properties, the flash DSC allows us to better understand the influence of nanofillers on the crystallization and the structure development under real processing conditions.

II. EXPERIMENTAL

A. Materials and processes

Polypropylene (PP HD120 MO, Borealis GmbH, Burghausen) was used as a matrix material and silicon dioxide (SiO_2) (Aerosil R8200, Evonik Degussa GmbH, Hanau-Wolfgang) and titatium dioxide (TiO_2) nanoparticles (Hombitec RM130 F, Huntsman, Duisburg) as inorganic nanofillers. The mean primary particle diameter of the SiO_2 and TiO_2 are 12 nm and 15 nm, respectively, according to the manufacturers.

Polypropylene nanocomposites were first prepared using an optimized co-rotating twin-screw extruder (Theysohn, Theysohn Extrusionstechnik GmbH, Salzgitter) followed by injection molding (Arburg Allround 420C, ARBURG GmbH + Co KG, Loßburg). The procedure is described in the literature [1] in detail. PP nanocomposites with 4 vol.% content of SiO_2 (PP-S-V4) and TiO_2 (PP-T-V4) were prepared.

B. Differential Scanning Calorimetry (DSC)

The specimens were prepared from the center of the injection-molded plate. The measurements were performed on a DSC (TA Q20, TA instruments, Eschborn) equipped with the Refrigerated Cooling Systems 90 (RCS90). The weight of each specimen ranged from 5 to 10 mg. For non-isothermal crystallization, the samples were heated to 220°C at a rate of 10 K/min and held for 3 minutes at this temperature. Then, the specimens were cooled to 0°C at various cooling rates (from 2 to 40 K/min).

For isothermal investigation, the specimens were heated to 220°C at a rate of 20 K/min. The holding time of 3 minutes at 220°C was selected to remove the processing history. The specimens were then quenched at a rate of 40 K/min to various isothermal crystallization temperatures (T_i) within the range of 128°C - 134°C, and kept constant at that temperature until the crystallization was completed.

C. Flash DSC

Non-isothermal crystallization experiments were performed using a power compensation differential fast scanning chip calorimetry (FLASH DSC 1, Mettler Toledo, Gießen) in combination with a Huber Intracooler TC100. Dry nitrogen was used as a purge gas at a rate of 30 mL/min. In order to provide a good thermal contact between the specimen and the sensor, the specimen is molten at a heating rate of 0.1 K/s before the actual experiments. Specimens were heated to 220°C at a rate of 1000 K/s, and were kept at this temperature for 0.1 s. Then, for non-isothermal crystallization, the specimen was cooled to -90°C at various cooing rates. The desired cooling rate range is between 0.5 to 4000 K/s.

To study the isothermal crystallization of the material, the test specimen was quenched at the rate of 4000 K/s to the target T_i within the range of 0°C - 150°C, and kept constant at that temperature until the crystallization was completed.

III. RESULTS AND DISCUSSION

A. Nonisothermal crystallization

Fig. 1 shows the cooling curves of neat PP obtained from flash DSC. The plots show that only one crystallization process can occur at slower cooling rates. In addition, two different crystallization processes can be observed at cooling rates above 100 K/s. The first crystallization process, high temperature, can be interpreted as the crystallization of the α-phase of PP. The second one, low temperature, is the formation of the mesophase [3-4].

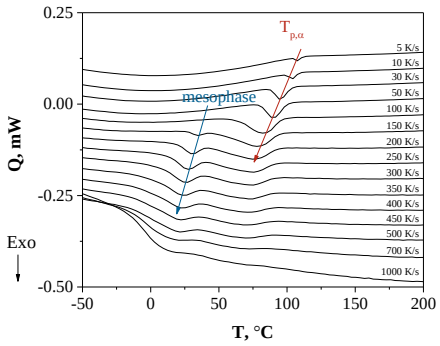


Fig. 1. Cooling curves of neat PP measured with FDSC

From the cooling curves in Fig. 1, the peak of crystallization temperature (T_p) can be determined and are shown in Fig. 2. As it is expected, an increase in the cooling rate results in the lowering of the T_p in all materials. The results show that the acceleration of the crystallization by an incorporation of TiO_2 nanoparticles, irrespective of the cooling rate. In case of PP/SiO_2, the nucleation effect can be observed only at the cooling rate lower than 1K/s.

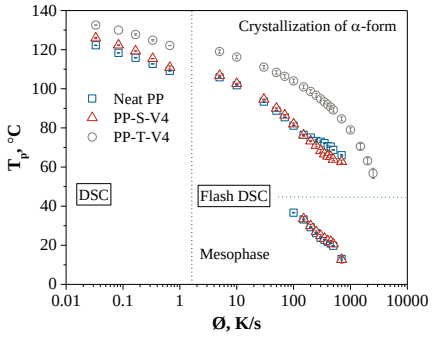


Fig. 2. T_p versus the cooling rate of neat and nano-filled PP

Furthermore, the mesophase of PP is not influenced by the SiO_2 nanoparticles, while it is obstructed by the presence of TiO_2 nanoparticles. Beyond a cooling rate of 700 K/s, the observable crystallization and mesophase formation cannot be seen in neat PP and PP filled with 4 vol.% SiO_2. However, the crystallization still occurs at a cooling rate up to around 2000 K/s in PP filled with 4 vol.% TiO_2.

B. Isothermal crystallization

The half-time of crystallization ($t_{1/2}$) defined as the time spent from the beginning of the crystallization to the completion of half of the final crystallization can be indicated the kinetics of crystallization and are shown in Fig. 3. A higher value of $t_{1/2}$ means a slower rate of crystallization, and vice versa. There is a bimodal behavior for neat PP and PP/SiO_2-4 vol.%, corresponding to maximum rates of crystallization of the α-phase (higher temperature) and of the formation of the mesophase (lower temperature) [3-4]. Fig. 3 also reveals that the presence of the nanofillers increases the crystallization rate compared to the neat PP. Moreover, the increase of the crystallization rate in PP/TiO_2-4 vol.% is much more pronounced than in the case of PP/SiO_2-4 vol.%.

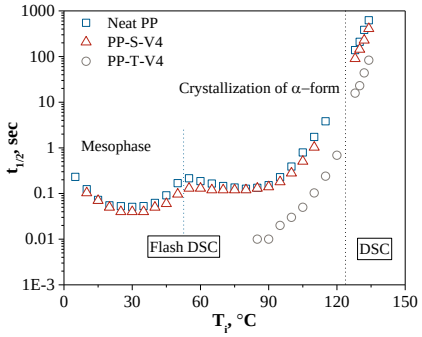


Fig. 3. $t_{1/2}$ of neat PP and nano-filled PP versus temperature

IV. CONCLUSIONS

This work shows the influence of nanofillers on the crystallization process of polypropylene by using conventional DSC and flash DSC analyses in order to verify a wide range of cooling rates and crystallization temperatures. The results reveal that SiO_2 nanoparticles only act as mild nucleating agent at cooling rates below 1K/s. At cooling rates higher than 100 K/s neat PP and PP-SiO_2 nanocomposites exhibit crystallization in a mesophase, too. In contrast, titania serves as a very effective nucleating agent by reducing the free energy needed for nucleation and increasing the nucleation density, consequently leading to a faster crystallization rate. In this case a mesophase cannot be observed.

ACKNOWLEDGMENT

The authors thank Borealis and Evonik for providing the materials needs for this study. We also thank Dr. L. Lin (CCe, Kaiserslautern) and Mr. V. Demchuk (PE, Hamburg-Harburg) for the compounding and injection molding of the nanocomposite samples.

REFERENCES

[1] A. K. Schlarb, D. N. Suwitaningsih, M. Kopnarski and G. Niedner-Schatteburg, J. Appl. Polym. Sci., vol. 131, pp. 1-8, 2014.

[2] B. Suksut and A. K. Schlarb, Journal of Plastic Technology,vol. 10, pp. 68-85, 2014

[3] R. Androsch, M. L. D. Lorenzo, C. Schick, B. Wunderlich. Polymer, vol. 51, pp. 4639-4662. 2010

[4] J. E. K. Schawe, J Therm Anal Calorim, vol. 116, pp. 1165-1173. 2014

Multiscale Model for Tumor Invasion and Therapy Approaches

Aydar Uatay
Technische Universtät Kaiserslautern
Felix-Klein-Zentrum für Mathematik
Paul-Ehrlich-Str. 31, 67663 Kaiserslautern, Germany
Email: uatay@mathematik.uni-kl.de

***Abstract*—Monoscale models for cancer invasion describe it to a certain extent. As the importance of subcellular dynamics and its influence on cancer cell population propagation is established, multiscale models gained attention. In this paper we present a multiscale model, that describes micro- and macroscale dynamics. On the microscopic level, the model accounts for integrin (a type of transmembrane receptors) binding to extracellular matrix (ECM) proteins (molecules, which provide structural and biochemical support). This subcellular dynamics condition cancer cell migration by modifying its motility through cell reshaping. We also consider heterogeneity of the tumor with respect to phenotype (proliferative and migrative) and interaction between both types as well. Therapeutic approaches, such as chemo- and radiotherapy are taken into account. The modeling will lead us to a system of partial and ordinary differential equations. The effect of potential concurrent targeted therapy aimed at integrin inhibition and subsequent impediment of cancer cell migration is assessed via numerical simulations. The simulations suggest that the inhibition can slow down the migration, but at the cost of additional tissue damage at the original tumor site.**

I. Introduction

Cancer is a general term for a large palette of diseases. There exist more than hundred types of cancer, all of which are characterized by abnormal cell growth and potential ability to invade adjacent tissues and spread to distal organs. Overall, there were 8.2 million cancer related deaths in the world in 2012 and more than 14 million new cases were registered in the same year [1].

A powerful non-invasive way to asses cancer growth, invasion, and its effects on healthy tissues (i.e., the patients response to some treatment), is to use mathematical modeling. With the help of mathematical models it is not only possible to construct the optimal treatment plan, but also to provide valuable insight for prospective treatment involving e.g. some chemotherapeutic agents and/or adaptive radiation dose prescriptions. Thus, modeling can be a valuable tool for both cancer researchers and practitioners.

A common approach to model evolution and invasion of cancer cells is to set up models on the population (macroscopic) level. However, this setting suffers from a major drawback. It is known that events on subcellular (microscopic) level greatly influence the behavior of cancer cells. For example, the latter tend to drastically modify their microenvironment by increasing acidity levels [2]. It is also known that the receptors on tumor cell membrane play a significant role in mediating proliferation, migration and invasion [3]. However, having a model set only on the microscopic level will not give us insight on the evolution of population dynamics of cancer cells or healthy tissues - the very quantities we are mostly interested in.

We will develop and analyze a multiscale model coupling the subcellular dynamics of the microscopic level with cell

population dynamics on the macroscopic level. Various therapy strategies, such as adjuvant or resection-free chemotherapy and/or radiotherapy, will also be studied in the framework of our model.

