Proceedings of the 4th bwHPC Symposium

October 4th, 2017

Alte Aula

Eberhard Karls Universität Tübingen

Edited by

Jens Krüger

University of Tübingen

and

Thomas Walter

University of Tübingen

Contents

bwHPC Symposium 2017
Editorial

Jens Krüger

High Performance and Cloud Computing Group
IT Center, University of Tübingen
Tübingen, Germany
jens.krueger@uni-tuebingen.de

Thomas Walter

Head of IT Center
IT Center, University of Tübingen
Tübingen, Germany
thomas.walter@uni-tuebingen.de

We had the great pleasure to host the bwHPC Symposium 2017 which brought together scientists from various disciplines and IT experts from the compute centers participating in the bwHPC project[1]. The event took place in Tübingen on October the 4th 2017. We were able to secure a very appealing event location, the 'Alte Aula' of the University of Tübingen. It is one of the oldest and most historic university buildings in Tübingen, offering an appropriate ambiance for our event. The main lecture hall gave us literally a great stage for the scientific talks from various disciplines. After a brief welcome to the some 70 participants, T. Walter, director of the IT Center of the University of Tübingen, gave an introduction to the bwHPC project. He highlighted the achievements of the past years as for example the specialization on certain disciplines by participating compute centers. By hosting specialized HPC resources in conjunction with support through competence centers, optimal research environments for different communities have been established. Currently there are four bwForClusters available in the state of Baden-Württemberg, namely the Justus cluster in Ulm supporting computational chemistry, the Nemo in Freiburg for neurosciences, micro systems engineering and particle physics, the MLS&WISO in Heidelberg/Mannheim for economics, social sciences, systems biology and structural biology, and the BinAC in Tübingen for bioinformatics and astrophysics. To offer an entry level and a generic platform to cope with the needs of smaller communities, the bwUniCluster is hosted by the KIT in Karlsruhe. A unique shareholder model is in place involving university and universities of applied science alike handling operation, maintenance and user support in a collaborative manner. Special emphasis was put on future developments as the actualization of the realization concept for bwHPC and the ambitious vision to establish a state-wide data federation. The keynote speech was delivered by A. Brinkmann, head of the Zentrum für Datenverarbeitung and of the Efficient Computing and Storage Group of the Johannes Gutenberg University Mainz. He shared his experiences with supporting communities with high performance computing and storage needs such as bioinformatics. Of particular interest was the argumentation about scaling of compute centers with

[1]www.bwhpc.de

respect to required personnel per operated compute resource. Afterwards, J. Rosskopf from the Institute of Theoretical physics gave a lively presentation about numerical methods for handling many-body-systems, clearly demonstrating that quantum effects are not only relevant in physics, but also for biology and medicine. K. Rabbertz concluded the morning sessions with insights about high precision predictions for particle collisions calculated on the NEMO cluster. He has a double affiliation at the KIT and the Cern, the later the location of the Large Hadron Collider (LHC) for which these predictions are made. For the lunch break we took a little walk through the historic city center of the beautiful town of Tübingen to have our meals at the 'Prinz Karl' university canteen. It was a pleasure to watch how the participants of the bwHPC mixed and grouped, continuing there lively discussions about scientific and HPC related topics.

The first afternoon session was started by M. Orzechowski and J. Kitowski from PL-Grid who gave an overview about federated HPC and data handling in Poland. The federate approach of their ONEDATA project addressed some challenges also present for the anticipated data federation in Baden-Württemberg. D. Thorwarth from the University Clinic in Tübingen illustrated how simulations may help to reduce the damage to healthy tissue for cancer patients undergoing radiation therapy. A. Gorska from the Algorithms in Bioinformatics Group at the University of Tübingen gave a fascinating talk about the dynamics of the human gut phageome and how it is affected during an antibiotics therapy. J. Kratzke spoke about his fluid dynamics simulations of human blood vessels which provide valuable insight for the design of aortic medical prostheses. During the following coffee break all participants had the chance to discuss various topics with the authors of the more than 16 posters presented at the bwHPC symposium. A topic drawn from life was addressed by P. Saini from the Service Operations Management Group at the University of Mannheim. Using HPC resources she explored the optimal location for small businesses such as convenient stores. I. Witte from the University of Hohenheim assessed the impact of climate change on agro-ecosystems. She employed multi-model ensemble simulations on the bwUniCluster. The final afternoon session was concluded by P. Renze head of the Institute of Energy and Drive Technologies of the Ulm Univer-

sity of Applied Sciences. His presentation about computational fluid dynamics addressed the impact of the surface structure of tubes and pipes for different cooling devices.

We thank the members of the program committee helping with reviewing the submissions for the proceedings of the bwHPC Symposium. Namely these are in unsorted order Karsten Siegmund, Jutta Oexle, Christian Mosch, Rainer Keller, Matthias Neuer, Dirk von Suchodoletz, Sven Siebler, Bernd Wiebelt, Werner Dilling, Felix Bartusch, Maximilian Hanussek Bärbel Grosse-Wöhrmann, Brigitte Wellenkamp, Heinz Kredel, Simon Raffeiner, Stefan Rösler, Sabine Richling and Michael Jancyk. Special thanks go to Jordan Denev and Volker Lutz who voluntarily reviewed additional contributions.

We look forward meeting all of you again on the next bwHPC symposium in 2018 in Freiburg.

Dynamics of the human gut phageome during the antibiotics therapy

Anna Górska*†, Silke Peter‡§, Matthias Willmann‡§, Robert Schlaberg¶‖ and Daniel Huson*

*ZBIT Center for Bioinformatics, University of Tübingen, Sand 14, 72076, Tübingen, Germany
†International Max Planck Research School From Molecules to Organisms,
Max Planck Institute for Developmental Biology and Eberhard Karls University Tübingen,
Spemannstr. 35 - 39, 72076, Tübingen, Germany
‡Institute of Medical Microbiology and Hygiene,
University of Tübingen, Elfriede-Aulhorn-Str. 6, 72076, Tübingen, Germany
§German Center for Infection Research (DZIF), partner site Tübingen, Tübingen, Germany
¶Department of Pathology, University of Utah, UT, Salt Lake City, USA
‖ARUP Institute for Clinical and Experimental Pathology, UT, Salt Lake City, USA

Abstract—**Antibiotics are widely used in all sorts of the medical procedures. Without them, modern medical practice is not possible. As much as they are inevitable, they have severe adverse effects in promoting the antibiotic resistance. Each antibiotic therapy has a profound impact on the human gut. Firstly antibiotics select resistant bacteria what results in rapid changes of the taxonomical composition of the human gut. Secondly, the antibiotic pressure prompts the horizontal gene transfer including transduction that is intermediated by the phages, what advances spread of the resistance. Although the taxonomical composition restores after a month from the therapy, the bacteria get enriched with the genes conferring resistance to antibiotics. We have analyzed changes in the microbiome and phageome of the two people throughout the antibiotic therapy and subsequent recovery time. Analysis of the phage genomes is bioinformatically tricky as phages have complex genomic structures and there are no reliable reference databases. Therefore, we developed a novel comprehensive pipeline for the time-series analysis of the metagenomic datasets focusing on the function and changes in the phage sequences. We observed the phages respond to the antibiotic therapy, and are enriched with antibiotic-resistant genes.**

I. INTRODUCTION

Antibiotics are amongst the most successful drugs in saving human life. They have been used in medicine since their discovery in 1928. Those 90 years of extensive antibiotics usage caused the emergence of antibiotic resistance [1] - i.e., there are more and more bacterial strains, including pathogenic, that cannot be eradicated from the patient's organism. In 2014 WHO classified antibiotic resistance as a severe threat to the human life [2].

No antibiotic is ideally specific towards the pathogen. Therefore, an antibiotic therapy affects a substantial portion of the person's gut bacteria and causes rapid changes in the composition of the gut microbiome. Although the taxonomical structure of the healthy microbiome can restore in a month, the therapy causes enrichment of the antibiotic-resistant genes within the genomes of the gut bacteria [3], [4]. Consequently, with each antibiotic therapy it is easier for the pathogenic bacteria to become resistant and it is harder for the microbiome to restore its taxonomic structure [5].

The antibiotic-resistant genes can pass between bacterial cells via means of the horizontal gene transfer (HGT), that includes the three mechanisms: natural transformation, conjugation involving plasmids and transduction involving bacterial viruses [6], [7].

The bacterial viruses (phages) infect bacteria and integrate their genomes into the bacterial chromosomes. Prompted by stress, phages can steer bacteria into the production of the phage particles loaded with the nucleic acid molecule of the phage genome [8]. The phages are often enriched with the bacterial genes located near to their integration side on the genome. This process is influenced by antibiotics as they select for resistant bacteria and consequently phages carrying resistant genes [9].

The human gut works as a bioreactor enabling faster HGT causing emergence and spread of the resistance. Therefore, analysis of the HGT is necessary to obtain a full picture of the dynamics within the gut microbiome and perhaps to control the emergence of the resistance in the future. This project aims to analyze in detail changes of the human gut phageome and genes transferred by phages during the antibiotic therapy and to lie down work for the comprehensive pipeline for analysis of the HGT within microbiomes.

II. STUDY

Two healthy volunteers took ciprofloxacin twice a day orally. Ciprofloxacin is a broad-spectrum fluoroquinolone antibiotic that inhibits DNA separation during the cell division via inhibiting the gyrases and topoisomerases [10]. The stool samples were collected six times: before the therapy, on the 1st, 3rd, and the last day of the therapy (6th), and on the subsequent +2 and +28 days after the therapy [11]. For each of the twelve samples (two participants and the six time-points), two types of sequencing protocols were carried out. First sequencing was a standard deep whole-genome sequencing of entire metagenome (MiSeq at GATC Biotech AG Constance,
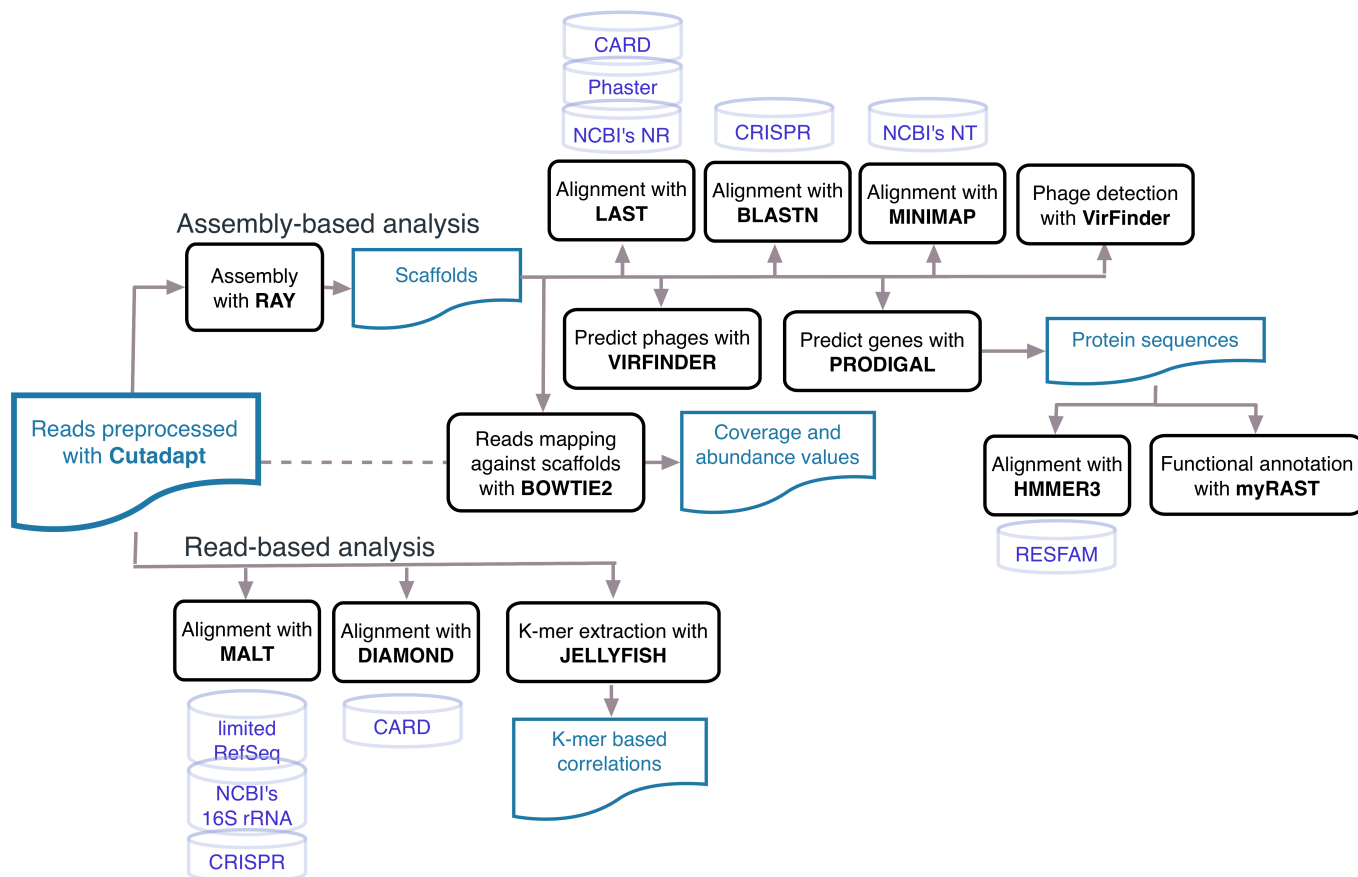
Figure 1: Data analysis pipeline. Blue elements denote inputs and results, black programs and violet the databases.

Germany). Second sequencing regarded only the viral-like particles as described in the [12]. It was performed on the NextSeq 500 System (mid-output kit, 2x150). Therefore, for each of the samples, two sequencing datasets are available: whole-metagenomic sequencing encapsulating both bacteria and phages (Microbiome set), and bacteriophage sequencing (Phage set).

III. Methods

Metagenomic sequencing generates large numbers of the short reads belonging to a wide range of the micro-organisms. The classical metagenome analysis pipeline relies on the alignment of the reads against the most suitable sequence database. Based on those alignments the taxonomic or functional profiles are computed. Therefore, the analysis is only as reliable as the reference databases.

In the case of the Phage set, the classical approach is not possible. There are only ∼2,000 complete phage genomes in the NCBI database what is an order of magnitude smaller than the number of the bacterial genomes (∼35,000 January 2018). Additionally, the phage genomes are believed to be much more versatile and variable. Therefore, rather than using the classical approach the reads were assembled and annotated. To ensure comparability, the samples of the Microbiome set underwent the same analysis pipeline.

The assembly and the subsequent annotation is a complex pipeline involving ten different databases and total of sixteen computational tools (Fig. 1). It is not only expensive in terms of the computational time, but also a complicated procedure regarding the installation of the programs with all their dependencies, and storage. The entire pipeline was deployed on the computational cluster. A bioinformatician's job besides the design and execution of the analysis pipeline includes visualization of the results. That is yet another complex computational task performed on the cluster. The plotting was done with python Matplotlib package [13].

A. Basic analysis

The analysis pipeline consists of two paths: read-based and assembly based analysis (Fig. 1). The first path provides a contamination assessment and outlook of the global changes in the microbiome and phageome. The second path enables high-resolution gene-specific analysis of the feature behaviors over the course of the antibiotic therapy and following recovery.

First, the reads were preprocessed with Cutadapt [14]. Next, based on the number of k-mer count vectors extracted from the preprocessed reads the correlation heatmaps (Jellyfish [15]) and PCA plots were computed. The alignment of reads against the 16S rRNA NCBI database was used to determine possible bacterial contaminations in the Phage samples (MALT [16]).
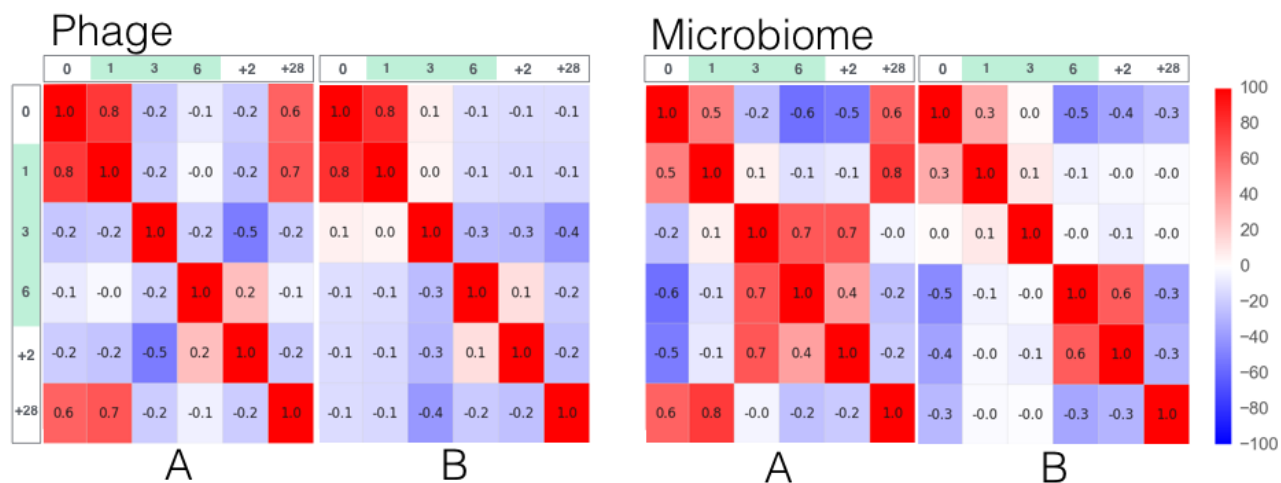
Figure 2: Correlation matrices for both datasets and both participants. Green color denotes days of the antibiotic therapy.

The assembled scaffolds were aligned against the multiple databases: CARD [17] to check if they carry the antibiotic-resistant genes, PHASTER [18] to find viral proteins, nucleotide database to enable MEGAN taxonomic annotation [19] and against the previously collected CRISPR-spacer database to identify phages [20], [21]. Next, genes were predicted, and the protein sequences were aligned with HMMER3 against the HMM profiles of the antibiotic resistance genes to enable sensitive detection of the antibiotic resistance. Finally, myRAST [22] was used for functional annotation of the extracted protein sequences.

Based on the reads mapping (Bowtie2 [23]), the coverage was computed, and then each scaffold was assigned an abundance value. That enables the abundance values of the scaffolds sharing particular annotations to be summed up. If we do it separately per each time-point, the result is an abundance trajectory. Therefore, abundance trajectories for different features in a context of the antibiotic therapy could be analyzed.

B. Phage detection

The Microbiome set comprises of the bacterial, plasmid and the phage scaffolds. We employed numerous strategies for the phage detection, e.g., based on the presence of CRISPR spacer, or of the phage protein. The methods pointed at disjoined sets of the scaffolds. Therefore, the Random Forest (RF) classifier was trained and used to enrich the selections. We used VirFinder [24] results (p-value ≤ 0.05) as the positive labeling, and training based on the other features such as GC content, gene density and a number of overlapping ORFs. In total, there were fourteen features used for the RF training.

IV. RESULTS AND DISCUSSION

K-mer based correlation plots (Fig. 2) show that both the phageome and the microbiome for both participants change as a result of the antibiotic pressure. In all variants, the 0 and 1st days are positively correlated with each other suggesting there is no change on the 1st day of therapy; and

negatively correlated with the following three days. Hence the structure of phageome and microbiome changes as a result of the therapy, starting on the 3rd day. In the case of the Microbiome set, participant A 3rd to +2 days correlate with each other, however, in both of the Phage sets, the 6 and +2 days are correlated. That means the gut microbiomes reach other stable states under the antibiotic pressure. Finally, for the Microbiome and Phage sets of the participant A, the last day correlates with the 0 and 1st days, i.e., microbiome and phageome structures restore after the therapy. That is not a case for the participant B, suggesting that only the A participant's microbiome is healthy.

As predicted, we observed the antibiotic influences foremost bacteria. Abundance of the both bacterial genomic and the plasmid scaffolds carrying antibiotics-resistant genes increase as a result of the therapy. We observed an increase of the abundance of the scaffolds with an alignment to the antibiotic-resistant HMM profile, suggesting phages pick up fragments of the unspecific resistant genes (Fig. 3). As showed by the k-mer profiles, the levels do return to the starting values in the case of the participant A but not B. Ovearll trajectory patterns for the phage sets are weaker. What is probably caused by a lower quality and depth of the sequencing of this dataset in a comparison to the Microbiome set.

V. CONCLUSION

The study has a pilot character, aimed at developing and testing of the pipeline as the two participants do not constitute a representative biological sample, the strong biological conclusions cannot be drawn. However, we have shown that the pipeline enables extensive analysis of the multiple aspects of such combined time series data. It is easily adaptable for analysis of other HGT mechanisms such as conjugation involving plasmids. Nevertheless, it is a complicated pipeline suitable for deployment only on the sizeable computational cluster such as BiNAC.
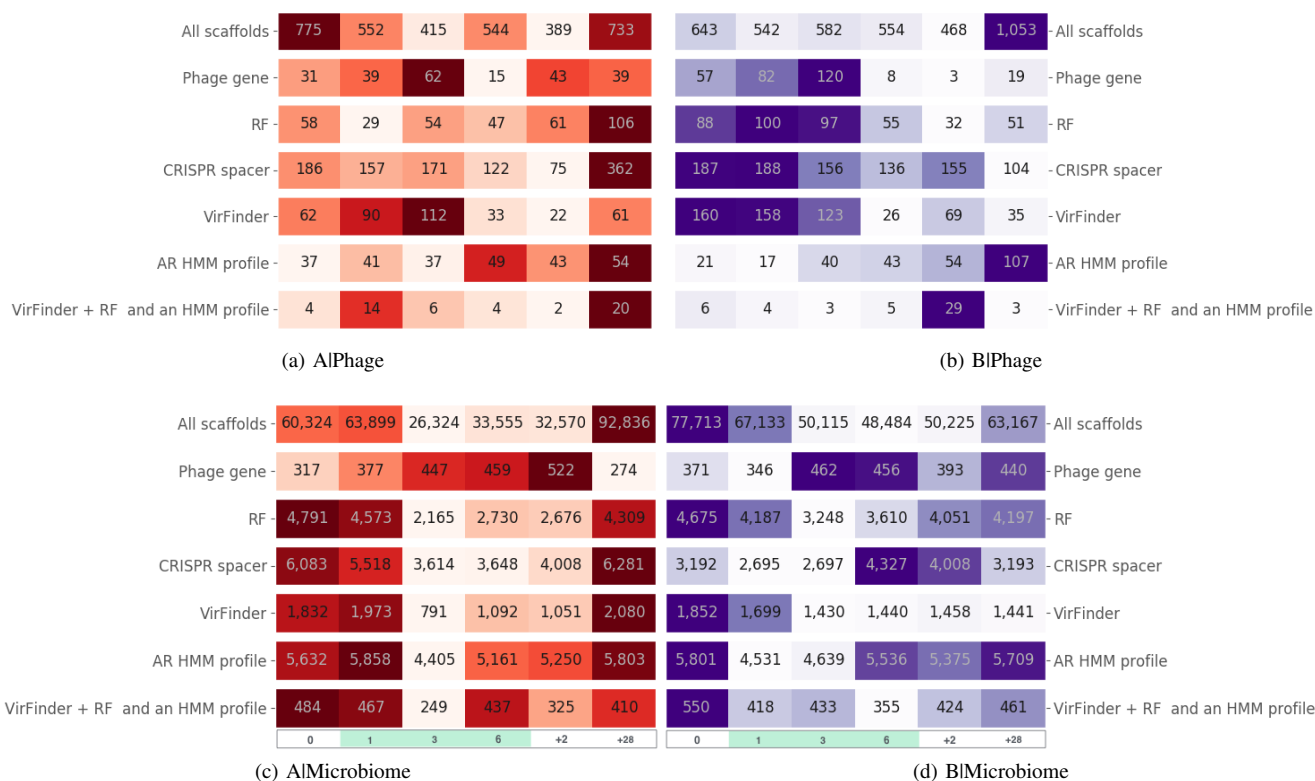
Figure 3: Feature abundance trajectories for different selections of the scaffolds: rows denote features and columns the consecutive sampling days. The green color marks the days of the antibiotic therapy. Four features are speaking to a possible phage origin of the scaffolds: Phage gene, CRISPR spacer, VirFinder and RF. HMM profiles of antibiotics informs about antibiotic resistance genes. Colors correspond to each horizontal trajectory separately.

REFERENCES

[1] H. Nikaido, "Multidrug Resistance in Bacteria," *Annu Rev Biochem.*, no. 2, pp. 119–146, 2009.

[2] WHO, "Antimicrobial resistance. Global Report on Surveillance," *Bulletin of the World Health Organization*, vol. 61, no. 3, pp. 383–94, 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22247201

[3] Y. Hu, X. Yang, J. Qin, N. Lu, G. Cheng, N. Wu, Y. Pan, J. Li, L. Zhu, X. Wang, Z. Meng, F. Zhao, D. Liu, J. Ma, N. Qin, C. Xiang, Y. Xiao, L. Li, H. Yang, J. Wang, R. Yang, G. F. Gao, J. Wang, and B. Zhu, "Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota." *Nature communications*, vol. 4, p. 2151, 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23877117

[4] A. A. Salyers, A. Gupta, and Y. Wang, "Human intestinal bacteria as reservoirs for antibiotic resistance genes," *Trends in Microbiology*, vol. 12, no. 9, pp. 412–416, 2004.

[5] L. Dethlefsen and D. A. Relman, "Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation." *Proceedings of the National Academy of Sciences of the United States of America,*

no. Suppl 1, pp. 4554–61, 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20847294

[6] P. J. Hastings, S. M. Rosenberg, and A. Slack, "Antibiotic-induced lateral transfer of antibiotic resistance," *Trends in Microbiology*, vol. 12, no. 9, pp. 401–404, 2004.

[7] W. V. Schaik, "The human gut resistome," *Royal Society Publisher*, vol. 370, pp. 1–9, 2015.

[8] T. A. Brown, "Mapping genomes," in *Genomes*, 2002, pp. 125–162.

[9] S. R. Modi, H. H. Lee, C. S. Spina, and J. J. Collins, "Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome." *Nature*, vol. 499, no. 7457, pp. 219–22, 2013.

[10] B. Jia, A. R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K. K. Tsang, B. A. Lago, B. M. Dave, S. Pereira, A. N. Sharma, S. Doshi, M. Courtot, R. Lo, L. E. Williams, J. G. Frye, T. Elsayegh, D. Sardar, E. L. Westman, A. C. Pawlowski, T. A. Johnson, F. S. Brinkman, G. D. Wright, and A. G. McArthur, "CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database," *Nucleic Acids Research*, vol. 45, no. D1, pp. D566–D573, 2017.

[11] M. Willmann, M. El-Hadidi, D. H. Huson, M. Schütz, C. Weidenmaier, I. B. Autenrieth, and S. Peter, "Antibiotic selection pressure determination through sequence-based metagenomics," *Antimicrobial Agents and Chemotherapy*, vol. 59, no. 12, pp. 7335–7345, 2015.

[12] R. V. Thurber, M. Haynes, M. Breitbart, L. Wegley, and F. Rohwer, "Laboratory procedures to generate viral metagenomes." *Nature protocols*, vol. 4, no. 4, pp. 470–483, 2009.

[13] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 99–104, 2007.

[14] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads," *EMBnet.journal*, vol. 17, no. 1, pp. pp. 10–12, 2011.

[15] G. Marçais and C. Kingsford, "A fast, lock-free approach for efficient parallel counting of occurrences of k-mers," *Bioinformatics*, vol. 27,

no. 6, pp. 764–770, 2011.