II. Model

A. Subcellular dynamics

In order to model subcellular dynamics, we focus on the reaction kinetics between cancer cell integrins and the ECM proteins. Moreover, we simplify the character of the interaction between the two, by assuming that there is no distinction between different types of integrins and consider only the total concentration of all types of integrins on the cell surface (as the subcellular dynamics is not fully understood). Thus, the subcellular interaction of cancer cells (i.e. adhesion and subsequent signal transduction) with the healthy tissue is determined only by integrins binding to ECM molecules.

Let v be the concentration of ECM-proteins, R_0 be the total concentration of free receptors of a single cancer cell (assumed to be constant). Then, the reversible reaction of binding the two, forming the complex R_0v of the bound integrins, can be described by the following chemical kinetics:

$$v + R_0 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R_0 v,$$

where k_1 and k_{-1} are the reaction rates for the forward and reverse reaction, respectively. Note that this reaction describes in a simplified way the attachment of a cancer cell to the ECM. Let y be the concentration of receptors bound to the ECM molecules, i.e. we have $y = R_0 v$. We consider chemotherapy, aimed at inhibiting cell-ECM adhesion. For example, the drug Heparin was found to inhibit αIIbβ3 integrin expression, which significantly reduced tumor cell adhesion [4] and consequently reduced metastatic potential, i.e. further cancer spread. Hence, we take the dose dependent binding and unbinding rates $k_1(d_c)$ and $k_{-1}(d_c)$ to be of the following form:

$$k_1(d_c) := \hat{k}_1 - b_d d_c$$

$$k_{-1}(d_c) := \hat{k}_{-1} + u_d d_c,$$

where $\hat{k}_1$, $\hat{k}_{-1}$, b_d, u_d are constants. By the law of mass action, the change of y in time can be modeled by the following ODE:

$$y_t = k_1(d_c)(R_0 - y)v - k_{-1}(d_c)y. \tag{1}$$

In response to changes in their environment, the cancer cells are able to reorganize their cytoskeleton and contract their shape in order to perform mesenchymal motion. As these changes in environment are transmitted through the transmembrane receptors, the contractivity is achieved as a result of series of chemical reactions on sub-cellular level [5]. Since cell reshaping is not instantaneous, there must be some delay between the time it takes for a cell to sense the change and reshape itself by contraction. We also assume that it is not advantageous for a cell to reshape its cytoskeleton, unless it received a signal to do so. Therefore, the contractivity function has the following form (see [6]):

$$\kappa_t := -\delta_\kappa \kappa + H(y(t-\tau)), \tag{2}$$

where $\tau > 0$ denotes the delay and the function H models the effect of the subcellular dynamics on a cell's ability to reshape itself, δ_κ is a constant, and the decay term models the preference of a cell to retain its shape. As the dependence of contractivity on subcellular dynamics is not completely elucidated, we assume for simplicity that $H := My$ where M is a positive constant.

B. Cell population dynamics

A tumor is not necessarily homogeneous with respect to its phenotype. It is known that the migrative phenotype is more resistant to irradiation induced apoptosis than theproliferative phenotype (see e.g., [7], [8]). Thus, we divide the cancer cell population into two subpopulations of moving and of proliferating cells, whose densities we denote by m and q, respectively. Furthermore, we assume that a cancer cell can either move or proliferate, but not both at the same time. This is the so-called "go-or-grow" dichotomy (see e.g., [9]), which in the context of tumor cell migration has also been used in [10].

Concerning moving cells, we assume there are two sources

of motion: the random motility, and the directed migration in the form of haptotaxis (see e.g., [11]). The random motility represents the diffusive nature of cancer cell movement through the ECM. On the other hand, the speed of the directed movement (haptotaxis) depends on the density of the surrounding tissue, as well as on the contractivity. Based on environmental cues, a cell either stays in the quiescent state or proceeds into the proliferative state. As these environmental cues induce changes on the subcellular level, and the triggered receptors transmit the signals to the cell, we assume that the transition rate between the proliferative and the moving cells depends on the concentration y of bound integrins. We assume the transition rate from the moving cells m to the proliferating cells q is given by a positive function γ. This function γ ought to be a decreasing one, since a lower concentration of the bound integrins indicates a weaker attachment to the ECM, which in turn inhibits the mesenchymal motion and hence, due to the "go-or-grow" dichotomy, causes phenotype switching from the migrating to the proliferating regime. The opposite holds true as well: the transition rate from the proliferating to the migrating subpopulations has to be proportional to the concentration of the bound integrins, as the increased binding promotes contractivity, in addition to amplified mesenchymal motion. These are not unreasonable assumptions, since it was shown that promoting migration can reduce the proliferative potential [12] and inhibiting proliferation enhances migration [13]. Therefore,the above consideration, the equation for the densities of moving cells m is given by:

$$\partial_t m = \nabla \cdot (\phi(\kappa, m, q, v)\nabla m) - \nabla \cdot (\psi(\kappa, v) m \nabla v) + \lambda(y)q - \gamma(y)m - \Gamma_m R_m(d_r)m. \quad (3)$$

The first term on the right hand side of equation (3) represents a tendency of cells to undergo random motion. Following [10], this motion is conditioned by a function ϕ:

$$\phi(\kappa, m, q, v) := \frac{D_c \kappa}{1 + \frac{(m+q)v}{K_c K_v}},$$

where D_c is a diffusion constant and K_c is the total carrying capacity (maximum density) of the cancer cells, K_v is carrying capacity of the normal cells. The second term represents directed movement in response to the gradient of ECM density (haptotaxis). Following [6], we assume haptotaxis is conditioned by function ψ:

$$\psi(\kappa, v) := \frac{D_H \kappa v}{K_v + v},$$

where D_H is a haptotactic constant. Functions λ and γ in equation (3) represent transition rates between proliferating and moving cancer cells. Due to lack of experimental data, we consider simple relations between the integrin concentration and the rates:

$$\gamma := \frac{\gamma_0}{R_0 + y}$$
$$\lambda := \lambda_0 (R_0 + y),$$

where γ_0 and λ_0 are positive constants. On the other hand, the evolution of the proliferative subpopulation q is given by (using the logistic growth model with competition for nutrients between the moving cells and healthy tissue):

$$\partial_t q = \mu_q q \left(1 - \frac{m+q}{K_c} - \eta_1 \frac{v}{K_v}\right) - \lambda(y)q + \gamma(y)m - \Gamma_q R_q(d_r)q, \quad (4)$$

where μ_q, η_1 are positive parameters characterizing the proliferation rate and the competition, respectively. Note that the ECM neither diffuses, nor undergoes any other significant motion when compared to the cancer cells. The ECM is also degraded by both the moving cancer cells (by releasing the degrading enzymes) and the proliferating cells (due to reinforced glycolysis). Therefore, we have the following ODE for the density v of normal cells:

$$\partial_t v = -\delta_v (m + \delta_q q)v + \mu_v v \left(1 - \eta_2 \frac{m+q}{K_c} - \frac{v}{K_v}\right) - \Gamma_v R_v(d_r)v, \quad (5)$$

where δ_v denotes the parameter characterizing degradation by the MDEs (matrix degrading enzymes), δ_q characterizes necrosis (cell destruction) due to increased acidity, and μ_v, η_2 characterize the proliferation rate and the competition, respectively. The radiotherapy in the population equations above is described by the terms $\Gamma_j R_j(d_r)j$, where $j \in \{m, q, v\}$ and Γ_j is sensitivity to irradiation of the population j. In practice,

the irradiation therapy is applied in multiple fractions. Thus, the decay rate is given by (see [14]):

$$R_j(d_r) = \sum_{i=1}^{l}(1 - S_j(d_r)),$$
$$S_j(d_r) = \exp(-\alpha_j d_r(1 + \hat{d_r}/(\alpha_j/\beta_j))),$$

where l is the number of fractions, d_r is the total dose and $\hat{d_r}$ is the dose (in Gy) per fraction. Here we employed the widespread Linear-Quadratic (LQ) model to quantify cell-death due to irradiation [15]. Thereby, α_j represents lethal lesions produced by single radiation track and β_j represents lethal lesions produced by two radiation tracks.

Using equations (1)-(5) we obtain the following system of differential equations:

$$\begin{cases} \partial_t m = \nabla \cdot (\phi(\kappa, m, q, v)\nabla m) - \nabla \cdot (\psi(\kappa, v)m\nabla v) \\ \qquad + \lambda(y)q - \gamma(y)m - \Gamma_m R_m(d_r)m \\ \partial_t q = \mu_q q\left(1 - \dfrac{m+q}{K_c} - \eta_1 \dfrac{v}{K_v}\right) - \lambda(y)q + \gamma(y)m \\ \qquad - \Gamma_q R_q(d_r)q \\ \partial_t v = -\delta_v(m + \delta_q q)v + \mu_v v\left(1 - \eta_2 \dfrac{m+q}{K_c} - \dfrac{v}{K_v}\right) \\ \qquad - \Gamma_v R_v(d_r)v \\ \partial_t y = k_1(d_c)(R_0 - y)v - k_{-1}(d_c)y \\ \partial_t \kappa = -\delta_\kappa \kappa + H(y(t-\tau)). \end{cases} \tag{6}$$

We assume that the cells do not leave through the boundary of a domain, where they are situated.

III. Mathematical analysis

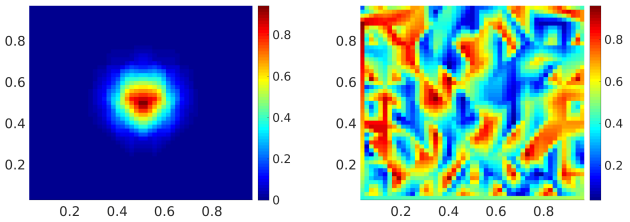
The local existence and uniqueness of the solution to (6) is proved by the author in [16].

IV. Numerical Results

Before we solve (6) numerically, we perform the nondimensionalization for computational convenience. The system in dimensionless form reads as follows (see [16] for rescaled variables and biologically relevant parameters used for simulations):

$$\begin{cases} \partial_t m = \nabla \cdot \left(\dfrac{D_c \kappa}{1 + (m+q)v}\nabla m\right) - \nabla \cdot \left(\dfrac{D_H \kappa v}{1+v} m \nabla v\right) \\ \qquad + \lambda_0(1+y)q - \dfrac{\gamma_0}{1+y}m - \Gamma_m R_m(d_r)m \\ \partial_t q = \mu_q q(1 - (m+q) - \eta_1 v) \\ \qquad - \lambda_0(1+y)q + \dfrac{\gamma_0}{1+y}m - \Gamma_q R_q(d_r)q \\ \partial_t v = -\delta_v(m + \delta_q q)v + \mu_v v(1 - \eta_2(m+q) - v) \\ \qquad - \Gamma_v R_v(d_r)v \\ \partial_\theta y = k_1(d_c)(1-y)v - k_{-1}(d_c)y \\ \partial_\theta \kappa = -\delta_\kappa \kappa + My(\theta - \tau), \end{cases} \tag{7}$$

where $\theta = \frac{t}{\chi}$ and χ is a time scaling constant, introduced to distinguish subcellular dynamics, which proceed much faster compared to cell population dynamics. The rescaled domain is a unit square. The plots of initial total cancer cell density $m_0 + q_0$ and initial ECM density can be found on Fig. 1. We assume that the initial density of the migrating cancer cells m_0 constitute 70% of the total initial cancer cell density. Due to absence of data, the initial ECM density is simulated by uniformly distributed random numbers on interval $(0, 1)$. We employ Finite Volume Method in order to obtain the numerical solution of the system above (see [16] for details on the scheme and discretization used).



(a) Cancer cells density (b) ECM density

Fig. 1: Initial conditions for ECM and total cancer cell density.

We consider the following treatment schedules:

- **Strategy 0**: No therapy. Simulation of the evolution of (7) for 9 weeks.
- **Strategy 1**: 3 weeks of neoadjuvant chemotherapy, followed by 6 weeks of concurrent chemo- and radiotherapy.

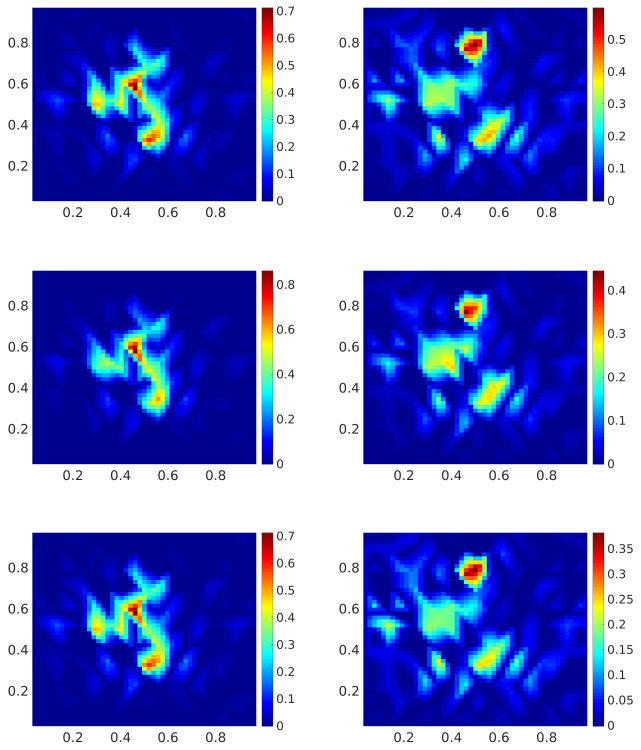


Fig. 2: Strategies 0 (top row), 1 (middle row) and 2 (bottom row). Moving cells m at three (left column) and nine (right column) weeks.

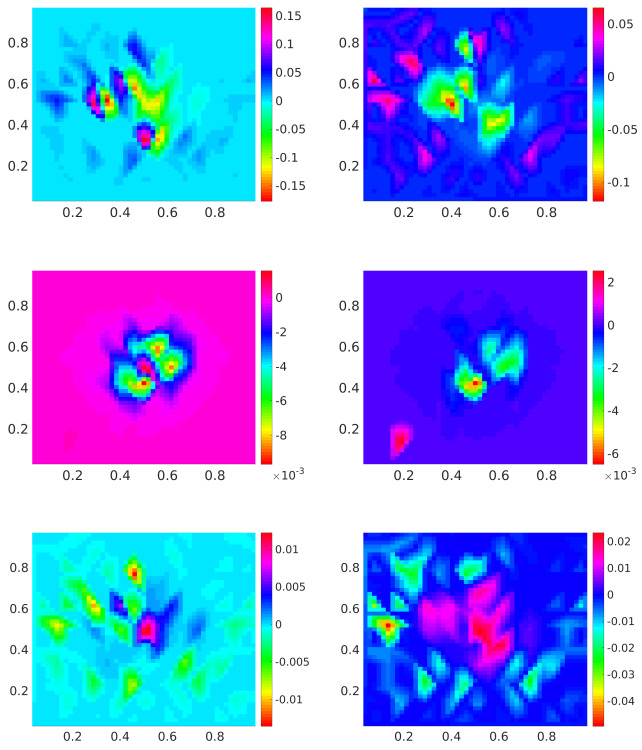


Fig. 3: Difference between Strategy 2 and Strategy 1. Moving cells m (top row), proliferating cells q (middle row) and ECM v (bottom row) at three (left column) and nine (right column) weeks.

- **Strategy 2**: 3 weeks of no therapy, followed by 6 weeks of radiotherapy.

Chemo- and radiotherapy are applied during weekdays with a break for the weekends. We assume a hyperfractionated radiotherapy with a daily dose of 2Gy. The type of chemotherapy under our consideration is motivated by clinical trials NCT01165333 and NCT00689221 at www.clinicaltrials.gov. We omitted chemotherapeutic effect of Cilengitide and Temozolomide directly aiming at cell kill and consider only the inhibiting effect on integrin binding.

V. Discussion

In this paper we considered multiscale model for tumor invasion. This model accounts for heterogeneity of cancer cells with respect to phenotype and, as a consequence, with respect to sensitivity to irradiation treatment. Due to importance of integrins in cancer invasion, these transmembrane receptors became an attractive target in cancer treatment (see e.g., clinical trial NCT01165333 of integrin inhibition drug Cilengitide at www.clinicaltrials.gov). In our model we included an example of such treatment in our therapeutic approach, which had two-fold aim: first, to slow down cancer cell migration by means of inhibiting cell-ECM binding and second, to achieve cell kill by irradiation therapy. By comparing the densities of cancer cells after three weeks of integrin binding inhibition against no treatment in Figure 3 (top and middle row), we can clearly see that inhibition therapy provides better tumor control on periphery (as fewer cells migrated away from tumor center) and worse control in the center. Furthermore, hindered migration causes cell to clump in the center, which in turn causes greater degradation of healthy tissue. Similar results due to application of integrin inhibition can be found in [17].

Therefore, when considering potential treatment strategies aimed at inhibiting cancer cell migration, one must also take into account such a side effect.

We plan to extend the scope of this model by considering action of two or more type of integrins, responsible for activating different cellular functions. We will also consider the effect of cadherins (type of transmembrane receptors, responsible for cell-cell adhesion) and the resulting collective behavior. Furthermore, numerical simulations can be extended to 3-D framework and we can modify the contractivity function in order to take into account the difference between cancer invasion in two and three dimensional setting, since the structure and dimensionality of the extracellular matrix influences both cellular and subcellular dynamics.

REFERENCES

[1] International Agency for Research on Cancer, *World Cancer Report 2014*, 2014.

[2] H. Kraahling, S. Mally, J. Eble, J. Noel, A. Schwab, and C. Stock, "The glycocalyx maintains a cell surface pH nanoenvironment crucial for integrin-mediated migration of human melanoma cells," *Eur. J. Physiol.*, vol. 458, pp. 1069–1083, 2009.

[3] J. Hood and D. Cheresh, "Role of integrins in cell invasion and migration," *Nature Rev. Cancer*, vol. 2, pp. 91–100, 2002.

[4] G. Bendas and L. Borsig, "Cancer cell adhesion and metastasis: selectins, integrins, and the inhibitory potential of heparins," *International Journal of Cell Biology*, vol. 2012, 2012.

[5] A. Huttenlocher and A. Horwitz, "Integrins in cell migration," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 9, p. a005074, 2011.

[6] G. Meral, C. Stinner, and C. Surulescu, "On a multiscale model involving cell contractivity and its effects on tumor invasion," *Discrete and Continuous Dynamical Systems-Series B*, vol. 20, pp. 189–213, 2015.

[7] F. Lefranc, J. Brotchi, and R. Kiss, "Possible future issues in the treatment of glioblastomas: special emphasis on cell migration and the resistance of migrating glioblastoma cells to apoptosis," *Journal of clinical oncology*, vol. 23, no. 10, pp. 2411–2422, 2005.

[8] R. Mangum and I. Nakano, "Glioma stem cells and their therapy resistance," *J Carcinogene Mutagene S*, vol. S1-002, pp. 1–8, 2011.

[9] P.-P. Zheng, L.-A. Severijnen, M. van der Weiden, R. Willemsen, and J. M. Kros, "Cell proliferation and migration are mutually exclusive cellular phenomena in vivo: implications for cancer therapeutic strategies," *Cell Cycle*, vol. 8, no. 6, pp. 950–951, 2009.

[10] G. Meral, C. Stinner, and C. Surulescu, "A multiscale model for acid-mediated tumor invasion: therapy approaches. In print." *Journal of Coupled Systems and Multiscale Dynamics*, 2015.

[11] A. Anderson, M. Chaplain, E. Newman, R. Steele, and A. Thompson, "Mathematical modeling of tumor invasion and metastasis," *J. Theoretical Medicine*, vol. 2, pp. 129–154, 2000.

[12] T. Celiá-assa *et al.*, "Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells," *The Journal of Clinical Investigation*, vol. 122, no. 5, pp. 1849–1868, 5 2012. [Online]. Available: http://www.jci.org/articles/view/59218

[13] V. Evdokimova, C. Tognon, T. Ng, and P. Sorensen, "Reduced proliferation and enhanced migration: two sides of the same coin? Molecular mechanisms of metastatic progression by YB-1," *Cell Cycle*, vol. 8, no. 18, pp. 2901–2906, 2009.

[14] A. Hunt and C. Surulescu, "A multiscale modeling approach to glioma invasion with therapy. Preprint." *Technische Universität Kaiserslautern*, 2015.

[15] J. F. Fowler, "The linear-quadratic formula and progress in fractionated radiotherapy," *The British Journal of Radiology*, vol. 62, no. 740, pp. 679–694, 1989, pMID: 2670032. [Online]. Available: http://dx.doi.org/10.1259/0007-1285-62-740-679

[16] A. Uatay, "Multiscale models for tumor invasion and therapy approaches," M. Sc. thesis, Technische Universität Kaiserslautern, Kaiserslautern, Germany, 2015.

[17] L. Oliveira-Ferrer, J. Hauschild, W. Fiedler, C. Bokemeyer, J. Nippgen, I. Celik, and G. Schuch, "Cilengitide induces cellular detachment and apoptosis in endothelial and glioma cells mediated by inhibition of fak/src/akt pathway," *Journal of Experimental & Clinical Cancer Research*, vol. 27, no. 86, 2008.