[16] A. Herbig, F. Maixner, K. I. Bos, A. Zink, J. Krause, and D. H. Huson, "MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman," *bioRxiv*, p. 50559, 2016. [Online]. Available: http://biorxiv.org/content/early/2016/04/27/050559.abstract

[17] A. G. McArthur, N. Waglechner, F. Nizam, A. Yan, M. A. Azad, A. J. Baylay, K. Bhullar, M. J. Canova, G. De Pascale, L. Ejim, L. Kalan, A. M. King, K. Koteva, M. Morar, M. R. Mulvey, J. S. O'Brien, A. C. Pawlowski, L. J. V. Piddock, P. Spanogiannopoulos, A. D. Sutherland, I. Tang, P. L. Taylor, M. Thaker, W. Wang, M. Yan, T. Yu, and G. D. Wright, "The comprehensive antibiotic resistance database," *Antimicrobial Agents and Chemotherapy*, vol. 57, no. 7, pp. 3348–3357, 2013.

[18] D. Arndt, J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, and D. S. Wishart, "PHASTER: a better, faster version of the PHAST phage search tool," *Nucleic Acids Research*, vol. 44, no. W1, pp. W16–W21, 2016. [Online]. Available: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw387

[19] D. H. Huson, S. Beier, I. Flade, A. Górska, M. El-Hadidi, S. Mitra, H.-J. Ruscheweyh, and R. Tappu, "MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data," *PLOS Comput Biol*, vol. 12, no. 6, pp. 4–12, 2016.

[20] R. Jansen, J. D. A. Van Embden, W. Gaastra, and L. M. Schouls, "Identification of genes that are associated with DNA repeats in prokaryotes," *Molecular Microbiology*, vol. 43, no. 6, pp. 1565–1575, 2002.

[21] A. F. Andersson and J. F. Banfield, "Virus population dynamics and acquired virus resistance in natural microbial communities." *Science (New York, N.Y.)*, vol. 320, no. 5879, pp. 1047–50, 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18497291

[22] E. M. Glass and F. Meyer, "The Metagenomics RAST Server: A Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes," *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, vol. 8, pp. 325–331, 2011.

[23] B. Langmead, "Aligning short sequecning reads with Bowtie," *Current Protocols Bioinformatics*, pp. 1–24, 2011.

[24] J. Ren, N. A. Ahlgren, Y. Y. Lu, J. A. Fuhrman, and F. Sun, "VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data," *Microbiome*, vol. 5, no. 1, p. 69, 2017.

Empirical Charge Scheme for Transition Metals and Lanthanoids: Development and Applications

Bodo R. Martin, Peter Comba
Inorganic Chemistry
Heidelberg University
Heidelberg, Germany
bodo.martin@aci.uni-heidelberg.de

Peter Comba
IWR
Heidelberg University
Heidelberg, Germany
peter.comba@aci.uni-heidelberg.de

Parameters for the fluctuating charge model are derived for In(III), La(III), Lu(III) and Bi(III), based on x-ray structural data and DFT single point calculations. A number of density partitioning schemes is tested.

atomic charges; transition metals; lanthanoids; force fields; fluctuating charge method;

I. Introduction

Atomic charges are a useful concept in computational chemistry. They are widely used to model electrostatic interactions in force field software. [1, 2] Even though polarization is known to be necessary for an adequate description of electrostatic interactions, few mayor software codes to date make use of these modified potentials [3]. Moreover, most force field programs are only parameterized for organic compounds and have poor or non-existent support for transition metal or lanthanoide complexes.

The aim of this work is to develop an accurate and fast model for atom charges, for elements up to and including actinides, implemented in the molecular mechanics program Momec [4]. This classical molecular mechanics software, developed in our group, is especially tailored for the calculation of coordination compounds. We have previously developed a fluctuating charge model, i.e. a geometry-dependent atomic charge model, for complexes containing copper and iron [5], and extend this here to complexes containing Indium, Lanthanum, Lutetium and Bismuth in the oxidation state III: In(III), La(III), Lu(III) and Bi(III). In a subsequent step, the atomic charges obtained for these elements can be the base for descriptors in a QSPR prediction of logD values, e.g. for complexes for PET imaging and radionuclide treatment [6]. Our focus here will be the determination of atomic charges from DFT calculations, based on approx. 200 compounds from the CSD [7].

II. The Fluctuating charge model

In the charge equilibration (QEq) method, the charge distribution is calculated via (1)

$$\chi_{eq} = \chi_i^0 + 2J_{ii}^0 q_i + \sum_{j>i} q_j J_{ij} \tag{1}$$

where χ_i^0 is the electronegativity and J_{ii} is the idempotential of atom i. The interaction term J_{ij} is calculated as a two-center Coulomb integral [8]. The valence state ionization energies of an element vs. its ionization can be approximated via a quadratic fit as in (2)

$$E^a(q) = E^a(q=0) + A^a q + B^a q^2 \tag{2}$$

Where the superscript "*a*" denotes parameters for the isolated atom. Assuming a similar relationship for atoms as part of a molecule leads to the equivalent equation (3)

$$E^m(q) = E^m(q=0) + A^m q + B^m q^2 \tag{3}$$

The expressions for the parameters A and B in the molecule are now described as the atomic parameters and linear increments (4)

$$A^m = A^a + \Delta A \quad B^m = B^a + \Delta B \tag{4}$$

First and second derivatives of (3) gives

$$\frac{dE^m(q)}{dq} = A^m + 2B^m q \quad \frac{d^2 E^m(q)}{dq^2} = 2B^m$$

In a molecule, the total electrostatic energy can be expressed as a function of partial atomic charges, (5)

$$E(q_1 \ldots q_N) = \sum_{i=1}^{N} \left(\begin{array}{c} E(0)_i^m + \left(\dfrac{\partial E_i^m}{\partial q} \right)_{q_i} q_i \\[2mm] + \dfrac{1}{2} \left(\dfrac{\partial^2 E_i^m}{\partial q^2} \right)_{q_i} q_i^2 \end{array} \right) \qquad (5)$$

$$+ \sum_{i=1}^{N} \sum_{j>i} q_i q_j D_{ij}$$

Substituting we get (6)

$$E(q_1 \ldots q_N) = \sum_{i=1}^{N} \left(E(0)_i^m + A_i^m q_i + 3 B_i^m q_i^2 \right)$$

$$+ \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i} q_i q_j D_{ij} \qquad (6)$$

Taking the first derivative of (6) with respect to partial charge q_i gives the Iczkowski-Margrave electronegativity of atom i (7)

$$\chi(q_i) = \left(\frac{\partial E}{\partial q} \right)_{q_i} = A_i^m + 6 B_i^m q_i + \frac{1}{2} \sum_{j \neq i} q_j D_{ij} \qquad (7)$$

With Sanderson`s electronegativity equalization principle (8),

$$\chi_1 = \chi_2 = \cdots \chi_N = \chi_{eq} \qquad (8)$$

conservation of the total charge Q_{tot} and by dropping the "m" superscripts we arrive at the matrix equation:

$$\begin{pmatrix} 6B_1 & \frac{1}{2}D_{12} & \cdots & \frac{1}{2}D_{1N} & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{2}D_{N1} & \frac{1}{2}D_{N2} & \cdots & 6B_N & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} q_i \\ q_2 \\ \vdots \\ q_N \\ \chi_{eq} \end{pmatrix} = \begin{pmatrix} -A_1 \\ -A_2 \\ \vdots \\ -A_N \\ Q_{tot} \end{pmatrix}$$

The interatomic term D_{ij} is expressed as (9) [5]

$$D_{ij} = \left(\sqrt{ r_{ij}^2 + \left(\frac{1}{2\sqrt{B_i B_j}} \right)^2 } \right)^{-1} \qquad (9)$$

The atomic parameter fits A^a and B^a can easily be calculated from ionization data; one example for Lu is shown in Fig. 1. The linear corrections for the atoms in a molecule, ΔA and ΔB, are the two parameters for the fluctuating charge model.
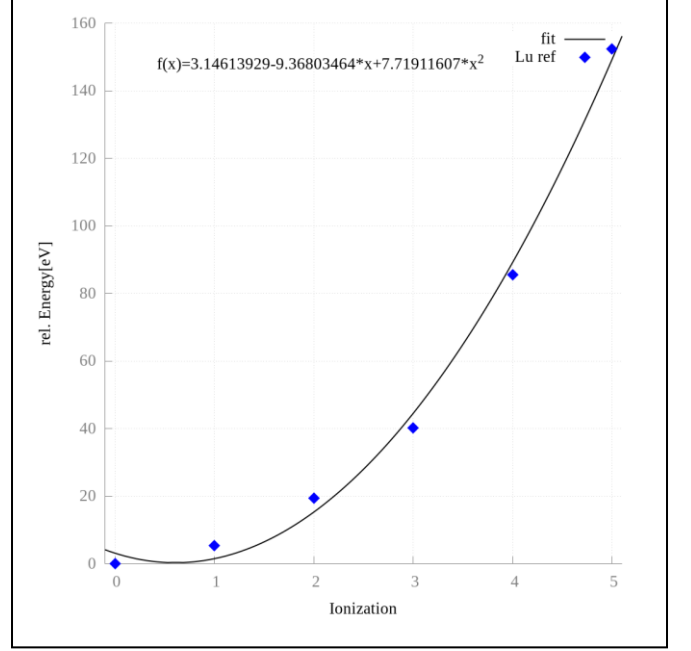
(5)

Fig. 1. Lu energy vs. ionization fit to experimental data [9]

III. Methodology

Starting from crystal structures, e.g. from structures obtained from the CSD [7], single point quantum mechanical calculations are followed by an iterative Hirshfeld treatment with Horton [10], or Chargemol [12], to obtain reference atomic charges. The reference charges are then approximated with our 2-parameter fluctuating charge model via a simplex minimization. Additionally, the geometries of the reference data set are reproduced by fitting the force field parameters of the Momec molecular mechanics program. The combined force field thus allows for a fast geometry optimization and charge prediction for complexes with elements up to Ac. The fluctuating charge method is of complexity $O(n^2)$; it is planned to reduce the computational cost by a divide and conquer algorithm. From an MM-calculated structure with atomic charges, one can then derive descriptors for e.g. logD or NMR isotropic shift predictions. A single point calculation with Gaussian typically takes a few hours on 16 cpu cores; Horton calculations take a similar time on one cpu core, but, dependent on the accuracy required, may require up to 100 GB of memory.
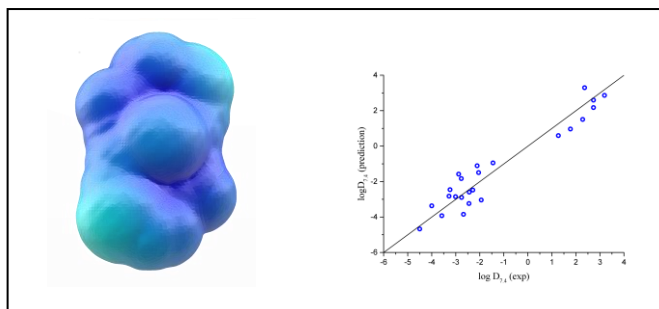
Fig. 2. SAS surface charge example (CSD code AGOVIB, left) and Cu complexes logD fit similar to [6].

IV. LOGD PREDICTION

We have previously demonstrated [6] that by using the above empirical charge scheme, it is possible to derive accurate predictions for logD values for a range of copper complexes, using a five-parameter model: the ovality O_v, the solvent accessible surface A, the local polarity π, the total variation from the mean positive and negative potentials σ_{tot}, and the polar surface area A_{pol}. With the parameterized charge model, we can now extend the logD prediction to all Transition Metal and Lanthanoide complexes, analogous to the surface charge distribution and logD fits from previous work [6] (cf. Fig. 2).
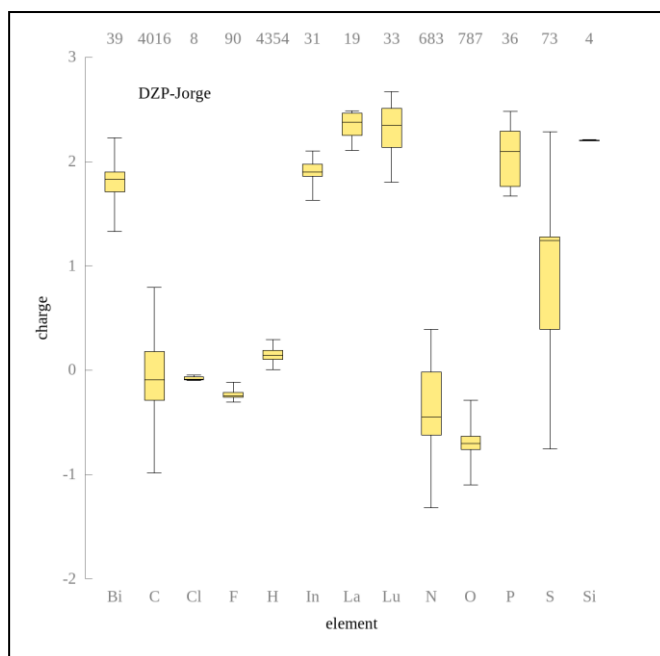

Fig. 3. Boxplots of reference atomic charges, calculated with Horton; Gaussian 09 with Jorge DZP [14] basis set. Numbers at top of figure are atom counts. Outliers (above/below 95%) not shown for clarity.

V. RESULTS

Fig. 3 shows iterative Hirshfeld atomic charges of a set of complexes of In(III), La(III), Lu(III) and Bi(III), for a range of coordination numbers, obtained via Horton 2.1.0b1, after calculating single points with B3LYP Gaussian 09 [11]; Fig. 4 shows DDEC6 [12] charges derived from the same basis set. Results for Orca 3.0.3 [13] with SVP/ZORA and TZP/ZORA basis sets are not shown for brevity.

The calculation of atomic charges depends on the DFT method, the basis set and the density partitioning scheme. As the data set for Fig 4 is larger than for Fig 3, the two cannot be directly compared; moreover, for the subsequent MM parameterization, consistency of the reference charges is more important than the absolute values, i.e. similar complexes should have similar atomic charges, and charge outliers should have a sufficiently different chemical environment.
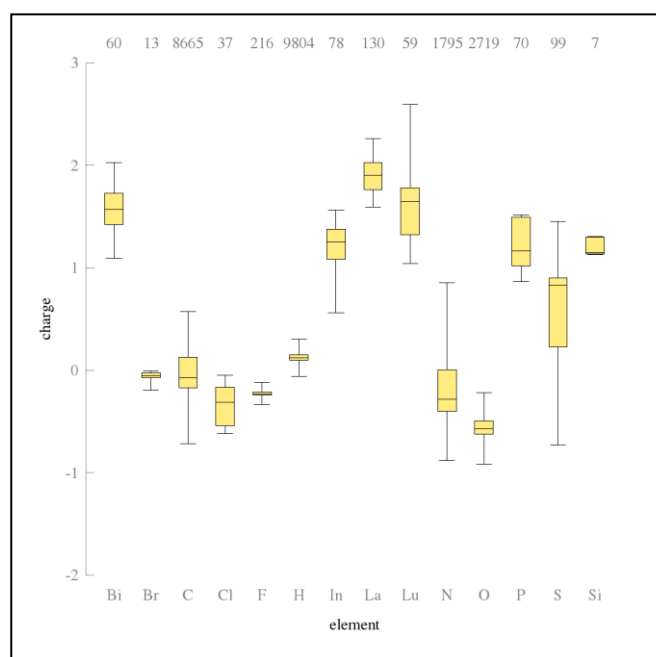
Fig. 4. Boxplot of reference atomic DDEC6 [12] charges; Gaussian 09 with Jorge DZP [14] basis set. Numbers at top of figure are atom counts. Outliers (above/below 95%) not shown for clarity.

VI. CONCLUSION/OUTLOOK

Calculation of reference atomic charges for elements in the 5th row and above is not straightforward; only a few general basis sets exist, and relativistic effects have to be accounted for either implicitly or via ZORA or DKH. Here we present two methods for obtaining charges, after post-processing via Horton [10] and Chargemol [12], to obtain iterative Hirshfeld and DDEC6 charges. We are still in the process of collecting data to decide on the best set of reference charges for our fluctuating charge model, which we have shown by a model parameterization and previous work to be able to reproduce atomic charges accurately. In a next step, we will extend the

logD prediction, based on descriptors obtained with this charge model. The NMR prediction based on charges may require higher moments; use of higher charge moments is currently under development.

REFERENCES

[1] D.A. Case, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2017), AMBER 2017, University of California, San Francisco.

[2] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers", SoftwareX, vol. 1-2, pp. 19-25, 2015.

[3] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J.W. Ponder, P. Ren, "The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins," J. Chem. Theory Comput. Vol 9(9), pp. 4046-4063, 2013.

[4] P. Comba, T. W. Hambley, B. Martin, "Molecular modeling of inorganic compounds", Wiley-VCH, Weinheim 2009.

[5] P. Comba, B. Martin, A. Sanyal, "An efficient fluctuating charge model for transition metal complexes," J. Comput. Chem. vol. 34, pp. 1598-1608, 2013.

[6] P. Comba, B. Martin, A. Sanyal, H. Stephan, "The computation of lipophilicities of Cu PET systems based on a novel approach for fluctuating charges," Dalton Trans. vol. 42(31), pp. 11066-11073, 2013.

[7] C.R. Groom, I.J. Bruno, M.P. Lightfoot, S. C. Ward, "The Cambridge Structural Database" Acta Cryst. vol. B72, pp. 171-179, 2016.

[8] A. K. Rappe, W. A. Goddard III, "Charge Equilibration for Molecular Dynamics Simulations," J. Phys. Chem. vol. 95, pp.3358-3363, 1991.

[9] A. Kramida, Yu. Ralchenko, J. Reader, NIST ASD Team [2017, August 23], "NIST Atomic Spectra Database (ver. 5.2)" https://physics.nist.gov/asd , 2014.

[10] T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, E. González-Espinoza, C. M. Chan, T. D. Kim, K. Boguslawski, S. Fias, S. Vandenbrande, D. Berrocal, P.W. Ayers, "HORTON 2.1.0", http://theochem.github.com/horton/

[11] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, and G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, and J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 09 , Revision D.01.

[12] T. A. Manz, N. G. Limas, "Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology," RSC Adv., vol. 6(53), pp. 47771-47801, 2016.

[13] F. Neese, "The ORCA program system", Rev.: Comput. Mol. Sci., vol. 2, pp. 73-78. 2002.

[14] F. E. Jorge, L. S. C. Martins, M. L. Franco, "All-electron double zeta basis sets for the lanthanides: Application in atomic and molecular property calculations," Chem. Phys. Lett. , vol 642, pp. 84-88, 2016.

Recent applications of the Tübingen-Vienna Smooth Particle Hydrodynamics code

Christoph M. Schäfer* , Thomas I. Maindl†, Christoph Burger†, and Oliver J. Wandel*

* Institute for Astronomy and Astrophysics, Eberhard-Karls-Universität Tübingen, Tübingen, Germany

{ch.schaefer,oliver.wandel}@uni-tuebingen.de

† Department of Astrophysics, University of Vienna, Vienna, Austria

{thomas.maindl,c.burger}@univie.ac.at

Abstract—Here, we present the latest improvements and applications of the Tübingen-Vienna Smooth Particle Hydrodynamics (SPH) code. By the use of modern graphics processing units (GPUs), we have increased the performance of astrophysical simulations in the field of hydrodynamics and solid mechanics by porting an OpenMP code to the GPU with CUDA™. Recently, we have added a porosity module and a soil module to our existing framework. The code is freely available upon request.

Index Terms—smooth particle hydrodynamics, modelling, computational astrophysics

I. INTRODUCTION

SPH was firstly introduced as a grid-free Lagrangian particle method for solving the system of hydrodynamic equations for compressible and viscous fluids by [7] and [13]. SPH is especially suited for boundary-free problems with high density contrasts. Rather than being solved on a grid, the equations are solved at the positions of the so-called particles, each of which represents a volume of the fluid with its physical quantities like mass, density, temperature, etc. and moves with the flow according to the equations of motion. The method was extended to solid mechanics in the beginning of the nineties by [10] and improved extensively in the following years [11], [12], [14]. The extension to model granular media and soil has been successfully added to SPH in the last years [2].

A recent development in high performance computing is to harness the power of GPUs for massive parallel applications. The most advanced interface for general purpose GPU programming is CUDA from Nvidia. We have successfully used the CUDA framework to accelerate our existing SPH code [16] and present here new modules that have been added to the framework lately. In the subsequent section, we will describe the physical models and equations and the numerical scheme to solve them. We will present three recent applications of our numerical framework in its subsequent section. Eventually, we will conclude in the last section.

II. PHYSICAL AND NUMERICAL MODELS

A. Basic equations

The system of partial differential equations that describe the dynamics of an elastic solid body is given by the three equations for the conservation of mass, momentum and (internal)

energy, respectively. The conservation of mass is given by the continuity equation, which reads

$$\frac{\mathrm{d}\varrho}{\mathrm{d}t} + \varrho \frac{\partial v^\alpha}{\partial x^\alpha} = 0, \tag{1}$$

where ϱ denotes the density and $\mathbf{v}$ the velocity and the Einstein summation rule is applied. Greek indices denote the spatial coordinates and run from 1 to 3. The second equation in the system accounts for the conservation of momentum

$$\frac{\mathrm{d}v^\alpha}{\mathrm{d}t} = \frac{1}{\varrho} \frac{\partial \sigma^{\alpha\beta}}{\partial x^\beta}. \tag{2}$$

The stress tensor σ is given by the pressure p and the deviatoric stress tensor $S^{\alpha\beta}$

$$\sigma^{\alpha\beta} = -p\delta^{\alpha\beta} + S^{\alpha\beta}. \tag{3}$$

In contrast to fluid dynamics, this set of partial differential conservation equations is not sufficient to describe the elastic body. Additionally the time evolution of the deviatoric stress tensor has to be specified. The missing relations are called the constitutive equations which describe the rheology of the body and relate the kinematic states of the body to the dynamic states. The time evolution of the deviatoric stress tensor can be expressed as

$$\frac{\mathrm{d}S^{\alpha\beta}}{\mathrm{d}t} = 2\mu \left[\dot{\varepsilon}^{\alpha\beta} - \frac{1}{3}\delta^{\alpha\beta}\dot{\varepsilon}^{\gamma\gamma} \right] + S^{\alpha\gamma}R^{\gamma\beta} + S^{\beta\gamma}R^{\gamma\alpha}, \tag{4}$$

where $R^{\alpha\beta}$ denotes the rotation rate tensor

$$R^{\alpha\beta} = \frac{1}{2}\left(\frac{\partial v^\alpha}{\partial x^\beta} - \frac{\partial v^\beta}{\partial x^\alpha} \right), \tag{5}$$

and $\dot{\varepsilon}^{\alpha\beta}$ the strain rate tensor

$$\dot{\varepsilon}^{\alpha\beta} = \frac{1}{2}\left(\frac{\partial v^\alpha}{\partial x^\beta} + \frac{\partial v^\beta}{\partial x^\alpha} \right). \tag{6}$$

The closure of this set of partial differential equations is given by the equation of state (eos). The eos relates the thermodynamic variables density ϱ, pressure p, and specific internal energy u. Currently, our framework allows to choose between several eos, which are described in more detail in the following.

B. Equation of State

1) Murnaghan eos: The Murnaghan eos is often used to model the behaviour of solids under high pressure conditions. The pressure depends non-linearly on the density

$$p = \frac{K_0}{n} \left[\left(\frac{\varrho}{\varrho_0} \right)^n - 1 \right], \tag{7}$$

with the zero pressure bulk modulus K_0 and the constant n. The Murnaghan eos is limited to isothermal compression.

2) Tillotson eos: The Tillotson eos distinguishes between two different domains depending on the specific internal energy u. In the case of compressed regions ($\varrho \geq \varrho_0$) and u lower than the energy of incipient vaporization e_{iv} pressure is given by

$$p = \left[a + \frac{b}{1 + u/(e_0 \eta^2)} \right] \varrho u + A\xi + B\xi^2, \tag{8}$$

with $\eta = \varrho/\varrho_0$ and $\xi = \eta - 1$. In case of expanded material (u greater than the energy of complete vaporization e_{cv}) the eos takes the form

$$\begin{aligned} p = a\varrho u + &\left[\frac{b\varrho u}{1 + u/(e_0\eta^2)} \right. \\ &+ A\xi \exp\left\{ -\beta\left(\frac{\varrho_0}{\varrho} - 1 \right) \right\} \right] \\ &\times \left. \exp\left\{ -\alpha\left(\frac{\varrho_0}{\varrho} - 1 \right)^2 \right\}. \right. \end{aligned} \tag{9}$$

The symbols ϱ_0, A, B, e_0, a, b, α, and β are material dependent parameters. In the partial vaporization regime $e_{\text{iv}} < u < e_{\text{cv}}$, the pressure is linearly interpolated between the pressures obtained via (8) and (9), respectively.

3) Ideal gas eos: For the simulation of gases, the framework includes the ideal gas equation

$$p = \left(\frac{c_p}{c_V} - 1 \right) \varrho u, \tag{10}$$

where c_p and c_V denote the specific heat capacities at constant pressure and volume, respectively.

C. Porosity model

In order to model subresolution porosity, we have implemented the p-α model. The basic idea of this porosity model is the separation of the volume change in a porous material into the collapse of the pores and the compression of the matrix material. The distention is a measure for the porosity of the material and defined as follows

$$\alpha = \frac{\varrho_s}{\varrho}, \tag{11}$$

where ϱ is the density of the porous material and ϱ_s the density of the corresponding matrix material. The pressure of the porous material is then determined by the experimentally found relation between the pressure p_s of the solid material and the porosity or distention [6]

$$p = \frac{1}{\alpha} p_s(\varrho_s, u_s) = \frac{1}{\alpha} p_s(\alpha\varrho, u). \tag{12}$$
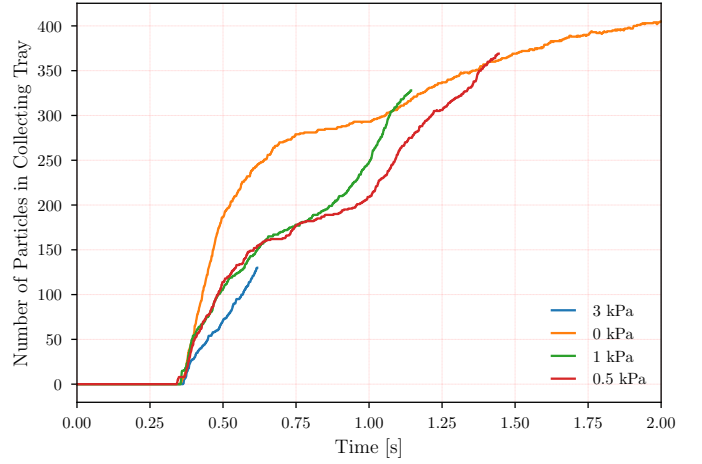

Figure 1. Sampling process on Phobos' surface for different material properties of regolith: The angle of internal friction is $\phi = 42\,\text{deg}$ and the cohesion c varies with $0\,\text{Pa}$, $500\,\text{Pa}$, $1000\,\text{Pa}$ and $3000\,\text{Pa}$.

Here, u_s denotes the specific internal energy of the solid matrix material. In addition to the basic conservation equations, the evolution of the distention is also integrated to determine the change of porosity during compaction [18].