Situation criteria as a component of the judgement of human competence

David Van de Water
Technical University of Kaiserslautern
Educational Science
Kaiserslautern, Germany
David.vandewater@sowi.uni-kl.de

***Abstract* — In Pedagogy the precise measurement of human competence is one of the greatest unresolved problems in our time. Required by current developments like the PISA-Studies or the Bologna-Process the competence of human beings is the heart of any educational program. Regarding this in current competence measurement systems situational criteria are not considered sufficiently. But each human action takes place in a specific situation e.g. the current physical environment which influences it inevitably. To judge the competence of a human being precisely it is necessary to observe his performance in specific situations in order to infer to the underlying competence. So in this article the basic idea of a situational based proceeding in the measurement of human competences is presented. To our knowledge there are occasional approaches but no universal model to integrate situation criteria in competence assessment systems.**

Keywords— Competence, Assessment, Situation, Technology, Models

I. MOTIVATION

The Assessment of human competencies is a fundamental requirement to optimize processes of learning, program planning or guidance. [1] Basically there is to distinguish between competence as dispositions of self-organized actions and performance as a direct application of competence in real situations. [15] Thereby, human competence is observable not directly but about its performance. Unfortunately, human actions depend highly on the current situation in which they happen. This phenomenon is pointed out by Khan et al. as follows: "In a hypothetical situation where two individuals have exactly same level of training and performance on a particular skill in a particular context, even then their performance might differ when facing unusual circumstances based on their inherent personality traits." [2, p. 2] Following from this it is very difficult to conclude from some single observations of human performance to the competence lying behind it with respect to a countless number of different situations which might influence it. So, on the one hand there is the requirement that any judgement of human competence has to consider the characteristics of the current situation to get a more precise and realistic view on this. On the other hand there is the challenge to assess human performance in all possible situations apart from this to describe them at all. As a first step to overcome this gap this work aims to provide a theoretical model of the relation between situational characteristics and human performance and to elicit potentialities to assess those conditions in order to encourage the complex process of competence measurements. In this way a first view on potential components of this model is presented.

II. RELATED WORK

The approach to consider situation criteria in measurement of competence is especially located in medicine. Boulet et al. pointed out that it is important "[…] to include a large sample of clinical situations in simulation-based assessments, as well as large numbers of judges." [16, p. 125] Burg et al. provide a description of competence including the specific situation a medical task has to be done in. [17] In addition, away from Pedagogy there are several approaches in relation to assess human performance by using technology. So this is not new or even utopistic. Quite the contrary it is part of the current world and used in different areas. Researchers at Fraunhofer Institute for Integrated Circuits (IIS) developed a complex sensor-based system to record the performance of soccer players in which every single player is equipped with several sensors to get e. g. data on their step length, step frequency or their body orientation in relation to the game situation and ball position. [13] These data are used to optimize the performance of the players by using data analysis in real-time. Hadjakos outlines in his doctoral thesis a sensor-based analysis of finger movements of piano students in order to compare the current movements of the students with the required movements by the teacher. [14] In principle, a main objective of this work is to find a method to transfer these techniques to pedagogical and vocational contexts.

III. RESEARCH QUESTIONS

For the most part the question appears to what extent information about the situation of a human activity might be used in methods of human performance assessment. Therefore, it is important to establish criteria which describe individual und situational properties in an adequate manner. Then, techniques are searched for with which those data can be captured objectively and holistically. Finally, a way to evaluate a large number of data has to be investigated.

IV. Theoretical Background

The complexity of the intention to assess human behavior in real environments call for an interdisciplinary approach in which knowledge of different subject areas is combined to capture and interpreted appropriate data. From the pedagogical point of view it has to be established which kind of competences are required by individuals to fit them to later professional activities. For this, educational standards are defined which know should be extended by situational aspects. To reach individual und situational characteristics psychological knowledge is available. The personality psychology is working on individual properties [3] and the educational psychology engages in situational properties. [4] To capture preferably objective and holistic data on human behavior sensor data which are integrated in great quantities into current commercially available devices like smartphones or wearable technologies might be promising. Moreover, these obtained data are evaluable by algorithmic processes e. g. using pattern recognition and artificial intelligence.

A. Definition and assessment of competence

With Erpenbeck competency is defined as the capability to act successfully in unexpected, open or sometimes chaotic situations in a self-organized and creative manner. [5] Immediately it follows that judging individual competence has to include environment characteristics as Kröll reaffirm: "In contrast to personality features, competences relate to a high degree to concrete situations which call for their application. The possession of competences can only be measured in context." [6, p. 547] The reasoning is that there are factors which influence the way someone retrieve his competence in concrete situations. Surely, this applies to any subject area and to any kind of competence. But in some areas the influence of the situation on human competence is highly important or even critical. A surgeon who is not able to operate in situations where e. g. the patient is in a critical condition might not be suitable to work in this job. A pilot who cannot control his air plane in a thunderstorm or with technical defects is not suitable, too. Therefore, these examples illustrate that expanding judging of competence by situation criteria is profitable in relation to decide if someone is able to show his competence even in desired environments.

B. Individual and situational characteristics

Knowledge of Differential Psychology provides a pool of personality traits which outline the diverse possibilities to describe individuals [3]: e. g. physical characteristics (body type, attractiveness), rating dispositions (set of values, attitude), capabilities (intelligence, creativity), action properties (needs, motivation) or temperament (extraversion, introversion). Speaking about situations has two dimensions [7]: First, an environment can be considered in an objective way by describing external conditions like physical data. Second, there is a subjective interpretation of situations by affected individuals. Both perspectives might differ considerably. Educational psychologists worked out several criteria to outline context criteria. [4] For example, there are genetical (body height, body shape) sociodemographic (age, income), individual (interests, capability) or socially (education, family) properties.

C. Sensor based data collection and algorithmic analysis

Judging individuals by individuals is always subjective. Observations of educational actives are influenced by their representations and thus there are selective. [1] Following from this, a rating of one identical situation will surely differ from one person to another. Therefore, an approach is needed, in which situation criteria are collected and evaluated as objective as possible. From a technical point of view sensors are predestined to assume this task because their recordings of environmental data are resistant over time. The sensors of devices like smartphones or wearable technologies are very varied in relation to their kind of recorded data. For example, devices with android systems provide motion sensors (accelerometers, gravity sensors, gyroscopes, rotational vector sensors), environmental sensors (barometers, photometers, and thermometers) and position sensors (orientation sensors, magnetometers). [8] Obviously, those sensors provide a large amount of data which cannot be handled by individuals. Algorithms are required which evaluate these data efficiently and in an adequate manner. It is desirable that algorithms also support for interpreting data because sensors are able to record data of many individuals and their situation at the same time. Methods of pattern recognition and artificial intelligence are basing on mathematical knowledge which deals with uncertainty. A sensor based system which supports an interpreting of recorded data has to decide in which category each kind of data belongs to. For example, there is an individual with an increased body temperature which can mean that this individual is ill, stressed or only influenced by a high room temperature. Using fuzzy sets is one way to deal with this difficulty. [9] The interpretation of temperature by individuals is subjective. The same temperature can be seen as low, high or very high. With fuzzy sets it is possible to decide automatically to which category a temperature value is presumably belonging to. In Fig. 1 a fuzzy partitioning of temperature is illustrated. With a degree of 0.8 the temperature of 18° C belongs to the category normal and with a degree of 0.2 to the category low.

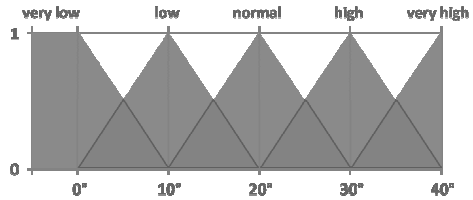


Fig. 1 Fuzzy-Sets to classify temperature (inspired by [9, p. 305])

A System becomes able to classify a data set in an uncertain context by considering its degree of belonging to a category.

V. Methodological Approach

A system for assessing human performance is typically composed of components which are pointed out by Stiggins: "The basic ingredients of a performance assessment may be described in three parts […]: (1) the specification of a perfor-

mance to be evaluated, (2) the development of exercises or tasks used to elicit that performance and (3) the design of a scoring and recording scheme for results.". [10, p. 2-3] In dependence on this, the first necessary step is to elicit and classify situational criteria in order to integrate them into usual descriptions of competence. Secondly, available sensor data have to be explored as well as systematized and it has to be examined in which way they are able to reflect characteristics of individuals in real environments. Thirdly, the required components of a situational based performance assessment system will be carved out. The model should supply the basis for implementing a sensor-based system and it should be applicable to arbitrary domains.

VI. Ethical Concerns

Sensors provide a comprehensive view on personal traits by recording detailed or even intimate data about e. g. biological or location-based parameter. At all times and locations human beings are observed and an unimaginable number of data are collected to catch a deep view on their activities, interests and habits. It is estimated that the digital universe is growing 40% a year and that in 2020 its data volume will reach 44 trillion gigabytes. [11] Thus, developing a sensor-based assessment system has to be controlled by discussing ethical issues. Radermacher refers to the character of violence which data collection could have to individuals. [12] It has to be reflected under which conditions a sensor-based system is ethically justifiable and at what point it would be unacceptable for affected individual's e. g. with respect to data protection.

References

[1] Ingenkamp, K., & Lissmann, U. (2008). Lehrbuch der pädagogischen Diagnostik (6., neu ausgestattete Auflage). Pädagogik. Weinheim, Basel: Beltz.

[2] Khan, K., & Ramachandran, S. (2012). Conceptual framework for performance assessment: competency, competence and performance in the context of assessments in healthcare - deciphering the terminology. Medical Teacher, 34(11), 920–928.

[3] Asendorpf, J. B., & Neyer, F. J. (2012). Psychologie der Persönlichkeit (5., vollständig überarbeitete Auflage). Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg; Imprint: Springer.

[4] Mienert, M., & Pitcher, S. (2011). Pädagogische Psychologie: Theorie und Praxis des Lebenslangen Lernens. Basiswissen Psychologie. Wiesbaden: VS Verlag für Sozialwissenschaften.

[5] Erpenbeck, J. (2012). Zwischen exakter Nullaussage und vieldeutiger Beliebigkeit: Hybride Kompetenzerfassung als künftiger Königsweg. In J. Erpenbeck (Ed.), Kompetenzmanagement in der Praxis: Bd. 6. Der Königsweg zur Kompetenz. Grundlagen qualitativ-quantitativer Kompetenzerfassung (pp. 7–42). Münster, New York: Waxmann.

[6] Kröll, M. (2009). E-Assessment: A Suitable Alternative for Measuring Competences? In J. A. Jacko (Ed.): Vol. 5613. Lecture Notes in Computer Science, Human-Computer Interaction. Interacting in various application domains. Proceedings, Part IV (pp. 543–550). Berlin [u.a.]: Springer.

[7] Fröhlich, W. D. (2010). Wörterbuch Psychologie (27. überarbeitete und aktualisierte Auflage). Dtv: Vol. 34625. München: Deutscher Taschenbuch-Verlag.

[8] Google Inc. Google Android Sensors Overview. Retrieved from http://developer.android.com/guide/topics/sensors/sensors_overview.html, viewed on 10.12.2015.

[9] Nahrstedt, H. (2005). Algorithmen für Ingenieure - realisiert mit Visual Basic: Eine anwendungsorientierte Einführung - Problemanalyse und Lösungsweg anhand konkreter Beispiele. Wiesbaden: Vieweg.

[10] Stiggins, R. J. (1995). Sound Performance Assessments in the Guidance Context (ERIC Digest).

[11] Turner, V., Reinsel, D., Gantz, J. F., & Minton, S. (2014). The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things: White Paper. Retrieved from IDC Analyse the future website: http://idcdocserv.com/1678, viewed on 13.12.2015.

[12] Radermacher, F. J. (2015). Algorithmen, maschinelle Intelligenz, Big Data: Einige Grundsatzüberlegungen [Algorithms, machine intelligence, big data : General considerations]. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz, 58(8), 859–865.

[13] Fraunhofer-Institut für Integrierte Schaltungen IIS. (n.d.). RedFIR® – Echtzeit Performance Analyse. Retrieved from http://www.iis.fraunhofer.de/content/dam/iis/de/doc/LN/Referenzprojekte/chip%20im%20ball%20bei%20redfir.pdf, viewed on 13.12.2015.

[14] Hadjakos, A. (2011). Sensor-Based Feedback for Piano Pedagogy (Dissertation). Technische Universität Darmstadt, Darmstadt. Retrieved from http://tuprints.ulb.tu-darmstadt.de/2797/, viewed on 13.12.2015.

[15] Erpenbeck, J. (Ed.). (2007). Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis (2., überarbeitete und erweiterte Auflage). Stuttgart: Schäffer-Poeschel.

[16] Boulet, J. R., & Swanson, D. B. (2004). Psychometric Challenges of Using Simulations for High-Stakes Assessment. In W. F. Dunn (Ed.), Simulators in critical care education and beyond (pp. 119–130). Mount Prospect, IL: Society of Critical Care Medicine.

[17] Burg, F. D., Lloyd, J. S., & Templeton, B. (1982). Competence in medicine. Medical Teacher, 4(2), 60–64.

Euler-Lagrange simulation of bubble hydrodynamics using a stochastic approach

Andreas Weber, Hans-Jörg Bart
Chair of Separation Science and Technology
Technische Universität Kaiserslautern
Kaiserslautern, Germany
www.uni-kl.de/tvt/

***Abstract*— For a better understanding of the dynamics inside bubble columns, several different types of simulations have been developed in the last three decades. Many computational approaches lack the detail on the bubble scale. Thus, an Euler-Lagrange approach is used to improve the prediction of bubble hydrodynamics due to the interaction of bubbles. Stochastic modeling of interaction between bubbles enables a fast solution even for a high number of bubbles.**

Keywords—Euler-Lagrange; bubble column; CFD; break-up; coalescence

I. INTRODUCTION

Bubble columns are multipurpose reactors, used for a wide set of operations in chemical and biological applications. The main design aim of a bubble column is to adjust the contact area between a liquid and a gas phase. Mass transfer and reactions will take place at the interface, which is why knowledge about the interactions between gas (bubbles) and liquid phase is essential. A large interface area, corresponding to small bubble diameters, leads to higher mass transfer but will also increase the retention time and the volume fraction of the gas in the reactor. Excess gas should not stay in the reactor after the reaction has finished, a long retention time is thereby not wanted. Thus, an optimal setting would avoid very large and very small bubbles [1].

One possibility of obtaining knowledge about the events inside a bubble column is by simulation. Bubbly flows in general can be simulated in various ways, from resolving detailed single bubble behavior up to large-scale reactors with billions of bubbles inside. Most of them are based on numerous different break-up and coalescence kernels (B&C) that have been developed in the last few years [2–5]. Those kernels were developed to describe gas bubbles in a flow merging together or splitting into smaller ones. Many computational approaches mainly concentrate on the bubble size distribution (BSD) and lack a level of detail when it comes to hydrodynamics on the bubble scale. Those models are using averaged values for flow and bubble position parameters, making them fast but less detailed [6, 7]. The reverse is the direct numerical simulation (DNS), where bubble and surrounding liquid are represented in full detail. Nevertheless, the DNS needs immense computational effort and is thereby limited to a small number of bubbles [8]. Thus, an Euler-Lagrange (EL) approach should be used to improve the prediction of B&C due to the local flow at bubble size level, while it is still possible to simulate a whole reactor. It can offer a more realistic solution of bubbly flows than an Euler-Euler (EE) approach does and is only limited due to its greater computational effort [9, 10].

The aim of this study is to validate an EL approach for a bubble column and to further improve the simulation's level of detail by merging and comparing existing models. Coupling is achieved by calculating balancing forces between the continuous phase and the dispersed bubbles. Collision of bubbles is modeled by a stochastic equation, reducing the computational effort in comparison to a deterministic collision model. Other effects on the behavior of bubbles are also modeled using a stochastic approach, such as break-up, collision and coalescence of bubbles. Implementation of new stochastic models is simplified, because no differential equations have to be solved, when using the EL approach. Thus, a variety of models for B&C have been adapted from literature. Through a high number of bubbles and stochastic interactions between the bubbles the resulting BSD is obtained similar to a Monte-Carlo simulation.

The simulation has been developed based on the papers of [11] and [12] using the OpenFOAM [13] framework. Kernels for break-up and coalescence have been adopted from those studies, which are namely the models of [14], [15], [16] and [17]. Several models for the drag force have been added to the existing ones in OpenFOAM, e.g. from [18], [19] and [20]. The collision model, taken from the original OpenFOAM code, is based on a stochastic process developed by [21]. The simulation has been validated using the case of [22].

II. MODELING

The EL simulation describes the link between the large-scale EE model and the detailed DNS. While the EE model resembles the bubbles as a density distribution, in the EL model every single bubble is calculated as a point volume acting under Newtonian dynamics while the movement of the surrounding fluid is solved via Navier-Stokes equations on an Eulerian grid.

This work was developed in collaboration with the DFG-GrK 1932 "Stochastic Models for Innovations in the Engineering Sciences"

A. Liquid Phase Hydrodynamics

The motion of the continuous phase is calculated through modified Navier-Stokes equations. Given the continuous velocity u_c, the pressure p, the density ρ, the viscosity μ and the source term f, it has to be solved on the Eulerian grid. The source term embodies the forces of the bubbles and acts as a coupling between the different phases.

$$\rho\left(\frac{\mathrm{d}u_c}{\mathrm{d}t}+(u_c\nabla)u_c\right)=-\nabla p+\mu\Delta u_c+f \tag{1}$$

A premade solver from OpenFOAM - namely sprayFoam - has been modified to resolve the bubbly flows. Within this model, the continuous fluid is assumed to be a compressible homogenous mixture, allowing later implementation of chemical reactions and heat transfer. Nevertheless, this work concentrates on EL simulation of a non-reactive, isothermal bubble column filled with water, which will be defined incompressible. A first-order numerical scheme has been set for the solution of the continuous phase, giving a stable but slow solution compared to second order schemes. Since the bubble size dictates the grid size for the EL simulation, a coarse mesh is used and the solution effort of the continuous phase equations is a minor problem. Turbulence is computed with RAS modeling using the standard k-epsilon model [23], which requires significantly less computational effort than solving for a LES approach.

B. Bubble dynamics

The bubbles are assumed to have their mass concentrated at its center of gravity at the position x_B. Mass can change due to break-up, coalescence, chemical reaction or mass transfer. In this work, chemical reaction and mass transfer have been turned off. The position is calculated using the velocity u_B of each bubble based on Newton's equation of motion, where Δt determines the elapsed time:

$$\Delta x_B = u_B\,\Delta t \tag{2}$$

Forces acting the bubble change will change its velocity. Here m_B resembles the bubble's mass and ΣF stands for the Sum of all acting forces on the single bubble.

$$m_B\,\Delta u_B = \Delta t\,\Sigma F \tag{3}$$

The sum of forces ΣF consists of the buoyancy and weight force F_B, the drag force F_D, the lift force F_L, and the virtual mass force F_{VM}. Here the subscripts B and C stand for the bubble and the continuous phase accordingly. Furthermore, g denotes the gravitational acceleration, r stands for densities, u stands for velocity and d_B for the bubble's diameter. Model constants are denoted with C_i.

$$\Sigma F = F_B + F_D + F_L + F_{VM} \tag{4}$$

These forces are in particular:

$$F_B = m_B g\left(1-\frac{\rho_C}{\rho_B}\right) \tag{5}$$

$$F_L = m_B\frac{\rho_C}{\rho_B}C_L\left(u_B-u_C\right)\times\nabla\times u_C \tag{6}$$

$$F_D = \frac{3}{4}C_D\frac{m_B\rho_C}{d_B\rho_B}\left|u_C-u_B\right|(u_C-u_B) \tag{7}$$

$$F_{VM} = -C_{VM}\rho_C V_B\left(\frac{\mathrm{D}_B u_B}{\mathrm{D}t}-\frac{\mathrm{D}_C u_C}{\mathrm{D}t}\right) \tag{8}$$

In (8), V_B stands for the bubble's volume and $\mathrm{D}_i/\mathrm{D}t$ denotes the material derivative, meaning that the derivative is made while following the bubble. Especially the drag force model is a critical parameter, because it influences not only the bubble rise velocity but also the volume fraction. The drag force coefficient C_D in (7) can be modeled in different ways, using different model assumptions, e.g. as a function of the Eötvös number [18], Reynolds-Number or both of them [20]. Other models are specifically designated to calculate the shape of the bubble and its influence on the drag force [19, 24]. For validation of the simulation, several of those different drag models have been implemented and compared. The lift force pushes the bubbles perpendicular to the slip velocity, its coefficient C_L in (6) is calculated using the model of [19]. The coefficient for virtual mass in (8) is set to $C_{VM} = 0.5$ according to [7]. The virtual mass is used to account for the liquid film which directly surrounds the bubble and has to be accelerated together with the bubble.

C. Phase coupling

Often, particle flows are simulated using only two-way coupling, which results in a less detailed but faster solution. Particles are seen as passive and do not change the continuous flow. This is the case for particles with a small Stokes number (St < 1). Since only the buoyancy force of bubbles drives the flow inside a bubble column, four-way coupling is needed fundamentally. This is achieved by coupling the phases in both directions, from bubble to continuous phase and vice versa.

The source term f in the Navier-Stokes equation (1) resembles all forces coming from the bubbles that act on the continuous phase. Here, only the *coupled* forces are taken into account while the *non-coupled* forces are left out, like the virtual mass force F_{VM}. For each computational cell the forces of the containing bubbles are summed up and divided by the cell's volume. Thus, each bubble can be seen as a source of momentum. Coupling in the other direction happens while calculating the bubble forces itself, because most forces depend on the slip velocity between bubble and continuous phase, see (6) and (7).

D. Stochastic modeling

In order to get a physically realistic picture of bubble movement in a reactor, several stochastic models have been used. Since bubble movement itself shows a random character it is natural to apply this to the numerical model. Thus, stochastic algorithms have been used for the collision, break-up, coalescence and stochastic diffusion of bubbles.

The stochastic collision equation by [21] has been used, giving the collision probability P_{coll}.

$$P_{\text{coll}} = \frac{\pi}{4}\left(d_{\text{B1}} + d_{\text{B2}}\right)^2 \left|u_{\text{B1}} - u_{\text{B2}}\right| \frac{n_{\text{B}}}{V_{\text{cell}}}\ \Delta t \tag{9}$$

Where d_{Bi} are the corresponding bubble diameters, u_{Bi} their velocities, V_{cell} the volume of the corresponding cell and n_{B} the number of bubbles inside that cell. For estimation of a collision, a uniformly distributed random number $\xi \sim U(0,1)$ is generated for each pair of bubbles positioned in the same cell for each time step. This random number is compared to the calculated collision probability P_{coll}, giving the constraint for collision.

$$\mathrm{P}_{\text{coll}} > \xi \tag{10}$$

When collision occurs, the coalescence probability P_{coal} will be evaluated using one of the models of [12, 15, 21]. Each model has its own principle of calculation, e.g. surface energy consideration or evaluation of film rupture time, and it has not been clarified which one leads to the most realistic coalescence behavior [25]. Regardless of the model, the comparison of P_{coal} with a random number $\xi \sim U(0,1)$ will again decide whether coalescence occurs or is neglected, analog to (10).