D. Rheology model for regolith

The plastic flow of regolith can be modelled by the Drucker-Prager yield criterion. In the Drucker-Prager model, the yield function f is given by the relation between the first and second invariant of the stress tensor

$$f(I_1, J_2) = \sqrt{J_2} + \alpha_\phi I_1 - k_c = 0. \tag{13}$$

The invariants are given by the following terms

$$I_1 = \text{tr}(\sigma) = \sigma^{\gamma\gamma} \quad \text{and} \quad J_2 = \frac{1}{2} S^{\alpha\beta} S_{\alpha\beta}. \tag{14}$$

The two material constants α_ϕ and k_c are called Drucker-Prager constants and are related to the material constants cohesion c and angle of internal friction ϕ in the following way (for plane strain conditions)

$$\alpha_\phi = \frac{\tan\phi}{\sqrt{9 + 12\tan^2\phi}} \quad \text{and} \quad k_c = \frac{3c}{\sqrt{9 + 12\tan^2\phi}}. \tag{15}$$

E. Self-gravity

The code implements a Barnes-Hut tree that allows for treatment of self-gravity. We follow the efficient implementation by [5].

III. Applications

In this section, we summarize three recently presented applications of the computational framework: modelling a possible sampling process on Phobos, collisions between planetary embryos and impacts of smaller objects into asteroids.
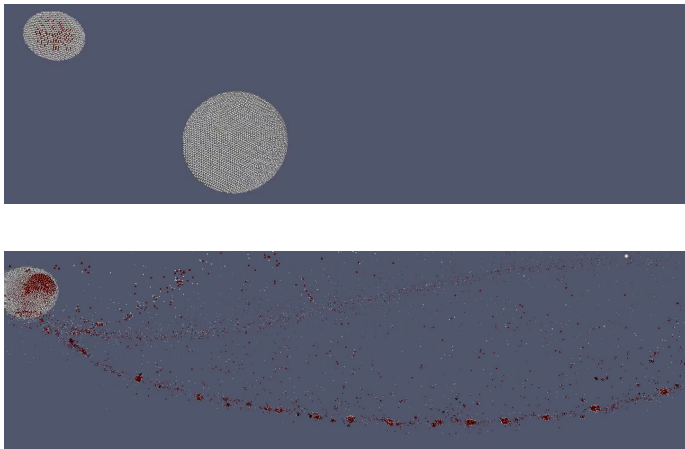
13

Figure 2. Chain forming simulation of two planetary embryos. The total mass of the system is 10^{23} kg and the mass ratio between projectile and target object is 1:9. The impact angle of this simulation is 30 deg and the impact speed was $2.5 \times$ the mutual escape velocity. The upper panel shows the initial setup of the collision and the lower panel the formed chain.

A. Sampling on Phobos

The surface of the Mars moon Phobos is covered with regolith and the thickness of the layer varies between $5\,\mathrm{m}$ and $100\,\mathrm{m}$. The European Space Agency (ESA) currently investigates the implementation of a sample return mission to Phobos. We developed a SPH model for one specific sampling tool design and performed several simulations with different conditions regarding the material properties of the regolith. Our implementation is based on the model developed by [2], a model for granular flow using SPH in the context of solid mechanics combined with a Drucker-Prager yield criterion for plastic flow. The sampling speed for different material properties of regolith is shown in figure 1. A detailed description of the project and the results of more simulations can be found in [17].

B. Collisions between planetary embryos

The Tübingen-Vienna SPH code has been successfully applied to the simulation of collisions between planetary embryos. Each of the two colliding planetary embryos consists of an iron core with a basaltic mantle and an outer ice shell as described in [3] and [4]. We trace the fate of the ice and water in these collisions. The outcome of a special chain-forming collision is shown in Figure 2.

C. Impacts of smaller objects into asteroids

Active asteroids are objects in the asteroid belt that are dynamically similar to asteroids, but show activity and mass loss like comets. The mass loss is triggered by impacts of meter-sized objects in these asteroids that can form craters up to $12\,\mathrm{m}$ in diameter [8]. The crater shaped by such an impact is shown in Figure 3.
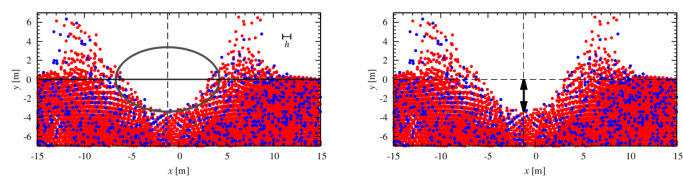

Figure 3. Impact crater formation on a Main-belt comet. Simulational outcome of a typical impact simulation. The diameter of the impactor was $1\,\mathrm{m}$ and the impact speed was $4.4\,\mathrm{km\,s^{-1}}$ [8].

IV. Conclusion

We presented three recent applications of the Tübingen-Vienna SPH code, where two of them featured our newly implemented porosity and regolith/soil modules. The bwFor-Cluster BinAC is specially suited for our purposes since it enables us to cover a vast material parameter space with high resolution. Often, in our kind of simulations, the material data is scarce and many different properties of the material have to be varied and respected, leading to a high number of simulations. The code is freely available to the astrophysical and geophysical community upon request by email to the authors.

Our intention is to extend the existing code to multi GPUs support, which will allow for even higher resolutions. The demand for higher resolution is prevailing in the field of planet formation, since collisions between planetary cores indicate that the existing simulations do not show convergence yet [15].

Acknowledgments

References

[1] U. Ayachit, *The ParaView Guide: A Parallel Visualization Application*, 2015.

[2] H. H. Bui, R. Fukagawa, K. Sako, and S. Ohno, "Lagrangian meshfree particles method (sph) for large deformation and failure flows of geomaterial using elastic–plastic soil constitutive model," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 32, no. 12, pp. 1537–1570, 2008.

[3] C. Burger and C. M. Schäfer, "Applicability and limits of simple hydrodynamic scaling for collisions of water-rich bodies in different mass regimes," *Proceedings of the First Greek-Austrian Workshop on Extrasolar Planetary Systems*, pp. 63–81, Mar. 2017.

[4] C. Burger and C. M. Schäfer, "Water transfer and loss in hit-and-run collisions," *accepted for publication in Celestial Mechanics and Dynamical Astronomy*, 2017.

[5] M. Burtscher, *GPU Computing Gems Emerald Edition*, W. Hwu, Ed., 2011.

[6] M. M. Carroll and A. C. Holt, "Static and dynamic pore-collapse relations for ductile porous materials," *Journal of Applied Physics*, vol. 43, no. 4, pp. 1626–1636, 1972. [Online]. Available: https://doi.org/10.1063/1.1661372

[7] R. A. Gingold and J. J. Monaghan, "Smoothed particle hydrodynamics: Theory and application to non-spherical stars," *MNRAS*, vol. 181, pp. 375–389, 1977.

[8] N. Haghighipour, T. I. Maindl, C. Schäfer, R. Speith, and R. Dvorak, "Triggering Sublimation-driven Activity of Main Belt Comets," *ApJ*, vol. 830, p. 22, Oct. 2016.

[9] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[10] L. D. Libersky and A. G. Petschek, *Smooth Particle Hydrodynamics with Strength of Materials*. Springer Verlag, 1990, p. 248.

[11] L. D. Libersky, A. G. Petschek, T. C. Carney, J. R. Hipp, and F. A. Allahdadi, "High strain lagrangian hydrodynamics," *J. Chem. Phys.*, vol. 109, p. 67, 1993.

[12] L. D. Libersky, P. W. Randles, T. C. Carney, and D. L. Dickinson, "Recent improvements in sph modeling of hypervelocity impact," *Int. J. Impact Eng.*, vol. 20, p. 525, 1997.

[13] L. B. Lucy, "A numerical approach to the testing of the fission hypothesis," *AJ*, vol. 82, no. 12, pp. 1013–1024, 1977.

[14] P. W. Randles and L. D. Libersky, "Smoothed Particle Hydrodynamics: Some recent improvements and applications," *Comp. Methods Appl. Mech. Engrg.*, vol. 139, p. 375, 1996.

[15] C. Reinhardt and J. Stadel, "Numerical aspects of giant impact simulations," *MNRAS*, vol. 467, pp. 4252–4263, Jun. 2017.

[16] C. Schäfer, S. Riecker, T. I. Maindl, R. Speith, S. Scherrer, and W. Kley, "A smooth particle hydrodynamics code to model collisions between solid, self-gravitating objects," *A&A*, vol. 590, p. A19, May 2016.

[17] C. M. Schäfer, S. Scherrer, R. Buchwald, T. I. Maindl, R. Speith, and W. Kley, "Numerical simulations of regolith sampling processes," *Planet. Space Sci.*, vol. 141, pp. 35–44, Jul. 2017.

[18] O. J. Wandel, C. M. Schäfer, and T. I. Maindl, "Collisional fragmentation of porous objects in planetary systems," *Proceedings of the First Greek-Austrian Workshop on Extrasolar Planetary Systems*, pp. 225–242, Mar. 2017.

Containerization of Galaxy Workflows increases Reproducibility

Felix Bartusch, Maximilian Hanussek and Jens Krüger

High-Performance and Cloud Computing Group, IT Center, University of Tübingen

Abstract—**Published scientific findings supported by computational experiments should be reproducible by following the methodical description in the publication. Because this is often not the case, we present a method to containerize Galaxy workflows within Singularity containers. These containers are executable such that other researchers can reproduce computational experiments or run their own experiments using the containerized workflows.**

I. INTRODUCTION

Scientific findings in fields like bioinformatics usually originate from processing primary data with computational methods to get results that agree or disagree with the initial assumption. The results are then published to make other researchers aware of the new findings and maybe new experimental methods. The applied methods are documented in the 'Methods' section of the paper and should ensure the reproducibility of the presented findings by other researchers. The content of this section differs from field to field but are often standardized. Examples for such standards in different fields are MIASE [1] for simulation experiments or MIAME [2] for microarray experiments. For computational methods it is consensus to state at least the used program version. Some publications state checklists of most important information needed to reproduce computational experiments [3].

But often the reproducibility is not guaranteed although the applied methods are described in detail in the corresponding publication [4], [5], [6]. Among others, the reason for irreproducibility is an insufficient description of used software, parameters and the data-processing workflow topology [6].

Scientific workflow engines like Galaxy [7], [8], [9], Taverna [10] and Knime [11] or portals like MoSGrid [12], [13] solve these problems by describing the computational workflow explicitly. In case of the mentioned workflow engines, the used software, parameters, and the workflow topology are described in a single file that can also be shared with other researchers via platforms like MyExperiment [14].

But also the use of workflow engines do not ensure full reproducibility because the shared workflow decays over time. A study based on a subset of Taverna workflows shared on the MyExperiment platform between the years 2007 and 2012 showed that around 80% of the workflows are not executable or do not produce the promised results any more [15]. At the time of the study the workflows had a maximal age of 5 years. The reasons for workflow decay are volatile third-party resources, missing example data, missing execution environment, and an insufficient description of the workflow.

Containerization techniques improved the mobility and usability of computations in the last years by encapsulating an operating system together with software and data. The most widely used containerization technique is Docker [16], but other techniques like Singularity [17] became more popular. Containerization of software increases reproducibility enormously [18]. One specific example is a Docker container providing a complex software stack for the simulation of spectroscopic fragmentation of small molecules with QCEIMS. The containerized software stack was used by UNICORE to execute the simulation protocol and to access the workflow conveniently [19].

In this work we combine the workflow engine Galaxy with Singularity to increase the reproducibility and mobility of scientific workflows. In the first part we describe how we achieve a containerization of the workflow. Then we show that the containerized workflow is straightforward executable on a local machine, on a HPC-system, and in the cloud. Last we show how our approach differs from related work that tries to make scientific workflows more reproducible.

II. CONTAINERIZATION OF GALAXY WORKFLOWS

The containerization is a fully automated process in two distinct steps as shown in Figure 1. The first step creates a template Singularity container with an operating system and installs a Galaxy instance in it. In the second step the actual workflow and its tool dependencies are installed in the container. Both encapsulation steps are explained in more detail in this section. The whole code base, configuration files and the example workflow are accessible via a GitHub repository [20]. We used Singularity 2.3.2, Python 2.7.13 and the Python package bioblend 0.9.0 for this work.

A. Create a template container

In this step a template Singularity container is created that is able to store one or more Galaxy workflows. We provide a shell-script that creates the container by bootstrapping with a so-called Singularity definition file that describes how a Singularity container is provisioned.

During the bootstrap we install CentOS 7.3 in the container, but also other operating systems would be possible. Subsequently several system packages are installed by the yum pack-
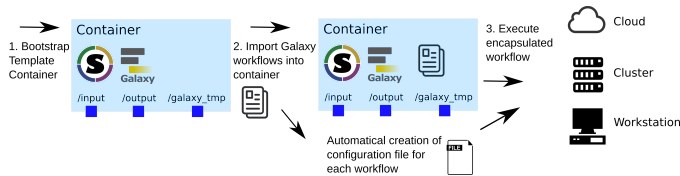
Fig. 1. Overview of the containerization workflow. **1.** Create the template Singularity container with an operating system and Galaxy. Also three folders are created to serve as mount points for the input, output and temporary Galaxy directory. **2.** Import the Galaxy workflows. For each workflow a configuration file is created **3.** Specify input/output mounts and workflow parameters using the execution configuration file. Then run the workflows in the cloud, on a cluster, or your own workstation.

age manager: `Development-Tools, epel-release, python-pip, python-devel, bzip2, git.`

Finally Galaxy is installed by cloning the GitHub branch corresponding to the user specified Galaxy version. We start Galaxy one time to initialize the Galaxy instance. After the initialization completes, we shut down the initialized Galaxy and the template container is ready to encapsulate Galaxy workflows.

B. Encapsulate Galaxy workflows in template container

The template container from above is now used to encapsulate Galaxy workflows. Because Singularity containers behave like normal files on the file system one can copy the template container and use the copy to encapsulate other workflows.

Galaxy workflows are explicitly described in .ga files that can be exported from other Galaxy instances or downloaded from platforms like myExperiment. Our import script will automatically install the workflow as described in the workflow file. It is also possible to obtain the workflow from a running Galaxy instance, but you need access to the workflow in the Galaxy instance beforehand.

We wrote a Python script to import the workflow in the container automatically. Prior to the workflow import, certain settings are specified in a configuration file. During the import, we generate another configuration file that is needed to execute the encapsulated workflow afterwards. First, we startup Galaxy in the container and use the Python module bioblend [21] to interact with the Galaxy-API.

a) Create new Galaxy user: We need a Galaxy user in the container in order to run the workflow and to use important Galaxy features like data libraries. Therefore a new Galaxy user is created in the containerized Galaxy instance. One has to specifiy beforehand the user name and user mail-address in the configuration file. We also generate an API-key as well as a random password for this user and store them in the execution configuration file. The API-key is used to run the containerized workflow from the command line whereas the password is needed to access the Galaxy instance in the browser.

b) Install tools and import workflow: Galaxy describes a workflow explicitly in a .ga-file that contains a json-dictionary. Among others, the file lists the inputs and computational steps of the workflow. For the input steps we parse the required input datasets together with their Galaxy datatype and an additional

description if provided. These input information are written to the execution configuration file. Later the user specifies the files that should serve as input when the workflow is executed on the command line.

For the computational steps we parse the specified tool version and install everything by using the Galaxy-API. The tools and their dependencies are then installed from the publicly available Galaxy ToolShed. Also so-called runtime parameters are handled explicitly. Runtime parameters are parameters whose values are not specified in the workflow file. The user has to specify them when invoking the workflow, e.g. number of iterations or cutoff values. The runtime parameters for a tool are also written to the execution configuration file.

III. EXECUTION OF THE ENCAPSULATED WORKFLOW

Additionaly to Singularity the Python code presented in this paper has to be installed on the remote system to run the containerized Galaxy instance. Whereas Singularity has to be installed by an administrator, the Python code can reside in user space as it does not need any special privileges. The user has to upload the container to the system that will execute the container.

We provide a Python script that can execute the containerized workflow in two modes, an automatic and an interactive mode. To execute the workflow automatically the user specifies the path to the input files and parameters of the computational steps in the execution configuration file. This file was automatically created in the encapsulation step and needs just small adjustments. The workflow is started by executing:

`python execute_workflow.py --conf <ini-file>`

The execution script starts Galaxy, uploads the input files, invokes the workflow, waits until the workflow has completed, and downloads the workflow results to a specified output folder. The input files are not duplicated when they are uploaded to Galaxy but a symbolic link to the location of each input file on the file system is created in the Galaxy input data library.

In the interactive mode, Galaxy is started by the script and the user can then access the Galaxy instance via the browser. The user can login to Galaxy, study the containerized workflows, upload datasets to the Galaxy instance, and run the workflow as well as just single tools. In the end the user can also download the resulting datasets.

Because Singularity container cannot be changed by a non-root user, we had to change the default Galaxy configuration in some points for the execution. Most of the actions in Galaxy cause changes of the Galaxy database, writes to logging files, or leads to the creation of some temporary data. Therefore we created a temporary directory for Galaxy in the container where the database and temporary files reside. Upon execution we create a unique temporary directory on the host and mount it to Galaxy's temporary directory in the container. All write operations of Galaxy are performed in this temporary directory on the host. This also ensures that the container itself is not

altered during the execution. The temporary directory on the host is deleted after the workflow was executed.

IV. CONTAINERIZATION INCREASES REPRODUCIBILITY

The created Singularity container encapsulates immediately executable workflows. This fact alone increases the reproducibility and reuse massively because it eliminates some of the most important reasons of workflow decay like missing execution environment or an insufficient description of the workflow [15].

Sharing the container together with the execution configuration files enables other researchers to reproduce published results or adjust parameters of the workflows to run their own analysis. Archiving the container ensures in principle reproducibility and reuse of the workflow a long time after creation. The CiTAR project [1] for example has the aim to archive Docker and Singularity containers and provide an infrastructure that secures the ability to execute the container in the future.

The proposed method increases also the portability of the computation. One can develop the workflow in a local or public Galaxy instance, encapsulate the workflow and run it on any computational resource. We did this for a 14-step Next Generation Sequencing workflow that is also included in the GitHub repository of this project [20]. We encapsulated the workflow in a Singularity container and were able to run the workflow on BinAC [22] and a VM on the de.NBI cloud site Tübingen[2].

V. RELATED WORK

Research environments and workflows implemented in Galaxy are often shared via virtual machine images [23], [24]. These virtual machine images contain all dependencies and are ready-to-use. But virtual machine images are usually not executable on HPC clusters because just few clusters installed hypervisors for the virtual machines on the compute nodes.

Galaxy is available as a Docker container [25] and the Galaxy team also provides the library Ephemeris to interact with the Galaxy-API via bioblend to install workflows and its dependencies[3]. Both, the Docker Galaxy container and Ephemeris, could be combined to encapsulate scientific workflows in a Docker container. This would be a subset of the functionality our approach provides. Besides we are using a different containerization technique, we also provide automatic execution of the workflow and handle input and output data. Singularity containers are more suitable for scientific HPC clusters because Singularity targets scientific applications and supports HPC resources like MPI (Message Passing Interface), Infiniband, and GPUs [17]. Another aspect is that Docker needs a root owned daemon process whereas Singularity tries to minimize the security risk with suid-executables that handle the privilege escalations needed for software containerization.

The project ViCE [26] works on sustainable virtual research environments like Galaxy. ViCE offers a image registry for virtual collaborative environments and an infrastructure to support accessibility and reproducibility of the environments.

VI. OUTLOOK

Singularity will be further developed. At the time of writing Singularity 2.4 was released. The new version offers some new features of which the containerization of workflows benefits, e.g. new container format, support of overlays and container instances.

Also Singularity could be used to encapsulate other workflow engines like Taverna or Knime. The created containers could be shared via newly platforms like Singularity Container Registry [4].

REFERENCES

[1] D. Waltemath, R. Adams, D. A. Beard, F. T. Bergmann, U. S. Bhalla, R. Britten, V. Chelliah, M. T. Cooling, J. Cooper, E. J. Crampin, A. Garny, S. Hoops, M. Hucka, P. Hunter, E. Klipp, C. Laibe, A. K. Miller, I. Moraru, D. Nickerson, P. Nielsen, M. Nikolski, S. Sahle, H. M. Sauro, H. Schmidt, J. L. Snoep, D. Tolle, O. Wolkenhauer, N. Novère, "Minimum information about a simulation experiment (MIASE)", *PLoS Computational Biology*, vol. 7, no. 4, pp. 1–4, 2011

[2] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, M. Vingron, "Minimum information about a microarray experiment (MIAME)-toward standards for microarray data", *Nature Genetics*, vol. 29, no. 2, pp. 365–371, 2001

[3] G. K. Sandve, A. Nekrutenko, J. Taylor, E. Hovig,"Ten Simple Rules for Reproducible Computational Research", *PLoS Computational Biology*, vol. 9, no. 10, pp. 1–4, 2013

[4] J. P. A. Ioannidis, D. B. Allison, C. A. Ball, I. Coulibaly, X. Cui, A. C. Culhane, M. Falchi, C. Furlanello, L. Game, G. Jurman, J. Mangion, T. Mehta, M. Nitzberg, G. P. Page, E. Petretto, and V. van Noort, "Repeatability of published microarray gene expression analyses", *Nature Genetics*, vol. 41, no. 2, pp. 149–155, 2008.

[5] K. A. Baggerly and K. R. Coombes, "Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology", *Annals of Applied Statistics*, vol. 3, no. 4, pp. 1309–1334, 2009.

[6] D. Garijo, S. Kinnings, L. Xie, L. Xie, Y. Zhang, P. E. Bourne, and Y. Gil, "Quantifying reproducibility in computational biology: The case of the tuberculosis drugome", *PLoS ONE*, vol. 8, no. 11, pp. 1–11, 2013.

[7] J. Goecks, A. Nekrutenko, J. Taylor, and T. G. Team, "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences", *Genome Biol*, vol. 11, no. 8, pp. 1–13, 2010.

[8] D. Blankenberg, G. V. Kuster, N. Coraor, G. Ananda, R. Lazarus, M. Mangan, A. Nekrutenko, and J. Taylor, "Galaxy: A web-based genome analysis tool for experimentalists", *Current protocols in molecular biology*, pp. 1–21, 2010.

[9] B. Giardine, C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. C. Miller, W. J. Kent, and A. Nekrutenko, "Galaxy: a platform for interactive large-scale genome analysis", *Genome research*, vol. 15, no. 10, pp. 1451–1455, 2005.

[10] K. Wolstencroft, R. Haines, D. Fellows, A. Williams, D. Withers, S. Owen, S. Soiland-Reyes, I. Dunlop, A. Nenadic, P. Fisher, J. Bhagat, K. Belhajjame, F. Bacall, A. Hardisty, A. Nieva de la Hidalga, M. P. Balcazar Vargas, S. Sufi, and C. Goble, "The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud.", *Nucleic acids research*, vol. 41, no. Web Server issue, pp. 557–561, 2013.

[11] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel, "KNIME: The Konstanz Information Miner", in *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*, Springer, 2007.

---

[1] https://www.alwr-bw.de/kooperationen/bwzwm/

[2] https://denbi.uni-tuebingen.de

[3] https://github.com/galaxyproject/ephemeris

[4] https://singularity-hub.org/

[12] J. Krüger, R. Grunzke, and S. Gesing, "The MoSGrid Science Gateway–A Complete Solution for Molecular Simulations", *Journal of Chemical Theory and Computation*, vol. 10, no. 6, pp. 2232–2245, 2014.

[13] L. Zimmermann, R. Grunzke, and J. Krüger, "Maintaining a Science Gateway – Lessons Learned from MoSGrid", *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 6233–6242, 2017.

[14] D. D. Roure, C. Goble, and R. Stevens, "The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflow", *Future Generation Computer Systems*, no. November 2007, pp. 561–567, 2009.

[15] J. Zhao, J. M. Gomez-Perez, K. Belhajjame, G. Klyne, E. Garcia-Cuesta, A. Garrido, K. Hettne, M. Roos, D. De Roure, and C. Goble, "Why Workflows Break -Understanding and Combating Decay in Taverna Workflows", (Washington, DC, USA), pp. 1–9, IEEE Computer Society, 2012.

[16] "Docker, https://www.docker.com/."

[17] G. M. Kurtzer, V. Sochat, and M. W. Bauer, "Singularity: Scientific containers for mobility of compute", *PLOS ONE*, vol. 12, no. 5, pp. 1–20, 2017.

[18] C. Boettiger, "An introduction to Docker for reproducible research", *ACM SIGOPS Operating Systems Review*, vol. 49, no. 1, pp. 71–79, 2015.

[19] M. Hanussek, F. Bartusch, J. Krüger, and O. Kohlbacher, "Efficent Mass Spectra Prediction through Container Orchestration with a Scientific Workflow", in *9th International Workshop on Science Gateways (IWSG)*, 2017.

[20] F. Bartusch, "fbartusch/galaxy2singularity: v.0.1.0, 10.5281/zenodo.1122568."

[21] C. Sloggett, N. Goonasekera, and E. Afgan, "BioBlend: Automating pipeline analyses within Galaxy and CloudMan", *Bioinformatics*, vol. 29, no. 13, pp. 1685–1686, 2013.

[22] J. Krüger, V. Lutz, F. Bartusch, W. Dilling, A. Gorska, C. Schäfer, and T. Walter, "Bioinformatics and astrophysics cluster (BinAC)", *3rd bwHPC Symposium*, pp. 1–1, 2017.

[23] S. J. Schultheiss, G. Jean, J. Behr, P. Drewe, N. Görnitz, A. Kahles, P. Mudrakarta, V. T. Sreedharan, G. Zeller, and G. Rätsch, "Oqtans: a Galaxy-integrated workflow for quantitative transcriptome analysis from NGS Data", *BMC Bioinformatics*, vol. 12, no. Suppl 11, pp. 1–2, 2011.

[24] R. L. Davidson, R. J. M. Weber, H. Liu, A. Sharma-Oates, and M. R. Viant, "Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data", *GigaScience*, vol. 5, no. 1, pp. 1–9, 2016.

[25] B. Grüning, M. van den Beek, B. Batut, J. Chilton, M. ISHII, D. Ryan, E. Afgan, H. Rudolph, K. Ellrott, C. Smith, P. Moreno, H.-R. Hotz, G. V. Kuster, R. Baertsch, M. Edwards, G. L. Corguillé, A. Azab, S. Hiltemann, Rdmorin, M. Chambers, T. Tanjo, R. Hernández, Jasper, and A. Petkau, "bgruening/docker-galaxy-stable: Galaxy Docker Image 17.05, 10.5281/zenodo.583723."

[26] C. B. Hauser, J. Domaschka, "ViCE Registry: An Image Registry for Virtual Collaborative Environments", *2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 82–89, 2017

Multi-model ensemble simulations to assess the impact of climate change on agro-ecosystems

Towards robust soil-crop models for better projections

Irene Witte and Thilo Streck

Institute of Soil Science and Land Evaluation
University of Hohenheim
Stuttgart, Germany
Irene.witte@uni-hohenheim.de

Global climate change will alter the water, nitrogen and carbon cycles of agroecosystems. To predict future agricultural production under climate change, numerical soil-crop models are used. These soil-crop models can represent the complex and coupled processes of agroecosystems in a deterministic manner for a given environment. The projections made by soil-crop models suffer from two kinds of uncertainty: (1) epistemic uncertainty and (2) parameter uncertainty. Additionally, it is assumed that the parameterization is applicable to other environments. Therefore, this study has two major aims. The first aim is to quantify the above-mentioned uncertainties simultaneously by combining two methods: multi-model ensemble modeling and Bayesian statistics. The multi-model ensemble allows to quantify epistemic uncertainty by comparing individual model outputs. This has been demonstrated in many studies. Bayesian methods are common to assess credible parameter intervals for highly nonlinear process models. The second aim of this study is to provide a framework for assessing the robustness of the parametrization of soil-crop models. Therefore, a preliminary numerical study was conducted to test different calibration schemes and to investigate parameters sensitivities in dependence of the environment. The soil-crop modelling software ExpertN 3.0 will be used to set up a multi-model ensemble with eight soil-crop models. The model output will be analyzed by comparison with data from two sites, five soil types and two crops gathered by the DFG Research Unit 1695 since 2010. To achieve the second aim a global sensitivity analyses was conducted to rank the input factors for each soil-crop model. The result of the global sensitivity analyses will clarify the impact of model input on model output in regard to environment, model combinations, and extent. Additionally, different calibration schemes will be tested to identify the method yielding the most robust parametrization. We used a Latin Hypercube sampling scheme. In total, the whole study requires 1,000,000 CPU hours. We expect that the results will enable us to develop a generally applicable and feasible strategy of how soil-crop models have to be set up to produce reliable predictions of agroecosystem behavior under climate change.