If the constraint for coalescence is accepted, the corresponding bubbles will merge to one larger bubble, with respect to the conservation of mass, momentum and energy. If rejected, bubbles will perform a fully elastic collision, bouncing off each other. Several simplifications have been made here, e.g. the bubbles' direction of movement is not concerned (only the magnitude), which can lead to collision and coalescence even though the bubbles are moving away from each other. Bubbles can interact only if their centers of gravity are located in the same cell of the Eulerian grid. In return, this simplifications lead to a significantly lower computation time than for deterministic approaches, where the exact trajectories of each bubble have to be calculated.

The mechanism of bubble break-up can be explained by looking at the dynamic forces acting on the bubbles. Pressure fluctuations caused by turbulent eddies inflict deformations of the bubbles and, in case they are strong enough, can tear them apart. Considering the turbulent eddies as an isotropic random event, it is natural to assume that break-up of bubbles can be modeled using a stochastic approach. For that, the break-up frequency ω_{b} is calculated using one of the models of [12, 14, 15] at each time step for each bubble. The break-up condition is then evaluated using a random number $\xi \sim U(0,1)$.

$$\omega_{\text{b}} \Delta t > \xi \tag{11}$$

When break-up occurs, a daughter bubble will be created with its properties copied from the mother bubble. The diameter of both will be evaluated with respect to the conservation of mass, but with its proportion being chosen randomly. The randomly generated mass fraction f defines the new bubble sizes.

$$\begin{aligned} m_{\text{daughter}} &= m_{\text{mother}}\, f \\ m_{\text{mother,new}} &= m_{\text{mother}}\ (1-f) \end{aligned} \tag{12}$$

The mass fraction f can be a random number of different distributions, e.g. normal or beta distributed, determined by the corresponding model assumptions. In all cases, the mass fraction f lies in the interval between 0 and 1. Additionally, it has been limited further, to ensure computability.

$$f_{\text{limit}} < f < 1 - f_{\text{limit}} \tag{13}$$

Where $f_{\text{limit}} = 0.01$, which prevents the formation of extremely small bubbles. If the chosen random number f lies outside the desired interval, it will simply be drawn again.

To achieve an even better natural bubble movement, stochastic diffusion of the bubbles has been implemented. The model comes with the OpenFOAM source code and provides particle movement based on the Brownian motion coupled to the turbulent energy. Thereby, bubbles in areas of high turbulence are randomly trembling around resolving a more chaotic movement. In sum, this random motion turns out as a diffusive effect, where the bubbles will spread out faster with increasing turbulence. The turbulent velocity u_{turb} is thereby a function of the turbulent kinetic energy k. The velocity's magnitude follows a normal distributed random value, the direction is spatially uniform distributed.

$$|\, u_{\text{turb}} \,| = N\left(0; \sqrt{\frac{2}{3}k}\right) \tag{14}$$

The stochastic modeling of B&C yields a simulation of the BSD without having to solve a differential equation, which provides the opportunity to switch between different B&C easily or to use more complex dependencies, e.g. on mass transfer. Using the EL simulation also allows usage of force, breakup and coalescence models regarding single bubble shape and mass transfer situation. The complexity of such a problem simplifies due to the independence of the stochastic steps, which is the main advantage of the presented EL approach, similar to Monte-Carlo simulations.

III. Simulation Setup

Like several other publications [11, 20, 26], the *Deen case* [22] was used as a reference to validate the simulation data of this work. The setup is described as a rectangular bubble column of 15 × 15 cm^2 base and a filling height of 45 cm. It is

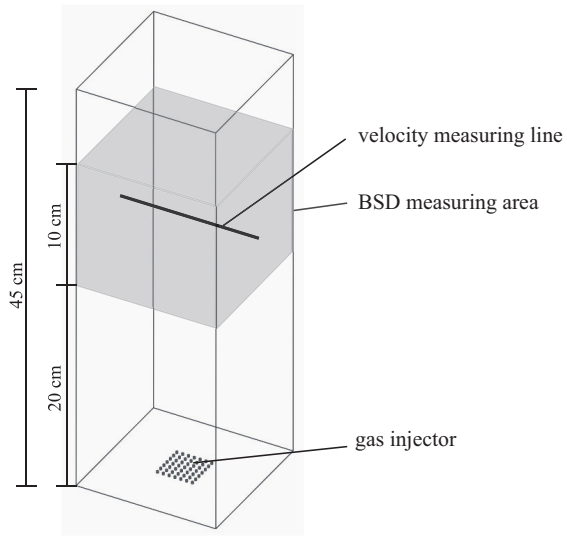


Fig. 1. Scematic view of the bubble column (*Deen case*) and measuring positions.

filled with water and air is injected trough 7 × 7 circular holes at the bottom end. The holes are positioned on a grid with 3.75 × 3.75 cm^2 area in the middle of the column. The superficial gas velocity is $u_{g,s} = 4.9$ mm s^{-1}, the initial bubble diameter at the inlet holes is $d_B = 4$ mm. The liquid velocity, which has been measured in the original experiment, will be compared with the simulation data. Second validation parameter is the BSD measured by [27] using interferometric particle imaging technique (PIV). Also, the data has been compared to other similar simulative work [11].

A. Mesh geometry

Particles/bubbles must not be larger than the computational cell they are in. Thus, the cell length is determined by the maximum bubble diameter, which is typically $d_{B,max} = 10$ mm. This leads to a mesh with a solution of 15 × 15 × 45 cells, like it has been used in comparable simulative setups [11]. Boundary conditions for the walls and bottom have been set to no-slip, the top boundary serves as an out/inlet with free-slip condition. The surface area will not be computed, the whole domain is filled with the continuous phase. Continuous fluid can flow in and out at the upper boundary but bubbles will only be able to leave the domain at this point. Analysis of the turbulence state near the walls have shown, that a finer mesh resolution would be suitable but have been omitted to maintain comparability with other studies.

TABLE I. DRAG FORCE MODEL EQUATIONS (EXCERPT)

Drag force model equations		
Name	***Model assumption***	***ref.***
Tomiyama	Deformed bubble (ellipsoidal shape)	[19]
Ishii & Zuber	Function of Eötvös number	[18]
Roghair	Function of Eötvös and Reynolds number	[20]
Tomiyama contaminated	Bubble in tap water (impurities)	[24]
Schiller & Nauman	Bubble as a solid sphere	[29]
B&C model equations		
Lee	Viscous thinning and film rupture Maxwell distributed turbulent energy	[14]
Coulaloglou & Tavlarides	Viscous thinning only Normal distributed turbulent energy	[15]

B. Model selection

Different models for drag force, breakup and coalescence have been implemented. Additionally the models itself contain various regulating parameters that have to be adjusted to the present physical system. The combination of models and parameters is just about endless, although not every combination fits the needs. The most promising combinations were selected, some of them adapted from literature, to fit the experimental data. In this study, a small selection of those will be presented. The different drag force and B&C model equations are shown in Table I. Furthermore, simplified model assumptions are given for each model. Later on, the names given in the table will be used in the results and discussion.

IV. Results

At the initiating time step, no movement and no bubbles are present in the bubble column. Similar to experimental observation, the first few bubbles injected will form a characteristic bubble mushroom (see Fig. 2). After a period of 20 seconds, the flow gets to a stable condition and measurements can be started. In almost every simulation case, the bubble plume was slowly oscillating from side to side. Anyhow, a main flow pattern could be noticed, showing an

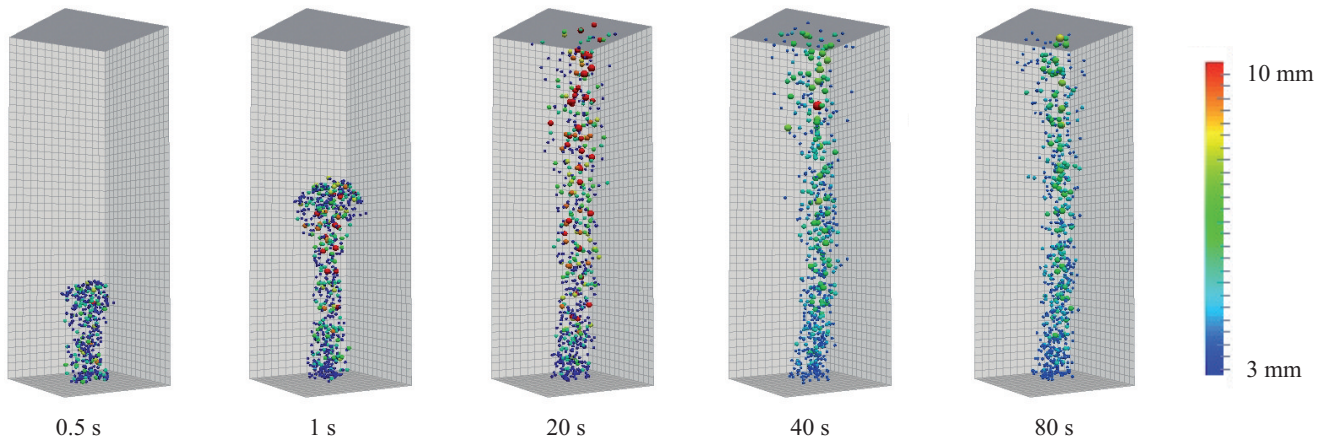


Fig. 2. Snapshots of the simulation at different points in time. B&C model: Lee [14], drag model: Ishii & Zuber [18].

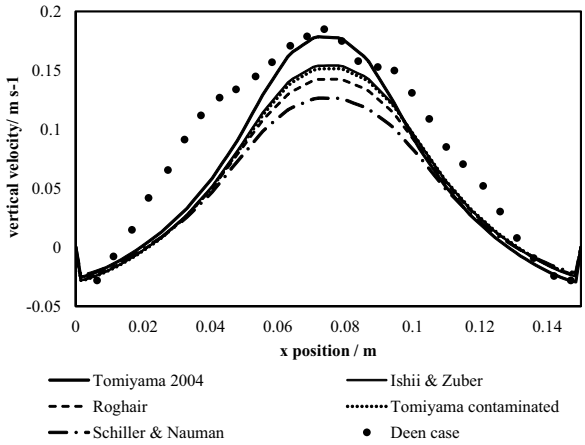


Fig. 3. Vertical velocity of the coninuous phase. See Fig. 1 for measuring positions. The different drag force models are resembled by the lines, points show the experimental data.

upward flow in the middle and downward flow at the walls. This can be found with all drag models and in the experiment, as well (see Fig. 3). It should be mentioned that all results were taken trough time interval averaging, which has not been done for the experimental values of the *Deen case*. Comparison of drag models indicates an advantage of the more complex models. The Schiller & Nauman [28] model, resolving the bubbles as solid spheres, shows significantly worse results compared to the detailed Tomiyama [19] model, which has the best congruency with the experimental data. Still, all simulations differ from the experiments, especially at the outer area of the bubble plume. This could be an indicator for insufficient diffusion modeling, as it seems that the bubbles have to spread out more widely from the middle. In addition, this could be caused by the used RAS model, because comparable studies based on LES turbulence models showed better results [11]. Expanding the RAS model with bubble induced turbulence (BIT) calculations could also improve the simulation without the computational cost for LES modeling [29].

The BSD obtained by the simulation have been compared to the experimental data and other simulative work (s. Fig.4). The obtained BSD (s. Fig. 4C) shows similarities to the other simulative work (s. Fig. 4B), which seems legit, because the same B&C models have been used. Nevertheless, our own values show a slightly higher bubble diameter, the width and shape of the distribution is almost the same. This could be explained due to differences in the turbulence models that have been used. In summary, it could be proofed that the software is capable of simulating a BSD, just like it has been done in comparable simulation studies. Comparison of the computation duration shows a significant benefit. While the reference [11] tells about a duration of several days, the simulation presented here can do the same calculation within 12 to 24 hours. Unfortunately, the information given in the reference does not tell much detail about the configuration of the machine used, thus a detailed comparison is not possible.

Comparison to experimental data still show a greater deviation. The experimental case (s. Fig. 4A) shows a significantly smaller bubble diameter and the width of the distribution is also much smaller. It should be possible to adjust the model parameters to find a better fit to the experimental data, because so far only the parameters from the reference have been used in the simulation. That is why the reference simulation data also shows an overrated BSD.

In summary, it has been shown that the software is capable of simulating a bubble column in a fast and detailed way using the EL approach. The main advantage lies in the great variety of selectable models and the calculation speed. Reproduction of reference data has not been a problem and with the right adjustment of the B&C model parameters, it should not be a problem, to gain a realistic solution for the BSD. In future, more models concerning the turbulence and the detail of bubble shape mapping will be implemented, which should give a further increased detail level of a EL simulation.

ACKNOWLEDGMENT

We want to thank the research training group (DFG GrK 1932) and all its members, through which we could obtain extensive knowledge about stochastics and random number generation. More thanks got to the Technomathematics group (University of Kaiserslautern), especially Anne Meurer, who was a great assistance in the algorithm development.

REFERENCES

[1] K. Akita and F. Yoshida, "Bubble Size, Interfacial Area, and Liquid-Phase Mass Transfer Coefficient in Bubble Columns," *Ind. Eng. Chem. Proc. Des. Dev,* vol. 13, no. 1, pp. 84–91, 1974.

[2] J. Lasheras, C. Eastwood, C. Martínez-Bazán, and J. L. Montañes, "A review of statistical models for the break-up of an immiscible fluid immersed into a fully developed turbulent flow," *International Journal of Multiphase Flow*, vol. 28, no. 2, pp. 247–278, 2002.

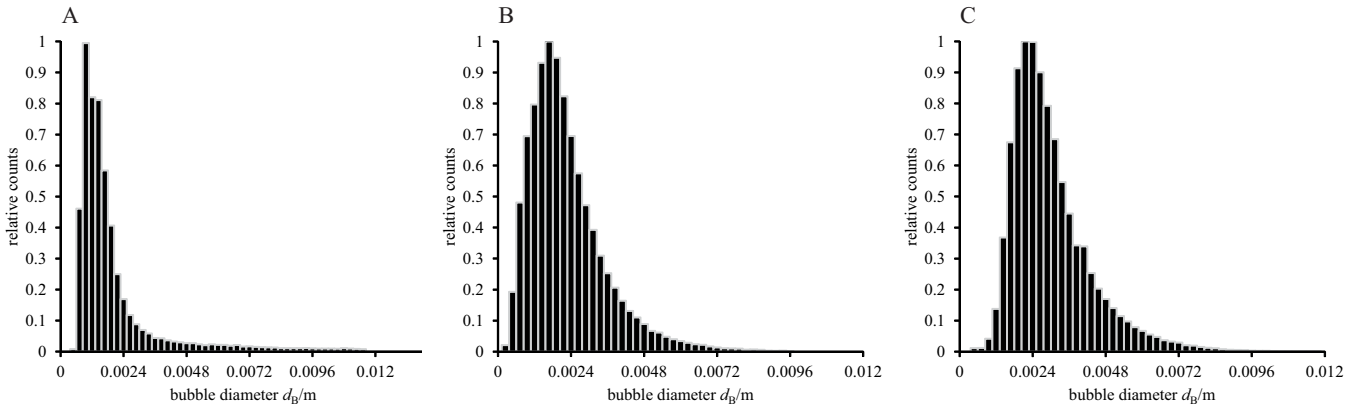


Fig. 4. Comparison of the bubble size distribution (BSD). A: experimental data from [27]; B: simulation data from [11], B&C models from Coulaloglou & Tavlarides; C: own simulation data, B&C models from Coulaloglou & Tavlarides. All measurements done in the predefined area (see Fig. 1).

[3] Y. Liao and D. Lucas, “A literature review of theoretical models for drop and bubble breakup in turbulent dispersions,” *Chemical Engineering Science*, vol. 64, no. 15, pp. 3389–3406, 2009.

[4] Y. Liao and D. Lucas, “A literature review on mechanisms and models for the coalescence process of fluid particles,” *Chemical Engineering Science*, vol. 65, no. 10, pp. 2851–2864, 2010.

[5] C. Martínez-Bazán, J. Rodríguez-Rodríguez, G. B. Deane, J. L. Montañes, and J. Lasheras, “Considerations on bubble fragmentation models,” *J. Fluid Mech,* vol. 661, pp. 159–177, 2010.

[6] A. Lapin and A. Lübbert, “Numerical simulation of the dynamics of two-phase gas—liquid flows in bubble columns,” *Chemical Engineering Science*, vol. 49, no. 21, pp. 3661–3674, 1994.

[7] E. Delnoij, F. Lammers, J. A. Kuipers, and W. van Swaaij, “Dynamic simulation of dispersed gas-liquid two-phase flow using a discrete bubble model,” *Chemical Engineering Science*, vol. 52, no. 9, pp. 1429–1458, 1997.

[8] W. Dijkhuizen, M. van Sint Annaland, and J. A. Kuipers, “Direct numerical simulation of the drag force in bubble swarms,” *6th International Conference on Multiphase Flow*, 2007.

[9] S. Laín, D. Bröder, M. Sommerfeld, and M. F. Göz, “Modelling hydrodynamics and turbulence in a bubble column using the Euler–Lagrange procedure,” *International Journal of Multiphase Flow*, vol. 28, no. 8, pp. 1381–1407, 2002.

[10] M. F. Göz, M. Sommerfeld, and S. Laín, “Instabilities in Lagrangian tracking of bubbles and particles in two-phase flow,” *AIChE J,* vol. 52, no. 2, pp. 469–477, 2006.

[11] M. Grünewald, H.-R. Lausch, M. Schlüter, M. C. Gruber, S. Radl, and J. G. Khinast, “Coalescence and Break-Up in Bubble Columns: Euler-Lagrange Simulations Using a Stochastic Approach,” *Chemie Ingenieur Technik*, vol. 85, no. 7, pp. 1118–1130, 2013.

[12] A. Vikhansky and M. Kraft, “Modelling of a RDC using a combined CFD-population balance approach,” *Chemical Engineering Science*, vol. 59, no. 13, pp. 2597–2606, 2004.

[13] OpenFOAM, *OpenFOAM. The Open Source CFD Toolbox. User Guide*: Free Software Foundation, Inc, 2014.

[14] C.-H. Lee, L. Erickson, and L. Glasgow, “Bubble breakup and coalescence in turbulent gas-liquid dispersions,” *Chemical Engineering Communications*, vol. 59, no. 1-6, pp. 65–84, 1987.

[15] C. Coulaloglou and L. Tavlarides, “Description of interaction processes in agitated liquid-liquid dispersions,” *Chemical Engineering Science*, vol. 32, no. 11, pp. 1289–1297, 1977.

[16] M. Sommerfeld, “Validation of a stochastic Lagrangian modelling approach for inter-particle collisions in homogeneous isotropic turbulence,” *International Journal of Multiphase Flow*, vol. 27, no. 10, pp. 1829–1858, 2001.

[17] C. Anh Ho and M. Sommerfeld, “Modelling of micro-particle agglomeration in turbulent flows,” *Chemical Engineering Science*, vol. 57, no. 15, pp. 3073–3084, 2002.

[18] M. Ishii and N. Zuber, “Drag coefficient and relative velocity in bubbly, droplet or particulate flows,” *AIChE J,* vol. 25, no. 5, pp. 843–855, 1979.

[19] A. Tomiyama, “Drag, Lift and Virtual Mass Force Acting on a Single Bubble,” *3rd International Symposium on Two-Phase Flow Modelling and Experimentation*, 2004.

[20] I. Roghair, Y. Lau, N. G. Deen, H. Slagter, M. Baltussen, M. van Sint Annaland, and J. A. Kuipers, “On the drag force of bubbles in bubble swarms at intermediate and high Reynolds numbers,” *Chemical Engineering Science*, vol. 66, no. 14, pp. 3204–3211, 2011.

[21] P. J. O'Rourke, “Collective drop effects on vaporizing liquid sprays,” Los Alamos National Lab, NM (USA), Jan. 1981.

[22] N. G. Deen, T. Solberg, and B. H. Hjertager, “Large eddy simulation of the gas–liquid flow in a square cross-sectioned bubble column,” *Chemical Engineering Science*, vol. 56, no. 21, pp. 6341–6349, 2001.

[23] B. E. Launder and D. B. Spalding, *Lectures in mathematical models of turbulence*. London, New York: Academic Press (London), 1972.

[24] A. Tomiyama, “Struggle with computational bubble dynamics,” *Multiphase Science and Technology*, vol. 10, no. 4, pp. 369–405, 1998.

[25] Y. Liao and D. Lucas, “A literature review of theoretical models for drop and bubble breakup in turbulent dispersions,” *Chemical Engineering Science*, vol. 64, no. 15, pp. 3389–3406, 2009.

[26] R. Sungkorn, J. Derksen, and J. G. Khinast, “Modeling of turbulent gas–liquid bubbly flows using stochastic Lagrangian model and lattice-Boltzmann scheme,” *Chemical Engineering Science*, vol. 66, no. 12, pp. 2745–2757, 2011.

[27] R. Hansen, “Computational and experimental study of bubble size in bubble columns,” PhD Thesis, Aalborg University, 2009.

[28] L. Schiller and Nauman A, “A drag coefficient correlation,” *VDI Zeitung*, no. 77, p. 318, 1935.

[29] R. Rzehak and E. Krepper, “Bubble-induced turbulence: Comparison of CFD models,” *Nuclear Engineering and Design*, vol. 258, pp. 57–65, 2013.