Keywords — soil-crop modelling; global sensitivity analysis; uncertainty quanitfication;multi-model ensemble;

I. INTRODUCTION

Soil-Crop models have been designed to coherently simulate crop growth, water and nitrogen dynamics in a given environment. The typical model outputs are soil water content, yield, evapotranspiration, groundwater recharge and nitrogen leaching. However, the state-of-the-art soil-crop models differ in complexity and representations of the various processes of plant growth and water regime as well as solute turnover and transport. Furthermore, these processes depend on environmental conditions such as climate, crop management and soil properties.

Soil-crop models represent the complex, coupled processes in agroecosystems in a deterministic manner. While they have been traditionally used to forecast yields and, later, nitrogen concentrations in seepage and ground water, they are now often used to predict the impact of environmental changes on agroecosystems. In the context of food security for a growing world population the Agricultural Model Intercomparison Project (AgMIP) set itself the goal to predict how global agricultural production will be affected by climate change [1]; [15]. Among others, [4]; [3] used a set of 30 crop models to predict how the grain yield of wheat - one of the world population's staple crops - will change with rising temperature in the future.

Taking a different approach, the DFG Research Unit 1695 has incorporated a soil-crop model into a land surface model [9]. The aim is to improve the accuracy of regional climate projections by capturing the dynamic feedbacks between the soil-crop system and the atmosphere. This is not possible with current climate models. It can be expected that the new approach will also lead to improved yield predictions.

In practice, soil-crop models are often used to assess the impacts of agricultural management in regard to fertilization or irrigation. This is especially important in regions where water is scarce or where the groundwater gets polluted by nitrate and pesticides - which is the case in some regions in Germany [7].

Because, as in the application areas listed above, soil-crop modeling serves as a building block for decision making, it is of utmost importance to achieve highest levels of confidence regarding both kinds of model uncertainty, i.e., (1) epistemic uncertainty and (2) parameter uncertainty [18].

Epistemic uncertainty emerges from differences in the representation of the underlying processes in different models. Parameter uncertainty provides information about credible parameter intervals. This is especially important when true parameter values are non-existent due to spatial and temporal natural heterogeneity and measurement inaccuracy. To quantify these uncertainties simultaneously it is planned to combine two methods: multi-model ensemble and Bayesian methods as suitable tools demonstrated by recent studies [11]; [3]; [20].

Equally important to the uncertainty quantification is the robustness of the models' parametrization. Put simply, robustness means that simulation results for environments the model was not trained in, are still in good agreement with additional measurements. As a first step in the quantification of the parameter uncertainty and the parameterizations' robustness, a global sensitivity analysis will be conducted. This will identify the influential model parameters for a set of different environments. Afterwards we plan to test different calibration schemes differing in the size of the training data sets (i.e., different sites and climate conditions) and in the targeted output variables being trained on. The drawback of all these methods is their computational cost. Soil-crop models easily depend on 50 parameters. To sample the parameter space representatively up to 1,000,000 model runs per method and environment are necessary.

In the next sections preliminary results are presented for the global sensitivity on a subset of the available data and for one model combination.

II. MATERIAL AND METHODS

A. Study Site and measurements

The data used in this study were measured in two agricultural fields - one located on Swabian Alb (EC6, 13 ha), the other one in Kraichgau (EC1, 15 ha). The sites are contrasting in climate and soils. The Swabian Alb (48.5°N 9.8°E, 690 m a.s.l.) has a cold, harsh climate with average temperature of 6.5 °C and annual precipitation of 962 mm. Mean temperature in Kraichgau (48.9°N 8.7°E, 319 m a.s.l.) is 9.3°C and annual precipitation 777 mm. At the EC1 site the soil is deeply developed and fertile (depth >165 cm) with a high storage capacity of plant available water. In contrast, at the EC6 site the soil profile is very shallow and clay rich (depth < 30 cm) why the storage capacity of plant available water is very low.

All measurements have been carried out within the DFG Research Unit 1695 since 2009. The fields are managed and cropped by local farmers (best practice). In this study we only used years when winter wheat was grown. Winter wheat is usually sown in October and harvested between the end of July and mid-August, depending on weather. At EC1 this was the case in the vegetation periods 2010/11, 2012/13 and 2014/15 and at EC6 in 2010/11 and 2013/14. At each site the energy and water fluxes are measured with the eddy-covariance technique. Volumetric soil water content is measured using FDR probes in 5, 15, 30, 45 and 75 cm depth at EC1 and in 5 and 15 cm depth at EC6. Nitrate and ammonium concentration in the soils are measured 2-4 times during the vegetation period, depending on the year under consideration, in 0-30 cm depth at both sites and in Kraichgau additionally in 30-60 cm and 60-90 cm. Leaf Area Index (LAI), vegetative and generative biomass, nitrogen content of the plant and development stage (BBCH) are measured as common plant characteristics in regular intervals. Weather stations record net radiation, rainfall, temperature, relative humidity and wind speed. For a detailed description of the measurement instruments and data processing see [8] and [19].

B. Model

Models were set up and run in the framework of the soil-crop modeling software ExpertN 3.1. As a first model combination we tested the CERES crop model [14], the SOILN [10] carbon and nitrogen turnover model and the Hydrus1D [17] model for the soil water regime. The FAO-Penman-Monteith [2] approach is used to estimate potential crop-specific evapotranspiration.

The CERES - SOILN model combination uses a diverse set of ordinary differential equations to simulate the crop e.g. growth, development stage and yield and the carbon and nitrogen turnover in the soil, e.g., nitrification and mineralisation. The soil water regime is described by the Richards equation (partial differential equation, PDE). Soil nitrogen dynamics are described by the Advection-Dispersion Equation (also a PDE). ExpertN 3.1 solves the set of coupled ordinary and partial differential equations using a fully implicit finite difference scheme with time step control. The CERES model has eight parameters which are typically adjusted (PD, PV, P1, P4, P5, PHINT, G1, and G2). The van Genuchten parameterization for the hydraulic functions was used to simulate the water regime. Since the hydraulic properties vary over the profile depth several sets of van Genuchten parameters are needed (n1, n2, a1, a2, KS1, KS2, l1 and l2). For assessing potential evapotranspiration we consider the crop factors (ki (kc_{ini}), km (kc_{mid}), ke (kc_{end})). Nitrification and denitrification rates in the topsoil (Ni, Dn) are parameters of the SOILN model. Additionally, the maximum root water uptake rate and the maximum root nitrogen uptake as parameters for the interrelations of the different submodels are considered.

C. Global Sensitivity Analyses

A global sensitivity analysis answers the question of what model input - if fixed - reduces the variance of the model output most [16]. The first-order sensitivity indices (S1) quantify how much of the total model output variance can be directly related to a specific parameter.

To calculate the S1 values for each parameter we conducted Latin Hypercube samples [12] with sizes ranging from 100,000 to 1,000,000 and followed the approach of [13]; [5]

as implemented in the python package SALib [7]. The approach has the advantage of not depending on a certain sampling design. Therefore, S1 can be estimated also in cases when the model sometimes crashes. Since this method needs a scalar as input and model outputs are time series we used the sum of squares (ssq) as a summarizing criteria. The ssq was calculated for each measurement point individually and then summed up separately for the four target model outputs (target variable): soil water content, nitrogen content in the soil and crop characteristics, and evapotranspiration. Based on ssq, we calculated S1 for each parameter, site, year and target variable. Replacing the model output through the ssq in the sensitivity analysis yields S1 values that give information on how the model performance changes with respect to one parameter. Notice that this is different from the traditional sensitivity analysis which quantifies how changes in the model output depend on changes in parameters.

All simulations and calculations are performed on the bwUniCluster. The big Latin Hypercube samples (1,000,000 model runs) are distributed among 560 cores and take 10-12 hours to finish. The resulting 150 GB of model output per sample are also analyzed in parallel on the bwUniCluster.

III. RESULTS & DISCUSSION

Figure 1 shows the S1 values for the different years and study sites for each target variable calculated for a sample size of 100,000. Depending on the site and year, 30 % to 80 % of the original sample can be used for the analysis. The other samples lead to crashes of ExpertN for numerical reasons.

The parameters shown in figure 1 have an S1 greater than 0.05. Hence, each parameter explains > 5 % of the variance in the output variable. Note that, across target variables, the van Genuchten parameters *a1, a2,* and *l1*, the interaction parameter *mW, P1* as a crop parameter and the *km* parameter, related to evapotranspiration explain most of the models output variance.

The sum of first-order sensitivities (S1 values) of all parameters shown ranges between 0.65 and 0.81 for the different output variables. This means that 65% - 81% of the variance can be explained by first-order effects. Only in 2014 at EC6 for the output variable N the sum is only 0.45.

Upon comparison of the two sites, one can see that at EC6 the simulated soil nitrogen content and the evapotranspiration also depend on a crop model specific parameter. In 2011 and 2015 the crop model performance is
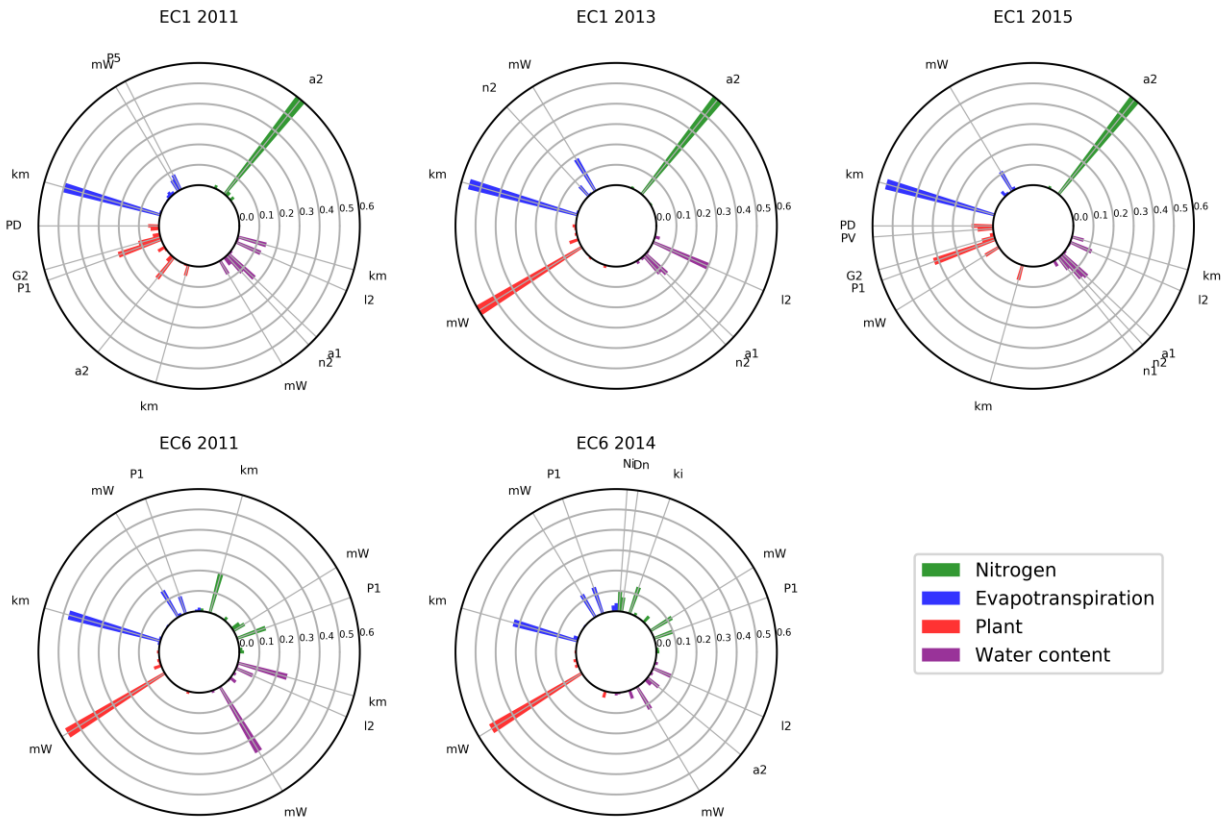

Figure 1. First order sensitivity indices (S1) for the two sites (EC1 and EC6) and the four years (2011, 2013, 2014 and 2015). Each circle plots the S1 for one year x site combination and the four target variables Nitrogen, Evapotranspiration, Plant and Water Content. The labels refer to parameters which have a S1 of higher than 0.05.

also related to crop specific parameters at EC1.

Rainfall during the vegetation period in these two years was 20-30% lower than in 2013 indicating that the amount of water entering the system affects the parameter sensitivities. In 2014 and 2015, vegetation periods were warmer than in 2011 and 2013, (+15% at EC1, 2015; +20% at EC6, 2014). Especially the winters were mild with mean temperatures over 0°C. However, this is apparently not affecting the parameter sensitivities. The shallower soil profile of EC6 results in an insensitivity to the van Genuchten parameter n as well as an insensitivity to the van Genuchten parameters in the first soil layers.

IV. CONCLUSIONS & OUTLOOK

We identified 6 parameters that explain most of the variance in soil-crop model performance. This result is independent from the targeted model output. However, the target variable affected the most by these six parameters is different between the two sites. Specific crop and carbon-nitrogen model parameters play a minor role although this result seems to be site- and year-specific. We conclude that the parameters belonging to the soil water regime are the most important for the overall model performance. Our planned next step is to calculate S1 independently from the measurements. S1 will be calculated for each simulation point and afterwards averaged in accordance with [6]. Besides, we will investigate if the S1s are affected by within-sample dependencies originating from the crashed simulations. Further, S1 values will be calculated for more sites and years to evaluate inasmuch the differences in the S1 values depend on year and site. We will also evaluate if our findings hold true with other popular crop models. We think that it is important to test to what extent parameters that control overall model performance affect also single target variables. The underlying hypothesis is that training soil-crop models with more comprehensive sets of target and non-target variables makes them more robust.

REFERENCES

[1] AgMIP, "The Agricultural Model Intercomparison and Improvement Project," http://www.agmip.org/, 20.03.2017.

[2] Allen, R. G.; Pereira, L. S.; Dirk, R.; Smith, Martin, "Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and Drainage Paper No. 56," FAO - Food and Agriculture Organization of the United Nations, 1998.

[3] Asseng, S. et al., "Uncertainty in simulating wheat yields under climate change," Nature Climate Change, vol. 3, pp. 827–832, 2013.

[4] Asseng, S. et al., "Rising temperatures reduce global wheat production," Nature Climate Change, vol. 5, pp. 143–147, 2014.

[5] Borgonovo, E., "A new uncertainty importance measure," Reliability Engineering & System Safety, vol. 92, pp. 771–784, 2007.

[6] Cuntz, M. et al., "Computationally inexpensive identification of noninformative model parameters by sequential screening," Water Resources Research, vol. 51, pp. 6417–6441, 2015.

[7] Finck, M., "Modellierung des N-Austrags unter Berücksichtigung regionaler N-Umsetzungsprozesse," Hohenheimer Bodenkundliche Hefte, 2010.

[8] Ingwersen, J. et al., "Comparison of Noah simulations with eddy covariance and soil water measurements at a winter wheat stand," Agricultural and Forest Meteorology, vol. 151, pp. 345–355, 2011.

[9] Ingwersen, J.; Högy, P.; Wizemann, H.-D.; Streck, Thilo, "Coupling the land surface model NOAHMP with the generic crop growth model GECROS: Model calibration and validation," Geophysical Research Abstracts, vol. 19, 2017.

[10] Johnsson, H.; Bergstrom, L.; Jansson, P.-E.; Paustian, Keith, "Simulated nitrogen dynamics and losses in a layered agricultural soil," Agriculture, Ecosystems & Environment, vol. 18, pp. 333–356, 1987.

[11] Martre, P. et al., "Multimodel ensembles of wheat growth: many models are better than one," Global change biology, vol. 21, pp. 911–925, 2015.

[12] McKay, M. D.; Beckman, R. J.; Conover, W. J., "Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics, vol. 21, pp. 239–245, 1979.

[13] Plischke, E.; Borgonovo, E.; Smith, Curtis L., "Global sensitivity measures from given data," European Journal of Operational Research, vol. 226, pp. 536–550, 2013.

[14] Ritchie, J., Godwin,D., "CERES Wheat 2.0," http://nowlin.css.msu.edu/wheat_book/, 01.12.2017.

[15] Rosenzweig, C. et al., "The Agricultural Model Intercomparison and Improvement Project (AgMIP)," Agricultural and Forest Meteorology, vol. 170, pp. 166–182, 2013.

[16] Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, Debora, "Global Sensitivity Analysis, 2008.

[17] Simunek, J.; K. Huang; M. van Genuchten, "The HYDRUS Code for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variables-Saturated Media," Tech. Rep. 144, U.S. Salinity Lab., United States Dep. of Agriculture, Agricultural Research Service, 1998.

[18] Wallach, D.; Mearns, L. O.; Ruane, A. C.; Rötter, R. P.; Asseng, Senthold, "Lessons from climate modeling on the design and use of ensembles for crop modeling," Climatic Change, vol. 139, pp. 551–564, 2016.

[19] Wizemann, H.-D.; Ingwersen, J.; Högy, P.; Warrach-Sagi, K.; Streck, T.; Wulfmeyer, Volker, "Three year observations of water vapor and energy fluxes over agricultural crops in two regional climates of Southwest Germany," Meteorologische Zeitschrift, vol. 24, pp. 39–59, 2015.

[20] Wöhling, T. et al., "Multiresponse, multiobjective calibration as a diagnostic tool to compare accuracy and structural limitations of five coupled soil-plant models and CLM3.5," Water Resources Research, vol. 49, pp. 8200–8221, 2013.

HPC-based uncertainty quantification for fluid-structure coupling in medical engineering

Jonas Kratzke

Engineering Mathematics and Computing Lab (EMCL)
IWR, Heidelberg University
Heidelberg, Germany
jonas.kratzke@iwr.uni-heidelberg.de

Vincent Heuveline

Engineering Mathematics and Computing Lab (EMCL)
IWR, Heidelberg University
Heidelberg, Germany

Abstract—**In recent decades biomedical studies with living probands (in vivo) and artificial experiments (in vitro) have been complemented more and more by computation and simulation (in silico). In silico techniques for medical engineering can give for example enhanced information for the diagnosis and risk stratification of cardiovascular disease, one of the most occurring causes of death in the developed countries. Other use cases for in silico methods are given by virtual prototyping and the simulation of possible surgery outcomes. High reliability is a requirement for cardiovascular diagnosis and risk stratification methods especially with surgical decision-making. Given uncertainties in the input data of a simulation, this implies a necessity to quantify the uncertainties in simulation results. Uncertainties can be propagated within a numerical simulation by methods of Uncertainty Quantification (UQ).**

For many cardiovascular applications, such as the simulation of blood flow through the aorta, the physiological understanding and mathematical modeling has still challenging aspects. Modeling the biomechanics of aortic blood flow by means of a three-dimensional fluid-structure interaction (FSI) problem, the simulation becomes computationally highly expensive. UQ usually multiplies the computational costs even further. We present an approach to compute patient-specific aortic FSI simulations incorporating UQ based on parallel computing. The coupled fluid- and elasto-mechanical problem is computed with a monolithic finite element (FEM) solver. The quantification of uncertainties is realized by a collocation method based on polynomial chaos (PC).

For validation purposes we consider an in vitro prototypical silicon phantom experiment and compare the simulation results with flow magnetic resonance images (MRI) of the phantom. Flow MRI is capable of visualizing the velocity field of the fluid flow. Taking the uncertainty into account, the simulation results are in well accordance with the MRI-measured velocity field. Furthermore, we present a patient-specific simulation method for simulating blood flow through the aortic bow based on high performance computing. Possible risk parameters can be evaluated in consideration of the uncertainty in the input data.

Keywords—Uncertainty Quantification, Fluid-Structure Interaction, Numerical Simulation, Blood Flow

I. INTRODUCTION

Cardiovascular diseases are one of the most occurring causes of death in the developed countries. Investigating the physiology of the cardiovascular system and the pathogenesis of its diseases leverages medical diagnosis and therapeutic possibilities. Greater knowledge can be acquired from medical studies with living patients and probands in vivo. Furthermore, investigations with artificial, i.e. in vitro experiments have been complemented more and more with virtual, so called in silico methods by means of computations and simulations. Overviews on the application of numerical simulations for investigating cardiovascular dynamics are given by [1,2,3]. For example, dynamics that are hard to be assessed non-invasively can be simulated supplementing pre-operative risk parameters. The outcome of different surgical procedures can be evaluated beforehand by means of simulations.

For the verification of cardiovascular numerical simulations, aortic phantom experiments can be used. In [4] a prototypical aortic phantom is presented. Fully consisting of non-metallic materials, the three-dimensional flow field can be measured in time by phase contrast magnetic resonance imaging (MRI) and compared to simulations of the same. Using a silicon material, the movement of the elastic vessel wall can also be observed in the phantom. As to the aorta, the elasticity of the complex vessel wall structure is crucial for damping the pressure waves induced by each heart beat. Though the layers of vessel wall soft tissue have been examined with respect to their biomechanics [5], it remains a challenging effort to simulate blood flow in elastic vessels in detail. One factor is given by the lack of non-invasive measurability of structural material parameters of a specific patient. The arising uncertainty in the soft tissue structure can be modeled by Uncertainty Quantification (UQ). In particular, UQ can provide a measure of the reliability of simulation results.

An overview on the modeling and numerical simulation of blood flow is given in [6]. Methods of UQ for flow simulations are reviewed in [7]. This work describes new developments in the simulation of blood flow coupled with elastic vessel wall movement by means of fluid-structure interaction (FSI) taking the uncertainty of input parameters into account. The presented framework is verified by means of MRI measurements of an aortic phantom experiment.

The paper is structured as follows: First, the underlying mathematical model is described. Section III gives an overview on the discretization and the utilized numerical methods. The simulation results for the phantom and a patient-specific aortic bow are presented in section IV. The conclusion gives a summary and an outlook to future work.

II. Mathematical Modeling

Figure 1 gives an overview on the considered prototypical aortic phantom. The general configuration is in accordance with a human aorta except for the presence of branching vessels and the complexity of the soft tissue structure.
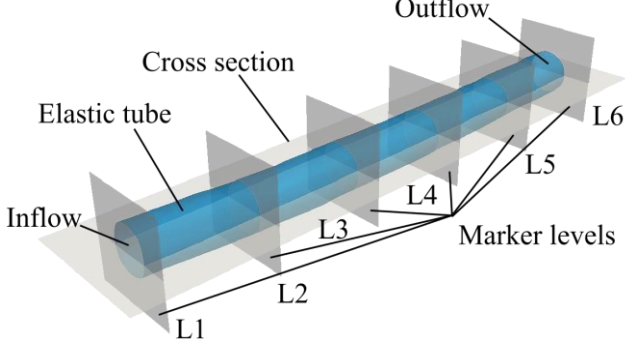

Fig. 1: Schematic overview on the aortic phantom used in [4]. A pulsatile flow is pumped through the inflow of the elastic vessel tube. The outflow is followed by a flow resistor. The flow was measured by phase contrast MRI at the cross-sectional planes of the marker levels.

A. Fluid flow

The flow dynamics of the glycerin-water mixture as used in the phantom experiment can be modeled by the incompressible Navier Stokes equations (NSE) for a Newtonian fluid. The NSE can also be used for modeling blood flow in large arteries, in which the shear-thinning influence of the red blood cells is relatively small.

$$\partial_t v + (v \cdot \nabla)v - \nu \nabla \cdot (\nabla v + \nabla v^T) + 1/\rho \nabla p = 0, \quad (1)$$
$$v \cdot \nabla = 0. \quad (2)$$

Equation (1) refers to the momentum conservation in the three-dimensional velocity and the one-dimensional pressure field v and p, respectively. (2) is derived from the conservation of mass. The glycerin (40%) to water (60%) proportions were chosen, such that the density ρ and kinematic viscosity ν is similar to blood. The inflow velocity of the simulation is set by means of a Dirichlet Poisseuille boundary condition scaled by the pulsatile flow rate measured at the marker plane L1. The inflow rate is modeled as an uncertain parameter of a uniform distribution with a maximal deviation of 15% from the measured value.

B. Elastic vessel wall structure

Aortic vessel wall soft tissue consists of three layers of cellular structures with different anisotropic biomechanical behavior. The phantom, however, can be modeled as a homogeneous hyperelastic Saint-Venant Kirchhoff material given by:

$$\rho \partial_{tt} u - \nabla \cdot (\lambda tr(E) I + 2\mu E) = 0, \quad (3)$$

The conservation of momentum, equation (3), governs the strain tensor E of the material which is defined by the three-dimensional displacement field u. Compressibility and stiffness are determined by the Poisson ratio γ and Young's modulus Y, respectively. They define the Lamé-constants $\lambda = Y\gamma(1 + \gamma)^{-1} (1 - 2\gamma)^{-1}$ and $\mu = 0.5Y(1 + \gamma)^{-1}$. The trace operator is denoted by tr and I is the identity matrix. To address the uncertainty in the generally hard to measure stiffness of vessel wall material, we model the Young's modulus Y as uncertain parameter of a uniform distribution with a maximal deviation of 30%.

C. Coupling of fluid flow and wall elasticity

In the mathematical problem formulation, coupling conditions model the interaction between fluid flow and vessel wall elasticity. At the interface, the displacement and the tension forces are assumed to coincide. The flow equations (1,2) are stated in the Eulerian frame of reference and the elasticity equations (3,4) in the Lagrangian frame of reference. A standard approach to couple these two perspectives is given by the Arbitrary Lagrangian-Eulerian (ALE) method. The ALE method introduces a mapping of the fluid domain from a reference state to the current configuration in time [8].

III. HPC Simulation

In order to numerically solve a complex stochastical boundary value problem as stated in section II, computationally efficient and parallel scalable methods have to be utilized. The large number of discrete unknowns requires the computation on high performance computing (HPC) clusters.

The FSI problem is solved in a monolithic way, resulting in a strong coupling of the fluid flow and vessel wall elasticity. The non-linearities in the mathematical problem are linearized by exact Newton-linearization. Finite elements of Taylor-Hood type discretize the spatial dimensions. The time steps are allocated by the one-step-θ-scheme. See for example [9] for details on the numerical methods. For the quantification of uncertainties, we use the generalized polynomial chaos expansion [10] with Legendre polynomials as a basis for the stochastic space. As a collocation method we use the Gauss-Legendre quadrature [11]. Each collocation point represents a single deterministic simulation run. This way, the code for the deterministic simulation does not have to be altered for the UQ study.

To numerically solve the discretized systems of equations, the open-source FEM framework HiFlow³ [12] is utilized on the HPC bwForCluster MLS&WISO production. Sufficient numerical accuracy is obtained with a polynomial chaos expansion degree of 3 resulting in 36 collocation points. Each collocation point is given by a number of approximately 1.6m FEM degrees of freedom for the aortic phantom and 1500 time steps. With a wall time of 32 hours on 64 CPU cores for each deterministic simulation, this leads to an overall computational demand of approximately 74k core hours.

$$\text{with } E = 0.5(\nabla u^T \nabla u + \nabla u^T + \nabla u). \quad (4)$$

IV. NUMERICAL RESULTS

A. Prototypical aortic phantom

Figure 2 shows a visualization of the aortic phantom simulation results. One can say that, with high standard deviation, the uncertainty of the results is highest at the time and location of maximal flow. The load stress in the vessel wall can be indicated by the von Mises stress as a comparative scalar field [13].

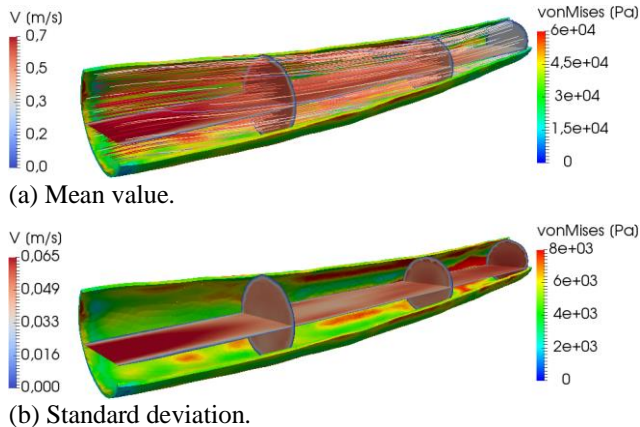
(a) Mean value.

(b) Standard deviation.

Fig. 2: Visualization of the aortic phantom simulation at the time step of highest inflow. The flow field is illustrated by streamlines and at cross-sectional planes. The vessel wall displacement is shown by the resulting von Mises stress.

The flow mean values in the cross-sectional planes indicated in figure 1 are compared to the MRI measurement in figure 3. The flow amplitude decreases with the distance to the inflow, which is an effect of the elasticity of the vessel wall. This effect cannot be observed in pure flow simulations with a rigid wall assumption, c.f. [14].
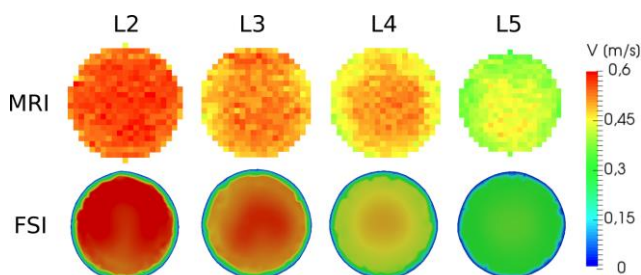

Fig. 3: Visualization of the flow through the cross-sectional planes at the marker levels depicted in figure 1 at the time of highest inflow. The corresponding MRI measurements and FSI simulation results are shown for comparison.

With regards to the volume flow rates through the cross-sectional planes, the simulation is in well accordance with the MRI measurement as can be seen in Figure 4.

B. Patient-specific aortic bow

We developed the framework for the simulation of an aortic phantom in a generic way, such that it can be directly applied to patient-specific geometries. Figure 5 shows a simulation result for the aortic bow of a healthy proband. Though the material models have to be adapted to human blood and to an
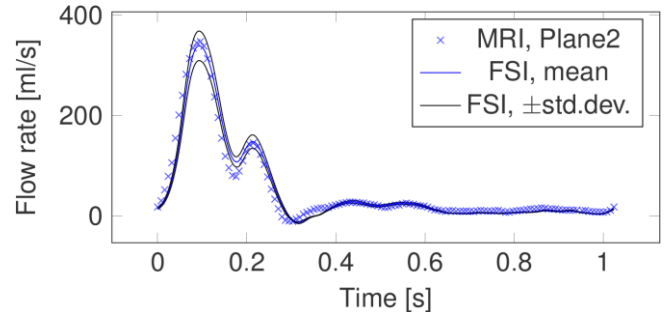

Fig. 4: MRI-measured volume flow rate through marker level 2 together with the simulated mean and standard deviation from the mean.

approximation of the complex structure of the vessel wall, the results reveal the feasibility and potential of uncertainty quantification for cardiovascular simulations. Hereby, the discrete geometry leads to 1.8m FEM degrees of freedom. The computational demand for a UQ simulation was 147k core hours.

V. CONCLUSION

This work presents a numerical framework for the patient-specific simulation of the fluid-structure interaction dynamics of aortic blood flow. The framework is verified by means of phase contrast MRI measurements of the flow field in a prototypical aortic phantom. Taking uncertainties into account, the simulation results obtain a measure of reliability, which is crucial for clinical decision-making.

Future work can concentrate on the incorporation of more realistic material properties of blood and vessel walls. Modeling the soft tissue as inhomogeneous structure, methods can be developed to include a spatially varying stochastic distribution of the material parameters. Furthermore the MRI-based flow measurements can be used not only to define boundary conditions but also to be assimilated by the simulation in order to utilize the simulation as enhancement of the medical imaging modality. The developed framework could finally be used in clinical trials to investigate risk factors for aortic diseases such as aortic stenosis, aneurysm development and vessel rupture.
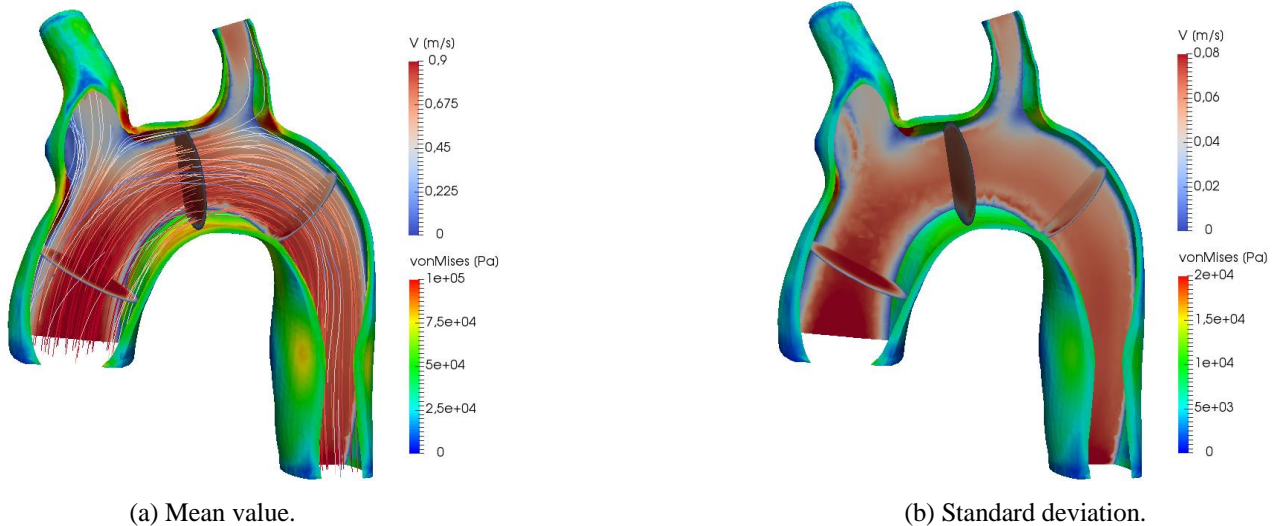
(a) Mean value.

(b) Standard deviation.

Fig. 5: Visualization of the FSI simulation of an human aortic bow at mid-systole. The flow field is illustrated by streamlines and at cross-sectional planes. The vessel wall displacement is shown by the resulting von Mises stress.

REFERENCES

[1] B. Chung, and J. Cebral, "CFD for evaluation and treatment planning of aneurysms: Review of proposed clinical uses and their challenges," Annals of Biomedical Engineering, vol. 43(1), pp. 122-138, 2015.

[2] J. B. Freund, "Numerical simulation of flowing blood cells," Annual review of fluid mechanics, vol. 46, pp. 67-95, 2014.

[3] A. L. Marsden, "Optimization in cardiovascular modeling," Annual review of fluid mechanics, vol. 46, pp. 519–546, 2014.

[4] M. Delles, F. Rengier, S. Ley, H. von Tengg-Kobligk, H. Kauczor, R. Unterhinninghofen, R. Dillmann, "Influence of imaging quality on magnetic resonance-based pressure gradient measurements," Proc. SPIE 7626, Medical Imaging 2010: Biomedical Applications in Molecular, Structural, and Functional Imaging, 762624, March 2010.

[5] G. A. Holzapfel, "Microstructure and mechanics of human aortas in health and disease," in Biomechanics: Trends in Modeling and Simulation, G. A. Holzapfel, and R. W. Ogden, Eds. Springer International Publishing, 2017, pp. 157-192.

[6] L. Formaggia, A. Quarteroni, and A. Veneziani, "Cardiovascular mathematics: Modeling and simulation of the circulatory system," vol. 1. Springer Science & Business Media, 2010.

[7] H. N. Najm, "Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics," Annual review of fluid mechanics, vol. 41, pp. 35-52, 2009.

[8] J. Hron and S. Turek, "A monolithic fem/multigrid solver for an ale formulation of fluid-structure interaction with applications in biomechanics," in Fluid-structure interaction, H. Bungartz and M. Schäfer, Eds. Springer, 2006, pp. 146–170.

[9] G. P. Galdi, R. Rannacher, A. M. Robertson, and S. Turek, "Hemodynamical flows," Birkhäuser Basel, 2008.

[10] D. Xiu, and G. E. Karniadakis, "The Wiener-Askey polynomial chaos for stochastic differential equations," SIAM journal on scientific computing vol. 24(2), pp. 619-644, 2002.

[11] F. Nobile, R. Tempone, and C. G. Webster, "A Sparse Grid Stochastic Collocation Method for Partial Differential Equations with Random Input Data," SIAM Journal on Numerical Analysis, vol. 46(5), pp. 2309-2345, 2008.

[12] S. Gawlok, P. Gerstner, S. Haupt, V. Heuveline, J. Kratzke, P. Lösel, K. Mang, M. Schmidtobreick, N. Schoch, N. Schween, J. Schwegler, C. Song and M. Wlotzka, "HiFlow[3] – Technical Report on Release 2.0", Preprint Series of the Engineering Mathematics and Computing Lab, vol. 06, 2017.

[13] A. Valencia, P. Burdiles, M. Ignat, et al., "Fluid Structural Analysis of Human Cerebral Aneurysm Using Their Own Wall Mechanical Properties", Computational and Mathematical Methods in Medicine, vol. 2013, Article ID 293128, 2013.

[14] J. Kratzke, F. Rengier, C. Weis, C. Beller, and V. Heuveline, "In vitro flow assessment: From PC-MRI to computational fluid dynamics including fluid-structure interaction," Proc. SPIE 9783, Medical Imaging 2016: Physics of Medical Imaging, 97835C, April 2016.

High precision calculations of particle physics at the NEMO cluster in Freiburg

C. Heidecker, M. Giffels, G. Quast, K. Rabbertz, M. Schnepf

Institut für Experimentelle Teilchenphysik (ETP)

Karlsruher Institut für Technologie (KIT)

Karlsruhe, Germany

Abstract— **At the Large Hadron Collider of CERN in Geneva particles such as protons are smashed onto each other at the highest man-made energies ever. Despite being gigantic in size, custom-designed detectors are capable of recording billions of such collisions with high accuracy. To accompany these measurements, predictions of highest precision are required, for some of which it took the theory community more than 20 years to develop the necessary methods and techniques. We explain how these CPU intensive calculations have been adapted to run efficiently on the computing resources available within the bwHPC project and present the required tools and infrastructure we developed for this purpose.**

Keywords—particle physics; software deployment; resource allocation

I. Introduction

To fully profit from precise measurements of high-energy particle collisions by modern particle detectors, these measurements must be matched with at least equally precise theoretical predictions. For the example of proton-proton collisions at the Large Hadron Collider (LHC) of CERN in Geneva, the theory of choice is quantum chromodynamics (QCD) [1], which is assumed to describe all phenomena involving the strong interaction or, in more common terms, nuclear interactions. Although solutions to the equations of QCD are not known in general, particle collisions at high energy, or equivalently very small distance, can be described by a perturbative approach to QCD (pQCD). Leading-order (LO) results in pQCD have been derived already in the seventies but provide only an order-of-magnitude estimate for the production rates of particular event types. More specifically, the example events used in this article contain at least one high-energetic "particle jet", which is the name given to a bundle of strongly interacting particles leaving a collision in roughly the same direction. Such jets are considered to be the visible manifestation of otherwise not freely observable proton constituents like quarks [2]. Next-to-leading order (NLO) predictions for jet production improve the relative precision to a level of about 5-10% and have become available in the nineties [3,4]. The computing time (CPU time) required for one such calculation is of the order of 1,000 hours. The pQCD calculations at next-to-next-to-leading order (NNLO) needed nowadays for an ultimate precision of 1-2% have recently been completed [5]. They require more than 100,000 hours of CPU time per desired observable.

In the following we describe the solutions we developed for the deployment of task-specific software to the worker nodes and for an efficient usage of the shared resources of the NEMO high-performance computing (HPC) cluster in Freiburg [6]. Their successful application is exemplified by means of a typical NNLO computation involving jet production.

I. Software environment

Large experimental collaborations in particle or high energy physics (HEP) develop and utilise their software within an environment adapted to their workflows. This implies the use of a particular operating system as well as numerous specific software packages. Computing resources like the NEMO HPC cluster that serve multiple user communities in the state-wide bwHPC concept cannot provide dedicated worker nodes without compromising their usability by all users. Therefore, a solution is sought to provide the software specific to the HEP community to the worker nodes of the NEMO cluster without perturbing other applications.

A. Static provisioning of software

One part of the solution is based on the virtualisation option of the HPC centre [7], which makes it possible to provide community-specific environments. Such a preconfigured environment can comprise i.a. a dedicated operating system with pre-installed HEP specific software packages.

B. Dynamic provisioning of software

Additionally, workflow- and user-specific software and tools that change frequently need to be provided. Here, only a more dynamical solution than virtualisation is feasible.

For purpose B we employ a remote file system called CERN Virtual Machine File System (CVMFS) [8] that is commonly used in HEP. This file system was developed to deliver software packages to all world-wide distributed LHC computing centres.

28

On the server side, files are hosted on web servers, which are organised in repositories. The utilisation of the HTTP protocol simplifies the world-wide distribution of files as well as the usage of caching proxies to reduce incoming traffic. In their standard configuration such proxies [9] cache all files with short life-cycles that are transferred via the HTTP protocol. Hence, software packages can either be provided directly via CVMFS or via other web services.

On the client side, CVMFS provides a common POSIX read-only file system in user space, which allows the user transparent access to files inside the repositories. To further reduce traffic, each CVMFS client uses an additional cache.

For usage within the NEMO HPC cluster, we installed two caching proxies, which serve as central cache and provide CVMFS to all worker nodes. Each one has a cache size of 240 GBytes and is directly connected to the internal worker node network. For even more flexibility, we have set up our own CVMFS server at KIT in addition to the CVMFS repositories provided by CERN.

This concept of static provisioning of the basic software environment via virtual machine images and the dynamic loading of workflow- and user-specific software packages via CVMFS has been operated successfully at the NEMO HPC cluster and is now in standard use for large-scale computing campaigns within our physics workflows. As a consequence, we were also able to adapt and test it for the deployment of our specific software environment to other shared resources like commercial cloud providers. Furthermore, other scientific communities with similar requirements could equally profit from the developed concept.

II. RESOURCE PROVISIONING

A. Usability requirements

For an efficient usage of the shared resources at the NEMO cluster, a dynamic setup is required that matches resource allocation to demand of computing power. Furthermore, a simplified and standardised system to submit workflows is necessary to make virtualised NEMO resources transparently accessible to end-users at the KIT.

Since the HEP community at the KIT manages computing resources via the batch system HTCondor [10], we have chosen to integrate the NEMO resources into our HTCondor pool. This necessitates an interface between the HTCondor system, which manages our workflow processing, and the Moab/Torque batch system [11] that is responsible for the resource management at NEMO. The life-cycle management of the virtual machines is performed by the OpenStack instance [12] at NEMO. For a successful operation, all three components must interact with each other.

B. Virtualisation management at NEMO

At the NEMO cluster, communication between the OpenStack instance and the Moab batch system is required to match virtual machines to allocated resources. Here, the assignment is handled by an automated system developed by
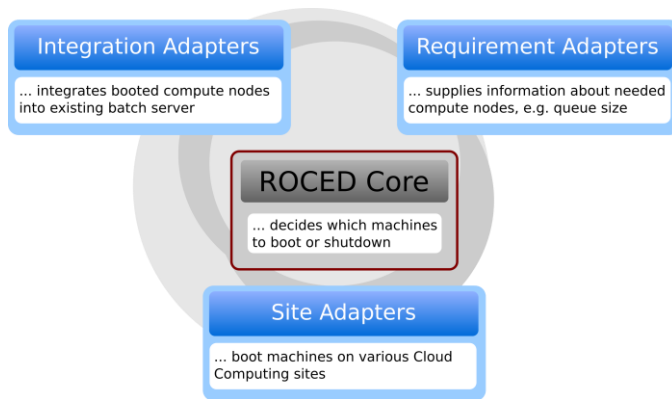

Fig. 1. Modular structure of ROCED based on three independent adapters connected via the ROCED core.

the NEMO HPC team. For the integration into our HTCondor pool, the demand of resources has to be translated into requests to the NEMO cluster. Furthermore, newly allocated resources must be added, while unused resources are deallocated and removed from our HTCondor pool.

C. Resource scheduler ROCED

As a solution, we developed the resource scheduler responsive on-demand cloud-enabled deployment (ROCED) [13], which is designed to interact with various batch systems and resource providers. ROCED monitors the demand for resources and the current state of the resource pool. If the demand is higher than the amount of free resources, ROCED sends a resource request to a provider such as an HPC centre or a cloud provider.

For that purpose, a modular structure of independent adapters was built around the ROCED core. These include requirement adapters, site adapters, and integration adapters as shown in Fig. 1. The requirement adapter monitors the HTCondor pool at the KIT and collects information about pending workflows waiting for free resources. Simultaneously, the site adapter manages the requested resources at the NEMO cluster. After detecting a resource demand, the ROCED core decides to request additional resources via the corresponding site adapter that communicates with the resource provider. In our setup, this is performed by sending a batch job with a request for booting an OpenStack virtual machine to the Moab batch system at NEMO as shown in Fig. 2. Subsequently, the integration adapter adds the new resources into the HTCondor pool at KIT. The fact that HTCondor worker nodes automatically connect to the configured pool at startup simplifies this step.

In our current setup, a maximum runtime of one day is configured for the virtual machines. Up to 8,000 CPU cores, the maximum that can be allocated to one user, can be used. Each of these virtual machines only accepts user workflows with a suitable runtime and automatically shuts down either when reaching its runtime limit or when being idle for some time. For testing reasons, the maximum was raised to over 11,000 CPU cores so we could demonstrate the capability of

our setup to dynamically request, boot up, and shut down such an amount of virtual machines within a short duration. This demonstrated flexibility is also observed during normal daily operation. Fig. 3 shows our usage of the NEMO cluster as a function of time for one month. The blue area represents the number of CPU cores running user workflows, while the red area corresponds to the number of idle CPU cores. For the latter, matching workflows could not be found because of memory, disk space, or runtime constraints resulting in a small share of allocated but unused resources. Red and blue numbers together represent the total number of CPU cores allocated at the NEMO cluster. The excess of idle CPUs that can be seen versus the end of the one-day runtime of the virtual machines is caused by the fact that jobs with a suitable runtime cannot be found anymore and new virtual machines must be started.
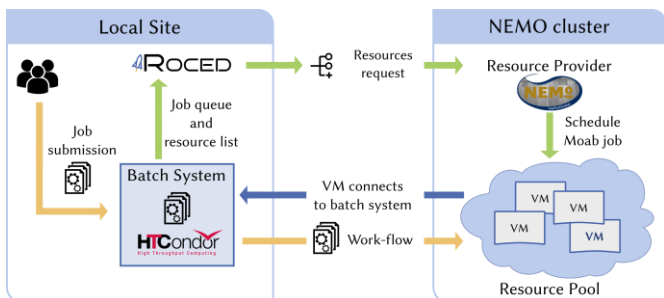

Fig. 2. Management cycle of ROCED for interaction between our local user group at KIT and the NEMO cluster.

The demand for computing resources varies over time. Hence, a highly dynamical setup like ours facilitates the efficient usage of shared resources, which are allocated only during processing of workflows and are given back afterwards. Also, the resources of large (commercial) computing providers can be accessed in a similar way opening up additional possibilities for peak demands of scientific projects.

III. Application workflow

A typical pQCD calculation at NNLO performed by us on the NEMO cluster proceeds in three steps. First, the setup for the numerical integration of many complex multidimensional integrals must be optimised. This is performed through direct submission with Moab of one multicore job per job type of the NNLOJET program [5] and needs about 2,000 hours of total CPU time. Secondly, interpolation grids for later fast reevaluations for varying input parameters such as assumptions on the proton structure are prepared by exploring the phase space in terms of proton momentum fractions and the energy scale accessible in the examined process. Two software packages are available for this technique, fastNLO and APPLGRID [13,14], which can save enormous amounts of computing time by avoiding to rerun the full NNLO computation in many cases. We apply the latest version of the fastNLO framework [15]. In the third step thousands of single-core jobs are submitted in parallel, each for about 20h of CPU time, in order to evaluate the prepared integrals by generating
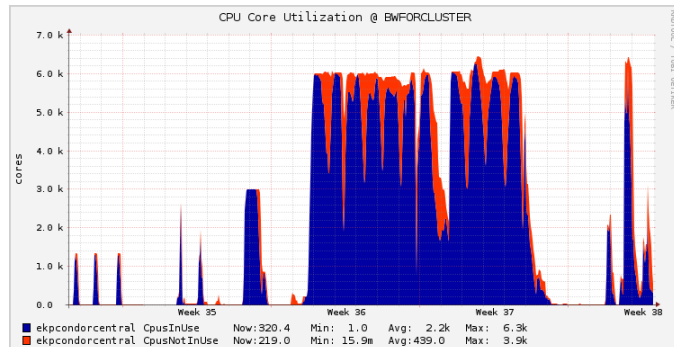

Fig. 3. Allocation and deallocation of virtual machines during one month.

pseudo-events and filling the interpolation grids. Table I gives an overview of the resources typically required in this last step. In a test campaign on the NEMO cluster we were able to finish one such computation in only two days with at maximum 7,800 cores working in parallel.

Table 1. Production Campaign of a typical pQCD calculation at NNLO accuracy. The job type corresponds to different parts of such a computation, which are evaluated separately.

Job type	# Jobs	Events / job / millions	CPU time / job [h]	# Events / billions	Total output [GBytes]	Total CPU time [h]
LO	10	140	20.6	1.4	0.024	206
NLO-R	200	6	19.0	1.2	1.3	3,800
NLO-V	200	5	21.2	1.0	1.2	4,240
NNLO-RRa	5,000	0.06	22.5	0.3	26	112,500
NNLO-RRb	5,000	0.04	20.3	0.2	27	101,500
NNLO-RV	1,000	0.2	19.8	0.2	6.4	19,800
NNLO-VV	300	4	20.5	1.2	2.0	6,150
Total	11,710	–	–	5.5	64	248,196

IV. summary

To cope with frequent changes in community- and user-specific software environments we use the virtualisation option of the NEMO cluster to provide a static virtual machine image and dynamically load workflow specific software via CVMFS. For an efficient and user-friendly utilisation, we integrate virtualised resources of the NEMO cluster into our HTCondor batch system. To manage these resources dynamically, we employ the resource scheduler ROCED developed at KIT.

Precision measurements at the forefront of physics require equally precise theoretical predictions leading to the most CPU intensive workflow we have successfully exercised with

the described setup and the large resources available to us thanks to the NEMO cluster in Freiburg. As demonstrated, the presented components allow us to run workflows with about 100,000 CPU hours in just days. Moreover, our presented concepts can be adapted to similar needs of other scientific communities and can facilitate the access to resources of commercial computing providers for example in times of peak demands.

REFERENCES

[1] R.K. Ellis, W.J. Stirling, B.R. Webber, "QCD and Collider Physics", Cambridge University Press, 1996.

[2] K. Rabbertz, "Jet Physics at the LHC", Springer, Berlin, 2016.

[3] S.D. Ellis, Z. Kunszt, D.E. Soper, "One-jet inclusive cross section at order alpha_s^3: Quarks and Gluons", Phys. Rev. Lett. 64 (1990), 2121.

[4] W.T. Giele, E.W.N. Glover, D.A. Kosower, "Higher order corrections to jet cross-sections in hadron colliders", Nucl. Phys. B 403 (1993) 633.

[5] J. Currie, E.W.N. Glover, J. Pires, "Next-to-Next-to Leading Order QCD Predictions for Single Jet Inclusive Production at the LHC ", Phys. Rev. Lett. 118 (2017) 072002.

[6] NEMO HPC Cluster https://www.hpc.uni-freiburg.de/nemo accessed 01.03.2018

[7] K. Meier et al., "Dynamic provisioning of a HEP computing infrastructure on a shared hybrid HPC system", J. of Phys., Conf. Ser. Vol. 762 (2016) 012012.

[8] P. Buncic et al., "CernVM – a virtual software appliance for LHC applications", *J. of Phys., Conf. Ser. Vol. 219 Part 04 (2010) 042003.*

[9] Squid proxy http://www.squid-cache.org/ accessed 10.12.2017

[10] HTCondor https://research.cs.wisc.edu/htcondor/ accessed 10.12.2017

[11] Moab HPC Suite http://www.adaptivecomputing.com/products/hpc-products/moab-hpc-basic-edition/ accessed 10.12.2017

[12] OpenStack https://www.openstack.org/ accessed 10.12.2017

[13] ROCED https://github.com/roced-scheduler/ROCED accessed 10.12.2017

[14] T. Kluge, K. Rabbertz, M. Wobisch, "fastNLO: Fast pQCD calculations for PDF fits", in Proc. of 14ᵗʰ International Workshop on Deep Inelastic Scattering (DIS 2006), doi:10.1142/9789812706706_0110.

[15] T. Carli et al., "A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project", Eur. Phys. J. C 66 (2010) 503.

[16] D. Britzger, K. Rabbertz, F. Stober, M. Wobisch, "New features in version 2 of the fastNLO project", in Proc. of 20th International Workshop on Deep Inelastic Scattering (DIS 2012), doi:10.3204/DESY-PROC-2012-02/165.

SDS@hd – Scientific Data Storage

Martin Baumann, Vincent Heuveline, Oliver Mattes, Sabine Richling, Sven Siebler

University Computing Centre (URZ)

Heidelberg University

Germany

Abstract. SDS@hd (Scientific Data Storage) is a central storage service for hot large-scale scientific data that can be used by researchers from all universities in Baden-Württemberg. It offers fast and secure file system storage capabilities to individuals or groups, e.g. in the context of cooperative projects. Fast data accesses are possible even in case of a high number of small files. User authentication and authorization are implemented in terms of the federated identity management in Baden-Württemberg allowing researchers to use their existing ID of their home institution transparently for this service. Data protection requirements can be fulfilled by data encryption and secure data transfer protocols. The service is operated by the computing center of Heidelberg University.

I. INTRODUCTION

In many fields of research, the capacity of data that is generated in scientific projects is enormous and continuously growing. This is a consequence of technical progress in data generating devices, such as high-throughput microscopes, telescopes and genome sequencers amongst others, and also for fast computer systems such as high-performance compute clusters or cloud systems that can be used e.g. for numerical simulations of complex processes. Suitable IT services for data storage and for data analysis for large data sets are key enabler. Besides the technical prerequisite, research projects have additional requirements. These include adequate group- and access management for co-operational setups and data protection for some projects with sensitive data. In the context of typical data life cycles for research data and of good scientific practise, a subset of the data has to be stored for long time (archive) or be accessible by a group of people or the public. Additional requirements for the collection of meta data and the transfer to corresponding long-term archive and repository services exist.

The second generation hardware of the "Large Scale Data Facility" (LSDF2) addresses this need of data storage in the context of research projects. LSDF2 is financed by the Ministry of Science, Research and the Arts (MWK) and the German Research Foundation (DFG). It is part of the state of Baden-Württemberg's concept for data-intensive services bwDATA [1]. LSDF2 is a storage system and at the same time a joint project by the Computing Center of the Heidelberg University and the Steinbuch Centre for Computing at Karlsruhe Institute of Technology. In the framework of this project, storage capacity is installed on both sites and operated by the two computing centers. The cooperation of the two institutions and also the use of almost identical technologies for the two systems - both on the hardware and software level – create synergies. The data distribution across university boundaries is possible through the respective storage services. This is described in detail in section III.C. In the following, the focus is on the details and use of the LSDF2 located in Heidelberg.

A new scientific data storage service "SDS@hd" [2] has been developed with the intention to improve the value of available data storage resources in the context of research projects. The service is tailored to those phases in the research data life cycle in which frequent accesses are present, i.e. for so-called 'hot data'. The data is stored on the LSDF2 in Heidelberg and can be accessed via local systems (e.g. by using a local high-performance compute cluster) or also from remote systems and even directly from the user's desktop computer. The service is open to all scientists at Baden-Württemberg's universities in the sense of a "Landesdienst".

Subsequently in this publication, the concept and technical specifications of the storage system LSDF2 in Heidelberg will be described in detail. This is followed by a presentation of the storage service SDS@hd including its most important features such as access protocols, management system and data security. After that, the connection to local and remote high-performance compute clusters and the typical usage scenarios are outlined. Finally, support and consulting activities are described.

II. LSDF2 HARDWARE

SDS@hd is a service which runs on the storage system LSDF2 located at Heidelberg University. The operational concept considers failover events. Therefore the storage system consists of two standalone systems – primary and secondary system – which are housed in different fire compartments of the Heidelberg University Computing Centre. The structure of the LSDF2 (light gray parts) as well as the connection to other systems at Heidelberg University and other external partners is shown in Fig. 1 LSDF2 was put into operation in 2016.

The primary LSDF2 storage system in Heidelberg is currently based in the back-end on a HPE Seagate G200 system built by a cooperation of the companies HPE and Seagate (nowadays Cray). The secondary system is based on Dell hardware. These main storage building blocks are extended by additional network components, as well as access- and administration servers. As file system software "IBM Spectrum Scale (Advanced Edition)" was chosen for both the primary

and secondary system. This file system provides high performance and good scaling characteristics and offers many valuable features like snapshotting, information lifecycle management, cluster export services for NFS, SMB, and S3 for concurrent accesses to the same namespace. By this, IBM Spectrum Scale allows for an efficient and reliable operation of the productive storage system LSDF2.

In addition to the separation into two fire compartments, both stand-alone system parts are high-available. There is no single point of failure due to a two- or many-times redundancy of all components – including the network, switches and external connection – depending on their criticality. The power supply is backed with an uninterrupted power supply (UPS). As internal storage network an Infiniband FDR network is utilized, to be able to provide the required bandwidth and
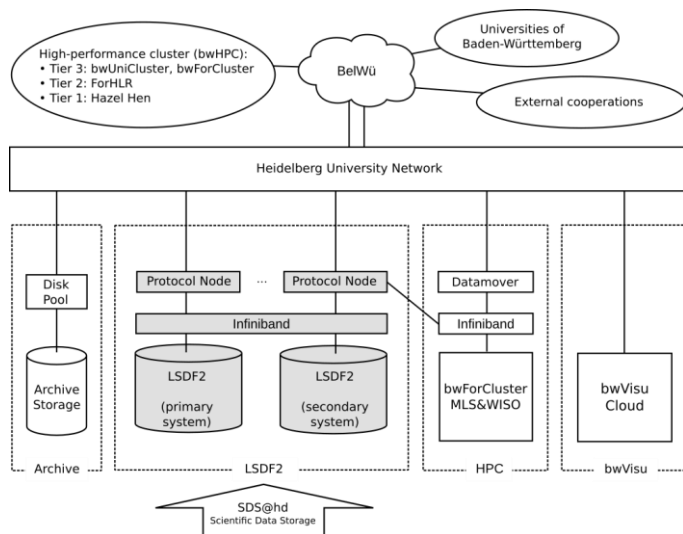
access times in the back-end.

Fig. 1. LSDF2 concept at Heidelberg University and its connection to other data-intensive services at URZ and Baden-Württemberg.

The storage system is highly scalable, designed for a staged gradual expansion. The current first expansion stage of the primary system is equipped with a usable capacity of 5.8 PB. As storage media in total more than 1000 hard disk drives (HDD) and solid-state drives (SSD) are installed. The SSDs are used to store the metadata and also small files to enhance the performance. The data storage is organized in a parity declustered RAID protection scheme called GridRAID [3], which distributes the parity information over groups of 42 drives. In addition, the spare space is distributed over all drives in a storage group. In case of a disk drive failure, the parity is automatically reconstructed on all remaining drives during the recovery phase. With the distribution of the reconstructed data across all surviving drives, the important MTTR (mean time to recover) is no longer depending on the I/O performance bottleneck of a single drive and is reduced to less then a quarter compared to the excessive rebuild times of classical RAID 6 on disk drives with multiple TB space [4]. After replacing the

failed disk drive, only a low-priority rebalancing needs to be done.

A scalable number of protocol nodes is used to provide the data access for the users via different file system protocols. These protocol nodes have Infiniband connections to the internal network as well as 40 Gbit/s Ethernet (40GE) connections directly to the backbone of the university and finally to the BelWü network [5] which connects the universities of the state of Baden-Württemberg. Currently the primary system can be reached via six 40GE network connections, so that data access with high bandwidth can be assured to users from Heidelberg as well as other universities of Baden-Württemberg. If needed, the aggregated bandwidth can be scaled up by increasing the number of protocol nodes and network connections.

For fail-over reasons data is stored on both systems, the primary and the secondary system, which are directly connected. In case of an outage of the primary system, the standalone secondary system can be accessed over a similar setup of network connections and protocol nodes. The secondary system provides a spatially separated copy of the data, which is available instantaneously without long rebuilding times from additional backup solutions. The throughput is reduced but the user data access can be directly enabled with a reduced performance. The purpose of the second storage system is not performance, but protection against loss of data and maintenance of service- and data-availability. The technical concept of the second storage includes one tier of online storage (ca. 50% of the primary system's capacity) and one tier of offline storage which leads to a good cost-effectiveness ratio. Since the second storage system contains only copies of data, its capacity does not lead to an increased usable total storage capacity of the service SDS@hd, but to a higher service quality.

III. SDS@HD SERVICE

A. Service description

The storage service "SDS@hd – Scientific Data Storage" serves as central and secure data storage of large-scale scientific research data. High performance data access in combination with high availability, data security and encryption for storage of "hot data" are the main properties of this service. "Hot data" means scientific data, which is often accessed and part of current active research, e.g. of data-driven sciences like life sciences, astrophysics or hydromechanics.

In the typical life cycle of scientific data, SDS@hd takes the role of the central storage which is accessed in different steps from data creation – e.g. from microscopy, gene sequencing or high performance computing (HPC) simulations – to data processing, analysis or visualization e.g. with compute clusters. When research activities with certain data are finished, the concerned data then can be handed over to repositories or archive solutions.

B. Access protocols

SDS@hd provides access via different data access protocols in order to enable data access from all commonly used operating systems which are used in the research community.

TABLE I.   DATA ACCESS PROTOCOLS

Protocol	Properties	Recommended OS
SMB 2/3	high availability, fault tolerance and load balancing	Windows, OS X
NFSv4	high availability, fault tolerance and load balancing	Linux, *nix
SSHFS / SFTP	easy access without special installation and configuration	all

As can be seen in TABLE I. currently the three protocol types SMB, NFS and SSHFS/SFTP are provided. Both NFSv4 as well as SMB are intended for a stable connection of data sources like microscopes or gene sequencers, compute clusters, servers and workstations. The connection is highly available, fault tolerant and of high performance. SMB is intended for access from Windows and OS X, NFS for connections from Linux (or other Unix-like) systems. For both protocols the installation needs a minimal effort and support of the network firewalls. In case of the NFS connection also a handling of Kerberos tickets is necessary. However, the usage of the SSHFS/SFTP protocol is intended for accesses from all operating systems especially from desktop or mobile systems and from systems in restrictive or external networks. Additional protocols like S3 will be complemented based on the user's demands.

As mentioned in Section II user access of SDS@hd is managed by the protocol nodes, which easily can be scaled in number to handle further throughput demands of the data storage service.

C. Management and registration

For the management of the storage resources and accesses, so-called Speichervorhaben (SV) are introduced as the organizational units of SDS@hd. One SV consists of one storage share with a defined quota and it corresponds to a dedicated group of users. Members of the same SV are able to share data easily while the access of non-members by default is not possible. There is a responsible person for each SV who takes responsibility for the data stored in the SV and also for the reporting (e.g. for evaluations by the DFG). The SV responsible can manage members and roles for his or her SV via a web management tool. A fine-grained access management within one SV on the level of user roles is currently in development.

The use of SDS@hd is in general possible for all bwIDM member organizations and involves a registration procedure on the user level. bwIDM [6] is the federated identity management of Baden-Württemberg's universities which is realized as sub-federation of the DFN-AAI [7]. This technology allows the researchers to use their ID of their home institution when using

SDS@hd. The access is controlled by the entitlements "sds-hd-user" and "sds-hd-sv" which are granted by each organization for their own staff and students.

Opening a new SV for a scientific project requires the entitlement "sds-hd-sv" that allows to initiate the creation of a contract between the SV responsible person and the SDS@hd operating institution which regulates usage, billing, and reporting issues. The entitlement "sds-hs-user" is required to use SDS@hd and to join an existing SV.

Before the first access to SDS@hd can be done, an additional registration step is needed in which a dedicated service password is set. At the same time, the user's account on the storage system is created.

D. Data security

In several data-intensive research areas like medicine and life science, there are high demands in data protection, especially when data from humans has to be stored (e.g. human gene sequencing or microscopy data). An overview of data protection concepts and technologies in the exemplary context of life science research is given in [8]. The authors conclude and summarize what is in principle well-known in the community: data security is related not only to technology or infrastructure – instead processes have to be designed in a way that the demands on data security can be fulfilled.

The service attributes of SDS@hd are such that particular data protection requirements can be met that might exist for various projects. In the following, relevant technologies, processes and concepts in that context are listed: The user access utilizes kerberized authentication and secure transport protocols. The owner of a share can independently manage the membership of a group of users to access the share and participate as a joint research project. In the backend of the storage service, an automatic and hardware assisted data encryption of all files is used, thus only encrypted data is stored on the disk drives. In case of failures, the defect drives remain and are shredded in the building. As usual, the access to the facilities and systems is restricted with personal access control. Specific demands on data security that users might have for certain projects can be checked individually by the operators of the service.

IV. ACCESS FROM BWHPC CLUSTERS

Many scientific communities need high performance computing (HPC) resources for data analysis. To facilitate this process, it is planned to provide access to SDS@hd from all bwHPC clusters [1] in Baden-Württemberg via datamover nodes. The connection can be established using any protocol that SDS@hd offers (see TABLE I.), whereby NFS is the recommended protocol for a production operation. Both the cluster file system and SDS@hd are connected with high bandwidth to the datamover nodes to allow fast copy processes and integration of the remote storage system into the cluster job management via data staging (yellow lines in Fig. 2). This concept is supported by the upgrade of the Baden-Württemberg scientific network BelWü to 100 Gbit/s. Datamover nodes are

already in production for the bwForCluster MLS&WISO (Production and Development) in Heidelberg [9] and the bwForCluster BinAC in Tübingen [10].
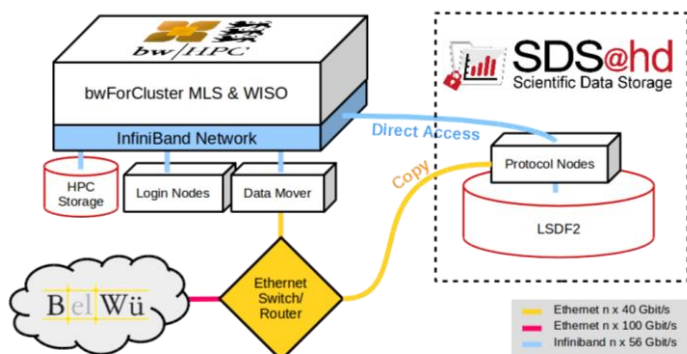

Fig. 2. Access to SDS@hd from bwForCluster MLS&WISO direct access (blue) and copy access (yellow). The latter is possible for other HPC clusters (e.g. bwHPC resources).

Due to the spatial proximity between the cluster and the storage system in Heidelberg, it is possible to provide a direct connection to SDS@hd on the compute nodes of bwForCluster MLS&WISO Production via InfiniBand (blue line in Fig. 2). This allows the direct analysis of SDS@hd data on the cluster without data staging or manual data synchronization. However, also for this direct network connection, currently only NFSv4 connections with Kerberos tickets are used. Other and potentially faster protocols for direct accesses are in preparation.

The management of Kerberos tickets on the cluster is done on the datamover nodes. Tickets are automatically renewed as long as possible. If security settings require manual ticket prolongation with password authentication, the user is notified by e-mail in time [11]. Direct access to SDS@hd on the compute nodes is used with growing interest.

The achievable performance depends on the workflow and the access patterns and is evaluated and improved in collaboration with the users. Currently, a series of performance measurements for different use cases and setups are made. The performance is analyzed systematically and compared to parallel scratch storage systems available in HPC clusters. The results will be publicated in a scientific contribution.

A close connection to SDS@hd is also in preparation for the remote visualization service bwVisu [12], which is located in Heidelberg. bwVisu will enter production state in 2018 and will provide resources for scalable and flexible data analysis and visualization.

V. User Meetings, Support and Consulting

The service SDS@hd is designed as generic as possible to meet the needs of a wide class of research projects. A flexible group, right and role management is possible that can be setup to fit to different workflows and multi-user and multi-role environments. However, to obtain the best result for specific problem classes, the capabilities of the storage service and the

characteristics and limitations of the underlying storage system have to be well-understood. In some cases, the way how the research data is managed needs to be adapted for increased performance (e.g. conversion of many small files into fewer larger files). In regular user meetings the features and characteristics of SDS@hd and LSDF2 are presented to users and people who are interested in these topics. Showcasing the use of SDS@hd for representative project setups seems to be particularly beneficial for the users and is one main focus in these meetings. The use cases range from a single-user direct access scenario to setups with multi-users having different status, sharing parts of their data and using specific external analysis systems. These meetings have proven to be very valuable both for the service providers (to get feedback from the users) and for users (to discuss needs for their specific data-intensive projects).

The feedback from users and the utilization statistics are considered in the planning of further expansions. As outlined before, the storage system has a modular structure where capacity and performance can be extended and increased on a demand-driven way. Optimizations on the level of the network connection to local systems on site and also of the uplinks for cross-locational use are possible and will be implemented when needed. For the further development of the storage service, several ideas exist (e.g. access via S3, utilization of user datagram protocol (UDP) or data staging for accesses from high-performance compute clusters) which are prioritized by the expected added value for the users.

VI. Summary

The service SDS@hd for hot large-scale scientific data was introduced. The technical aspects of the underlying storage system with its components and characteristics were outlined. Subsequently, the storage service was described including the available access protocols, the management concept, the registration process, and data security aspects. The connection and interplay between Baden-Württemberg's HPC resources and SDS@hd was described and also the dedicated support that is given to users.

With the storage system LSDF2, the storage service SDS@hd and the accompanying activities the operators hope to provide a valuable service and support for the data-intensive scientific community in Baden-Württemberg.

VII. Acknowledgments

References

[1] Rahmenkonzept der Hochschulen des Landes Baden-Württemberg für datenintensive Dienste – bwDATA (2015-2019): http://dx.doi.org/10.15496/publikation-21187

[2] SDS@hd – Scientific Data Storage. https://sds-hd.urz.uni-heidelberg.de

[3] Seagate. "Seagate ClusterStor G200 with IBM Spectrum Scale" (Datasheet), 2015

[4] Roskow, M. "The Unique Technical Benefits of an Engineered Solution for GPFS" (Presentation), 2015

[5] BelWü - das Landeshochschulnetz. https://www.belwue.de/

[6] Föderiertes Identitätsmanagement der baden-württembergischen Hochschulen. http://bwidm.de/

[7] DFN-AAI - Authentication and authorization infrastructure. https://www.aai.dfn.de/en/

[8] Helvey, T., R. Mack, S. Avula and P. Flook. "Data security in Life Sciences research", Drug Discovery Today: BIOSILICO, Volume 2, Issue 3 (2004): 97-103, https://doi.org/10.1016/S1741-8364(04)02403-5

[9] bwForCluster MLS&WISO, http://www.bwhpc-c5.de/wiki/index.php/Category:BwForCluster_MLS&WISO

[10] bwForCluster BinAC. http://www.bwhpc-c5.de/wiki/index.php/Category:BwForCluster_BinAC

[11] Richling, S., M. Baumann, S. Friedel and H. Kredel. "bwForCluster MLS&WISO" in Proceedings of the 3rd bwHPC-Symposium: Heidelberg 2016, eds. S. Richling, M. Baumann, M. and V. Heuveline, heiBOOKS, http://dx.doi.org/10.11588/heibooks.308.418

[12] Schridde, D., M. Baumann and V. Heuveline. "Skalierbare und flexible Arbeitsumgebungen für Data-Driven Sciences" in E-Science-Tage 2017: Forschungsdaten managen, eds. J. Kratzke and V. Heuveline, heiBOOKS, http://dx.doi.org/10.11588/heibooks.285.377

Virtualized Research Environments on the bwForCluster NEMO

Michael Janczyk, Bernd Wiebelt, and Dirk von Suchodoletz
eScience Department, Computer Center, University of Freiburg, Freiburg, Germany
{michael.janczyk, bernd.wiebelt, dirk.von.suchodoletz}@rz.uni-freiburg.de

Abstract—**The bwForCluster NEMO offers high performance computing resources to three quite different scientific communities (Elementary Particle Physics, Neuroscience and Microsystems Engineering) encompassing more than 200 individual researchers. To provide a broad range of software packages and deal with the individual requirements, the NEMO operators seek novel approaches to cluster operation [1]. Virtualized Research Environments (VREs) can help to both separate different software environments as well as the responsibilities for maintaining the software stack. Research groups become more independent from the base software environment defined by the cluster operators. Operating VREs brings advantages like scientific reproducibility, but may introduce caveats like lost cycles or the need for layered job scheduling. VREs might open advanced possibilities as e.g. job migration or checkpointing.**

Index Terms—**Virtualized Research Environments, VRE, HPC, Virtualization, ViCE, bwHPC, bwHPC-C5**

I. VIRTUALIZED RESEARCH ENVIRONMENTS

The author of [3] defines Virtual Research Environments to "[...] comprise digital infrastructure and services which enable research to take place. [...] A VRE is best viewed as a framework into which tools, services and resources can be plugged. VREs are part of infrastructure, albeit in digital form, rather than a free-standing product." While this more than a decade old paper emphasizes the general shared electronic infrastructure aspect, the term Virtualized Research Environment shifts the focus more to the technological viewpoint. Virtualization became a mainstream technology over the last decade as it allows the operators to host more than one operating system on a single server and to strictly separate users of software environments. While widespread in computer center operation, virtualization is still seen skeptically in the field of scientific computing and thus seldom applied in HPC.

The filesystem of a VRE is a container presented as a single file (or a couple of files linked together to present the container). From the operator's perspective this container is seen as a black box requiring no involvement or efforts like updates of the operating system or the provisioning of software packages of a certain version. From the researcher's perspective the VRE is an individual (virtual) node where everything from the hardware level – at least to a certain degree like CPU or RAM – up to the operating system, applications and configurations can be controlled solely by the research groups. Individual research groups become independent of the software and update strategy of the cluster operator. The caveat is that the old HPC performance mantra "every cycle counts"

has to be sacrificed to a certain extent [5]. However, the advantages of this approach makes HPC attractive for broader scientific communities in the first place.

A. Reproducible Science

Digital research data management requires the conservation and publication of raw data as well as computed results. However, this is only half the story. It is equally important to preserve the computational tool chain used in the acquisition of new scientific insights. In the classic HPC approach, digital research environments are tightly coupled to the hardware and base operating system and therefore no longer usable when the HPC system is decommissioned. Instead, a lot of effort has to be put into reestablishing similar research environments on the successor system. By applying well-known techniques used in Cloud computing to HPC, deploying digital software environments to new HPC systems becomes a straightforward, simple and reproducible procedure. Even better, these *virtualized* digital research environments can be conserved in the state they were in when the actual scientific research was conducted. Establishing such certified research environments (CREs) – tested, certified and immutable software stacks – could be the next logical step. A CRE could thus help to complement electronic lab books.

B. ViCE Project on VREs

The eResearch initiative on Virtualized Research Environments (VREs) sponsored by the state of Baden-Württemberg provided the perfect framework to bring together both the infrastructure providers like the computer center of the University of Freiburg and research communities (e.g. Elementary Particle Physics from Karlsruhe and Freiburg) in the ViCE project. The research groups have various demands for computing in scientific workflows which are tightly coupled to well defined software environments. The project helped the involved parties to understand each others requirements to utilize virtualization and containerization technologies and to separate the software-hardware stack to foster more independence and flexibility for both of the parties. In daily operations, VREs help to disentangle the responsibilities regarding the digital research environment required for a certain task. The computer center focuses on the provisioning of a scalable hardware base including all necessary components to run virtualization and containerization technologies like OpenStack or Singularity.

II. Visions and Goals

In the classic HPC approach, research environments are meticulously handcrafted to squeeze out every last cycle count of the compute resources. How this is achieved is often hidden deep inside software module definitions created by software administrators or hidden inside source code modifications done by the user. It is easy to lose track of all these changes, especially in case there has been no agreed upon standard procedure put into place to record and conserve them.

When we look at software deployment on individual workstations or notebooks, the situation becomes even more inextricable. Hardware and software installation merge and converge into a monolithic masterpiece, fully understood solely by its creator – if at all. In any case, once the creator loses interest, his creation is doomed.

In contrast, a successful strategy in Cloud computing has been the orchestration of software environment deployment. The scientists create or modify recipes to generate Virtualized Research Environments tailored to their scientific workflows. These can be shared and improved upon by co-workers and collaborators. To this end, software packages (source or binary) are requested from well-defined and permanent repositories. VRE images (templates) get built on dedicated VRE build hosts. These VRE build hosts can be run as virtual machines on the user's workstation or in a Cloud service. After that, VRE instances can be deployed in the needed number on the Cloud system. This workflow is demonstrated in Fig. 1.

Ideally, these VRE instances should be agnostic of the environment they are started in. So they should be runnable in either Cloud or HPC environments. In the case of HPC, VRE instances should be considered as "just another job".
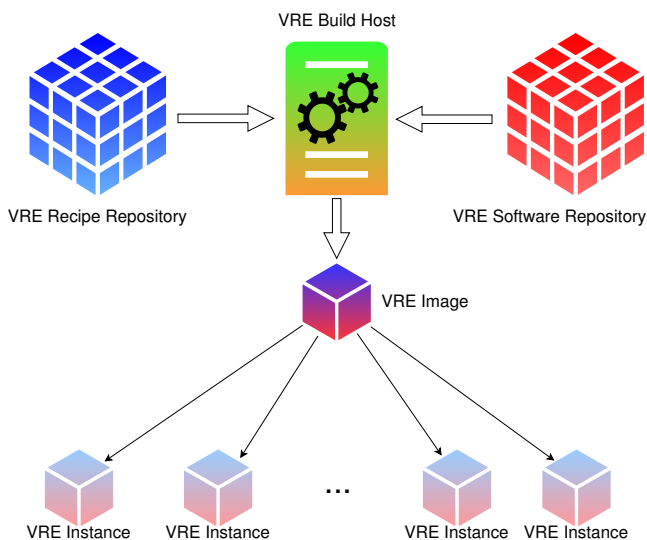

Fig. 1. Recipe and software repositories for automatic building of reproducible VREs.

III. Running VRE Instances on NEMO

On a HPC cluster users don't start their computations on the worker nodes themselves. This is done by the scheduler, who decides which job to put on which worker node. It can employ different strategies to do so and can take priorities and different fairness strategies into account. Compute Clouds use scheduling as well. Usually a set of system flavors – pre-configured virtual hardware sets – is given, which describes the resources one can get in the Cloud. In HPC it is custom that users get exactly the resources they asked for. If requirements can not be satisfied instantly, the job will have to wait in a queue. In Clouds the hardware resources usually are overbooked so that queues are processed within minutes and waiting time is minimized to the start of a machine. Commercial Clouds additionally steer their queues by monetary incentives.

Given these two system variants, HPC scheduling on the one side and Cloud scheduling on the other, competing for hardware resources in the cluster it is necessary to orchestrate both systems through a mediator. This mediator is called VRE service. It runs as a service daemon and waits for requests from users and administrators and translates it to the OpenStack API. In this hybrid approach one needs to define a primary scheduler who controls which jobs should run on which worker node and instruct the other scheduler to run within the boundaries of the other scheduler, especially respecting the scheduling decision of the other. The problem is, that neither the HPC scheduler, in our case the combination of the workload manager Moab and the resource manager Torque, nor the Cloud scheduler are aware of each other and no APIs exist on both sides to integrate a different scheduler. Since we provide a hybrid Cloud use case on a HPC cluster and users are familiar with submitting jobs from the console and Moab has the more sophisticated scheduling features compared to OpenStack we define Moab as the default scheduler. The OpenStack scheduler is still running but is told where to start its virtual machines. This way we do not need to integrate the Moab scheduler into OpenStack and that way this should work with every HPC / HTC scheduler on the market.

Starting virtual machines through the OpenStack dashboard is still allowed but only in a special service environment which does not interfere with the worker nodes of the HPC cluster. The VRE service is monitoring all virtual machines started and destroys every machine which is started outside of the service environment when no corresponding Moab job ist found.

Starting Virtualized Research Environments on the worker nodes is done through the Moab scheduler by submitting special tags as variables. The scheduler queues and routes this VRE job as a standard HPC job respecting policies like fairshare, usage limits and the requested resources like memory, CPU and wall-time. Once the job is started on the node the torque client (mother superior) filters it for the special VRE tags and submits this request to the VRE service. The VRE service then translates this information to the corresponding tuple of OpenStack project, VRE name, flavor for resources and binds it to the corresponding node. This information is

then send to the OpenStack service which then starts the VRE (a virtual machine) on the pre-selected worker node. If the Moab job ends or is canceled by the system or user the mother superior sends this information to the VRE service which then destroys the instance. During the lifetime of the job the VRE service is monitoring the virtual machine and reports back utilization when requested by users or services. No direct interaction with the OpenStack service is necessary by users or administrators. The principle can be seen in Fig. 2.
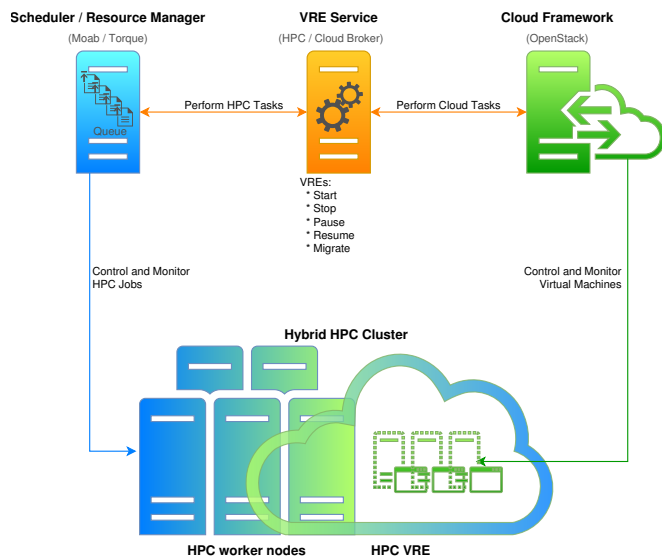

Fig. 2. VRE service acts as broker between the two schedulers.

The helper tool `cloud-init`, which is used to customize virtual machine images on boot, is used to inject scripts to execute within the VRE, if provided. The job description file, which defines the resources needed and the commands to execute during the job's run-time, and usually is used on compute clusters is replaced by a simple wait script to ensure that the HPC job is not ended before the VRE job has finished.

The VRE service is developed in a cooperation between the HPC team in Freiburg and Adaptive Computing, the company behind Moab. The concept is still in development, so future results can differ from this description. The following lines show an example for submitting VREs on NEMO:

```
msub -l nodes=1:ppn=20 \
    -l pmem=6GB \
    -l walltime=1:00:00:00 \
    -l var=vmscript:$script+vmimage: ↩
        $image+vmproject:$project+ ↩
        vmsshkeyname:$key \
    job_dummy.sh
```

IV. BIOINFORMATICS AND HIGH ENERGY PHYSICS VRES ON NEMO

A typical concern for many scientific workflows, especially in (geographically) distributed large projects, is the strict requirement for the software stack involved [1].

Therefore the huge CERN projects like ATLAS and CMS usually are very conservative with changes in their environment and usually test future environments over months or even years. When this paper was written the ATLAS and CMS project still used Scientific Linux 6.x as operating system while the bwForCluster NEMO used the newer operating system CentOS 7.4. The complete software and service stack of the High Energy Physics (HEP) is tailored around this Scientific Linux 6. Driven by the fact that Elementary Particle Physics is one of the research fields on the bwForCluster NEMO (the others being Neuroscience and Microsystems Engineering) a solution was needed to provide the HEP research environments on the cluster. By using VREs ATLAS and CMS [6] users have now the possibility to run their jobs in their standard HEP research environments and it is guaranteed, that the operating system and the software stack is compatible with the HEP environments used on bare-metal clusters.

The Bioinformatics department uses the the Galaxy Portal for their computations. To support this environment on NEMO a VRE including Galaxy has been created and the integration into the NEMO cluster is in development currently.

V. ADVANCED CONCEPTS

Using virtualization inside a HPC system opens up possibilities for several interesting features. While their implementation requires tighter integration between HPC scheduler and virtualization framework as shown in Fig. 2, they could solve several classic problems with HPC systems, especially those designated for novice HPC users.

Using the virtualization features "snapshot and migration" for a virtualized cluster like e.g. described in [4], [5], would enable operators and users to stop running processes and move them to a different node in the cluster where they could be resumed. For an HPC system, this would be practical for two use cases. The first one concerns long running monolithic jobs. These are, for very practical reasons, non-favored jobs in HPC environments, assuming they are permitted in the first place. However, the costs to adapt a particular workflow based on such monolithic tasks to a HPC system, e.g. by parallelizing and partitioning it manually, may sometimes exceed the practical use of the resulting solution. If the monolithic job could automatically be stopped, check-pointed and resumed at regular intervals, this might very well constitute a more economic procedure. In the second use case, if there is a mix of pleasingly parallel high throughput jobs (using only single cores or nodes) and massively parallel high performance jobs (using several nodes), the second class of jobs should be concentrated on nodes that share optimal high performance network communication paths. Typically this is accomplished by high investments in the network topology or sophisticated tuning of the job queue. However, if jobs could be moved around the physical machines (i.e. "de-fragmented"), optimal high performance network communication paths can be guaranteed by concentrating massively parallel jobs on the same or adjacent high performance network switches.

VI. Outlook

There still exist a couple of limitations on the presented concepts and workflows. The use cases considered feature embarrassingly parallelized workloads which do not require interaction between individual cluster nodes. This simplifies the setup and operation of virtual machines as the high bandwidth node interconnect could be ignored in this scenario. Singularity might overcome that limitation in the future but would make the application of the advanced concepts more complex as explained in [4]. A further challenge arises from the provisioning of the data in the form of parallelized high performance storage. If users become root in their VREs than the traditional means of privilege separation will not work any more. A similar issue arises from the heightened mobility of VREs. While in static cluster configurations IP based security with its limitations made sense, this becomes more challenging when VREs get involved. Depending on the way virtualization or containerization is implemented, the HPC scheduling setup has to be aware of the more dynamic nature of resources.

Pre- and post-processing in many workflows often profit from interactive intervention instead of batch-driven automatic processing. Here, VREs could help to use the same working environment for Cloud (pre- and post-processing) and HPC systems (main computational task). Beside the organizational and administrative benefits VREs might allow to gain more flexibility through e.g. job checkpoint-restart or job migration. Because of the containerization of VREs these could get extended to Certified Research Environments in the future.

References

[1] K. Meier, G. Fleig, T. Hauth, M. Janczyk, G. Quast, D. von Suchodoletz, and B. Wiebelt, "Dynamic provisioning of a HEP computing infrastructure on a shared hybrid HPC system," 17th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2016), Valparaiso Chile, Proceedings, p. 18–22, January 2016

[2] D. von Suchodoletz, B. Wiebelt, K. Meier, and M. Janczyk, "Flexible HPC: bwForCluster NEMO," Proceedings of the 3rd bwHPC-Symposium: Heidelberg 2016, editors: S. Richling, M. Baumann, and V. Heuveline, doi: 10.11588/heibooks.308.418, heiBOOKS, 2017

[3] M. Fraser, "Virtual research environments: overview and activity," num. 44, url: http://www.ariadne.ac.uk/issue44/fraser, Ariadne, 2005

[4] P. Anedda, S. Leo, S. Manca, M. Gaggero, and G. Zanetti, "Suspending, migrating and resuming HPC virtual clusters," Future Generation Computer Systems, vol. 26, num. 8, p. 1063–1072, Elsevier, 2010

[5] A. Younge, R. Henschel, J. Brown, G. von Laszewski, J. Qiu, and G. Fox, "Analysis of virtualization technologies for high performance computing environments," 2011 IEEE International Conference on Cloud Computing (CLOUD), p. 9–16, IEEE, 2011

[6] T. Hauth, G. Quast, M. Kunze, V. Büge, A. Scheurer, and C. Baun, "Dynamic extensions of batch systems with cloud resources," Journal of Physics: Conference Series, vol. 331, num. 6, p. 062034, IOP Publishing, 2011

Numerical Simulation of Heat Transfer

in Turbulent Pipe Flow with

Structured Wall Surfaces

Peter Renze

Institute of Energy and Drive Technology
University of Applied Sciences Ulm
Germany
renze@hs-ulm.de

Abstract— **Innovative heat transfer technology is the key to the optimization of many processes in the power or process industry. Operational costs and usage of valuable resources can be reduced, if the heat transfer efficiency is increased and pressure loss is reduced. Therefore, the current work is focused on heat transfer enhancement at tubes with micro-structured walls with turbulent flow. In this kind of geometry modern optical measurement technology cannot be applied and the analysis of the turbulent transport is only possible with numerical flow simulations. A large-eddy turbulence model is applied to account for turbulence closures. First, the simulation setup is validated with data from the literature and then several micro-structured geometries are investigated. The simulations are computationally costly and depend on high performance computing (HPC). The open-source software library OpenFOAM® is applied to perform massively parallel simulations.**

Keywords—computational fluid dynamics; heat transfer enhancement; turbulence; micro-structured tubes

I. INTRODUCTION

Modern high performance heat exchangers are of great importance in many industrial applications like power generation, cooling, refrigeration, chemical processes and other production processes. In these industries the necessity to increase production at lower consumption of resources is a demanding task. Optimizing heat transfer equipment is a crucial part in the overall process optimization. Therefore, innovative heat transfer technologies are required.

The term *heat transfer enhancement* is used in the literature to describe the improvement of the heat transfer coefficient k. Thus, the heat transfer is increased or the heat transfer area can be reduced while keeping the heat transfer rate constant. Either way, the efficiency of the process is increased. This is a classical challenge for the engineering and many researches have published on this topic, see [1, 2].

In the scope of this paper the heat transfer enhancement at tubes with micro-structured wall surfaces are investigated at turbulent flow conditions. Numerical flow and heat transfer simulations are performed using the open-source software library OpenFOAM® in massively parallel simulations.

II. METHODS

A. OpenFOAM

OpenFOAM® (**O**pen **S**ource **F**ield **O**peration and **M**anipulation) is an object-oriented software library programmed in C++ and designed for the numerical solution of differential equation from continuum mechanics, see Jasak and Weller [3]. The main usage is in the area of computational fluid mechanics. OpenFOAM® was declared open-source in 2004 and is distributed free of charge. It is licensed under the GPL, which makes it ideal for simulations in the area of academic as well as industrial research.

OpenFOAM® is distributed with a variety of pre-designed solvers. Usually the finite volume method (FVM) is applied, but there are applications of the finite-element method (FEM) available as well. New solvers and algorithms can be easily generated by adaption and combination of available physical model implementations. These are structured in the form of object oriented classes, e.g. within the turbulence class several approaches are available like Reynolds-averaged-Navier-Stokes (RANS), large-eddy simulation (LES), or direct numerical simulation (DNS) as well as hybrid approaches.

The top-level structure of OpenFOAM® makes it easy for engineering students or unexperienced programmers to begin the design of their own code. The strictly object-oriented programming of OpenFOAM® puts a strong focus on objects instead of functions. The pre-processing is done classically via the command line terminal, but in the online community a large

number of scripts or graphical user interfaces (GUI) are available today. The computational meshes can be generated within the software, they are unstructured and based on hex-dominant cells. For graphical post-processing of the simulation results integrated modules in ParaView [4] are provided.

B. Physical Modeling

The physical modeling of flows with heat transfer is based on conservation laws for mass, momentum, and energy. The continuity equation for an unsteady flow of a compressible fluid can be formulated as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}), \tag{1}$$

with the nabla operator ∇, the density ρ, and the vector of velocity $\mathbf{u}$. Thus, the conservation of momentum is formulated

$$\frac{\partial (\rho \mathbf{u})}{\partial t} + \nabla \cdot (\rho \mathbf{u}\mathbf{u}) = -\nabla p + \rho g + \nabla \cdot \left(2\mu_{eff}\mathbf{S}(\mathbf{u})\right) - \nabla \left(\frac{2}{3}\mu_{eff}(\nabla \cdot \mathbf{u})\right), \tag{2}$$

where p is the static pressure, $\mathbf{g}$ the constant of gravity, and μ_{eff} is the sum of the molecular and turbulent dynamic viscosity. The deformation tensor $\mathbf{S}(\mathbf{u})$ is defined as $\mathbf{S}(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + (\nabla \cdot \mathbf{u})^T)$.

The conservation of energy is formulated using the variable of inner energy e

$$\frac{\partial (\rho e)}{\partial t} + \nabla \cdot (\rho \mathbf{u} e) + \frac{\partial (\rho k)}{\partial t} + \nabla \cdot (\rho \mathbf{u} k) + \nabla \cdot (p \mathbf{u}) = \nabla \cdot \left(\alpha_{eff}\nabla e\right) + \rho \mathbf{u} \cdot g, \tag{3}$$

where $k = |\mathbf{u}|^2/2$ is the kinetic energy per mass unit. The effective thermal diffusivity α_{eff} is defined as

$$\alpha_{eff} = \frac{\rho v_t}{Pr_t} + \frac{\mu}{Pr} = \frac{\rho v_t}{Pr_t} + \frac{\lambda}{c_p} \tag{4}$$

where λ is the thermal conductivity, c_p the heat capacity at constant pressure, Pr_t is the turbulent Prandtl number and v_t the turbulent kinematic viscosity. The turbulent transport parameters in the equations (1) to (4) require a turbulence closure with a corresponding modeling technique.

In the literature, an isotropic concept for the turbulent Prandtl-number is pre-dominantly found, but recently tensor-based models are applied more often. A thorough discussion is not possible here, but recent improvement has been made in RANS-development like the v2-f-model [5, 6] or the scale-adaptive model [7]. In recent years the effort for the development of RANS-models has been strongly reduced due to lack of general success as well as the rapidly increasing computational resources that allow the application of more costly methods.

Here, the LES technique is applied, i.e. only the large-scale vortices that contain most of the turbulent kinetic energy are directly resolved and all smaller scale vortices are modeled [8, 9, 10]. Like this, the accuracy compared to RANS is increased substantially and at the same time the numerical effort compared to DNS is reduced up to a level that makes this approach interesting for the industry. A detailed overview about the LES techniques is given in [11]. These techniques have been demonstrated to perform well for technically relevant turbulent flows [12, 13, 14] and even for multiphase flows [15, 16, 17].

In LES a filter function with a characteristic filter width Δ is applied to equation (1)-(3), so that the field variables are decomposed into a resolved and a non-resolved part. Thus, the velocity vector becomes $\mathbf{u} = \hat{\mathbf{u}} + \mathbf{u}'$. The expression for the effective viscosity is reformulated to $\mu_{eff} = \mu + \mu_{SGS}$, i.e. the sum of molecular and apparent viscosity of the subgrid scales. Following the Boussinesq approximation the subgrid scale viscosity is then modeled as $\mu_{SGS} = \rho C_k \Delta \sqrt{k_{SGS}}$ with $C_k = 0.07$ and k_{SGS} as kinetic energy of the subgrid scales. The Smagorinsky model [8] is chosen with

$$k_{SGS} = \frac{C_k}{C_\varepsilon}\Delta^2 |\mathbf{S}|^2 \tag{5}$$

with $C_\varepsilon = 1.048$ and $|\mathbf{S}| = (\mathbf{S}:\mathbf{S})^{\frac{1}{2}}$.

In the current work an additional transport equation model for the kinetic energy is solved for some cases and the increased numerical effort will be juxtaposed to the gained accuracy. The equation reads

$$\frac{\partial k_{SGS}}{\partial t} + \nabla \cdot (k_{SGS}\mathbf{u}) - \nabla \cdot \left(\frac{\mu_{eff}}{\rho}\nabla k_{SGS}\right) = 2v_{SGS}|\mathbf{S}|^2 - \frac{C_\varepsilon k_{SGS}^{\frac{3}{2}}}{\Delta}. \tag{6}$$

TABLE I.   INFORMATION ABOUT HPC RUNS

HPC Details	Simulation Cases		
	Smooth Tubes (outer forced convection) Re_D=6900, Pr=7	*Smooth tubes (inside flow) Re_τ=180-314, Pr=7*	*Structured Tubes (inside Flow) Re_τ=180, Pr=7*
Execution time	96	144	192
Simulated time	10 flow cycles	10 flow cycles	10 flow cycles
Cores/Nodes	56 / 2	56 / 2	56/ 2
Number of time steps	~100000	~110000	~130000
Cell number	~8 millions	~8 millions	~26 millions

42

III. COMPUTATIONAL DETAILS

Most of the simulations have been performed within the scope of the bwHPC-C5 project using the bwUniCluster. Selected costly computations have been performed at the forHLR II cluster, where access has been granted to the author.

More detailed information about the HPC runs are given in Table I.

IV. RESULTS

In this chapter the preliminary results of an ongoing project are presented. Before the heat transfer enhancement at tubes is investigated, the simulation setup as well as workflow are validated and the results are compared with literature.

A. Simulation of heat transfer at the turbulent flow around tubes

In Fig. 1 a visualization of the instantaneous flow and temperature field at forced convection around single tubes is given. The Reynolds number is calculated with the outer tube diameter and with the outer flow velocity $\mathbf{u}_D$. It yields $Re_D = \frac{\rho \mathbf{u}_D D}{\mu} = 6900$.
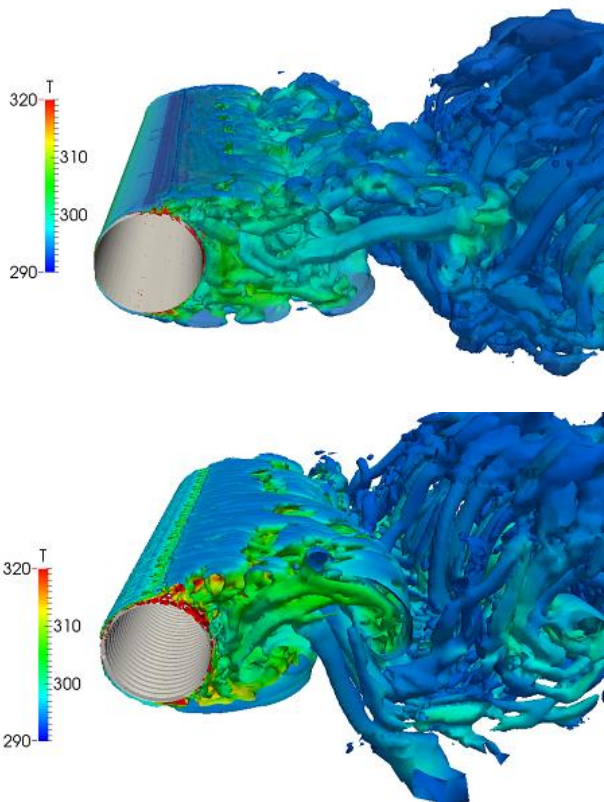

Fig. 1. Results of a large-eddy simulation (LES) of turbulent flow around tubes at $Re_D = 6900$ with a smooth heat transfer surface (top) and micro-structures (bottom).

Here, a constant wall temperature is applied as a boundary condition and the heat transfer to the fluid is calculated. In Fig. 1 (top) the coherent flow structures in this turbulent flow with periodic separation are emphasized using the Lambda2 method.

In Fig. 1 (bottom) the results of the simulation with identical setup but a different wall structure are presented. Here, the outer tube surface is micro-structured with fins. The separation phenomena are clearly effected as well as the heat transfer. The time averaged heat transfer is enhanced by around 20%. The variation of the heat transfer with time is about 10%

B. Simulation of heat transfer at turbulent pipe flow

The accurate calculation of heat transfer at the turbulent flow inside of tubes with a large-eddy simulation technique requires a far more complex numerical setup compared to the work in the previous section. The main challenge is the generation of turbulent inflow data at the inlet section of the pipe. Usually, the computational costs do not allow for the simulation of long pipe sections with a fully developed flow. That is why, only a small section is simulated, i.e. about 12 inner tube diameters.
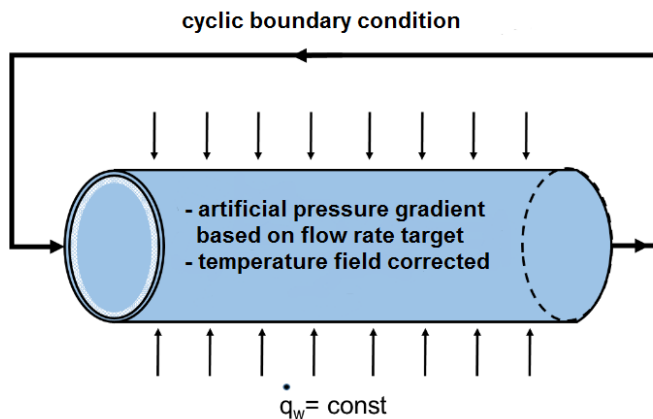

Fig. 2: Sketch of the numerical setup and the boundary condition for the LES of the turbulent flow through tubes.

Thus, at the flow inlet of the domain unsteady flow data has to be generated artificially. Assuming a fully developed turbulent flow field this can be done by applying cyclic boundary conditions between inlet and outlet. The pressure gradient that drives the flow is artificially introduced to the momentum equations. The numerical setup is shown in Fig. 2. The heat flux through the tube wall is usually defined by either a constant temperature or a constant heat transfer boundary condition at the wall surface depending on the validation case. In both cases the temperature data has to be corrected by the introduced heat flux if a cyclic boundary between inlet and outlet is used.

The computational meshes for all cases are generated with the grid generation tool *snappyHexMesh* distributed with OpenFOAM®. A sufficient wall resolution of around $y_+ = 0.1$ is reached by the application of prism layers near the wall. The unstructured meshes are predominantly cubical without stretching which results in a large number of total grid cells compared to literature cases. For the low Reynolds number cases this number is about 6 million. For most literature cases structured meshes with high grid stretching reduces the size of the mesh by a factor of up to 10. But with the current automated mesh generation setup complex wall surface structures can be meshed, which is a great advantage. Like this exactly the same setup can be used to simulate varying surface geometries eliminating the influence of mesh topology substantially.

In Fig. 3 snapshots of the 3D flow field of the simulations with a smooth tube surface are shown at an arbitrary time step. The coherent vertical structures (top) in the flow field and the instantaneous thermal boundary layer (bottom) are clearly visible.
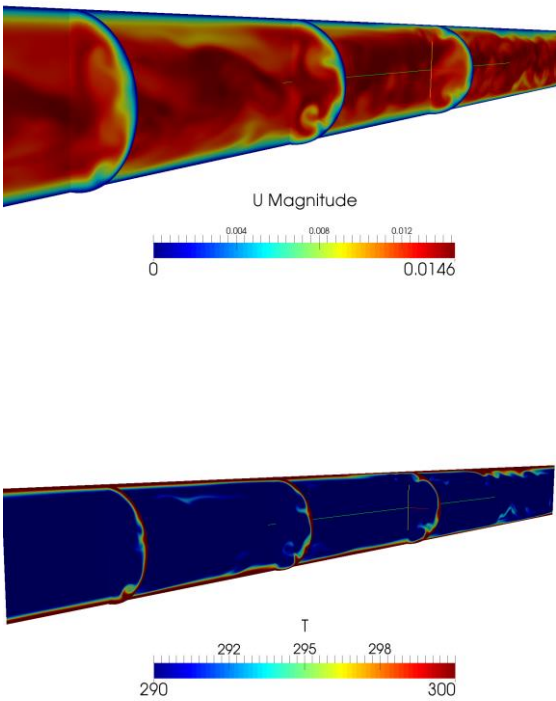

Fig. 3. Contours of the flow and temperature field at an instantaneous time step at a smooth tube surface, $Re_\tau = 180$, $Pr = 7$; top: magnitude of velocity; bottom: temperature.

Time averaged results of the simulations are shown in Fig. 4. The flow field has been averaged over 20 flow cycles. The velocity profile in radial direction that is shown in Fig. 4 has been spatially averaged as well as time averaged. The velocity profile is shown in the classical logarithmic scale for two different Reynolds numbers, $Re_\tau = 180$ and $Re_\tau = 314$. The results match literature data very well considering the aforementioned design of the computational mesh.

An analysis of the heat transfer process is given in Table II. For the case of $Pr = 7$ and two Reynolds numbers the resulting Nusselt numbers are calculated from the simulations and compared with classical engineering correlations for heat transfer in turbulent pipe flow, see Gnielinski [18]. The results match very well and the simulation workflow proves to be capable of predicting turbulent heat transfer inside of tubes with reasonable accuracy.
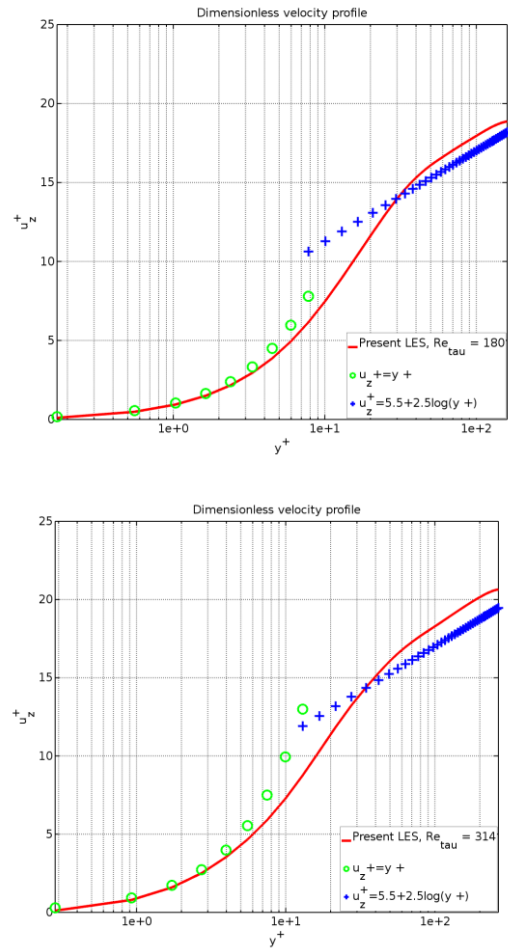

Fig 4. $Re_\tau = 180$, Pr = 7 (top), $Re_\tau = 314$, Pr = 7 (bottom)

TABLE II: Comparison of the simulated heat transfer (Nusselt number Nu) with literature data from Gnielinski [18]

Case	Nu$_{simulated}$	Nu$_{literature}$
$Re_\tau = 180$, $Pr = 7$	45.2	48.8
$Re_\tau = 314$, $Pr = 7$	86.9	81.7

In Fig. 5 preliminary results are shown for a simulation of turbulent flow inside of tubes with micro-structured surfaces at $Re_\tau = 180$. The computational mesh has been generated with exactly the same setup as described above. The increased complexity of the near wall mesh leads to a total number of 30 million grid cells. The transport processes will be analyzed in the ongoing work.
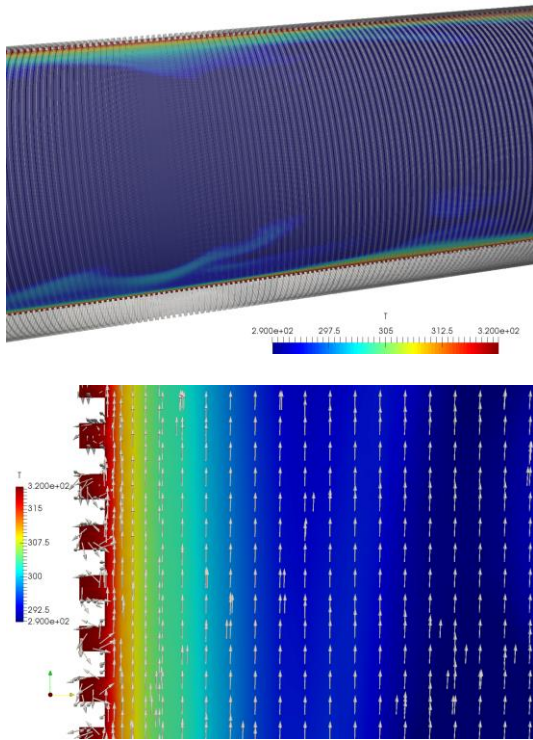

Fig. 5. Preliminary results of turbulent heat transfer through tubes with micro-structured surfaces at $Re_\tau = 180, Pr = 7$

V. Summary

The software library OpenFOAM® is increasingly used for computational fluid dynamics in academic and industrial research. Here, it is applied to calculate the heat transfer enhancement by micro-structured surfaces in turbulent flows. The numerical setup has been validated by comparison of the results with literature data. Especially, the simulation of turbulent heat transfer inside of tubes is an intricate task. The current methods based on OpenFOAM® proved to be capable for this challenge. Preliminary results of simulations with complex surfaces have been shown. The physics of heat transfer enhancement will be thoroughly analyzed with the help of these methods in the scope of the ongoing work. So far it can be concluded, that the method can be readily applied to develop new surface structures for heat transfer enhancement. The numerical costs are not prohibitive and a later transfer of the method into the industry seems feasible.

Nomenclature	
ρ	density
$\boldsymbol{u}$	velocity vector
μ	dynamic viscosity
k	turbulent kinetic energy
$\boldsymbol{S}$	shear rate tensor
g	gravity constant
p	static pressure
α	heat transfer coefficient
λ	thermal conductivity
e	specific energy
Re_τ	Reynolds number based on friction velocity
Pr	Prandtl number or turbulent Prandtl number with subscript t

References

[1] A. E. Bergles, M. K. Jensen, B. Shome, Bibliography on Enhancement of Convective Heat and Mass Transfer. Heat Transfer Laboratory Report No HTL-23 Rensselaer Polytecnic Institute, 1995.

[2] R. L. Webb, N. H. Kim, Principle of enhanced heat transfer. 2nd Edition.Taylor Francis: New York, NY, USA, 2005.

[3] H. Jasak H., H.G. Weller, Interface tracking capabilities of the InterGamma differencing scheme, Technical Report, Imperial College, University of London, 1995.

[4] J. Ahrens, B. Geveci, C. Law, ParaView: An End-User Tool for Large Data Visualization, Visualization Handbook, Elsevier, 2005, ISBN-13: 978-0123875822.

[5] Laurence, D.R., Uribe J.C., Utyuzhnikov, S.V. A Robust Formulation of the v2-f Model, Flow Turbulence and Combustion, 73, 169-185, 2004.

[6] Popovac, M., Hanjalic, K., Compound Wall Treatment for RANS Computations of Complex Turbulent Flows ans Heat Transfer, Flow Turbulence and Combustion, Nr. 78, S. 177-202, 2007.

[7] Menter, F. R., and Y. Egorov. The scale-adaptive simulation method for unsteady turbulent flow predictions. Part 1: theory and model description. Flow, Turbulence and Combustion 85.1 (2010): 113-138.

[8] J. Smagorinsky. General circulation experiments with the primitive equations, part I: The basic experiment. Monthly Weather Review, 91:99–164, 1963.

[9] Germano, M., Piomelli, U., Moin, P. and Cabot, W. H. (1991), A dynamic sub-grid scale eddy viscosity model, Physics of Fluids, A(3): pp 1760-1765, 1991.

[10] You, D. and Moin, P. (2007), A dynamic global-coefficient subgrid-scale eddy-viscosity model for large-eddy simulation in complex geometries, Physics of Fluids, 19(6): 065110, 2007.

[11] Sagaut, Pierre (2006). Large Eddy Simulation for Incompressible Flows (Third ed.). Springer. ISBN 3-540-26344-6.

[12] P. Renze, W. Schröder, M. Meinke, Large-eddy simulation of film cooling flows at density gradients, International Journal of Heat and Fluid Flow 29.1; 18-34, 2009.

[13] P. Renze, W. Schröder, M. Meinke, Large-eddy simulation of film cooling flows with variable density jets, Flow, Turbulence and Combustion 80.1 (2008): 119-132, 2008.

[14] P. Renze, W. Schröder, M. Meinke, Large-eddy simulation of film cooling flow ejected in a shallow cavity, AIAA Paper 927, 2007.

[15] P. Renze, K. Heinen, M. Schönherr. Experimental and Numerical Investigation of Pressure Swirl Atomizers. Chem Eng Tech, 34 (7):1191–1198, 2011.

[16] P. Renze, A. Buffo, D. L. Marchisio, M. Vanni, Simulation of Coalescence, Breakup, and Mass Transfer in Polydisperse Multiphase Flows, Chemie Ingenieur Technik, 86(7):1088–1098, 2014.

[17] A. Buffo, D. L. Marchisio, M Vanni, P. Renze, Simulation of polydisperse multiphase systems using population balances and example application to bubbly flows, Chem. Eng. Res. Des., 91(10):1859–1875, 2013.

[18] Verein Deutscher Ingenieure, VDI-Gesellschaft Verfahrenstechnik und Chemieingenieurwesen (GVC). VDI-Wärmeatlas: Berechnungsblätter für den Wärmeübergang. 11. Auflage, 2013.

[19] Steinbuch Center for Computing. Forschungshochleistungsrechner ForHLR 2. http://www.scc.kit.edu/dienste/forhlr2.php,

Planets in turbulent Disks

Wilhelm Kley
Institute for Astronomy and Astrophysics
Eberhard-Karls-Universität Tübingen
Tübingen, Germany
wilhelm.kley@uni-tuebingen.de

Moritz H.R. Stoll
Institute for Astronomy and Astrophysics
Eberhard-Karls-Universität Tübingen
Tübingen, Germany
moritz.stoll@uni-tuebingen.de

Giovanni Picogna
Universitätssternwarte
University of Munich
München, Germany
picogna@usm.lmu.de

Abstract—**Planets form in protoplanetary accretion disks around young protostars. These disks are driven by internal turbulence and the gas flow is in general not laminar but has stochastic components. For weakly ionized disks the turbulence can be generated purely hydrodynamically through the vertical shear instability (VSI). Planets, that are forming in such turbulent disks, experience random components in the torques acting on them, that will modify their migration behaviour.**

In this contribution we present results of our study of the impact that a turbulent disk has on an embedded planet. For that purpose we performed three-dimensional hydrodynamical simulations of a locally isothermal disk with embedded planets of various masses. For the lower mass planets (5 and 10 $M_\oplus$) we find that the inward migration rate can be faster than in the laminar case, while for the more massive planets (30 and 100 $M_\oplus$) the results agree well with the laminar results.

Index Terms—planet formation, accretion disks, computational astrophysics

I. INTRODUCTION

To simulate planets embedded in turbulent accretion disks our computational domain consists of a 3D section of a disk in spherical polar coordinates (r, θ, ϕ). We will use also cylindrical coordinates (R, Z, ϕ) to describe our setup and the results. We perform hydrodynamical simulations using a locally isothermal equation of state where the temperature is a fixed function of radius

$$T(R) = T_0 \left(\frac{R}{R_\mathrm{p}} \right)^q ,$$

and vertically isothermal. Here we use $q = -1$ which implies a constant aspect ratio H/R of the disk, where H denotes the vertical pressure scale height of the disk. The radial temperature gradient and the action of gravity by the star results in the following hydrostatic disk configuration for the density

$$\rho(R, Z) = \rho_0 \left(\frac{R}{R_\mathrm{p}} \right)^p \exp \left[\frac{GM_*}{c_\mathrm{s}^2} \left(\frac{1}{r} - \frac{1}{R} \right) \right]$$

and angular velocity

$$\Omega(R, Z) = \Omega_\mathrm{K} \left[(p+q) \left(\frac{H}{R} \right)^2 + (1+q) - \frac{qR}{\sqrt{R^2 + Z^2}} \right]^{\frac{1}{2}}$$

In these equations, $M_* = 1 M_\odot$, c_s = sound speed, R_p = planet position, $p = -1.5$, Ω_K is the Keplerian angular velocity, and ρ_0 the density at the reference radius, $R_\mathrm{p} = 5.2\,\mathrm{AU}$.
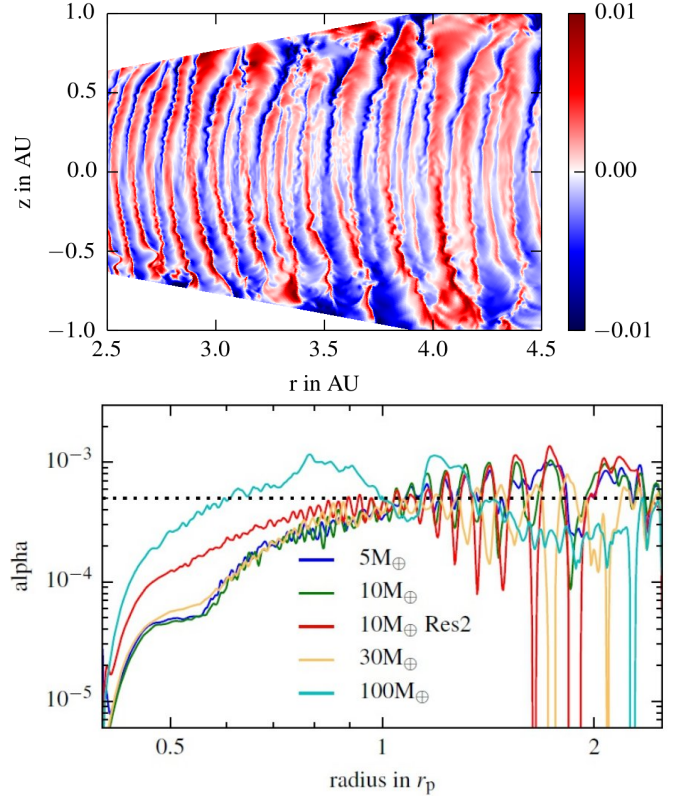

Figure 1. The structure of a VSI-turbulent disk. **Top**: Vertical velocity v_θ (at the $\phi = 0$ slice), red positive, blue negative. In equilibrium state, after a few 100 orbits. Adapted from [7]. **Bottom**: Radial α value for disks for different planet masses. The model labeled 'Res2' has twice the resolution. Adapted from [9].

In this setup we embed planets of different masses and evolve the system in time until a quasi-stationary equilibrium is reached where the turbulence in the disk, driven by the VSI, has saturated. Then we measure the turbulent stresses, quantify the effectiveness of the turbulence, and obtain an effective value of $\alpha = 5 \cdot 10^{-4}$. Using this value for α in a disk setup with explicit viscosity we run comparison simulations to estimate the effect of the turbulence on the embedded planet. The planet is held on a fixed orbit and the torques are measured. The disk section that is modeled extends in radius from $0.4 R_\mathrm{p}$ to $2.5 R_\mathrm{p}$, where $R_\mathrm{p} = 5.2\,\mathrm{AU}$ is the orbital distance of the planet to the star. Vertically
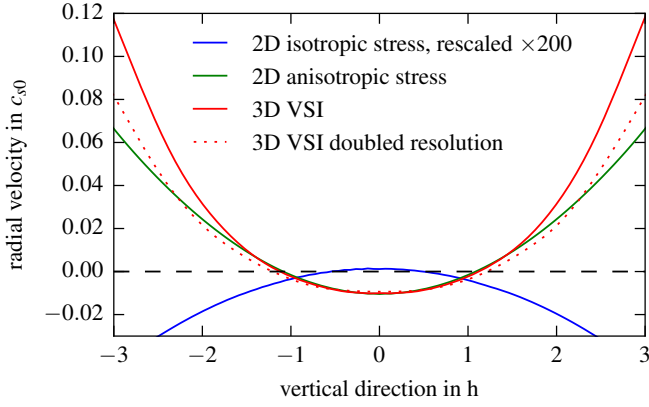
Figure 2. Radial velocity profile, averaged over 50 orbits. Shown are two VSI turbulent disks in 3D simulations in different resolutions, for which the velocity has been azimuthally averaged. The 2D anisotropic model has been calculated for cylindrical coordinates in axisymmetry. The viscous case with isotropic viscosity (standard α-disk) has been rescaled to better visualize the difference. Adapted from [8].

the disk section covers 5 pressure scale heights, H, above and below the equator, and a full annulus $[0, 2\pi]$ in the azimuthal direction. This computational domain is covered with $N_r = 600, N_\theta = 128$, and $N_\phi = 1024$ gridcells that are spaced logarithmically in the radial and equidistant in the other two directions. We ran a comparison model where we doubled the spatial resolution, to check convergence of the results on α.

II. THE TURBULENT DISK

The structure of the VSI-turbulent disk with locally isothermal background is displayed in Fig. 1. The vertical velocity is shown in the top panel in the central (radial) part of the disk. The turbulent elements stretch over the complete vertical direction and show signs of Kelvin-Helmholtz instabilities between upward and downward motions. The effective α-value is plotted in the lower panel where the radial distribution, $\alpha(r)$, is plotted for various planet masses. To estimate the α-parameter as a function of the radius we first calculate the Reynolds stress in cylindrical coordinates (R, Z, ϕ)

$$R_{R,\phi} = \rho u_R \delta u_\phi \,, \tag{1}$$

where u_R is the radial velocity and δu_ϕ is the deviation of the azimuthal velocity from its mean. The α-value is obtained by

$$\alpha(r) = \frac{\langle R_{R,\phi} \rangle_{t,\theta,\phi}}{\langle P \rangle_{t,\theta,\phi}} \,, \tag{2}$$

where P is the pressure and $\langle \rangle_{t,\theta,\phi}$ denotes the average over time, here 140 orbits of the planet, and a spatial average over the whole vertical and azimuthal domain.

The average value found in our simulations lies at $\alpha = 5 \cdot 10^{-4}$ (bottom panel in Fig. 1), relatively independent of the planet mass, where the planet is located at $r = 1$. In the inner parts of the disk the effective α is lower, partly because the inner regions are damped for stability reasons with an
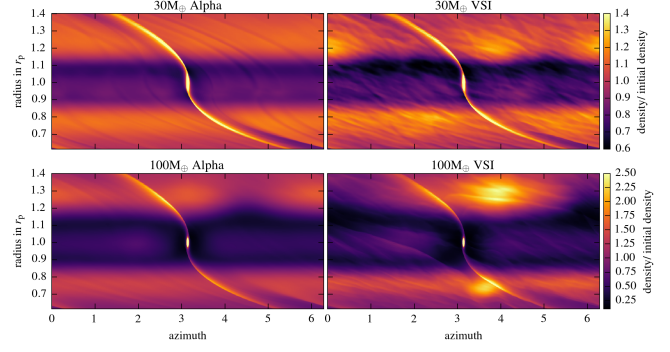

Figure 3. 2D surface density distribution for disks with planets of 30 and 100 $M_\oplus$. **Right**: VSI turbulent disk. **Left**: corresponding viscous disk with $\alpha = 5 \cdot 10^{-4}$. Adapted from [9].
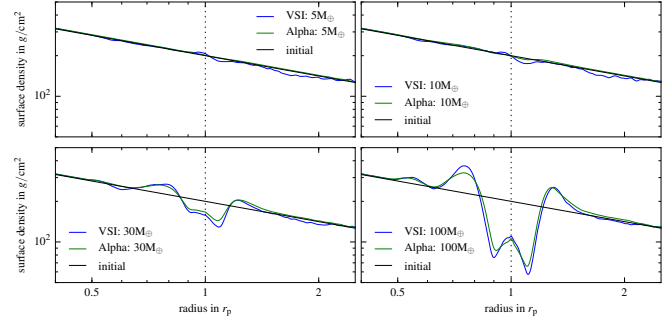

Figure 4. Radial surface density profile for the 4 different planets after 200 orbits. Shown are the profiles for the VSI disk (in blue), the corresponding α-disk (in green) and the initial profile (in black). Adapted from [9].

explicitly added viscosity. The model with twice the spatial resolution yields the same α-value.

In Fig. 2 we display the mean radial velocity profile for the turbulent and laminar disk as a function of vertical distance. Surprisingly, the radial flow within the disk is directed inwards near the midplane of the disk and outward near the surface layers. This behaviour is exactly opposite to what is seen in laminar viscous disks, as shown by the blue (scaled) curve. In [8] we show that this opposite flow structure is a direct consequence of an anisotropic turbulence, where the $Z - \phi$ component of the turbulent stress tensor is much larger than the $R - \phi$ part, see also [6].

III. THE DENSITY DISTRIBUTION

Having calculated the efficiency of the VSI-turbulence, we ran comparison models for viscous disks using exactly this α-value. In Fig. 3 we compare the resulting 2D surface density structure (obtained by integrating the density vertically over the disk) of the VSI turbulent case with the corresponding α-disks. Shown is the configuration near the end of the simulations after about 200 orbits of the planet. One notices that for the VSI disk vortices are visible at the outer and inner edge of the gap. These are created by the action of the planet in an inviscid disk. For the 30 $M_\oplus$ case there are two relatively weak vortices visible on either side of the planet. For the more
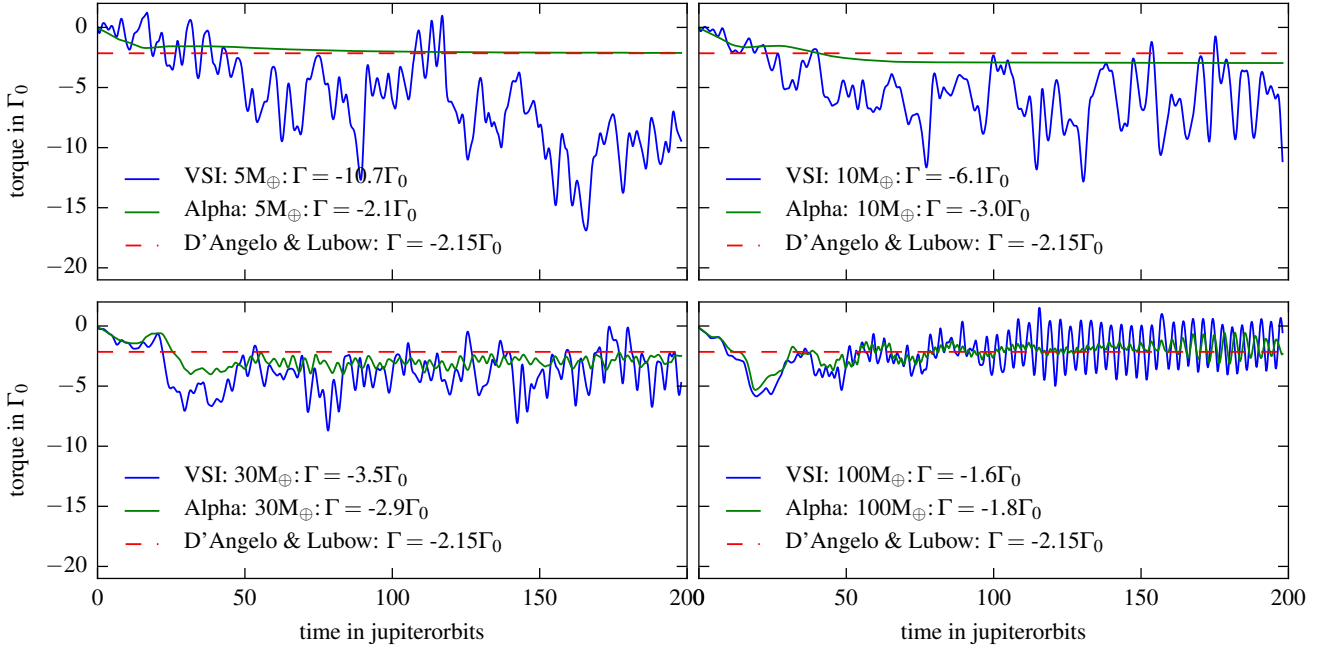
Figure 5. Time evolution of the total torque acting on the planets of different mass. Blue: VSI model, green: corresponding α disk, red dashed: results of [1] for low mass planets. The quoted torque values are averaged over the last 50 orbits of the simulations. Adapted from [9].

massive planet there is only one single vortex remaining on each side.

In Fig. 4 we compare the radial surface density distributions, obtained by additional averaging over the azimuthal direction. Clearly, the gap structure of VSI- and α-disk agree very well with each other. This is an additional indication that the radial angular momentum transport generated by the turbulence can be captured nicely by a corresponding viscous flow using an appropriate value for the effective viscosity.

IV. THE TORQUES ACTING ON THE PLANETS

The time evolution over the last 50 orbits of the total torque acting on the planets with different masses is shown in Fig. 5. Displayed is the torque evolution for the VSI turbulent disk (blue), and the laminar α-disk (green). We compare our results to the studies by [1] and we normalize the total torque to

$$\Gamma_0 = \Sigma(R_{\mathrm{p}})\Omega^2(R_{\mathrm{p}})R_{\mathrm{p}}^4 \left(\frac{M_{\mathrm{p}}}{M_*}\right)^2 \left(\frac{R_{\mathrm{p}}}{H}\right)^2, \qquad (3)$$

where M_* denotes the mass of the central star. The torques acting on the planets in the viscous disk are significantly smoother than those in the turbulent disk, and they match the findings of previous studies [1] very well. Those were obtained for low mass planets in the linear regime and hence the normalized values all agree. The torques for the larger planets (30, 100 $M_\oplus$) agree well with the expectation but the values for the smaller masses are on average more negative that the laminar values which leads to a faster inward migration towards the central star. The results come about because in the turbulent disk case there is *on average* a density enhancement just inside and behind the planet, creating a negative torque
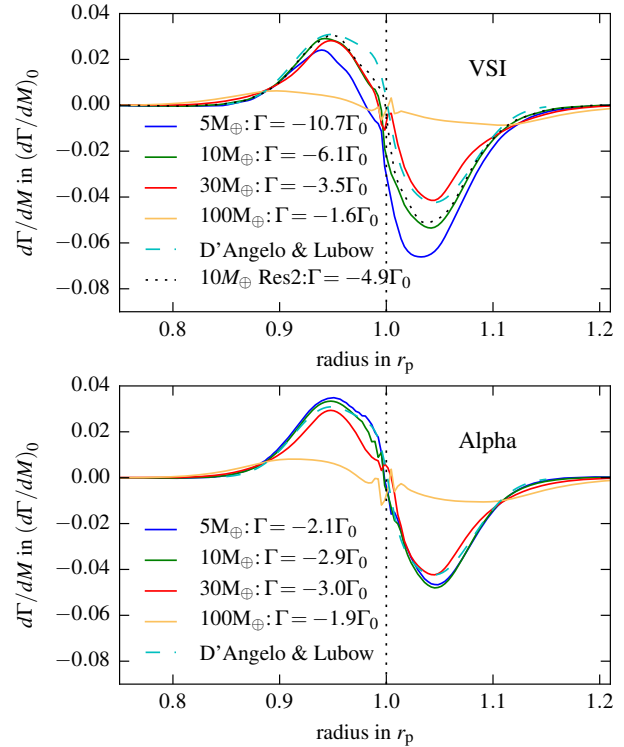

Figure 6. Radial torque density for planets of different masses for the VSI case (top) and α-disk (bottom), averaged over the last 50 orbits. For comparison the analytic profile, that has been obtained by fitting numerical results by D'Angelo & Lubow (2010) [1], is shown. The label RES2 refers to a model with double spatial resolution. Adapted from [9].

contribution. For larger planet masses this density structure is no longer visible. The stochastic fluctuations (in the first 3 panels for the lower masses) are due to the turbulent motions in the disk. These are typically smaller than those found in MRI-turbulent disks as the turbulence is more vigorous in ideal MHD simulations, where they can be much larger than the mean values leading to extended phases of outward migration [5], [10]. The very regular variations in last panel ($100 M_\oplus$ planet) resembles the beat period with the outer vortex (see Fig. 3, bottom right). The torque average over the last 50 orbits is quoted in the legend for each of the four cases.

To analyse our results further, we study the radial distribution of the torque density, which we express in units of

$$\left(\frac{d\Gamma}{dm}\right)_0 = \Omega^2(R_\mathrm{p}) R_\mathrm{p}^2 \left(\frac{M_\mathrm{p}}{M_s}\right)^2 \left(\frac{R_\mathrm{p}}{H}\right)^4 . \qquad (4)$$

The radial torque density is defined such that the total torque is given by the integral over all contributions of the disk

$$\Gamma_\mathrm{tot} = 2\pi \int_\mathrm{disk} \frac{d\Gamma}{dm} \Sigma r\, dr . \qquad (5)$$

The results are displayed in Fig. 6 for the VSI disk (top) and α-disk (bottom). All model show the generic feature of the Lindblad torques where inside of the planet the torque density is positive and outside negative. This structure is created by the action of the two spiral arms originating at the planet location, see Fig. 3. For the laminar α-disk (lower panel) the normalized torque densities agree very well with the linear expectation of [1], while for the 100 $M_\oplus$ planet the onset of gap formation reduced the torques next to the planet, but in such a way to maintain approximately the same total torque as seen above. For the turbulent disk (upper panel) the torque density changes such that inside of the planet the torque is less positive while on the outer disk the torque is more negative. This leads altogether to a significantly increased inward migration speed for the two smaller planets, while the massive planets migrate at approximately the speed that is expected. This is indicated by the quoted total torques, Γ, that are obtained by integrating the individual components of each cell over the whole disk, see eq. (5).

V. Summary

In this work we have studied the interaction of embedded planets in turbulent disk driven by the vertical shear instability (VSI). For the analysed masses (ranging from 5 to 100 $M_\oplus$) the planets have only a very limited effect on the turbulence level in the disk. In our locally isothermal models with $H/R = 0.05$ the angular momentum transport created by the VSI corresponds to an effective turbulent viscosity $\alpha = 5 \cdot 10^{-4}$. Low mass planets have only a small impact on the disk surface density but planets more massive than about 10 $M_\oplus$ begin to carve gaps in the disk. The gap edges become unstable to the Rossby wave instability and vortices develop. Starting from several smaller vortices in the initial phase, eventually they merge to one large vortex one on each side of the planet. The lifetime of the vortices is much longer

in the VSI-active disk due to vorticity perturbations generated by the VSI.

For planets with masses smaller than about $10 M_\oplus$ the time averaged torques are negative throughout. For the lowest mass we find inward migration which is about 5 times faster than for the corresponding α-disk model. This effect is due to a density bump behind the planet, that becomes noticeable in the time averaged surface density distribution. From our results we can infer that inviscid disks will generate turbulence driven by the VSI with a finite turbulence level. Hence in realistic 3D discs, the depth of the gap created by the planet is limited, such that planet migration will always occur. We do not expect that planets will stall, even in inviscid disks, as has been suggested for example by [2].

VI. HPC Resources

The simulations for this project were performed in the vast majority on the ForHLR-I high performance computer system installed at the Steinbuch Centre for Computing (SCC) in Karlsruhe. For the simulations the 3D hydrodynamic code `PLUTO` [4] was applied using MPI-paralllization. Each run used 400 processors (20 processors per node and 20 nodes). They lasted on average 7 days and used in total 67200 core hours per different run

References

[1] G. D'Angelo and S. H. Lubow, "Three-dimensional Disk-Planet Torques in a Locally Isothermal Disk," *ApJ*, vol. 724, pp. 730–747, Nov. 2010.
[2] J. Fung and E. Chiang, "Save the Planet, Feed the Star: How Super-Earths Survive Migration and Drive Disk Accretion," *ApJ*, vol. 839, p. 100, Apr. 2017.
[3] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
[4] A. Mignone, G. Bodo, S. Massaglia, T. Matsakos, O. Tesileanu, C. Zanni, and A. Ferrari, "PLUTO: A Numerical Code for Computational Astrophysics," *ApJS*, vol. 170, pp. 228–242, May 2007.
[5] R. P. Nelson, "On the orbital evolution of low mass protoplanets in turbulent, magnetised disks," *A&A*, vol. 443, pp. 1067–1085, Dec. 2005.
[6] A. A. Philippov and R. R. Rafikov, "Radial Transport and Meridional Circulation in Accretion Disks," *ApJ*, vol. 837, p. 101, Mar. 2017.
[7] M. H. R. Stoll and W. Kley, "Vertical shear instability in accretion disc models with radiation transport," *A&A*, vol. 572, p. A77, Dec. 2014.
[8] M. H. R. Stoll, W. Kley, and G. Picogna, "Anisotropic hydrodynamic turbulence in accretion disks," *A&A*, vol. 599, p. L6, Mar. 2017.
[9] M. H. R. Stoll, G. Picogna, and W. Kley, "Planet-disc interaction in laminar and turbulent discs," *A&A*, vol. 604, p. A28, Jul. 2017.
[10] A. L. Uribe, H. Klahr, M. Flock, and T. Henning, "Three-dimensional Magnetohydrodynamic Simulations of Planet Migration in Turbulent Stratified Disks," *ApJ*, vol. 736, p. 85, Aug. 2011.