Influence of the density of states on nonequilibrium dynamics

Sebastian Weber*, Anika Rämer, Nils Brouwer and Baerbel Rethfeld
Department of Physics and Research Center OPTIMAS, University of Kaiserslautern
Erwin-Schrödinger-Str. 46, 67663 Kaiserslautern, Germany
*weber@physik.uni-kl.de

***Abstract*—On ultrashort timescales, laser-excited electrons in a solid are in a strong nonequilibrium, and cannot be described by temperature-based models. We study the microscopic relaxation processes, in particular thermalization and relaxation. We determine the electron-phonon-coupling strength for copper.**

I. Introduction

Since the invention of the laser by Theodore Maiman in 1960, laser-matter interaction has become an important topic in research and daily life. The question of how to describe the laser excitation is frequently discussed in the scientific community. We want to focus on electron dynamics on an ultrashort timescale of about 100 fs. Here, the electrons are driven out of thermal equilibrium and no temperature can be defined. As a consequence of this nonequilibrium, temperature-based models, like the two temperature model [1], are questionable [2]. With the help of kinetic models, their applicability can be studied.

II. Theory

A solid consists of a regular arrangement of atoms, called lattice. These atoms consist of ions and electrons. Two types of electrons can be identified in a solid. The localized electrons with low energy are bound in states similar to the atomic levels. Higher energetic electrons can move inside the lattice and their energy dispersion is influenced by the periodic potential of the atoms. The levels overlap and broaden into bands [3], [4]. The energy dispersion depends on the direction of the momentum of the electron. The description of this direction-resolved band-structure is very complex and requires a high numerical effort. Thus, density of states (DOS)-based models are often applied. An example for the DOS $D(E)$ of copper is shown in fig. 1. It represents the number of electronic states in a small energy interval averaged over all bands and directions [3], [4]. With an effective one-band model [5], an isotropic dispersion relation can be calculated to connect the momentum

$$k(E) = \sqrt[3]{\frac{3 \cdot 2\pi^2}{\sigma} \int_0^E \mathrm{d}\varepsilon \, D(\varepsilon)} \tag{1}$$

of the electron with its energy E [5], as shown in fig. 1. Here, the DOS of a certain material is reflected. However, the numerical effort of the implementation is strongly reduced as compared to the implementation of a complete momentum-resolved band-structure.

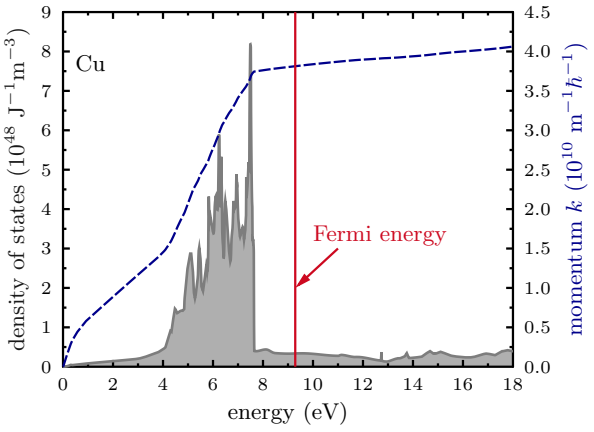


Figure 1. Density of states (DOS, left axis) [6] and the momentum following the corresponding one-band dispersion relation (right axis) of copper. As for other noble metals, the DOS has an area of high density, dominated by the d-electrons. For copper, this area lies between 4 eV and 7.5 eV.

Because of the huge amount of free electrons in the lattice, it is not possible to trace every single electron, so a statistical description with a distribution function f must be used. In equilibrium, the electrons follow a Fermi-Dirac statistics. Due to Pauli-blocking, the electron distribution can not be larger than one. At $T = 0\,\mathrm{K}$ only states below the Fermi energy are occupied by electrons. For elevated temperatures, electrons can also occupy states above this Fermi edge.

Besides the electrons, the behavior of the lattice is important for the dynamics in the solids. It can be described by phonons, which represent collective excitations of the lattice by the quantization of the vibration modes [3], [4]. In difference to the electrons, phonons follow a Bose-Einstein distribution which can be larger than one [3], [4]. The higher the temperature, the more the lattice is in vibration and the more phonons are excited.

If a solid is excited by an ultrashort laser pulse, only the electrons absorb the energy. They are driven out of equilibrium and no temperature can be defined. Due to electron-electron collisions the system thermalizes again on a short timescale

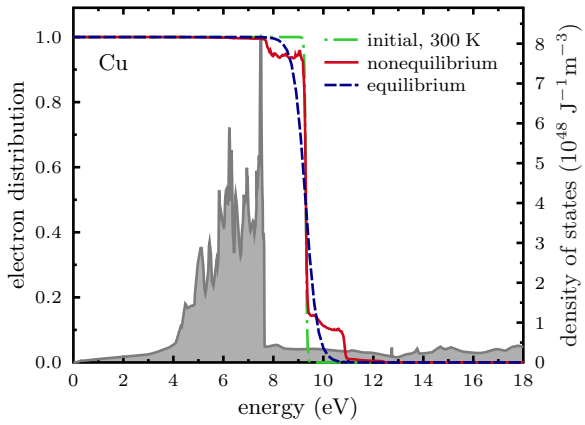


Figure 2. Electron distribution before and after laser excitation. Due to Pauli-blocking, the electrons are driven out of equilibrium and the distribution has a step-like structure (red solid curve). Characteristics of the DOS are represented in the nonequilibrium distribution. Later the distribution thermalizes to a new equilibrium (blue dashed curve).

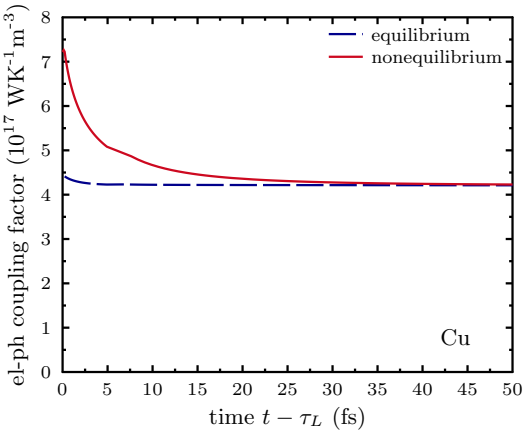


Figure 3. Electron-phonon coupling factor after the end of the laser pulse. The nonequilibrium factor is at first much higher than the one in the equilibrium case and approaches the latter within about 25 fs.

[5]. In the meantime, the electrons couple to the phonons and energy is transfered to the lattice [2].

III. Results

We trace the nonequilibrium dynamics of the electrons after an excitation with a laser pulse of $\tau_L = 10\,\mathrm{fs}$ duration, a rectangular shape, a fluence of $0.65\,\mathrm{mJ/cm^2}$ and a wavelength of $800\,\mathrm{nm}$. Due to Pauli-blocking, only the electrons at the Fermi edge can be excited and a step-like structure develops in the distribution function [2], [7]. Figure 2 shows the nonequilibrium distribution function for copper at the end of the laser pulse. One photon energy below the Fermi edge, the electrons are able to absorb one photon and can be exited above the edge. Thus two steps appear around it. Now electrons with energies two photon energies below the Fermi edge can absorb photons and a new step can develop. The probability for this excitation is much lower as for the first step, and so is the depth of the second step [2], [5]. Looking closer at the step, the characteristics of the DOS can be found in the distribution function one photon energy above each distinct peak [5]. At these peaks in the DOS, there are much more electrons than in the surrounding energy area, so the probability to excite an electron from the peak-energy is much higher and therefore more electrons are excited.

As a consequence of the different electron distributions, the coupling with the lattice is different in equilibrium and nonequilibrium [5]. In both cases, the coupling factor

$$\alpha = \frac{\frac{\partial u_e}{\partial t}\Big|_{\text{el-ph}}}{T_p - T_e} \tag{2}$$

can be defined analogously to the two temperature model (TTM) [1]. It divides the rate of energy transfer $\frac{\partial u_e}{\partial t}$ to the lattice by the difference of the temperatures in equilibrium case.

Figure 3 shows the transient behavior of the coupling strength after the end of the laser pulse. For the nonequilibrium case, the areas of high density of states can couple more easily to the phonons, thus the coupling strength is higher. After a few tens of femtoseconds the system is mostly thermalized again and the electron-phonon coupling factor obtained for nonequilibrium is almost identical to the equilibrium value [5], [8].

IV. Conclusion

After the excitation with an ultrashort laser pulse, the electrons in a metal are in a strong nonequilibrium. Thus no temperature is defined. This nonequilibrium has a strong influence on the electron-phonon coupling. However, an equilibrium is re-established within a few tens of femtoseconds. Within this timescales, the applicability of temperature-based descriptions is questionable.

Acknowledgment

The authors thank Benedikt Y. Mueller for his preliminary work and the Allianz für Hochleistungsrechnen Rheinland-Pfalz for providing computing resources through project LAINEL on the Elwetritsch high performance computing cluster. Financial support by the Deutsche Forschungsgemeinschaft through the Heisenberg program (grant No. RE 1141/15-1) is gratefully acknowledged.

References

[1] S. Anisimov, B. L. Kapeliovich, and T. L. Perel'Man, "Electron emission from metal surfaces exposed to ultrashort laser pulses," *Soviet Journal of Experimental and Theoretical Physics*, vol. 39, pp. 375–377, 1974.

[2] B. Rethfeld, A. Kaiser, M. Vicanek, and G. Simon, "Ultrafast dynamics of nonequilibrium electrons in metals under femtosecond laser irradiation," *Physical Review B*, vol. 65, no. 21, p. 214303, may 2002.

[3] J. Ziman, *Electrons and Phonons*. Oxford: Oxford University Press, 1999.
[4] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, 1st ed. New York: Holt, Rinehart and Winston, 1976.
[5] B. Y. Mueller and B. Rethfeld, "Relaxation dynamics in laser-excited metals under nonequilibrium conditions," *Physical Review B*, vol. 87, no. 3, p. 035139, jan 2013.
[6] Z. Lin, L. Zhigilei, and V. Celli, "Electron-phonon coupling and electron heat capacity of metals under conditions of strong electron-phonon nonequilibrium," *Physical Review B*, vol. 77, no. 7, p. 075133, feb 2008.
[7] W. S. Fann, R. Storz, H. W. K. Tom, and J. Bokor, "Electron thermalization in gold," *Physical Review B*, vol. 46, no. 20, pp. 13 592–13 595, nov 1992.
[8] B. Y. Mueller and B. Rethfeld, "Nonequilibrium electron–phonon coupling after ultrashort laser excitation of gold," *Applied Surface Science*, vol. 302, pp. 24–28, may 2014.

YRS 2016

Organizers

The YRS 2016 is organized by Innovationszentrum Applied System Modeling for Computational Engineering (ASM4CE) and TU-Nachwuchsring.

Innovationszentrum Applied System Modeling
www.applied-system-modeling.de

TU-Nachwuchsring
https://nachwuchsring.uni-kl.de

innovationszentrum
applied system
modeling

TU NACHWUCHSRING

ISBN 978-3-8396-1010-7

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN