2016

Annual Report
Jahresbericht

BEYOND THE LIMITS

Heidelberg Institute for
Theoretical Studies

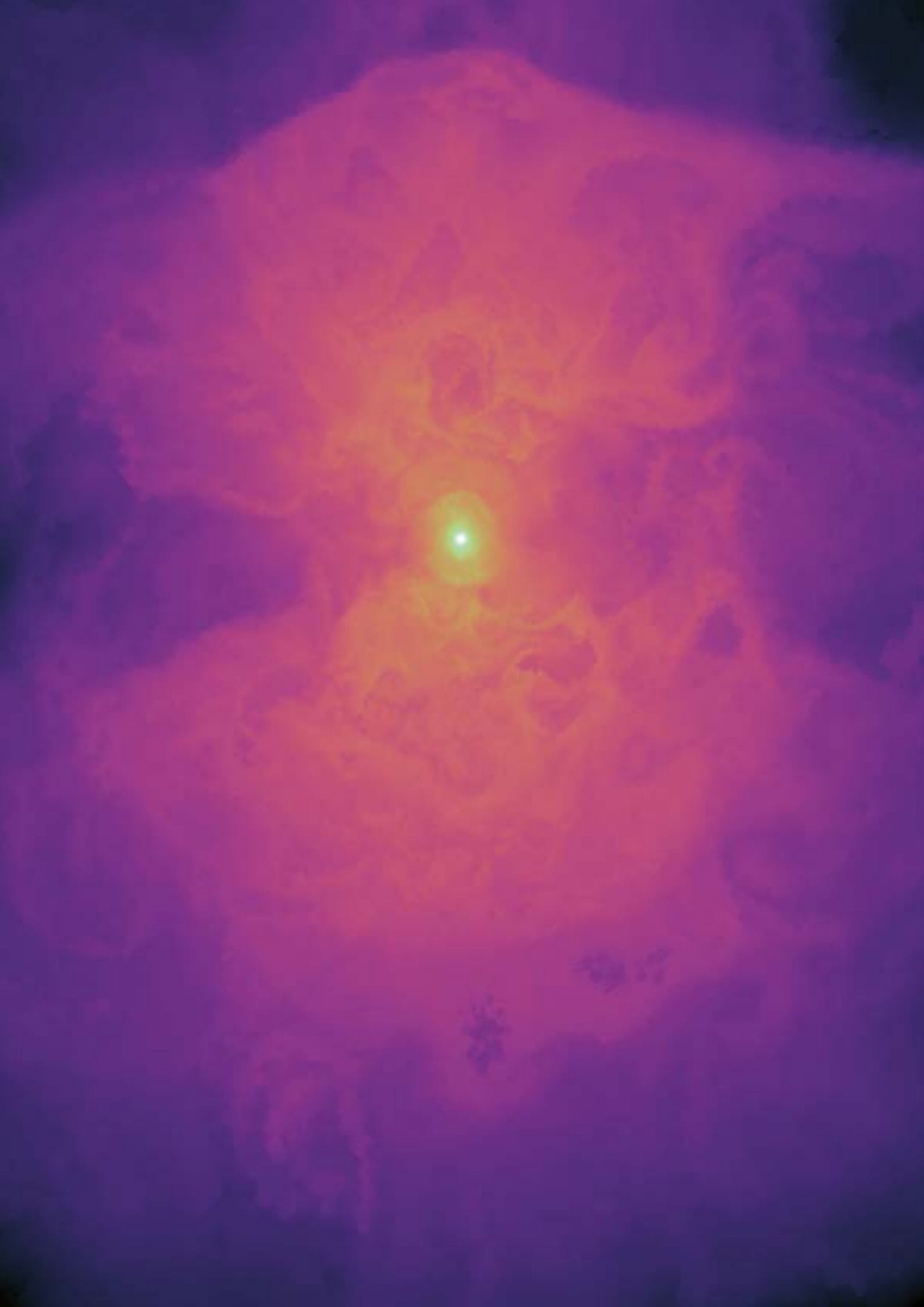
HITS

Turbulent times: When stars approach

Screenshot from a computer simulation of the common-envelope phase of binary stars. HITS researchers discovered dynamic irregularities that may help to explain how binary stars evolve to become supernovae (see *Chapter 2.10*, p. 96 ff).

Turbulente Zeiten: Wenn sich Sterne näher kommen

Ausschnitt einer Computersimulation von Doppelsternsystemen in der Phase der gemeinsamen Hülle. HITS-Forscher entdeckten dabei dynamische Unregelmäßigkeiten, die helfen können, die Entwicklung von Doppelsternsystemen hin zu Supernovae besser zu erklären (siehe Kapitel 2.10, S. 96 ff).



Inhalt
Table of Contents

1
Think Beyond the Limits!

Prof. Dr. Rebecca Wade
(Scientific Director / Institutssprecherin)

Dr. Gesa Schönberger
(Managing Director / Geschäftsführerin)

This is a foreword to the annual report for 2016 but let's first look back at the foreword to the 2014 annual report. It began with "The first start-up phase of HITS has been completed" and went on to "Now that the institute has reached its 'cruising altitude'". So, in 2016, what did the "cruising altitude" look like?

On the management side, Gesa Schönberger began as Managing Director of HITS in January 2016 and Andreas Reuter stepped down in April following his retirement (see *Chapter 10*). Andreas Reuter, along with Klaus Tschira, built up HITS from its very beginning. His leaving is thus a significant loss for HITS but we are glad that he continues to advise in his role in the management of the HITS Stiftung. Rebecca Wade completed her two-year term as Scientific Director at the end of 2016 and Michael Strube took over this role in January 2017. The HITS management is supported by the Shareholders Board and the Scientific Advisory Board (see *Chapter 10* for a Who's Who). Jeannette Wing stepped down from the Scientific Advisory Board after her three-year term and we thank her for her constructive and enthusiastic engagement.

At the beginning of the year, we were struggling to find desks for all the scientists and students who wanted to do research at HITS. In the summer, we were at last able to increase the number of desk spaces available to HITSters by expanding to use the whole of Villa Reiner and by opening a second site at the Mathematikon. We expect that the Mathematikon site will facilitate interactions between HITSters and scientists at the university and other institutes on the Neuenheimer Feld campus. The Scientific Databases and Visualization (SDBV, see *Chapter 2.12*) group led by Wolfgang Müller has moved to the offices in the

Mathematikon where they are closer to some of their cooperation partners and next door to other institutions founded by Klaus Tschira or his foundation: EML European Media Laboratory GmbH, the Heidelberg Institute für Geoinformation Technology (HeiGIT) and Heidelberg Mobil International GmbH. At year end, the HITS working space had certainly reached "cruising altitude" but, with the planned addition of a group in scientific visualization, as well as a gradual increase in the number of junior research groups, the empty desks can only be temporary.

At "cruising altitude", we can expect research groups to come and go. Indeed, Christoph Pfrommer started his ERC-supported junior group in High Energy Astrophysics and Cosmology (HAC, see *Chapter 2.6*) in April 2016 but will move in April 2017 to take up a Professorship at the University of Potsdam and establish a research group at the Leibniz Institute for Astrophysics in Potsdam.

During the year, HITS continued to build its connections in a multitude of ways. HITS became a scientific partner of the Heidelberg Laureate Forum Foundation, thereby paving the way for a long-term synergistic cooperation. The HITS-Yale program in Astrophysics was established; it supports two postdoctoral "Tschira fellows" and exchange visits between HITS and Yale University in the USA (see *Chapter 6*). New research cooperations extended as far as Australia, with HITS scientists participating in the Australian Centre of Excellence for All Sky Astrophysics in 3 Dimensions (CAASTRO-3D) project. Amongst many outreach activities, the HITS open house event in July was very well attended and HITS scientists contributed to a successful series of lunch-time talks in St Peter's church in the Heidelberg Old Town on maths-related topics. For the first time, a radio science reporter, Michael Stang, was this year's "Journalist in Residence", and we also hosted Alexander Mäder as a visiting science journalist (see *Chapter 4*). Most importantly, the HITS scientists continued to carry out excellent cutting-edge research, some of which was recognized by a remarkable number of awards in 2016.

While we leave you to decide from reading this report, we think it is fair to conclude that HITS has actually reached its "cruising altitude" but we won't be surprised when the "altitude" changes in the coming years! Finally, we thank the Klaus Tschira Foundation for their continual support without which it would never have been possible to reach "cruising altitude" or to pursue our goal, to "think beyond the limits!"

1
Think Beyond the Limits!

Dieses Vorwort führt in den Jahresbericht 2016 ein, doch bevor wir damit beginnen, erlauben Sie uns einen kurzen Blick in das Vorwort zum Jahresbericht 2014. Dort heißt es ... „die erste Aufbauphase des HITS ist in allen wesentlichen Punkten abgeschlossen“ und weiter „das Institut [hat] nun seine „Reiseflughöhe“ erreicht“. Mit Blick auf 2016 stellt sich deshalb die Frage: Wie sieht diese „Reiseflughöhe“ aus?

Auf der Seite des Institutsmanagements gab es wesentliche Veränderungen: Gesa Schönberger begann im Januar 2016 ihre Tätigkeit als Geschäftsführerin des HITS und übernahm diese Aufgabe vollständig, als Andreas Reuter im April 2016 sein Amt als Geschäftsführer niederlegte. Andreas Reuter hat HITS gemeinsam mit Klaus Tschira geplant und aufgebaut. Sein Ausscheiden ist ein großer Verlust. Deshalb sind wir froh darüber, dass er HITS weiterhin beratend als Vorstand der HITS Stiftung zur Verfügung steht (siehe auch Kapitel 10). Zum Ende des Jahres beendete Rebecca Wade außerdem ihre zweijährige Tätigkeit als Institutssprecherin und übergab das Amt zum Januar 2017 an Michael Strube. Die Institutsleitung wird unterstützt von den HITS-Gesellschaftern und dem Wissenschaftlichen Beirat (zu beiden Gremien siehe auch Kapitel 10). Jeannette Wing, seit drei Jahren Mitglied des wissenschaftlichen Beirats, legte ihr Amt Ende 2016 nieder. Das HITS dankt ihr sehr für ihr Engagement und ihre enge Verbundenheit mit dem Institut.

Zu Beginn des Jahres standen wir vor dem Problem, ausreichend Arbeitsplätze für die vielen neuen Wissenschaftler und Studierenden am HITS bereitzustellen. Das änderte sich Mitte des Jahres, als HITS die Villa Reiner komplett übernahm und zusätzlich einen zweiten Standort in Heidelberg eröffnete. Wir gehen davon aus, dass unser neuer Standort im Mathematikon die Zusammenarbeit mit den wissenschaftlichen Instituten auf dem Campus der Universität Heidelberg im Neuenheimer Feld erleichtert und intensiviert. HITS ist dort mit der Gruppe „Scientific Databases and Visualization“ (SDBV, Kapitel 2.12) und ihrem Leiter Wolfgang Müller zu finden, in enger Nachbarschaft zu einigen Kooperationspartnern sowie zu anderen Einrichtungen, die Klaus Tschira oder seine Stiftung ins Leben gerufen haben, wie der EML European Media Laboratory

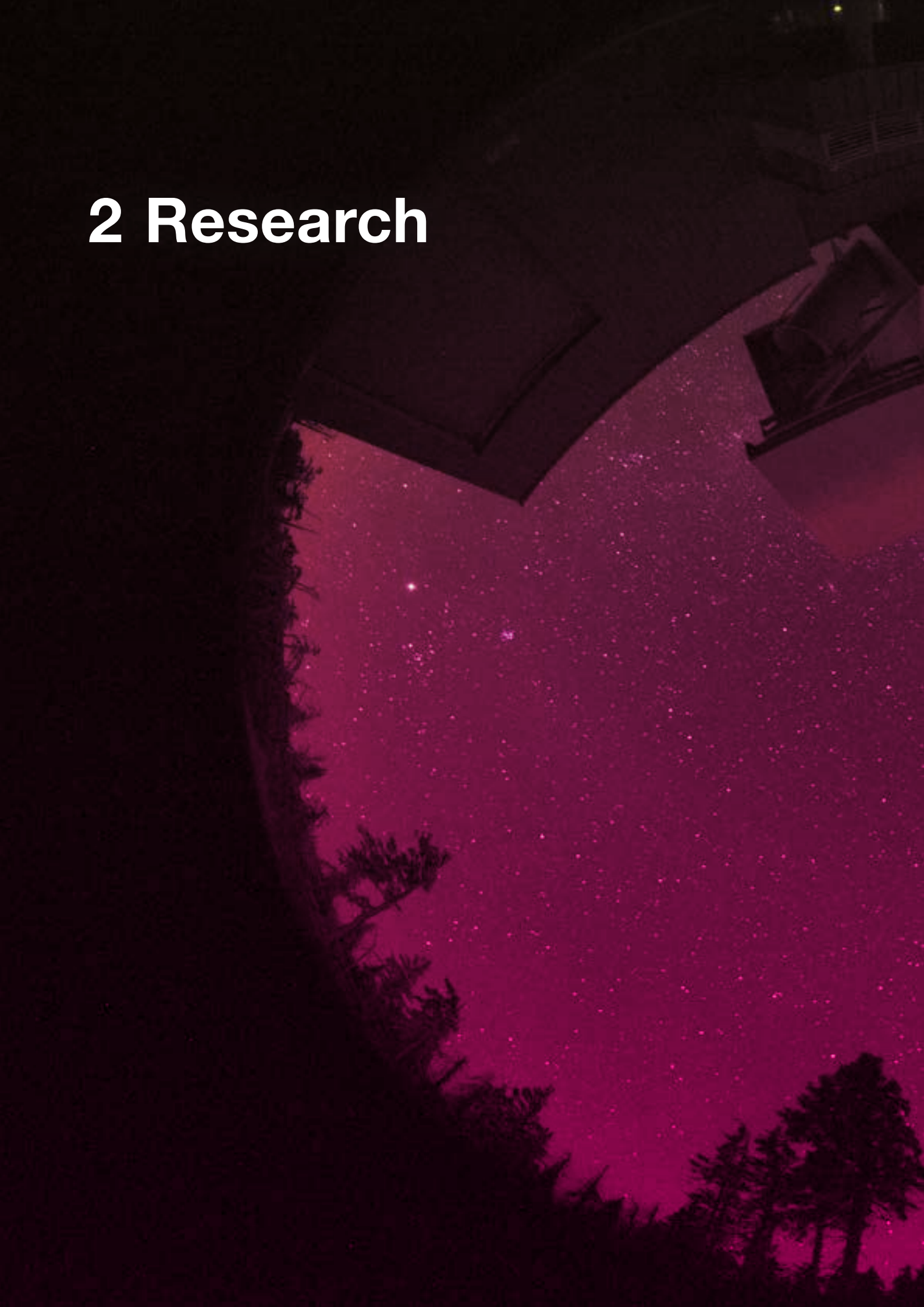
GmbH, dem Heidelberg Institute for Geoinformation Technology (HeiGIT) und der Heidelberg Mobil International GmbH. Was die Zahl der Arbeitsplätze im HITS angeht, haben wir mit Sicherheit inzwischen Flughöhe erreicht. Und die durch die räumlichen Erweiterungen freien Plätze werden der geplanten Gruppe für wissenschaftliche Visualisierung sowie der ein oder anderen neuen Juniorgruppe zur Verfügung stehen.

Um im Bild zu bleiben: Trotz „Reiseflughöhe" werden Gruppen am HITS immer kommen und gehen. Dass das schneller geht als erwartet, bewies die Gruppe „High-Energy Astrophysics and Cosmology" (HAC, siehe Kapitel 2.6). Christoph Pfrommer begann mit dem Aufbau seiner Gruppe unterstützt durch ERC-Fördermittel im April 2016. Bereits ein Jahr später nimmt er den Ruf auf eine Professur an der Universität Potsdam an und wird eine Forschungsgruppe am dortigen Leibniz Institut für Astrophysik aufbauen.

Das Jahr 2016 hat HITS erneut genutzt, um seine Verbindungen zu anderen Institutionen weiter zu vertiefen und auszubauen: Es wurde wissenschaftlicher Partner der Heidelberg Laureate Forum Foundation und schuf damit die Basis für eine langfristige Zusammenarbeit. Das HITS-Yale-Programm für Astrophysik wurde begonnen; es umfasst zwei Postdoktoranden, die als sogenannte „Tschira Fellows" je zwei Jahre am HITS und in Yale (USA) arbeiten, sowie eine kontinuierliche Zusammenarbeit beider Institute (siehe Kapitel 6). Neue Wissenschaftskooperationen reichten bis Australien: eine davon ist das „Australian Centre of Excellence for All Sky Astrophysics in 3 Dimensions" (CAASTRO-3D). Darüber hinaus waren HITS-Wissenschaftler an zahlreichen Veranstaltungen für die Öffentlichkeit beteiligt, wie einer Vortragsserie über Mathematik in der Heidelberger Peterskirche. Die Vorträge wie auch der Tag der offenen Tür des HITS stießen auf große öffentliche Resonanz. Erstmals kam mit Michael Stang ein Hörfunkjournalist als „Journalist in Residence" ans Institut. Außerdem konnten wir den Wissenschaftsjournalisten Alexander Mäder als Gast begrüßen (siehe Kapitel 4). Trotz aller Kommunikationsaktivitäten waren auch im Jahr 2016 die HITS-Wissenschaftler vor allem damit beschäftigt, ihre hoch innovativen Forschungsprojekte voranzutreiben, und nicht wenige erhielten dafür eine Auszeichnung.

Wir laden Sie ein, sich mit diesem Jahresbericht selbst ein Bild von HITS zu machen. Sie dürften dabei ebenso wie wir feststellen, dass HITS in der Tat seine „Reiseflughöhe" erreicht hat. Aber wir würden uns nicht wundern, wenn sich diese Höhe in den nächsten Jahren immer wieder ändern wird. Abschließend ist es uns ein großes Anliegen, der Klaus Tschira Stiftung für ihre ständige Unterstützung zu danken. Ohne diese wäre es nicht möglich, die aktuelle „Reiseflughöhe" zu erreichen, oder unser wichtigstes Ziel: Think beyond the limits!

2 Research



2.1 Astroinformatics (AIN)

2.1
Astroinformatics (AIN)

The Astroinformatics Group develops new methods and tools for dealing with the complex, heterogeneous, and large datasets currently available in astronomy.

Over the past two decades, computers have revolutionized astronomy. Advances in technology have given rise to new detectors, complex instruments, and innovative telescope designs. These advances enable today's astronomers to observe more objects than ever before and with high spatial / spectral / temporal resolution. In addition, there are new, untapped wavelength regimes still to be investigated. Dedicated survey telescopes map the sky and constantly collect data. The task we have set for ourselves is to enable scientists to analyze this increasing amount of information in a less biased fashion.

The Astroinformatics Group is interested in the development of improved methods for time-series analysis and redshift models based on photometric measurements. These are key tools for the analysis of data in upcoming large survey projects such as SKA, Gaia, LSST, and Euclid. Another scientific concern is the development of methods and tools for the extraction and filtering of rare objects (outliers) for detailed follow-up analysis with 8-m-class telescopes. With estimated occurrences of only a few objects per million, manual inspection of the existing catalogs is out of the question. The Astroinformatics Group's other interests include the morphological classification of galaxies based on imaging data as well as measuring similarity in high-dimensional data spaces.

Die Astroinformatik Gruppe entwickelt neue Methoden und Werkzeuge, um eine Analyse, der heutzutage verfügbaren komplexen, heterogenen und großen Daten im Bereich der Astronomie zu ermöglichen.

In den letzten zwanzig Jahren hat der Einsatz von Computern die Astronomie stark beeinflusst. Durch technologische Fortschritte wurde es möglich, neue Detektoren sowie innovative Instrumente und Teleskopdesigns zu realisieren. Dadurch können Astronomen nun Objekte mit bisher unerreichtem Detailreichtum und in neuen Wellenlängenbereichen beobachten. Mit speziell dafür vorgesehenen Teleskopen wird der Himmel wiederholt durchmustert und die so gewonnen Daten werden frei zur Verfügung gestellt. Durch unsere Forschung ermöglichen wir es Wissenschaftlern, diese riesigen Datenmengen durch neue Analysemethoden explorativ und unvoreingenommener zu erschließen und somit effizienter zu nutzen.

Unsere Gruppe beschäftigt sich mit der Zeitreihenanalyse sowie der Entwicklung photometrischer Rotverschiebungsmodelle. Dies wird für die neuen Generationen von Himmelsdurchmusterungen, benötigt. Des Weiteren beschäftigen wir uns mit der Suche nach astronomischen Objekten, die mit einer Häufigkeit von ein paar wenigen pro Million vorkommen. Um solch seltene Objekte für detaillierte Untersuchungen zu finden, scheidet die manuelle Selektion aus. Die morphologische Klassifikation von Galaxien sowie hoch-dimensionale Ähnlichkeitsmaße sind weitere Forschungsbereiche der Astroinformatik Gruppe.

Group Leader

Dr. Kai Polsterer

Staff Members

Dr. Nikos Gianniotis
Dr. Sven Dennis Kügler *(until May 2016)*
Dr. Jennifer Wagner *(since September 2016)*

Scholarship Holder

Antonio D'Isanto *(HITS Scholarship)*

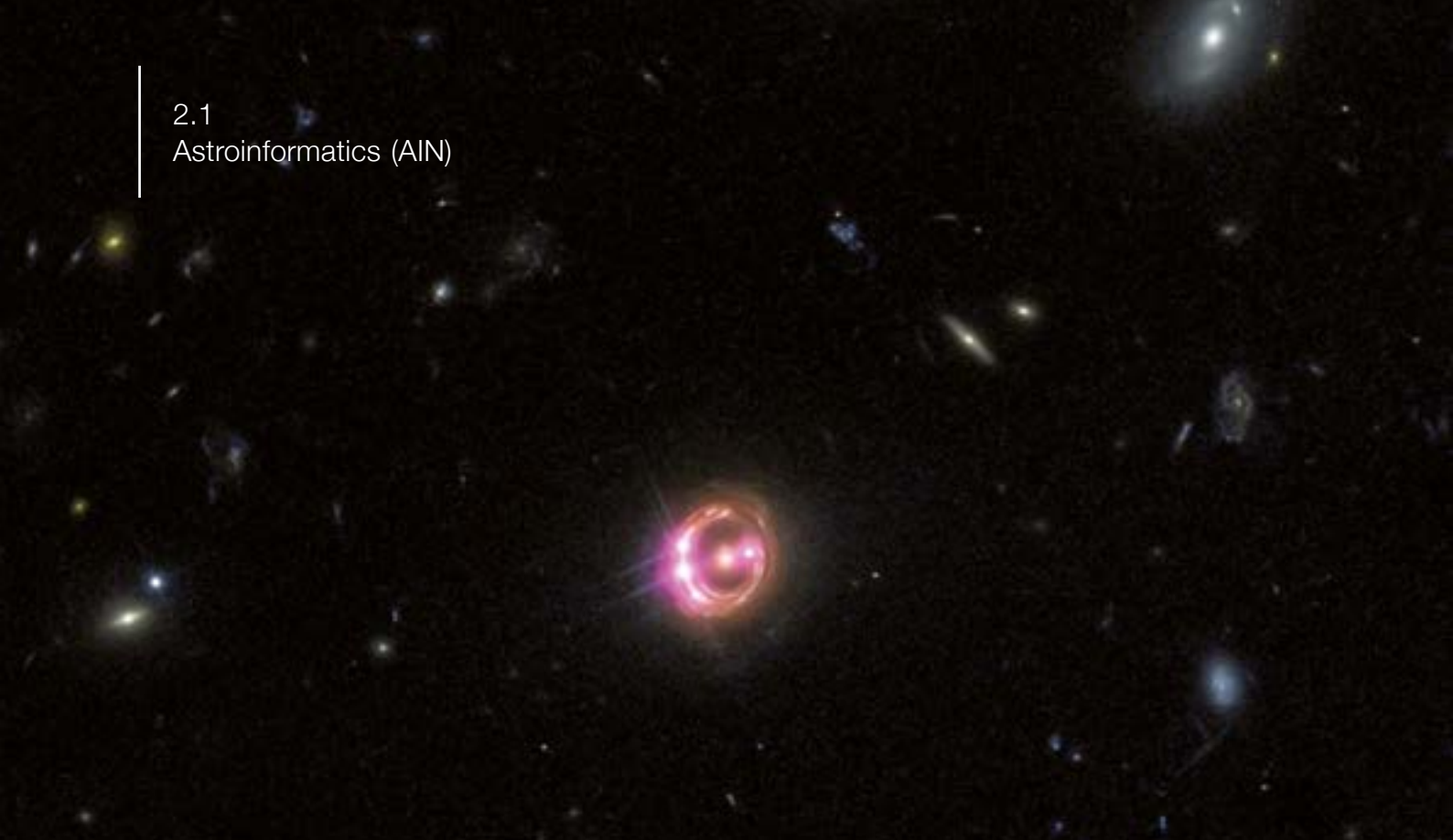


Figure 1: Quasars are believed to be one of the most distant types of objects known to us in the Universe. Studying these objects offers a glimpse of the early Universe. Image by NASA/CXC/Univ of Michigan/R.C.Reis et al; Optical: NASA/STScI.

Redshift estimation

Determining the distance of astronomical objects through redshift is an important task in astronomy due to its fundamental role in cosmological research, and estimating the redshift of quasars is of special interest. Quasars are super-massive black holes that accrete material and thereby create extreme luminosities *(Figure 1)*. Quasars are among the brightest objects in the universe, and their redshift is of mainly cosmological origin, i.e., related to the expansion of the universe. Hence, detecting highly redshifted quasars provides us with a glimpse into the early universe. However, estimating redshift via spectroscopic analysis is extremely demanding due to long integration times and costly instrumentation requirements. Consequently, the use of pure photometry has been proposed as a practical alternative.

Redshift estimation has been a major focus of the Astroinformatics Group. As is detailed in the next sections, we have pursued research along two avenues: the development of a probabilistic approach based on first principles and, complementarily, the development of a data-driven approach that operates directly on raw image data.

Bayesian photometric redshift estimation

We approached the problem of photometric redshift estimation by formulating a probabilistic model based on first principles. Working with photometry brings up a number of issues, namely that (a) the set of photometric measurements often consists of only a few values, (b) the measurements are subject to considerable noise, and (c) the observed measurements may originate from a wide variety of physical mechanisms. These issues render the estimation of redshift via photometric data problematic in the sense that they may admit a number of distinct solutions rather than one single solution. We postulate a probabilistic model that explains how observed photometry arises. In a first stage, the model constructs a representative set of spectra that are intended to capture the behavior of all possible quasar spectra. In a second step, the model explains how – given a particular spectrum and redshift value z – observed photometry is produced by integrating over filter curves. The generation of photometry is the "forward" mode that takes us from a spectrum and redshift z to a set of observed magnitudes. In the "backward" mode, we

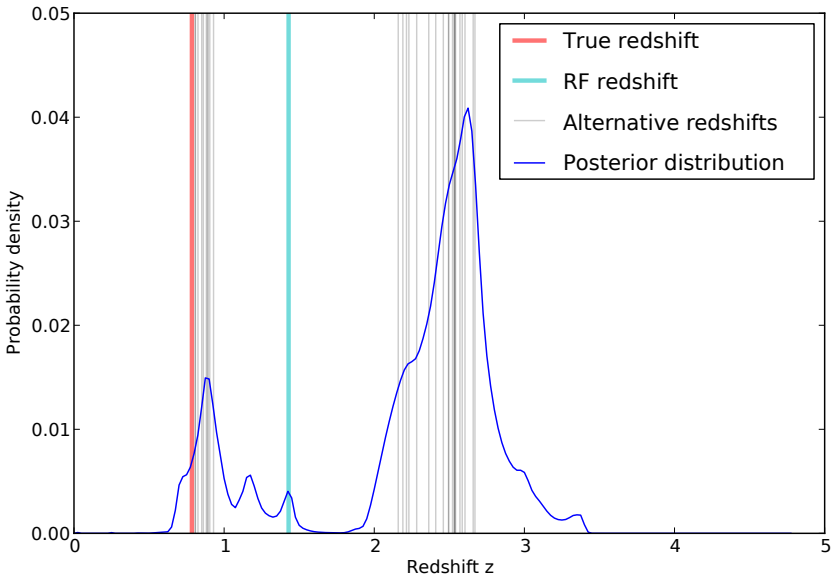


Figure 2: The red and cyan lines represent the true redshift and the predicted value by the random forest, respectively. In addition, redshifts of objects that are similar in terms of color are shown as grey lines. In blue, we plot the density predicted by our model, which exhibits two dominating modes: One mode explains a smaller cluster of alternative redshifts (which includes the true redshift), while the other explains a larger, distinct group of alternatives redshifts.

present the model with observed magnitudes, which produces a probability density for redshift.

The proposed approach requires that spectra be positioned in their rest frames. However, in practice, not all wavelengths are observed for all spectra in the rest frame. This means that spectra appear incomplete, which prevents us from extracting photometry by integrating over the wavelength range. Hence, in the first step of the algorithm, we fill in the unobserved wavelengths of each spectrum, and we thus project all quasar spectra onto a common subspace that captures the main characteristics of the physical spectral behavior. We then generate complete spectra by mapping the projected spectra back onto the original spectrum space.

In the second part of the approach, we model how a spectrum shifted by a redshift z produces flux in a particular filter band and, hence, in different magnitudes. In order to model observation errors, we assume Gaussian noise on the magnitudes, which gives rise to a likelihood function. The probabilistic model is completed with prior distributions on the redshift z and spectra.

With all probabilistic elements in place, we can reason "backwards" and infer the most likely redshift z that underlies a given set of observed magnitudes. In probabilistic terms, this involves calculating the posterior distribution of z. As previously mentioned, redshift estimation is a tricky issue and may admit multiple solutions. The posterior distribution deals with this issue in a very natural manner by exhibiting multimodality, i.e., multiple peaks.

The multimodality of the redshift posterior distribution is illustrated in *Figure 2* for an object taken from the BOSS catalogue. The chosen object admits multiple redshifts as possible solutions. However, these multiple solutions are not readily available. Here, we assume that these solutions can be approximately recovered as the redshifts that belong to objects (in the BOSS catalogue) that are closest to the test object in terms of color. We plot these alternative redshifts as grey lines (see *Figure 2*). We observe that the grey lines cluster around two locations. This tells us that objects that are very similar in color may be associated with different redshfits. We also plot the true redshift as a red line, which is known in this case. The model posterior distribution for z is plotted as a blue curve. We observe that the dominating peaks of the model posterior overlap with the two clusters of grey lines. The fact that the true redshift appears closer to the smaller posterior peak as opposed to the larger one does not mean that our prediction is wrong; it merely means that our model assigns less probability mass to this redshift. Indeed, the fact that the number of grey lines is fewer at the smaller than at the larger peak justifies this behavior. The plot clearly reveals that both dominating modes of the model posterior are justified since grey lines appear close to both of them. Moreover, we plot the prediction of the random forest as a cyan line. The random forest (RF) cannot cope with multimodality; hence, its prediction is a compromise of the multiple modes, leading to a prediction in a region in which no grey lines are found. In numerical experiments on the BOSS catalogue, we demonstrate that our approach generally produces solutions of higher likelihood than does the RF.

Uncertain Photometric Redshifts via Deep Learning Methods

In this work, we pursue a data-driven approach to estimating redshift. We make use of two neural network models: a Mixture Density Network (MDN), designed to use photometric features as inputs to generate probability density functions (PDFs), and a Deep Convolutional Network (DCN) that extracts information from raw input images. We show that this approach performs better – in contrast to using pre-defined features – since it learns to make use of the information present in the raw data.

A Mixture Density Network is the combination of a feed-forward neural network and a Gaussian mixture model. The output of the network parametrizes a Gaussian mixture in order to define the means, variances, and weights that are capable of producing multimodal predictive densities. The means, variances, and weights are thus obtained directly by the outputs z of the network. Instead of using the usual negative log-likelihood as a loss function, we employ the continuous rank probability score (CRPS) in order to obtain a trained MDN, which produces both, well-calibrated and sharp PDFs as measured by the CRPS itself. The probability integral transform (PIT) is used to check the calibration of the generated PDFs.

A DCN is a model composed of several convolutional and sub-sampling layers coupled with a fully connected network. This type of architecture is geared towards learning from raw image data since our main purpose is to estimate redshifts directly from images without needing to extract photometric features. We couple a DCN with an MDN (Deep Convolutional Mixture Density Network – DCMDN) in order to produce photo-z PDFs directly from SDSS images, and we alternate convolutional and pooling layers that generate feature maps and a hierarchically compressed representation of the input data. The output of the convolutional network is used as input for the MDN to obtain a multi-modal predictive density for photo-z. Thus, the extraction of the feature maps is completed automatically by the network. Our network exploits GPU computing and runs on a cluster equipped with Nvidia Titan X.

In the experiments, we take data from the Sloan Digital Sky Survey Quasar Catalog V based on the Sloan Digital Sky Survey (SDSS)'s 7th data release, which consists of 105,783 spectroscopically confirmed quasars in a redshift range between 0.065 and 5.46. A random subsample of 50,000 patterns is used for every experiment. For each pattern, we take the five ugriz magnitudes as input features and the respective images in the same bands. Finally, the performances of MDN and DCMDN are compared with the widely used RF. The RF, in its original architecture, is not meant to provide results in the form of distributions. In order to obtain distributions, we first collect the predictions of each individual decision tree in the forest for every data item. Then, the PDF is defined for the RF by fitting a mixture of 5 Gaussian components to the individual tree outputs.

In the case of experiments using the RF and the MDN, we chose the 5 magnitudes plus all possible color combinations as input, thereby obtaining a 15-dimensional feature vector. The generated training and test set contain 25,000 patterns each. The DCMDN is trained on the images, which are obtained using the Hierarchical Progressive Survey's data partitioning format and performing a proper cutout on the client side in order to obtain the desired dimensions (16 x 16). Each pattern is originally a stack of 5 images in the ugriz filters. As with the usual features, we additionally form the color images from the ugriz images by taking all possible pairwise differences, thus obtaining a stack of 15 images; every object / pattern is then represented by a tensor of dimensions 15 x 16 x 16.

In order to obtain a rotational invariant network, we apply data augmentation, taking rotations of each image at 0, 90, 180, and 270 degrees. Thus, the training set is composed of 100,000 images, the validation set of 50,000 images, and the test set of 50,000 images. We also apply dropout to control overfitting.

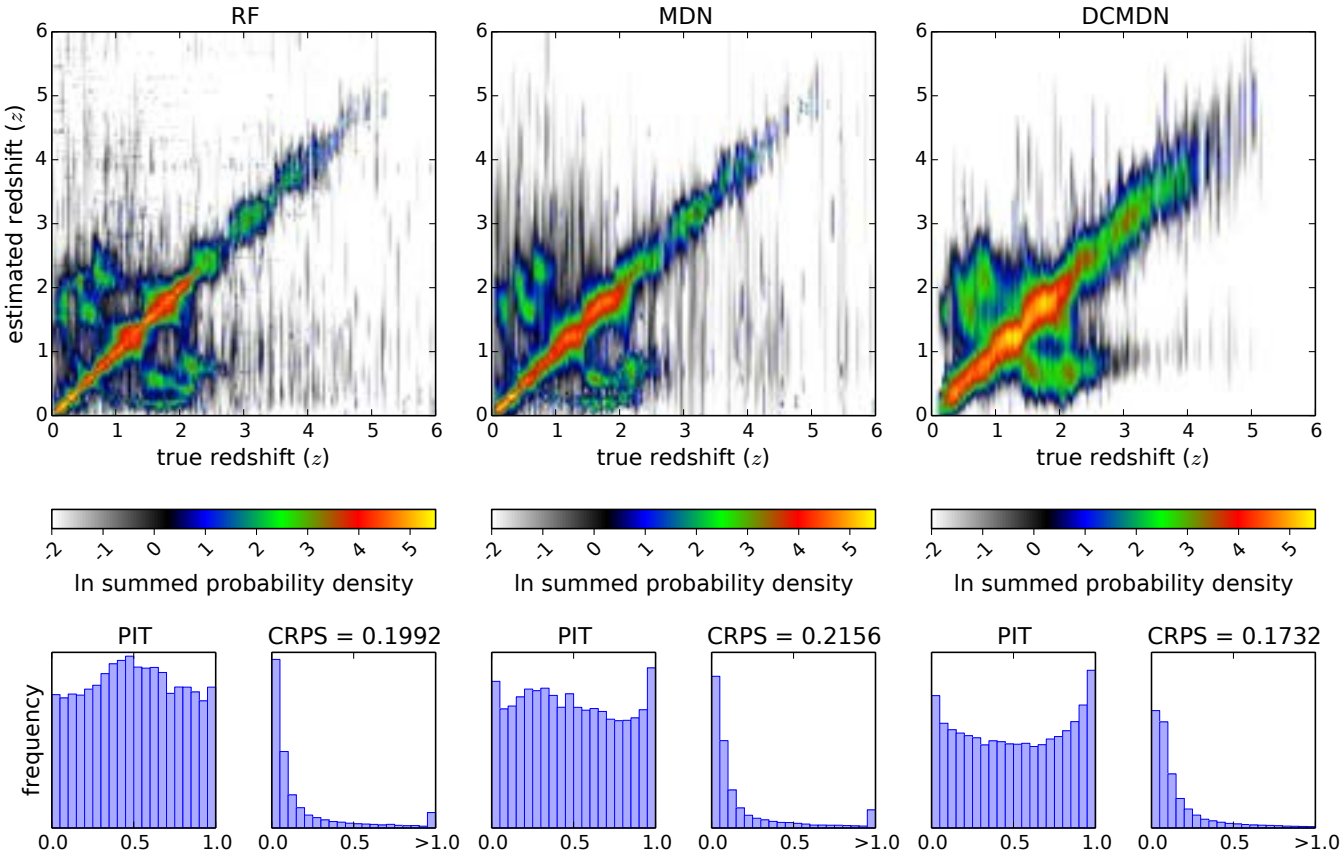


Figure 3: Results of the prediction obtained with the MDN and the DCMDN compared with the RF results. For each experiment, three plots are present. The upper plots compare the spectroscopic redshift with the predictive density produced by the model with the color indicating the logarithm of the summed probability density of the distributions. We can thereby observe the DCMDN's better generalization capability and the significantly reduced number of outliers produced by this architecture. In the two lower plots, the histogram of the PIT values and the histogram of the individual CRPS values are shown. The mean CRPS value is also reported.

In the RF experiment, the model reaches a CRPS score of 0.20 and the PIT shows a bit of overdispersion. The performance of the MDN (with a score of 0.21) is a bit worse than that of the RF in terms of the CRPS but exhibits a better calibrated PIT. Using the DCMDN architecture, we achieve the best results in terms of the CRPS, namely a score of 0.17. The resulting PIT is acceptable, although it continues to display some underdispersion. The reason for the DCMDN's better overall performance lies in the fact that the features-based approach uses only a fraction of the available information. In fact, a lot of information is lost in the process of feature extraction. Instead, using images, the DCMDN is able to automatically determine thousands of features, thereby providing a better prediction of the photo-z PDFs. Overall, the proposed architectures display better performances in the comparison carried out with an RF-based approach. In particular, it is clear that the proposed DCMDN provides the best performance since it is able to use all of the information contained in the images (*Figure 3*).

Model-independent, data-driven information retrieval

Standard, state-of-the-art approaches for characterizing a gravitationally lensing distribution of (dark) matter rely on model assumptions and mostly use χ^2-parameter estimation to retrieve the model parameters that optimally describe the observed (multiply-)imaged sources. As the observed ensembles of gravitationally lensed sources usually cannot be reproduced by simple lens models, the mass reconstruction is an iterative refinement process that often requires detailed fine-tuning by adding local substructures. Despite yielding a consistent picture, the resulting model is subject to a high degree of degeneracy because the same observational configuration can be described by various models with different parameters and substructures. In addition, finding such a fine-tuned lens model that consistently reproduces all observations cannot be feasibly automated because the parameter estimation by model fitting is so time consuming that a set of feasible models and parameter ranges has to be selected beforehand and because adding substructures to locally adjust the lensing mass is performed on individual observations and does not follow general principles. To overcome these issues, this research project focuses on the question of what can be learned about gravitational lens systems from multiply imaged sources without assuming a lens model (*Figure 4*).

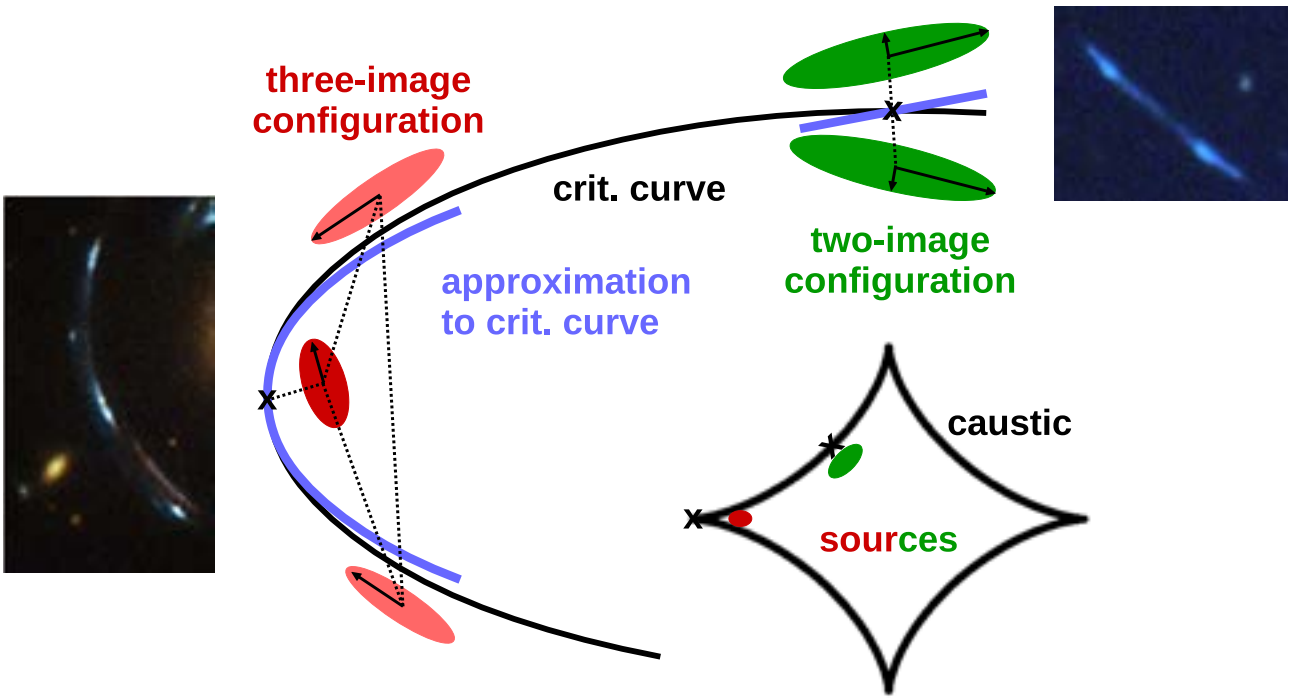


Figure 4: The model-independent characterization of gravitational lenses employs observables from two- and three-image configurations (large green and red ellipses, respectively; details from telescope pictures right and left) close to the critical curve (elliptical black line) in order to reconstruct ratios of potential derivatives at critical points (black crosses on the critical curve) and the shape of the critical curve (blue line and parabola) in their vicinity. The image configurations occur when sources are either close to a caustic segment (small green ellipse) or close to a junction of two caustic segments (small red ellipse).

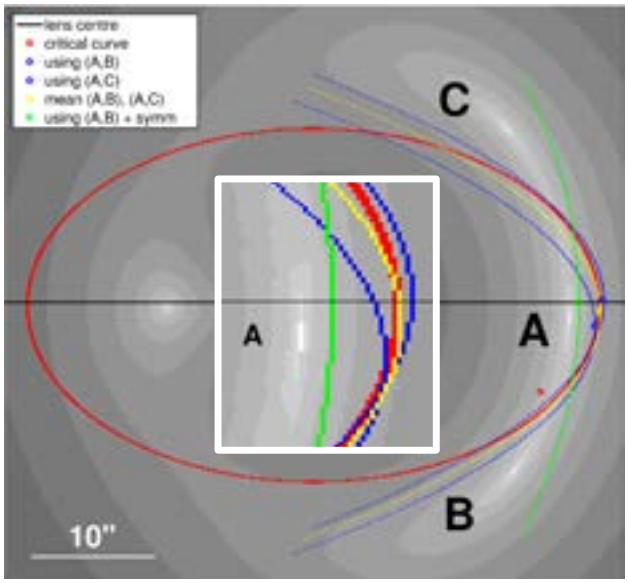
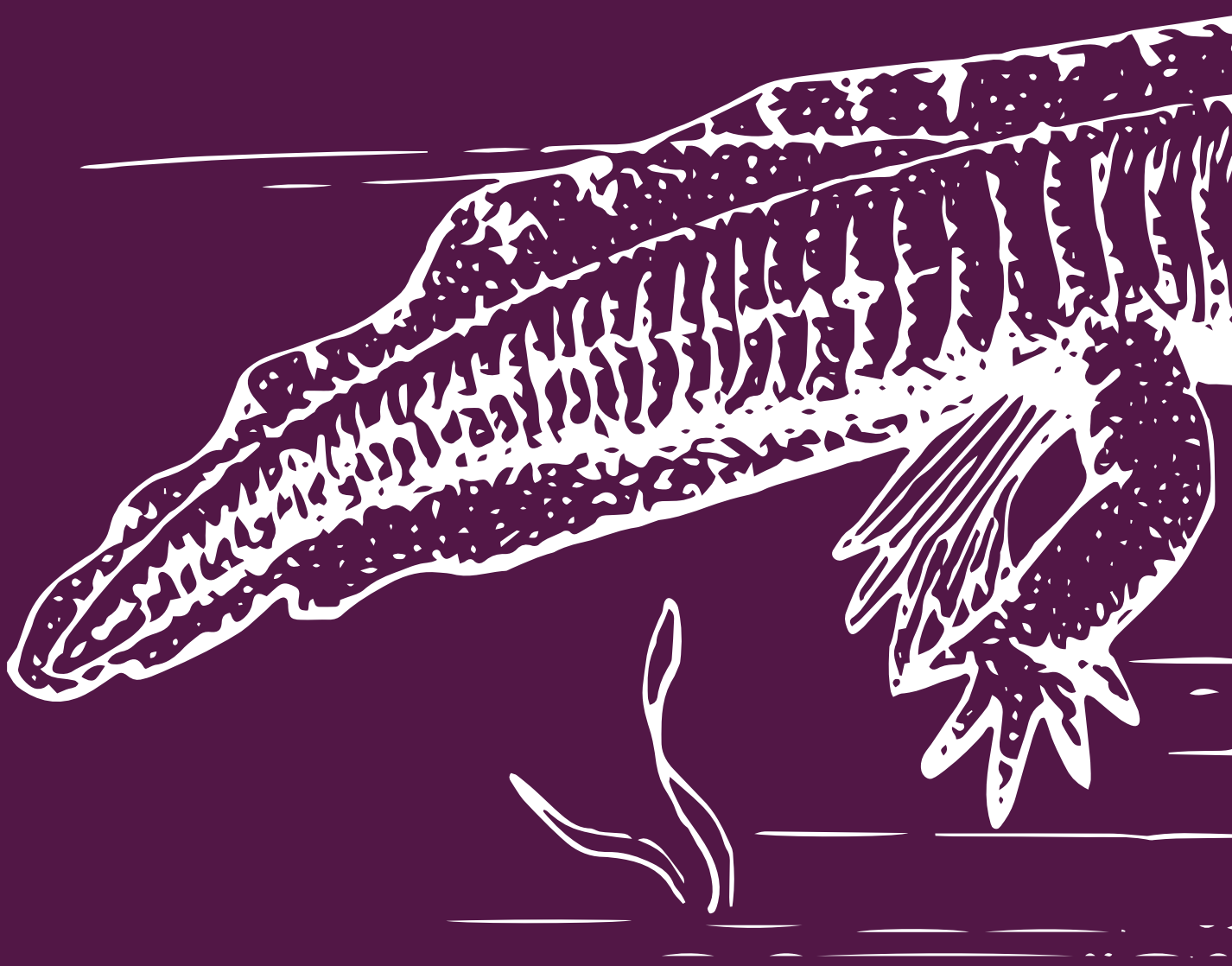
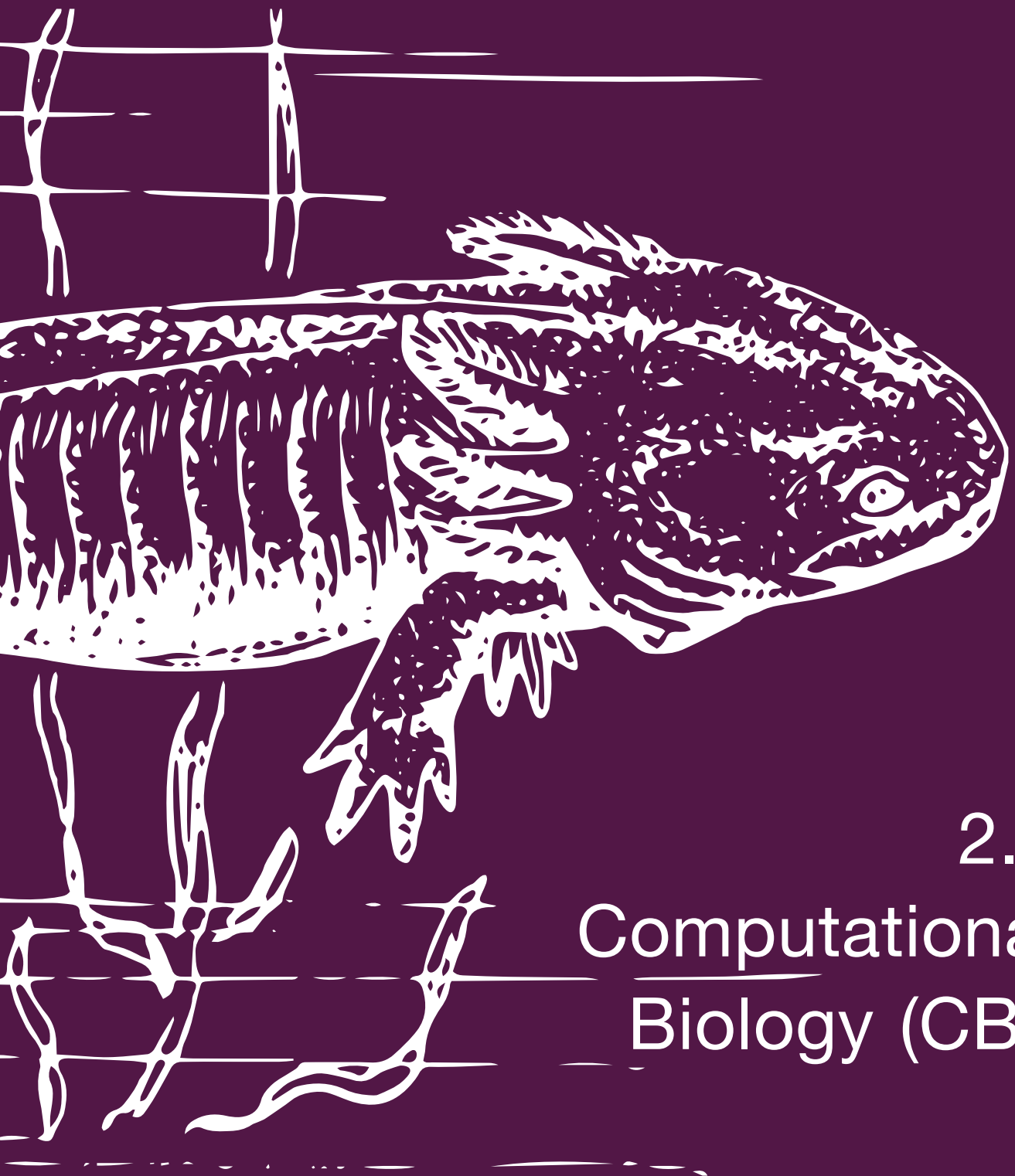


Figure 5: Reconstruction of the critical curve for a simulated three-image configuration (images shown as white spots; denoted by A, B, C) in a singular isothermal elliptical lens from observables of different pairs of images. The critical curve of the lens is given by an ellipse (red curve). The symmetry axis through the lens center (black line) is drawn to better enable the visualization of the location of the closest critical point (the crossing of the black line and the red curve) and to compare this point with the vertex positions (blue, yellow, and green dots) for the different reconstructions, which mark the reconstructed cusp point. The reconstruction of the fold point between A and B is also shown (red dot). On top, a detail of the cusp position reconstruction is shown (white square).

This approach is based on the generic properties of multiple images near critical curves of the lens mapping. The observables used are the relative image positions, the magnification ratios, ellipticities and relative orientations of extended images, and time delays between images with temporally varying intensity. These observables constrain derivatives and ratios of derivatives of the gravitational lensing potential near a critical curve, thereby revealing local information about the dark matter distribution in galaxies and galaxy clusters in a model-independent manner. Experiments based on a simulated three-image configuration indicate promising performance in recostructing the critical curve (*Figure 5*).

2 Research





2.2 Computational Biology (CBI)

2.2
Computational Biology (CBI)

The CBI group lies at the interface(s) between computer science, mathematics, and the biological sciences. Our research focuses on the computational and algorithmic foundations of genome biology. Of the multitudes of issues encountered in our field, we are especially interested in whole-genome assembly, i.e., the reconstruction of a genome's sequence from the data produced by a DNA sequencer. The basic principle applied for assembly is to randomly (over-)sample overlapping fragments from the genome, sequence them, and computationally reconstruct the full sequence from these fragments.

The complexity of this task is largely dependent on two characteristics of the fragments: average length and accuracy. The current generation of sequencers produces very long fragments but maintains high error rates; therefore, new approaches to the problem of efficient assembly under such conditions are needed. The development of such algorithms and their efficient implementation and application in genome sequencing projects thereby represent the group's main goals.

Die CBI Gruppe arbeitet an der Schnittstelle von Information, Mathematik und Biologie, mit Fokus auf die informatischen und algorithmischen Grundlagen der Genombiologie. Von der Vielzahl an Problemen in diesem Feld sind wir besonders an der Assemblierung von Genomsequenzen interessiert. Darunter ist die Rekonstruktion der Sequenz (Folge der Nukleotide) eines Genoms, basierend auf Daten, die durch einen DNA-Sequenzierer produziert wurden, zu verstehen. Das Prinzip hinter Assemblierung ist, aus dem Genom zufällig (überlappende) Fragmente auszulesen, diese zu sequenzieren und anschließend aus der Sequenz dieser Fragmente die komplette Genomsequenz mit computergestützten Verfahren zu rekonstruieren.

Die Komplexität dieses Ansatzes wird primär von der Länge der Fragmente und der Fehlerrate des DNA-Sequenzierers bestimmt. Die aktuelle Generation an Sequenzierern, welche sehr lange Fragmente, aber mit einer hohen Fehlerrate produzieren, erfordert neue algorithmische Ansätze, um Genome effizient unter solchen Bedingungen rekonstruieren zu können. Die Entwicklung solcher Verfahren und deren Anwendung in Genomsequenzierungsprojekten stellen die Hauptaufgaben der Gruppe dar.

Group Leader

Dr. Siegfried Schloissnig

Staff members

Martin Pippel
Thomas Sean Powell

Scholarship holders

Philipp Bongartz *(HITS Scholarship)*
Philipp Kämpfer *(HITS Scholarship)*

Introduction

The group's main focus is on de-novo genome assembly, the reconstruction of a genome's sequence from reads produced by a DNA sequencer. The complexity of the reconstruction is dependent on the genome to be reconstructed (size, ploidy, and repetitiveness) and the characteristics of the sequencer used. Three properties of the sequencer, namely speed, error rate, and (average) read length, are usually of interest, but speed only impacts the cost associated with the sequencing. The error rate and read length are tied to the underlying approach, which is employed by the sequencer. The sequencers of the first generation were able to produce reads with lengths in the 100 s of base pairs and a relatively low error rate < 2 %. Unfortunately, this was a very expensive, slow, and only semi-automated process (adding labor costs as a result) and was replaced by NGS (Next Generation Sequencing – the 2nd generation) in the late 1990s for mostly economic reasons. Consequently, the cost per base pair dropped significantly in the 2nd generation, and the error rate was held at the levels of the previous method, but the average read length dropped by an order of magnitude for the first market-ready models. The disadvantages of this technology became apparent when attempting to assemble anything beyond bacterial genomes. Even though the technology has improved over the years, with read lengths now reaching 200 – 300 bp and error-rates below one percent, the results when applying this technology to genomes of higher organisms have been less than satisfactory. The third generation, which came to fruition in the last couple of years, can be largely summarized as single-molecule sequencing. As the name suggests, the passing or processing of a single strand of DNA through a protein(-complex) can be observed, and its sequence can be inferred from the signals caused during this process (e.g., fluorescence or electrical current). For example, the SMRT technology from Pacific Biosystems (PacBio) observes the light emitted when fluorescently labeled nucleotides are incorporated into a DNA strand by a polymerase. These approaches pushed read lengths into the tens of kilo base pairs, but at the cost of a high error rate (> 10 %) due to the stochasticity of the process. Large changes in technology also drive changes in the software used the deal with the generated data. Initially, the overlap-consensus (OLC) method was employed, which (in a naïve approach) compared each read to every other read, thereby calculating overlaps between them and resolving unique paths in the graph structure induced by all overlaps, which thus produced longer stretches of sequence.

Longer reads improve the chances of correctly reconstructing repetitive regions due to their potential of spanning a repetitive element and thereby causing fewer junctions in the overlap graph.

Obviously, 100 s of base pairs are insufficient for most real-world repeat structures, so a hierarchical approach was employed. First, the genome was decomposed into overlapping clones (fosmids, cosmids, or BACs) ranging from 20 KB up to 300 KB that were sequenced and assembled independently, thereby reducing the problem significantly. The assembled clones were subsequently used for assembling the complete genome.

The OLC approach broke down computationally with the advent of NGS, which was able to produce millions of short reads (around 50 bp initially) very quickly. After abandoning the overlap-consensus model, assembly moved on to de Bruijn graphs (dBG), which are directed cyclic graphs that represent overlaps between k-mers. Each vertex corresponds to a k-mer (a stretch of k base pairs), and an edge indicates an overlap of length k-1. This rigid structure allows for the very quick creation of a k-mer overlap graph by decomposing all reads into k-mers and adding edges between k-mers that are shifted by one base (e.g., k = 3 and ACTG would result in vertices ACT and CTG with an edge from the former to the latter). A dBG can be built efficiently and implicitly "computes" overlaps between k-mers. Of course, the decomposition of reads into k-mers results in a loss of information, and errors in the reads induce up to k incorrect k-mers. Touring the graph (i.e., reading out the unique paths) is coupled with a plethora of heuristics to resolve junctions and deal with graph structures induced by read errors, thus leading the different dBG-based assemblers to produce vastly different result sets when given the same input data.

Figure 6: Schmidtea mediterranea (S.med) has the remarkable capacity to regenerate complete animals from small body parts within a week after amputation. (Image: Alejandro Sánchez Alvarado, (Creative Commons Lic.)).

Due to long reads and high error rates from single-molecule sequencing, two properties that make the reads incompatible (for most applications) with a dBG-based approach, the overlap-consensus model is making a comeback.

The research projects outlined below summarize our current work on MARVEL, our noisy long-read assembler, including its application in the assembly of various genomes and hybrid dBG / OLC assembly.

The MARVEL Assembler

The long reads from single molecule sequencers, which can be upwards of 10 k in length, are impressive, but with an error rate of 15 %, the resulting computational complexity of such reads is exuberant. However, truly random error positioning and near-Poisson single-molecule sampling imply that reference-quality reconstructions of gigabase genomes are in fact possible with as little as 30X coverage. Such a capability would resurrect the production of true reference genomes and enhance comparative genomics, diversity studies, and our understanding of structural variations within a population.

We have built an assembler that we call MARVEL, which we have used to assemble repetitive genomes of up to 32 Gb directly from a shotgun, long-read dataset that is otherwise currently only producible with the PacBio RS II sequencer. MARVEL is based on the string graph / overlap-consensus paradigm, its most important attributes being scalability, repeat suppression, and assembly using reads at their native error rate of 15 %. In collaboration with Gene Myers (Max Planck Institute of Molecular Cell Biology and Genetics in Dresden), we have developed a local alignment finder, called Daligner, for noisy reads.

One of the impediments to scaling the overlapper to large genomes lies in the genomes' repetitiveness. If every fragment of the genome were unique, then only true overlaps would be calculated. However, this is only partly the case for even the simplest bacterial genomes. The repetitiveness, which is ever-present at varying degrees, induces partial matches (i.e., local alignments) of the reads that are largely responsible for the overlapper's runtime and storage requirements. In order to alleviate this problem, we coupled the overlapper with an adaptive repeat finder, which begins tagging regions of a read as repeats as soon as enough local alignments have been calculated. This occurs parallel to overlapping and results in the continuous exclusion of parts of a read from the overlapping process. This has resulted in savings of an order of magnitude in disk space and runtime. The repeat finder is not specific to a single overlap job, but instead gathers local alignment statistics from all overlapper jobs and distributes repeat interval annotations for all reads to all of these overlapper jobs.

Scrubbing detects and corrects read artifacts that would lead to partial alignments that would subsequently be classified as local alignments (e.g., repeat induced) and

discarded during the assembly process. Artifacts can be either a sudden spike in insertions, lost bases, missed adapters, or polymerase strand jumps. Quality information about a read is derived from the alignments produced by the overlapper. We use this information a) to detect and trim low-quality regions, b) to repair regions in a read containing large numbers of insertions or to add missing sequencing lost through polymerase jumps, and c) to split reads at missed adapters and polymerase jumps.

The pipeline was developed with the common overlap-correct-overlap paradigm in mind, meaning that we overlap the initial raw reads, scrub and correct them, and overlap the corrected reads again. However, we soon realized that the correction can actually be skipped and the assembly can be made directly from the raw reads after scrubbing. This realization necessitated a more thorough scrubbing phase that is thorough enough to guarantee not only a < 1 % error rate after correction, but also proper assembly. The move to an uncorrected assembly was motived by the savings in compute time and the fact that very long reads (spanning a repeat) can often not be corrected reliably. The main realization here is that when restoring large-scale errors with uncorrected parts of other reads, the sequence used for the correction only needs to be good enough to allow the alignment to continue at the native error rate; it does not actually have to be the correct sequence.

Uncorrected assembly suffers from missing overlaps, which is mostly due to low-quality regions in the reads that lead the overlapper either to miss overlaps completely or to terminate the alignment early when a bad-quality stretch is reached, thereby causing problems with the transitive reduction (the elimination of as many edges as possible while still maintaining the same reachability) of the overlap graph. We therefore decided to forego transitive reduction and directly tour the overlap graph.

> 2016 marked the completion of the first version of the MARVEL assembler.

With a full front-to-end assembler on hand, we have now shifted our focus towards improving algorithmic aspects and the scalability of the separate modules, notably through the adaption of the assembly process to cloud-based computing environments.

The Assembly of the Drosophila Histone Complex

Histones are DNA-packaging proteins that allow for a significant DNA-length reduction and play an important role in gene regulation. During cell division, both the number of chromosomes and the number of histone proteins have to be doubled to package the newly created DNA. This is assumed to be the reason for the dozens of copies of the histone genes found within almost every eukaryotic genome. These copies are often spread throughout the genome, but they are sometimes clustered in one particular region.

In the Drosophila genome, five histone genes are encoded in one 5 kbp-long sequence. More than one hundred copies of this histone sequence are clustered on chromosome IIR. Assembling this cluster with previous short read technology or even with the longer Sanger reads has been inconceivable. However, with the advent of long read sequencing machines, the assembly of this highly repetitive region may finally prove feasible. Apart from the insights into the evolutionary genesis of a gene cluster that could be derived from a complete assembly, this project is an ideal playground when it comes to trying out ideas for resolving highly repetitive regions.

The longest PacBio reads span five to six copies of the histone sequence. Unfortunately, these copies are highly conserved, as copies of essential genes often tend to be. Nevertheless, even highly conserved genes allow for alternative codons (mutations that do not affect the peptide sequence), and the spacers between the five coding regions are able to accumulate a certain number of mutations, as well. In order to detect these discriminating variations between different copies of the histone sequence, we calculated a multi-alignment out of all histone sequences that occurred in the PacBio Drosophila Melanogaster dataset.

Based on further analyses of the (multi-)alignments, the detection of unique anchor points in the histone complex, clustering, and deep learning approaches, we were able to assemble the complete complex. We are now further investigating the possibility of performing the assembly of the complex in a completely automated fashion. This has proven to be rather difficult due to the noisy nature of the dataset and the often only subtle differences in the histone repeat copies. In order to effectively deal with these issues, we are using deep learning (i.e., neural network) approaches to perform targeted selection and the correction of discriminating read features that are needed to automate the assembly process of the histone complex and that are general enough to be used for similar problem settings.

The Sequencing and Assembly of the Axolotl

The axolotl (Ambystoma mexicanum) belongs to the order of Caudata (tailed amphibians) and is known as the Mexican salamander or Mexican walking fish. Its natural habitat (the lakes around Mexico City) is slowly disappearing due to urban expansion, water pollution, and the continued draining of lakes and wetlands. Therefore, the IUCN classified the axolotl as critically endangered in 2006 and included it on the Red List of endangered species.

The axolotl can reach a length of up to 30 cm (12 inches), can live for up to 10 years, and comes in various colors, which may explain its popularity as an exotic pet. Furthermore, due to its ability to regenerate, the axolotl is a model organism in regeneration and evolutionary research.

Its strong regenerative capabilities enable the axolotl to regenerate entire body structures, such as limbs, its tail, and its jaw, as an adult. Another characteristic of axolotl is its neoteny, meaning that it reaches sexual maturity without undergoing metamorphosis, i.e., it retains its juvenile features during maturation. In addition to the axolotl's regenerative abilities, the size of the genome, estimated at 32 gbp base and distributed among 14 haploid chromosomes, makes this species an interesting as well as challenging model for de-novo genome assembly.

Initial sequencing was based on BAC libraries and aimed at constructing an expressed sequence tag (EST) library. These approaches indicated that the axolotl's genes are 5x larger than human genes on average due to increased intron lengths and that the overall genic component of the salamander genome is approximately 2.8 gbp.

The genome sequencing was performed by our collaborators at the Max Planck Institute for Molecular Cell Biology and Genetics, the Systems Biology Center in Dresden, and the Research Institute for Molecular Pathology in Vienna. The creation of the primary dataset required 1.5 years of sequencing and consisted of a 28-fold random oversampling of this genome.

We performed an initial assembly using the MARVEL assembler, which resulted in roughly 37 billion local alignments, took 350,000 CPU hours to compute, and revealed a repeat content of at least 60 %. The assembly contained 28 gbp, the longest contig was 3 mbp, and the N50 was 250 kbp. We are currently investigating whether the repeat structure of the genome prohibits a long continuous assembly and whether the jump in one order of magnitude in genome size resulted in problems with our assembly pipeline.

The Sequencing and Assembly of Schmidtea mediterranea

Schmidtea mediterranea (S.med) is a freshwater flatworm and a member of the invertebrate Plathyhelminthes phylum. The full genome is estimated at between 800 mbp and 1 gbp and is distributed among four chromosome pairs with a nucleotide distribution of 70 % AT-rich. S.med has the remarkable capacity to regenerate complete animals from small body parts within a week after amputation. Furthermore, due to a large number of pluripotent stem cells spread throughout its body and the low cost of keeping it in a laboratory setting, S.med. has become a model organism in many research areas, such as stem cell biology, germ line formation, and regeneration.

Owing to the lack of physical maps, high AT-content, and high repeat density, not one single high-quality planarian genome of S.med is currently available. We are thus working on a draft assembly of S.med., which has

defied previous assembly attempts for many years now. The assembly is based on reads that have been produced with the current single molecule real time (SMRT) sequencing technology from PacBio. Our primary dataset consists of a 52-fold random oversampling of this genome.

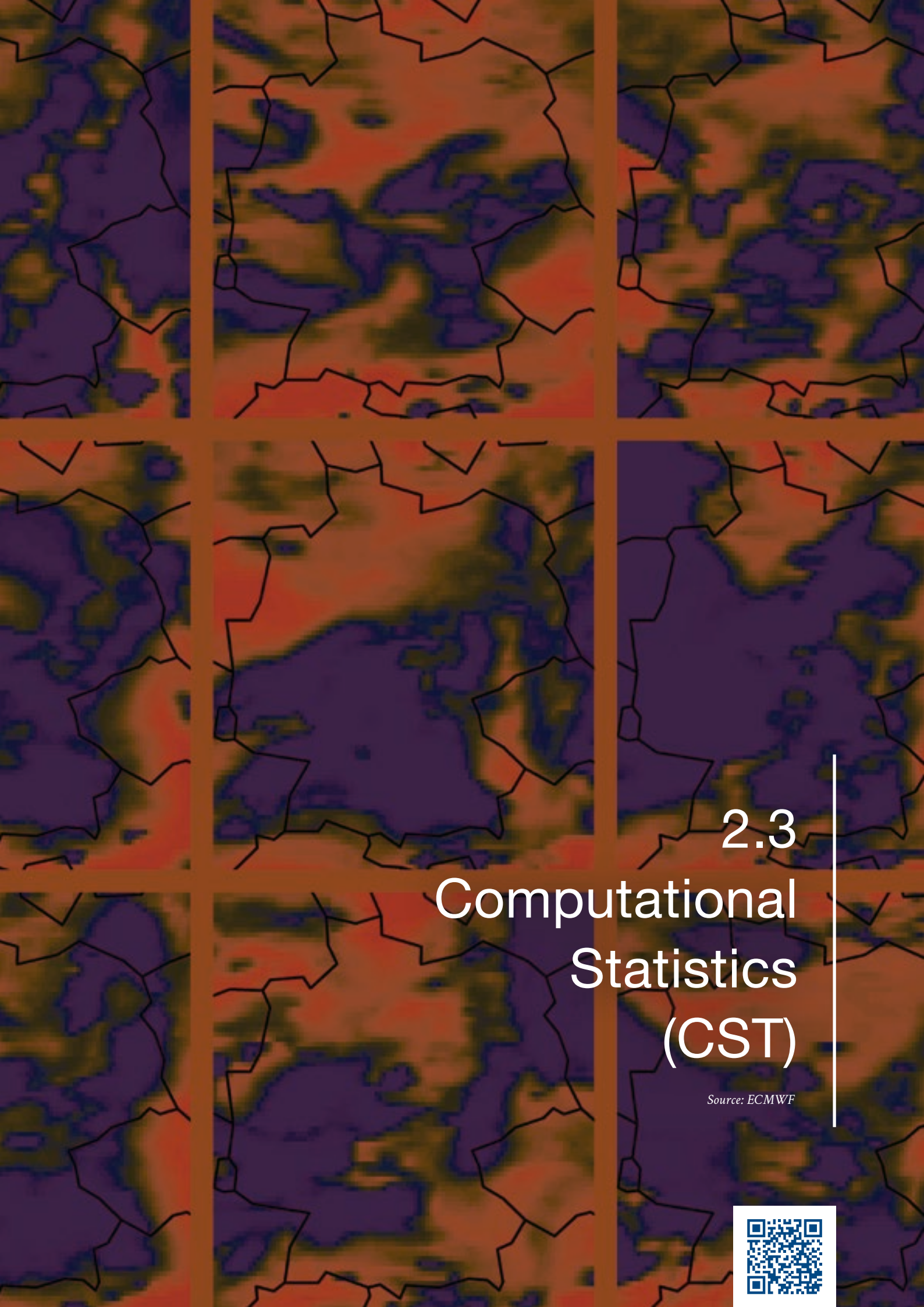
The S.med. assembly is based on our most recent assembly workflow, which included optimizations for an uncorrected assembly. Overlapping required 3,372 CPU hours with adaptive repeat masking and resulted in 61 % of the read-mass's being annotated as repetitive. The scrubbing and assembly were performed as previously explained and resulted in an initial assembly containing 726 mbp and with an N50 of 0.7 mbp.

The Assembly of the Round Goby

Gobiids are a diverse and fascinating group of small, predominately bottom-dwelling fish species with a global distribution. The Gobiidae family contains more than 1,700 species in more than 200 genera and is therefore one of the world's largest vertebrate families. Gobiids display a wide range of very special adaptations. The round goby's natural habitat is the Sea of Marmara, the Black Sea, the Caspian Sea, and the Sea of Azov, as well as the rivers adjacent to these seas. Round gobies are euryhaline (salt-tolerant) and are found in both freshwater and marine ecosystems. Since 1990, round gobies have been rapidly expanding in new habitats and have been found in the North American Great Lakes as well as in several rivers in Europe, where they have been rapidly displacing native species. Therefore, the invasive round goby Neogobius melanostomus has become a model species for invasion ecology and invasion genetics.

DNA sequencing was performed by the Center for Regenerative Therapies at the TU Dresden. Based on the estimated genome size of 1 Gb, a coverage of 50x of PacBio long reads (P6C4 chemistry tool kit) was sequenced. We used our MARVEL assembler and produced a final haploid assembly consisting of 1,364 contigs. The largest contig is 19.3 Mb, the NG50 is 2.8 Mb, and the assembled content amounts to 1004 Mb.

2 Research



2.3 Computational Statistics (CST)

Source: ECMWF

2.3
Computational Statistics (CST)

The Computational Statistics group at HITS was established in November 2013 when Tilmann Gneiting was appointed group leader and Professor of Computational Statistics at the Karlsruhe Institute of Technology (KIT). The group's research focus is on the theory and practice of forecasting, with secondary interests in spatial and spatio-temporal statistics and related areas.

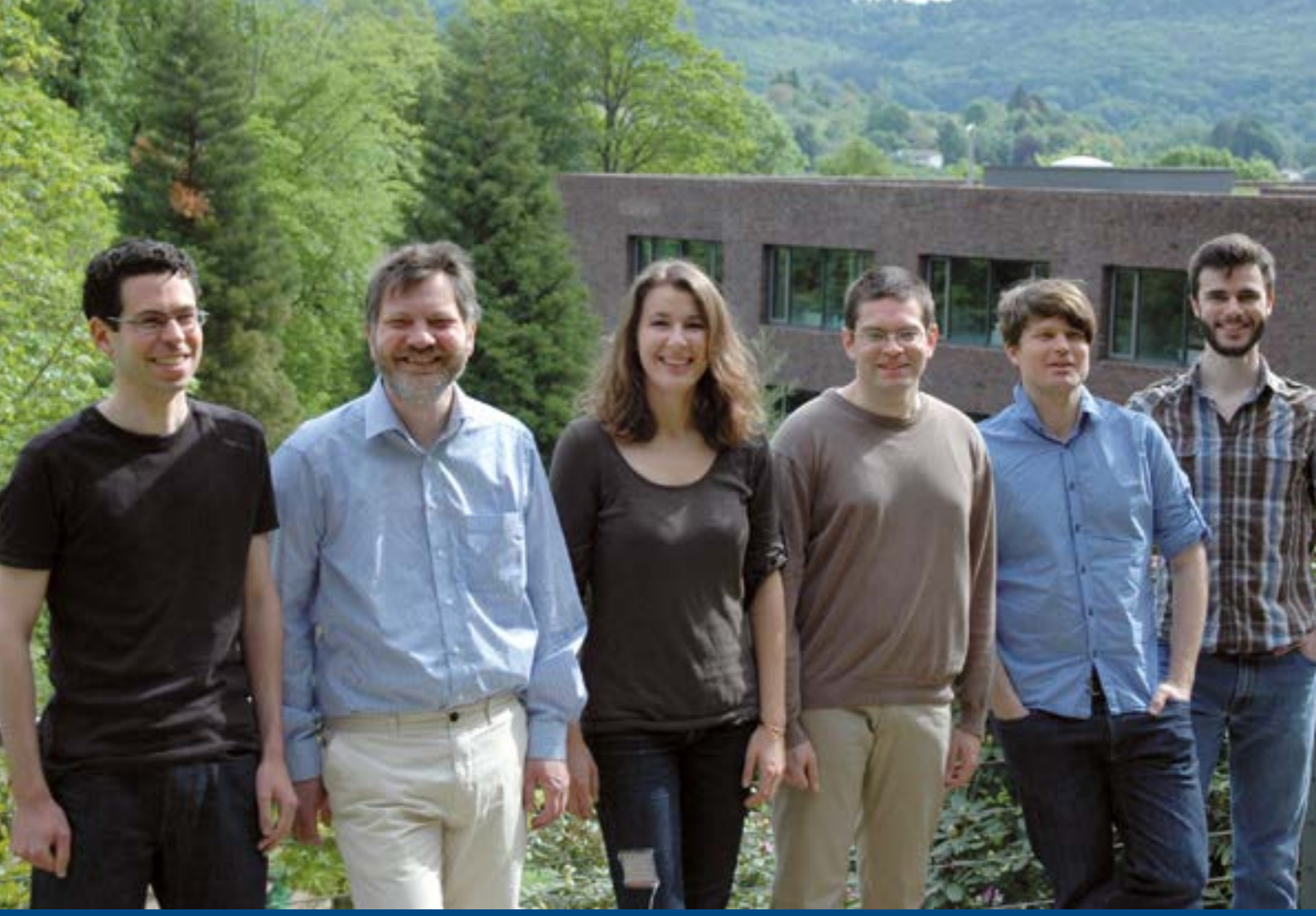
As the future is uncertain, forecasts should be probabilistic in nature, i.e., they should take the form of probability distributions over future quantities or events. Accordingly, we are currently in the midst of a trans-disciplinary shift of paradigms from deterministic or point forecasts to probabilistic forecasts. The CST group seeks to provide guidance and leadership in this transition by developing both the theoretical foundations for the science of forecasting and cutting-edge statistical methodology, notably in connection with applications. Our work in this area is supported by an Advanced Grant from the European Research Council in addition to other sources of external funding.

Weather forecasting is a key example of environmental prediction in real-time. The research group thereby maintains contacts and collaborative relations with national and international hydrologic and meteorological organizations, including the German Weather Service, the German Federal Institute of Hydrology, and the European Centre for Medium-Range Weather Forecasts.

Die Computational Statistics Gruppe am HITS besteht seit November 2013, als Tilmann Gneiting seine Tätigkeit als Gruppenleiter sowie Professor für Computational Statistics am Karlsruher Institut für Technologie (KIT) aufnahm. Der Schwerpunkt der Forschung der CST Gruppe liegt in der Theorie und Praxis der Vorhersage; weitere Interessen schließen unter anderem die räumliche und die Raum-Zeit-Statistik mit ein.

Im Angesicht unvermeidbarer Unsicherheiten sollten Vorhersagen probabilistisch sein, das heißt Prognosen sollten die Form von Wahrscheinlichkeitsverteilungen über zukünftige Ereignisse und Größen annehmen. Dementsprechend erleben wir aktuell einen transdisziplinären Paradigmenwechsel von deterministischen oder Punktvorhersagen hin zu probabilistischen Vorhersagen. Ziel der CST Gruppe ist es, diese Entwicklungen nachhaltig zu unterstützen, indem sie theoretische Grundlagen für wissenschaftlich fundierte Vorhersagen entwickelt, eine Vorreiterrolle in der Entwicklung entsprechender statistischer Methoden einnimmt und diese in wichtigen Anwendungsproblemen, wie etwa der Wettervorhersage, zum Einsatz bringt. Wir sind dankbar für externe finanzielle Unterstützung unter anderem durch den Europäischen Forschungsrat.

In diesem Zusammenhang bestehen Kontakte und Kooperationen mit nationalen und internationalen hydrologischen und meteorologischen Organisationen, wie etwa dem Deutschen Wetterdienst, der Bundesanstalt für Gewässerkunde und dem Europäischen Zentrum für mittelfristige Wettervorhersagen.



Group Leader

Prof. Dr. Tilmann Gneiting

Staff Members

Dr. Werner Ehm
Kira Feldmann
Dr. Stephan Hemri
Dr. Alexander Jordan
Dr. Fabian Krüger *(until September 2016)*
Dr. Sebastian Lerch
Dr. Roman Schefzik *(until May 2016)*
Patrick Schmidt *(HITS Scholarship)*

Visiting Scientists

Prof. Dr. Sándor Baran *(July 2016)*

General news

The CST group's research continues to focus on the theory and practice of forecasting, with external funding provided by the European Research Council via the Advanced Grant ScienceFore, the German Research Foundation, the Volkswagen Foundation, and the German Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde, BfG). Ongoing and fruitful interdisciplinary collaborations include joint work with hydrologists at the BfG in Koblenz and meteorologists at the Karlsruhe Institute of Technology (KIT) and the European Centre for Medium-Range Weather Forecasts (ECMWF) in Reading, United Kingdom.

As always, intense disciplinary and interdisciplinary scientific exchanges we have on many occasions represent an important and very enjoyable aspect of our work. In 2016, the Computational Statistics group was again happy to welcome guests from all over the world who work in a diverse range of scientific disciplines, including mathematics and statistics, economics and finance, meteorology, and seismology. Sándor Baran of Debrecen University in Hungary and Donald Richards of Pennsylvania State University in the United States visited over the summer. In June, we hosted a workshop on "Advances in Economic Forecasting," and in July, we held a theoretically oriented mini-symposium on "Positive Definite Functions and Related Topics."

The following sections describe facets of our work on the statistical postprocessing of ensemble weather forecasts. In the final section, we turn to research on the "forecaster's dilemma" that arises when public discussions of forecasts focus attention on predictive performance in cases of extreme events.

Statistical postprocessing of ensemble weather forecasts

In order to provide weather forecast information to the general public, national and international weather centers draw on highly sophisticated numerical models of the atmosphere that are run in real time on supercomputers and produce point forecasts of future atmospheric states. In a strong move towards probabilistic forecasts, these efforts have been transformed through the operational implementation of so-called ensemble systems. An ensemble forecast consists of multiple runs – typically between 10 and 50 – of numerical weather prediction (NWP) models, which differ from each other in terms of the two major sources of uncertainty, namely the initial conditions of the atmosphere and the mathematical representation of the respective physical processes.

Figure 7 displays the predictions from eight randomly selected members of the 50-member European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble system for precipitation over Germany, issued June 26, 2016, and valid in the late evening of June 27, 2016. The ECMWF ensemble is the most skillful and most widely used NWP forecast worldwide. The CST group maintains close links with the ECMWF facilitated by group leader Tilmann Gneiting's appointment as an ECMWF Fellow and extended research stays at the center's premises in Reading (United Kingdom) by Kira Feldmann and Stephan Hemri. Our work in this area focuses on the statistical postprocessing of NWP ensemble output with grant support from the ScienceFore project, the Volkswagen Foundation, and two subprojects in the German Research Foundation (DFG) funded Collaborative Research Center (SFB / Transregio) "Waves to Weather". In addition to funding PhD students and PostDocs, the projects allow for data access and structured collaboration with meteorologists at the Karlsruhe Institute of Technology (KIT), the ECMWF, and the German Weather Service (Deutscher Wetterdienst, DWD), thereby creating an inspiring interdisciplinary research environment.

Despite their undisputed successes, ensemble systems are subject to systematic errors, such as biases (predictions are systematically too high or too low) and dispersion errors (predictions are systematically spread out too much or too little). It is therefore common practice to statistically postprocess the NWP output. State-of-the-art techniques for doing this include the nonhomogeneous regression or ensemble model output statistics (EMOS) technique and the Bayesian model averaging (BMA) approach. The EMOS predictive distribution is a single distribution from a parametric family, such as the Gaussian

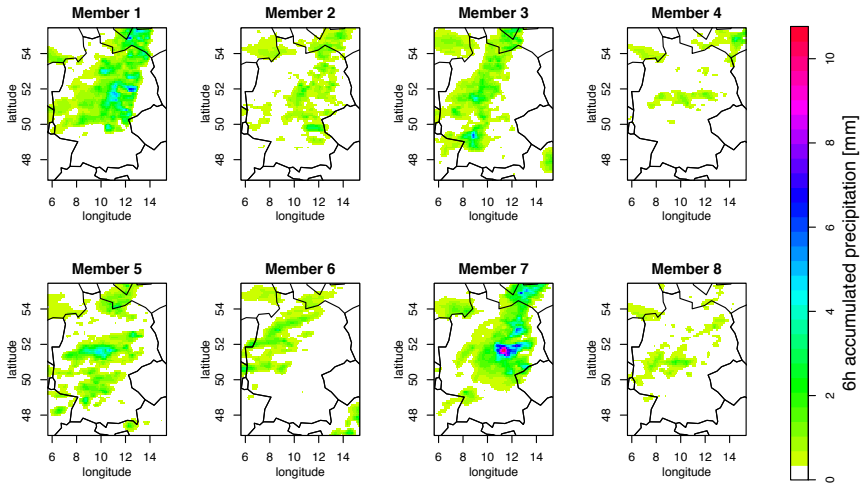


Figure 7: Members of the ECMWF forecast ensemble for six-hour accumulated precipitation over Germany at a lead time of 42 to 48 hours, valid June 27, 2016.

or normal family, in which the parameters depend on the ensemble forecast at hand in physically meaningful ways. In contrast, the BMA method employs a mixture distribution in which the mixture components depend on the ensemble member forecasts in suitable ways, with the mixture weights reflecting the corresponding member's skill over a training period. The choice of the parametric family used in the EMOS approach and of the mixture components employed by the BMA technique depends on the weather variable of interest. The simplest case arises in the EMOS model for temperature or air pressure, which uses the Gaussian family. In contrast, EMOS and BMA models for nonnegative weather quantities, such as wind speed or precipitation accumulation, require more complex statistical models that involve the use of gamma distributions and generalized extreme value distributions, among others.

Ensemble model output statistics for cloud cover

End-users in various areas, including but not limited to energy demand and renewable energy production, agriculture, astronomy, and tourism, benefit from skillful predictions of future cloud cover. Formally defined as the "portion of the sky cover that is attributed to clouds," observed cloud cover is typically reported as an integer between 0 and 8 (henceforth called oktas). Zero oktas are referred to as clear sky, 4 oktas to a half-cloudy sky, and 8 oktas to completely overcast conditions. In contrast, the European Centre for Medium-Range Weather Forecasts (ECMWF) cloud cover forecasts are issued as continuous values between zero and one hundred percent. *Figure 8* depicts the cloud cover predictions from eight randomly selected members of the 50-member ECMWF ensemble issued on June 26, 2016, and valid for June 28, 2016, at 01:00 am over Germany.

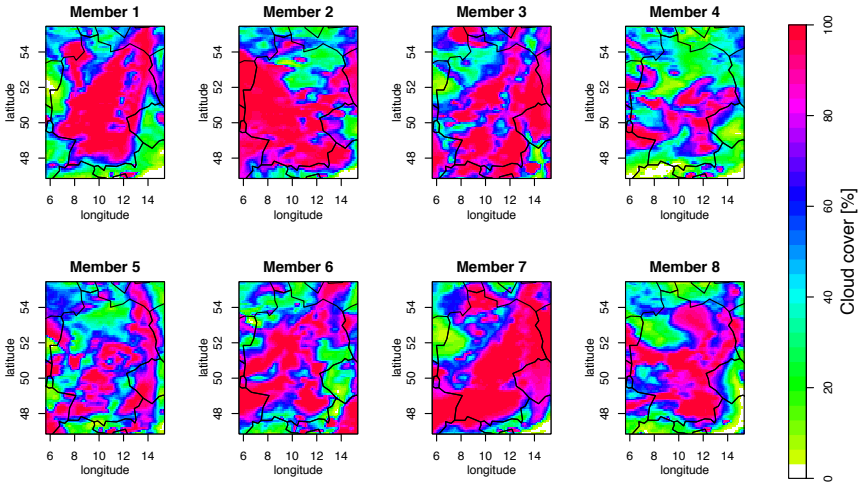


Figure 8: Members of the ECMWF forecast ensemble for cloud cover over Germany at a lead time of 48 hours, valid June 28, 2016, at 1:00 am.

The categorical nature of cloud cover observations calls for discrete postprocessing approaches. In joint work with colleagues at the ECMWF, Stephan Hemri has developed an ensemble model output statistics (EMOS) method for cloud cover [Hemri et al., 2016]. Based on the proportional odds logistic regression (POLR) technique, an EMOS model has been constructed that maps the continuous ECMWF raw ensemble forecasts to the nine different cloud cover classes given in oktas. As it turns out, POLR not only removes dispersion errors but is also able to sharpen the predictions. Sharpness can be understood as a measure of forecast confidence, i.e., the more confident a forecast gets, the sharper it is. Technically, sharpness corresponds to the width of prediction intervals. The improvement in terms of sharpness has been a pleasant surprise but can be explained by noting that the ECMWF ensemble tends to assign too much weight to the extreme cloud cover categories of zero and eight oktas, i.e., clear sky or completely overcast conditions. Statistical postprocessing can correct for this, as illustrated in *Figure 9* for cloud cover predictions over Mannheim Neuostheim, about 20 km west of the HITS premises.

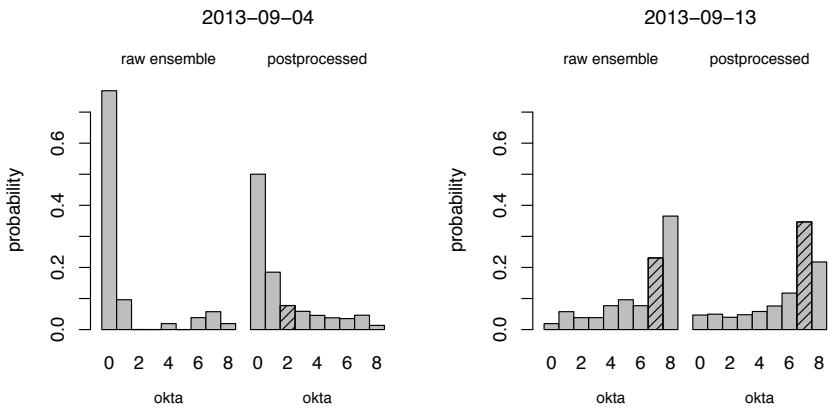


Figure 9: Examples of raw ECMWF ensemble and postprocessed categorical cloud cover forecasts at a lead time of 6 days over Mannheim Neuostheim, valid at 13:00 pm on the day indicated. The cloud cover that realized is shaded.

Multivariate postprocessing

Statistical postprocessing techniques, such as ensemble model output statistics (EMOS) and Bayesian model averaging (BMA), apply to a single weather variable at a single location and a single look-ahead time. They thus yield a postprocessed predictive distribution for each univariate weather quantity individually. When applied to the full output of a numerical weather prediction (NWP) model, which comprises multiple weather variables on regional or global scales and for multiple look-ahead times, a key challenge is retaining physically realistic and coherent joint dependence structures. This critical task requires multivariate or joint predictive distributions that are in line with the univariate predictive distributions when reduced to their univariate margins while also modelling dependencies in physically realistic ways. As expressed by the celebrated Sklar's theorem, every multivariate distribution can be decomposed into two essential components, namely the univariate marginal distributions and a so-called copula function that represents multivariate dependencies.

Sklar's theorem demonstrates that univariate approaches to statistical postprocessing can accommodate any type of joint dependence structure provided that a suitable copula is specified. If the dimension is small or if the specific structure can be exploited, such as with a temporal or spatial structure, parametric families of copulas can be fitted. For example, recent work by Kira Feldmann and her coauthors [Feldmann, K., Scheuerer, M., Thorarinsdottir, T. L. (2015), Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression, Monthly Weather Review, 143, 955-970] employs Gaussian copulas in a spatial stochastic process setting for surface temperature over Germany.

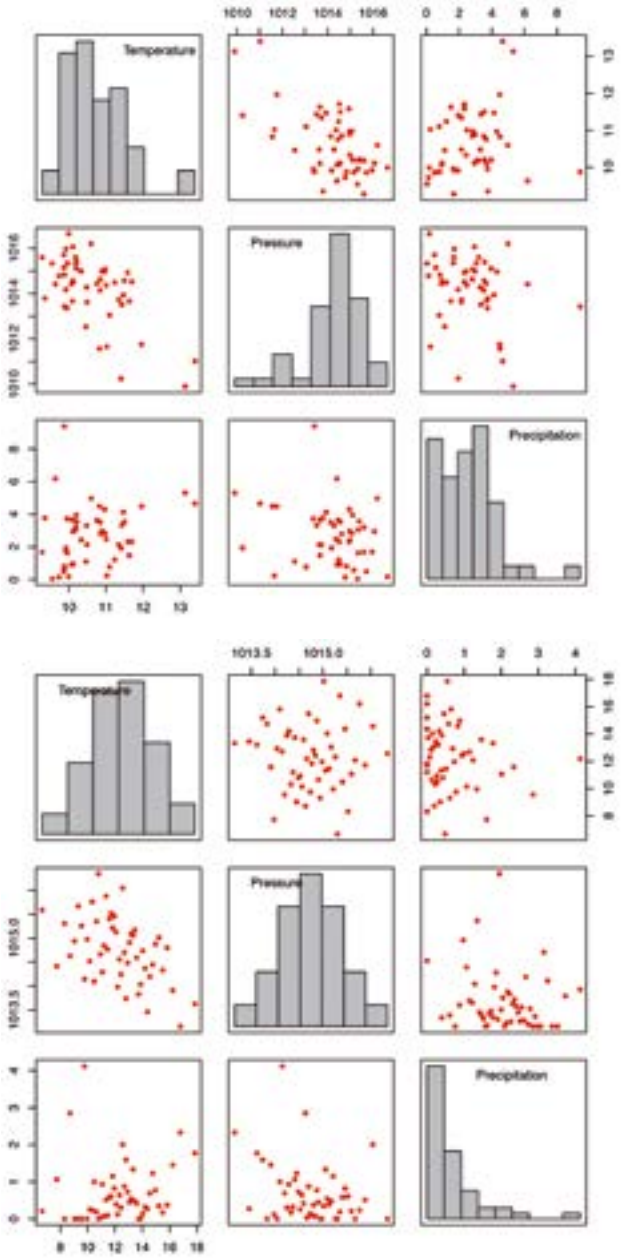


Figure 10: Twenty-four-hour ahead ensemble forecasts valid April 4, 2011, at 2:00 am local time at Frankfurt, Germany, for temperature, pressure, and six-hour accumulated precipitation, in units of degrees Celsius, Pascal, and mm. The panels on the diagonal indicate marginal histograms, and the off-diagonal panels indicate bivariate scatterplots. Top: Raw 50-member ECWMF ensemble. Bottom: Above the diagonal, statistical samples of size 50 from independently BMA postprocessed marginal predictive distributions. Below the diagonal, the samples have been subject to ensemble copula coupling (ECC).

If the dimension is large and no specific structure can be exploited, it is necessary to resort to nonparametric approaches and adopt dependence structures from records of historical weather observations (or from the ensemble forecast at hand) as embodied in so-called empirical copulas. One technique, called the Schaake shuffle, borrows the rank order structure from suitably chosen past weather records, whereas the ensemble copula coupling (ECC) approach draws on rank dependence information supplied by the ensemble forecast at hand. The ECC technique then generates a postprocessed ensemble forecast with the same number of members and the same rank order structure as the raw ensemble, as illustrated in *Figure 10*.

Roman Schefzik has developed these approaches in various directions. One of his significant developments is the SimSchaake approach, which is a specific implementation of the Schaake shuffle [Schefzik, 2016a]. In this method, the dependence structure is adopted from past weather observations, as in the original Schaake shuffle, but is now restricted to observations that correspond to dates during which the ensemble forecast resembled the forecast at hand based on carefully crafted similarity measures. The SimSchaake technique thus exploits both observational and raw ensemble information for an informed choice of an empirical copula and multivariate dependence pattern.

Furthermore, Roman Schefzik has combined parametric multivariate postprocessing methods with empirical copula techniques, aiming to exploit the strengths of either notion [Schefzik, 2016b]. This technique can be viewed as an extension of tools, such as ECC and the Schaake shuffle, to settings in which the postprocessed predictive distributions are not necessarily univariate.

Lastly, Roman Schefzik has compared ECC to so-called member-by-member postprocessing (MBMP) methods. The two seemingly unrelated concepts can be unified in that MBMP can be considered a special case within the general framework of ECC [Schefzik, R. (2017), Ensemble calibration with preserved correlations: Unifying and comparing ensemble copula coupling and member-by-member postprocessing, Quarterly Journal of the Royal Meteorological Society, 143, 999-1008].

Forecaster's dilemma: Forecast evaluation and extreme events

Public discussions of forecasts often focus attention on the predictive performance in cases of extreme events, particularly if forecasters have failed to predict an event with high economic or societal impact. *Figure 11* gives examples from newspapers, magazines, and broadcasting corporations that demonstrate the focus on extreme events. The international financial crisis of 2007/08 and the L'Aquila earthquake of 2009 are two noteworthy examples. After the financial crisis, a great deal of attention was paid to economists who had correctly predicted the crisis, and a superior predictive ability was attributed to them. In 2011, against the protest of many scientists around the world, a group of Italian seismologists was put on trial for not warning the public of the devastating L'Aquila earthquake of 2009 that caused 309 deaths. Six scientists and a government official were found guilty of involuntary manslaughter in October 2012 and sentenced to six years each in prison. In November 2015, the scientists were acquitted by the Supreme Court in Rome, whereas the sentence of the deputy head of Italy's civil protection department, which had been reduced to two years in 2014, was upheld.

Year	Headline	Source
2013	Forecast failure: How the Met Office lost touch with reality	The Spectator
2015	National Weather Service head takes blame for forecast failures	NBC News
2015	Blizzard 2015: How did the forecasts get it so wrong?	CBS NEWS
2016	Why isn't the U.S. better at predicting extreme weather?	New York Times
2014	An astonishing record – of complete failure	Financial Times
2017	Chief economist of Bank of England admits errors in Brexit forecasting	The Guardian
2017	I was one of the only economists who predicted the financial crash of 2008 – in 2017 we need to make urgent changes	The Independent
2014	Italian scientists cleared of failing to predict L'Aquila earthquake	The Telegraph

Figure 11: Media coverage illustrating the focus on extreme events in public discussions of the quality of forecasts in areas ranging from weather prediction (upper block) to economic forecasts (middle block) and earthquake prediction (final entry).

At first sight, the practice of selecting extreme observations while discarding non-extreme ones and proceeding using standard evaluation tools appears to be natural. Intuitively, accurate predictions on the subset of extreme observations may suggest superior predictive ability. However, the restriction of the evaluation to selected subsets of the available observations has undesirable effects that may discredit even the most skillful forecast available. In a nutshell, if forecast evaluation proceeds conditionally on a catastrophic event's having been observed, predicting a calamity every time becomes a worthwhile strategy. Given that media attention tends to focus on extreme events, skillful forecasts are bound to fail in the public eye, and it becomes tempting to base decision making on misguided inferential procedures. We refer to this critical issue as the "forecaster's dilemma."

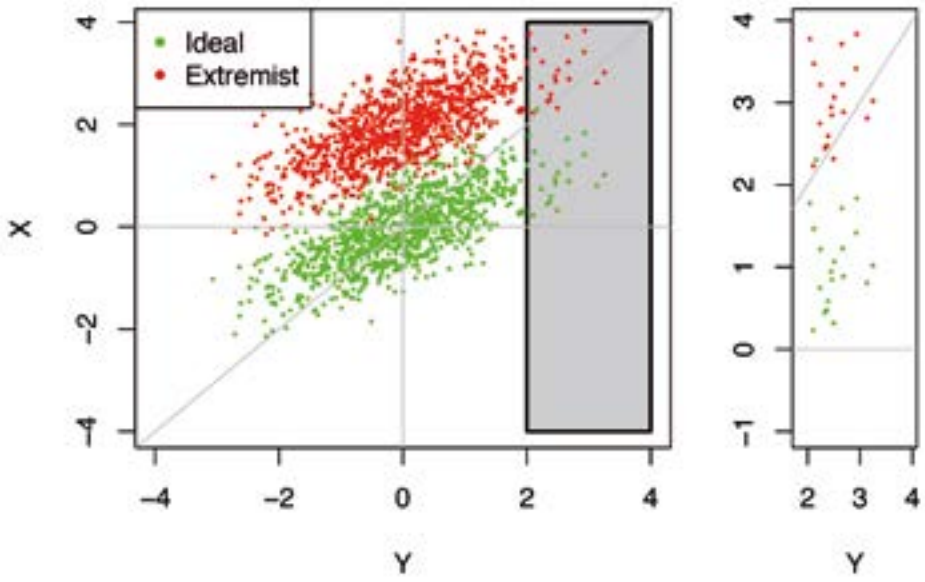


Figure 12: Illustration of the joint distribution of the ideal point forecast and the deliberately misspecified extremist forecast (x-values) and the observation (y-value). The vertical stripe, which is enlarged at the right, corresponds to cases for which the observation exceeds the threshold of 2.

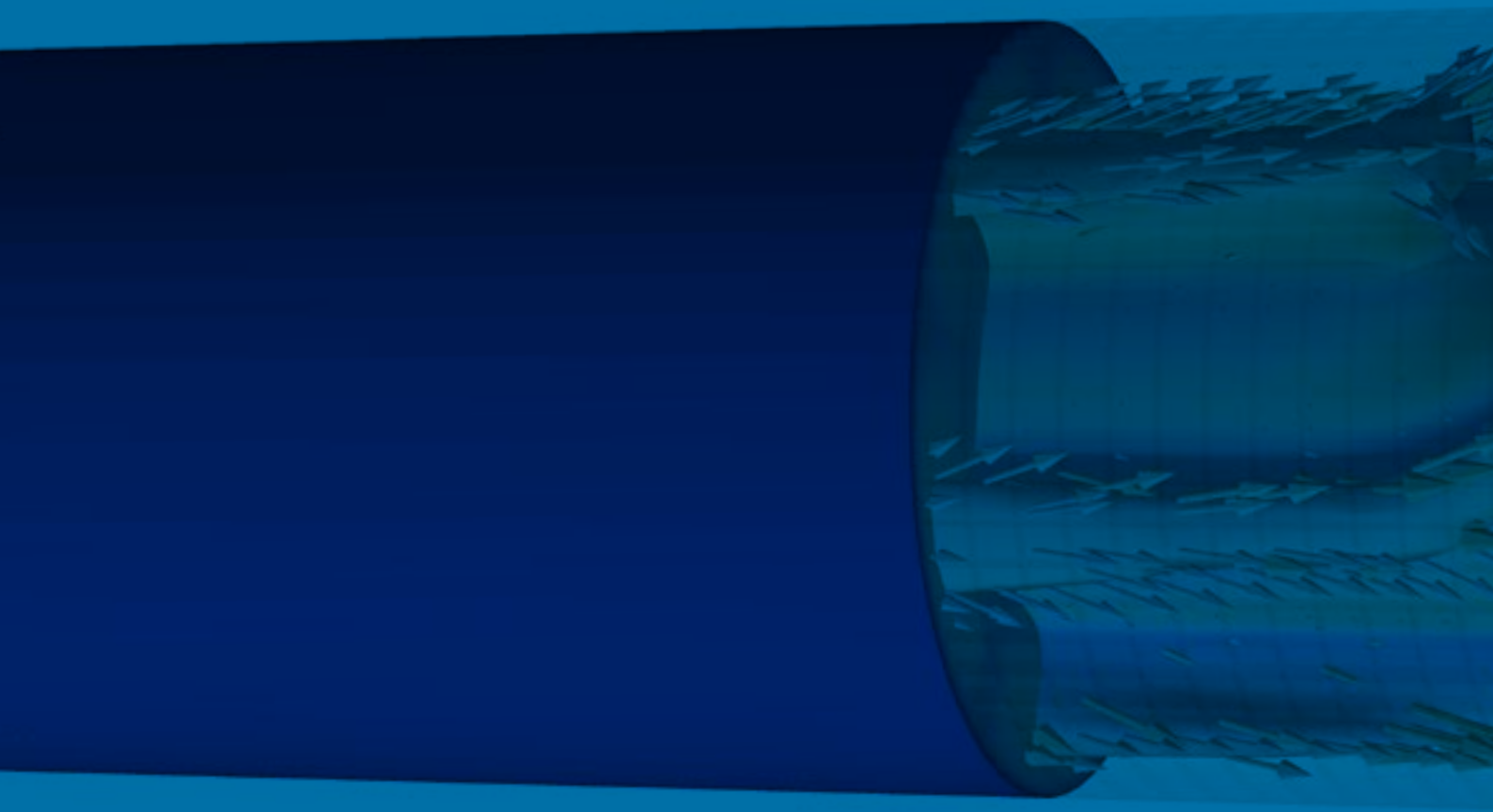
To demonstrate the phenomenon, consider point forecasts for observations that are randomly drawn from a normal distribution with an unknown mean value. The ideal forecaster has magical knowledge of the unknown mean value and uses it as the (theoretically optimal) point forecast, whereas the deliberately misspecified extremist forecaster shows a constant, systematic bias.

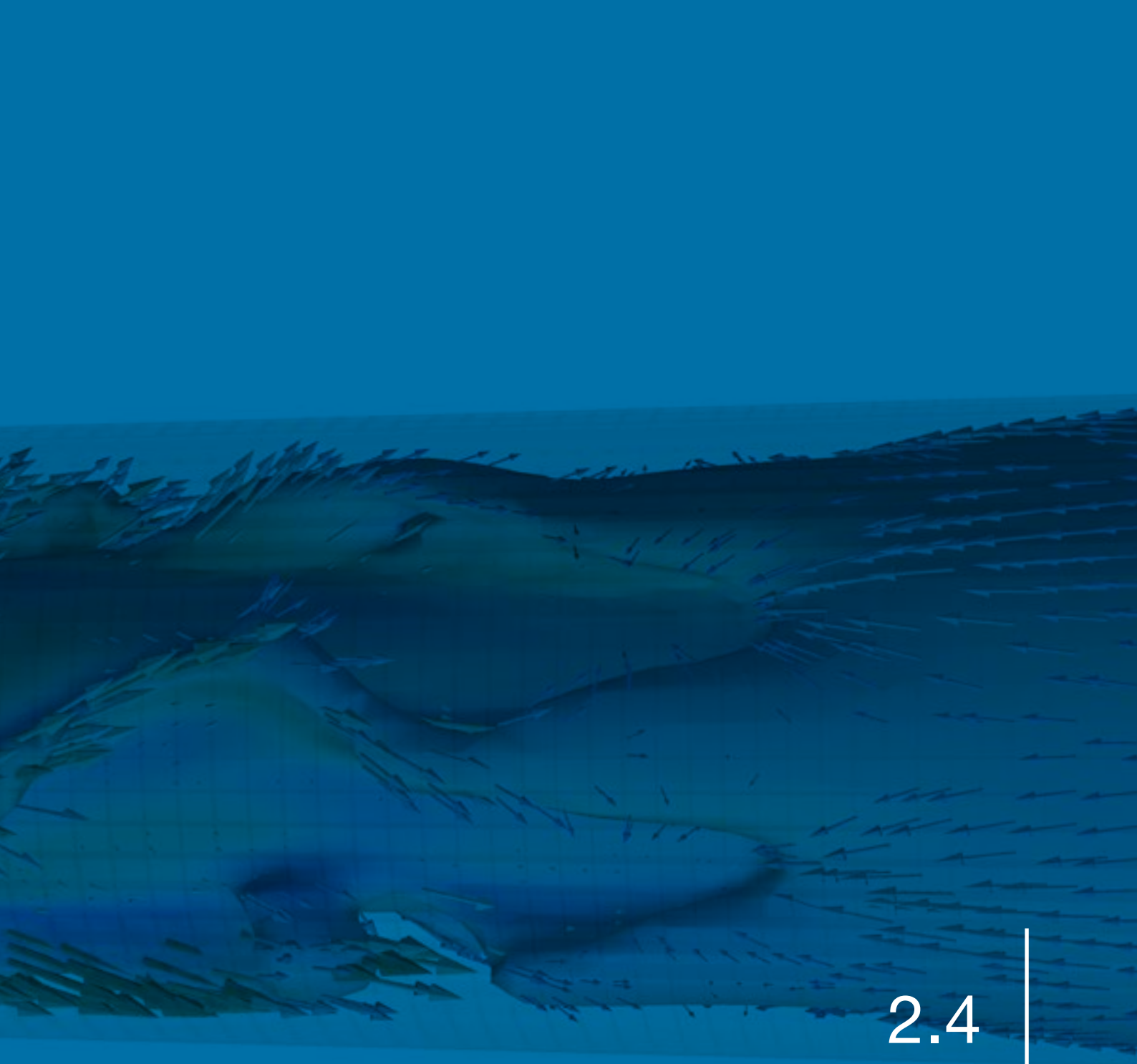
Figure 12 illustrates the joint distribution of the corresponding point forecasts and observations in this setting. The bias of the extremist forecast is readily seen when all forecast cases are taken into account. However, if we restrict our attention to cases for which the observation exceeds a high threshold of 2, it is not obvious whether the ideal or the extremist forecast is preferable. In this simple example, we have seen that if we stratify by the value of the realized observation, a deliberately misspecified forecast may appear appealing, while an ideal forecast may appear flawed.

In the case of probabilistic forecasts that take the form of predictive distributions, extant methods of forecast evaluation can be adapted to place emphasis on extremes in theoretical principled ways. In particular, suitably weighted proper scoring rules allow for the comparative evaluation of probabilistic forecasts with emphasis on extreme events and also facilitate interpretation. These tools share many of the desirable properties of standard evaluation methods and are currently being tested for adoption by national and international weather services, among many other uses.

In a recent expository article, Sebastian Lerch, Tilmann Gneiting, and their coauthors analyze and explain the forecaster's dilemma along with potential remedies using theoretical arguments, simulation experiments, and a real data study on forecasts of U.S. inflation and gross domestic product growth [Lerch, S., Thorarinsdottir, T. L., Ravazzolo, F., Gneiting, T. (2017), Forecaster's dilemma: Extreme events and forecast evaluation, Statistical Science, 32, 106-127].

2 Research





2.4 Data Mining and Uncertainty Quantification (DMQ)

2.4
Data Mining and Uncertainty Quantification (DMQ)

The Data Mining and Uncertainty Quantification (DMQ) group, headed by Prof. Dr. Vincent Heuveline, began its research in May 2013. The group works in close cooperation with the Engineering Mathematics and Computing Lab (EMCL) at the Interdisciplinary Center for Scientific Computing (IWR) at Heidelberg University, which is also headed by Prof. Heuveline.

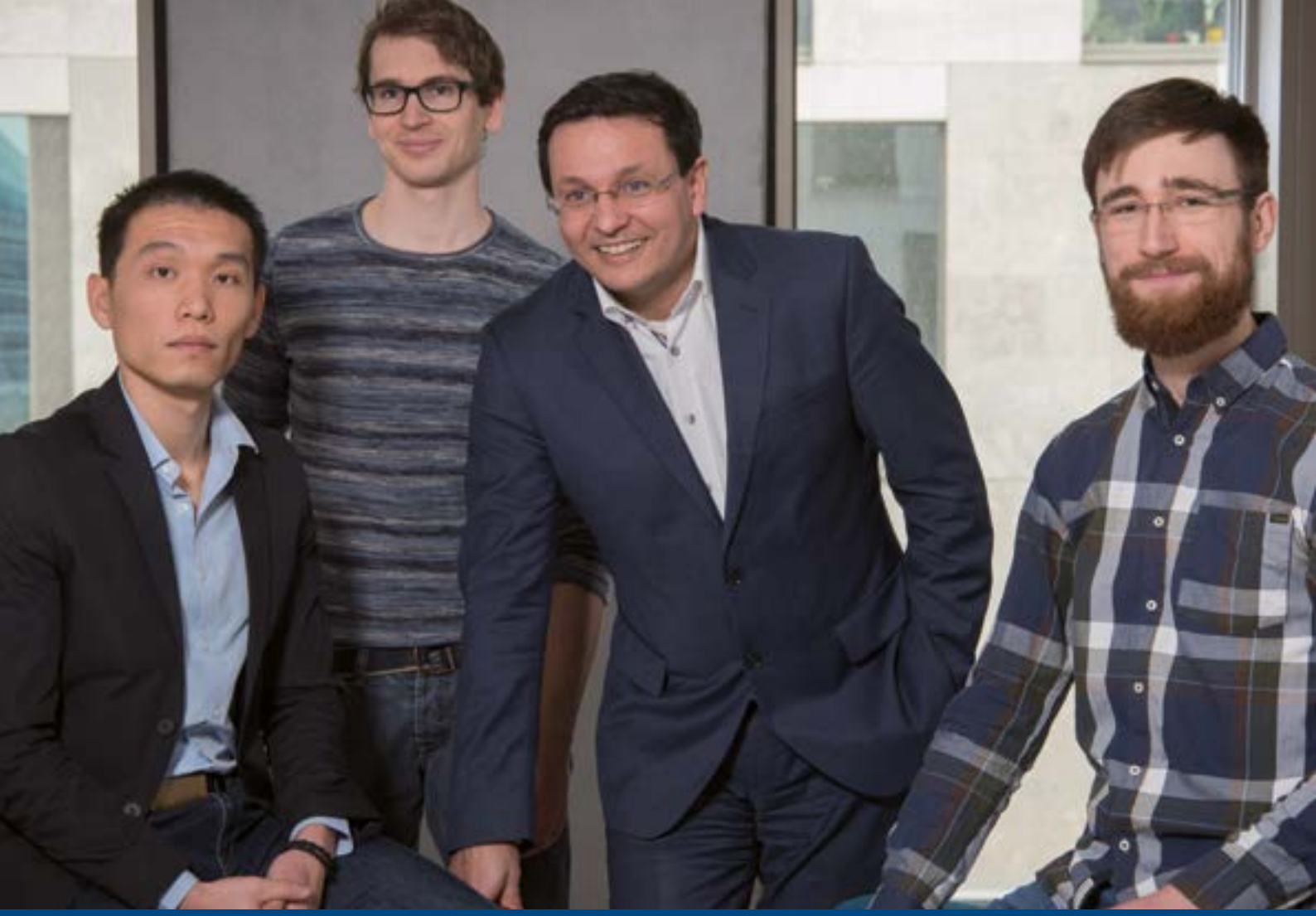
DMQ's research focus lies on gaining knowledge discovery from extremely large and complex datasets through data mining technologies. Reliability considerations with respect to these datasets are addressed by uncertainty quantification methods. Both fields – data mining and uncertainty quantification – require a decidedly interdisciplinary approach to mathematical modeling, numerical simulation, hardware-aware computing, high-performance computing, and scientific visualization.

In 2016, DMQ focused on research activities in the following areas: Uncertainty Quantification in medical engineering, "Biomedisa" (an app for large image data), algorithms for data assimilation, and High Performance Computing for Complex Computational Fluid Dynamics (CFD) problems.

Die Forschungsgruppe „Data Mining and Uncertainty Quantification" unter der Leitung von Prof. Dr. Vincent Heuveline besteht seit Mai 2013. Sie arbeitet eng mit dem „Engineering Mathematics and Computing Lab" am Interdisziplinären Zentrum für Wissenschaftliches Rechnen der Universität Heidelberg zusammen, welches auch von Prof. Heuveline geleitet wird.

Im Fokus der Forschungsarbeit steht ein zuverlässiger und strukturierter Wissensgewinn aus großen, komplexen Datensätzen, der mittels Data-Mining Technologien erreicht und mit Methoden der Uncertainty Quantification validiert wird. Beide Themenfelder – Data Mining und Uncertainty Quantification – erfordern Interdisziplinarität in den Bereichen mathematische Modellierung, numerische Simulation, hardwarenahe Programmierung, Hochleistungsrechnen und wissenschaftliche Visualisierung.

In 2016 wurde dazu in der Gruppe in folgenden Anwendungsbereichen gearbeitet: Uncertainty Quantification in der Medizintechnik, „Biomedisa" – Eine App für große Bild-Daten, Algorithmen für Daten-Assimilation und Hochleistungsrechnen für numerische Strömungssimulationen.



Group Leader

Prof. Dr. Vincent Heuveline

Scholarship holders

Michael Bromberger
(HITS Scholarship – until March 2016)
Chen Song *(HITS Scholarship)*

Staff members

Philipp Gerstner
Dr. Maximilian Hoecker *(until June 2016)*
Fabian Kißler *(since July 2016)*
Dr. Peter Zaspel

Visiting Scientist

Dr. Stefanie Speidel *(until September 2016)*

Uncertainty Quantification (UQ) of a blood pump

Cardiovascular diseases (CVDs) are group of disorders of the heart and blood vessels and are the leading cause of death globally, resulting in more than 17.5 million deaths in 2014. Heart failure is one of most severe complications of CVDs, and the problem with most surgical behavior is that the heart muscle is not capable of pumping sufficient blood and oxygen to organs. Despite the fact the modern transplant techniques are already very advanced, the lack of heart donors and missing knowledge on how to prolong a patient's life during the waiting period are becoming more challenging. The blood pump could therefore play an important role in saving lives.

The implantable supporting heart systems – the blood pumps – are an effective temporary solution for patients suffering from heart failure, especially in the case of long-time waiting periods for transplantation, which is when the blood pump's performance is most crucial. Ventricular Assist Devices (VADs) have been widely used for the last thirty years and are essentially a mechanical pump that can support the function of the left, right, or both heart ventricles. Moreover, the centrifugal pump (the one we study in our project) demonstrates stable performance in different situations.

The Finite Element Method (FEM) is an advanced numerical technique used to study complex mechanical devices. For our blood pump application, it provides us with information about wall shear stress, the velocity of blood flow, pressure, etc. We could thus use this knowledge to improve the functionality of the blood pump device. In addition to modeling the rotating process and blood flow within the pump, the real-world applications of the pump are plagued by uncertainties. Accordingly, UQ provides a framework for characterizing uncertainties of simulations and is also an important tool in examining the variability and probabilistic behavior of engineering systems. By considering the stochastic model, we are able to gain more information, which thus improves the reliability of the machine.

In our work, we incorporate the stochastic Galerkin projection technique and incompressible Navier-Stokes in rotating system to simulate the propagation of uncertainties in a blood pump. The simulation is achieved via our open source Finite Element library Hiflow3. Three uncertainties (i.e., the inflow boundary condition, dynamic viscosity, and rotating angular speed) are considered in this work. Spatial, temporal, and stochastic spaces are then coupled using the intrusive UQ technique (stochastic Galerkin projection) since an efficient solving method and strategy is indeed required for dealing with a large coupled system (e.g., *Figures 13* and *14* display a numerical result of a system with 40 million unknowns). In this model, we apply a multi-level preconditioning technique to overcome this challenge.

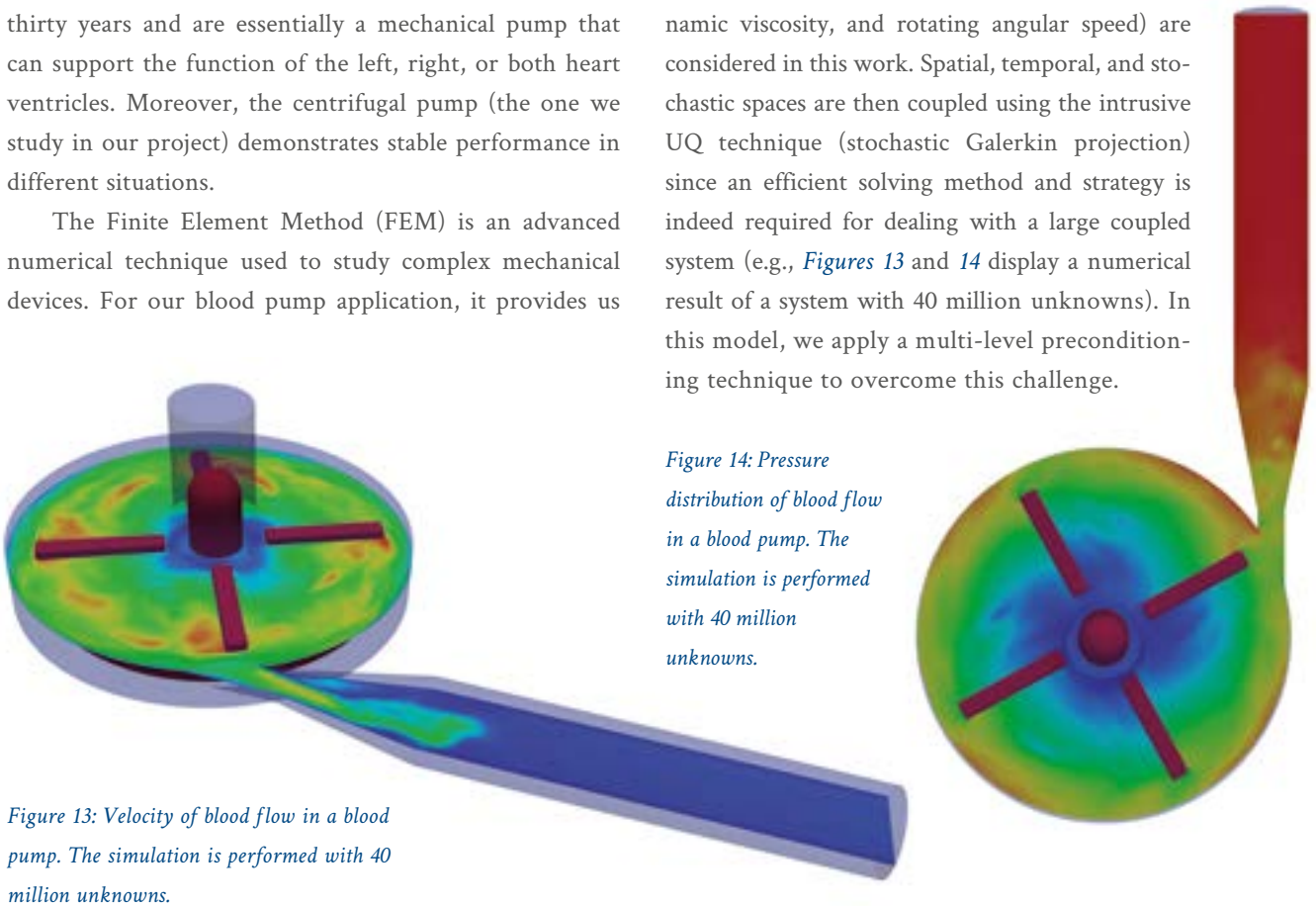


Figure 13: Velocity of blood flow in a blood pump. The simulation is performed with 40 million unknowns.

Figure 14: Pressure distribution of blood flow in a blood pump. The simulation is performed with 40 million unknowns.

Biomedisa – The biomedical image segmentation app

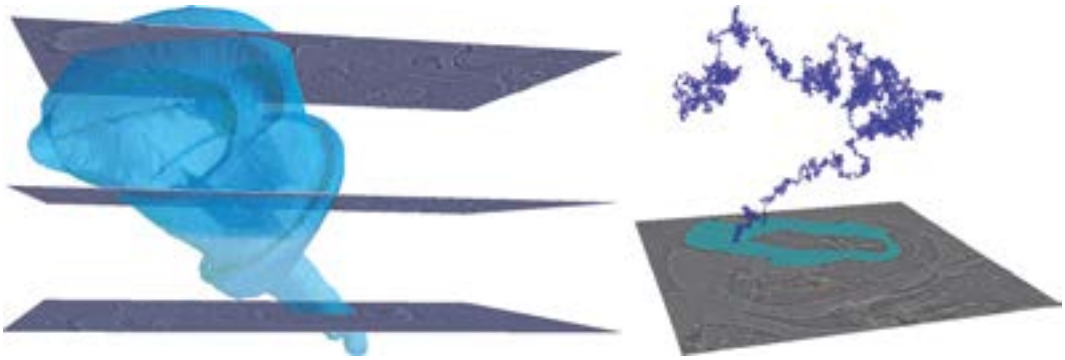


Figure 15: Left Final segmentation of a beetle's hip joint based on three pre-labeled reference slices. The right weighted random walk begins in a reference slice.

X-ray microtomography (μCT) has established itself as a tool for the investigation of morphological questions. However, with the analysis of the data and their ever-increasing volume, new problems are arising. The segmentation of tomographic images remains one of the most challenging tasks in computer vision. In medicine, this segmentation is important for simulation, operation planning, and decision-making, and in biology, it is important for phylogenetic and functional analysis. Manual segmentation is still considered the "gold standard" in many cases.

The web application Biomedisa (https://biomedisa.de) was developed by the associated researcher Philipp Lösel as a semi-automatic tool for the improvement and acceleration of manual segmentation within the framework of the BMBF projects ASTOR and NOVA. It is based on a parameter-free and highly scalable diffusion algorithm. The segmentation is performed using weighted random walks, which begin at a few manually labeled reference slices (*Figure 15*) and extrapolate the information contained in these slices to the remaining volume. Over time, the voxels are hit by numerous random walks. Based on the respective number of hits, it is possible to approximatively determine the probability that a voxel can be assigned to one particular segment. Thus, segmentation can be performed by assigning each voxel to the region from which most of the hits were made. The weights are defined without the use of hyperparameters. Thus, it is not necessary to search for proper values, which eliminates the need for an elaborate and tedious configuration. Due to the random walks' independence, their calculation is highly scalable, which allows for the use of massively parallel computer architectures (e.g., graphics processors) and thus for an evaluation of large image data in a short time. In addition, a quick and easy correction of the segmentation results is possible by adding additional reference slices. This semi-automatic approach has already proven to be extremely effective within the ASTOR project. Since the input data can be created with any common segmentation software, such as Amira or ImageJ, the users can continue to work in their familiar environment. However, the segmentation process itself is significantly accelerated, and the quality of the resulting segmentation is improved. In order to provide users with the computing power needed, the web application Biomedisa was established. Users can upload their tomograms and labeled slices, run the segmentation process, and visualize both the image data and the results using 3D rendering software and a 2D slice viewer.

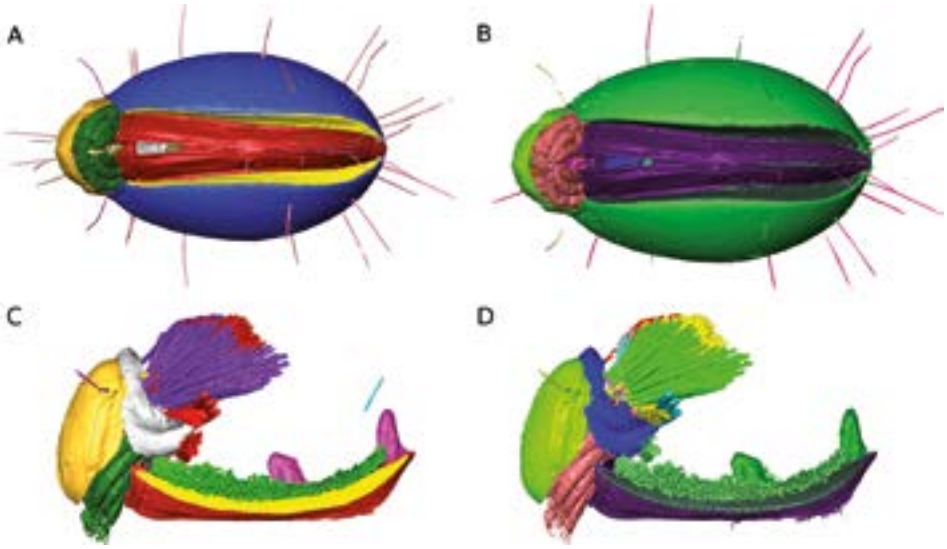


Figure 16: Comparison of manual segmentation (A, C) and automatic segmentation (B, D) based on the manual labeling of every 50th slice.

Data assimilation

The term data assimilation refers to methods for the combination of different sources of information to estimate the state of a system, such as the temperature distribution in a room. These sources of information include model equations, noisy or partial observations of the system state (e.g., temperature measurements), and a priori information (e.g., summary statistics in case of uncertain initial conditions). Historically, data assimilation has mainly been used in meteorology to improve the quality of weather forecasts. Today, other fields also use methods that were developed in the context of data assimilation: In medical applications, data assimilation algorithms are used to personalize biomechanical models by means of estimating parameters that characterize the properties of soft tissues (e.g., density or stiffness). Another important application area is glaciology. Here, data assimilation methods provide an invaluable technique for inferring quantities that are not measurable, such as basal friction for grounded ice. A typical data assimilation cycle is shown in *Figure 17*. Here, we assume that we have a dynamical system that models a physical process, such as the temporal evolution of the above-mentioned temperature distribution in a room. The system state (e.g., the temperature) is observed at equally spaced points in time, and the measurements are represented by the red crosses. Using the dynamical system and the initial state at time 1, we produce forecasts of the system state at times 2, 3, and 4, which is indicated by the solid line. We next aim to improve the forecast by taking the observations into account. To this end, we find an initial state at time 1 such that the forecast produced by the dynamical system is closest to the observed data at times 1 to 4. We call this improved forecast the analysis, which is indicated by the dashed line in the first assimilation interval. Using the state of the analysis at time 4 as the new initial value for the dynamical system, we produce a new forecast for the next assimilation interval from times 4 to 7. As soon as observations of the system state at these times become available, it again becomes possible to improve the forecast by finding a new initial condition at time 4 and producing a new analysis. This procedure can be repeated as soon as new observations become available. The methodology can also be used in a stochastic framework in which the initial conditions or the model are not deterministic but are instead represented by random variables or a stochastic model, respectively. The challenge then becomes finding approximations of the distribution of the system state given observation data. In the DMQ group, we aim to develop algorithms to assimilate data into complex large-scale systems modeled by stochastic partial differential equations in order to make medical applications more patient-specific and reliable.

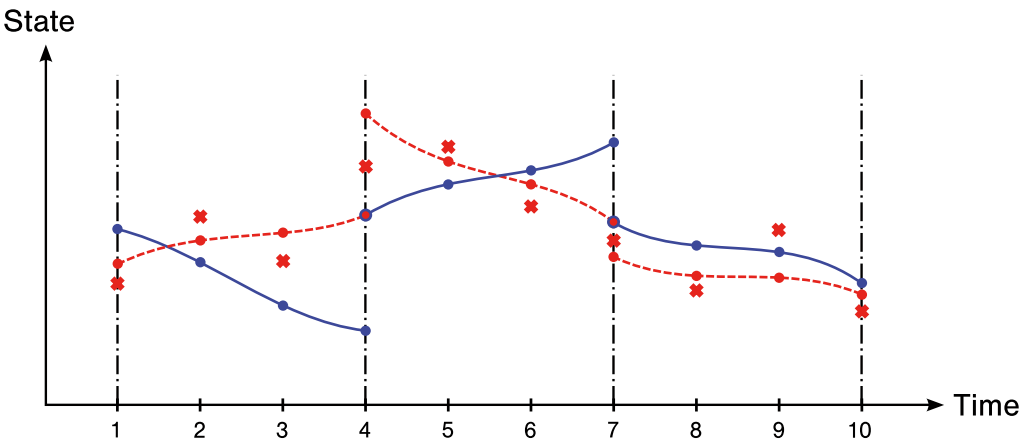


Figure 17: Typical data assimilation cycle.

High-performance computing for complex computational fluid dynamics (CFD) problems

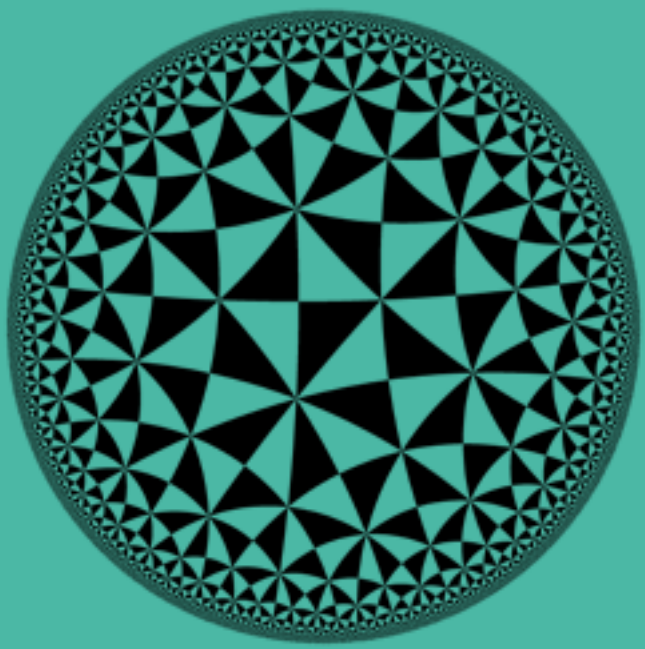
Containment in engineering often provides thermal insulation, as in heat exchanger systems, where an improvement of heat transfer via an efficient enhancement is of general interest due to its benefits of having low operational costs and maintaining a sustainable usage of energy. One possible heat transfer enhancement technique is given by applying electric fields, which is known as thermal electro-hydrodynamic (TEHD) driven heat transfer augmentation. In cooperation with our project partners at BTU Cottbus, we aim at gaining further insight into the hydrodynamical behavior of dielectric fluids in a cylindrical enclosure with applied electric field- and temperature gradients by combining numerical simulations with experimental data. To this end, we consider to a multi-physics model based on the well-known Boussinesq equations for natural convection combined with an additional electrical body force.

To obtain an approximate solution for this set of partial differential equations, we use the so-called Space-Time Finite Element Method. In so doing, it is necessary to solve linear systems of equations of a large scale, i.e., equations on the order of 10^6 to 10^7. To address this problem, we have developed a specialized and scalable method that is capable of solving these systems by making use of large computing clusters with several hundreds to thousands of CPUs. This linear solver is based on a Krylov subspace solver in combination with a preconditioner that decouples different physical aspects of the model at an algebraic level. Moreover, we study hydrodynamic behavior by considering the corresponding dual problem. The solution to this problem characterizes the sensitivity of the previously calculated simulation with respect to some given quantity of interest. For instance, it is possible to reveal non-obvious and complex correlations in the different variables, such as fluid velocity, pressure, and temperature. As a next step, our aim is to develop an a posteriori error estimator that will allow for the quantification of the discretization error in the computed solution and thus indicate the quality of the simulation.

Figure 18 illustrates the considered geometry: The fluid is contained inside the gap of two concentric cylinders. The inner wall is heated, whereas the outer wall is cooled. In addition, an electric field is applied between the two of them. Furthermore, the isosurface of some specific temperature is shown.

Figure 18: Temperature isosurface and velocity field of dielectric fluid in a cylindrical enclosure.

2 Research



2.5 Groups and Geometry (GRG)

2.5
Groups and Geometry (GRG)

The "Groups and Geometry" research group works together closely with the "Differential Geometry" research group at Heidelberg University. Both groups are headed by Prof. Dr. Anna Wienhard.

Symmetries play a central role in mathematics as well as in other natural sciences. Mathematically, symmetries are transformations of an object that leave this object unchanged. These transformations can be composed (i.e., applied one after another) and form what is called a group.

In the 19th century, mathematician Felix Klein proposed a new definition of geometry as the study of all properties of a space that are invariant under a given group of transformations. In short: Geometry is Symmetry. This concept unified classical Euclidean geometry, the then recently discovered hyperbolic geometry, and projective geometry, which has its origins in the study of perspective in art and is not based on the measurement of distances, but rather on incidence relations. Even more importantly, Felix Klein's concept fundamentally changed our view of geometry in mathematics and theoretical physics and continues to influence it today.

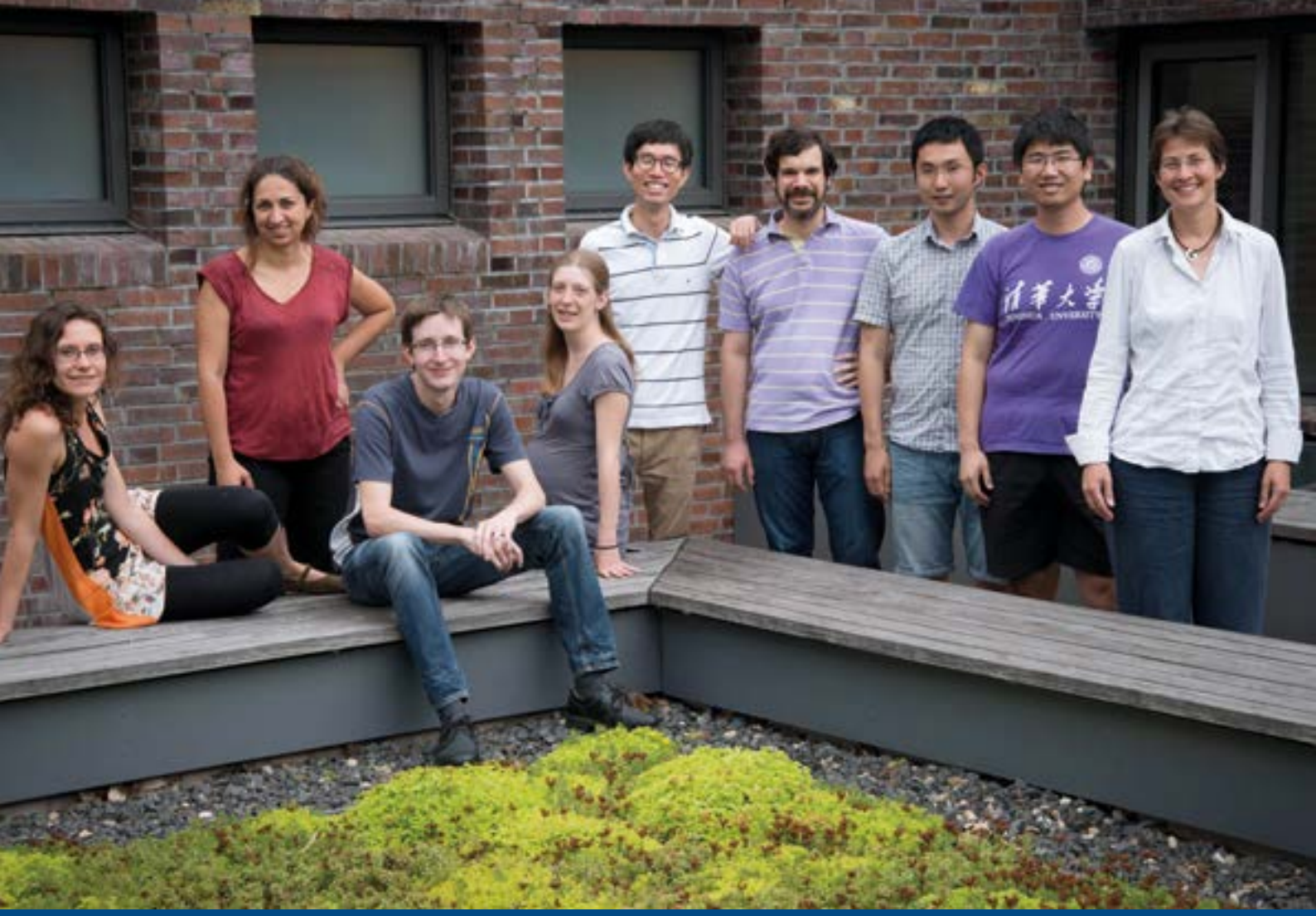
Our research group investigates various mathematical problems in the fields of geometry and topology that involve the interplay between spaces, such as manifolds and metric spaces, and groups acting as symmetry groups on them. A special focus lies on investigating all possible manners by which a given group can be realized as a symmetry group of a geometric space.

Die Arbeitsgruppe „Gruppen und Geometrie" arbeitet eng mit der Arbeitsgruppe „Differentialgeometrie" an der Uni Heidelberg zusammen. Beide Gruppen werden von Prof. Dr. Anna Wienhard geleitet.

Symmetrien spielen eine zentrale Rolle sowohl in der Mathematik als auch in vielen Naturwissenschaften. In der Mathematik verstehen wir unter Symmetrien die Transformationen eines Objektes, die dieses invariant lassen. Solche Transformationen lassen sich verknüpfen, d.h. hintereinander ausführen, bilden so eine Gruppe.

Im 19. Jh. entwickelte der Mathematiker Felix Klein einen neuen Begriff der Geometrie: Geometrie ist das Studium der Eigenschaften eines Raumes, die invariant sind unter einer gegebenen Gruppe von Transformationen. Kurz gesagt: Geometrie ist Symmetrie. Mit diesem Konzept vereinheitlichte Klein die klassische Euklidische Geometrie, die damals gerade neu entdeckte hyperbolische Geometrie, sowie die projektive Geometrie, die aus dem Studium der perspektivischen Kunst erwuchs und die nicht auf dem Messen von Abständen, sondern auf Inzidenzrelationen beruht. Noch wichtiger ist, dass Felix Kleins Konzept unser Verständnis von Geometrie in der Mathematik und der theoretischen Physik grundlegend verändert hat und bis heute prägt.

Unsere Arbeitsgruppe beschäftigt sich mit verschiedenen mathematischen Forschungsfragen aus dem Gebiet der Geometrie und Topologie, in denen das Zusammenspiel zwischen Räumen und Gruppen, die auf diesen als Symmetriegruppen wirken, zentral ist. Ein besonderer Forschungsschwerpunkt liegt darauf, für eine gegebene Gruppe alle möglichen Realisierungen dieser Gruppe als Symmetriegruppe eines geometrischen Raumes zu untersuchen.



Group Leader:

Prof. Dr. Anna Wienhard

Staff members:

Dr. Ana Peón-Nieto *(until December 2016)*
Florian Stecker

Visiting scientists:

Dr. Daniele Alessandrini
Dr. Shinpei Baba
Dr. Federica Fanoni *(since October 2016)*
Dr. Sourav Ghosh
Johannes Horn *(since October 2016)*
Christoph Karg *(since October 2016)*
Dr. Gye-Seon Lee
Dr. Andreas Ott
Evgenii Rogozinnikov *(since October 2016)*
Dr. Andy Sanders *(since October 2016)*
Anna Schilling
Nicolaus Treib

Students:

Tim Adler *(since November 2016)*
Sven Grützmacher *(since November 2016)*

Geometries

Our research deals with different geometries in various dimensions. Most of these geometries appear rather exotic to the layperson. However, even the layperson is quite used to Euclidean geometry in the plane (or in three-space), where we can measure lengths and angles and compute the area of geometric shapes, such as triangles and more general polygons. On the other hand, most of us are not very familiar with other types of geometries, for which lengths and angles no longer have meaning. One classical example of such a geometry is projective geometry, which is based on incidence relations. In the projective plane, the only relation between geometric objects (e.g., a point and a line) is incidence: Either the point is on the line, or it is not on the line. Projective geometry is related to perspective art and drawings, and it is one of the geometric structures we investigate. There are many more exotic geometries that we usually do not encounter in our daily lives. They arise in high-dimensional spaces, and some of them play an important role in theoretical physics.

Even if we stick to two-dimensional geometries in which we can measure length and angles and compute the area of geometric shapes, Euclidean geometry is not the only option. There are, in fact, two other model geometries: spherical geometry, which is geometry on the surface of a ball, and hyperbolic geometry, which is geometry on a surface called a "hyperboloid".

Spherical geometry is of course relevant to everyone since we all live on the surface of a round Earth, but hyperbolic geometry is also surprisingly common, for we encounter it in our daily lives even though we usually do not realize it.

In order to illustrate the different two-dimensional geometries (and particularly the lesser-known hyperbolic geometry) to the interested general public, we purchased several 3D-printed models with which to illustrate interesting features of hyperbolic geometry and that allow the learner to hold a "piece of the hyperbolic plane" in his / her hands. We feature some of these objects in this report while also describing the corresponding features of the hyperbolic plane.

Tilings from reflections

Some of the differences between Euclidean, spherical, and hyperbolic geometry can be illustrated by looking at tilings.

If we take a square in the Euclidean plane and reflect it along its sides and then reflect the resulting copies again and keep repeating this, the squares we obtain never overlap (unless they are the same), and they eventually fill out the entire plane (see *Figure 19*). This process is known as a tiling of the plane by squares. Of course, this procedure also works with other polygons. For example, we can use an equilateral triangle. If all the side lengths of the triangle are equal, all its angles are also equal, and since the sum of the inner angles of a triangle (in the Euclidean plane) is 180°, each angle is 60°. If we take such a triangle and reflect it along its sides, they again do not overlap and fill out the entire plane. We thus obtain a tiling of the plane by equilateral triangles. If we take six of these triangles together around a vertex, they form a regular hexagon (i.e., a hexagon in which all the side lengths are equal). The tiling by equilateral triangles also creates a tiling of the Euclidean plane by equilateral hexagons. We will return to this tiling later.

Can this procedure of tiling the Euclidean plane also work with other polygons? Yes, but there are certain conditions we have to satisfy: At every corner, the interior angles of the

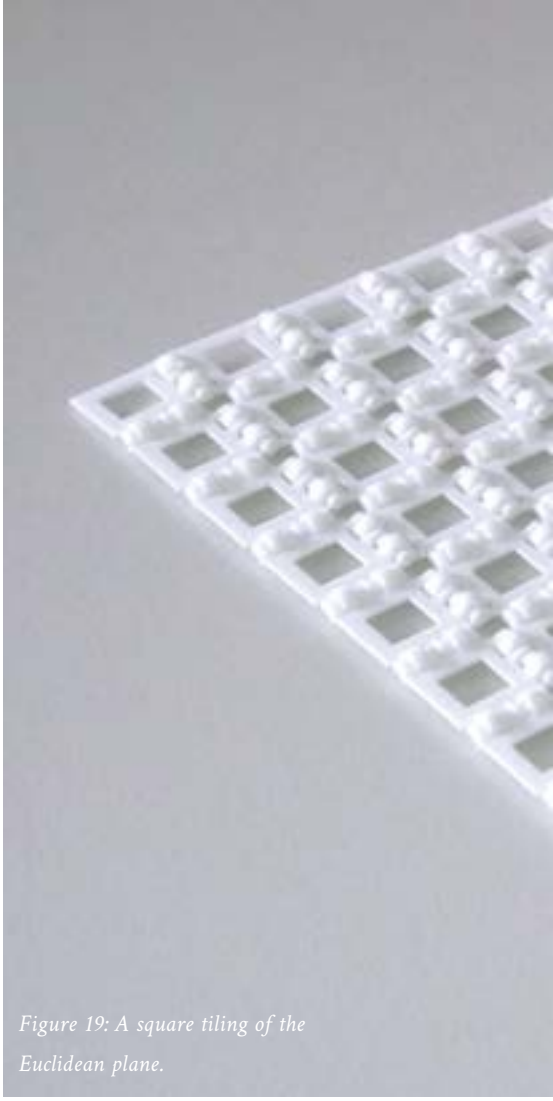


Figure 19: A square tiling of the Euclidean plane.

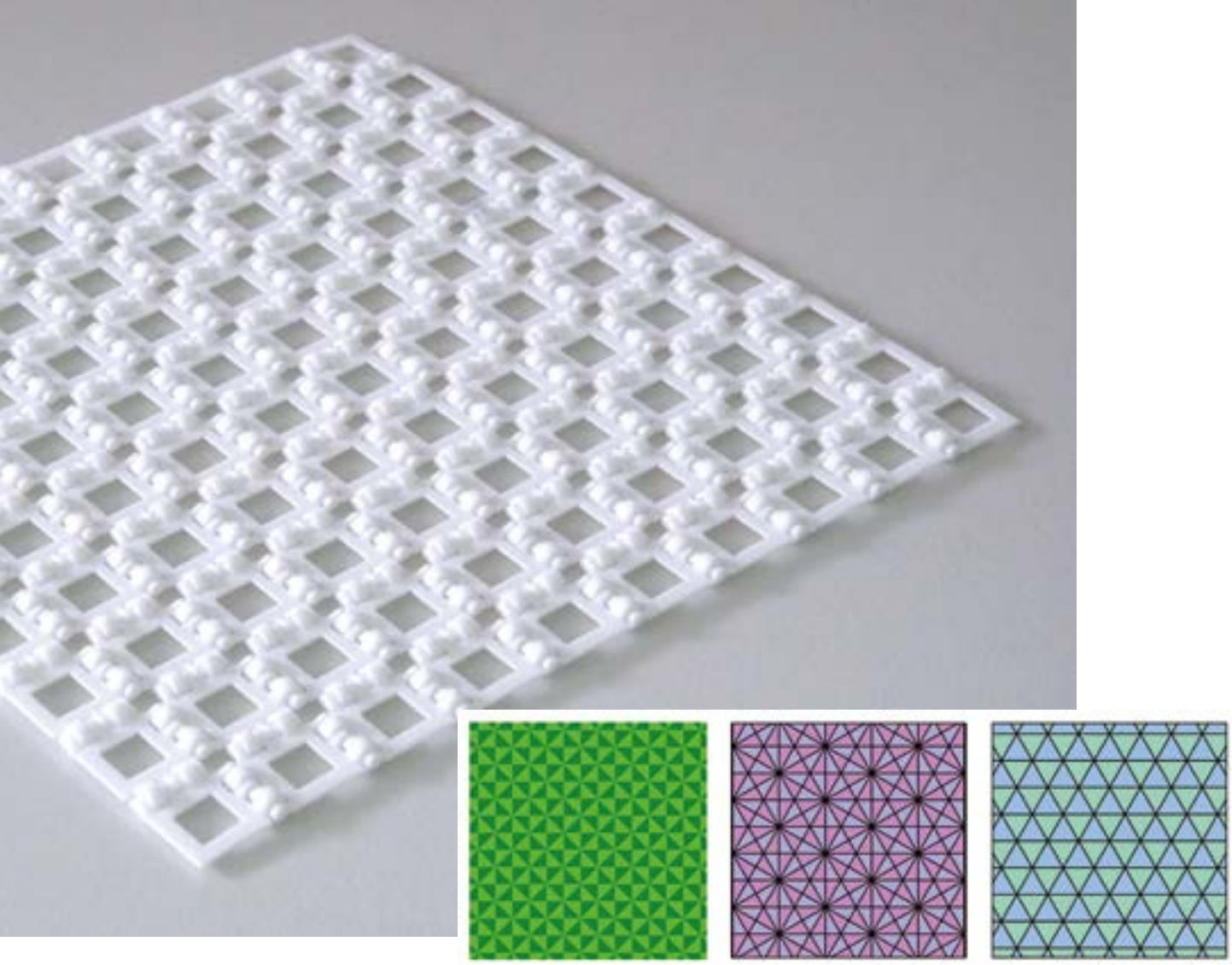


Figure 20: Euclidian plane (f.l.t.r: Tom Ruen, Public Domain; R.A. Nonenmacher (2), GNU Free Documentation License).

meeting polygons have to add up to 360°. Thus, if we try to tile the Euclidean plane with a regular polygon (i.e., one for which all the side lengths and all the angles are equal), the interior angle of our regular polygon has to be an integer divisor of 360°. A short calculation reveals that this is only possible for triangles, squares, and hexagons, but not, for example, for pentagons. This fact explains why five-fold symmetries cannot be found in crystals, but only in quasi-crystals. Dan Schechtman received the Nobel Prize in Chemistry in 2011 for his discovery of the quasi-crystal. Mathematically, quasi-crystals are related to so-called a-periodic tilings. A-periodic tilings with five-fold symmetry were discovered by Sir Richard Penrose in the 1970s.

What would happen if we glued together these polygons that do not tile the Euclidean plane? This is not possible on the flat Euclidean plane because if the sum of the interior angles around a corner is less than 360°, the plane begins to bend together. If the sum of the interior angles around a corner is greater than 360°, the plane begins to crumple. This fact can be nicely illustrated with the soccer-ball pattern.

Figure 21: A negatively curved plane out of equilateral triangles.

Euclidean, spherical, and hyperbolic soccer-ball patterns

Figure 22 displays the honeycomb-like tiling of the Euclidean plane by regular hexagons, where some hexagons are colored black and the others white. There are always six white hexagons around each black hexagon. Here, the angle sum around every corner is $3 \cdot 120° = 360°$, so the pattern fits nicely into the Euclidean plane. However, if we take out the black hexagons, replace them with pentagons, and place five hexagons around each black pentagon, we create the usual pattern found on a soccer ball. What happens to the angles? The angle sum around the corner drops to $108° + 120° + 120° = 348°$, and the plane curves into a spherical shape. This is the pattern of leather panels that has commonly been used for the production of soccer balls since the 1970s. We thus obtain a combinatorial model of the sphere, namely the surface of a (soccer) ball, i.e., a surface with positive curvature. If we take out the black hexagons, replace them with black heptagons, and place seven white hexagons around each black heptagon, the angle sum around a corner is approximately $129° + 120° + 120° = 369°$. This is greater than the angle around a point in the flat Euclidean plane, which causes the plane to crumple up into a structure resembling the form of a lettuce leaf, as seen in *Figure 21*. This is a surface with negative curvature.

Figure 22: Euclidean (copyright: 123 freevectors), spherical (Alex Y), and hyperbolic (Frank Sottile).

Area growth

The different behavior of the Euclidean, spherical, and hyperbolic soccer-ball patterns manifests itself in the growth of surface area. In each of the three soccer-ball-like patterns, we can see a circle around one of the corners with a radius equal to the side length of a hexagon. Let's say that this side length is a fixed constant l. In the Euclidean pattern, the area of the interior of this circle is $\pi \cdot l^2$. In the spherical case of the real soccer-ball pattern with pentagons, because the radius runs around the vertex, it amounts to an angle of only 348°, and the circle contains an area of only $^{348}/_{360} \cdot \pi \cdot l^2$. This leads to a surface that curves "inwards" in an effort to minimize the area consumption.

Figure 23: (2, 3, 7) tiling on a hemisphere.

However, for the hyperbolic soccer-ball pattern in which we use heptagons, the area of the circle increases to $^{369}/_{360} \cdot \pi \cdot l^2$. While this seems to be a small change, it results at every single vertex, thereby leading to an exponential growth of the area as the radius of the circle grows larger and larger. Because the area increases so rapidly, it cannot find enough space in the flat plane to fill and instead folds up into the crumpled shape seen in *Figure 21*.

Geometry on the hyperbolic plane

Just as we can study Euclidean geometry on the flat plane, we can also study geometry on the sphere or the hyperbolic plane, though the laws are a bit different for each approach. For example, in the Euclidean plane, the sum of the interior angles of a triangle is always 180°. On the sphere, the precise value of the sum of the interior angles depends on the triangle but is always greater than 180°. On the hyperbolic plane, however, the sum of the interior angles of every triangle is less than 180°. Even more surprising, the area of the triangle on the hyperbolic plane is determined by the triangle's angles and is proportional to 180° minus the sum of the angles.

To imagine and study geometry on the sphere (the surface of a ball), we can realize it in the three-dimensional Euclidean space and study its properties there. Unfortunately, the hyperbolic plane curves more than the 3-dimensional Euclidean space can support, so it cannot exist as an embedded surface in our 3-dimensional world, but only as an abstract object. The crumpled shape (*Figure 21*) provides a tangible model of a finite piece of the hyperbolic plane, but to understand the entire hyperbolic plane and its geometry, mathematicians have developed various smooth models for the hyperbolic plane.

In order to visualize the hyperbolic plane, different types of projections onto the Euclidean plane can be used that capture some features very well while distorting others. The same problem arises when creating an atlas of the Earth. To create an atlas, we project pieces of the surface of the Earth onto the Euclidean plane, but in order to do this, we always distort some geometric features (e.g., distances and lengths). Two of these projections of the hyperbolic plane are named after the famous mathematician Henri Poincaré: the "Poincaré half-space model" and the "Poincaré disk model". Both of these projections can be obtained from a 3D model (*Figure 23*) of the hyperbolic plane as the surface of a hemisphere for which the equator is infinitely far away from every point on the hemisphere.

Using a flashlight to illuminate the 3D model from different directions, the shadow in one case creates the Poincaré disc model (when the flashlight is above the pole), and in another case, it creates the Poincaré half-space model (when the light source is at the equator).

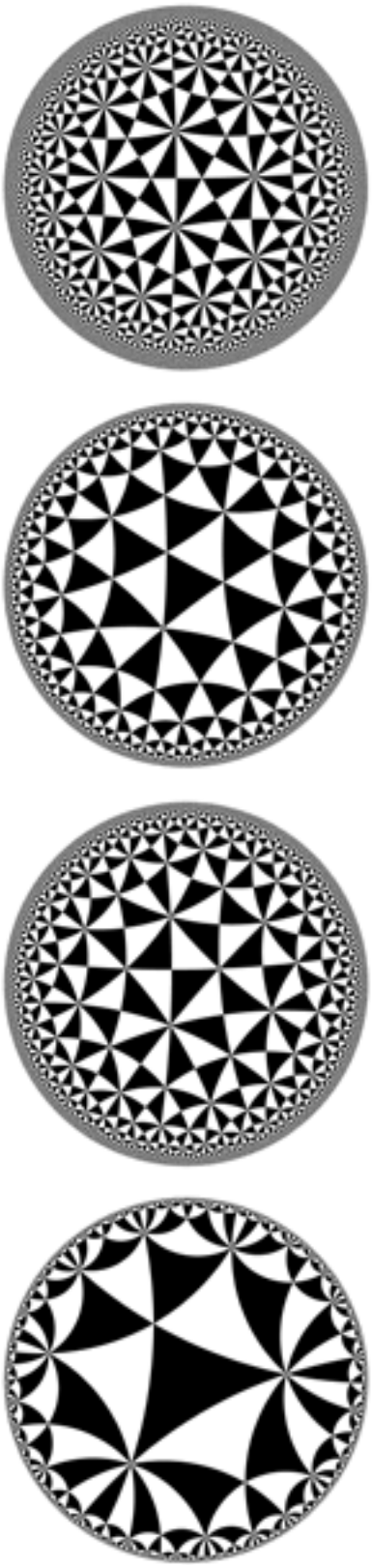


Figure 24: Triangular tilings of the hyperbolic plane (pictures: Anton Sherwood, Wikimedia commons).

Tilings of the hyperbolic plane

In the same way that we filled out the Euclidean plane by reflecting a triangle, we can generate tilings of the hyperbolic plane via reflections of a hyperbolic polygon, such as a triangle. As mentioned above, the angle sum in a hyperbolic triangle is less than 180°, which provides us with many more options when tiling. The model in *Figure 23* exhibits a tiling via a triangle with interior angles of 180° / 2 = 90°, 180° / 3 = 60°, and 180° / 7 = 26° (i.e., a total of only 176°). This is the smallest hyperbolic triangle by area that enables such a tiling, but there are infinitely many other possibilities (see *Figure 24*). This fact contrasts with the Euclidean case we examined previously, where only very few triangular tilings were possible (see *Figure 20*).

Using a 3D printer to visualize mathematical objects

We have a large collection of 3D-printed objects to help us visualize hyperbolic geometry and other kinds of geometric structures. Several of these objects were produced by Henry Segerman from Oklahoma State University, and we use the objects at outreach events (such as the 2016 Open House at HITS) to provide the general public with a glimpse of the geometric structures that we investigate. As of 2016, we also have our own 3D printer to experiment with printing 3D models of mathematical objects (*Figure 25*). While it is not a high-quality printer like the ones with which the models seen in *Figures 21* and *23* were made, it allows us to directly print whatever object comes to mind. The printer operates by moving a printing head along a programmed pattern in the plane. While doing so, a filament of plastic (e.g., PLA) is fed to the printing head, where it is melted and then ejected onto a heated floor, which slowly moves downwards while the printed object develops layer by layer. Unfortunately, this 3D-printer design does not usually enable us to print objects with steeply overhanging faces or roofs since every new layer needs a supporting lower layer on which to be printed. In the future, we would like to

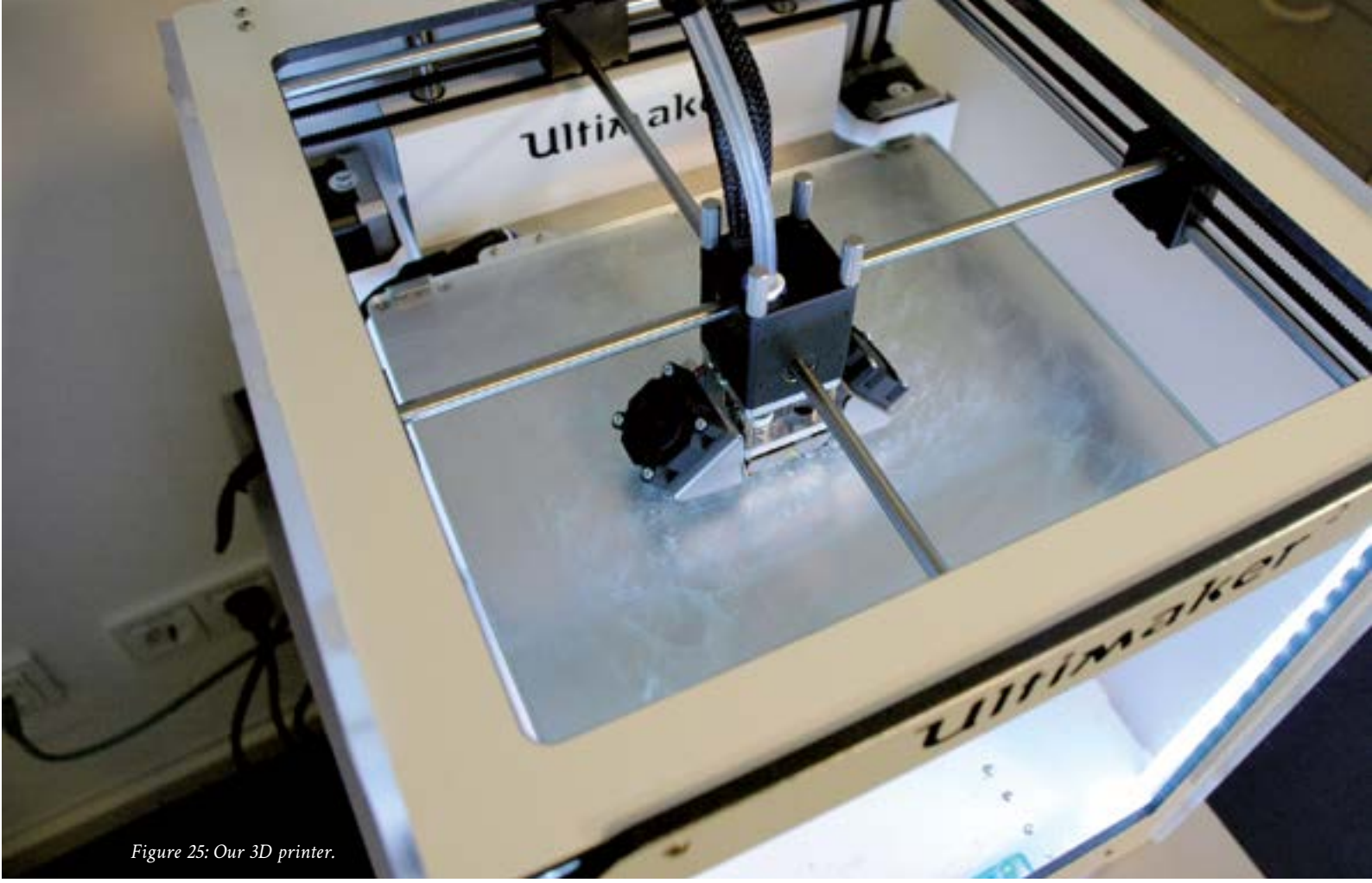


Figure 25: Our 3D printer.

make use of a second printing head to print support structures consisting of a special plastic that dissolves in water. Tim Adler and Sven Grützmacher operate our printer. One of the first models we created with it was the Menger sponge (*Figure 26*), which is a fractal object constructed by taking a cube and slicing it into 3 equal parts along every axis until it is divided into 27 smaller cubes. Then, the center cube as well as the 6 cubes that lie in the center of every face are removed, and the other small cubes remain. This procedure is repeated for each of the 20 remaining small cubes. Recursively completing these steps yields the Menger sponge, a fractal object that is neither a 2-dimensional nor a 3-dimensional object but has a fractal dimension in between 2 and 3 dimensions. In the future, we hope to use the printer to visualize aspects of more exotic geometric structures (e.g., projective structures) and in working with students. Of course, when using the 3D printer, we are bound to geometric objects and shapes that can be visualized in three dimensions. To study higher-dimensional and more exotic geometric structures, we have to rely on our imagination.

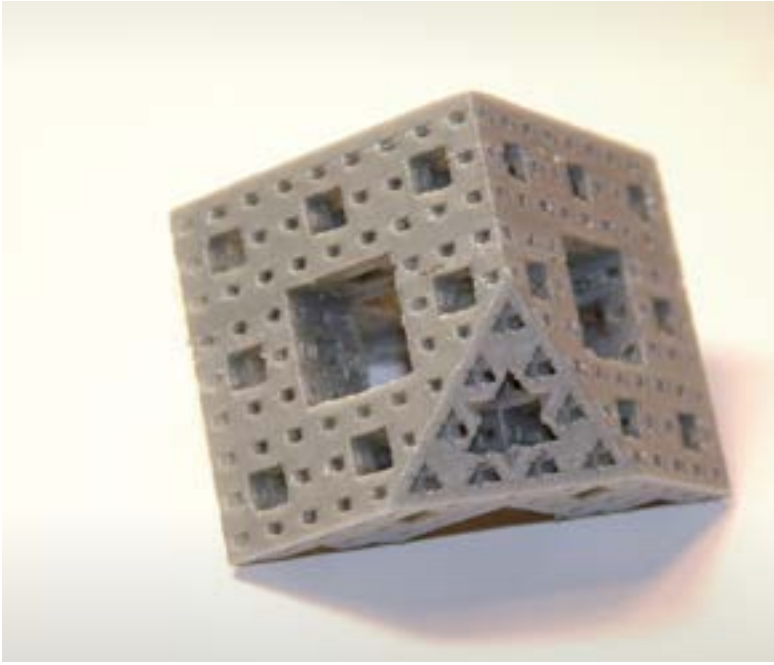
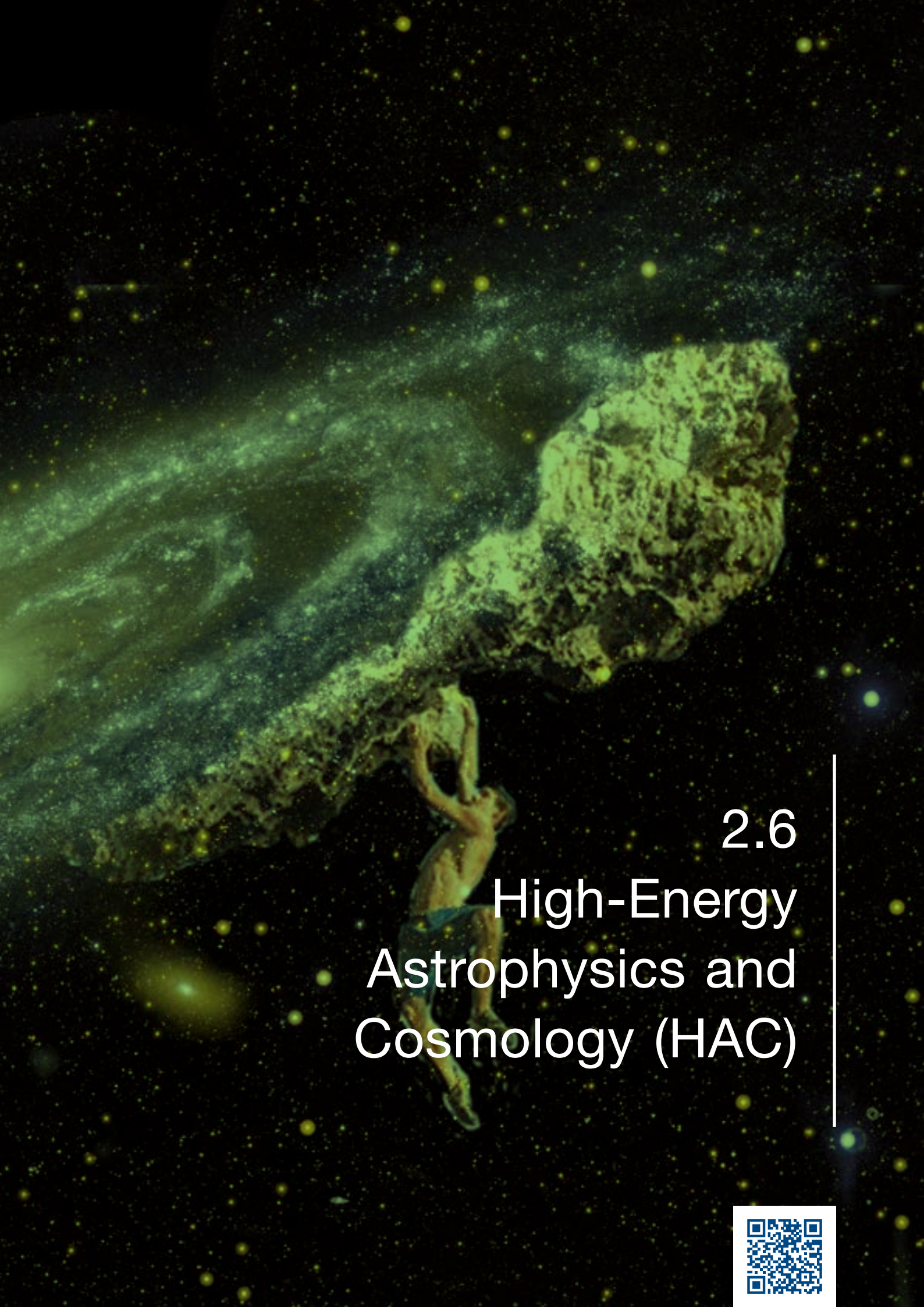


Figure 26: The Menger sponge.

2 Research



2.6 High-Energy Astrophysics and Cosmology (HAC)

2.6
High-Energy Astrophysics and Cosmology (HAC)

Cosmic rays consist of charged elementary particles that move very quickly, close to the speed of light. Now, one century after their discovery, we are only just beginning to understand the astrophysical mechanisms that underlie the acceleration of ordinary elementary particles to these incredible velocities: powerful shock waves driven by exploding stars or powered by mass accretion onto supermassive black holes. Are these cosmic rays merely an astrophysical curiosity, or do they play a decisive role in the evolution of the cosmic structure?

This is the central question that we hope to answer through our research in the "High-Energy Astrophysics and Cosmology" group. First indications point to the possibility that cosmic rays could be responsible for driving powerful gaseous outflows during the formation of galaxies, which would explain the low star formation observed and help solve a major problem in our understanding of galaxy formation.

Cosmic rays could even play a critical role during the evolution of galaxy clusters, the largest bound objects in existence and an important touchstone in understanding the formation of the structures in our Universe. However, the thermal history of clusters remains mysterious; some central regions of these clusters should have long cooled and collapsed since their formation, a famous conundrum known as the "cooling flow problem". Instead, the most massive black holes in the Universe, which are situated at the center of these galaxy clusters, appear to heat the cooling gas at just the right rate. But how does this mechanism work? Again, elusive cosmic rays could come to the rescue and provide the necessary stable heating mechanism.

In our group, we tackle these challenging issues via a combination of paper-and-pencil theory and an advanced simulation technique that employs a moving mesh as the underlying numerical structure, thereby enabling an unprecedented combination of accuracy, resolution, and physical completeness. Our theoretical efforts are complemented by a focused observational program on the non-thermal emission of galaxies and clusters that takes advantage of new capabilities in radio to gamma-ray wavelengths.

Kosmische Strahlung besteht aus geladenen Elementarteilchen, die sich sehr schnell bewegen, mit Geschwindigkeiten nahe der Lichtgeschwindigkeit. Ein Jahrhundert nach ihrer Entdeckung fangen wir nun langsam an, den astrophysikalischen Mechanismus zu verstehen, der gewöhnliche Elementarteilchen derartig beschleunigen kann: Mächtige Stoßwellen, die von Sternexplosionen verursacht werden oder – alternativ – beim Wachstum eines supermassiven Schwarzen Loches entstehen. Ist die kosmische Strahlung nur eine astrophysikalische Kuriosität oder spielt sie eine entscheidende Rolle bei der Entwicklung kosmologischer Strukturen?

Dies ist die zentrale Frage, welche wir in unserer Forschungsgruppe „Hochenergie-Astrophysik und Kosmologie" gern beantworten wollen. Erste Anzeichen sprechen für die Möglichkeit, dass die kosmische Strahlung für die gewaltigen Gasausflüsse während der Galaxienentstehung verantwortlich sein könnte. Diese Ausflüsse könnten dann die beobachtete verringerte Sternentstehungsrate innerhalb der Galaxien erklären und somit ein großes Rätsel der Galaxienentstehung lösen.

Die kosmische Strahlung könnte sogar eine wichtige Rolle bei der Entwicklung von Galaxienhaufen spielen, welche die größten gebundenen Objekte im Universum sind, und ein Prüfstein für die kosmologische Strukturentstehung darstellen. Jedoch bleibt die thermische Entwicklung von Galaxienhaufen mysteriös: die Zentralbereiche einiger Haufen sollten sich schon längst wesentlich stärker abgekühlt haben und kollabiert sein. Das ist das berühmte Kühlproblem in Galaxienhaufen. Stattdessen scheinen die schwersten Schwarzen Löcher im Universum, welche sich genau in den Zentren dieser Galaxienhaufen befinden, das kühlende Gas zu heizen. Aber wie funktioniert dieser Mechanismus genau? Hier könnte wiederum die flüchtige kosmische Strahlung des Rätsels Lösung sein und den nötigen stabilen Heizmechanismus bereitstellen.

In unserer Gruppe versuchen wir, uns diesen interessanten Herausforderungen zu stellen, indem wir Papier und Bleistift Theorie mit einer fortschrittlichen Simulationstechnik verknüpfen. Die Simulationsmethode verwendet ein sich bewegendes Gitter als zugrunde liegende numerische Struktur und ermöglicht Simulationen mit bisher unerreichter Genauigkeit, numerischer Auflösung und physikalischer Vollständigkeit. Diese theoretischen Anstrengungen werden von einem Beobachtungsprogramm komplementiert, das die nicht-thermische Emission von Galaxien und Galaxienhaufen untersucht, und dabei neue Beobachtungsmethoden im Radiowellenlängenbereich bis hin zur Gammastrahlung ausnützt.

Group leader

PD Dr. Christoph Pfrommer

Postdoc

Dr. Philipp Girichidis

PhD Students

Matteo Pais
Georg Winner *(HGSFP Scholarship)*

Master Student

Kristian Ehlert

Visiting Scientist

Balaji Venkat S. Narayanan *(June – July 2016)*

The impact of cosmic rays on galaxy formation

Understanding the physics of galaxy formation is arguably one of the most interesting elements of modern astrophysics. According to our current cosmological standard model, galaxies form in (slowly rotating) dark matter halos. Since the radiative cooling time of gas is shorter than the age of the galaxies, the gas cools and diminishes its pressure support at the center.

As a result, it collapses while conserving its specific angular momentum, thus settling into a rotationally supported cold disk. In the disk, the gas is compressed by self-gravity to densities that are sufficiently high for star formation. In the absence of any feedback process, we would expect all baryonic gas to be converted to stars. However, observations suggest that at most 20 % of the available gas is transformed into stars on the scale of the Milky Way and that the star conversion efficiency plummets towards smaller dwarf galaxies and larger giant elliptical galaxies. Understanding this enigma is of central importance to studies of cosmological galaxy formation.

While recent progress in cosmological simulations of galaxy formation models is remarkable, it is important to note that the involved feedback has typically been modeled empirically and tuned to match observed galaxy scaling relations, thereby weakening the predictive power of the corresponding calculations. The physics behind galactic winds and outflows remains especially unclear and has thus far been treated in a purely phenomenological manner in hydrodynamical simulations of galaxy formation. More problematic, most of these simulations have neglected cosmic rays and magnetic fields, which are known to couple dynamically and thermally to gas. Cosmic rays consist of very energetic particles that move at almost the speed of light and are extremely rare: There is only one cosmic ray particle (with a typical energy of a giga volt) per one billion gas particles in our Galaxy. Despite their small number, cosmic ray particles provide comparable pressure support to that of all gas particles in our Galaxy, thus making them a prime candidate for providing feedback on galaxies and driving powerful outflows.

In close collaboration with the TAP group at HITS, we have developed and implemented a numerical framework that follows cosmic ray physics in the cosmological hydrodynamic code AREPO [Pfrommer C, Pakmor R, Schaal K, Simpson CM, Springel V. Simulating cosmic ray physics on a moving mesh. Mon. Notices Royal Astron. Soc. (2017) 465:4500–4529]. This framework accounts for the sites of acceleration, particularly at resolved shock waves, at supernova remnants, and in collimated outflows that emerge from super-massive black holes. Cosmic rays gyrate around a given magnetic field line; they are thus transported alongside the plasma as it moves the magnetic field around.

We also follow the diffusive transport of cosmic rays along individual field lines [Pakmor et al., 2016a], which enables cosmic rays to move ahead of the thermal plasma and to dominate the gas dynamics. Cosmic rays slowly lose their energy as they collide inelastically with ambient gas. This process enables us to detect traces of them in the radio and gamma-ray sky and delivers a unique signal in the form of ghost particles, the elusive neutrinos.

As a massive star ends its life, it explodes in a supernova, which drives a shock wave into the interstellar medium. Our novel algorithm can identify the shock wave in our simulation, measure its strength, and accelerate cosmic rays to almost the speed of light (see *Figure 27*). Upon transferring a significant fraction of the shock kinetic energy to cosmic rays, the shock wave compresses the ambient gas to much higher densities and advances the shock at a slower rate, thereby demonstrating the immediate impact of cosmic rays on hydrodynamics.

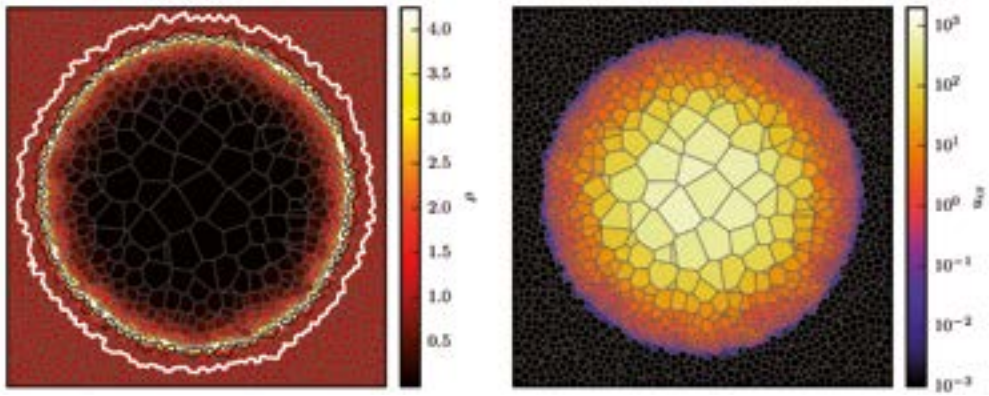
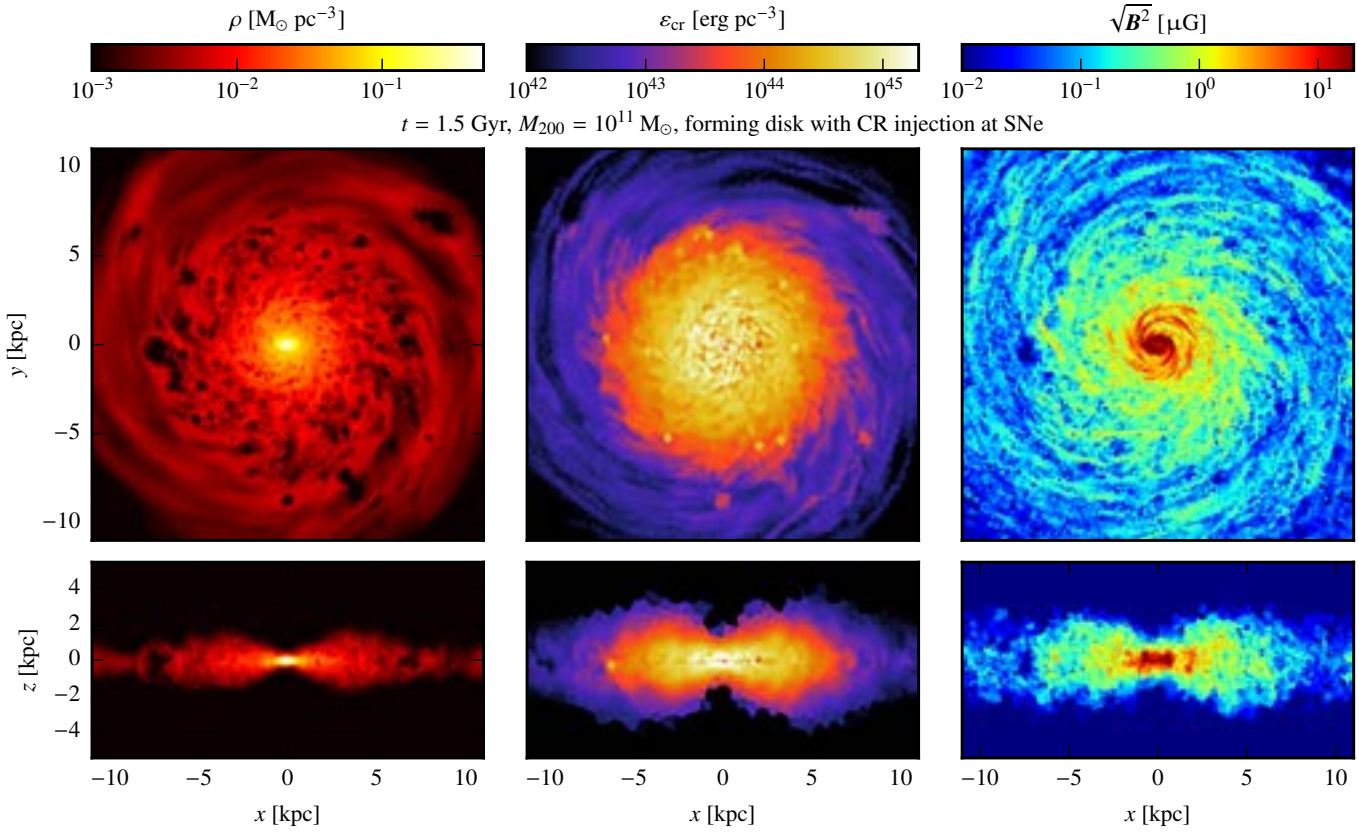


Figure 27: Expanding shock with cosmic ray acceleration. We display cross-sections through the center of the 3D simulation of the density (left) and the cosmic ray energy per unit mass (right).

The force of these shocks causes the interstellar medium to expand around the location of the explosion, which becomes visible as low-density "holes" in the face-on view of our simulated galaxy (shown as dark patches in the left panel of *Figure 28*, which correspond to bright spots in the central panel depicting the cosmic ray energy density). After the shock has sufficiently slowed down as a result of the interaction with the interstellar gas, the cosmic rays are set free and spread throughout the galaxy. Their additional pressure lowers the gas density and increases the galactic scale height, thereby slowing down the rate at which stars form.

Figure 28: Properties of the galactic gas disk in our simulation, which accounts for cosmic ray acceleration in supernovae and follows cosmic ray advection with gas. We depict cross-sections in the mid-plane of the disk (face-on views) and vertical cuts through the center (edge-on views) of the gas density (left-hand panels), the cosmic ray energy density (middle panels), and the magnetic field strength (right-hand panels).



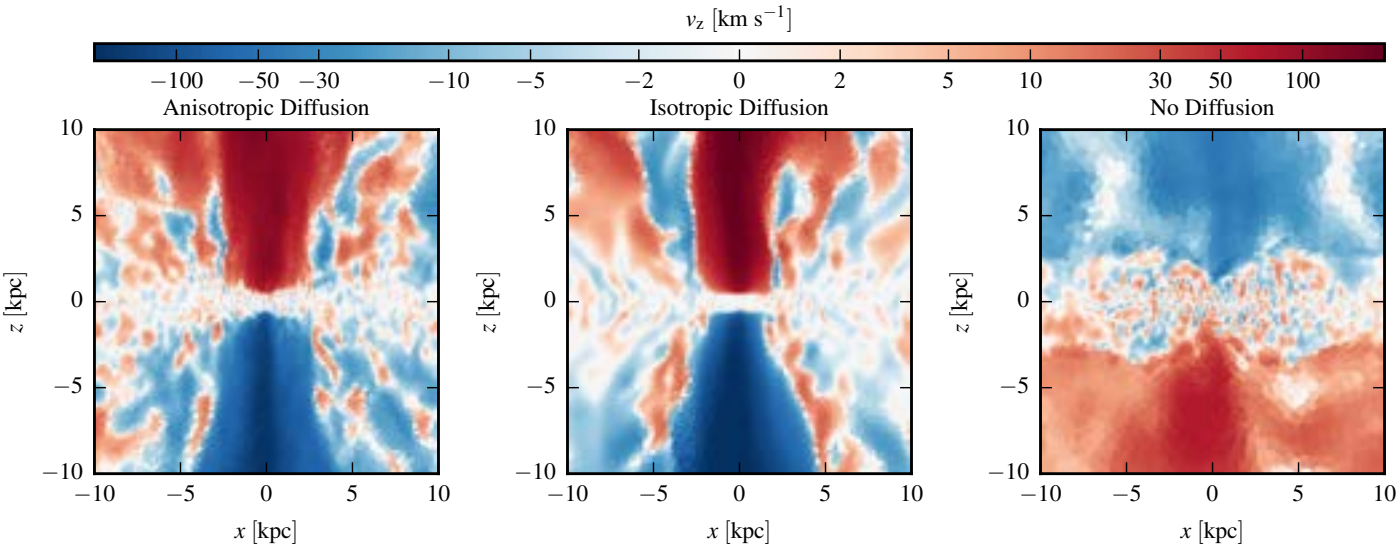


Figure 29: Slices through the center of the disk (edge-on view) showing the vertical velocity component after 1.5 Gyrs. The columns from left to right correspond to the simulations with anisotropic cosmic ray diffusion, with isotropic cosmic ray diffusion, and without cosmic ray diffusion, respectively.

The winding of the forming galactic disk stretches the magnetic field and amplifies it. As the cosmic rays propagate along this disk field, they begin to rise from the disk (as they are lighter than the heavy medium that surrounds them). This process drags the magnetic field up into the halo and also enables the cosmic rays to diffusely escape from the disk. At this point, the rays' pressure dominates over the thermal gas, accelerating it and forming a galactic super-wind [Pakmor et al., 2016b]. The formation of such an outflow merely requires that the cosmic rays be transported relative to the gas (irrespective of whether they are constrained to follow magnetic fields). Artificially suppressing this process causes small-scale fountain flows but results in no outflows (see right panel of *Figure 29*).

The physics of feedback by active galactic nuclei

Galaxy clusters are the largest gravitationally collapsed objects in the universe that have formed in recent cosmic history. Most of the ordinary matter in clusters is composed of hot gas that is emitted in X-rays, thereby cooling it. Cooling gas becomes denser, further increasing the cooling rate. This process creates a snowball effect, which should form stars at rates of up to several hundred solar masses every year. However, the predicted amount of cold material has not been observed, a conundrum known as the "cooling flow problem." Over the past few decades, it has become clear that the giant elliptical galaxies at the cluster centers host super-massive black holes with masses exceeding billions of solar masses (so-called active galactic nuclei). When gas cools, it loses pressure support and falls to the center. There, it joins an accretion disk that swirls around the black hole and eventually collapses onto it. A fraction of the collapsing gas does not accrete onto the black hole but instead becomes ejected in collimated outflows (so-called jets) of magnetic fields and energetic cosmic rays. While the total available energy in these jets is more than enough to offset the cooling, it is far from clear whether this fact represents the long-sought-after solution to the "cooling flow problem" or how this heating process would work exactly.

As we move to larger radii in the cluster, we encounter hotter gas that has not yet had time to cool. The

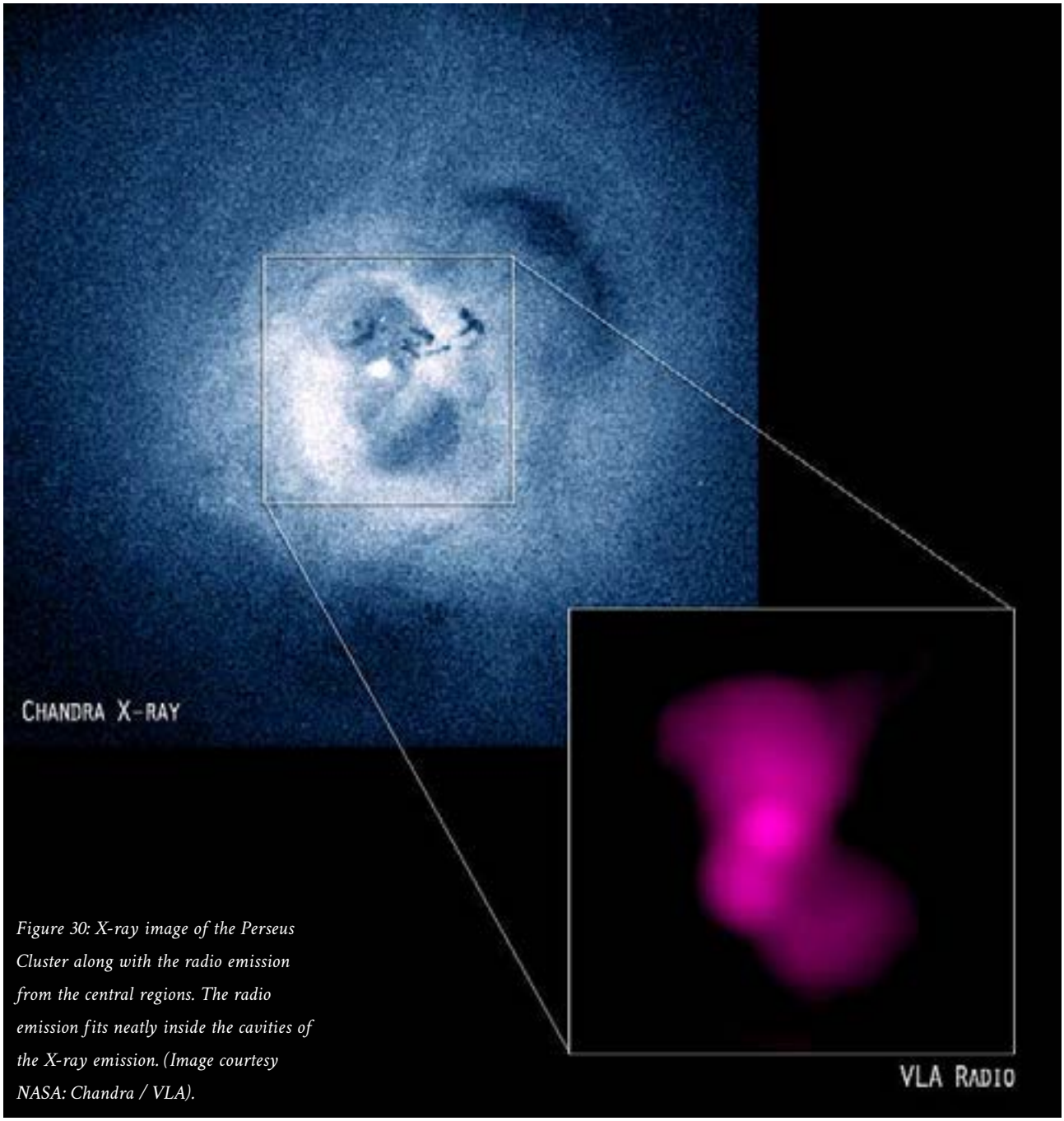


Figure 30: X-ray image of the Perseus Cluster along with the radio emission from the central regions. The radio emission fits neatly inside the cavities of the X-ray emission. (Image courtesy NASA: Chandra / VLA).

hotter gas thus acts as an enormous heat bath that can transfer heat to the cooler central regions by means of thermal conduction. However, the transport rate of heat steeply declines towards the colder regions, thereby calling for another process to provide a source of heat.

A promising solution could be provided by feedback through the active galactic nucleus at the cluster center. As the collimated outflows from black holes interact with the ambient, X-ray emitting cluster gas, they push it away and create spectacular, radio-emitting lobes, which potentially represent the signposts of a self-regulated feedback loop (as observed in the Perseus cluster; see *Figure 30*).

The light radio lobes begin to rise in the gravitational cluster potential and finally break up due to hydrodynamic instabilities. This releases the previously confined cosmic rays, which spread towards the cluster exterior. The propagating cosmic rays excite magnetic waves that are damped and thermalized, thereby heating the surrounding thermal plasma at a rate that scales with the number of cosmic rays. Analyzing a large sample of 39 cooling clusters, we find stable steady-state solutions that all match the observed gas density and temperature profiles very well [Jacob S, Pfrommer C. Cosmic ray heating in cool core clusters I: diversity of steady state solutions.

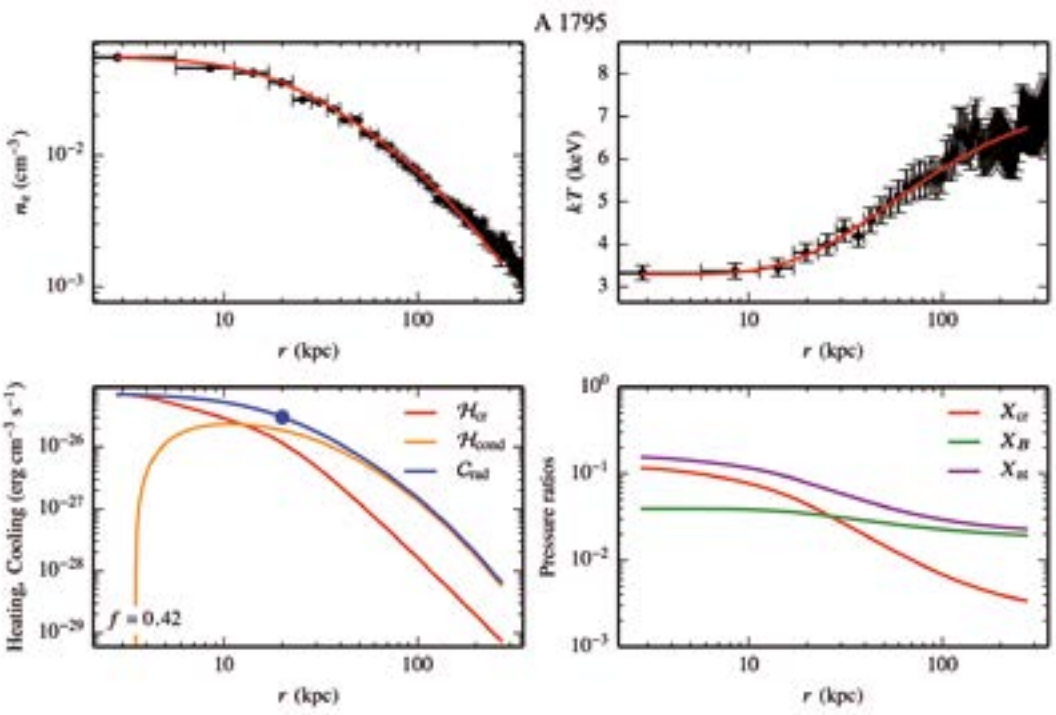


Figure 31: Steady state solution for cluster A 1795. The modelled electron number density and temperature profiles match observational data very well (top panels). We present the cooling- and heating-rate profiles (bottom left) created by cosmic rays (red) and conduction (orange) as well as ratios of CR-to-thermal pressure (red), magnetic-to-thermal pressure (green), and total non-thermal-to-thermal pressure (purple) as a function of the radius (bottom right).

Mon. Notices Royal Astron. Soc. (2017), I and II)]. The fact that cosmic ray heating dominates in the central, fast-cooling region is common to all these solutions. This process is replaced by conductive heating at larger radii (see lower left panel of *Figure 31*). In this cluster, the pressure contained in cosmic rays and magnetic fields is less than 10 % of the thermal pressure, which is in agreement with observational limits (see lower right panel of *Figure 31*).

How can we test this model? As these cosmic ray protons propagate, they collide inelastically with the ambient thermal gas protons. As a result of this hadronic reaction, pions are produced, which decay into gamma rays, "secondary" electrons, and positrons. The latter emit radio synchrotron radiation upon gyrating around the ubiquitous magnetic fields in clusters. Interestingly, the predicted radio emission exceeds the observed level of emission in those galaxy clusters that show an unusually high cooling- and star-formation rate. Lowering the cosmic ray pressure in order to bring the predictions into agreement with the observations would imply an imperfect balance of the heating and cooling, thus ruling out the cosmic ray heating solution. On the other hand, the model is perfectly in agreement with the observations of clusters in which cooling and star formation are most strongly suppressed. This apparently conflicting result can be resolved by realizing that galaxy clusters are not static but dynamically evolving systems. The novel solution to the "cooling flow problem" lies in a heating-cooling cycle with the following properties: As long as cosmic ray heating effectively balances cooling, only little gas can cool and precipitate from the hot cluster atmosphere, which leads to the observed slowly simmering star formation. Eventually, the supply of gas accretion onto the black hole is starved. As a consequence, the jets become weaker, which also reduces the amount of freshly injected cosmic rays. Once the previously injected cosmic rays have streamed out, their associated heating rate drops below the gas cooling rate, resulting in the observed abundant star formation. Consequently, for these systems, we expect to find a diffusely glowing radio emission that is excited by the cosmic rays that have previously heated the cluster cores and now stream outwards. Interestingly, this is observed as "radio mini halos", which only occur in these rapidly cooling clusters.

The origin of radio halos in galaxy clusters

According to the standard model of cosmology, the structure in the Universe grows hierarchically from small to large scales. Galaxies form first and merge into groups.

Finally, giant galaxy clusters that contain several hundreds to a few thousand galaxies form. In the process, galaxy clusters inevitably collide with each other to form even larger entities. The gravitational energy associated with these mergers – the most energetic events in the Universe after the Big Bang – is dissipated in enormous shock waves.

These giant mergers are associated with cluster-sized diffuse structures of radio synchrotron emission (called "giant radio halos") that provide evidence for relativistic electrons and magnetic fields that permeate entire galaxy clusters. However, these relativistic electrons have relatively short cooling times, rendering them quickly invisible in the radio band. Hence, they have to be accelerated close to the sites where they are observed, which challenges theoretical models to find an efficient cluster-wide acceleration process that can explain radio halos.

One promising model for radio halos is the acceleration of seed relativistic electrons by compressible turbulence in the intracluster medium. We explore the origin of these seed electrons by following their spectral evolution throughout cosmic history via high-resolution cosmological simulations of the formation of galaxy clusters [Pinzke A, Oh SP, Pfrommer C. Turbulence and particle acceleration in giant radio haloes: the origin of seed electrons. Mon. Notices Royal Astron. Soc. (2017) 465:4800–4816]. We begin with the initial acceleration of protons and electrons at strong cosmological shocks during the formation of a proto-cluster and follow the cosmic journey of these particles into present-day clusters. For standard assumptions, this leads to a seed electron population that produces an overly centrally concentrated radio emission. By modifying properties of the cosmic ray population (rapid streaming to the cluster periphery, enhanced electron acceleration at shocks) or turbulence (increasing turbulent-to-thermal energy density with cluster radius), we can match observations, but only at the expense of fine-tuning (see *Figure 32*), which displays how the simmering emission (dashed) is energized via turbulent re-acceleration).

To overcome this problem with fine-tuning, we perform a parameter study. Under naïve assumptions, radio properties are exponentially sensitive to the amplitude of turbulence, which is inconsistent with small scatter in the scaling relations of X-ray and radio luminosities. This sensitivity can be removed if we relate the acceleration time to the time it takes turbulence to dissipate motion into thermal energy. In this case, turbulence above a threshold value required to overcome the fast cooling of electrons provides a fixed amount of amplification; observations could then potentially constrain the unknown cosmic ray seed population. Understanding the small scatter in radio halo scaling relations may prove a rich source of insight into plasma physics in the intracluster medium.

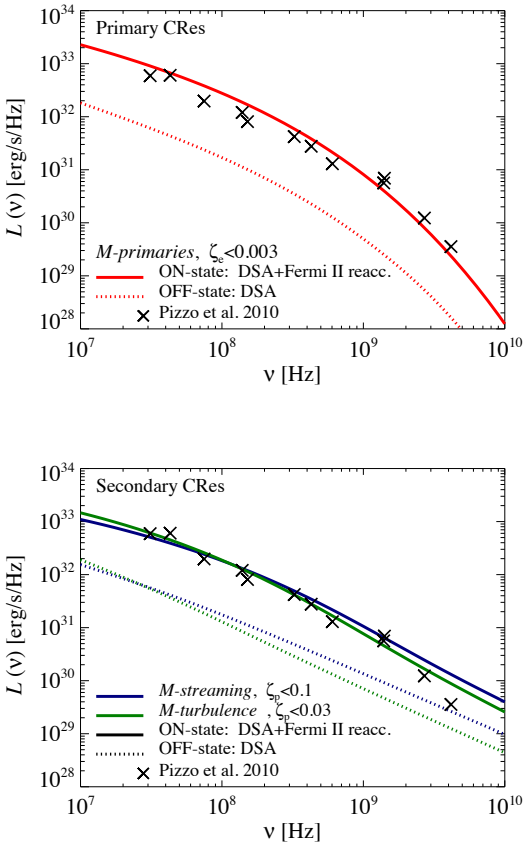
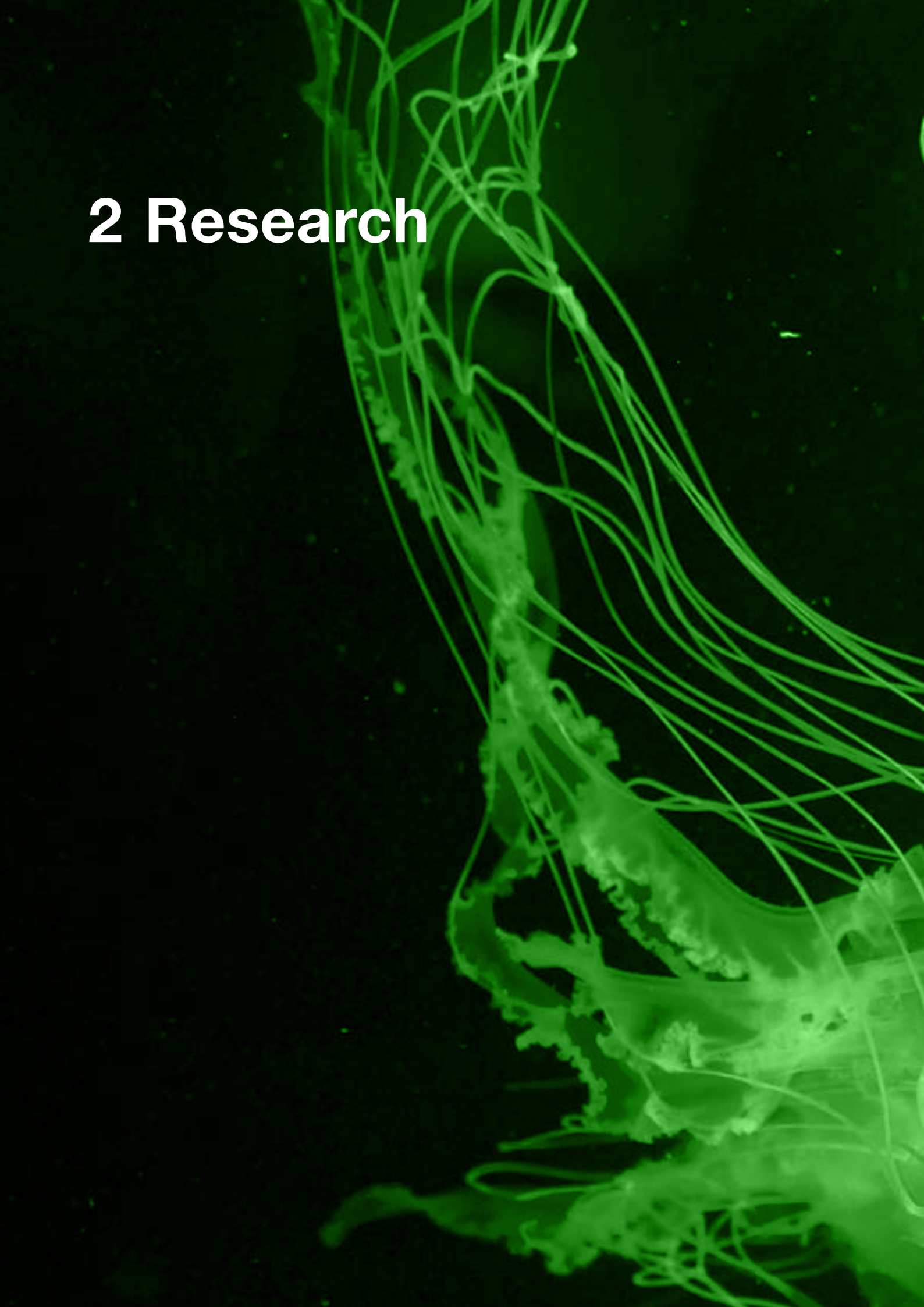
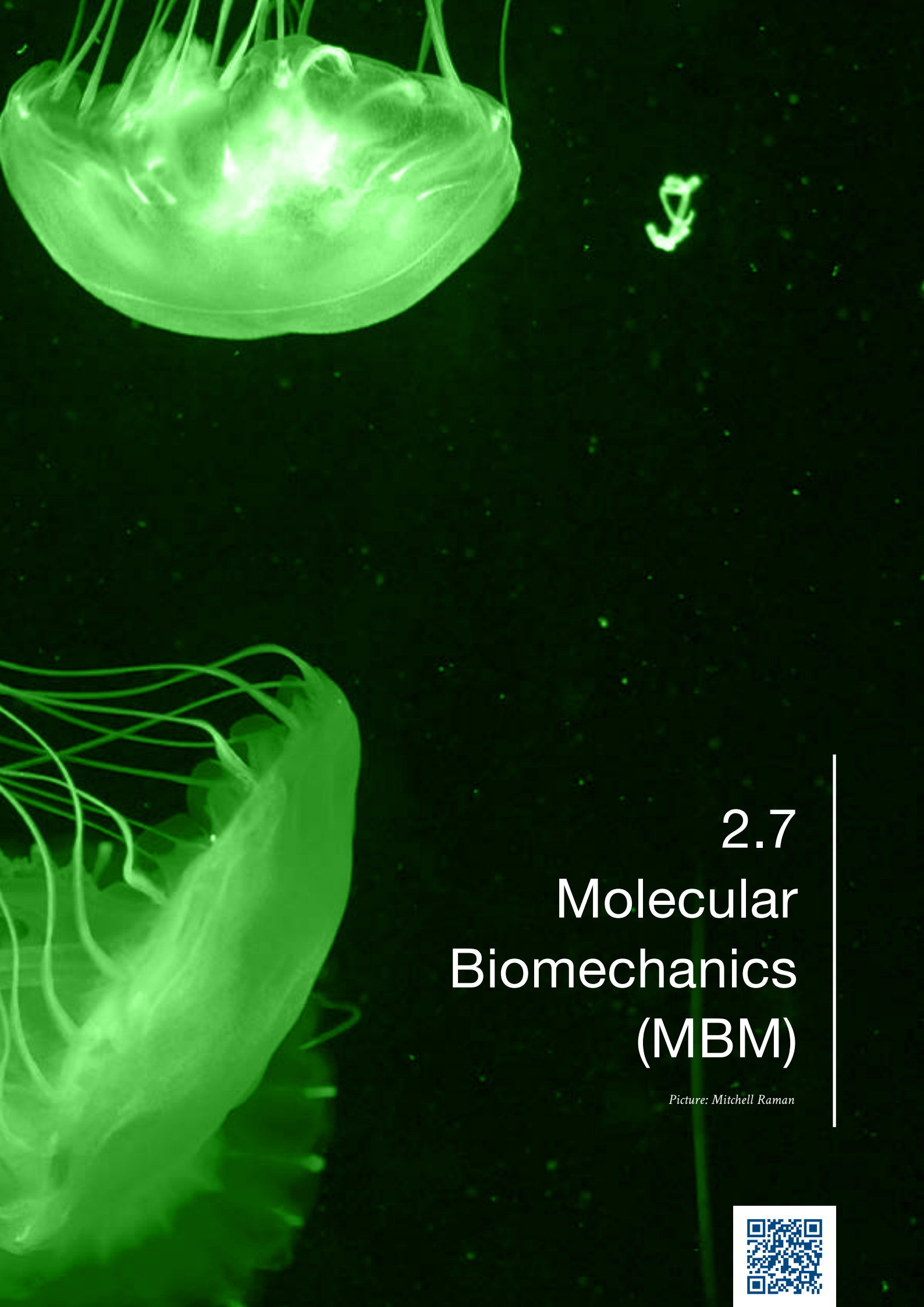


Figure 32: Radio synchrotron spectra. The dotted lines display models that account for the shock acceleration of cosmic rays and their successive transport (off-state of the radio halo), while the solid lines account for turbulent reacceleration during the merger phase (on-state). We depict the radio emission induced by primary electrons (top) and the emission from secondary electrons (bottom). Crosses are compiled from different observations.

2 Research





2.7 Molecular Biomechanics (MBM)

Picture: Mitchell Raman

2.7
Molecular Biomechanics (MBM)

Proteins are the workhorses of living systems. In past decades, we learned a great deal about how proteins look, move, and work. A rather recent discovery is that many proteins show very surprising properties once mechanical force acts upon them, which has important consequences for their function in the biological cell. The aim of the "Molecular Biomechanics" group is to decipher the mechanical function of proteins. We use Molecular Dynamics simulations, coarse-grained simulations, and continuum mechanics approaches to address the protein in multiple time- and length scales.

As one of our major achievements in 2016, we obtained very detailed insight into the basic principles of how proteins unfold under mechanical force. We were able to unfold the very same protein over and over again – more times than ever before – at atomistic detail on a computer, enabling us to statistically analyze the resulting large set of data. We concluded that the complex nature of protein dynamics leads to a heterogeneous ensemble of unfolding pathways, which explains previous experimental observations. We also helped to further understand the mechanical function of a number of specific proteins, including one that constitutes the elastic yet stable wall of the animal hydra and one that plays a key role in the first steps of blood coagulation.

Proteine sind die Arbeitspferde lebender Organismen. In den letzten Jahrzehnten haben wir vieles darüber gelernt, wie Proteine aussehen, sich bewegen und ihre Arbeit verrichten. Ein verhältnismäßig neuerer Aspekt ist, dass viele Proteine sehr überraschende Eigenschaften zeigen, wenn Kraft auf sie einwirkt. Dies kann wichtige Konsequenzen für ihre Funktion in der biologischen Zelle haben. Das Ziel der Molecular Biomechanics Gruppe ist, die mechanische Funktion von Proteinen zu verstehen. Wir benutzen hierfür atomistische Molekulardynamik-Simulationen, gröbere Proteinmodelle sowie Kontinuumsmechanik, um Proteinmechanik auf verschiedenen Zeit- und Längenskalen zu studieren.

Ein wichtiger Meilenstein im Jahr 2016 war unsere Arbeit zu den grundsätzlichen Prinzipien von Proteinentfaltung unter Ziehkräften. Wir konnten ein und das gleiche Protein immer und immer wieder entfalten, häufiger als das bei der atomistischen Auflösung jemals am Computer gemacht wurde. Dies erlaubte eine statistische Analyse der großen generierten Datensätze und bisher nicht mögliche Einblicke in die Komplexität und Vielfalt der Wege, die ein Protein bei seiner Entfaltung beschreiten kann. Darüber hinaus haben wir die spezifischen mechanischen Funktionen von bestimmten Proteinen mit materialwissenschaftlicher oder medizinischer Relevanz erforscht. Dazu gehört ein Bestandteil von Nesseltieren, der sowohl elastisch als auch sehr belastbar ist, sowie ein Molekül, dessen Kraftaktivierung essentiell für Blutgerinnung ist.

Group Leader:

Prof. Dr. Frauke Gräter

Staff members:

Dr. Camilo Aponte-Santamaría
Dr. Csaba Daday
Ana Maria Herrera-Rodriguez
Dr. Katra Kolšek
Dr. Davide Mercadante
Dr. Vedran Miletić
Dr. Agnieszka Obarska-Kosinska
Johannes Wagner
Christopher Zapp

Visiting Scientist:

Dr. Lipi Thukral

Students:

Daniela Barone
(March – June 2016)
Sara Becker *(December 2016)*
Florian Franz
Julia Greiwe
(May – August 2016)
Laura Mateyka
(January – March 2016)

The unique polymerization of minicollagens plays a crucial role in the biomechanics of sea anemones and jellyfish

Davide Mercadante, Frauke Gräter

Cnidarians are among the oldest organisms on Earth and include sea anemones and jellyfish. Their ability to prey and defend themselves in the marine environment is unique and particularly interesting from a biomechanical perspective. Cnidarians are equipped with a balloon-shaped organelle (a nematocyst) that contains an expellable needle used for preying or defense. The needle's ejection occurs on remarkably fast timescales since the nematocyst is under high pressure (approximately 5 GPa) before discharge. Thus, nematocyst walls are particularly interesting in ability to withstand such pressure. We therefore investigated the molecular determinants responsible for the biomechanical properties of the nematocyst.

Different folds evolved for different functions in minicollagens!

From a structural point of view, minicollagens are subdivided into two distinct folds that define the N- and C-terminal minicollagen molecules according to their position along the collagen chains that form the nematocysts [Dames, S. A., One short cysteine-rich sequence pattern – two different disulfide-bonded structure – a molecular dynamics simulation study. J Pept Sci (2015) 21:480-494]. The presence of two distinct folds has raised the question of whether the association of minicollagens is polarized towards homopolymerization (two associated minicollagens of the same type) or heteropolymerization (associated N- and C-terminal minicollagens). The association between minicollagens occurs through the formation of inter-molecular S-S bonds after the breakage of intra-molecular disulfide bonds, which are responsible for the observed 3D structure. We employed a strategy featuring molecular dynamics (MD) simulations and protein-protein docking to analyze the tendency of the two differently folded minicollagens to form dimers, which, on a larger

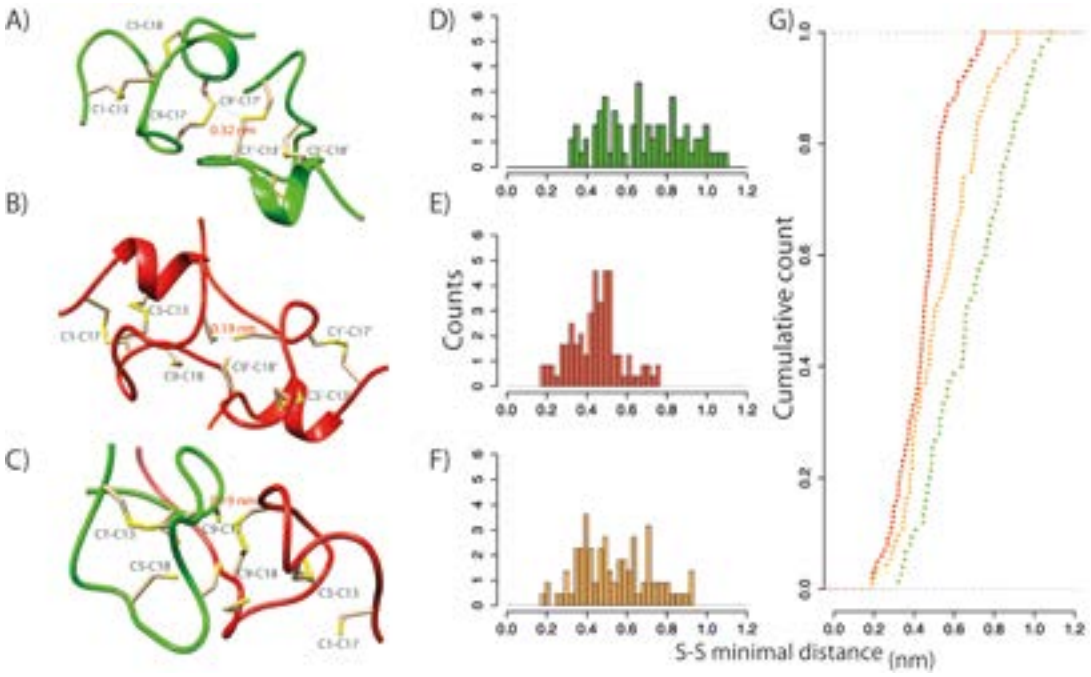
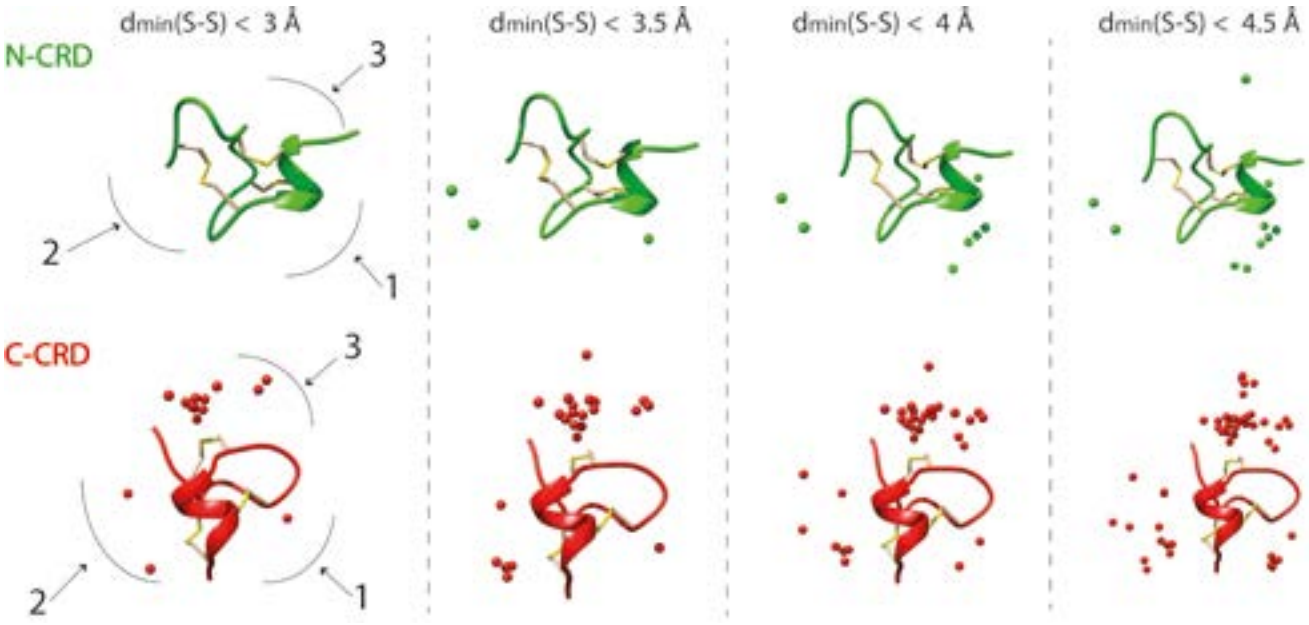


Figure 33: Preferential association of minicollagens. (A-C) Representative structures of N- (green) and C-terminal (red) minicollagens as retrieved from protein-protein docking. (D-F) Distance distributions between inter-molecular S atoms that can re-shuffle upon minicollagen polymerization. (G) Cumulative distributions of the histograms are reported in D-F.



scale, could lead to the polymerization of the proteins in nematocysts that form the outer wall of the organelle. In order to combine the knowledge acquired from computational studies and to test the predictive ability of simulations at different length scales, we worked with experimentalists at Heidelberg University and discovered the structural basis leading to minicollagen polymerization and nematocyst formation in the fresh-water cnidarian Hydra magnipapillata. Simulations and protein-protein docking revealed a preferential tendency of C-terminal minicollagen domains to react faster and form polymers (*Figure 33*).

Moreover, simulations determined the molecular basis for the observed tendency by reporting particularly high pre-stress in one of the C-terminal minicollagen's S-S bonds that tends to rupture with higher frequency and can more likely lead to a re-shuffling of intra-molecular S-S bonds into an inter-molecular disulfide formation. More importantly, we discovered that whereas N-terminal minicollagens can only form linear polymeric chains upon association, C-terminal minicollagens can engage in cross-linked polymers that have a critical impact on these molecules' ability to form a 3D cross-linked net to surround the nematocyst capsule (*Figure 34*).

This property is crucial during nematocyst maturation and presents minicollagens as a crucial polymerization unit in biology and potentially important elements for future biotechnological applications. The ability of minicollagen domains to polymerize in such a specific manner according to their 3D fold can be used to create conjugative blocks for polymers with tunable properties for a wide range of applications. As in the case of the nematocyst outer wall, thanks to polymerizing minicollagens, meshes of single or multiple polymers can be assembled in a highly controlled manner, a discovery that is useful for understanding both the evolution of cnidarians and the mechanism behind such highly mechanical biostability in nature.

Figure 34: Linear and branching abilities of minicollagens to polymerize. N- (N-CRD – green) and C-terminal (C-CRD – red) minicollagens' abilities to form homopolymers, investigated by MD. The relative frequencies of association between the monomers and the position of the second docked partner around the structure are reported as spheres surrounding the central structure using different distance cut-off values. C-CRDs show a higher tendency to polymerize (even at the lowest cut-off value) at three distinct sites around the structure, which reveals a cross-linking ability that is not present for N-terminal minicollagens.

Moving towards a functional understanding of mechanosensitive proteins during blood coagulation

Camilo Aponte-Santamaría, Frauke Gräter

The von Willebrand factor (VWF) is a glycoprotein in the blood that plays a central role in hemostasis. Among other functions, VWF is responsible for platelet adhesion at sites of injury. Using a multi-disciplinary approach that combines simulations and experiments, we have significantly contributed to the functional understanding of a number of VWF domains and to the comprehension of VWF genetic disorders for the development of diagnostic tools.

Disentangling the molecular strength of mutual domain interactions in VWF

VWF A1 / A2 interactions have recently been shown to inhibit the binding of platelets to VWF domain A1 in a force-dependent manner prior to A2 cleavage. Here, we provided relevant atomistic details of this protein-protein interaction via molecular dynamics (MD) simulations, atomic force microscopy (AFM), and Brownian dynamics (BD) simulations. The strength of the A1-A2 interaction was quantified by rupture force, which was in the order of 100 pN at typical experimental loading rates (*Figure 35A*). Extrapolation of the AFM rupture forces to very high loading rates, using the MD data as constraints, predicts the forces at rupture that will be observed in force-probe MD simulations. Force values derived from MD simulations are close to the predicted range but generally tend to be underestimated, presumably due to the variation and incomplete stability of conformations adopted by the complex before pulling (*Figure 35C*). Nevertheless, their relative differences are consistent with larger rupture forces observed exclusively for the native construct in which the A2 domain is able to unfold (*Figures 35A* and *35B*). When combined with force-probe MD simulations, AFM has the potential to validate predicted structural models not only of A1-A2 but also of any protein complex (in principle).

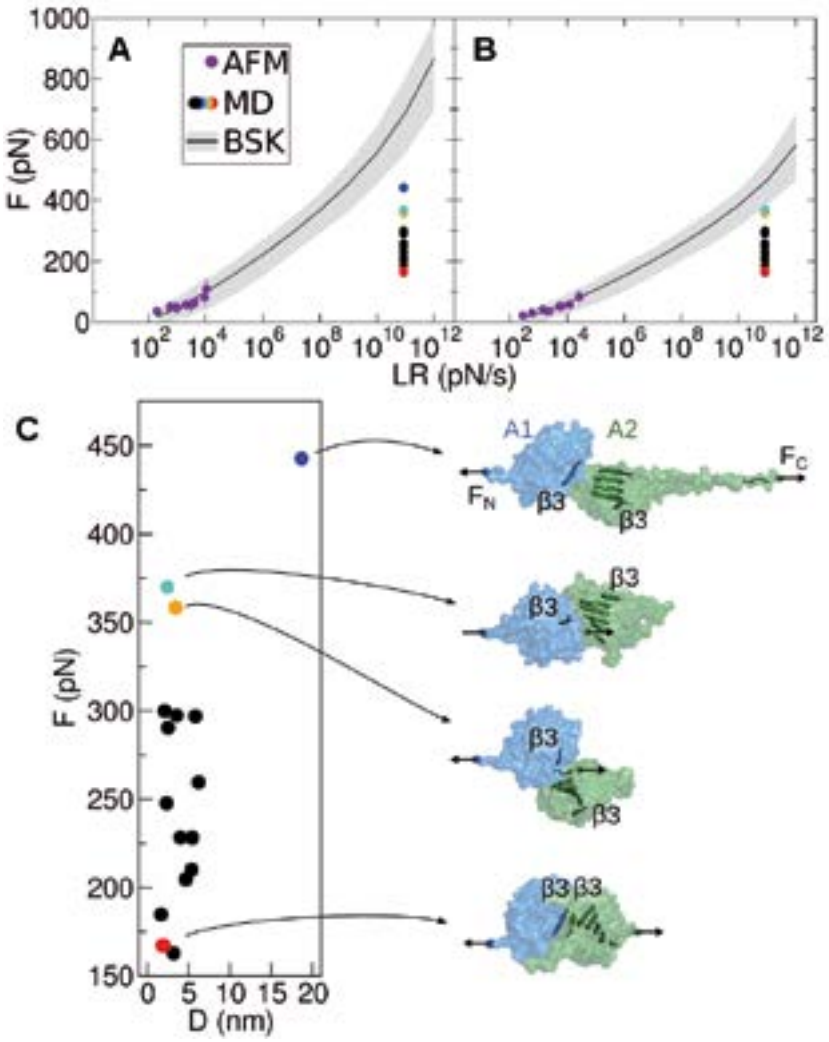


Figure 35: Force response of the VWF A1-A2 complex. (A-B) Rupture forces of the A1-A2 complex as a function of the applied loading rate for the A1-A2 complex (A: wild-type; B: locked mutant). Forces measured by AFM and computed from MD simulations (dots) were fitted using the BSK model. (C) Rupture force as a function of the distance between the pulled ends, recovered from MD simulations. Cartoons illustrate the conformation of the two domains for the indicated forces at the moment of rupture.

Fast but not too fast: protein cleavage in blood

VWF is cleaved by the protease ADAMTS13 as a down-regulatory mechanism to prevent excessive platelet aggregation. The ADAMTS13 cleavage site is buried deep inside the VWF A2 domain. External forces in vivo and denaturants in vitro trigger the unfolding of this domain, leaving the cleavage site solvent-exposed and ready for cleavage. Thus, by using fluorescence correlation spectroscopy, we determined – for the first time – that a mutation in the A2 domain (G1629E, *Figure 36A*) drastically accelerates the cleavage by almost 20 times, thereby resulting in an unwanted bleeding condition (*Figure 36B*). MD simulations and free energy calculations concluded that the pronounced cleavage activity of these disease-related VWF mutants can be readily explained by an increased availability of the ADAMTS13 cleavage site through a thermodynamic structural destabilization of the A2 domain (*Figure 36C*). Despite these adverse effects in vivo, these types of VWF mutants can be used as efficient and sensitive enzyme substrates for the detection of ADAMTS13 in clinical laboratories.

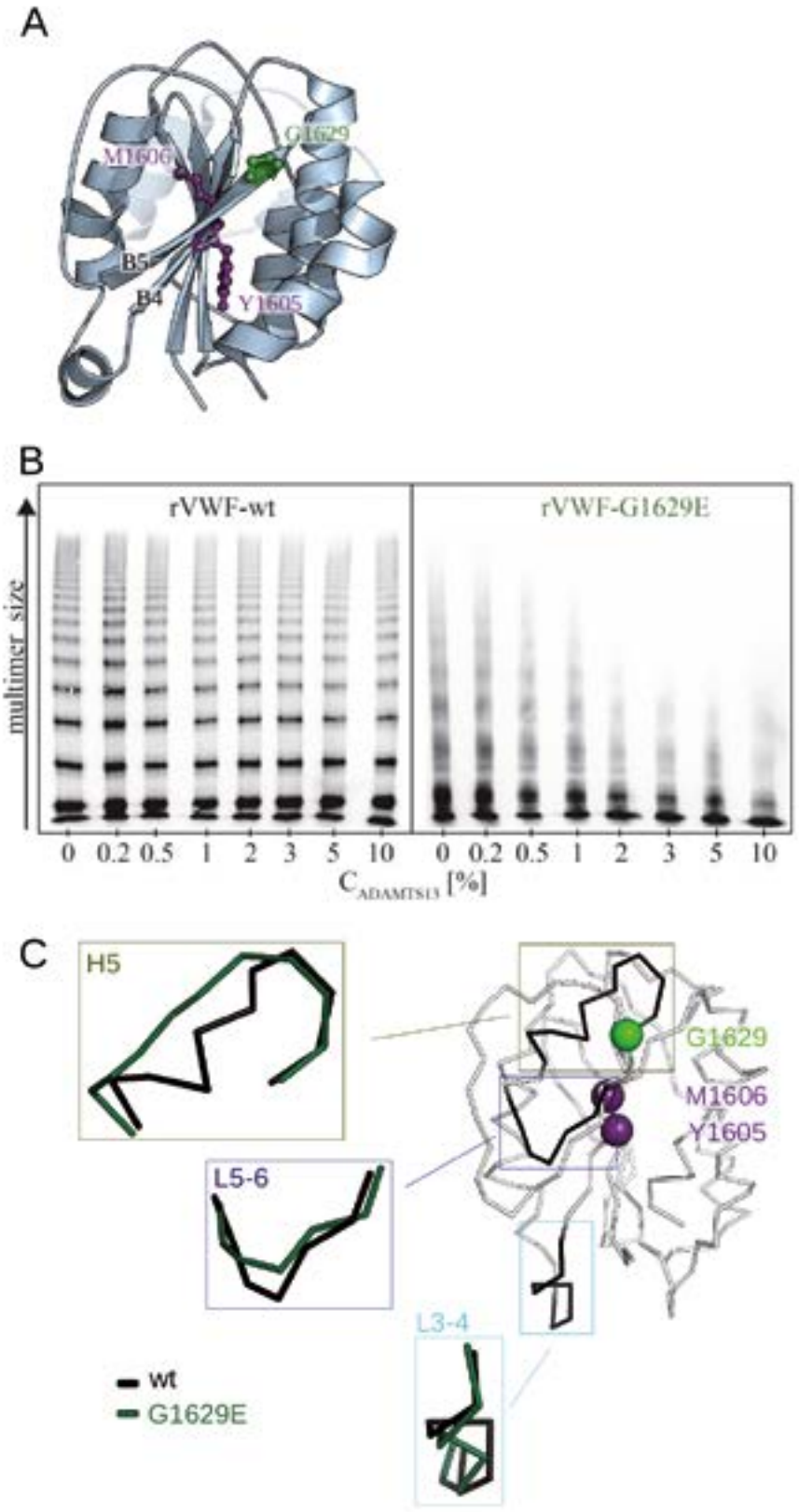


Figure 36: Mutation G1629E increases VWF cleavage via a cooperative destabilization mechanism. (A) VWF A2 domain is cleaved by ADAMTS13 after flow-induced exposure of the cleavage-site residues (Y1605 and M1606). (B) Disease-related mutation G1629E dramatically increases the cleavage as revealed by a multimer analysis. Large wild-type multimers are nearly absent for the mutant rVWF-G1629E compared with native VWF rVWF-wt. (C) Structural destabilization of VWF A2 domain upon G1629E mutation, recovered from MD simulations. Helix H5 and loops L5-6 and L3-4 display the largest structural differences.

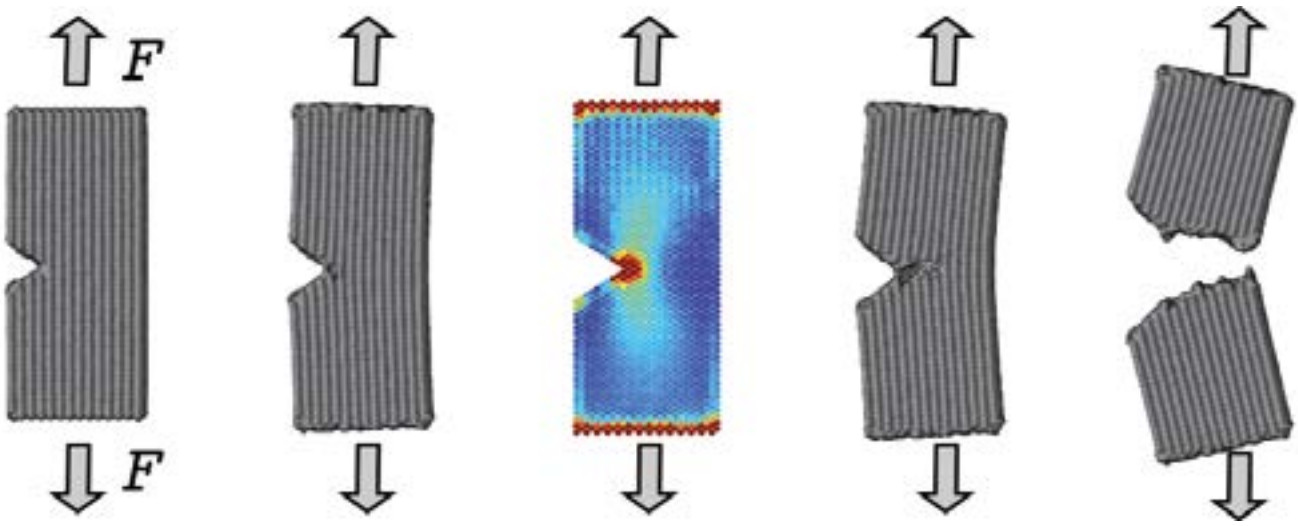


Figure 37: Diagram of nacre aragonite nanoflaws subjected to force.

Nacre aragonite is riddled with nanoflaws, but why is it still so strong?

Eduardo Cruz-Chú, Frauke Gräter

The ability to engineer materials at the nanoscale holds great promise for producing new materials with extraordinary properties. Nature provides us with many examples of mechanically robust materials composed from nanoscale components, many of which have mechanical properties far superior to their synthetic counterparts. Nacre, composed of 95 % CaCO3 crystal tablets integrated with 5 % organic layers in elaborate hierarchical nanostructures, is one such biological material and is often found on the inner layer of seashells and the outer coating of pearls.

The ordered CaCO3 tablets have previously been shown to be present in an aragonite polyform at the atomic level. An unusual feature of nacre aragonite is the presence of nanoflaws [Gao H, Ji B, Jäger IL, Arzt E, et al. Materials become insensitive to flaws at nanoscale: lessons from nature. Proc Natl Acad Sci USA (2003) 100:5597–5600], which are believed to be the points of crack initiation and hence material weakening. However, contrary to this belief, nacre still proves to have remarkable toughness and a high rupture strength. The organic layer and its various molecules are thought to contribute to these mechanical properties, but this function remains poorly understood.

Flaw tolerance: Rupture strength and stress distribution

We studied rupture strength and stress distribution under load via Molecular Dynamics (MD) simulations using minimalistic computational models of flawed aragonite single crystals (with and without nacre proteins). These very detailed simulations were then compared with macroscopic predictions from continuum mechanics. By introducing flaws of differing depths into aragonite to closely emulate the experimental data available and then applying force, we were able to compare the mechanical strength of the flawed nanocrystals. Surprisingly, flaws of only a few atoms resulted in a drop in strength, thereby calling into question the flaw-insensitivity put forward earlier for such systems.

Dynamic disorder can explain the non-exponential kinetics of fast protein mechanical unfolding

Dr. Bogdan Costescu, Sebastian Sturm
Prof. Dr. Frauke Gräter

The unfolding kinetics of proteins are well-described by mathematical models that point towards stretched exponentials rather than simple two-state models. At the molecular level, the basis of this observed behavior is still

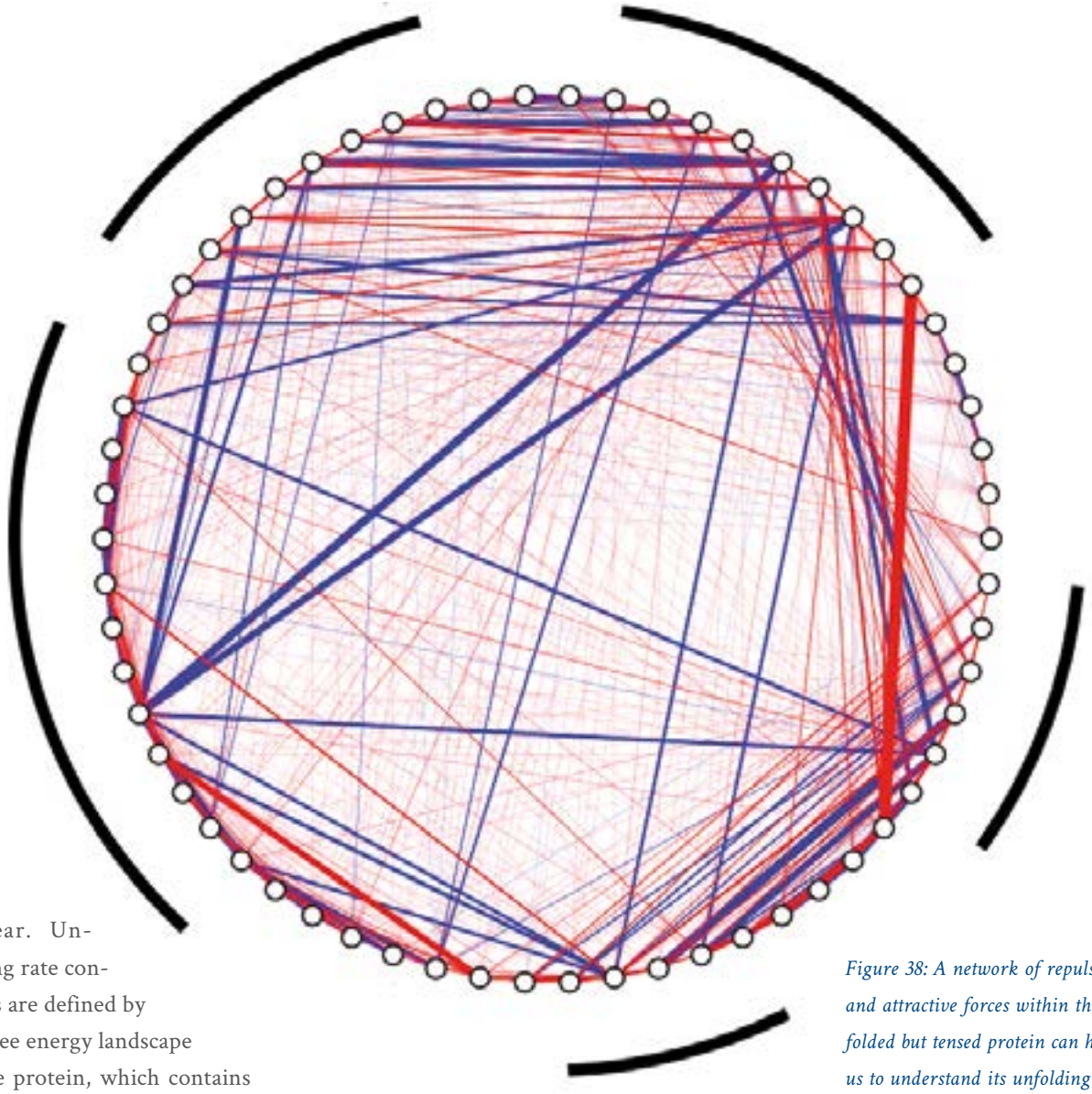


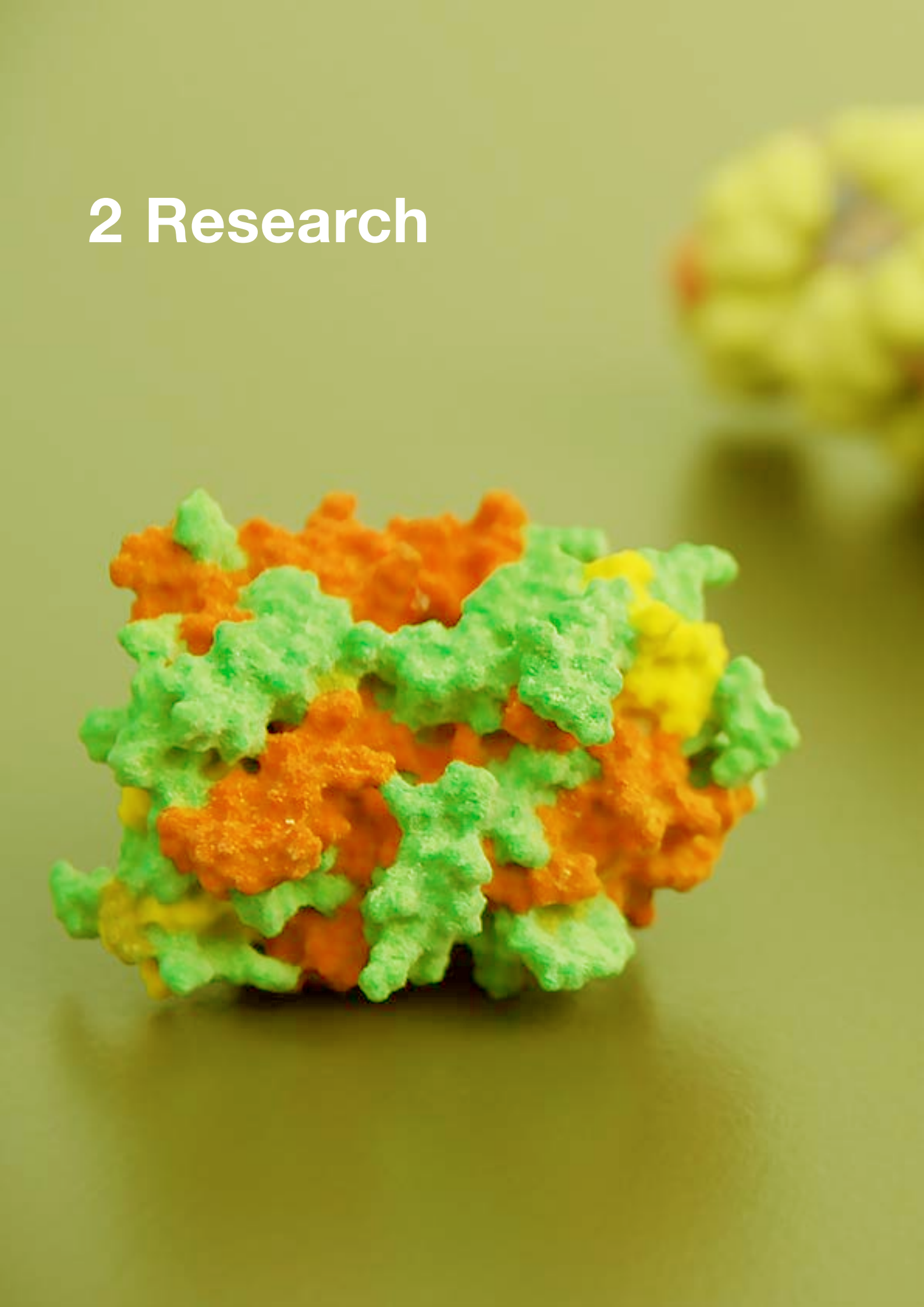
Figure 38: A network of repulsive and attractive forces within the folded but tensed protein can help us to understand its unfolding pathway.

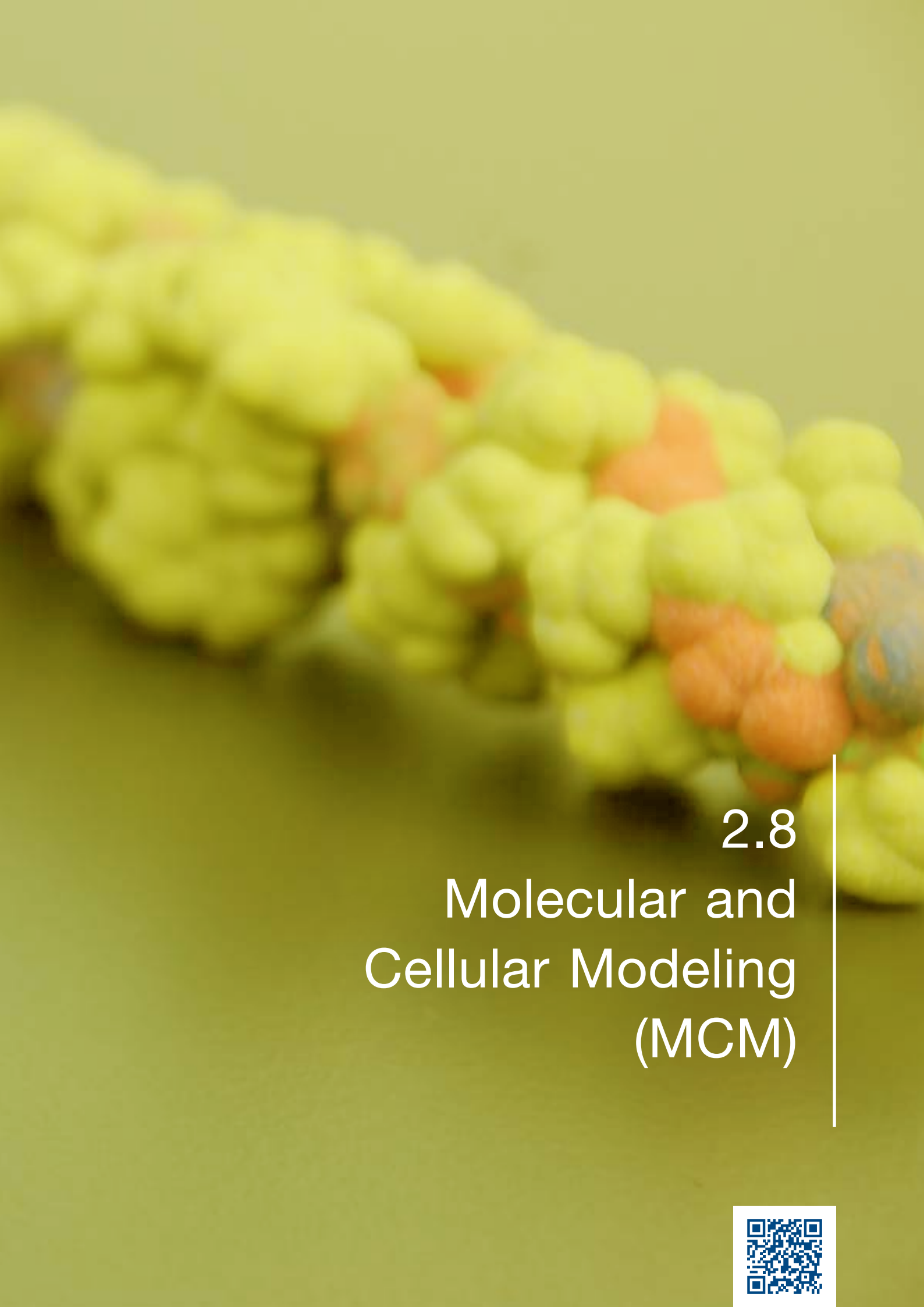
unclear. Unfolding rate constants are defined by the free energy landscape of the protein, which contains numerous barriers that link the unfolded and folded states. Slow conformational fluctuations can give rise to static disorder [Kuo TL, Garcia-Manyes S, Li J, et al. Probing static disorder in Arrhenius kinetics by single-molecule force spectroscopy. Proc. Natl. Acad. Sci. USA (2010) 107:11336–11340.], while fast nanosecond conformational fluctuations of the protein result in dynamic disorder. Both types of fluctuations can explain stretched exponential models describing unfolding kinetics equally well.

However, when studying protein unfolding experimentally, it is nearly impossible to differentiate between dynamic and static disorder because the small-scale atomistic fluctuations cannot be followed over time. However, molecular dynamics simulations can apply a constant force and retrieve conformations explored within tiny timescales all the way down to a millionth of a millionth of a second. Using such computer simulations to study the mechanical unfolding of two small and well-studied proteins, we were able to determine if non-exponential kinetics were present in the ns timescale in addition to those found in the s timescale of experiments. We were further able to address the question of whether the origin of stretched-exponential kinetics in protein unfolding is due to dynamic or static disorder.

By sampling 500 unfolding events of both proteins at a number of different constant forces, we were able to unequivocally verify (and, to our knowledge, for the first time) that the mechanical unfolding of the protein proceeds along stretched exponential kinetics, even in the very short ns timescale of the simulations. This finding supports previous ideas that proteins appear to be dynamically disordered, regardless of the timescale, due to the large dynamic range of fluctuations to which they are subjected [Hu X, Hong L. Smith MD, et al. The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time. Nature Phys (2015) 12;2:, 171–174.]

2 Research





2.8 Molecular and Cellular Modeling (MCM)

2.8 Molecular and Cellular Modeling (MCM)

Molecular recognition, binding, and catalysis are fundamental processes in cell function. The ability to understand how macromolecules interact with their binding partners and participate in complex cellular networks is crucial to the prediction of macromolecular function and to applications such as protein engineering and structure-based drug design.

In the MCM group, we are primarily interested in understanding how biomolecules interact. What determines the specificity and selectivity of a drug-receptor interaction? How can proteins assemble to form a complex, and what shape can the complex take? How is the assembly of a complex influenced by the crowded environment of a cell? What makes some binding processes quick and others slow? How do the motions of proteins affect their binding properties?

These questions are illustrative of the types of problem we address in our projects via the development and application of computational approaches to the study of biomolecular structure, dynamics, interactions, and reactions. We take an interdisciplinary approach that entails collaborations with experimentalists and makes concerted use of computational approaches based on physics and bio-/chemo-informatics. The broad spectrum of techniques employed ranges from interactive, web-based visualization tools to atomic-detail molecular simulations.

In this report, we detail some of the results achieved this year. These achievements demonstrate the types of methods we develop to study macromolecular interactions and their application to problems in biology, biotechnology, and drug design. Following a general overview, we turn our focus to projects on (i) exploring protein binding sites and their dynamics for ligand design and (ii) understanding macromolecular association and protein adsorption to surfaces through multiscale simulation approaches.

Molekulare Erkennung, Bindung und Katalyse sind grundlegende Prozesse der Zellfunktion. Die Fähigkeit zu verstehen, wie Makromoleküle mit ihren Bindungspartnern interagieren und an komplexen zellulären Netzwerken teilnehmen, ist entscheidend für die Vorhersage von makromolekularen Funktionen und für Anwendungen wie beispielsweise Protein-Engineering und strukturbasiertes Wirkstoffdesign.

In der MCM-Gruppe sind wir in erster Linie daran interessiert zu verstehen, wie Moleküle interagieren. Was bestimmt die spezifische und selektive Wirkung beim Zusammenspiel von Wirkstoff und Rezeptor? Wie werden Proteinkomplexe gebildet und welche Formen können sie annehmen? Welche Wirkung hat die beengte Zellumgebung auf die Bildung eines Proteinkomplexes? Warum verlaufen einige Bindungsprozesse schnell und andere langsam? Welche Auswirkungen haben Proteinbewegungen auf ihre Bindungseigenschaften?

Diese Fragen sind beispielhaft für die Art von Problemen, die wir in unseren Projekten durch die Entwicklung und Anwendung rechnerischer Methoden zur Untersuchung biomolekularer Strukturen, Dynamik, Wechselwirkungen und Reaktionen behandeln. In enger Zusammenarbeit mit Experimentatoren verwenden wir in interdisziplinären Ansätzen rechnerische Methoden aus den Bereichen der Physik-, Bio- und Chemoinformatik. Das breite Spektrum unserer Methoden reicht dabei von interaktiven webbasierten Visualisierungswerkzeugen bis hin zu Molekularsimulationen auf atomarer Ebene.

In diesem Bericht beschreiben wir einige der Ergebnisse unserer diesjährigen Arbeit. Sie demonstrieren einerseits die Methoden, die wir entwickeln, um makromolekulare Interaktionen zu modellieren und zu simulieren, und andererseits ihre Anwendungen in Biologie, Biotechnologie und Medikamentenforschung. Die Projekte beschäftigen sich (i) mit der Erforschung von Protein-Bindungs-Stellen und ihrer Dynamik für das Wirkstoffdesign und (ii) mit der Untersuchung von makromolekularer Assoziation und Protein-Adsorption an Oberflächen mittels Multiskalensimulation.

Group Leader

Prof. Dr. Rebecca Wade

Staff members

Dr. Neil Bruce
Dr. Daria Kokh
Dr. Prajwal Nandekar
Dr. Joanna Panecka
Ina Pöhner
Dr. Stefan Richter
Dr. Kashif Sadiq *(since March 2016)*
Antonia Stank

Scholarship holders

Gaurav Ganotra
Ghulam Mustafa
Musa Özboyaci *(until August 2016)*
Mehmet Öztürk

Visiting scientists

Dr. Hirdesh Kumar *(until November 2016)*

Students

Max Horn
Kai Horny *(since November 2016)*
Leo Kiss *(July – August 2016)*
Rebecca Neil *(since October 2016)*
Martin Reinhardt
Astrid Stubbusch *(February – May 2016)*
Sören von Bülow *(until February 2016)*

2.8 Molecular and Cellular Modeling (MCM)

General news

This year, we welcomed Dr. Kashif Sadiq from Barcelona as a new member of the group to work on the K4DD (Kinetics for Drug Discovery) project, a European IMI (Innovative Medicines Initiative) project. Additionally, Dr. Hirdesh Kumar visited the group to work on a project on structure-function relationships of actin filament stability that began in 2016. This project is a collaboration with Dr. Ross Douglas at the Center for Infectious Diseases at Heidelberg University Medical School and is supported by the Heidelberg University Innovation Fund Frontier program.

Dr. Musa Özboyaci completed his doctoral studies on the simulation of protein-solid surface interactions, and Astrid Stubbusch successfully carried out her thesis project for her Bachelor's degree in Biological Science at Heidelberg University. Four master's students from Heidelberg University carried out internships in the group: Soren von Bülow (Molecular and Cellular Biology), Leo Kiss (Molecular Biotechnology), Martin Reinhardt (Physics; he has continued on with his master's thesis studies), and Kai Horny (Biochemistry). Rebecca Neil joined us in October as an Erasmus student for a year abroad as part of her studies in Biochemistry with German for Science at Imperial College, London, U.K.

We completed the FP7 ramp-up phase of the EU-funded Human Brain Project this year and began on the next phase in the Horizon 2020 program. In the ramp-up phase, we collaborated with Paolo Carloni (Forschungszentrum Jülich, Germany), Richard Lavery (CNRS Lyon, France), and Jeanette Hallgren-Kotaleski (Stockholm, Sweden) to investigate the interactions of adenylate cyclase, a key enzyme in signaling cascades in the brain, and we published a comparative analysis of the interaction properties of different isoforms of this enzyme using our PIPSA methodology [Tong et al., 2016]. We are continuing to develop Brownian dynamics simulation methodology in our SDA software to investigate molecular interactions in crowded and confined cellular environments in this project. The EU FP7-supported NMTrypI (New Medicines for Trypanosomatidic Infections) consortium, in which we perform molecular modelling, published its first paper, which describes a new class of compounds with antiparasitic activity [Borsari et al., 2016]. Together with Marc Bianciotto (Sanofi, Vitry-sur-Seine, France) and Steffen Wolf (Ruhr-University, Bochum), Daria Kokh and Rebecca Wade were awarded 33 million core hours on the supercomputer at CINECA (Italy) by PRACE (the Partnership for Advanced Computing in Europe) for "Molecular Dynamics Simulations for Computing Kinetic Constants for Drug Discovery" in the K4DD IMI project.

Exploring protein binding sites and their dynamics for ligand design

The dynamics of protein binding pockets are crucial for their interaction specificity. Structural flexibility allows proteins to adapt to their individual molecular binding partners and facilitates the binding process. This implies the necessity to consider protein internal motion in determining and predicting binding properties and in designing new binders. Although accounting for protein dynamics presents a challenge for computational approaches, it expands the structural and physicochemical space for compound design and thus offers the prospect of improved binding specificity and selectivity.

A cavity on the surface or in the interior of a protein that possesses suitable properties for binding a ligand is usually referred to as a binding pocket. The influence of protein flexibility on binding pockets can vary from small changes to an already-existent pocket to the formation of a completely new pocket. In a

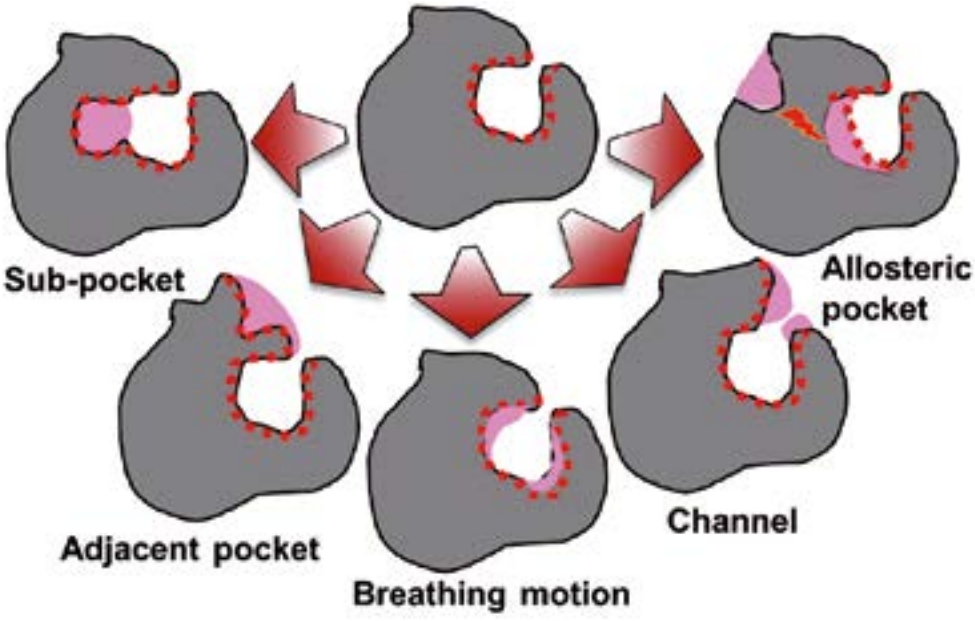
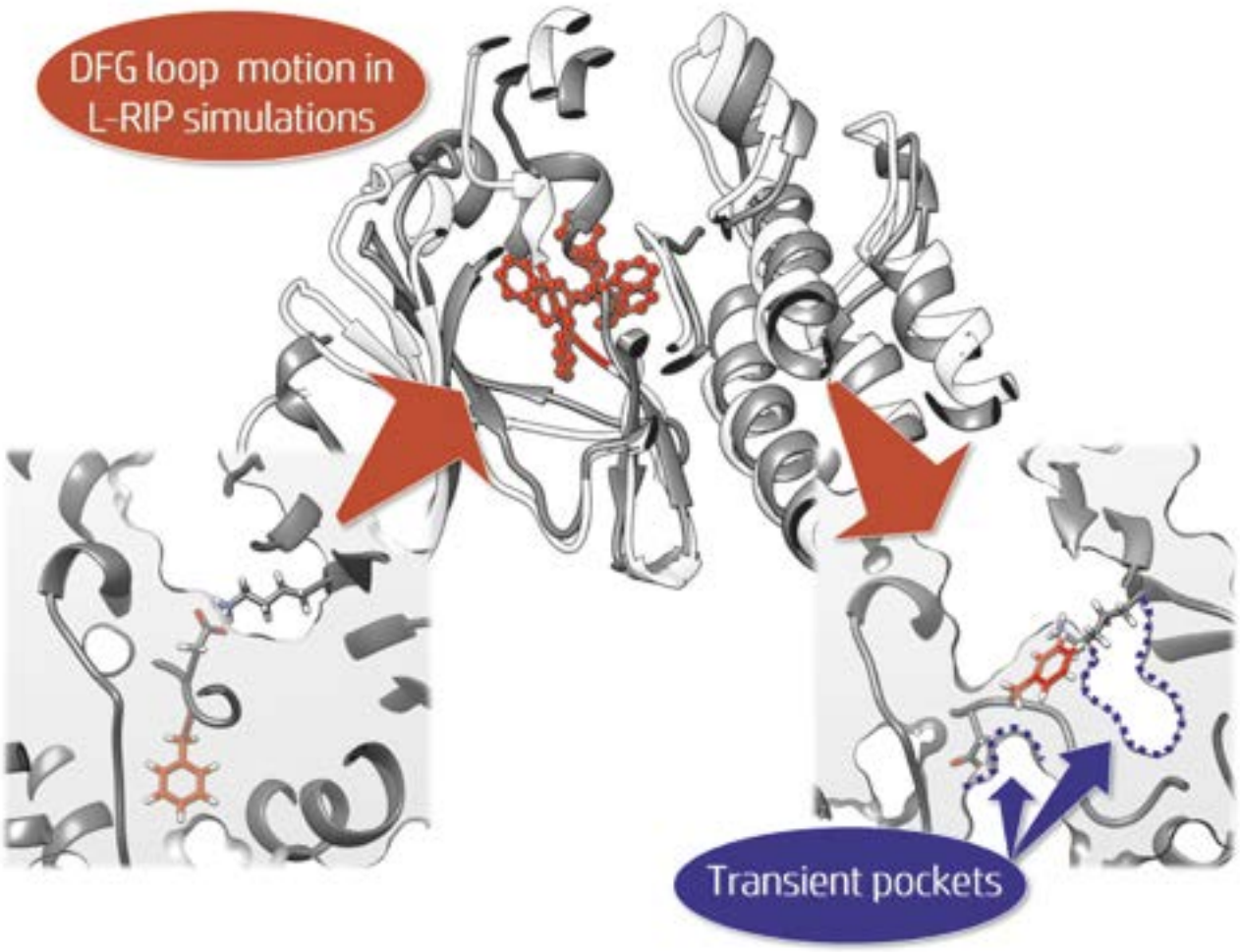


Figure 39: Illustration of five different classes of protein pocket dynamics [Stank et al., 2016a]: sub-pocket appearance/disappearance; adjacent pocket appearance/disappearance; pocket breathing motion, which may be caused by side-chain fluctuations, and backbone or inter-domain vibrational motion; channel/tunnel opening/closing connecting a pocket inside the protein with solvent, including lid motion; allosteric pocket appearance/disappearance at a site on a protein distinct from an already-existing pocket, with the binding of a ligand to the allosteric binding site affecting the original pocket. Regions colored in pink indicate pocket variation relative to the reference structure (shown in the center); the red dotted lines show the pocket shapes.

review of recent developments in computational methods to detect binding pockets and to study their dynamics [Stank et al., 2016a], we introduced five different classes of protein pocket dynamics (see *Figure 39*). We suggest that the class of pocket dynamics – as well as the type and extent of protein motion affecting the binding pocket – should be factors considered in choosing the most appropriate computational approach to study a given binding pocket.

One of the challenges in studying protein binding pocket dynamics is that the slow motions of proteins that can affect pocket shape are difficult to access using standard molecular dynamics (MD) simulations. Therefore, we have developed two non-equilibrium MD approaches to identify conformational changes of a binding site and to detect transient pockets associated with these motions [Kokh et al., 2016]. These methods are based on the Rotamerically Induced Perturbation (RIP) MD approach, which employs a perturbation of side-chain torsional motion to initiate large-scale protein movement. The first approach, Langevin-RIP (L-RIP), entails a series of short Langevin MD simulations, each beginning with a perturbation of one of the side-chains lining the binding site of interest. L-RIP provides extensive sampling of conformational changes of the binding site. In less than 1 ns of MD simulation with L-RIP, we observed distortions of the alpha-helix in the ATP bind-



ing site of HSP90 and a flipping of the DFG loop in Src kinase (see *Figure 40*). In the second approach, RIPlig, a perturbation is applied to a pseudo-ligand placed in different parts of a binding pocket, which enables flexible regions of the binding site to be identified in a small number of 10 ps MD simulations. We evaluated the methods for four test proteins displaying different types and degrees of binding site flexibility. Both methods revealed all transient pocket regions within a total of 10 ns of simulations even though many of these regions remained closed in 100ns conventional MD. The proposed methods provide computationally efficient tools for exploring binding site flexibility and can aid in the functional characterization of protein pockets and in the identification of transient pockets for ligand design.

Figure 40: Illustration of the application of the L-RIP method to identify transient pockets in Src tyrosine kinase. Beginning with the crystal structure on the left, L-RIP perturbation initiates motion of the DFG loop, which contains the sequence Asp(D)-Phe(F)-Gly(G). This motion (top image) is accompanied by flipping of the Phe405 residue (shown in red), which opens a cryptic pocket adjacent to the ATP binding site (right image). The shapes of the binding pockets before and after simulation are shown on a cross-section through the protein. The image was generated using Chimera.

To facilitate the study of transient pockets in proteins, we developed the TRAPP webserver (http://trapp.h-its.org), which provides an automated workflow that allows users to explore the dynamics of a protein binding site and to detect pockets or sub-pockets that may transiently open due to protein internal motion. Conformations of pockets can be generated by several methods, including L-RIP and

RIPlig. We have recently incorporated tools to visualize binding pocket sequence conservation and sequence annotations along with other transient pocket properties. Residue-based annotations are displayed with ProSAT+ [Stank et al., 2016b]. ProSAT+ (Protein Structure Annotation Tool – plus) is a new web server available at http://prosat.h-its.org for mapping protein sequence annotations onto a protein structure and visualizing them simultaneously with the structure. ProSAT+ incorporates many of the features of our preceding ProSAT and ProSAT2 tools but also provides new options for the visualization and sharing of protein annotations. Data are extracted from the UniProt KnowledgeBase, the RCSB PDB, and the PDBe SIFTS resource, and visualization is performed using JSmol. User-defined sequence annotations can be added directly to the URL, thus enabling visualization and easy data sharing. An example of usage is shown in *Figure 41*.

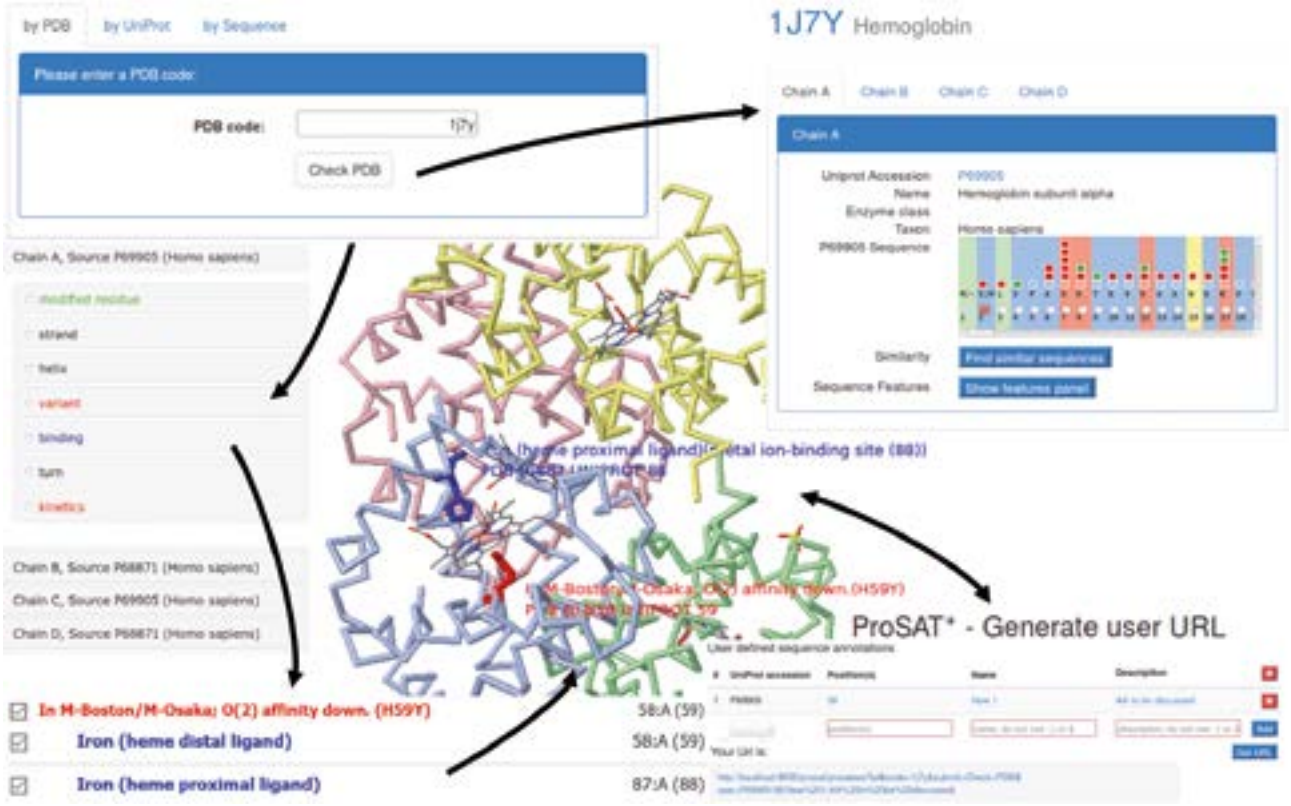


Figure 41: Screenshots illustrating the use of ProSAT+ to explore functional sites in the hemoglobin heterotetramer and why a specific variant is associated with a disease phenotype affecting oxygen transport. The user enters the protein data bank (PDB) code of a protein structure and is provided with an annotated display of the protein. The sequence scroll bar and the sliding side panel can be used to select specific functional residues and find sequence annotations. The heterotetramer has four subunits (chains A – D, distinguished by color): two alpha subunits and two beta subunits. The sequence annotations of the two corresponding Uniprot entries are listed in the side panel. The variant shown leads to a decrease in oxygen binding affinity and has a single point mutation at a residue near the heme cofactor in the 3D structure. The annotation and its visualization can be saved and shared in a user-generated URL. [Stank et al., 2016b, Figure 40, reproduced by permission of Oxford University Press].

Understanding macromolecular association and protein adsorption to surfaces through multiscale simulation approaches

The combination of Brownian and molecular dynamics simulations provides an efficient approach for the study of macromolecular association processes. Two different ways in which these techniques can be combined are illustrated by the two studies described below. For studying the formation of a chromatosome, we generated structures of a nucleosome using atomic-detail molecular dynamics simulations. We then used these structures in implicit solvent, rigid-body Brownian dynamics of the diffusional association of a linker histone protein with a nucleosome [Öztürk et al., 2016]. To investigate the mechanism of protein adsorption on a gold surface, we used structures of the encounter complexes between a protein molecule and a gold surface generated by Brownian dynamics simulations as starting points for atomic-detail, explicit solvent molecular dynamics simulations of the further adsorption of the protein onto the surface [Ozboyaci et al., 2016a].

For the Brownian dynamics simulations, we used our SDA (Simulation of Diffusional Association) software. This software has long been applied to simulate the diffusional association of two proteins in order to compute bimolecular association rate constants and to generate structures of diffusional encounter complexes. Some applications published this year are described in [Gdynia et al., 2016], [Gundic et al., 2016], [Lensink et al., 2016], [Subramanian et al., 2016], [Wild et al., 2016], and [Yu X et al., 2016a]. In 2016, we introduced a new treatment of interaction grids into SDA to allow large molecular systems and big high-resolution grids to be handled [Ozboyaci et al., 2016c]. We demonstrated that the DT-Grid data structure can be adapted and applied to efficiently represent macromolecular interaction grids by exploiting the non-uniformity of information on the grid while at the same time ensuring fast random data access. Furthermore, we demonstrated the use of DT-Grid for computing the interaction of a viral capsid with a gold surface [Ozboyaci et al., 2016c].

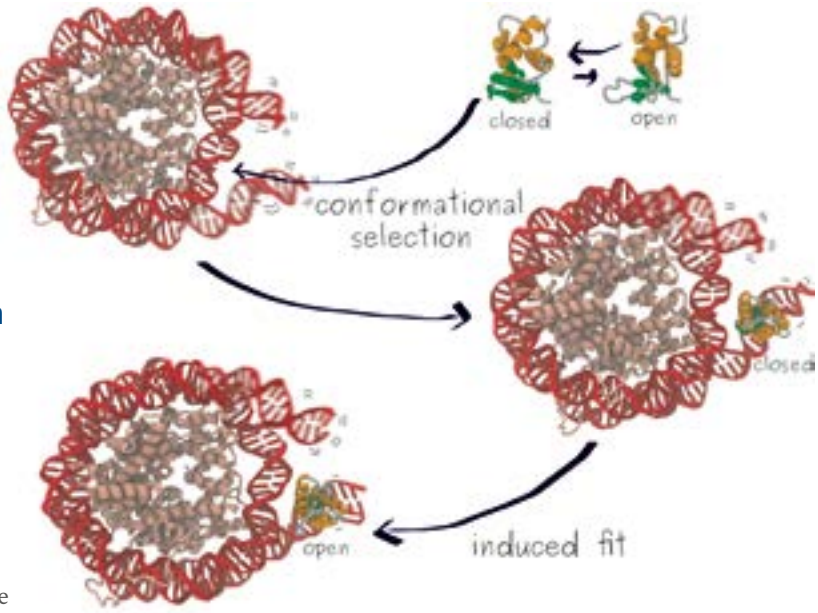


Figure 42: Illustration of the mechanism of linker histone – nucleosome binding to form a chromatosome (DNA is dark red; core histones are pink). The linker histone H5 can take closed and open conformations; the closed conformation is preferred in solution (upper-right corner, colored by secondary structure: orange = alpha helix, green = beta sheet, gray = loops and disordered regions). Guided by conformational selection, the closed linker histone preferably binds to the nucleosome in an off-dyad position, and the motion of the DNA is suppressed by linker histone binding. Upon linker histone binding to the nucleosome, the linker histone adopts an open state via an induced fit mechanism see [Öztürk et al., 2016].

Molecular dynamics involved in DNA compaction in the cell nucleus

DNA is packaged in the cell nucleus by wrapping itself around proteins called histones to form chromatin fibers. Nucleosomes, the structural units of these fibers, are formed by wrapping 146 base pairs of DNA around eight histones. Additional histone proteins, known as linker histones, are essential for DNA compaction in chromatin; they bind to nucleosomes in a 1:1 ratio to form chromatosomes. We studied the dynamics of chromatosome formation together with Vlad Cojocaru (Max Planck Institute for Molecular Biomedicine, Münster) [Öztürk et al., 2016], using classical and accelerated molecular dynamics simulations and Brownian dynamics simulations to investigate how the linker histone known as H5 binds to nucleosomes. We found that chromatosomes do not have a single defined geometry but multiple geometries depending on the dynamic shapes (conformation) and sequences of the interacting nucleosome and linker histone. This concept represents an important contribution towards understanding how DNA compacts in the cell nucleus.

Three steps to gold: The mechanism of protein adsorption revealed by Brownian and molecular dynamics simulations

Protein-gold interactions are one of the most widely studied types of protein-surface interaction due to gold's unique optical, magnetic, and chemical properties (see [Ozboyaci et al., 2016b] for a review of computational studies). Applications of protein-gold interactions include biological imaging, drug delivery, and biosensors. Whereas cysteine-gold bond formation has been studied extensively, the effects of weaker interactions on protein-gold binding are still poorly understood due to experimental limitations in revealing the binding dynamics and the adsorption conformation and orientation of a protein. We employed a multi-scale simulation approach and free energy calculations to study the mechanism of binding to a gold surface of a model protein, the β-lactamase inhibitor protein, with three N-terminal histidines added. The N-terminal tail substantially increases the protein's binding potency to an Au(111) surface. We found that adsorption is a three-step process: (i) the recognition of the surface predominantly by the histidine fusion peptide and the formation of an encounter complex facilitated by a reduced dielectric screening of water in the interfacial region, (ii) the adsorption of the protein on the surface and the adoption of a specific binding orientation, and (iii) the adaptation of the protein structure on the metal surface accompanied by induced fit (see *Figure 43*). We anticipate that the mechanistic features of protein adsorption to an Au(111) surface revealed in this work can be extended to other inorganic surfaces and proteins and will therefore aid in the design of specific protein-surface interactions.

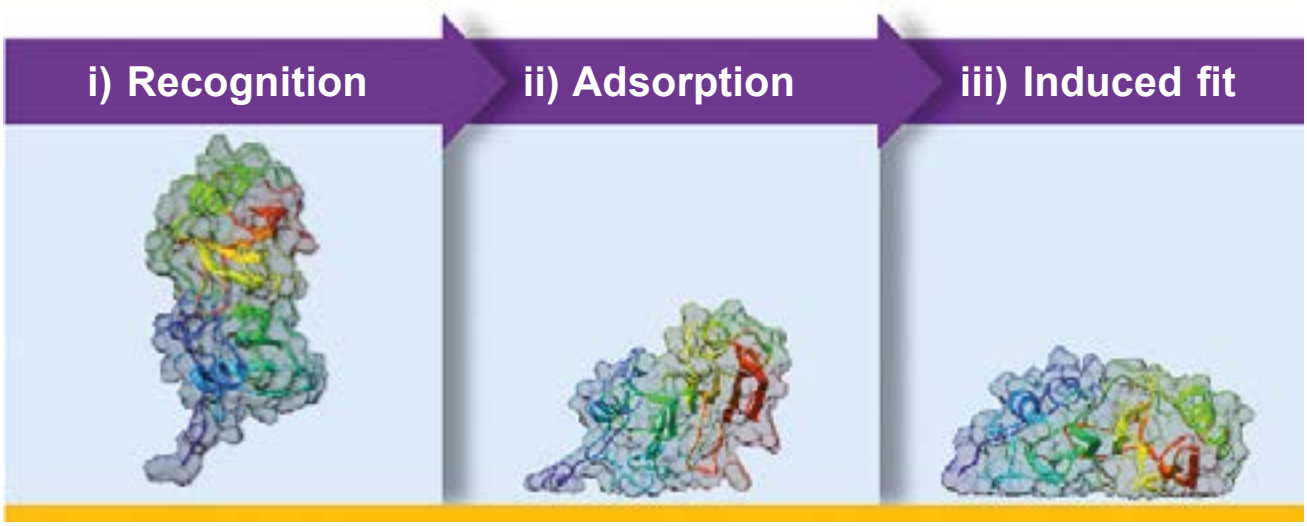
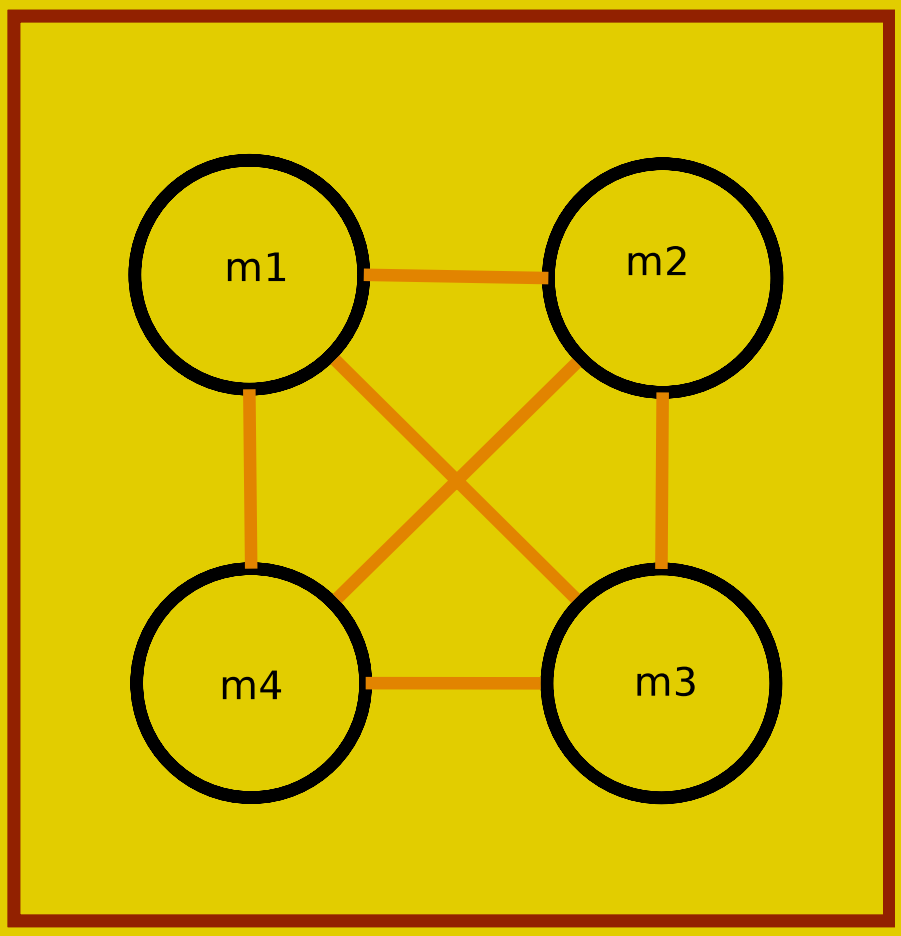
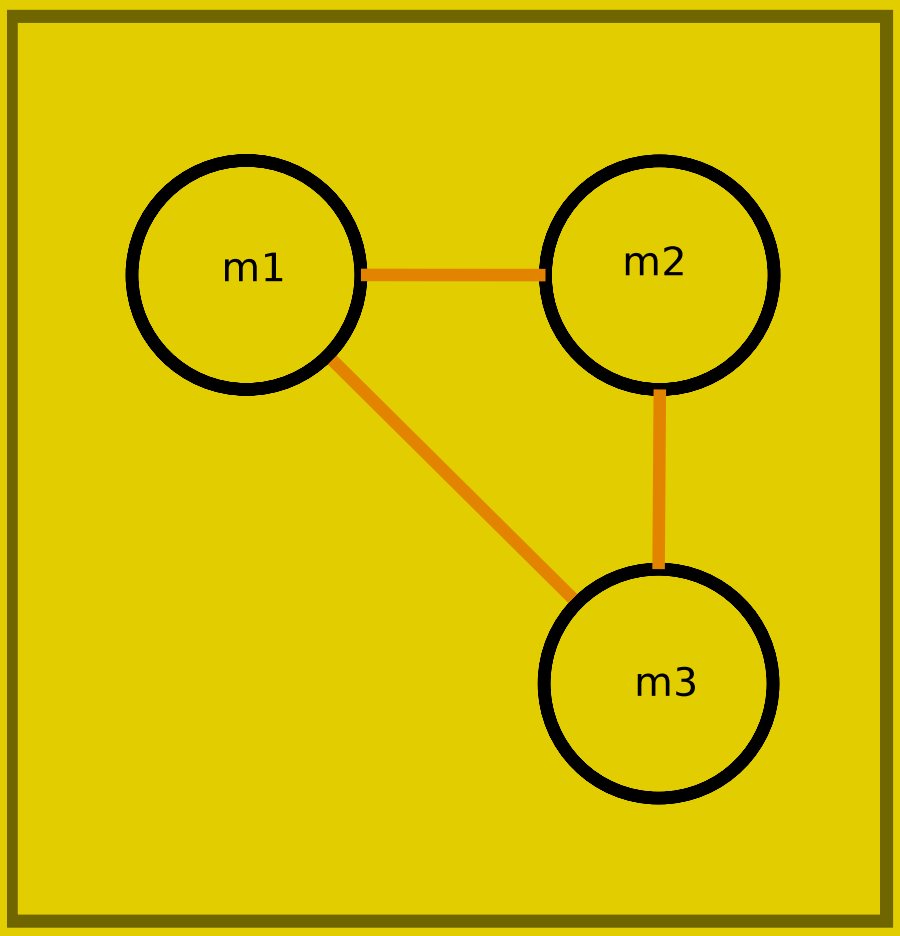


Figure 43: Brownian and molecular dynamics simulations reveal how a model protein adsorbs to a gold surface in three distinct steps [Ozboyaci et al., 2016a].

2 Research



Gold entity



System entity

2.9 Natural Language Processing (NLP)

2.9 Natural Language Processing (NLP)

Natural Language Processing (NLP) is an interdisciplinary research field that lies at the intersection of computer science and linguistics. The NLP group develops methods, algorithms, and tools for the automatic analysis of natural language. The group focuses on discourse processing and related applications, such as automatic summarization and readability assessment.

The year 2016 began with Yufang Hou's successful defense of her dissertation on "Unrestricted Bridging Resolution." Yufang now works at IBM Research in Dublin, Ireland. Later in the year, Sebastian Martschat submitted his thesis entitled "Latent Structures for Coreference Resolution," and Daraksha Parveen submitted her thesis on automatic summarization. Daraksha left HITS in December 2016. In August 2016, we welcomed Feifei Peng, a new PhD student who joined us after obtaining her master's degree at the University of the Chinese Academy of Sciences in Beijing, China.

The other members of the NLP group continued their excellent research, which led to the publication of papers at all major conferences in the field (ACL, NAACL, EMNLP, and COLING). In addition, Alex Judea and Benjamin Heinzerling participated most successfully in shared tasks on event extraction and entity linking at the Text Analysis Conference. PhD students in the NLP group are also encouraged to complete internships in industry. Benjamin Heinzerling spent four months at Microsoft Research Asia in Beijing, China, while Mohsen Mesgar began a six-month internship at SAP in Walldorf, Germany.

Group leader Michael Strube still contemplated the risks of NLP with regard to privacy issues in social media, the abuse of NLP techniques to process unstructured data for spying on an entire population, and the dual use of NLP methods.

Natural Language Processing (NLP) ist ein interdisziplinäres Forschungsgebiet, das mit informatischen Methoden linguistische Fragestellungen bearbeitet. Die NLP Gruppe entwickelt Methoden, Algorithmen und Tools zur automatischen Analyse von Sprache. Die Gruppe konzentriert sich auf die Diskursverarbeitung und verwandte Anwendungen wie automatische Zusammenfassung und Lesbarkeitsbewertung.

Am Anfang des Jahres 2016 verteidigte Yufang Hou ihre Dissertation zum Thema „Unrestricted Bridging Resolution". Yufang arbeitet jetzt bei IBM Research in Dublin, Irland. Im Spätsommer reichte Sebastian Martschat seine Doktorarbeit mit dem Titel „Latent Structures for Coreference Resolution" ein, und Daraksha Parveen stellte ihre Dissertation im Bereich der automatischen Zusammenfassung zum Ende des Jahres hin fertig. Sebastian arbeitet jetzt an der Universität Heidelberg als PostDoc, Daraksha verließ HITS im Dezember. Schon im August 2016 hießen wir Feifei Peng als Doktorandin willkommen. Feifei erwarb ihren Master an der University of the Chinese Academy of Sciences in Peking, China.

Die anderen Mitglieder der NLP Gruppe setzten 2016 ihre erfolgreiche Arbeit fort, was wir mit Publikationen auf allen wichtigen Konferenzen in unserem Gebiet (ACL, NAACL, EMNLP und COLING) dokumentiert haben. Außerdem nahmen Alex Judea und Benjamin Heinzerling wieder mit Erfolg an den Shared Tasks über Event Extraction und Entity Linking der Text Analysis Conference (TAC) teil. Doktoranden der NLP Gruppe sind ermutigt, Praktika in der Industrie zu machen. Benjamin Heinzerling verbrachte vier Monate bei Microsoft Research Asia in Peking, China, und Mohsen Mesgar begann im Herbst ein sechsmonatiges Praktikum bei SAP in Walldorf, Germany.

Der Leiter der Gruppe, Michael Strube, pflegte weiterhin über die Risiken von NLP nachzusinnen, insbesondere über Datenschutz in sozialen Medien, den Missbrauch von NLP-Techniken mit dem Zweck, die Bevölkerung auszuspionieren, und im Allgemeinen die Verwendung von Ergebnissen der NLP-Forschung für militärische und geheimdienstliche Aktivitäten (dual use).

Group leader

Prof. Dr. Michael Strube

Staff members

Sebastian Martschat *(until March 2016)*
Dr. Mark-Christoph Müller

Scholarship holders

Alex Judea
Mohsen Mesgar
Nafise Moosavi
Daraksha Parveen
Feifei Peng

Visiting scientists

Benjamin Heinzerling *(Research Training Group AIPHES)*
Sameer Pradhan *(August – October 2016)*

Students

Raphael Schumann
Julia Suter *(since August 2016)*
Isabell Wolter

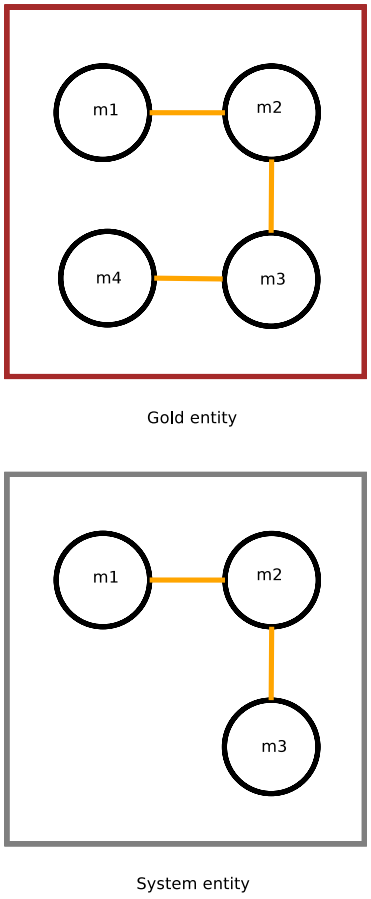


Figure 44: MUC evaluation metric.

LEA: A Reliable Evaluation Metric for Coreference Resolution

Understanding the meaning of the sentence "John tried to call Harry, the CEO of the company, on the phone, but he wasn't available," requires recognizing that "Harry," "the CEO of the company," and "he" refer to the same person. The task of detecting various referring expressions that refer to the same real-world entity is called coreference resolution.

An important step towards developing better coreference resolvers is having a reliable evaluation metric, which is required to determine which model leads to better results and what kind of knowledge can be used to develop more accurate systems. The most common metrics for evaluating coreference outputs developed between 1995 and 2005 are MUC, B-cubed, and CEAF. MUC evaluates coreference outputs based on the minimum number of links that are required to connect the mentions of one entity. Since MUC only considers the minimum number of links in each entity, it is known to be the least discriminative metric for evaluating coreference resolution.

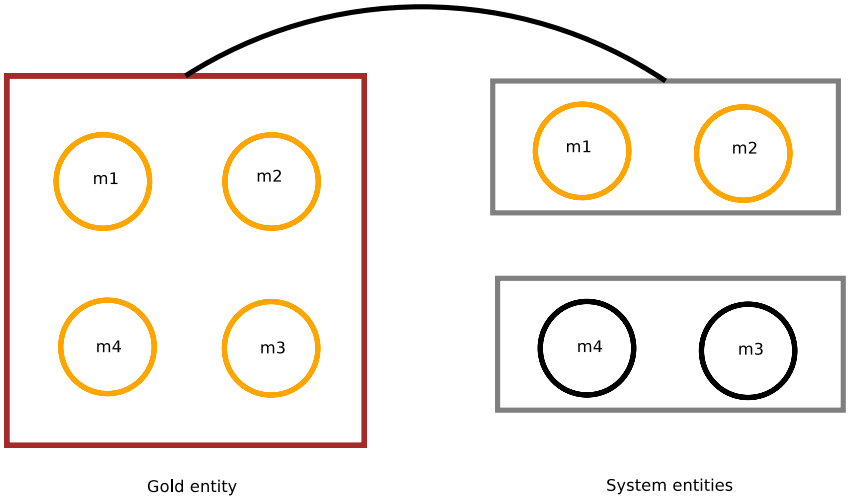


Figure 45: CEAF evaluation metric.

The B-cubed evaluation is based on the number of common mentions between gold and output entities. B-cubed incorrectly rewards system output if one mention appears several times in different entities or if a coreferent mention appears in an incorrect entity.

The CEAF metric maps each gold entity to only one system-generated entity. Therefore, if some coreference relations are missing in the output and a gold entity is split into two or more system entities, only one will be rewarded by CEAF.

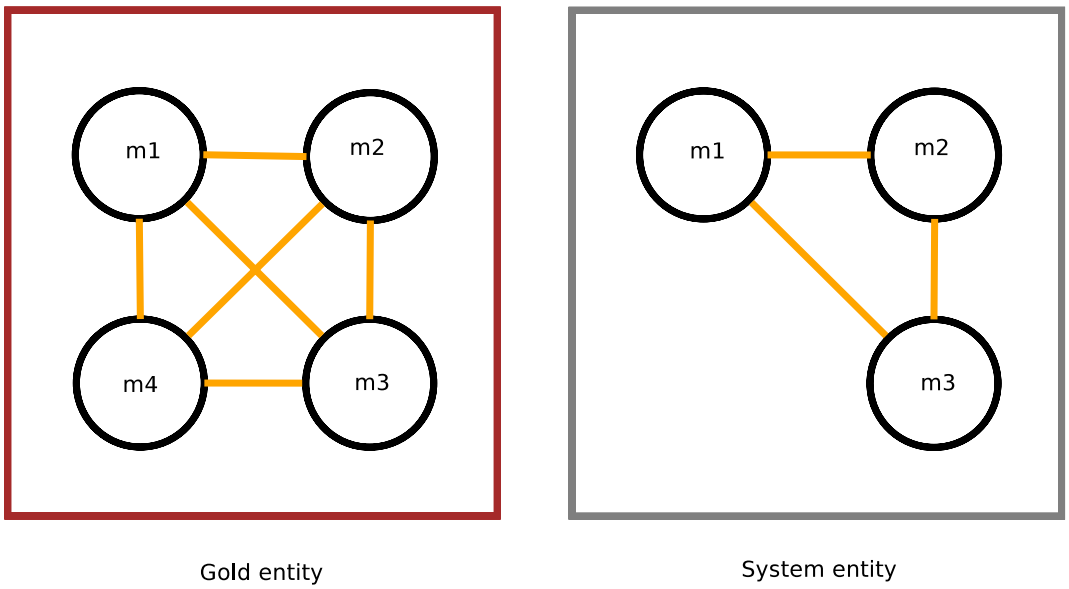


Figure 46: LEA evaluation metric.

We have introduced a new coreference evaluation metric, the Link-based Entity Aware (LEA) metric [Moosavi and Strube, 2016b]. LEA overcomes all known drawbacks of the current evaluation metrics by considering all coreference links in each entity and rewarding a system entity based on the number of resolved coreference links. The number of times that an entity is mentioned in a text is an indicator of its prominence. Therefore, LEA assigns higher weights to the resolution of larger entities when it computes the overall score of all entities.

Coherence Modeling

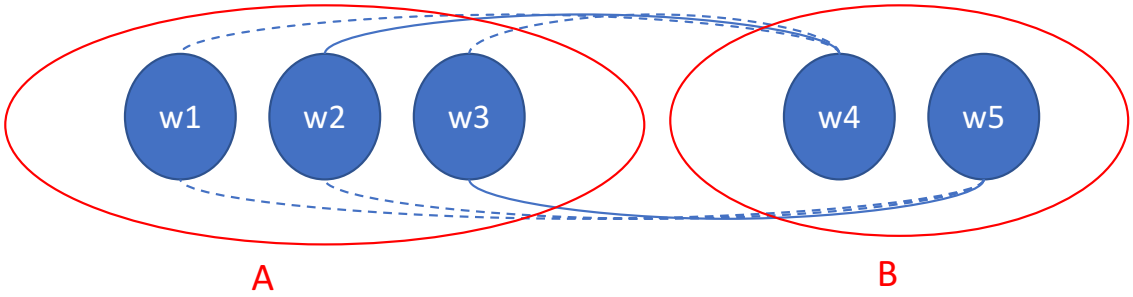
In 2015, we developed a graph-based coherence model based on named entities shared by sentences in a text. Although this model performed very well on the readability ranking task, it suffers from a lack of understanding of the other types of semantic relations between all words in the sentences. A pair of sentences can be semantically related because both sentences contain entities that are related to each other, as in the following example:

"Why does the little boy wriggle all the time? Girls don't."

In this example, the words "boy" and "girls" are semantically related. Although they do not refer to the same entity, they still connect these two sentences. The issue is here is how we can find these types of relations among words. Recent improvements in embedding representations of words allow us to efficiently compute semantic relations among words in the vector space. These representations use a vector of numbers to encode the meanings of words, and this vector representation has the interesting property of being able to store semantically related words close to each other in the vector space.

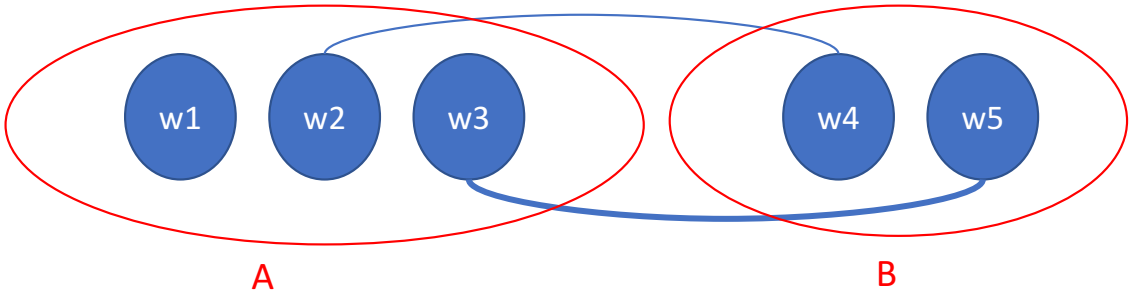
The distance between two word vectors can be measured by the cosine function. If the absolute value of the cosine function is close to 1, the two words are strongly related; if it is close to 0, the words are only weakly related. Given a pair of sentences like A and B, we use pre-trained word embeddings to find semantically related words. We can represent these relations via a graph (*Figure 47*).

Figure 47: Semantically related words.



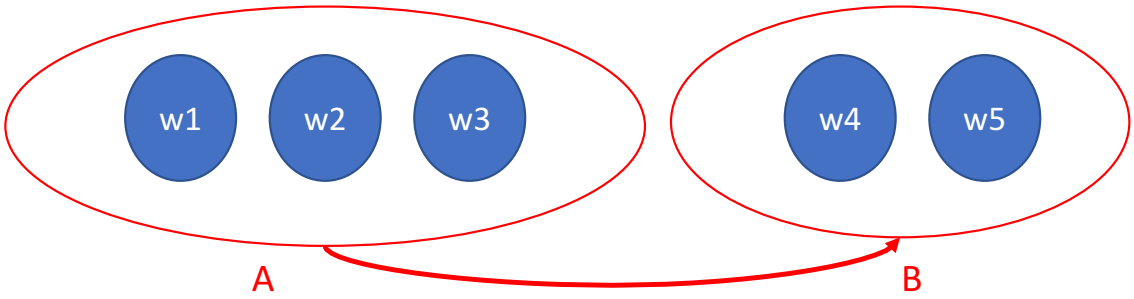
As shown in *Figure 47*, we measure the semantic relatedness between each word of sentence B to that of all words of sentence A with the cosine function. We remove all edges of each word of B except for the one with the maximum cosine value (*Figure 48*).

Figure 48: Maximally semantically related words.



Finally, the word pair with the maximum weight connects the sentences (*Figure 49*). The edge direction here reveals the order of the sentences in the text. If the edge weight is less than a specified threshold, it is eliminated.

Figure 49: Semantically related sentences.



We apply this procedure to all sentence pairs in the text to obtain a graphic representation of the text. The graph's nodes are sentences, and its edges are semantic relations among sentences. We represent the connectivity structure of each graph via a vector whose elements are the frequencies of all possible subgraphs, and we interpret these frequencies as coherence features. It is thereby possible to compare the coherence of a pair of texts using a machine learning model, such as support vector machines.

We evaluate this model on readability ranking, and the task is to rank two texts based on their readability. Texts are collected from the Wall Street Journal and labeled by people who are asked to consider how well sentences are connected. We rank texts based on these people's labels and ask the model to predict the ranks. Our model obtains up to 98 % accuracy on this task, which represents an improvement to the state of the art by a fair margin [Mesgar and Strube, 2016].

Better Entity Linking through Automatic Verification

(Part of this work was completed during a research internship at Microsoft Research Asia.)

Entity Linking (EL) is the task of linking mentions of entities, such as "Dublin", to their corresponding entry in a knowledge base (KB), such as Wikipedia. Since entities mentioned in a text are an important part of what a text is about, EL is an important part of understanding a text.

In a coherent text, entities tend to show high semantic relatedness. For example, a text about a horse race might mention race horses, jockeys, or the racetrack. EL systems exploit coherence to find a mention's correct KB entry by preferring KB entries that are related to other entities mentioned in the text over other candidate KB entries that are less related. For computational reasons, current EL systems only consider a generic notion of coherence, which can be misleading. For example, consider the beginning of this news article:

"DUBLIN 1996-08-31 Result of the Tattersalls Breeders Stakes, a race for two-year-olds run over six furlongs at The Curragh…"

While this article is clearly about a horse race in Ireland, current entity linking systems are keen to link the entity mention "Breeders Stakes" to the Wikipedia entry about a Canadian horse race of the same name since the Irish horse race does not have a Wikipedia entry and the Canadian horse race shows high semantic relatedness to "Tattersalls" (a race horse auctioneer based in the UK and Ireland) and "The Curragh" (a famous horse racetrack in Ireland). To prevent these kinds of errors, we introduce previously ignored aspects of coherence, such as geographic coherence (i.e., entities mentioned in a text tend to be geospatially close) and temporal coherence (i.e., entities mentioned in a text tend to be temporally close).

In order to integrate these specific aspects of coherence into existing EL systems and to avoid intractable optimization, we propose a two-step solution: In the first step, a given EL system links entity mentions in a text to the KB. In the second step, we automatically verify these results by considering specific aspects of coherence. In the example above, geographic outlier detection would identify the link to the Wikipedia entry for the Canadian horse race as a potential error.

Due to this two-step process, our automatic verification method avoids the computational issues that limit current systems and allows for incorporating knowledge-rich aspects, such as geographic coherence, which would otherwise be infeasible (to be published at EACL 2017).

Event Extraction

Event Extraction is a difficult information extraction task in which information about (important) events is sought in texts and is typically specified in the following manner: Given a sentence, find the words that indicate an important event (called "triggers") and find the times, places, and entities that play important roles in these events (the "arguments"). Consider the following example taken from a newswire text:

"We know the Iraqis have fired sand missiles."

Here, one event is expressed through the Trigger "fired", namely an ATTACK event. "Iraqis" is the Attacker in this case (we say the entity fills the Attacker role of the respective event), and "sand missiles" is the weapon being used. An automatic event extraction system would have to recognize the Trigger and all of its arguments. Such a task can quickly become complicated because sentences may contain multiple events, and event arguments may be shared between them, potentially filling different roles in different events. In many cases, looking at one sentence alone is not sufficient when extracting events:

"But the strikes prove controversial."

In this example, "strikes" has no potential event arguments, and the context is not sufficient to disambiguate it correctly: Is it the trigger of an ATTACK event because it actually means bomb strikes, or is it not a trigger at all because it refers to an industrial action?

"Soon after dawn on this fourth day, confirmation of the ship's first strike arrived... This is the first of an unknown number of strikes we'll conduct during our watch in operation enduring freedom... But the strikes prove controversial."

Given the other sentences, it is easier to infer that "strikes" is indeed the trigger of an ATTACK event when analyzed in its surrounding context. Looking at a broader context also benefits argument detection. For example, entities can play coherent roles in different events within one document. Consider the following text:

"Sam Waksal was sentenced to seven years and three months in federal prison… He's being released on his own cog in a sans before he's to report to jail."

Looking only at the individual sentences, it would be difficult to predict that "report" should trigger an ARREST-JAIL event, which in turn would make it difficult to predict that "he" should fill the Person role of this event. However, if the system looks at the entire document and knows that "he" and "Sam Waksal" refer to the same entity, it can better infer that the Defendant of a SENTENCE event can be the Person of an ARREST-JAIL event, which in turn would make it easier to infer that "he" is this Person.

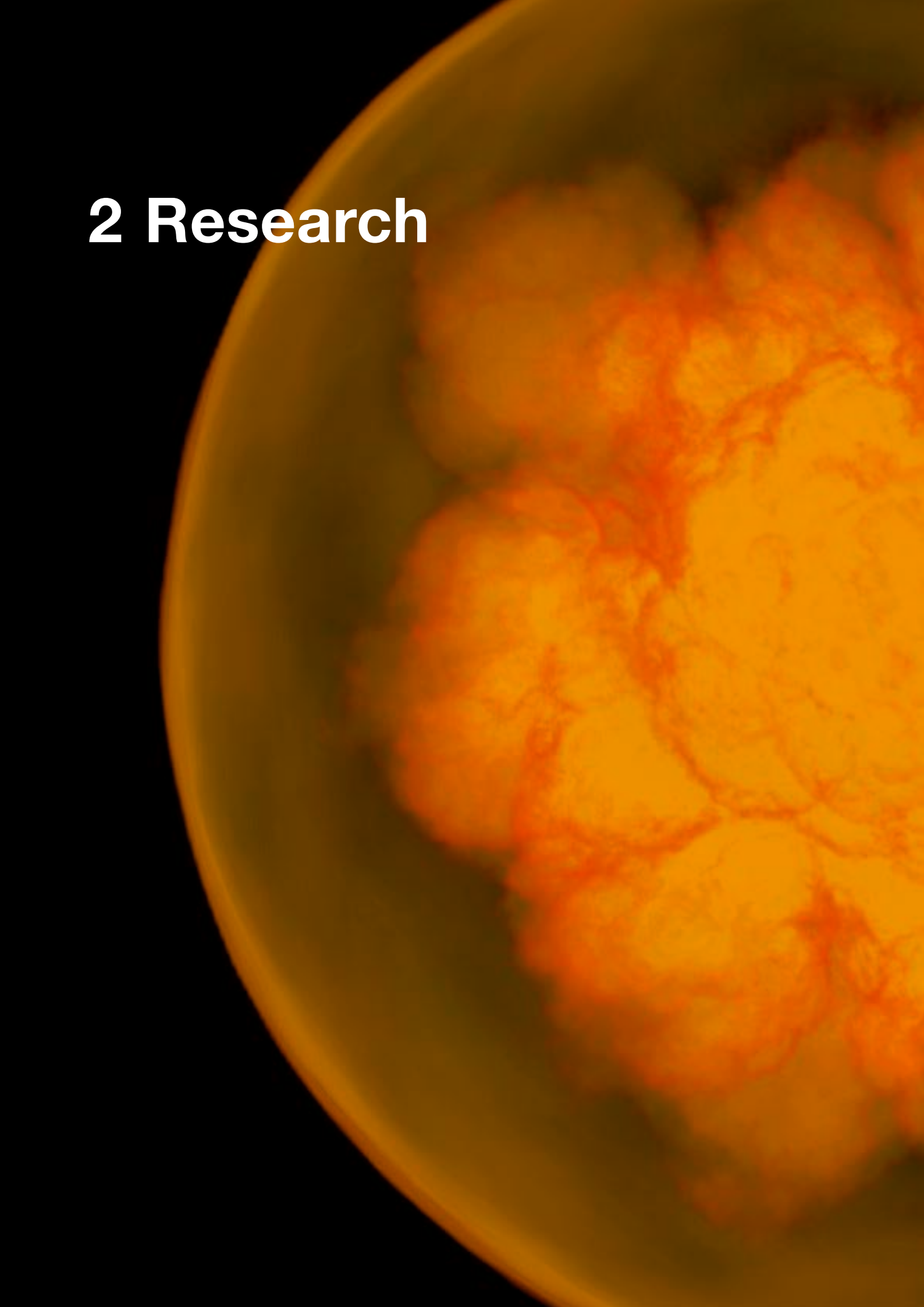
Many state-of-the-art systems focus on intra-sentential extraction. The best extractors jointly infer the entire trigger-argument structure present in a sentence; however, they lack information present in other sentences. Recently, we have developed a method to extend such a joint yet local state-of-the-art extractor to global, document-wide inference [Judea and Strube, 2016]. Our idea is simple: We incorporate information present in other sentences into the inference process of the sentence at hand. For example, when we look at words like "fired" or "strikes" in the above sentence, we ask the system if the same word occurred as a trigger anywhere else in the document, and if so, which event type it triggered; if not, we ask which event types occur in the document. These clues help to refine the event extraction of the actual sentence, which may in turn help to refine extraction in other sentences. Because of this inter-dependency in the inference process, we have to complete multiple inference passes over one document.

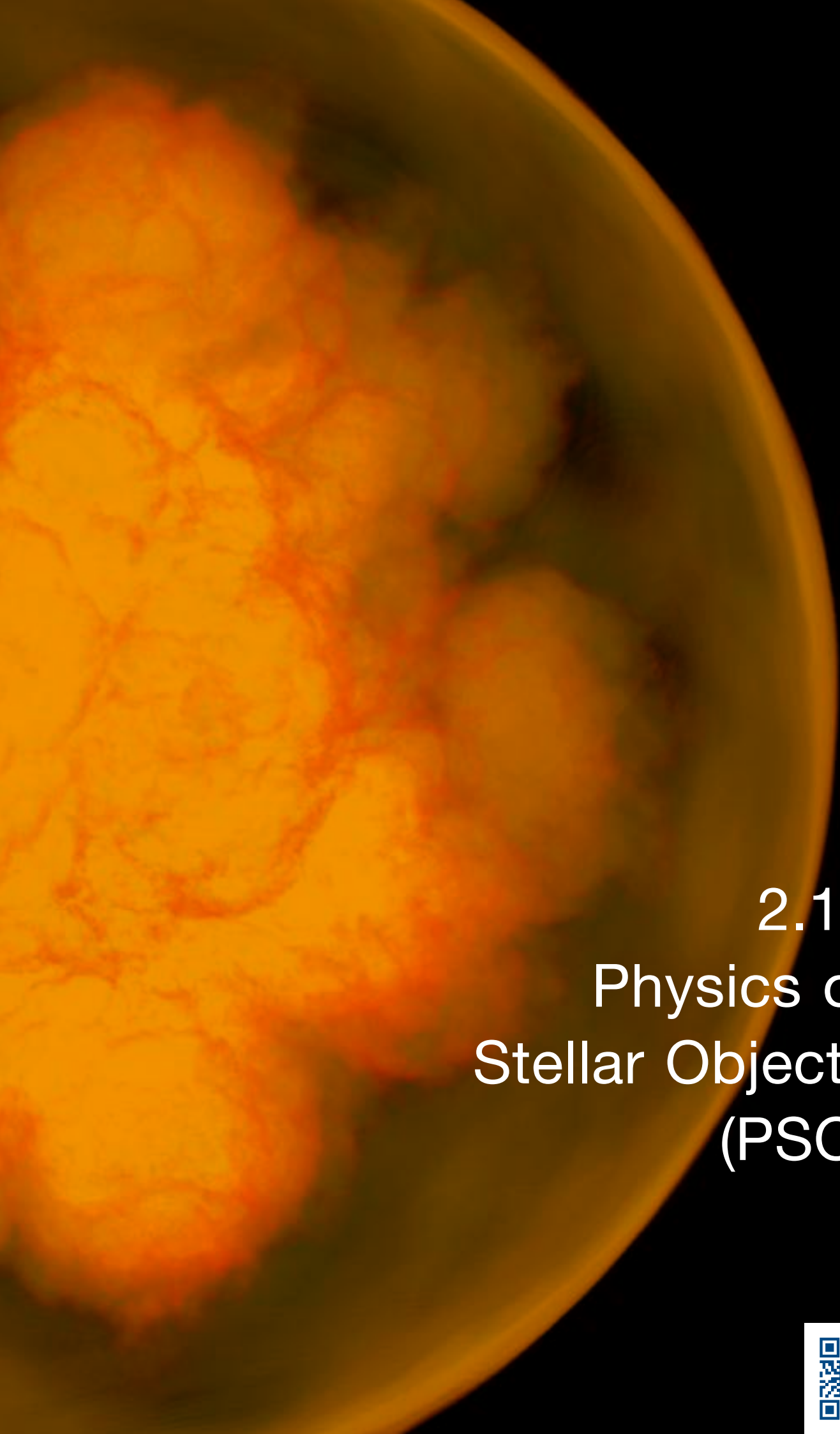
Author Disambiguation

Research on author disambiguation at HITS was conducted under the project "Scalable Author Disambiguation for Bibliographic Databases" (SCAD) in cooperation with DBLP and zbMATH and funded by the Leibniz-Gemeinschaft. Author name ambiguity can be observed when two or more authors bear or publish under the same name such that the author's identity cannot be established on the basis of the name alone. Author disambiguation is a well-established research topic in information and digital library science, computer science, and NLP. Current research is dominated by approaches that deal with author disambiguation in publications from the computer science domain. On the one hand, this is justified given the rapid growth of that field and the ensuing number of publications, which is also the case for our partner DBLP. On the other hand, however, these approaches induce a bias that leads to the development of systems that rely on properties of that domain. One of these properties is rich coauthor information, which is only available in domains with many multi-author publications. In domains such as mathematics, however, researchers have quite different publication habits than in computer science, resulting in a higher proportion of single-author publications. For these publications, author disambiguation methods based on coauthor information do not work well, as has been observed by our partner zbMATH.

Another aspect of author ambiguity that is underrepresented in current research is name variability, which is present when one author appears under several name variants. These variants can comprise full vs. initialized first names, full vs. initialized vs. dropped middle names, and the like. An important source of name variability is different transliterations of names from languages with special alphabets, such as Chinese and Russian. In order to advance author disambiguation research in these directions, the SCAD project has created a new, openly available dataset (https://doi.org/10.5281/zenodo.161333) based on zbMATH data that is meant to be more representative of these aspects. The dataset is the result of our empirical study on existing datasets (to appear in Scientometrics). Other efforts in the project have been directed towards the development of an initial author disambiguation system prototype. In line with the argumentation given above, we want to emphasize the contribution of content-aware NLP methods (as opposed to coauthor-based methods). Based on promising results of state-of-the-art Deep Learning methods in comparable tasks, we have employed methods involving LSTM (Long Short-Term Memory) networks for capturing the semantic similarity between publications that can provide important evidence for author disambiguation beyond coauthor matching. Apart from its proven suitability for NLP-related tasks, Deep Learning also lends itself to our author disambiguation system because it is a general-purpose ML paradigm that can integrate the various facets of the task.

2 Research





2.10 Physics of Stellar Objects (PSO)

2.10
Physics of Stellar Objects (PSO)

"We are stardust" – the matter we are made of is largely the result of processing the primordial material formed in the Big Bang. All heavier elements originate from nucleosynthesis in stars and gigantic stellar explosions. How this material formed and how it is distributed over the Universe is a fundamental concern for astrophysicists. At the same time, stellar objects make the Universe accessible to us by way of astronomical observations. Stars shine in the optical and other parts of the electromagnetic spectrum. They are the fundamental building blocks of galaxies and larger cosmological structures. With the help of extensive numerical simulations, our research group "Physics of Stellar Objects" seeks to understand the processes that occur in stars and stellar explosions. Newly developed numerical techniques and the ever-growing power of supercomputers facilitate the modeling of stellar objects in unprecedented detail and with incomparable precision.

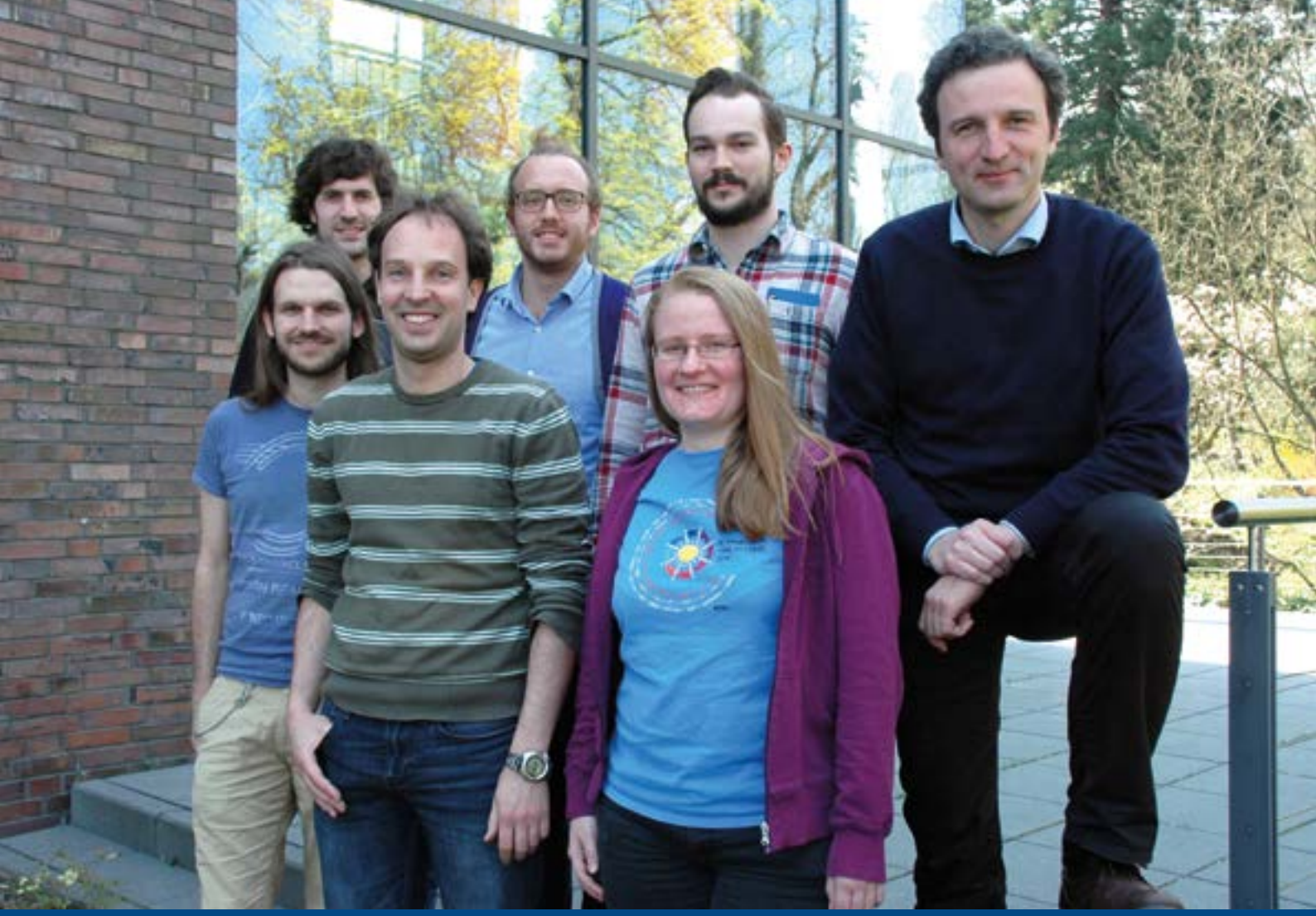
A primary goal of our group is to model the thermonuclear explosions of white dwarf stars that lead to the astronomical phenomenon known as Type Ia supernovae. These are the main sources of iron in the Universe and have been instrumental as distance indicators in cosmology, leading to the spectacular discovery of the accelerating expansion of the Universe. Multi-dimensional fluid-dynamic simulations in combination with nucleosynthesis calculations and radiative transfer modeling provide a detailed picture of the physical processes in Type Ia supernovae but are also applied to other kinds of cosmic explosions.

Classical astrophysical theory describes stars as one-dimensional objects in hydrostatic equilibrium. This approach has been extremely successful, for it explains why stars are observed in different configurations and provides a qualitative understanding of stellar evolution. However, simplifying assumptions limits the predictive power of such models. With newly developed numerical tools, our group explores dynamic phases in stellar evolution in three-dimensional simulations. Our aim is to construct a new generation of stellar models based on an improved description of the physical processes taking place in stars.

„Wir sind Sternenstaub" – die Materie, aus der wir geformt sind, ist zum großen Teil das Ergebnis von Prozessierung des primordialen Materials aus dem Urknall. Alle schwereren Elemente stammen aus der Nukleosynthese in Sternen und gigantischen stellaren Explosionen. Eine fundamentale Frage ist, wie dieses Material geformt wurde und wie es sich im Universum verteilt. Gleichzeitig machen stellare Objekte das Universum für uns in astronomischen Beobachtungen überhaupt erst sichtbar. Sterne scheinen im optischen und anderen Teilen des elektromagnetischen Spektrums. Sie sind fundamentale Bausteine von Galaxien und aller größeren kosmologischen Strukturen. Unsere Forschungsgruppe „Physik stellarer Objekte" strebt mit Hilfe von aufwändigen numerischen Simulationen ein Verständnis der Prozesse in Sternen und stellaren Explosionen an. Neu entwickelte numerische Techniken und die stetig wachsende Leistungsfähigkeit von Supercomputern ermöglichen eine Modellierung stellarer Objekte in bisher nicht erreichtem Detailreichtum und mit großer Genauigkeit.

Ein Hauptziel unserer Gruppe ist die Modellierung von thermonuklearen Explosionen weißer Zwergsterne, die zum astronomischen Phänomen der Supernovae vom Typ Ia führen. Diese sind die Hauptquelle des Eisens im Universum und wurden als Abstandsindikatoren in der Kosmologie eingesetzt, was zur spektakulären Entdeckung der beschleunigten Expansion des Universums führte. Mehrdimensionale strömungsmechanische Simulationen kombiniert mit Nukleosyntheserechnungen und Modellierung des Strahlungstransports ergeben ein detailliertes Bild der physikalischen Prozesse in Typ Ia Supernovae, werden aber auch auf andere Arten von kosmischen Explosionen angewendet.

Die klassische astrophysikalische Theorie beschreibt Sterne als eindimensionale Objekte im hydrostatischen Gleichgewicht. Dieser Ansatz ist extrem erfolgreich. Er erklärt, warum wir Sterne in verschiedenen Konfigurationen beobachten, und liefert ein qualitatives Verständnis der Sternentwicklung. Die hierbei verwendeten vereinfachenden Annahmen schränken jedoch die Vorhersagekraft solcher Modelle stark ein. Mit neu entwickelten numerischen Hilfsmitteln untersucht die Gruppe dynamische Phasen der Sternentwicklung in dreidimensionalen Simulationen. Unser Ziel ist es, eine neue Generation von Sternmodellen zu schaffen, die auf einer verbesserten Beschreibung der in ihnen ablaufenden physikalischen Prozesse basiert.



Group Leader

Prof. Dr. Friedrich Röpke

Staff Members

Dr. Andreas Bauswein
Dr. Philipp Edelmann
Dr. Markus Kromer *(Gliese Fellowship, since October 2016)*
Dr. Sebastian Ohlmann

Scholarship Holders

Leonhard Horst
Dr. Samuel Jones *(Alexander von Humboldt Scholarship)*

Visiting scientist

Aron Michel

Students

Alberto Botto Poala *(since March 2016)*
Sabrina Gronow
Christian Sand *(since September 2016)*

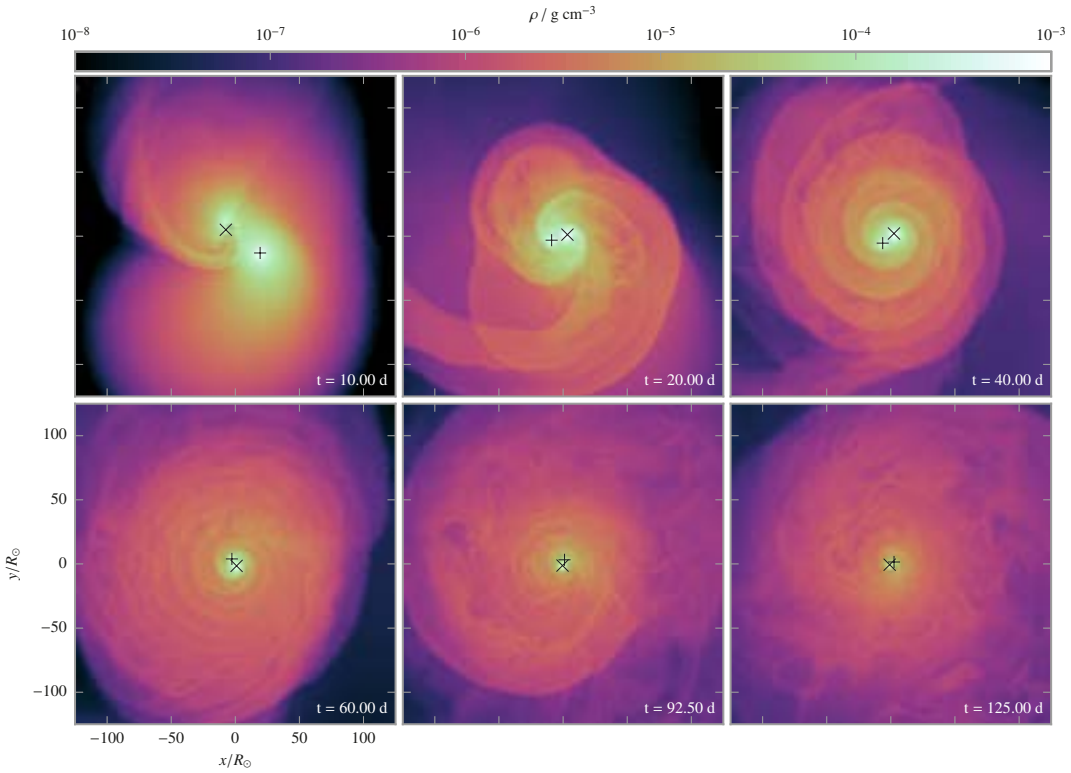


Figure 50: The gas density in the orbital plane is plotted at different times during the simulation. The symbols mark the positions of the companion and the core of the red giant. First, a spiral density wave structure is caused by the spiral-in of the companion. At around 60 days, large-scale instabilities set in, changing the structure of the envelope.

A tight stellar dance

In the night sky, stars appear as single dots of light at large distances from Earth. Stellar solitude, however, is an illusion caused by our eye's limited angular resolution. In fact, more than half of the stars we know have a companion, a second nearby star that can have a major impact on the evolution of the binary star system. The interaction within these binary systems is particularly strong when the two stars involved are passing through a phase in which they are surrounded by a common stellar envelope. Compared with the overall time taken by stars to evolve, this phase is extremely short, so it is difficult to observe and is hence not well understood. After evolving through this phase, the system may give rise to a number of stellar events, such as Type Ia supernovae.

How a common envelope phase emerges

Stars are approximately spherical bodies of gas that radiate light into space. They are held together by their own gravity. The energy of their luminosity is largely generated in the nuclear fusion of hydrogen at their core. However, after a certain point, the star's hydrogen fuel is exhausted. This happens first in the heavier of the two stars. As a consequence of the ceasing nuclear energy production, its core shrinks. At the same time, a highly extended and dilute stellar envelope develops: The star becomes a red giant. As the envelope of the red giant continues expanding, the gravity of the companion star pulls on it, and part of the envelope flows towards it. In the course of this process, the two stars approach each other. Finally, the companion star may fall into the envelope of the red giant, and both stars become surrounded by a "common envelope." As the core of the red giant and the companion star draw closer together, they release gravitational binding energy that is transferred to the common envelope. The envelope is thus ejected and mixes with interstellar matter in the galaxy, leaving behind a close binary star system consisting of the core of the giant and the companion star. Details of the physical processes inside the common envelope as well as the mechanism of envelope expansion remain enigmatic. Because the common envelope phase is dynamically the most important stage in binary star evolution, this severely limits the reliably of theoretical predictions.

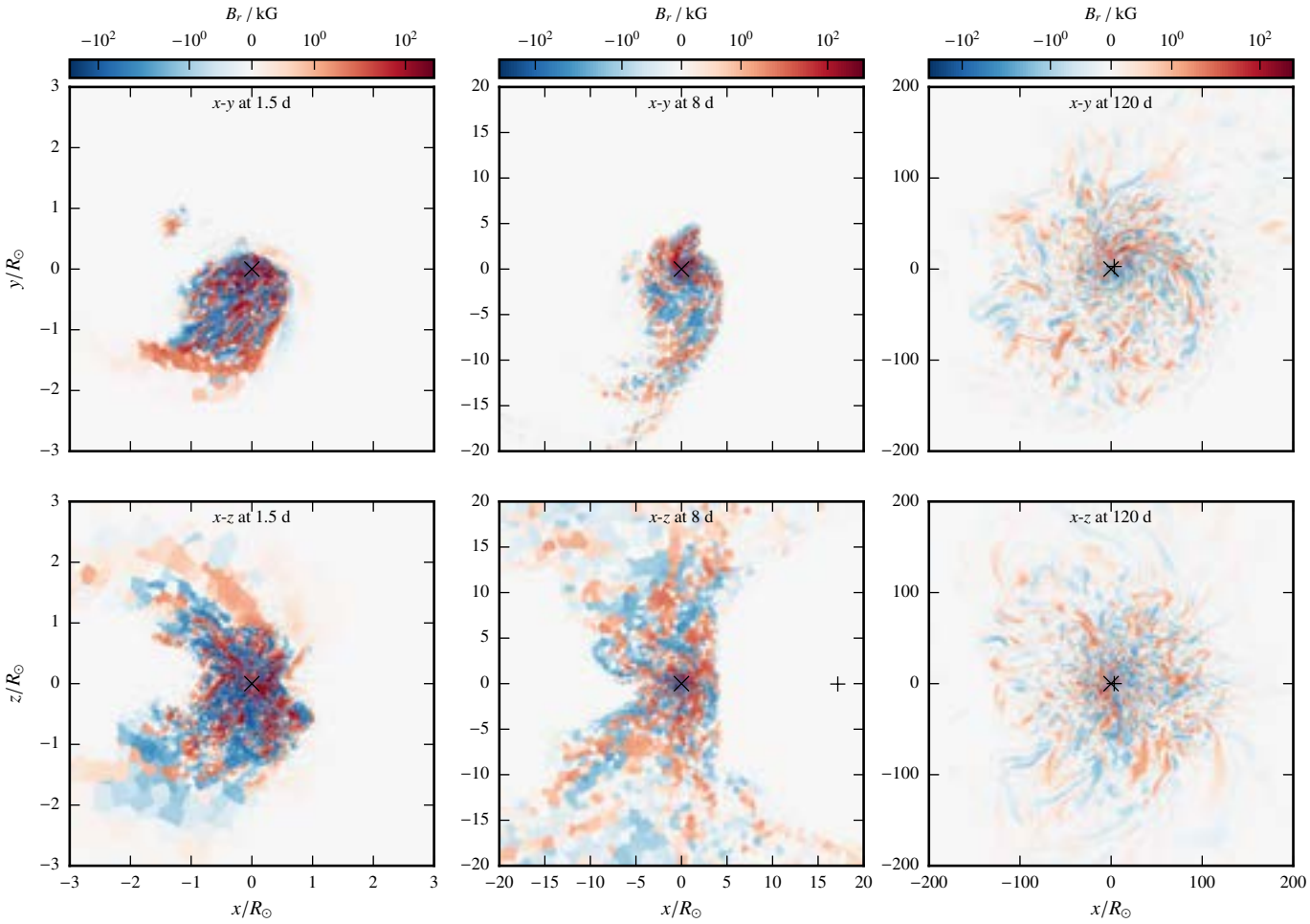


Figure 51: The structure of the radial magnetic field is displayed at different times (from left to right) in the orbital plane (upper row) and perpendicular to the orbital plane (lower row). The symbols mark the positions of the companion and the core of the red giant. Strong magnetic fields appear near the companion in the beginning and are dispersed throughout the envelope after 20 days.

Modeling challenges

Sebastian Ohlmann and Friedrich Röpke from the PSO group tackle this problem from a modeling perspective in collaboration with Rüdiger Pakmor and Volker Springel from the Theoretical Astrophysics group (TAP). They use computationally-intensive simulations to complement observations, but modeling the common envelope phase is difficult because of the wide range in scales: The core of the red giant, for instance, may be up to a million times smaller than its envelope. Utilizing special methods, the astrophysicists adapt and apply the hydrodynamics code AREPO, authored by Volker Springel, to this problem. AREPO solves the equations of fluid dynamics on a moving mesh. This method is well suited for following the dynamics of the gas in the rotating binary star system and thus enhances the accuracy of the modeling.

3D simulations improve flow modeling

With this improved modeling approach, Sebastian Ohlmann's team was able to demonstrate its ability to simulate the dynamic spiral-in of the companion through the envelope of the primary giant star in unprecedented detail [Ohlmann et al., 2016a]. Similar to previous results of other researchers, Ohlmann's team found that only a small part of the envelope is ejected during the simulation. The evolution of the density during the 3D hydrodynamic simulation is shown in *Figure 50*. The accuracy of the simulation code, however, is dramatically improved compared with previous approaches. For the first time, dynamic instabilities in the flow structure can be identified and resolved in exquisite detail. These indicate the onset of convective motions in the envelope that alter the mechanism of energy is transport through the envelope. Hence, they influence the further evolution of the system. The simulations performed by

the HITS team mark a new quality of modeling common envelope phases and set the current standard in the field. In a related research project, HITS scientists collaborated with Ewald Müller from the Max Planck Institute for Astrophysics in Garching to study the behavior of magnetic fields during the common envelope phase [Ohlmann et al., 2016b]. The world's first magneto-hydrodynamic simulations of the common envelope phase show a strong amplification of magnetic fields by many orders of magnitude. As can be seen in *Figure 51*, the magnetic fields concentrate mainly in the region around the in-spiraling companion star. Although the resulting field strengths are not large enough to explain the creation of highly magnetic white dwarfs, they may be detected in future observations of stars undergoing a common envelope phase.

The two projects are part of Sebastian Ohlmann's PhD thesis [Ohlmann 2016] that was successfully defended at Heidelberg University in July 2016 under the supervision of Friedrich Röpke. In addition to the main results described above, the thesis treats approximations of giant star models in more detail and covers thorough tests of the numerical method with different mesh resolutions. Ohlmann's thesis also includes results from simulations with the AREPO code that treat the ionization state of the stellar gas: Recombination from ions in the plasma state to atoms can release energy when the envelope expands and cools. This energy may help to unbind and finally eject the envelope, thus potentially explaining the long-debated ejection mechanism.

A journey to the "Bermuda Triangle" of stellar evolution

The superbly energetic explosions of stars, supernovae represent a final rite of passage in astrophysics. The archetypal supernova explodes with about 10^{44} J of kinetic energy, equivalent to about 2×10^{25} gigatons of TNT. However, the source of the supernova's energy depends on the type of supernova.

Two mechanisms of stellar explosions

The most common types of supernovae are by far type Ia and type IIP supernovae, which are spectral classifications pertaining to the most common thermonuclear and core-collapse supernovae, respectively. The former are powered by energy liberated from the atomic nucleus via the fusion of carbon and oxygen into iron and nickel. Precisely which kinds of stellar systems are predominantly responsible for type Ia supernovae remains unclear, but it is known that these systems consist of compact white dwarf stars composed primarily of carbon and oxygen and supported against the force of their own gravity by the repulsive quantum properties of electrons at high density and low temperature (degeneracy pressure). Such objects are the end-products of low- and intermediate-mass stars, of which our own Sun is an example. The conversion of fuel to ashes and, hence, the liberation of nuclear binding energy propagates from its ignition spot as a nuclear flame travelling at either sub-sonic (deflagration) or super-sonic (detonation) speeds. Simulations of deflagrations that transition to detonations have been most successful in explaining the observed characteristics of type Ia supernovae.

Core-collapse supernovae, on the other hand, are the deaths of massive stars with more than about ten times the mass of the Sun. The cores of these stars become hot enough that their initial composition of hydrogen and helium is eventually converted entirely into iron-group elements by the process of nuclear fusion. The process of nuclear fusion liberates energy, enabling the star to maintain its internal pressure gradient that supports it against the force of its own gravity. Fusion ceases to produce a positive energy yield after the formation of the iron core because iron-group elements have the most tightly-bound nuclei. The core instead relies on the degeneracy pressure of the electrons for support against gravity. At the extreme densities in the iron core (about 10^{13} kg m^{-3}), however, electrons are forced inside of nuclei (inverse β-decay), which convert protons into neutrons. Thus, the electrons become fewer and the core implodes, collapsing

into a proto-neutron star and liberating about 10^{46} J of gravitational energy, the majority of which is converted into neutrinos. A fraction of this energy, roughly 1 %, is thought to be re-absorbed into the layer of high-density matter surrounding the proto-neutron star, thereby yielding the canonical 10^{44} J (or, as it is more commonly referred to, 10^{51} erg) supernova.

The unknown fate of "light massive" stars

Around the same time that a deflagration wave was proposed as the mode of nuclear energy release in type Ia supernovae, it was also discovered that a similar deflagration wave could be ignited in compact and electron-degenerate stars composed of oxygen and neon. These are the big brothers of carbon-oxygen white dwarfs and are produced in stars with initial masses roughly 7 to 10 times the mass of the Sun. The twist is that this deflagration wave is ignited by electron capture by ^{20}Ne (about 35 % of the composition by mass), which occurs at a characteristic density of about 10^{13} kg m^{-3}, similar to that encountered in core-collapse supernovae. Owing to the density's being so high, the fate of these stars lies with the victor of a delicate balance between nuclear energy liberation at the deflagration front and electron capture by its ashes. It is this balance that determines whether the star "implodes" or explodes, i.e., whether the star will die as a core-collapse supernova and a neutron star or via a thermonuclear explosion.

Three-dimensional simulations with unexpected results

The last few decades have seen these so-called electron-capture supernovae or accretion-induced collapses attributed to quite a number of observational phenomena, all of which require that they indeed implode rather than explode. The devil, however, is always in the details. The PSO group has conducted simulations of the deflagration wave in high-density, compact oxygen-neon stars at HITS in order to determine whether electron-capture superno-

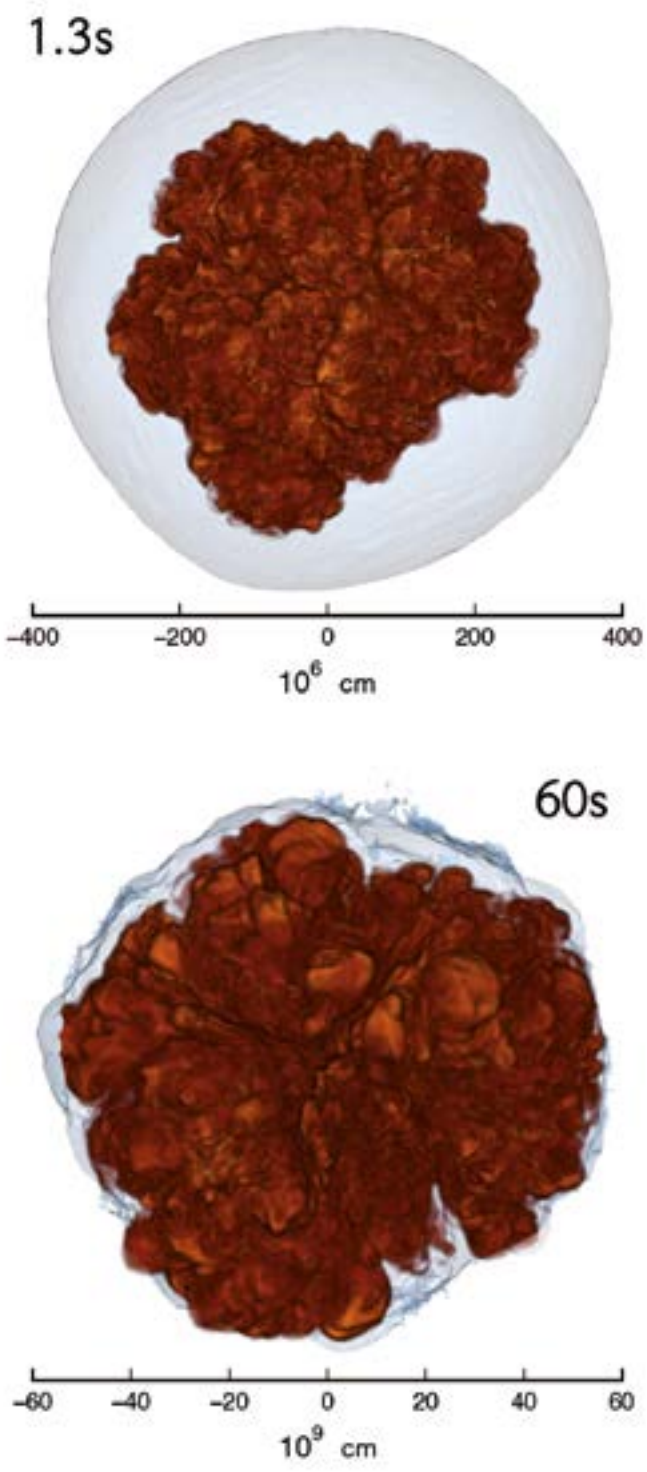


Figure 52: Simulation of a deflagration wave in an oxygen-neon white dwarf star. The light grey surface is a density isocontour demarcating the surface of the star. The iron-rich ashes of the deflagration are colored red.

vae and accretion-induced collapse events are indeed implosions in which a neutron star is produced or whether the liberation of nuclear binding energy by the flame is enough to evade this fate and yield a thermonuclear explosion [Jones et al., 2016a]. The group found that converse to the current paradigm, the oxygen deflagration wave is able to gravitationally unbind up to 70 % of the star's mass and that gravitational collapse only occurs when the central density of the progenitor star is twice as high as is currently thought to be necessary.

These findings raise several questions regarding the contribution of electron-capture supernovae to the neutron star population, galactic chemical evolution, and a subclass of especially faint, low-energy type IIP supernovae. In actual fact, it is possible that accretion-induced ignition of fuel in oxygen-neon white dwarfs could instead comprise a subset of faint, peculiar type Ia supernovae, of which a growing number are being observed. In order to study this possibility further, synthetic observables must be computed. This computation requires detailed radiative transfer calculations, which is a work-in-progress for the PSO group.

The deflagration wave initially propagates by conduction, which is very efficient at transporting heat in degenerate media owing to the large mean free path of the electrons. This is a consequence of the high-density quantum effects alluded to earlier.

In this regime, the flame is said to be laminar, and its speed is determined purely by the local thermodynamic properties of the fuel into which it is burning. Nuclear fusion at the deflagration front injects a vast amount of energy (6×10^{16} J kg^{-1}) into the fluid, which feeds back into the hydrodynamic flow, cascading down to the smallest scales of turbulence. The interaction of turbulence with the flame front can, depending upon its severity, accelerate the rate of fuel consumption by the flame by increasing its surface area. *Figure 53* displays the average, maximum, and minimum laminar and turbulent flame speeds in our simulations over the first 1.5 seconds, which is when burning takes place. The flame burns in the laminar regime for the first 0.4 seconds, where turbulent effects begin to become significant, until its speed is completely dominated by turbulent feedback from 0.6 seconds onwards.

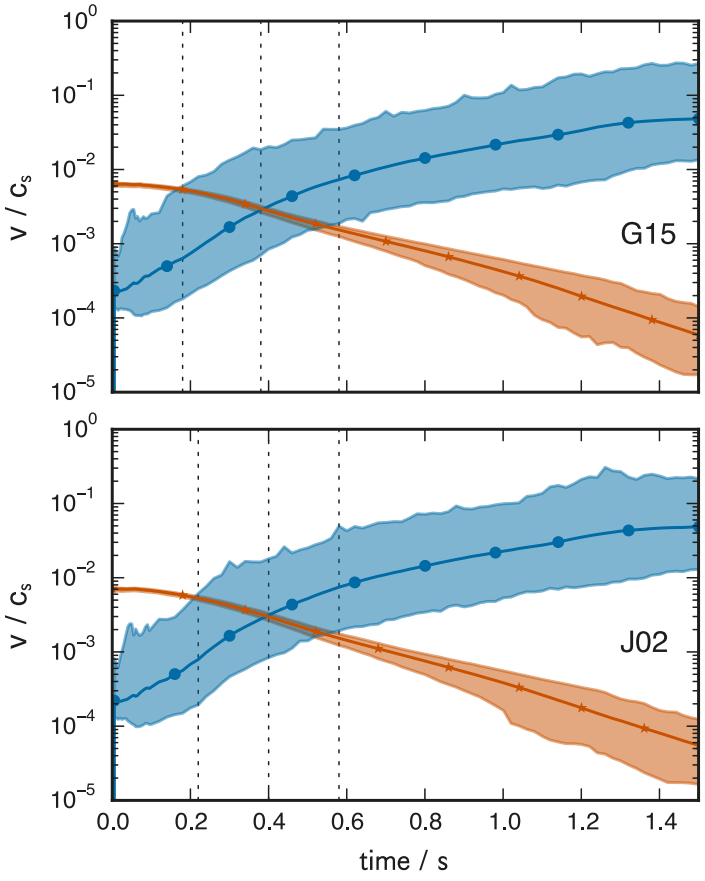


Figure 53: Average laminar (red line) and turbulent (blue line) flame speeds over the oxygen-neon deflagration front as a function of time from ignition; cs is the local speed of sound. The shaded areas provide the range.

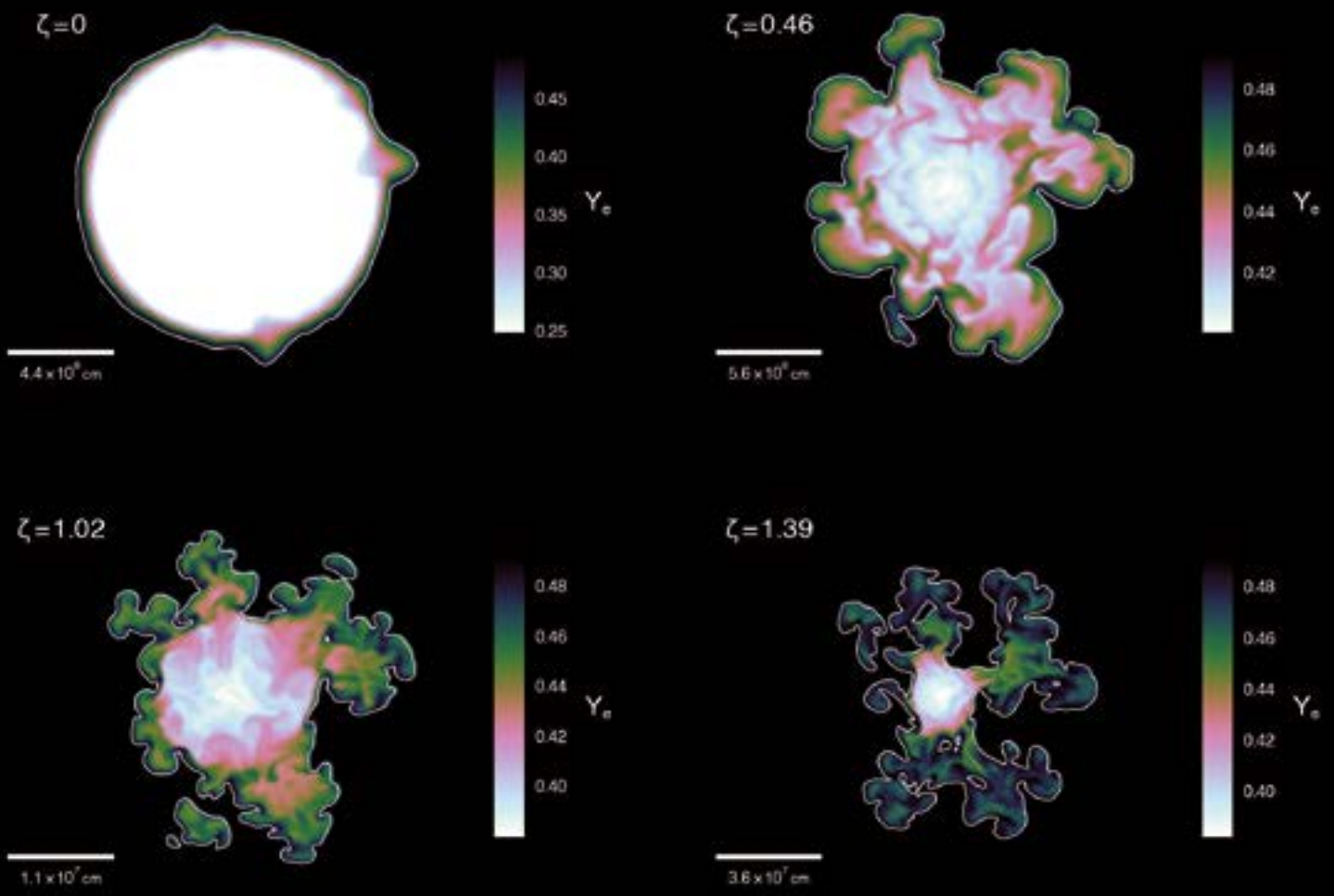
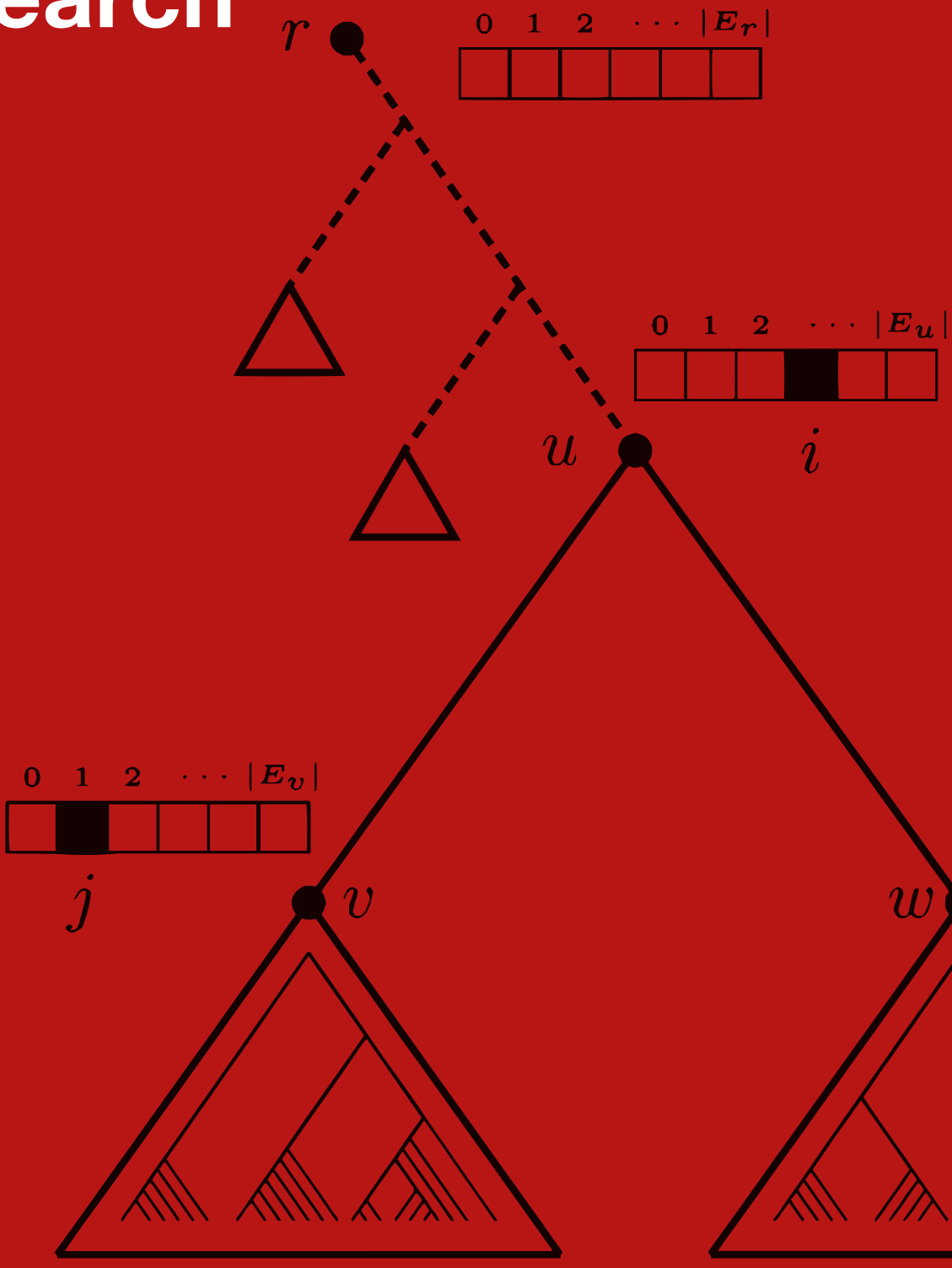


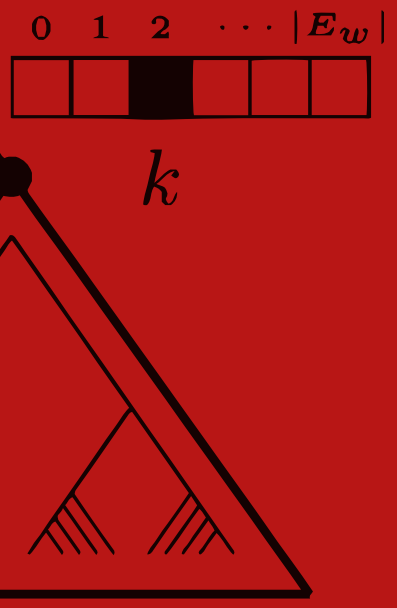
Figure 54: Varying degrees of flame asymmetry in simulated deflagration waves and the value of a diagnostic asymmetry parameter ζ to which they correspond; Y_e is the number ratio of electrons to baryons. Lower Y_e and the suppression of the Rayleigh-Taylor instability (symmetrical flame front) favor the collapse of the white dwarf into a neutron star, which is the fate of the simulation in the top left panel.

A second but equally important facet of feedback from nuclear energy liberation at the flame front into the fluid flow is the activation of the Rayleigh-Taylor instability. This occurs when the increased pressure resulting from the internal energy gain of the flame's ashes produces a density inversion across the flame front, i.e., the ashes "beneath" the fuel have a lower density than does the overlying fuel. This situation is unstable and leads to the upwards acceleration of ash pockets by buoyancy forces if the density inversion persists long enough. The growth of this instability also enhances the rate of energy liberation by increasing the surface area of the flame. Whether or not the instability can grow in this situation is determined by two factors: First, if the laminar flame speed is very fast, density perturbations that seed the instability will quickly be erased (so-called "fire-polishing"); second, electron captures by nuclei in the ashes reduce the gas pressure significantly, thereby suppressing the potential expansion due to the addition of thermal energy. Both factors, which can inhibit the growth of the Rayleigh-Taylor instability, are more pronounced at high density.

Figure 54 displays four slices through 3D simulations of the flame corresponding to various degrees of flame asymmetry. The Rayleigh Taylor instability has grown in all but the upper left panel, in which the electron number fraction is also lowest. Indeed, this is the simulation in which the progenitor white dwarf had the highest central density (2×10^{13} kg m^{-3}) and is the only one of the simulations that collapses into a neutron star, doing so within a prompt 350 ms. Such high densities are likely not achieved in these stars, and we have therefore proposed that the collapse into a neutron star will likely be avoided. ■

2 Research





2.11 Scientific Computing (SCO)

2.11
Scientific Computing (SCO)

The Scientific Computing Group (SCO) focuses on developing algorithms, computer architectures, and high-performance computing solutions for bioinformatics.

We mainly focus on
- computational molecular phylogenetics,
- large-scale evolutionary biology data analyses,
- supercomputing,
- quantifying biodiversity,
- next-generation sequence-data analyses, and
- scientific software quality & verification.

Secondary research interests include but are not limited to
- emerging parallel architectures (GPUs, Xeon PHI),
- discrete algorithms on trees, and
- population genetics.

In the following section, we outline some highlights of our research activities in 2016. Our research is situated at the interface(s) between computer science, biology, and bioinformatics. Our overall goal is to devise new methods, algorithms, computer architectures, and both freely available and accessible tools for molecular data analysis. In other words, our overarching goal is to enable research. One aim of evolutionary biology is to infer evolutionary relationships between species and the properties of individuals within populations of the same species.

In modern biology, evolution is a widely accepted fact and can nowadays be analyzed, observed, and tracked at the DNA level.

A famous dictum widely quoted in this context comes from evolutionary biologist Theodosius Dobzhansky: "Nothing in biology makes sense except in the light of evolution."

Die Gruppe wissenschaftliches Rechnen (SCO) beschäftigt sich mit Algorithmen, Hardware-Architekturen und dem Hochleistungsrechnen für die Bioinformatik.

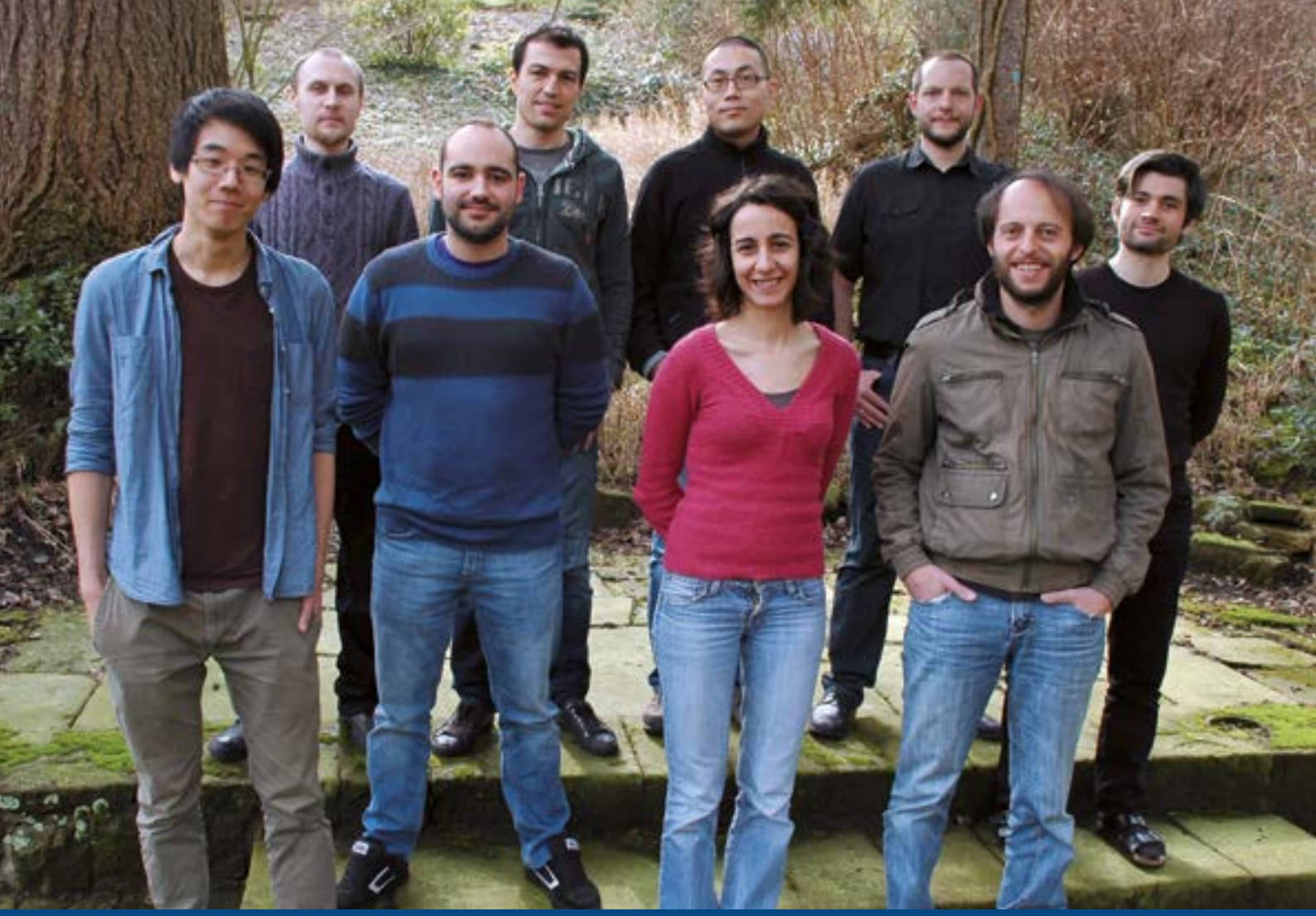
Unsere Hauptforschungsgebiete sind:
- *Rechnerbasierte molekulare Stammbaumrekonstruktion*
- *Analyse großer evolutionsbiologischer Datensätze*
- *Hochleistungsrechnen*
- *Quantifizierung von Biodiversität*
- *Analysen von „Next-Generation" Sequenzdaten*
- *Qualität & Verifikation wissenschaftlicher Software*

Sekundäre Forschungsgebiete sind unter anderem:
- *Neue parallele Rechnerarchitekturen (GPUs, Xeon PHI)*
- *Diskrete Algorithmen auf Bäumen*
- *Methoden der Populationsgenetik*

Im Folgenden beschreiben wir unsere Forschungsaktivitäten. Unsere Forschung setzt an der Schnittstelle zwischen Informatik, Biologie und Bioinformatik an. Unser Ziel ist es, Evolutionsbiologen neue Methoden, Algorithmen, Computerarchitekturen und frei zugängliche Werkzeuge für die Analyse molekularer Daten zur Verfügung zu stellen. Unser grundlegendes Ziel ist es daher, Forschung zu ermöglichen. Die Evolutionsbiologie versucht die evolutionären Zusammenhänge zwischen Spezies sowie die Eigenschaften von Populationen innerhalb einer Spezies zu berechnen.

In der modernen Biologie ist die Evolution eine weithin akzeptierte Tatsache und kann heute anhand von DNA analysiert, beobachtet und verfolgt werden.

Ein berühmtes Zitat in diesem Zusammenhang stammt von Theodosius Dobzhansky: „Nichts in der Biologie ergibt Sinn, wenn es nicht im Licht der Evolution betrachtet wird."



Group Leader

Prof. Dr. Alexandros Stamatakis

Staff Members

Dr. Diego Darriba
Dr. Tomáš Flouri
Dr. Paschalia Kapli
Alexey Kozlov

Scholarship Holders

Pierre Barbera *(HITS Scholarship)*
Lucas Czech *(HITS Scholarship)*
Benoit Morel *(HITS Scholarship)*

Visiting Scientists

Khouloud Madbouh
(July / August & November / December 2016)
Nikos Psonis
(September – December 2016)

Students

Sarah Lutteropp
Nico Schweiger

What happened at the Lab in 2016?

In the winter of 2015/2016, Alexis, Tomas, Alexey, Kassian, Diego, and Lucas taught the "Introduction to Bioinformatics for Computer Scientists" class at the Karlsruhe Institute of Technology (KIT). As in previous years, we received positive teaching evaluations from the students (with a learning quality index of 100 out of 100; see http://sco.h-its.org/exelixis/web/teaching/courseEvaluations/Winter15_16.pdf).

In July 2016, Tomas and Alexis received an award based on student evaluations for teaching excellence from the Dean of the Faculty of Computer Science at KIT for the "Hands-On Bioinformatics Practical" taught in summer 2015.

Three students of the preceding practical programming course in winter 2014/2015 received a prize for student research at the 2016 KIT Campus Day for the peer-reviewed journal paper based on the practical [Hoff et al., 2016].

Finally, in summer 2016, we also taught our main seminar, "Hot Topics in Bioinformatics," with a record number of participants. Student numbers in our course at KIT further increased in the winter term 2016/17.

The year was also very important for our former PhD student Kassian Kobert, who successfully defended his PhD thesis at the computer science department of KIT in May 2016. By that time, Kassian had already begun working as a postdoc in Lyon, France. Our current postdoc, Diego Darriba, also successfully defended his PhD thesis in April 2016 at the computer science department of the University of A Coruna in Spain.

In May 2016, Pierre Barbera submitted an outstanding master's thesis in computer science at KIT on the efficient implementation of phylogenetic placement methods and joined our lab as new PhD student in the fall. We were also happy to welcome our new PhD student Benoit Morel from France, who had previously worked as a scientific programmer in the industry.

In 2016, we hosted two PhD students via our visiting PhD student program. Nikos Psonis, a biologist from the University of Crete, stayed with us from September to December 2016. Khouloud Madbouh, a computer scientist from the University of Tunis, worked with us in July and August and returned for a second visit in mid-November 2016.

Another highlight in 2016 was the course on computational molecular evolution, for which Alexis again served as the main organizer. The course took place for the 8th time at the Hellenic Center for Marine Research in Heraklion, Crete. Paschalia and Alexey also played significant roles as teaching assistants (see *Chapter 5.1.1*). At the end of the year, we had to say goodbye to Paschalia, who went on to begin a second postdoc in London.

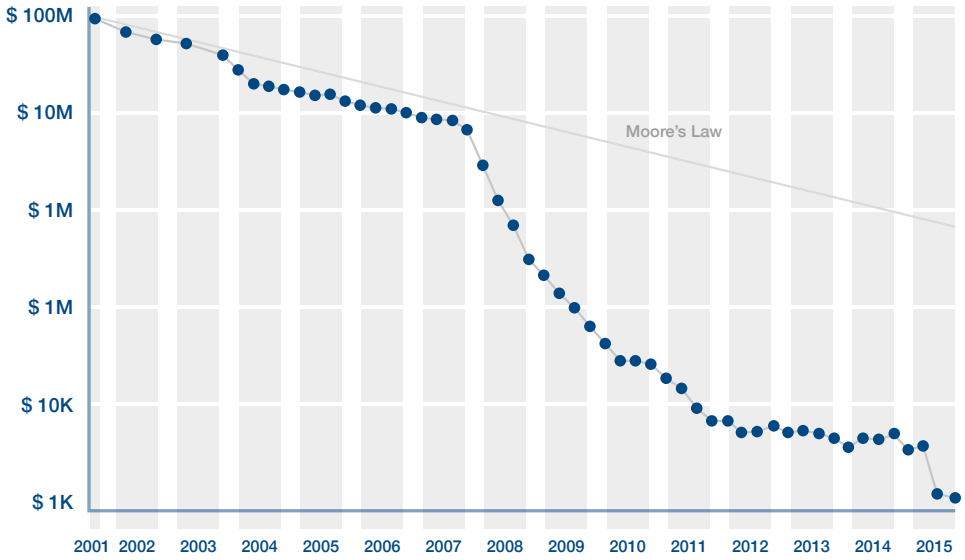
Introduction

The term "evolutionary bioinformatics" refers to computer-based methods for reconstructing evolutionary trees from DNA or from, say, protein or morphological data. The term also refers to the design of programs for estimating statistical properties of populations, that is, for disentangling evolutionary events within a single species. The very first evolutionary trees were inferred manually by comparing the morphological characteristics (traits) of the species under study. Nowadays, in the age of the molecular data avalanche, the manual reconstruction of trees is no longer feasible. Evolutionary biologists thus have to rely on computers for phylogenetic and population-genetic analyses.

Following the introduction of so-called short-read sequencing machines (machines used in the wet-lab by biologists to extract DNA data from organisms) that can generate over 10,000,000 short DNA fragments (each containing between 30 and 400 DNA characters), the community as a whole is facing novel challenges. One key problem that needs to be tackled is the fact that the amount of molecular data available in public databases is growing at a significantly faster rate than the computers capable of analyzing the data can keep up with.

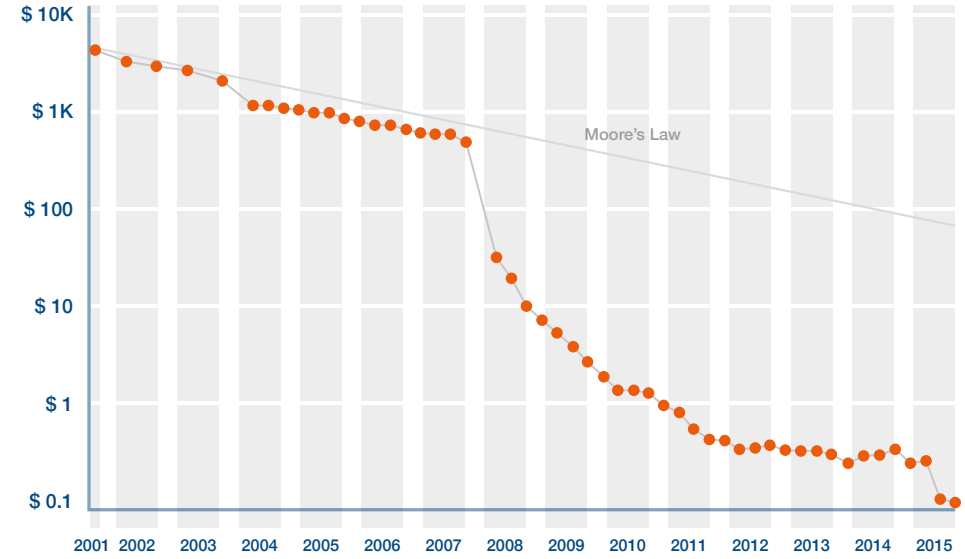
Cost per Genome

Source: National Human Genome Research Institute (NHGRI).



Cost per Raw Megabase of DNA Sequence

Source: National Human Genome Research Institute (NHGRI).



In addition, the cost of sequencing a genome is decreasing at a faster rate than is the cost of computation. This gap further widened in 2016 (see https://www.genome.gov/sequencingcosts/).

Accordingly, as computer scientists, we are facing a scalability challenge, that is, we are constantly trying to catch up with the data avalanche and make molecular data analysis tools more scalable with respect to dataset sizes. At the same time, we also want to implement more complex and hence more realistic and computer-intensive models of evolution. Another difficulty of evolutionary bioinformatics is that next-generation sequencing technology is changing rapidly. Accordingly, the output of these machines in terms of the length and quality of the sequences they can generate is also changing. This requires the continuous development

of new algorithms and tools for filtering, puzzling together, and analyzing these molecular data. For instance, Sarah is currently working on a novel, unified framework for correcting errors in sequence data.

Phylogenetic trees (evolutionary histories of species) are important in many domains of biological and medical research. The programs for tree reconstruction that were developed in our lab can be deployed to infer evolutionary relationships among viruses, bacteria, green plants, fungi, mammals, etc. In other words, they are applicable to all types of species.

In combination with geographical and climate data, evolutionary trees can be used, e.g., to disentangle the geographical origin of the H1N5 viral outbreak, determine the correlation between the frequency of speciation events (species diversity) and climatic changes in the past, and analyze microbial diversity in the human gut.

Based on the prolegomena, one key challenge for computer science is scaling existing analytic methods to the huge new datasets produced by next-generation sequencing methods. Moreover, these codes need to be capable of leveraging the computational resources provided by supercomputers. In 2016, we focused on re-designing and re-implementing some of our most widely used tools from scratch, namely RAxML and Modeltest. These re-engineered versions of state-of-the-art codes for phylogenetic inference and phylogenetic model selection are now substantially more efficient and have been designed so as to be able to scale from a laptop with merely a few cores to supercomputers with thousands of cores. This vast re-engineering effort will pay off with easier maintainability and improved software quality in the next couple of years.

Along the same lines, Pierre has completely re-engineered our evolutionary placement algorithm [Barbera, 2016]. This algorithm allows, for instance, for the evolutionary identification of bacterial community samples. The current sequential code is already 30 – 40 % faster than existing software, and Pierre is currently working on a scalable streaming pipeline parallelization approach.

Another exciting new direction of research is cancer cell evolution. Here, we do not reconstruct phylogenies of different species, but instead infer the phylogeny of the mutated genomes of different body cells in a patient. The reconstruction of the evolutionary history of cancer mutations is expected to aid medical diagnostics and predictions. Diego and Alexey have already adapted respective models in our phylogeny reconstruction tools, and we are currently testing these by means of simulated data in collaboration with David Posada at the University of Vigo.

Molecular Species Delimitation Reloaded

The field of molecular species delimitation addresses the following question by computational means: Given a set of sequences from several individuals of potentially closely related organisms, how many species are contained therein? Of course, the question of what a species actually is has been the subject of long-lasting, intense and even controversial debates among biologists. At present, there are around 30 definitions of how a species might actually be defined. As long as there is sexual reproduction, defining a species is relatively straight-forward, but when it comes to bacteria, for instance, the situations is simply a mess.

As computer scientists, we desire to mathematically capture what a species is. Previous approaches for molecular species delimitations have deployed straight-forward clustering approaches based on sequence similarity. However, such approaches do not take into account the evolutionary history of the sequences and do not use an explicit statistical model of evolution. In late 2013, we developed a model called Poisson Tree Processes that delimits species based on a statistical model on a given phylogenetic tree. The model is a classifier that lumps the branch lengths of the tree or rate of evolution that is thus defined into two classes: within- and among-species evolution.

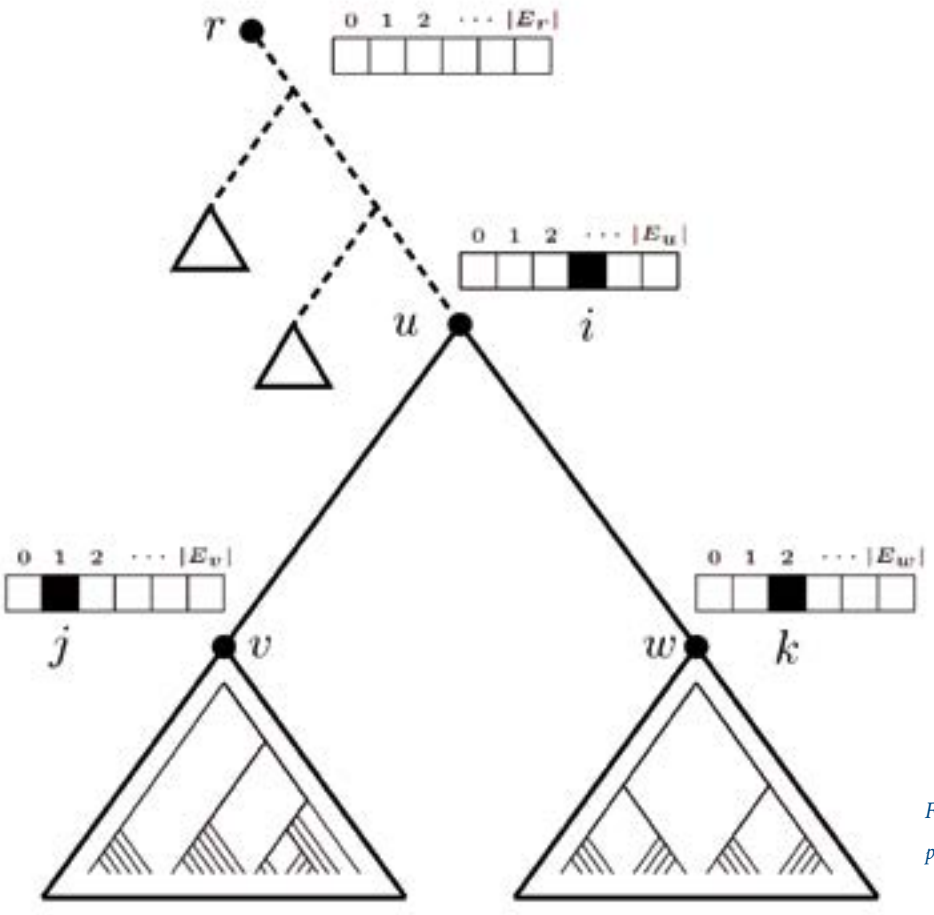


Figure 55: Schematic outline of the dynamic programming-like mPTP algorithm.

The initial PTP paper and corresponding web-service have already become a popular and widely used tool in evolutionary biology. However, the initial model and implementation were rather simplistic and slow. To this end, we developed multi-PTP, a novel, faster, more flexible, and more accurate method for molecular species delimitation that can do pretty much anything except make a cup of coffee [Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T. – Multi-rate Poisson Tree Processes for singlelocus species delimitation under Maximum Likelihood and Markov Chain Monte Carlo. bioRxiv 063875 (2016)]. Multi-PTP's key contributions are a substantially more efficient dynamic programming-like algorithm for searching for alternative delimitations (mPTP is at least five orders of magnitude faster than PTP) and an extension of the statistical model that can now encompass more than two parameters. The new model not only distinguishes between among- and within-species evolutionary processes but can also assign distinct evolutionary rates to individual species, hence the term multi-(parameter) PTP. The mPTP tool also automatically determines the most appropriate – in the statistical sense number of parameters to explain the data. A noteworthy implementation detail is that apart from the maximum likelihood version, we have also designed a Bayesian version of the tool that uses MCMC (Markov Chain Monte Carlo) sampling to explore the space of possible delimitations on trees.

We have thus designed a tool that can perform molecular species delimitation on thousands of sequences within 2 – 3 minutes on a standard laptop computer.

The Site repeats Technique

We also continued work in our established field of research, that is, improving the efficiency of phylogenetic likelihood calculations. Since the evaluation of the likelihood function on trees accounts for 90 – 95 % of overall run-time in programs for Maximum Likelihood-based and Bayesian phylogenetic inference, this is the computational kernel that needs to be optimized. We revisited an idea for accelerating likelihood calculations that Alexis had already explored in his PhD thesis back in 2002. The main idea is to detect identical DNA patterns in subtrees of the tree for which we intend to calculate the likelihood. If we have two such identical DNA patterns, we can omit calculating and storing the intermediate likelihood values for one of the two patterns. Thus, detecting such patterns can help to reduce both memory requirements and floating point operations. The main algorithmic challenge is to efficiently detect, store, and update the information about repeating DNA patterns, which we refer to as site repeats. If implemented naively, the computational cost for detecting patterns can be higher than the cost for calculating intermediate likelihood results, especially when the tree topology changes during a maximum likelihood tree search or MCMC topological sampling procedure. Hence, the key contribution of our work [Kobert K, Stamatakis A, Flouri T. – Efficient detection of repeating sites to accelerate phylogenetic likelihood calculations. Systematic Biology, https://doi.org/10.1093/sysbio/syw075 (2016)] was the design of an efficient algorithm and data structure for detecting site repeats. A prototype implementation of our method yielded up to 12-fold speedups and consumed up to 78 % less memory than our current, own implementations, which are among the fastest and most highly tuned implementations of the phylogenetic likelihood function available. This method is generic and can seamlessly be integrated into any likelihood-based phylogeny software. Looking ahead, the site repeats technique generates a novel pattern of computational imbalance for parallel likelihood calculations. Thus, designing novel data distribution algorithms is on our research agenda for 2017. Some initial experiments were conducted in a conference paper that emerged from a bachelor thesis [Scholl et al, 2016].

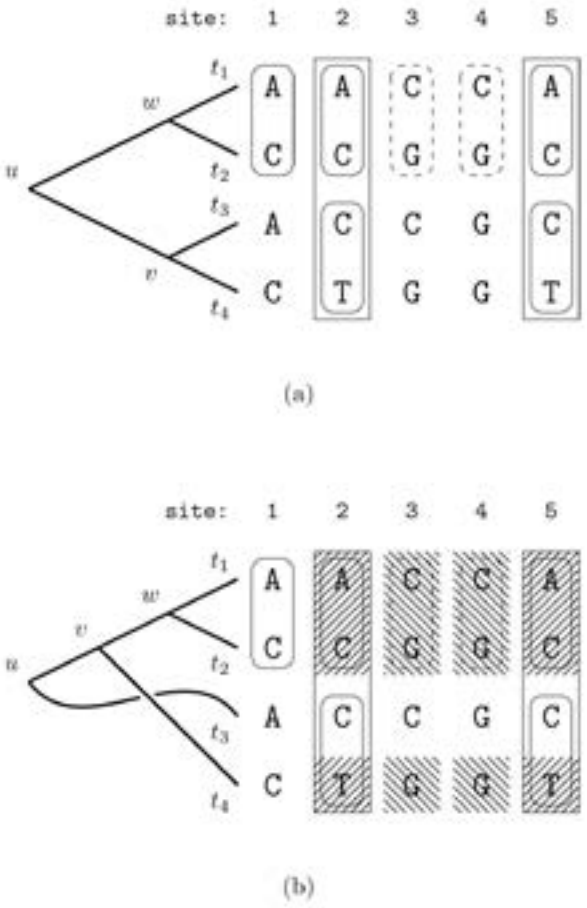


Figure 56: (a) Sites 1, 2, and 5 form repeats at node w; they share the same DNA pattern AC. Another repeating pattern is located at Sites 3 and 4 (CG) for the same node. Analogously, Sites 2 and 5 are site repeats for node v as they have the same pattern CT, and the intermediate likelihood is hence the same for these two sites. Finally, Sites 2 and 5 form repeats for node u (ACCT). (b) For this particular tree topology, node v has two sets of repeats: Sites 2 and 5 (ACT) and Sites 3 and 4 (CGG).

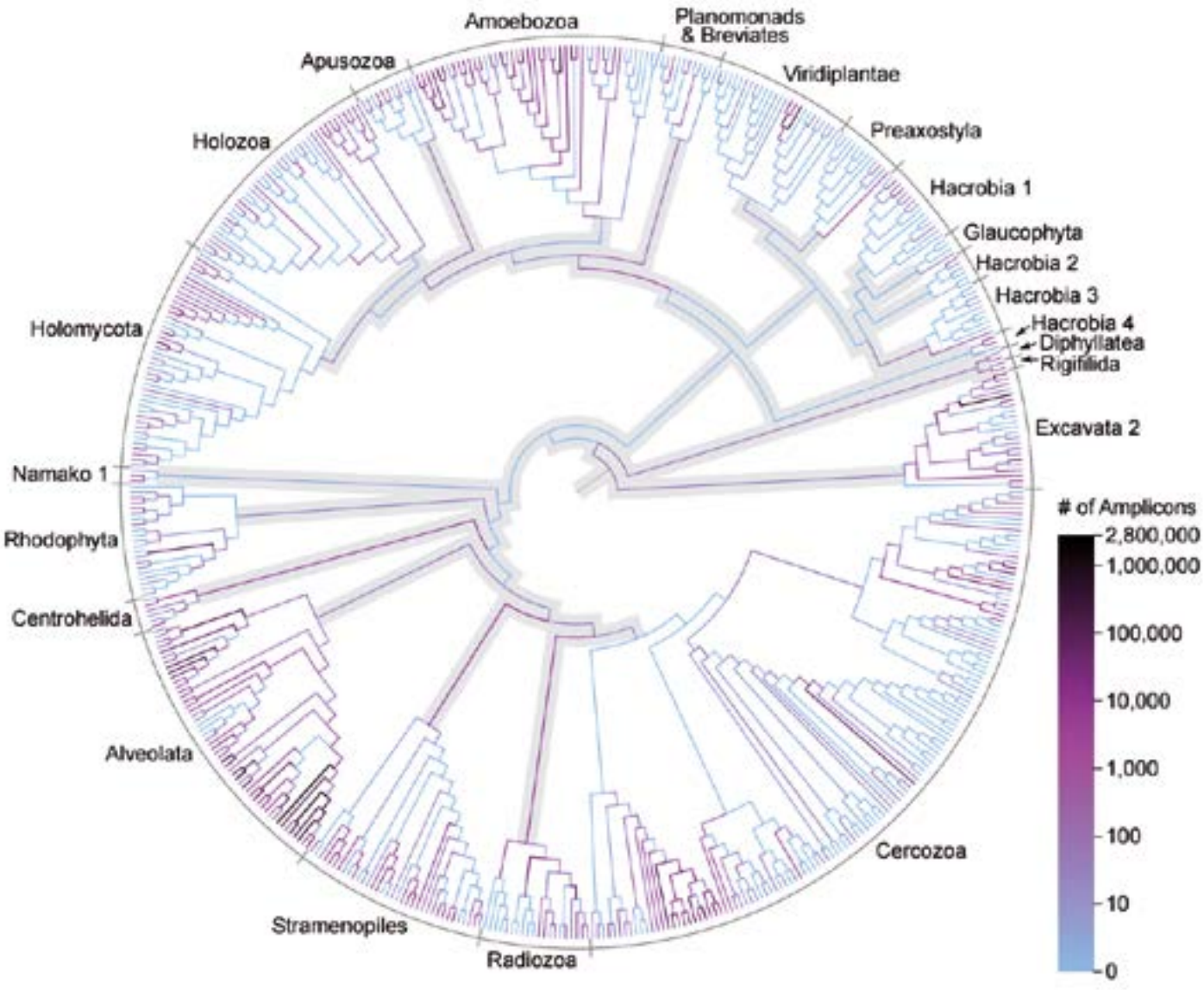


Figure 57: Phylogenetic placement of protist sequences on a reference phylogeny.

Disentangling the Diversity of Protists

In a collaborative empirical data analysis project with evolutionary biologists, we analyzed large samples of protist communities from rain forests in Costa Rica, Panama, and Ecuador by means of the aforementioned evolutionary placement algorithm [Mahe et al., 2016] (accepted for publication in Nature Ecology & Evolution in 2017). We found that protists inhabiting neotropical rainforest soils are hyperdiverse and dominated by the parasitic Apicomplexa, which infect arthropods and other animals. We suspect that these host-specific protist parasites potentially contribute to the vast animal diversity in the forests by reducing population growth in a density-dependent manner.

This study has shed light not only on protist diversity but also on future computational and methodological requirements for evolutionary placement algorithms. While the evolutionary placement runs have already been executed on the Munich supercomputing system, it turns out that datasets have become so large that the re-implementation and novel parallelization of the algorithm that Pierre is working on is urgently needed to improve throughput. Furthermore, the study has revealed challenges and future directions for the post-analysis tools for phylogenetic placements that Lucas is currently working on.

2 Research


D-3-HYDROXYACYL-ENZYME
D-3-HYDROXY-BUTYRYL-ENZYME
trans-2,3-DEHYDRO-ACYL-ENZYME
CROTONYL-ENZYME
FATTY A
(PALMITOYL-CoA ENZYME PALMI-TOYL TRANS-FERASE)
for chain elongation (mitochondrial)

2.12 Scientific Databases and Visualization (SDBV)

2.12
Scientific Databases and Visualization (SDBV)

The work at SDBV revolves around collecting data and preparing it to share. In our longest-running project, SABIO-RK, we enter data on a daily basis and curate it to conform to the standards and be well-identified and well-interlinked with related data sources. SABIO-RK is known and respected in the community and was celebrated in a workshop organized at HITS in 2016. Curating data for SABIO-RK is both costly and complicated. Having central facilities like SABIO-RK curate data is not scalable, which presents a challenge that modern data management needs to solve.

The data management projects we are part of aim at enabling users to self-curate using the SEEK (developed in collaboration with Manchester) system. Associated tools, such as RightField (developed in Manchester) and Excemplify (HITS), are designed to help users make data FAIR (findable, accessible, interoperable, and reusable) by themselves.

We employ a variety of methods to help make users self-sufficient, including actually visiting scientists and talking with them (training and consulting); more modern and remote versions of keeping contact and exchanging views, methods, and information; and participating in standardization efforts for software engineering methodology, semantic web, information retrieval, and user experience.

A key property of SDBV is our mix of scientists with a biology / biochemistry / chemistry background (Anjani Drangadhariya, Martin Golebiewski, Renate Kania, Olga Krebs, Hadas Leonov, Maja Rey, Ulrike Wittig, Andreas Weidemann) and those who focus on computer science (Steffen Brinkmann, Anjani Drangadhariya, Hadas Leonov, Quyen Nguyen, Marcel Petrov, Andreas Weidemann, Jill Zander). Projects work best when biologists with an understanding of computer science and developers with an interest in traditional science communicate closely to solve users' problems. For us, this is both a challenging and very rewarding task.

Im Zentrum der Arbeiten der SDBV steht das Sammeln und Teilen von Daten. In unserem ältesten laufenden Projekt arbeiten wir jeden Tag daran, Daten einzugeben, diese zu kuratieren, damit sie Standards entsprechen, eindeutig identifiziert sind und gut mit mit anderen Daaten vernetzt. SABIO-RK ist in der Community bekannt, respektiert und wurde in diesem Jahr mit einem Workshop zum zehnjährigen Jubiläum am HITS gefeiert. Die Kuratierung von SABIO-RK-Daten ist kompliziert und kostenträchtig. Wenn man sich die Situation genau ansieht, skaliert die Kuratierung von Daten in zentralen Eindrichtungen nicht. Dies ist eine Herausforderung, die vom modernen Datenmanagement adressiert werden muss.

Die Datenmanagement-Projekte, deren Teil wir sind, versuchen Nutzern zu ermöglichen, ihre eigenen Daten mithilfe der Werkzeuge SEEK (Ko-Entwicklung mit Uni Manchester), Rightfield (aus Manchester), Excemplify (HITS) zu kuratieren, und so die Daten FAIR zu machen, d.h. suchbar, zugreifbar, interoperabel und wiederverwendbar zu machen.

Zu diesem Zweck nutzen wir viele Methoden, von klassischen Besuchen für Training und Consulting, elektronischer Kommunikation zum Austausch von Ansichten, Methoden und Daten, über Teilnahme an Standardisierungsinitiativen bis zu Software-Engineering, Semantischem Web, Information Retrieval und Mensch-Maschine-Interaktion.

Eine wesentliche Eigenschaft unserer Gruppe ist die Mischung von Wissenschaftlern mit einem biologischen/ biochemischen/ chemischen Hintergrund (Anjani Drangadhariya, Martin Golebiewski, Renate Kania, Olga Krebs, Hadas Leonov, Maja Rey, Ulrike Wittig, Andreas Weidemann), und Wissenschaftlern mit einem Hintergrund aus der Informatik (Steffen Brinkmann, Anjani Drangadhariya, Hadas Leonov, Quyen Nguyen, Marcel Petrov, Andreas Weidemann, Jill Zander). Projekte funktionieren am besten, wenn Biologen mit Verständnis für Belange der Informatik und Informatiker mit naturwissenschaftlichem Interesse eng miteinander zusammenarbeiten und im Sinne der Nutzer an Problemlösungen arbeiten. Für uns ist dies eine schwierige, aber sehr lohnende Aufgabe.

Group leader

Priv.-Doz. Dr. Wolfgang Müller

Staff members

Dr. Steffen Brinkmann *(since July 2016)*
Anjani Dhrangadhariya *(since September 2016)*
Martin Golebiewski
Ron Henkel
Renate Kania
Dr. Olga Krebs
Dr. Hadas Leonov *(since February 2016)*
Quyen Nguyen
Dr. Maja Rey
Dr. Andreas Weidemann
Dr. Ulrike Wittig

DHBW Students

Marcel Petrov *(since October 2016)*
Jill Zander *(until October 2016)*

Scientific Databases and Visualization (SDBV)

2016 bore witness to a variety of interesting developments at SDBV. We started LiSyM, the new, exciting, follow up of the Virtual Liver Network, and at the same time the project SBEpo came to ist end. However, let us first describe the project structure of SDBV.

Three of our projects are dedicated to providing data-management services to large collaborative consortia: LiSyM, NMTrypI, and SBEpo. Then, there are two more general infrastructure projects: FAIRDOM is a project that serves the European ERASysAPP consortium and works to reach the wider European community of systems biology and related fields. De.NBI, the German Network for Bioinformatics Infrastructure consortium, is the German service provider for bioinformatics services and teaching. With our Curated Reaction Kinetics Database SABIO-RK and our data management services and courses, we represent an integral part of the network. One essential ingredient of data management systems are standards that prescribe how to store data, models, and SOPs (standard operating protocols). The NormSys consortium on modelling standards has a very broad visibility in the field and facilitates the discussion of an ISO standard. Finally, in 2016, we worked on the OperationsExplorer, our software designed to help journalists discover stories in data about medical procedures in Germany. This tool is in active use, and we are looking forward to reading stories in which OperationsExplorer has proven useful.

While we have tried to provide a complete overview over our yearly work in the past, in this report, we focus on 3 projects. We decided to feature Excemplify (a tool for collecting experimental data in a standardized manner), the NMTrypI (New Medicines for Trypanosoma Infections) project's data management, and finally, NormSys.

Acknowledgements: Some of the development of both NMTrypI and NormSys was performed by the IT Services of HITS. Our collaboration with them proved both enjoyable and fruitful. In NMTrypI, we closely collaborated with the Wade group,

Figure 58: Interactive graphical representation of spreadsheet data as line charts in Excemplify.

consisting of Rebecca Wade, Ina Pöhner, and Joanna Panecka. We are grateful for this very enjoyable and fruitful collaboration, including the discussions we had and the journeys we shared on the way to EU project meetings.

Excemplify

Excemplify is a web-based application that was developed to collect and store experimental data, particularly spreadsheets and immunoblot-related data. The guiding principle behind Excemplify is that immunoblot experiments are accompanied by a series of spreadsheets, each dependent on the sheet of the previous experimental stage. Moreover, if tools do not provide guidance and help, it is very difficult to maintain self-standardization, i.e., it is difficult for a single person to maintain standardization in experiment documentation. With Excemplify, however, self-standardization becomes very simple.

Excemplify is able to parse initial experimental setup descriptions and generate the subsequent spreadsheet stages for the experimental workflow.

In addition to Excemplify's data-storage capabilities, experimentalists are freed from time-consuming data-handling procedures and error-prone manual data input. Additionally, after uploading data to one or several SEEK instances, the data can be shared within larger projects and consortia. Excemplify was developed in close collaboration with the department for Systems Biology of Signal Transduction at the German Cancer Research Institute (DKFZ).

In 2016, we developed an extended version of Excemplify with enhanced visualization capabilities that aid in deciding what data should be shared using a data-sharing tool like SEEK (http://www.seek4science.org). The graphical display of spreadsheet data in an interactive manner (*Figure 58* and *59*) is one such tool. A public demo version is available at http://sabiork.h-its.org/excemplify/.

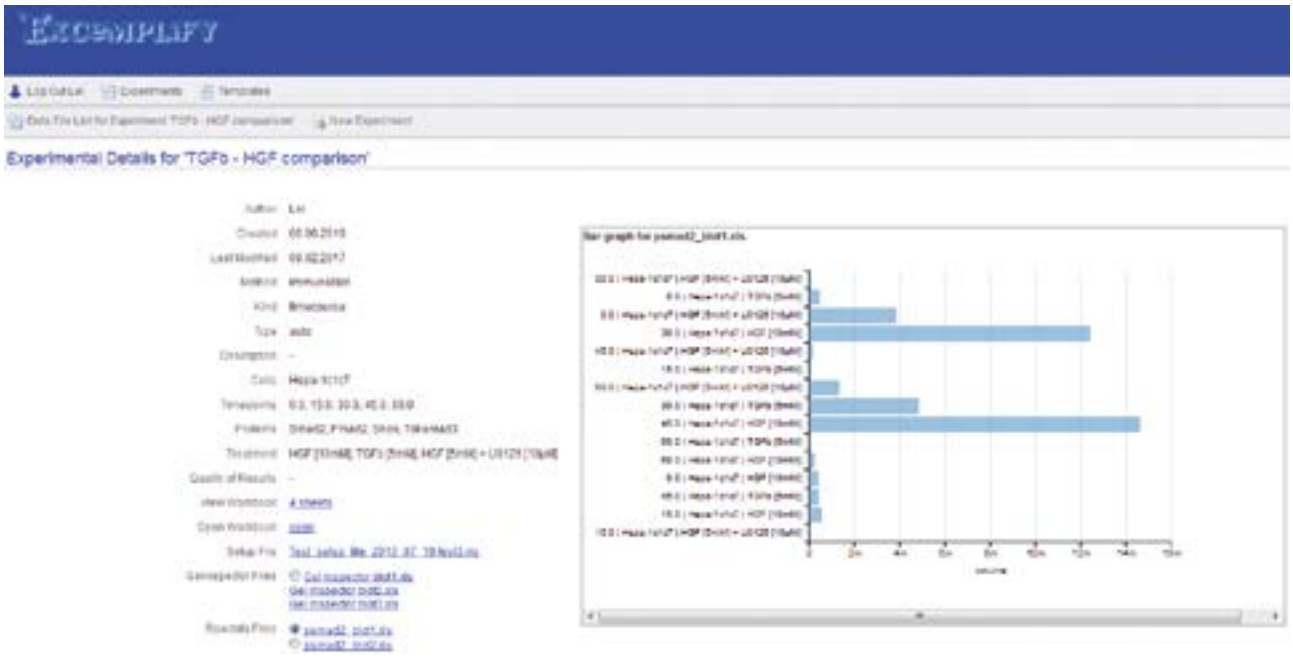


Figure 59: Display of spreadsheet data as bar charts in Excemplify.

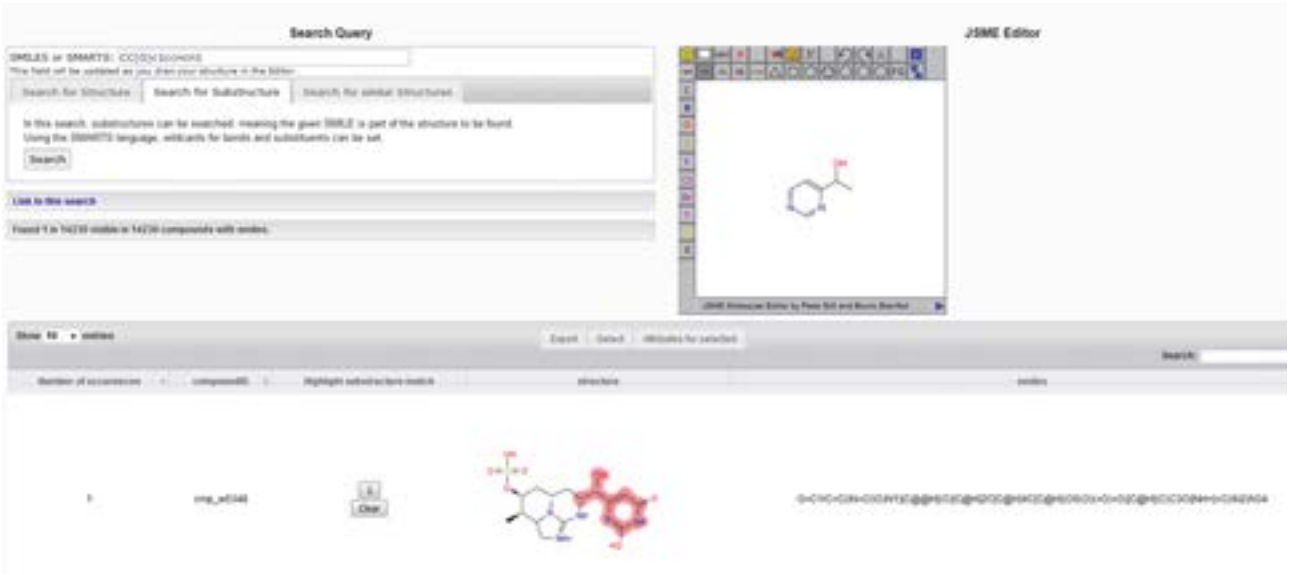


Figure 60: Compound substructure search interface in NMTrypI SEEK. Matched compound substructure is highlighted in red.

Data Management for NMTrypI

NMTrypI (New Medicines for Trypanosomatidic Infections) is a project funding 14 EU partners running up till January 2017. The goal of the project was to search for new drugs for Trypanosomatidic infections (i.e., sleeping sickness, leishmaniasis, and Chagas disease). Some of these diseases are lethal, all are painful, and all of them occur to a large extent in poverty-stricken areas of the world, making cheap medication a precondition for successful therapy.

Our data management platform SEEK was used for central data management within the project, and additional features were developed and implemented to meet user requirements for the exploration and visualization of the research data and to handle information specific to chemical compounds.

One of the main requirements of NMTrypI project partners was to be able to search through all compound data in SEEK by based on chemical structure. Using the available SMILES (Simplified Molecular-Input Line-Entry System) strings within the Excel tables in SEEK, we implemented a graphical representation for drawing 2D molecules as input for substructure (see *Figure 60*) and similarity search tools in SEEK. The results allow the user to view all the files in SEEK that contain the resulting compounds and provide an exploration of the files and data relating to those particular compounds.

With an emphasis on data visualization, we expanded the system to allow the user to select and visualize data from different sheets of the same file with an Excel-like selection, i.e., sheet1:sheet3!A1:F200 would select cells A1 to F200 from sheets 1–3. Past and present visualization functionalities were expanded to include data from different sources, which served as a middle step in selecting and plotting data from different Excel files.

Interactive exploration of the data can be achieved through the newly introduced Parallel Coordinates plot and the improved Heat Map function. A parallel coordinate plot (shown in *Figure 61*) allows the user to visualize and analyze multivariate data to draw insights from easily observed 2D patterns. The user could choose to filter the data within the plot and rearrange it in different ways. As often occurs in experiments, not all parameters were al-

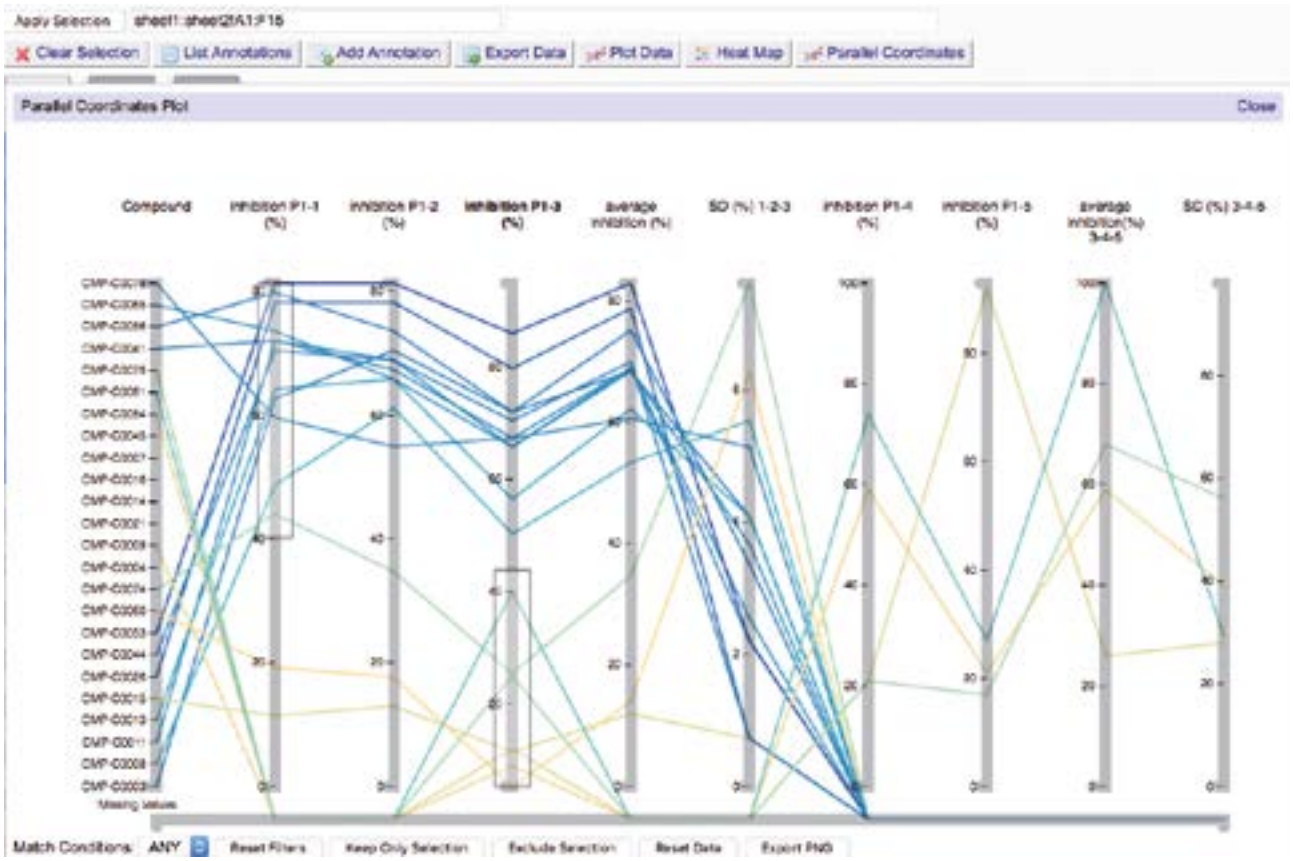


Figure 61: Visualization of table data using parallel coordinates in NMTrypI SEEK. The lines are colored by the 4th axis (column) and filtered by values from both the 2nd and 4th parallel axes. The missing value axis is at the bottom.

ways measurable. Our plot allowed us to account for these cases and learn of their distribution in the data through the horizontal "missing values axis". Both plotting functions allowed for immediate data extraction by hovering on the data, interactive filtering, and coloring, as well as exporting the plot as an image.

The NMTrypI project consortium used two different SEEK instances for its research-data management: NMTrypI SEEK (https://nmtrypi.h-its.org/) was password-protected and for project-internal use only. Synergy SEEK (https://fp7h-synergy.h-its.org/) was used for the publication and dissemination of research data for sharing the NMTrypI project results with other drug-discovery projects and the scientific community. After ending the project, HITS offered to maintain both SEEK instances for at least the next ten years (until January 31st, 2027).

Modeling standards in systems biology (NormSys)

Norms and requirements for objects, methods, processes, and practices shape our everyday lives. They are often represented in formal documents, each defining a standard for a specific field in a particular domain. In systems biology and other fields of the life sciences, the consistent structuring and description of produced data and computer models – with their underlying experimental or modeling processes – is crucial and only possible by applying dedicated standards for formatting and describing data and corresponding metadata (data that describe the data and its context), as well as workflows and models. Such consistent documentation based on standards is mandatory for reproducing, re-using, und integrating data and models, which are major tasks in the life sciences, particularly in highly interdisciplinary fields such as systems biology, systems medicine, and synthetic biology. The standards in these fields are not static but develop and evolve with the progress of science and technology. A current major challenge is to harmonize standardization efforts in

the different fields that refer to different approaches and technologies and to make the corresponding standards interoperable.

The NormSys consortium (http://www.normsys.de), which was coordinated by Martin Golebiewski (SDBV), aimed at bringing different stakeholders together to further harmonize and unify standardization in systems biology. These stakeholders included researchers from academia and the industrial world (with their grass-roots standardization initiatives, scientific journals, and research-funding agencies) as well as national and international standardization institutions, such as the DIN (German Institute for Standardization) and the ISO (International Organization for Standardization). The German Federal Ministry for Economic Affairs and Energy (BMWi) funded the NormSys project, which began in 2014 and ran until the end of 2016. To factor in requirements from different stakeholders, we collaborated with partners from academia (University of Potsdam) and a private enterprise (LifeGlimmer GmbH, Berlin), who were our project partners. The major aim was to enhance and promote the formal standardization of existing community standards for modeling by bridging existing gaps between stakeholder groups.

To achieve this goal, both during and beyond the runtime of NormSys, we actively contributed to scientific standardization initiatives, such as the Computational Modeling in Biology Network (COMBINE) [Schreiber et al., 2016], for which we are part of the board of coordinators. COMBINE is a "community of communities" initiative to coordinate the development of various community standards and formats for computational models in biology and comprises widespread modeling standards:

- CellML
- BioPAX (BIOlogical Pathway Exchange)
- NeuroML for neuroscience models
- SBGN (Systems Biology Graphical Notation)
- SBML (Systems Biology Markup Language)
- SBOL (Synthetic Biology Open Language)
- SED-ML (Simulation Experiment Description Markup Language)

As part of these activities, we organized the 1-day COMBINE & de.NBI Tutorial "Modelling and Simulation Tools in Systems Biology" as a satellite to the International Conference on Systems Biology (ICSB) in Barcelona, Spain, on September 16th, 2016 (see Events), and we were co-organizers of the COMBINE meeting series HARMONY and COMBINE forum. Beyond the COMBINE community, we actively sought to bring together stakeholders for standardization, as in the NormSys round table talk about standardization in the life sciences on April 6th, 2016, which was a part of the Conference on Systems Biology of Mammalian Cells (SBMC 2016) in Munich, as well as at the NormSys Conference "Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences," which we organized together with our project partners in Potsdam (December 5th – 6th, 2016).

In addition to this involvement in scientific grass-roots standardization initiatives, we also participated in the committee work of standardization bodies, such as the Technical Committee for Biotechnology Standards at ISO (ISO / TC 276) and its national German mirror committee at the DIN. As an important step for reaching the project goals of NormSys, HITSter Martin Golebiewski was appointed to head the international work group for "data processing and integration" (WG5) at the ISO / TC 276 (which comprised more than 60 experts from 15 countries) and chaired several of the working group's meetings in 2016 in France, Ireland, and the US. In this context, we are currently working at the ISO to create an international framework standard (ISO / PWI 20691: "Downstream data processing and integration workflows") that refers to existing standards for data processing and modeling in the life sciences with the specific goal of establishing an ISO standard as a hub for such community standards and providing a guideline for these standards' application.

A prerequisite for this ISO standard is a comprehensive inventory of community standards for modeling in systems biology and related fields [Schreiber, Bader, Golebiewski et al., 2015]. To survey these standards, we developed the NormSys registry for modeling standards (http://normsys.h-its.org), which was first released in October 2015 (*Figure 62*). This registry provides a single access point for consistent information about model-exchange formats, such as the COMBINE standards SBML,

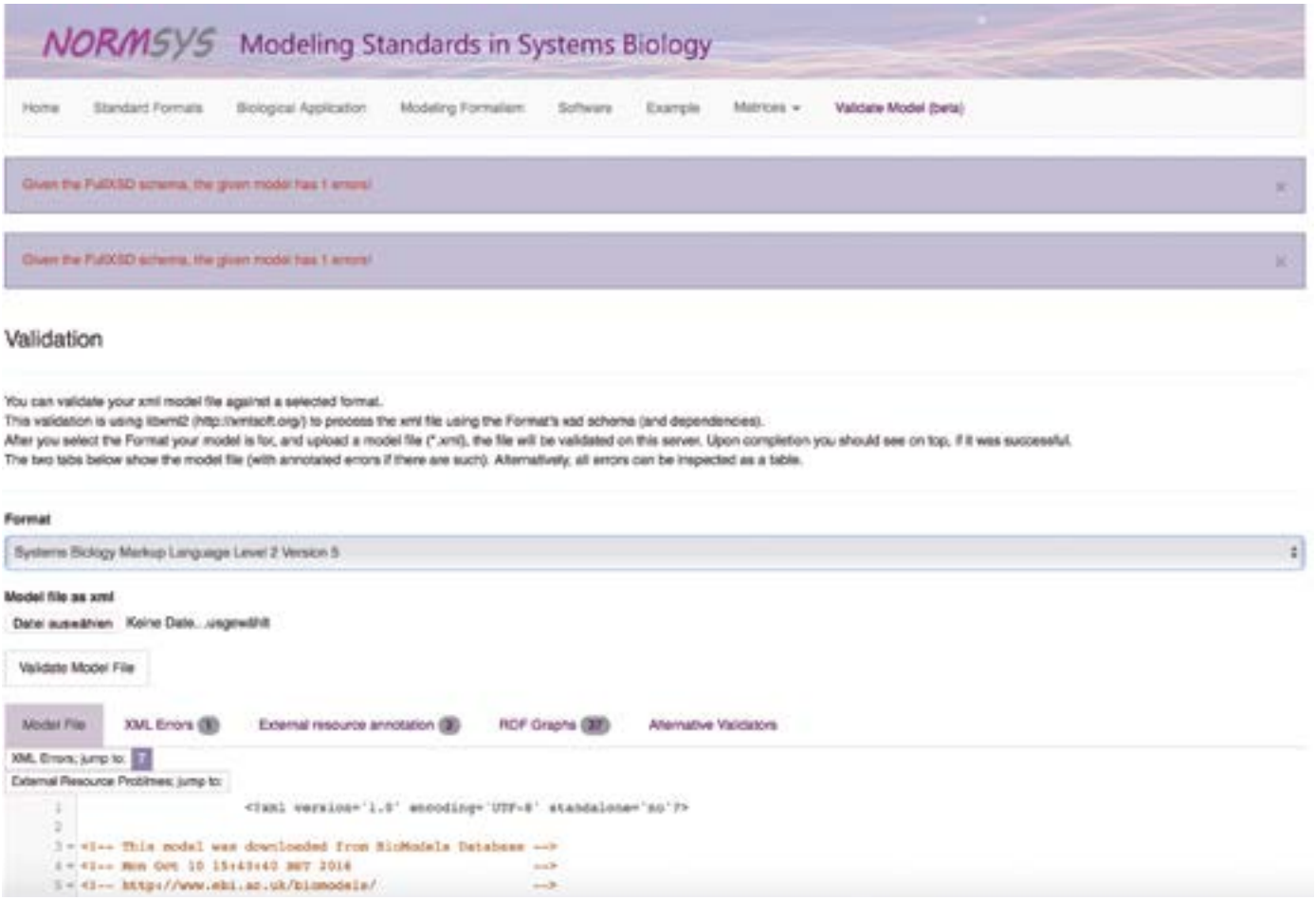
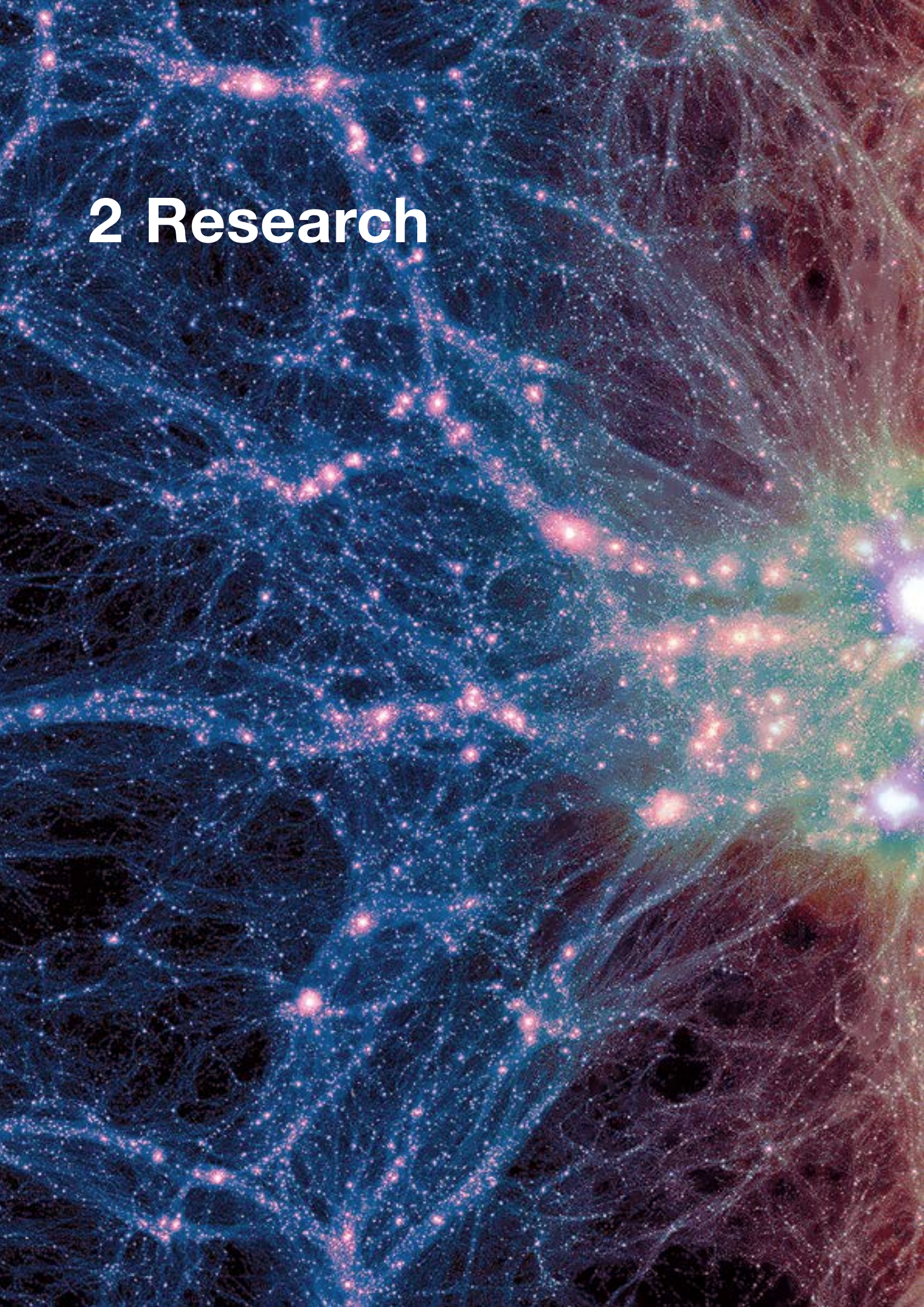


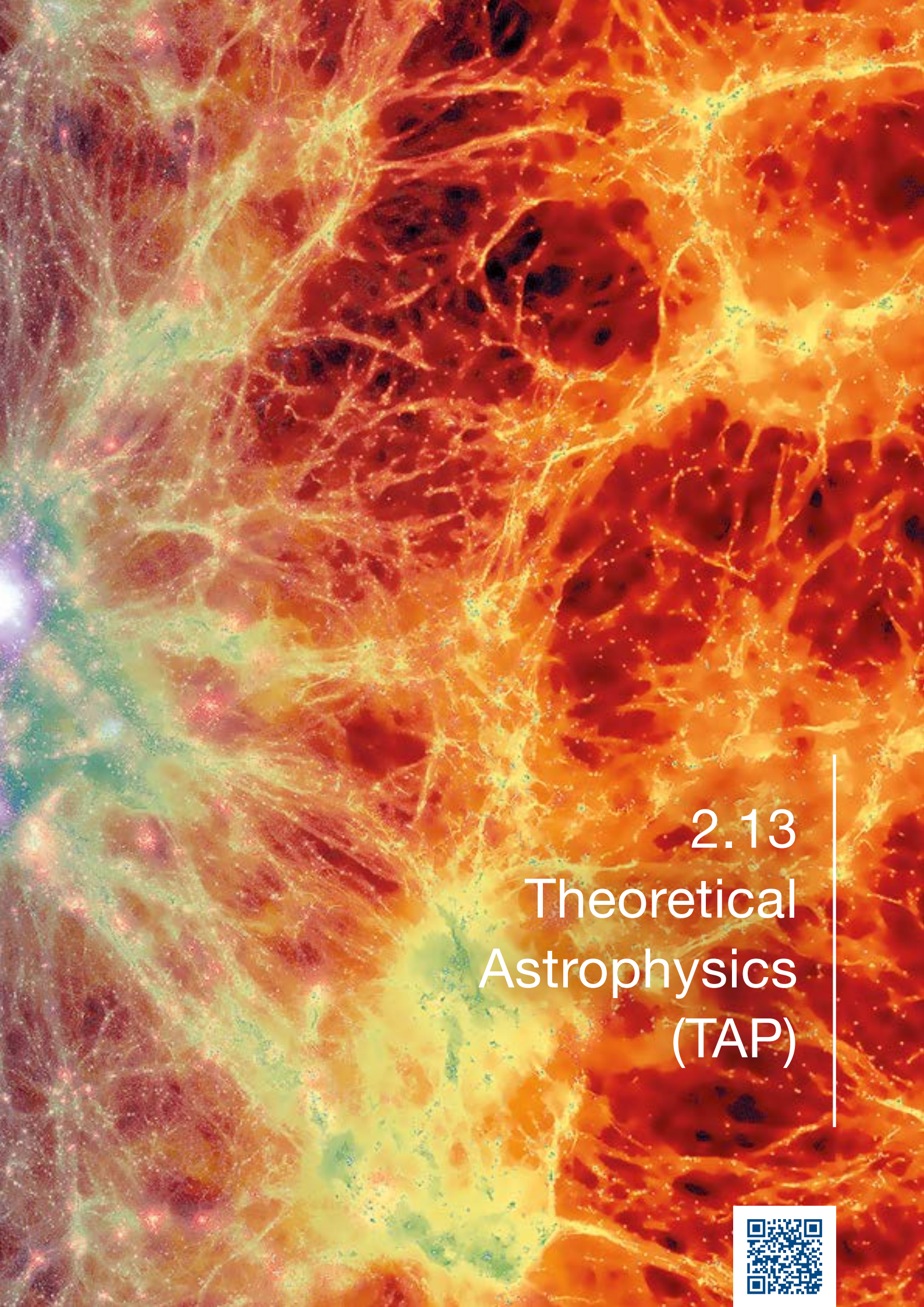
Figure 62: The NormSys model validator for syntax and semantic checks of biological models, implemented in the NormSys registry for modeling standards (http://normsys.h-its.org).

CellML, SBGN, SED-ML, NeuroML, SBOL, and the Pharmacometrics Markup Language (PharmML), among others. The publicly available platform not only lists the standards but also compares their major features, their potential fields of biological application, and potential use cases (including model examples), as well as their similarities, relationships, commonalities, and differences. This NormSys registry is a convenient information source, especially for experimentalists, modelers, and software developers planning to apply the standard formats for their respective purposes. The registry provides these individuals with detailed information as well as links to webpages, specifications, and web services associated with the formats. In 2016, we implemented website analytics based on Piwik to track website usage (statistics on website clicks, single users, geographic locations, etc.).

Based on the NormSys registry, we developed and implemented an extension of the existing system via a validation and certification platform that is accessible online for computer models of biological systems that are described in the standard formats. This NormSys model validator, which was released at the end of 2016, includes both a formal syntax check (XML-based) and a semantic check of the models and the integrity and consistency of their content, their entities, and their annotations. The implementation was performed in close collaboration with Nils Wötzel from the IT Services group (ITS) of HITS. For syntax validation, a generic framework was implemented that makes use of the XML schema information on the corresponding standards stored in the NormSys registry. The consistency and validity of model entities and their corresponding annotations are checked for semantic validation. This semantic annotation test of an uploaded model is performed by a compliance check according to the MIRIAM guidelines, which provide a checklist catalogue for model annotations in biology (Minimal Information Required In the Annotation of Models: http://co.mbine.org/standards/Miriam).



2 Research



2.13 Theoretical Astrophysics (TAP)

2.13
Theoretical Astrophysics (TAP)

The Theoretical Astrophysics group at HITS seeks to understand the physics of cosmic structure formation over the last 13.5 billion years – from briefly after the Big Bang until today. We are especially interested in how galaxies form and ultimately produce magnificent systems like our own Galaxy, a busy metropolis of more than one hundred billion stars. We also aim to constrain the properties of dark matter and dark energy, the two enigmatic matter and energy components that dominate today's universe and contribute to some of the most fundamental problems in modern physics.

Numerical simulations on a variety of scales, including both the collisionless and hydrodynamical type, play a prominent role in our work. To this end, we develop novel numerical schemes that can be used efficiently on very large supercomputers with the goal of exploiting them at their full capacity in order to link the initial conditions of the universe with its complex and evolved current state. The simulation models are indispensable for the interpretation of observational data and their comparison with theoretical models.

Using simulations, we are able to study how diverse physical processes relevant in structure formation interact in a complex and highly non-linear fashion. A current priority in our group is the incorporation of improved physical models for supermassive black hole formation, cosmic rays, and radiative transfer. This report highlights a few results from our work over the past year.

Die Theoretische Astrophysik Gruppe am HITS versucht die Physik der kosmischen Strukturentstehung während der letzten 13,5 Milliarden Jahre, vom Urknall bis heute, zu verstehen. Unser besonderes Interesse gilt der Entstehung von Galaxien, welche schließlich zur Bildung von großartigen Systemen wie unserer Milchstraße führt, einer geschäftigen Metropole mit mehr als einhundert Milliarden Sternen. Wir arbeiten auch an einer Bestimmung der Eigenschaften der Dunklen Materie und der Dunklen Energie, jenen rätselhaften Komponenten, die den heutigen Kosmos dominieren und die zu den fundamentalsten Problemen der modernen Physik gehören.

Eine besonders wichtige Rolle in unserer Arbeit spielen numerische Simulationen auf verschiedenen Skalen. Zu diesem Zweck entwickeln wir neue numerische Verfahren, die effizient auf sehr großen Supercomputern eingesetzt werden können, mit dem Ziel, deren volle Kapazität für eine Verknüpfung der Anfangsbedingungen des Universums mit seinem heutigen komplexen Zustand auszunutzen. Die Simulationen sind für die Interpretation von Beobachtungen und deren Vergleich mit theoretischen Modellen unverzichtbar.

Mit der Hilfe von Simulationen sind wir insbesondere in der Lage, das komplexe und nichtlineare Zusammenspiel verschiedener physikalischer Prozesse zu studieren. Eine aktuelle Priorität in unsere Gruppe besteht darin, Physik in unsere Modelle einzubauen, die zwar als wichtig erachtet wird, die aber bisher vernachlässigt wurde, etwa superschwere Schwarze Löcher, kosmische Strahlen oder Strahlungstransport. In diesem Bericht stellen wir beispielhaft einige Ergebnisse unserer Arbeit im vergangenen Jahr vor.

Group Leader

Prof. Dr. Volker Springel

Postdocs

Dr. Robert Grand
Dr. Thomas Guillet *(since September 2016)*
Dr. Rüdiger Pakmor
PD Dr. Christoph Pfrommer *(until April 2016)*
Dr. Christine Simpson
Dr. Martin Sparre
Dr. Freeke van de Voort *(since October 2016)*
Dr. Dandan Xu

Students

Christian Arnold
Andreas Bauer *(until March 2016)*
Sebastian Bustamante
Conrad Chan *(since July 2016)*
Felipe Goicovic *(since August 2016)*
Svenja Jacob
Lee Rosenthal *(until June 2016)*
Kevin Schaal *(until July 2016)*
Rainer Weinberger
Jolanta Zjupa

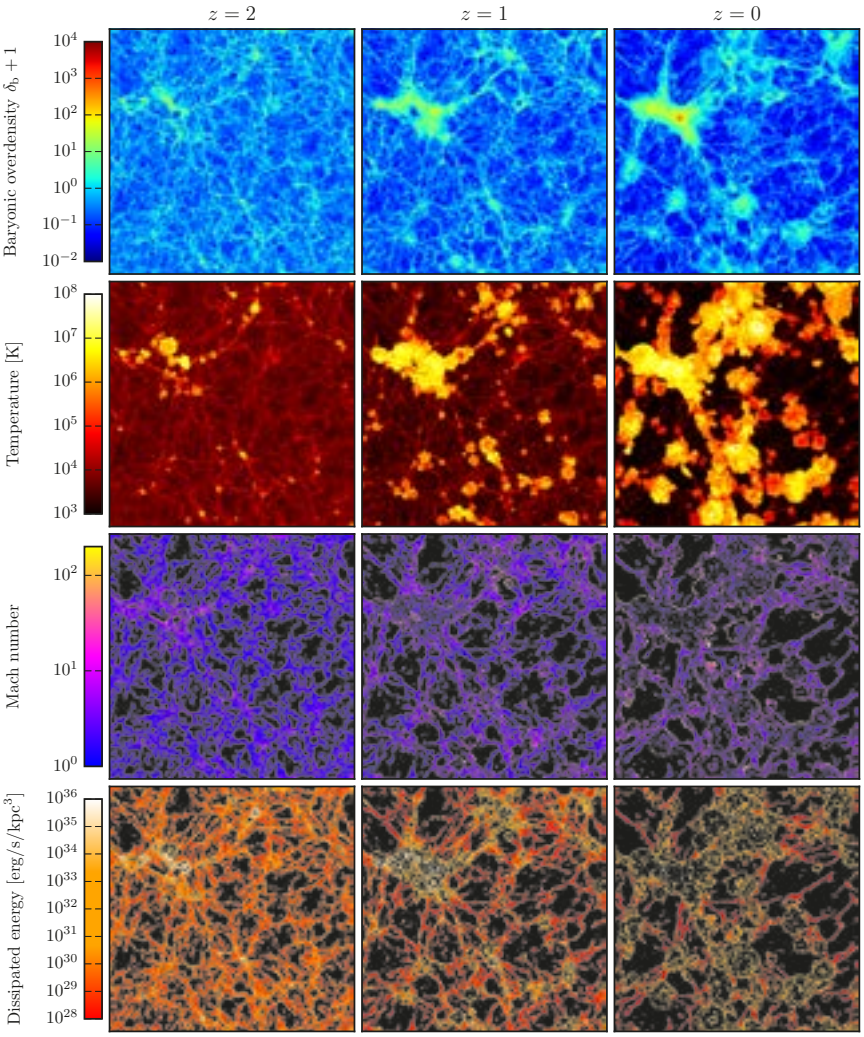


Figure 63: Projections of the baryonic overdensity, the temperature, the Mach number field, and the energy dissipation rate density for the full Illustris simulation at different redshifts. On large scales, accretion shocks on to the cosmic web dominate the early universe, whereas at late times, the radio mode feedback of BHs is omnipresent and creates hot bubbles that shock heat the ambient medium. A movie of the shock dynamics in a sub-box of Illustris is accessible online at https://youtu.be/1CXCRv3i3uw

Shocks in the Illustris Universe

Hydrodynamic shock waves play an important role in the evolution of the baryonic component of our Universe. During the process of hierarchical structure formation, shocks dissipate and thermalize kinetic energy in supersonic gas flows, thereby heating the cosmic web and the structures therein. Gas inside dark matter haloes is compressed and virialized by means of shocks before radiative cooling in the densest regions allows it to collapse further, eventually leading to the formation of stars and accretion onto black holes (BHs). Supersonic gas motions are also created by feedback processes, such as stellar winds, supernovae (SNe) explosions, and relativistic jets from active galactic nuclei (AGN). The conversion of the released mechanical energy to thermal energy in the ambient gas is again mediated by thermalization processes that ultimately involve hydrodynamic shocks.

Unlike in classical hydrodynamic shocks in which the kinetic energy of a fluid is thermalized via particle collisions, shocks on cosmological and galactic scales are collisionless. Interestingly, collisionless shocks are also sites of diffusive shock acceleration, which create relativistic cosmic rays. In this process, ions inside the converging flow can cross the shock multiple times by scattering off the magnetic field fluctuations in the pre- and post-shock regions, thereby gaining more and more energy with every crossing. The net result is a power-law momentum spectrum of energetic particles, which are called cosmic rays.

Directly observing cosmological shocks is generally very difficult. While the central high density regions of galaxy clusters can be observed in X-rays, the gas temperature in these environments is high, as well. This allows only for low Mach number shocks, which are difficult to detect since the density contrast between the shocked and unshocked gas is small. Still, a number of bow shocks in merging clusters have been discovered in XMM-Newton and Chandra observations, the most famous being the Mach number $M \approx 3$ shock in the bullet cluster. Moreover, pressure jumps associated with

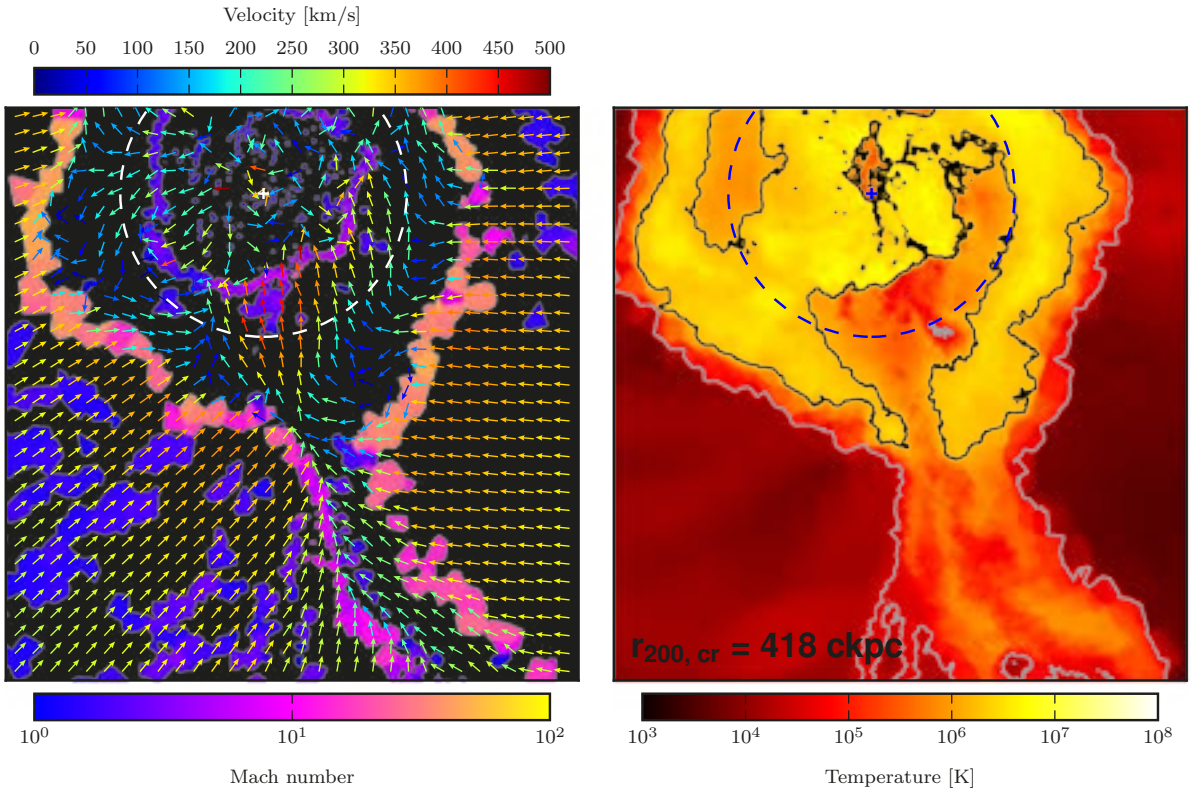


Figure 64: Zoomed in image of a halo with a double accretion shock. The left-hand panel shows a thin projection of the Mach number field as well as the direction and magnitude of the peculiar gas velocity field. The mass-weighted mean temperature is projected on the right-hand side panel. A vertical filament penetrates the outer accretion shock and channels cold gas into the interior. Since the cold phase mixes with gas that is shock heated at the outer accretion shock, a roughly spherical inner accretion shock is formed.

shocks have been found with the Planck space telescope at the outskirts of the Coma Cluster. Fortunately, the synchrotron emission of (re-)accelerated electrons can reveal the locations of cosmological shocks more easily, thus providing another channel for important observational constraints on shocks.

In a recent paper [Schaal et al., 2016], we analyzed shock waves in our most advanced cosmological simulation of galaxy formation, the Illustris simulation. Our analysis utilized a new shock-finding algorithm that we developed and that operates directly on the Voronoi mesh of the AREPO simulation code. This allowed us to be the first to investigate shocks in a state-of-the-art cosmological simulation that includes gas and dark matter in addition to stars, BHs, and their associated feedback processes. These feedback sources are expected to create powerful additional shocks, making it particularly interesting to compare the resulting shock statistics with those obtained for the non-radiative runs.

Figure 63 provides an overview of the global shock patterns we identified in the Illustris simulation in the form of the mean baryonic overdensity, the mass-weighted temperature, the dissipated energy weighted Mach number, and the mean dissipated energy density at different redshifts. The cosmic web and the halos at the intersections of filaments are clearly visible in the baryons. Both are surrounded by accretion shocks; however, at low redshift, additional strong shocks emerge from the largest halos. These are powerful feedback shocks driven by AGN activity in the centers of massive halos.

The Illustris simulation has enough dynamic range to also allow a detailed study of the gas flows within halos, as illustrated in *Figure 64*, in which we depict a zoomed-in image of the 4th-largest halo. The halo is surrounded by a strong outer accretion shock with Mach numbers around $M = 40 - 50$. Such shock strengths are expected for relatively cool accretion flows from voids and weak filaments. Moreover, the gas inside the highlighted filament is dense and fast enough to penetrate this outer shock and thermalize further inside, where the halo pressure is higher. The accretion flow from the filament is accelerated towards the potential minimum and reaches a velocity of around 500 km/s relative to the halo center. Interestingly, the fact that the filament crosses the outer shock in the vertical direction gives rise to the formation of an inner accretion shock, something that has been able to be demonstrated for the first time in our study.

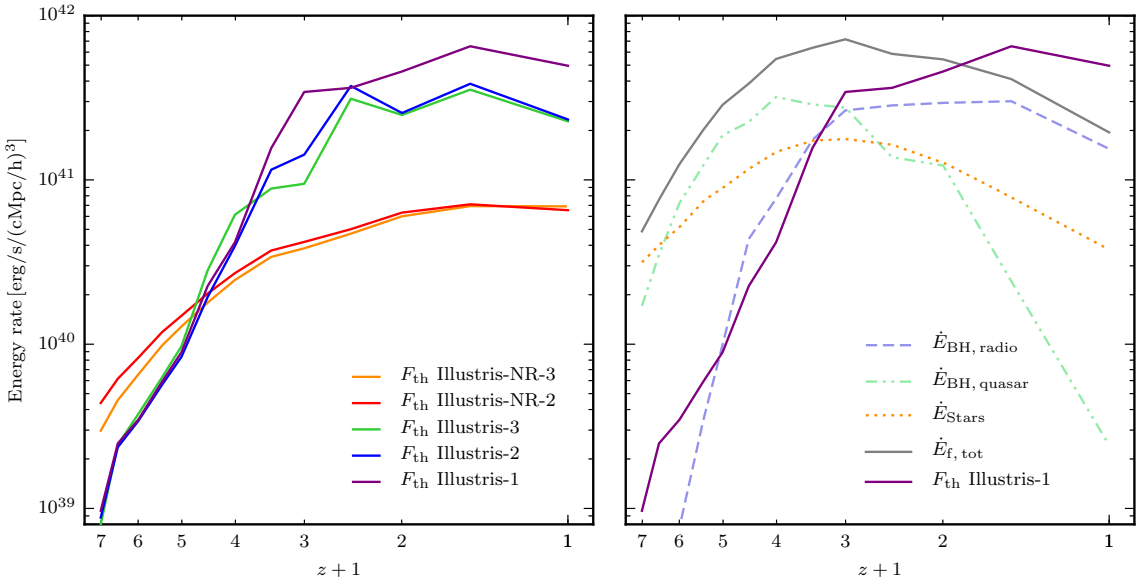


Figure 65: Total energy dissipation (left panel) at shocks as a function of redshift for different Illustris runs compared with the energy output of different feedback channels (right panel) in the Illustris-1 simulation. For redshifts $z > 1$, the dissipation stays well below the total energy dumped into feedback processes by stars and BHs, indicating that not all feedback channels produce energetic shocks. On the other hand, there is a strong correlation at redshifts $z < 4$ between the energy released by BHs in the radio-mode and the energy dissipated at shocks.

By measuring the energy dissipation rate at shocks, the impact on the thermal history of the gas can be inferred. In non-radiative runs, the overall energy dissipation grows with cosmic time, and a bimodality in the Mach number dependence of the energy dissipation statistics can be seen; internal low Mach number shocks dissipate most of the energy, whereas external shocks have high Mach numbers but are much less energetic. Remarkably, for Illustris, we found that shocks with Mach numbers smaller and larger than $M \approx 10$ contribute roughly equally to the overall dissipation, and no bimodality can be observed. We attribute the large population of high Mach number shocks to feedback processes present in this full physics simulation. In fact, at low redshift, we can directly account for the energy dissipation in strong shocks by the energy released through black hole feedback (see *Figure 65*).

Our results underline the importance of feedback processes for shock statistics and the large impact they have on the shocks that are present. As we expect the strongest shocks to be efficient accelerators that produce non-thermal particle distributions, shock statistics combined with upper limits on cosmic ray densities could be developed into a powerful constraint for viable feedback models.

The role of cosmic rays in accelerating galactic outflows

Stellar feedback plays a critical role in galaxy- and star-formation through its regulation of the interstellar medium (ISM) and the powering of galactic winds. The sources of stellar feedback are varied and impart different types of energy on different timescales and in different environments. SNe are a particularly important feedback source, and their energy likely combines with other stellar feedback effects (e.g., UV radiation from young stars) in a non- linear manner to impact the ISM. The acceleration of CRs at shock fronts in supernova remnants is a potentially crucial aspect of SNe feedback. Observations of local SNe remnants suggest that approximately 10 % of the explosion energy is converted to CRs. In contrast to cooling processes that operate for thermal energy, CR energy does not dissipate quickly once created. In addition, CRs are transported through both advection and diffusion

processes. The diffusion process, in particular, has the ability to transport significant amounts of CR energy independent of bulk gas motions to distances far from CR acceleration sites, thereby creating potentially significant pressure imbalances that can drive large-scale gas flows.

In one of our central studies last year [Simpson et al., 2016], we investigated the formation of galactic outflows from supernova explosions (SNe) with the moving-mesh code AREPO in a stratified column of gas with a surface density similar to the Milky Way disk at the solar circle. We compared different simulation models for SNe placement and energy feedback, including cosmic rays (CRs). In our fiducial default model, the SNe energy is all thermal and distributed locally over the explosion cells, with the sites for SNe chosen probabilistically in accordance with the local star formation rate. Since the latter increases with density, SNe tend to explode in dense gas in this model. As a first variant, we considered a model differing only in that 10 % of the SNe energy was put into CR energy while the remaining 90 % was added as thermal energy. The CR energy here is assumed to be locally locked into the gas by a tangled magnetic field so that it essentially cannot move relative to the gas and is always advected with it. A much more realistic treatment is obtained if the CRs are allowed to diffuse along the direction of the magnetic field. This is possible at a technical level thanks to a robust anisotropic diffusion solver that we have recently developed. Finally, we contrast these models with a scenario in which no CRs are included and the locations of SNe are chosen randomly within the star-forming layer of the gas. In this manner, SNe often explode in gas of quite low density.

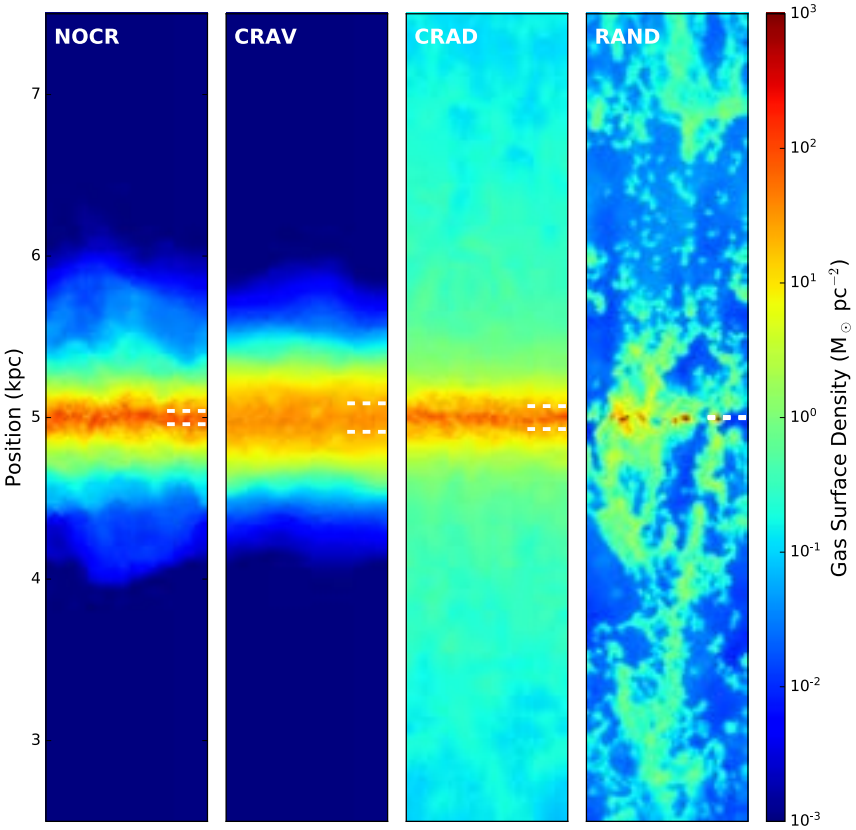


Figure 66: Unweighted projections of gas density for four of our models: no cosmic rays (NOCR), cosmic rays with advection (CRAV), cosmic rays with diffusion (CRAD), and a random placement of supernovae (RAND) after 100 Myr of evolution. The projections show the central 5 kpc of our simulated tall box and are 1 kpc wide and 1 kpc deep.

A visual overview of the behavior of these four primary models can be seen in *Figure 66*. If SNe are placed into dense gas, no outflows are produced, mostly because the injected energy is then quickly radiated away. If, instead, cosmic rays are also created, the gas is puffed up a bit because of the longer dissipation timescale of the cosmic rays, but otherwise, no significant change occurs. This is very different when we also account for CR diffusion. In this case, we are able to drive outflows with similar mass loading obtained from a random placement of SNe with no CRs, as seen in *Figure 67*. However, despite this similarity, CR-driven outflows differ in several other key properties, including their overall clumpiness and velocity. Moreover, the forces that drive these outflows originate in different sources of pressure, with the CR diffusion model relying on non-thermal pressure gradients to create an outflow driven by internal pressure and the random-placement model depending on kinetic pressure gradients to propel a ballistic outflow.

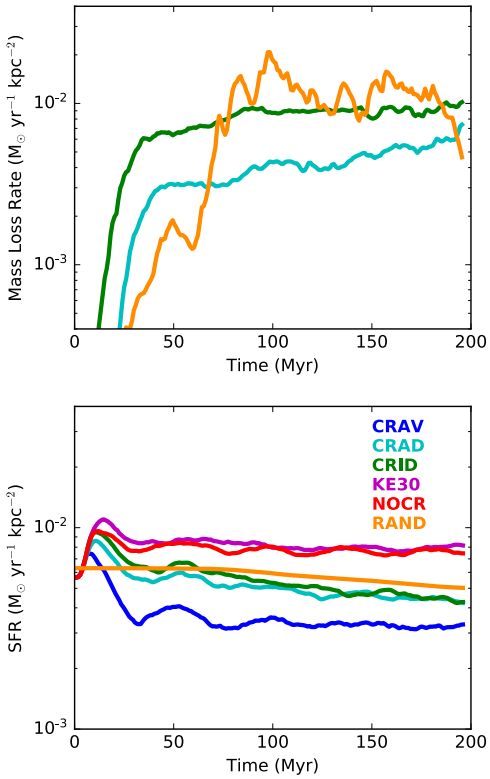


Figure 67: Time evolution of the mass loss rate in the outflow (top) and the total star formation rate (bottom). Only the models with isotropic or anisotropic cosmic ray diffusion (or a random placement of SNe) have sufficient mass loss to be included in the outflow rate panel.

Thus far, the physical nature of the driving mechanism of galaxy outflows remains unclear. Our results strongly suggest that cosmic rays represent non-negligible physics in the formation of outflows from the interstellar medium, and in fact, we may have identified one of the key mechanisms here. It is also tantalizingly interesting to note that the efficiency of CR-driven outflows increases rapidly towards smaller galaxies, a feature that is required in any successful theory of galaxy formation within the current standard model of cosmology.

Bias in Hubble constant measurements from lensing time delays

Strong gravitational lensing is a major tool in modern extragalactic astrophysics. Ever since its first discovery, it has been used as a natural telescope to magnify the distant Universe, as a scale to weigh galaxies, and as a ladder to measure the Hubble constant H_0 – the expansion rate of the Universe. With the current and upcoming space missions, including Gaia and Euclid, and ground-based facilities such as the Large Synoptic Survey Telescope and the Square Kilometer Array, strong lensing studies will experience an unprecedented opportunity to exploit several thousands of lensing galaxies, though only hundreds are currently known. This transition into the big data era demands a good understanding of and control over systematic errors that are introduced by lens modeling techniques.

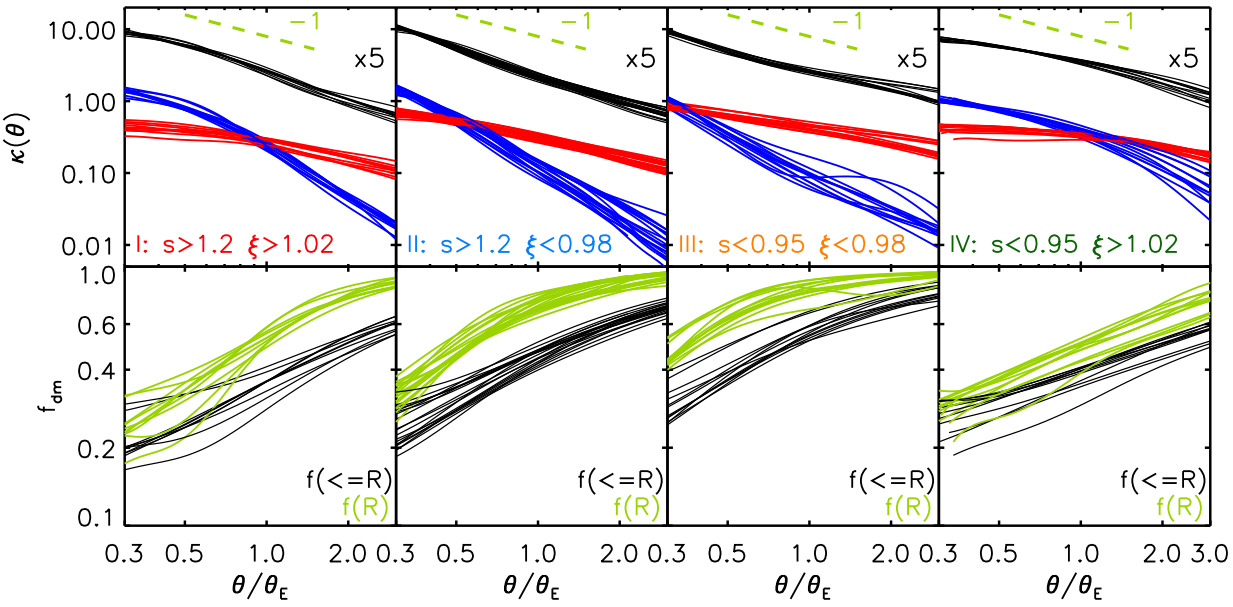
The observational properties of galaxy-scale strong lens systems (i.e., positions and flux ratios of unresolved images, brightness distribution of lensed extended components, and time-delays) depend mainly on the mass distribution inside and near the Einstein radius of the lens, which corresponds in general to a few half-light radii. Within this radius, both dark and baryonic matter are believed to co-exist in roughly similar amounts. However, because of the interplay between cooling and heating mechanisms, the unknown efficiency of star formation, and a wide variety of feedback processes, a large scatter in the galaxy total mass profiles is naively expected. Observationally, X-ray emissions of elliptical galaxies as well as results from galaxy dynamics combined with strong lensing measurements suggest a limited diversity of profiles and indicate that the inner profile is almost isothermal. This surprising result, also known as the "bulge-halo conspiracy," sets important constraints on the formation history of galaxies. In addition, it has an important impact on phenomenological studies of galaxies as it motivates simple power-law models of the volume mass density profile and of the corresponding convergence profile $\kappa(\theta)$.

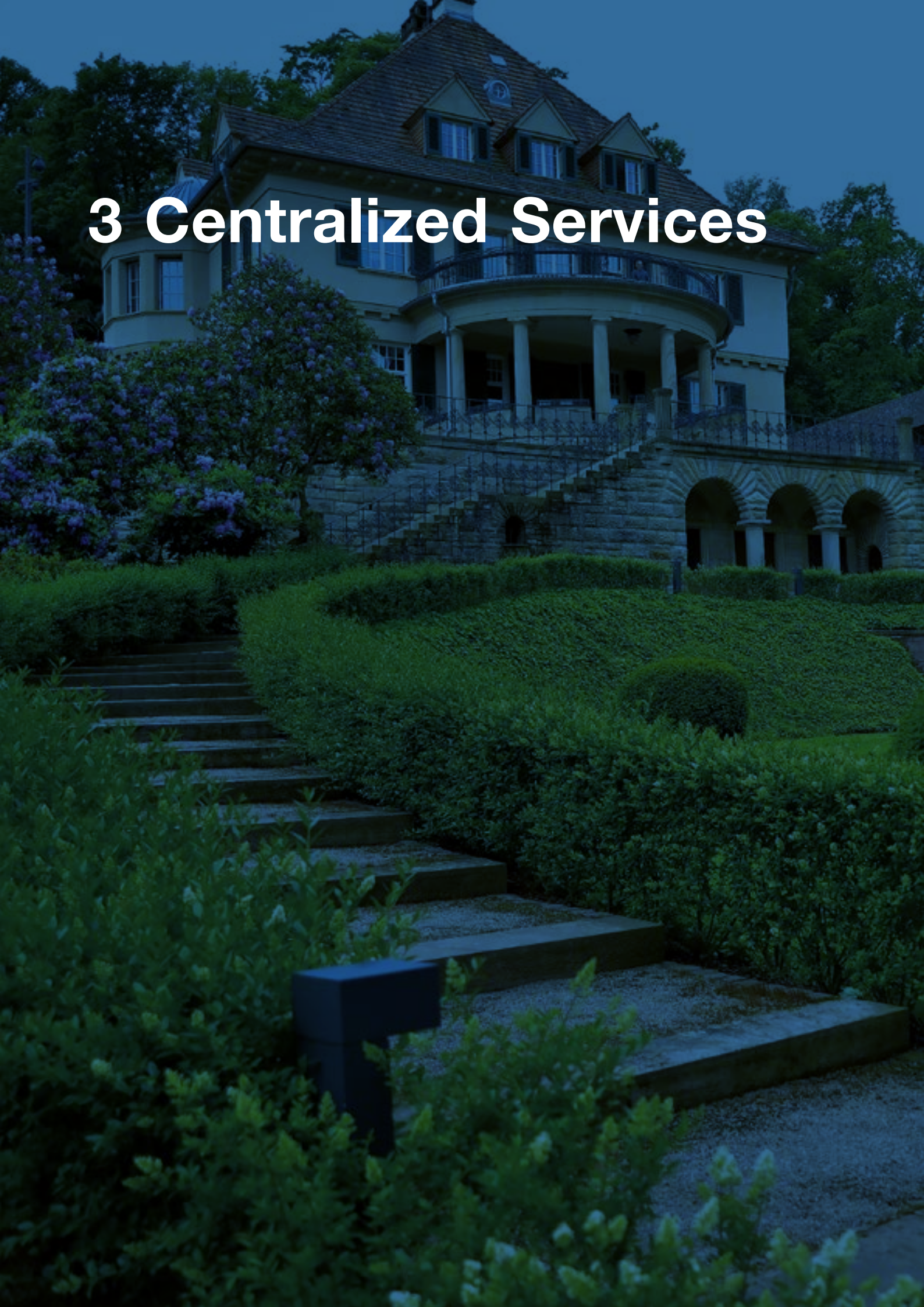
The power-law density profile, however, is no more than a convenient "working model." This family of density profiles does not uniquely represent the mass distribu-

tion of the inner regions of galaxies due to the so-called mass sheet degeneracy. Any given convergence field $\kappa(\theta)$ can be transformed into a new convergence profile $\kappa\lambda(\theta)$ via $\kappa\lambda(\theta) = \lambda\kappa(\theta) + (1 - \lambda)$. Together with an isotropic rescaling of the source plane, this transformation leaves all observable quantities, such as image positions and flux ratios, invariant – except for the time delays. Hence, the only observable quantity modified under the "mass sheet transformation" is the product of the Hubble constant H_0, and the time delay Δt in a lens that is transformed such that $H_0\Delta t \rightarrow \lambda H_0 \Delta t$. Because of the invariance of ordinary observables, information from gravitational lens properties cannot distinguish between any of the members of the family of mass profiles $\kappa\lambda(\theta)$. However, if a particular parameterized mass model is chosen, such as a (local or global) power-law profile, the mass sheet transformation yields that member of the family that is closest to the chosen mass profile parameterization. This artificial breaking of the mass sheet degeneracy can lead to a systematic bias in the determination of H_0.

In [Xu D et al., 2016], we studied the expected distribution of surface density profiles of galaxies drawn from the cosmological hydrodynamic simulation Illustris. By selecting galaxies that most resemble the observed lensing galaxies in terms of mass/velocity dispersion, we tested the validity of the power-law assumption in the central regions of galaxies and evaluated the expected distribution of the multiplicative bias λ on H_0. We found that the simulation reproduces numerous observational properties of real galaxies very well. In particular, we find that the cumulative dark matter fraction of the simulated lensing galaxies ranges from 30 % to 70 %, which complies well with observations (see *Figure 68*). The average isothermality of the profiles derived for the simulated galaxies is also in agreement with observational studies. We find that the bias on H_0 introduced by the power-law assumption can reach 20 – 50 %, with a rms-scatter of 10 – 30 %. However, we also found that by selecting galaxies with an inferred power-law model slope close to isothermal, it is possible to reduce the bias on H_0 to less than 5 %, suggesting a promising method to construct less biased statistical samples for H_0 measurements in the upcoming large survey era.

Figure 68: Ten typical surface density profiles and projected dark matter fraction distributions from each subgroup of our lens sample. In the top panels, the black curves show the total surface density distribution $\kappa(\theta)$ scaled up by a factor of 5 for clarity; the red and blue curves represent profiles of the projected dark matter and projected baryonic mass, respectively; the dashed green line indicates the logarithmic slope of s = 1. In the bottom panels, the distributions of the cumulative and local dark matter fraction are given in black and green, respectively.





3 Centralized Services

3.1 Administrative Services

3.2 IT Infrastructure and Network (ITS)

3.1
Administrative Services

The mission of HITS' administrative services is to provide administrative support and relief to the scientific groups. It handles a variety of human resource and finance tasks, clarifies legal issues, organizes events, and operates the buildings.

The HITS administration had a very "athletic" 2016. Not only was the institute scheduled to reach its final size in the third quarter – an opportunity for the team to demonstrate to the leadership that it could maintain its customary pace even at its new size – but during the first half of the year, the department was also tasked with smoothly passing the baton between the leaving and the succeeding Managing Director. Andreas Reuter, mastermind and creator of the HITS gGmbH (and formerly EML Research gGmbH; see *Chapter 10*), retired in mid-April of 2016, thereby leaving the group to Gesa Schönberger, who had joined the team as Managing Director as well as new group leader in January 2016. We thank Andreas Reuter for his many years of trusting cooperation with the admin team, which he had headed as a provisionary group leader since 2013. At the same time, we are delighted that Andreas will continue his association with HITS: As member of the board of directors of the HITS Stiftung he is a representative of the HITS majority shareholder, lending his extensive and diverse expertise in an advisory capacity.

Towards the end of the year, another baton was passed: Rebecca Wade (see *Chapter 2.8*), HITS' first Scientific Director, retired from her position upon the expiration of her two-year term. Michael Strube (see *Chapter 2.9*) took over, and Wolfgang Müller (see *Chapter 2.12*) became his deputy. In addition to Gesa Schönberger's settling into her new role, the team also used these dynamics of change to hand over established and proven routines as well as future tasks without any major loss of productivity.

As had been planned for some time, HITS opened a second site at the Mathematikon at the edge of Heidelberg University Campus "Im Neuenheimer Feld" to take advantage of the proximity to the research institutions based there. One of HITS' most long-standing groups, Scientific Databases and Visualization (SDBV; see *Chapter 2.12*), moved into the new site in September 2016.

HITS on the move: one of the labeled boxes bound to be moved to the Mathematikon.

HITS opened a second site at the Mathematikon, Berliner Strasse.

Group Leader

Dr. Gesa Schönberger

Staff members

Christina Blach *(Office)*
Christina Bölk-Krosta *(Controlling)*
Benedicta Frech *(Office)*
Ingrid Kräling *(Controlling)*
Kerstin Nicolai *(Controlling)*
Thu Phuong *(Student assistant since August 2016)*
Rebekka Riehl *(Human Resources and Assistant to the Managing Director)*
Stefanie Szymorek *(Human Resources)*

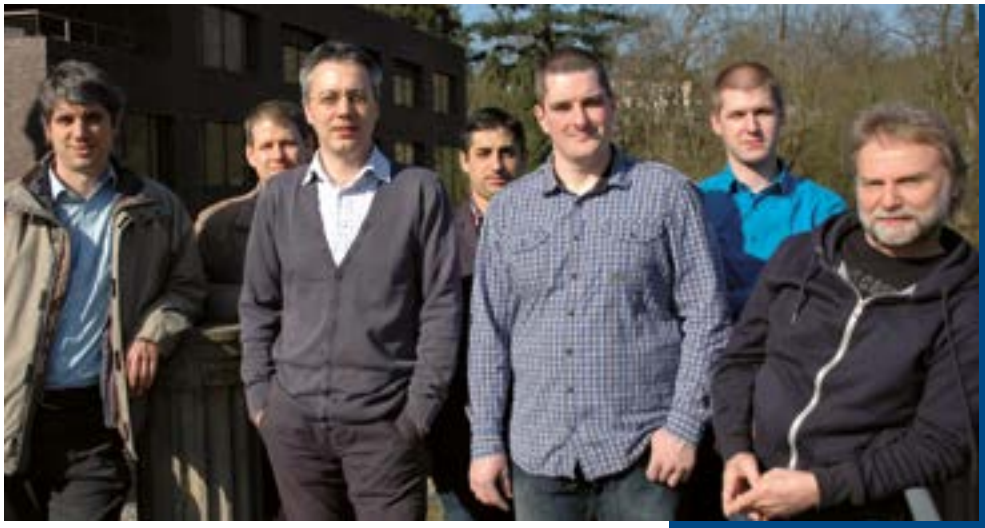
At the housewarming party: Managing Director Gesa Schönberger (left) with DMQ group leader Vincent Heuveline.

Scientific Director Rebecca Wade during her welcome address.

All HITSters and selected guests were invited to a small housewarming party in November to learn more about the work environment and conference facilities available at the new location. Every move, no matter how small, involved a great deal of planning and organization, execution, and follow-up activities, which the admin team mastered well despite temporary bottlenecks regarding facility maintenance.

Another major change that required planning and execution occurred in July when EML European Media Laboratory GmbH moved out of Villa Reiner, where it had been renting premises at HITS. The freed-up workspace was renovated and has since become available to the steadily growing HITS groups.

In addition to integrating new leadership, the administration faced additional challenges: Since a colleague in controlling was on extended sick leave, tasks had to be reassigned, and processes had to be renegotiated and tested. Fortunately, we were able to hire student assistant Thu Phuong, who provided extra support beyond the end of the year. In a traditional relay, each team runs a predetermined distance, with each runner carrying the baton during his or her leg of the race. Comparatively, in December of 2016, 146 people worked at HITS in 13 research groups and two service groups. 66 researchers, 24 student assistants, and 18 administrative, communications, and IT staff members were employed. HITS also hosted 18 HITS-scholarship recipients and 20 visiting scientists. They all played their part in close coordination with one another to ensure that the metaphorical HITS baton made it across the finish line. In 2016, the administrative services team put their heart and soul into the effort and managed to jump all upcoming hurdles in the process.



Group Leader

Dr. Ion Bogdan Costescu

Staff members

Dr. Bernd Doser *(Software Developer)*
Dr. Christian Goll *(System Administrator)*
Cristian Huza *(System Administrator)*
Norbert Rabes *(System Administrator)*
Andreas Ulrich *(System Administrator)*
Dr. Nils Wötzel *(Software Developer)*

Students

Stevan Bratic
Philipp Grüner *(until August 2016)*

2016 was marked by a simultaneous expansion in two directions: the opening of a new HITS site in the Mathematikon and the hosting of a part of our HPC resources at the Computing Center of the Heidelberg University.

The continued growth of HITS over the past years has led to a shortage of desks, which was dealt with by expanding in Villa Reiner and by opening a new HITS site in the Mathematikon. Together with our colleagues from Villa Bosch, we have expanded the network infrastructure and made the associated services available in the new locations, while aiming to provide the HITS groups moving there with a "plug-and-play" experience. To cope with the growing traffic and the number of internal high-speed connections, we have upgraded the core of our network infrastructure and have switched the main Internet connection to a 10 Gbit/s link. The HITS offices from Mathematikon have complete WiFi coverage and offer the same network names available on the HITS campus, thus allowing for a seamless transition between the two sites.

Due to the increased needs of the scientific groups for computational and data-storage resources and the limited capacity of the HITS computer room, new cluster nodes and storage servers are hosted at the Computing Center of the Heidelberg University, a HITS shareholder. The cluster nodes offer a total of 3584 Intel Haswell cores and more than 14 TB RAM. The storage servers provide 1.3 PB in a BeeGFS file system. All nodes and servers are interconnected with FDR Infiniband in a full-bisection-bandwidth network, enabling optimal communication performance. They are also linked to the HITS computer room through a 10 Gbit/s dedicated connection, for fast data transfers and easy access to the remote nodes.

We have migrated the cluster home directories to new, highly-available servers, which use an internally developed software solution that enables a two-to-five-time reduction in costs compared with similar commercial solutions. The servers use flash-based storage and connect to the rest of the cluster through Infiniband and 10 Gbit/s links, giving users quick access to their files and thus addressing the main complaint regarding the old storage system. To optimize the workflows and, in particular, to reduce (mis)communication around managing user accounts, involving the ITS group, the administration, and the scientific group leaders, we have developed a web service that centrally stores user-related data, keeps it synchronized to our Active Directory, and sends notifications to the relevant individuals.

With changes affecting the institute's cluster, network, storage, and web apps already planned for 2017, maintaining and extending the institute's IT infrastructure and services will continue to keep us busy throughout the coming year.

4 Communication and Outreach


KLAUS TSCHIRA
Heidelberg Institute
for Theoretical Studies

4 Communication and Outreach

HITS Communications team in 2016 (f.l.t.r.): Anna Giulia Deutsch, Peter Saueressig, Isabel Hartmann.

Head of Communications

Dr. Peter Saueressig

Staff member

Isabel Hartmann *(public relations)*

Student

Anna Giulia Deutsch *(until December 2016)*

The HITS communications team pursues two main goals. On the one hand, we are engaged in establishing "HITS" as a name for a small but excellent research institute with an international, interdisciplinary, and inspiring atmosphere. On the other hand, we draw the attention of the media to the excellence of our scientists in their respective fields.

In 2016, the HITS researchers helped us significantly in achieving this twofold goal. Some highlights are presented below. We were proud to announce once again that some of our scientists rank among the group of most highly cited researchers worldwide. According to the "Highly Cited Researchers" report by Clarivate Analytics (formerly part of the Thomson Reuters group), Tilmann Gneiting, Computational Statistics (CST), Volker Springel, Theoretical Astrophysics (TAP), and, for the first time, Alexandros Stamatakis, Scientific Computing (SCO), ranked among the scientists most cited in their subject field and year of publication. These citations are important indicators of the scientific impact of a publication.

HITS Scientific Director Rebecca Wade, Molecular and Cellular Modeling (MCM), was awarded the International Society of Quantum Biology and Pharmacology (ISQBP) Award in Computational Biology. This award is given biannually to a researcher who has performed high-quality research in a field of interest to the ISQBP and has served the society or the interests of the society.

Volker Springel was elected to the German Academy of Sciences Leopoldina. The Leopoldina is one of the oldest academies of science in the world and is dedicated to the advancement of science for the benefit of humankind and to the goal of shaping a better future. Volker Springel is also one of the coordinators of the new HITS-Yale program in astrophysics, funded by the Klaus Tschira Stiftung (see *Chapter 6*).

Media relations: Journalist in Residence program

We firmly believe that an important prerequisite for successful science communication is the development of reliable and sustainable journalistic contacts. The "Journalist in Residence" program thereby represents an important project for HITS. The program is addressed to science journalists and offers them a paid sojourn at HITS. During their stay, they can learn more about data-driven science and get to know researchers and new research topics without the pressure of the "daily grind."

In February 2016, HITS welcomed its fifth Journalist in Residence, radio journalist Michael Stang from Cologne (Germany), for a four-month stay. Michael Stang held an internal seminar for the HITSters on radio science journalism and described the opportunities and challenges for journalists and scientists alike when it comes to reporting on science topics in an adequate manner for radio listeners. Moreover, he gave a public speech on the current problems radio science journalists are facing.

Michael Stang (left) and Alexander Mäder.

Michael Stang used his stay for intensive topical research, interviewing members of all 13 scientific groups. Moreover, he developed a concept for a new program that would enable science journalists to stay at institutes and at editorial offices or broadcast stations, combining features of our Journalist in Residence program with ideas he had developed with his colleagues at the German Science Journalists Association WPK. In the spring, HITS also hosted Alexander Mäder, the designated chief editor of the "Bild der Wissenschaft" magazine at the time, for a two-week research stay. During his time at HITS, Alexander Mäder collected information for a book on astronomy and astrophysics by interviewing HITS astro scientists Volker Springel, Friedrich Röpke, Kai Polsterer, and Christoph Pfrommer. He also gave a public speech on online journalism and how journalists are not only gatekeepers nowadays but are also increasingly responsible for directing a spotlight on relevant science issues.

In addition, we were thrilled that Michele Catanzaro, HITS Journalist in Residence 2014, was named European Science Writer of the year 2016. He received the prize for a cross-border journalism project that he had begun during his time here in Heidelberg.

Speaking of crossing borders: Our next Journalist in Residence will be Padma TV, a science journalist from Delhi (India), who writes for "Nature", "Physics World", and "New Scientist", among other journals. One of her goals for her stay is to develop a fruitful dialog between the Asian and the European perspectives on science and science communication. She will commence with her stay at HITS in February 2017.

Outreach activities

Another important part of our work consists of outreach events throughout the year that are targeted at different groups. Every two years, HITS organizes an open-house event for the public at large. The diverse program of the event included hands-on stations for kids, presentations by HITS research groups, and talks in addition to a science slam (see *Chapter 5.3*).

Do you speak Mathematics?

"Do you speak Mathematics?" was the title of the so-called Academic Lunch Break (Akademische Mittagspause) during the summer term of 2016 and consisted of 61 talks from different scientists in the fields of mathematics and computer science. HITS organized the event together with the Interdisciplinary Institute for Scientific Computing (IWR) and the Mathematics Center Heidelberg (MATCH). HITS scientists from almost all groups contributed talks to this event series, presenting different mathematical and computational aspects of their scientific fields, which included geometry, probabilistic forecasting, de-novo genome sequencing, and number-crunching cosmological simulations. The talks were held in St. Peter's Church (Peterskirche), the traditional church of Heidelberg University and the city's oldest house of prayer.

Anna Wienhard "speaking mathematics" in St. Peter's church.

30 girls visited HITS at the "Girls' Day" 2016.

We need more STEM girls

On April 28, HITS once again participated in the national "Girls' Day." During this event, which takes places once a year, various institutions offer workshops for girls to introduce them to jobs in which women are still underrepresented. The goal of the event is to broaden the minds of young girls and to get them interested in, say, a STEM subject, such as research at HITS. In 2016, thirty girls between the ages of ten and sixteen visited us to participate in the workshops and find out what it's like to be a HITS researcher. Scientists from four research groups (CST, MCM, SCO, and TAP) offered small-scale, hands-on workshops to show the girls what the daily work and life of a researcher looks like in fields ranging from bioinformatics and molecular biology to mathematics and astrophysics. After the workshops, the girls had lunch with the scientists and enjoyed a guided tour through the HITS server room.

Science for everyone

As in previous years, HITS participated in the Explore Science event, which once again took place in Mannheim's Luisenpark from June 15 – 19. The event is geared towards children, secondary-school students, and their families. Organized by the Klaus Tschira Foundation, Explore Science offers various hands-on stations, exhibitions, and presentations designed to get youngsters interested in the natural sciences. In 2016, Explore Science's motto was "The Human Universe" ("Universum Mensch"), and it attracted more than 50,000 visitors from all over the region. HITS researchers from three groups (MBM, MCM, and SDBV) presented the topic in three interactive hands-on stations. In a board game, for instance, the visitors had to collect all the nutrients that the human body needs – be they proteins, fat, or carbon – by rolling the dice and moving their pawns. Another hands-on station explained how the membranes in the body function via the use of matches, and at the third station, young and old visitors could participate in a picture quiz game and learn that the human brain is an excellent image-processing computer.

Another event conceived by the Klaus Tschira Stiftung is the Heidelberg Laureate Forum (HLF), the 4th installment of which took place in September. As of this year, the institute, together with Heidelberg University, is a scientific partner of the HLF. Among other activities, we again hosted a group of young researchers from all over the world (see *Chapter 6*).

Explore Science: Csaba Daday (MBM) explaining membrane functions using matches.

Christopher Zapp (MBM) talking to the freshmen in physics.

The Mathematikon (Photo: Mathematikon Heidelberg GmbH (n. Schw. R.) & Co KG).

Raising the next generation

School students and University newcomers with an interest in science and computing are also an essential target group for our institute. In August, the institute again participated in the International Summer Science School Heidelberg, where Dr. Camilo Aponte (MBM) gave a scientific talk on Molecular Biomechanics for the 20 selected school students from Europe, the USA, and Australia.

In October, HITS invited a group of 20 freshmen in physics from our shareholder, Heidelberg University. The event was organized with the student union of the Department of Physics and Astronomy. Junior HITS researchers Christopher Zapp (MBM), Rainer Weinberger (TAP), and Martin Reinhardt (MCM) discussed their experiences at HITS and the importance of programming skills when studying physics.

Finally, in November, HITS had the chance to host the series "Physics on Saturday Morning" ("Physik am Samstagmorgen"), organized by the Max Planck Institute for Nuclear Physics in Heidelberg. 90 school students between 15 to 17 years old from schools in Heidelberg and the surrounding towns listened to talks given by HITS astrophysicists Christoph Pfrommer, High-Energy Astrophysics and Cosmology (HAC), and Volker Springel (TAP).

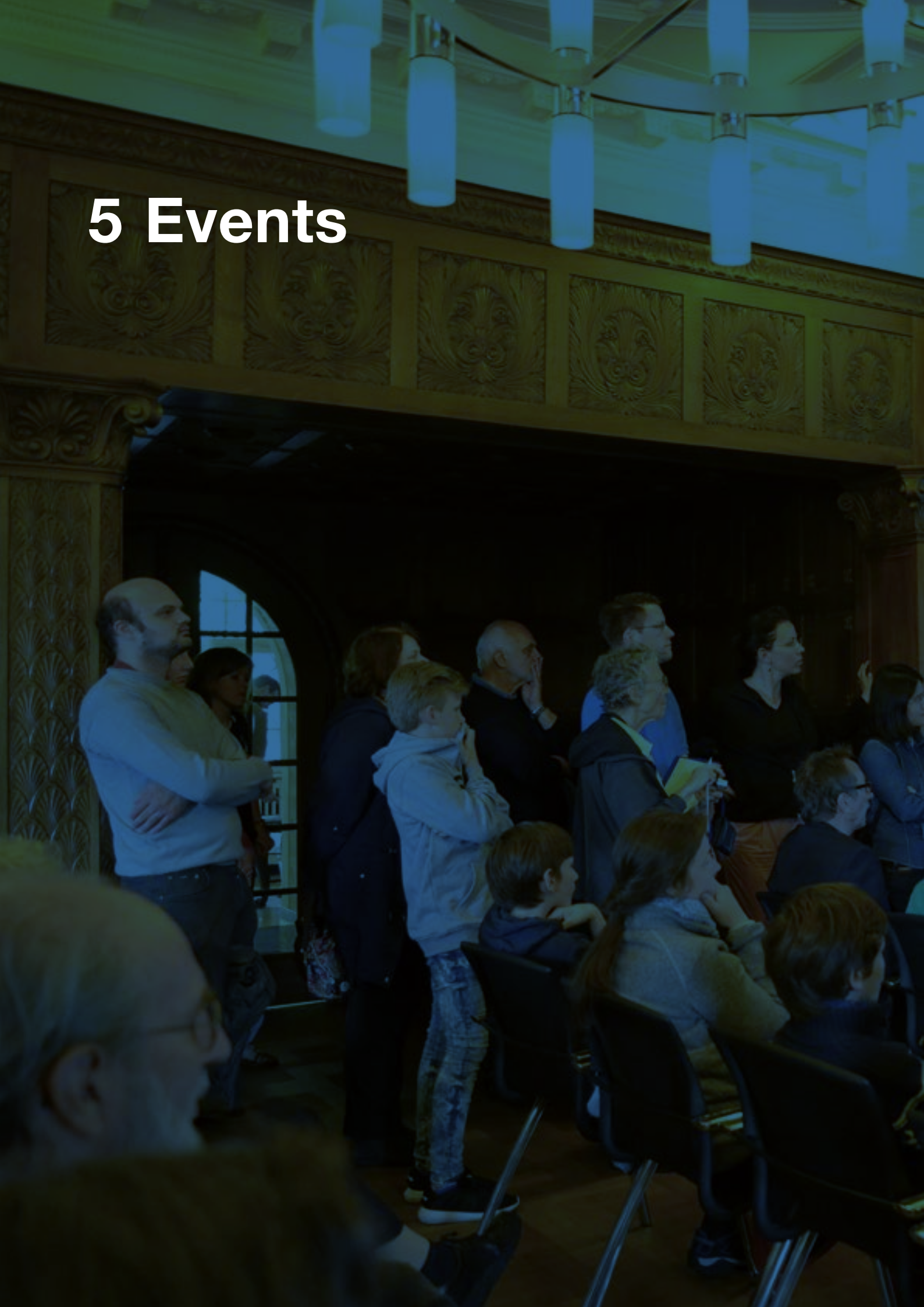
Mathematikon: HITS' second site

The foundation of excellent external communication has always been internal communication. If you don't reach your colleagues, you won't succeed in reaching out at all. In September 2016, HITS opened a second site on Heidelberg's university campus in Neuenheimer Feld. We now have offices in the "Mathematikon" at Berliner Str. 45, which strengthens our presence on the University campus. The Scientific Databases and Visualization (SDBV) group, which works in close collaboration with scientists at Heidelberg University and the German Cancer Research Center (DKFZ), has also relocated to the Mathematikon. Moreover, there is a conference room for HITS internal meetings and for an exchange with partners from the University and other institutes on the campus. The HITS communications team joined forces with the administration to ease the relocation for the scientists involved and to make them feel at home in their new environment. This includes a second communication office at the Mathematikon to ensure that the communication both within HITS and with the outside community flows smoothly.

Old ties and companions

HITS is also eager to keep in touch with its alumni and friends via various online and social media channels and by providing former employees with printed materials, including "The Charts" newsletter and the Annual Report. Among other activities, there will be a two-day alumni meeting in the summer of 2017 that will focus on interactions between alumni and current HITSters and will feature career talks by alumni as well as social activities and group reunions.

5 Events



5.1 Conferences, Workshop & Courses

5.2 HITS Colloquia

5.3 HITS Open House Event

Participants, teaching assistants, and lecturers of the 2016 school on computational molecular evolution in front of the main entrance to the Hellenic Center for Marine Research in Crete.

5.1.1 EMBO Practical Course on Computational Molecular Evolution

May 8–19, 2016, Heraklion / Greece

The EMBO Practical Course took place for the 8th time at the Hellenic Institute of Marine Research near Heraklion, Crete, and provided graduate and postgraduate researchers with the theoretical knowledge and practical skills to carry out molecular evolutionary analyses on sequence data. The course offered the opportunity for direct interaction with some of the world-leading scientists and authors of famous analysis tools in evolutionary bioinformatics, including Olivier Gascuel, Nick Goldman, Bruce Rannala, and Ziheng Yang. Alexandros Stamatakis (SCO group leader) was the main organizer of this event, for which we received 200 applications for 35 available spaces. Paschalia Kapli and Alexey Kozlov as well as former SCO postdoc Pavlos Pavlidis contributed substantially as teaching assistants. HITS was the main co-sponsor of this course.

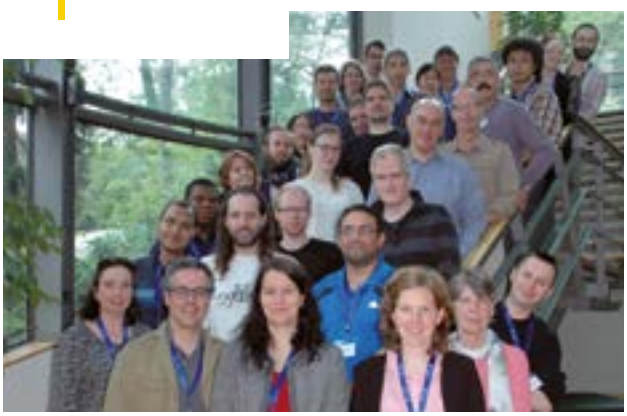
5.1.2 Kinetics on the Move: Workshop on Data for Computational Modelling

May 30–31, 2016, Heidelberg / Germany

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

de.NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE



For the 10th Anniversary of SABIO-RK (Biochemical Reaction Kinetics Database), the Scientific Database and Visualization (SDBV) group of HITS hosted the "Kinetics on the Move" Workshop in the Studio Villa Bosch. It was organized as a training course for experimentalists and modelers by the Data Management Node (de.NBI-SysBio) of the "German Network for Bioinformatics Infrastructure – de.NBI" and supported by both the FAIRDOM project and HITS.

During her keynote, Ursula Kummer, professor at Heidelberg University and former group leader at EML Research, commemorated Isabel Rojas (1968 – 2013), who founded the SDBV group in 1999 and began SABIO-RK 10 years ago. After her overview of SABIO-RK and Modelling of Biological Systems, the following talks and hands-on sessions focused on the publication, curation, retrieval, and usage of kinetic data from the reaction kinetics database SABIO-RK as well as on the incorporation of data in modelling. The second day was dedicated to data management, including best practice in data and model storage and reusability by introducing the SEEK system and incorporated tools. The talks and hands-on sessions were enriched by questions, lively discussions, and feedback given by the approximately twenty workshop participants.

CST group member Dr. Fabian Krüger welcoming all participants to the workshop in the Studio Villa Bosch.

5.1.3 Workshop Advances in Economic Forecasting

June 6, 2016, Heidelberg / Germany

When talking about forecasting, most of us associate it with the weather, come rain or shine. However, forecasting is also a big issue in the economic sector, where forecasts inform the decisions of policymakers, businesses, and consumers alike. On June 6, 2016, the Computational Statistics (CST) group at HITS (led by Prof. Tilmann Gneiting) hosted a workshop on "Advances in Economic Forecasting" with support by the European Research Council (ERC) Advanced Grant "ScienceFore".

After a welcome by Dr. Fabian Krüger (CST group member) in the Studio Villa Bosch, 17 experts discussed different aspects of forecasting, focusing on recent techniques for combining and evaluating forecasts. One of the crucial topics was how probabilistic forecasting can help to predict macroeconomic developments and provide realistic assessments of the associated forecast uncertainty.

5.1.4 COMBINE & de.NBI Tutorial: Modelling and Simulation Tools in Systems Biology at the International Conference on Systems Biology (ICSB)

September 16th, 2016, Barcelona / Spain

Systems biologists from all around the world gathered at the 17th International Conference on Systems Biology (ICSB) that took place in Barcelona, Spain, from September 16th to 20th, 2016 (http://www.icsb2016barcelona.org). At the conference, HITSter Martin Golebiewski (SDBV) organized the one-day COMBINE & de.NBI Tutorial Workshop "Modelling and Simulation Tools in Systems Biology" as a satellite (http://co.mbine.org/events/tutorial2016). The tutorial showed over 60 attendees, comprised of mainly young scientists from many different countries, how to set up computer models of biological networks and simulate these models in different systems biology platforms. Lectures, software demonstrations, and hands-on sessions provided the attendees with the necessary skills to create, simulate, and handle such models. An international team of tutors instructed the scientists how to use modelling and simulation tools as well as databases and demonstrated how to use standardized formats to describe and exchange the models. The SDBV group (Andreas Weidemann, Ron Henkel, Olga Krebs, and Martin Golebiewski) presented their SABIO-RK database for reaction kinetics data and the SEEK system (FAIRDOMHub) for integrated data and model management. Other software tools that were introduced included COPASI for the simulation and analysis of biochemical networks and their dynamics, the BioModels database for computational models, as well as CellDesigner and JWS Online, two widespread modelling and simulation platforms.

As an important prerequisite for the information and data exchange amongst the scientists, domain-specific community standard formats were demonstrated to be crucial for the data and model exchange. Such community standards for modelling in the life sciences are defined by the COMBINE network (http://co.mbine.org), which co-hosted this tutorial workshop.

5.1.5 de.NBI / FAIRDOM Systems Biology Developer's Foundry

December 1–2, 2016,
Frankfurt am Main / Germany

The de.NBI / FAIRDOM Systems Biology Developer's Foundry is an innovative workshop format organized by PIs from de.NBI-SysBio and FAIRDOM that brings together practitioners from the interdisciplinary fields of systems biology data management and tool development. After successful workshops in Vienna in 2012 and Heidelberg in 2014, 16 developers met at the Frankfurt am Main train station for two days packed full of discussions about development. On the first day, we showed each other the results of our respective development efforts in interactive sessions that combined the advantages of scientific speed dating and classical demo sessions. On the second day, we had follow-up discussions about a topic chosen by the participants (Scientific Workflow Systems) followed by a tutorial about Docker (lightweight) containers for the deployment of scientific software delivered by Bérénice Batut from the University of Freiburg. After two days of intense discussion, the participants went home with plenty of ideas, knowledge about new tools, and as well as new collaboration partners.

Bérénice Batut, University of Freiburg, during her lecture.

5.1.6 NormSys Conference "Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences"

December 5–6, 2016, Potsdam/Germany

The NormSys conference aimed at exploring the relation between standardization and innovation in the life sciences, with special focus on the role of standards for the translation of scientific results into commercial products for industries. Representatives from academia (universities and research institutes), industry (Big Pharma and service providers), European research infrastructures and funding organizations, and standardization bodies (DIN) that have been making exceptional contributions in the development and dissemination of standards all discussed with the participants the pros and cons and importance of the use and implementation of standards in the daily work flow. The conference was organized by all partners of the German NormSys project, coordinated by Martin Golebiewski (SDBV), and HITSters Ron Henkel and Olga Krebs also presented talks. It was held in connection with the tutorial "Managing Big Data" that provided attendees with practical insights into Standard Operating Procedures (SOPs), good experimental practices, modelling techniques, and tools for data and model management.

Supported by:

Federal Ministry
for Economic Affairs
and Energy

on the basis of a decision
by the German Bundestag

NORMSYS

5.1.7 Workshop on Supernovae and Related Topics

December 8–9, 2016,
Würzburg/Germany

On December 8th and 9th, 2016, the Physics of Stellar Objects (PSO) group organized the XI Würzburg Workshop on Supernovae and Related Topics in collaboration with Prof. Christian Klingenberg's Fluid Dynamics Group at the Mathematics Department of Würzburg University. This series of biannual meetings provides a forum to discuss the latest developments in the PSO group's research areas and is intended to bring together its national and international collaborators. The meetings focus on giving the younger members of the group the opportunity to present their projects. More than 20 researchers from HITS, the University of Würzburg, the Max Planck Institute for Astrophysics Garching, Stockholm University, and Queen's University Belfast participated in this workshop.

Prof. Christian von Mering

Institute of Molecular Life Sciences, University of Zürich, Switzerland

February 15, 2016: The "Species Problem" in Microbiology – Limitations and Improvements of Operational Taxonomic Units as a Practical Solution

Prof. Tom Jordan

Southern California Earthquake Center, Los Angeles, USA

March 21, 2016: The Prediction Problems of Earthquake System Science

Dr. Pavlos Protopapas

Harvard School of Engineering and Applied Sciences, Cambridge, USA

April 18, 2016: Classifying Everything in the Sky

Dr. Alexander Mäder

Science Journalist, Bild der Wissenschaft, Stuttgart, Germany

April 20, 2016: Quotenjournalismus – Wie Redaktionen mit den Daten über ihr Publikum umgehen sollten

Prof. Iain D. Couzin

Max Planck Institute for Ornithology, Konstanz, Germany

May 23, 2016: Collective Sensing and Decision-Making in Animal Groups: From Fish Schools to Primate Societies

Michael Stang

Freelance Science Journalist, Köln, Germany

June 1, 2016: Dudelfunk statt Inforadio? Qualität im Hörfunk in Zeiten des Spardiktats

Prof. Tim Palmer

Department of Physics, University of Oxford, UK

June 20, 2016: More Accuracy with Less Precision: Inexact Computing and Hard Science

Prof. Achim Zeileis

Faculty of Economics and Statistics, University of Innsbruck, Austria

July 18, 2016: Forecasting Sports Tournaments by Ratings of (Prob)abilities

Prof. Janet Thornton

EMBL-EBI The European Bioinformatics Institute, Hinxton, Cambridge, UK

October 26, 2016: The Evolution of Enzyme Function

Prof. Matthias Rarey

Centre for Bioinformatics, University of Hamburg, Germany

November 21, 2016: Novel Tools for Browsing Large Protein Structure Collections

Prof. Alexander Szalay

The Johns Hopkins University, Baltimore, USA

December 5, 2016: Data Driven Discoveries in Science: The Fourth Paradigm

Prof. Christian von Mering

Prof. Tom Jordan

Dr. Pavlos Protopapas

Dr. Alexander Mäder

Prof. Iain D. Couzin

Michael Stang

Prof. Tim Palmer

Prof. Achim Zeileis

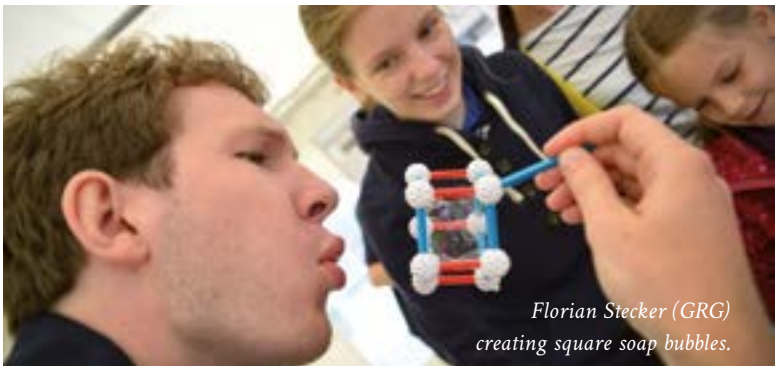
Prof. Janet Thornton

Prof. Matthias Rarey

Prof. Alexander Szalay

On July 4, 2016, HITS opened its doors to the public and showed interested people from Heidelberg and the surrounding area the life and work of an international research institute. The program consisted of hands-on stations at several locations on the HITS campus from twelve different research groups, guided tours through the HITS main building, three research talks, as well as one so-called science slam. Despite the rainy summer day, the HITS campus was filled with people eager to get to know and interact with the HITS researchers and to learn about their scientific work.

The hands-on stations couldn't have been more diverse: From galaxy travels with the Oculus Rift virtual reality glasses to 3D molecules, geometrical soap bubbles, and a statistical forecast game, the Open House event offered a plethora of activities for families. In addition to the hands-on stations, Volker Springel (Theoretical Astrophysics group), Christoph Pfrommer (High-Energy Astrophysics and Cosmology group), and Christopher Zapp (Molecular Biomechanics group) presented talks on their research activities. The Klaus Tschira Library, where the talks took place, was constantly filled with visitors, some of whom listened to all the talks in a row. In their presentations, Volker and Christoph took the visitors on a journey through the history of our universe and shed light on the mysteries of dark matter, dark energy, and black holes. Whereas both astrophysicists painted the big picture, Christopher Zapp explained how he and his research group conduct research on the very small elements of life, such as proteins. In his talk, Christopher revealed how external forces, such as gravity, influence how the proteins in our body work. Lucas Czech (Scientific Computing group) had yet another approach to the origins of life and held the first "science slam" at HITS to deal with how scientists decipher the tree of life from tropical soil. Science slams are short but witty science talks with many jokes and "geeky" references that enable spectators to learn about "hard" science in a delightful and entertaining way. All talks proved to be fun for visitors and speakers alike.



Florian Stecker (GRG) creating square soap bubbles.

Kai Polsterer (AIN) taking a visitor on a galaxy travel.

Creating molecules with origami paper.

Lucas Czech (SCO) giving his science slam.

To "round off" of the event, Rebekka Riehl (Administrative Services) led several groups around the HITS building in half-hourly guided tours, explaining the history of HITS and its campus. As in years past, the culinary offers were another of many of the event's highlights. HITS chef Ralf Westermann prepared his famous mini pizzas and other tasty dishes and ensured that no one stayed hungry throughout the event.

The HITS-Yale connection: The two fellows Freeke van de Voort (left) and Nir Mandelker (right) gather round Volker Springel.

HITS-Yale Program in Astrophysics

In the summer of 2016, HITS and Yale University, New Haven / U.S., began a joint initiative to foster a collaboration between astronomers and astrophysicists on both sides of the Atlantic: the HITS-Yale Program in Astrophysics, funded by the Klaus Tschira Stiftung (KTS). The program is jointly run by Volker Springel, head of the Theoretical Astrophysics group, and Frank van den Bosch, Associate Professor of Astronomy at Yale University. It is initially funded for a 4-year period, which includes funding for two "Tschira" postdoctoral fellows, a position named after the physicist Klaus Tschira (1940 – 2015), the founder of the KTS and HITS. The foundation promotes the advancement of natural sciences, mathematics, and computer science. Each fellow will spend two years at Yale and two years at HITS for a research stay. The two Tschira fellows selected for this program are Freeke van de Voort and Nir Mandelker.

Across the Universe, Across the Atlantic

Freeke van de Voort, who is currently a postdoctoral fellow at University of California Berkeley, began her fellowship at HITS in October 2016. She will go to Yale University in the summer of 2018. Freeke earned her PhD in 2012 from the University of Leiden and is an expert in hydrodynamical simulations of galaxy formation.

Nir Mandelker finished his PhD Thesis at the Hebrew University of Jerusalem with Prof. Avishai Dekel. He began his fellowship at Yale University concurrent with Freeke at HITS, and he will move to Heidelberg in the summer of 2018. Nir is an expert in galaxy formation and has worked on violent disk instabilities, compressible turbulence, and instabilities in cold streams.

In addition to the fellows and their research stays, the collaboration will also include collaboration meetings during which astrophysicists and astronomers from HITS and Yale visit each other at their respective institutions to exchange ideas and work together.

After signing the contract (left to right): Bernhard Eitel (president of Heidelberg University), Prof. Andreas Reuter (Scientific Chairperson Heidelberg Laureate Forum Foundation), Beate Spiegel (Chairperson Heidelberg Laureate Forum Foundation), Gesa Schönberger (HITS Managing Director), and Rebecca Wade (HITS Scientific Director). (Photo: Heidelberg Laureate Forum Foundation / B. Kreutzer)

Group picture with the HLF young researchers and their HITS hosts.

Heidelberg Laureate Forum

Since 2013, the Heidelberg Laureate Forum (HLF) has taken place annually in Heidelberg. At this networking event, award-winning scientists from mathematics and computer science come together to exchange ideas with each other and selected young scientists and students. HLF is organized by the Heidelberg Laureate Forum Foundation (HLFF). As of September 2016, HITS and Heidelberg University are scientific partners of the foundation, and both partners contribute their scientific expertise. The partnership agreements were signed shortly before the opening of the 4th HLF on the 18th of September, 2016, in the auditorium of the New University of Heidelberg.

HITS: From co-initiator to scientific partner

The HLF resulted from a joint initiative of the Klaus Tschira Foundation, which supports both the HLF and HITS. Prof. Andreas Reuter, founding director and longtime manager of HITS, has been involved from the beginning. The Heidelberg Laureate Forum Foundation (HLFF) was founded by the Klaus Tschira Foundation in 2013. The HLFF organizes the event, and Andreas Reuter serves as scientific chairperson. According to the newly signed agreement, HITS is in charge of continuing the scientific support of the networking event and will continue to contribute its world-renowned expertise in mathematical and computational topics – just as intended by its founder, Klaus Tschira.

Young researchers are welcome

As in previous years, HITS hosted a group of young researchers from HLF 2016, from undergraduate students to postdoc scientists. They enjoyed five group presentations and a poster session, where members of four HITS groups presented their current research topics and publications.

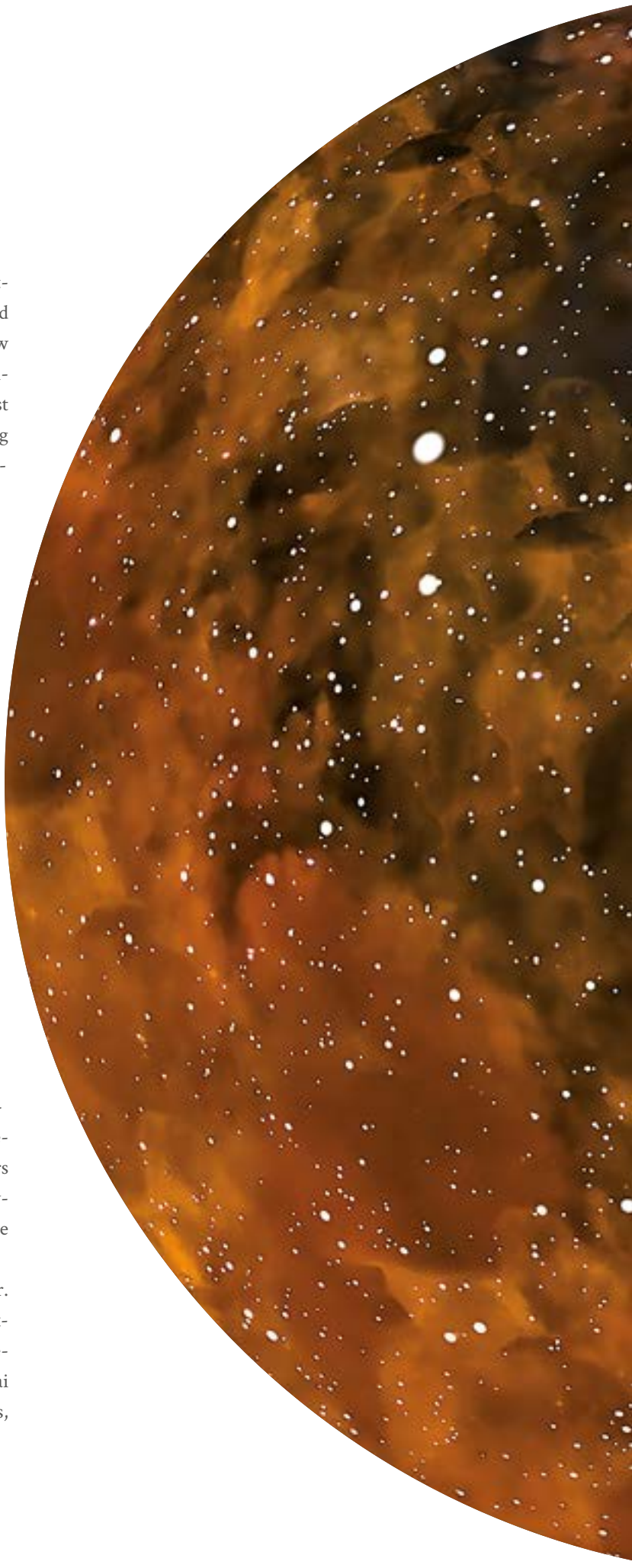
ESO Supernova

Space, stars, and the Universe hold an almost magical attraction for many people. While astronomy is a vivid and rapidly evolving science, transferring current and new knowledge from researchers to the public is often difficult, and daily work limits it to rare occasions for most scientists. The ESO Supernova visitor center in Garching near Munich is intended to bridge the gap between astronomical research and the interested public, allowing visitors to become updated on the current status of astronomy and on recent discoveries and to become and remain inspired by the fascination of the Universe in which we live.

The center is the fruit of a collaboration agreement between HITS and ESO, the foremost intergovernmental astronomy organization in Europe and the world's most productive ground-based astronomical observatory by far. The Klaus Tschira Stiftung (KTS) funds the construction of the premises, and ESO will run the facility. The ESO Supernova is due for completion in the spring of 2018 and will open its doors to the public in the course of that same year.

The Supernova project group here at HITS contributes its scientific expertise to the development of the content for the exhibition. Moreover, the group explores and develops interactive exhibits specifically for the ESO Supernova that will allow visitors to dive into a broad range of topics by means of interactive computer simulations, virtual reality, and state-of-the-art computer graphics. This will enable the visitors to discover astronomy on an individual and personal level, deepen their understanding of astronomy, and share their experience and fascination with others.

The two developers, Dr. Dorotea Dudaš and Dr. Volker Gaibler, cover a broad range of professional expertise, from computer graphics and numerical mathematics to astronomy and theoretical astrophysics. Dr. Kai Polsterer, leader of the Junior Group Astroinformatics, serves as the Project Manager. ■



Still image from the planetarium show From Earth to the Universe, showing the Eagle Nebula in 3D. (Credit: ESO/Theofanis N. Matsopoulos.)

7
Publications

Ahmadzadeh F, Flecks M, Carretero MA, Böhme W, Ihlow F, Kapli P, Miraldo A, Rödder D. Separate histories in both sides of the Mediterranean: phylogeny and niche evolution of ocellated lizards. Journal of Biogeography (2016) 43:1242–1253.

Ahnen ML, Ansoldi S, Antonelli LA, Antoranz P, Babic A, Banerjee B, Bangale P, Barres de Almeida U, Barrio JA, Becerra González J, Bednarek W, Bernardini E, Biasuzzi B, Biland A, Blanch O, Bonnefoy S, Bonnoli G, Borracci F, Bretz T, Buson S, Carmona E, Carosi A, Chatterjee A, Clavero R, Colin P, Colombo E, Contreras JL, Cortina J, Covino S, Da Vela P, Dazzi F, De Angelis A, De Lotto B, de Oña Wilhelmi E, Delgado Mendez C, Di Pierro F, Domínguez A, Dominis Prester D, Dorner D, Doro M, Einecke S, Eisenacher Glawion D, Elsaesser D, Fernández-Barral A, Fidalgo D, Fonseca MV, Font L, Frantzen K, Fruck C, Galindo D, García López RJ, Garczarczyk M, Garrido Terrats D, Gaug M, Giammaria P, Godinovic N, González Muñoz A, Gora D, Guberman D, Hadasch D, Hahn A, Hanabata Y, Hayashida M, Herrera J, Hose J, Hrupec D, Hughes G, Idec W, Kodani K, Konno Y, Kubo H, Kushida J, La Barbera A, Lelas D, Lindfors E, Lombardi S, Longo F, López M, López-Coto R, Lorenz E, Majumdar P, Makariev M, Mallot K, Maneva G, Manganaro M, Mannheim K, Maraschi L, Marcote B, Mariotti M, Martínez M, Mazin D, Menzel U, Miranda JM, Mirzoyan R, Moralejo A, Moretti E, Nakajima D, Neustroev V, Niedzwiecki A, Nievas Rosillo M, Nilsson K, Nishijima K, Noda K, Orito R, Overkemping A, Paiano S, Palacio J, Palatiello M, Paneque D, Paoletti R, Paredes JM, Paredes-Fortuny X, Pedaletti G, Persic M, Poutanen J, Prada Moroni PG, Prandini E, Puljak I, Rhode W, Ribó M, Rico J, Rodriguez Garcia J, Saito T, Satalecka K, Schultz C, Schweizer T, Sillanpää A, Sitarek J, Snidaric I, Sobczynska D, Stamerra A, Steinbring T, Strzys M, Takalo L, Takami H, Tavecchio F, Temnikov P, Terzic T, Tescaro D, Teshima M, Thaele J, Torres DF, Toyama T, Treves A, Vazquez Acosta M, Verguilov V, Vovk I, Ward JE, Will M, Wu MH, Zanin R, Pfrommer C, Pinzke A, Zandanel F. Deep observation of the NGC 1275 region with MAGIC: search of diffuse -ray emission from cosmic rays in the Perseus cluster. Atron. & Astrophys. (2016) 589:A33.

Aramburu IV, Mercadante D, Milles S, Ringkjøbing M, Banterle N, Koehler C, Tyagi S, Clarke J, Shammas SL, Blackledge M, et al. Plasticity of nucleoporin nuclear transport receptor interactions-molecular description of a highly dynamic, ultrafast interaction mechanism. Biophys. J. (2016) 110:357a.

Arnold C, Springel V, Puchwein E. Zoomed cosmological simulations of Milky Way-sized haloes in f(R) gravity. Mon. Notices Royal Astron. Soc. (2016) 462:1530–1541.

Baran S, Lerch S. Mixture EMOS models for calibrating ensemble forecasts of wind speed. Environmetrics (2016) 27:116–130.

Bauswein A, Stergioulas N, Janka HT. Exploring properties of high-density matter through remnants of neutron-star mergers. Eur. Phys. J. A (2016) 52:56.

Blecha L, Sijacki D, Kelley LZ, Torrey P, Vogelsberger M, Nelson D, Springel V, Snyder G, Hernquist L. Recoiling black holes: prospects for detection and implications of spin alignment. Mon. Notices Royal Astron. Soc. (2016) 456:961–989.

Borsari C, Luciani R, Pozzi C, Poehner I, Henrich S, Trande M, Cordeiro-da Silva A, Santarem N, Baptista C, Tait A, Di Pisa F, Dello Iacono L, Landi G, Gul S, Wolf M, Kuzikov M, Ellinger B, Reinshagen J, Witt G, Gribbon P, Kohler M, Keminer O, Behrens B, Costantino L, Tejera Nevado P, Bifeld E, Eick J, Clos J, Torrado J, Jiménez-Antón MD, Corral MJ, Alunda JM, Pellati F, Wade RC, Ferrari S, Mangani S, Costi MP. Profiling of flavonol derivatives for the development of antitrypanosomatidic drugs. J. Med. Chem. (2016) 59:7598–7616.

Bray AD, Pillepich A, Sales LV, Zhu E, Genel S, Rodriguez-Gomez V, Torrey P, Nelson D, Vogelsberger M, Springel V, Eisenstein DJ, Hernquist L. Modelling galactic conformity with the colour-halo age relation in the Illustris simulation. Mon. Notices Royal Astron. Soc. (2016) 455:185–198.

Broderick AE, Tiede P, Shalaby M, Pfrommer C, Puchwein E, Chang P, Lamberts A. Bow Ties in the Sky. I: The Angular Structure of Inverse Compton Gamma-Ray Halos in the Fermi Sky. Astrophys. J. (2016) 832:109.

Bulla M, Sim SA, Kromer M, Seitenzahl IR, Fink M, Ciaraldi-Schoolmann F, Röpke FK, Hillebrandt W, Pakmor R, Ruiter AJ, Taubenberger S. Predicting polarization signatures for double-detonation and delayed-detonation models of Type Ia supernovae. Mon. Notices Royal Astron. Soc. (2016a) 462:1039–1056.

Bulla M, Sim SA, Pakmor R, Kromer M, Taubenberger S, Röpke FK, Hillebrandt W, Seitenzahl IR. Type Ia supernovae from violent mergers of carbonoxygen white dwarfs: polarization signatures. Mon. Notices Royal Astron. Soc. (2016b) 455:1060–1070.

Chang P, Broderick AE, Pfrommer C, Puchwein E, Lamberts A, Shalaby M, Vasil G. The Linear Instability of Dilute Ultrarelativistic $e^{\pm}$ Pair Beams. Astrophys. J. (2016) 833:118.

Clark JA, Bauswein A, Stergioulas N, Shoemaker D. Observing gravitational waves from the post-merger phase of binary neutron star coalescence. Class. Quantum Gravity (2016) 33:085003.

Côté B, Ritter C, O'Shea BW, Herwig F, Pignatari M, Jones S, Fryer CL. Uncertainties in Galactic Chemical Evolution Models. ApJ (2016) 824:82.

Crawford E, Norris RP, Polsterer KL. WTF? Discovering the Unexpected in next-generation radio continuum surveys. In "Astronomical Data Analysis Software an Systems XXV (ADASS XXV)", Eds. Taylor AR, Rosolowsky E (2016) vol. 496 of Astronomical Society of the Pacific Conference Series.

Cruz-Chu ER, Xiao S, Patil S, Gkagkas K, Gräter F. Organic filling mitigates flaw-sensitivity of nanoscale aragonite. ACS Biomater. Sci. & Eng. (2016).

Cyr-Racine FY, Sigurdson K, Zavala J, Bringmann T, Vogelsberger M, Pfrommer C. ETHOS – an effective theory of structure formation: From dark particle physics to the matter distribution of the Universe. Phys. Rev. D (2016) 93:123527.

Czech L, Huerta-Cepas J, Stamatakis A. A Critical Review on the Use of Support Values in Tree Viewers and Bioinformatics Toolkits. bioRxiv, doi: https://doi.org/10.1101/035360 (2016).

Darriba D, Weiß M, Stamatakis A. Prediction of missing sequences and branch lengths in phylogenomic data. Bioinforma. (2016) 32:1331–1337.

D'Isanto A. Uncertain Photometric Redshifts with Deep Learning Methods. IAU Symp. Astroinformatics (2016a) 325.

D'Isanto A, Cavuoti S, Brescia M, Donalek C, Longo G, Riccio G, Djorgovski SG. An analysis of feature relevance in the classification of astronomical transients with machine learning methods. Mon. Notices Royal Astron. Soc. (2016b) 457:3119–3132.

Dugard CK, Mertz RA, Rayon C, Mercadante D, Hart C, Benatti MR, Olek AT, SanMiguel PJ, Cooper BR, Reiter WD, McCann MC, Carpita NC. The cell wall arabinose-deficient arabidopsis thaliana mutant murus5 encodes a defective allele of reversibly glycosylated polypeptide2. Plant Physiol. (2016) 171:1905–1920.

Ehm W. Reproducibility from the perspective of meta-analysis. In "Reproducibility: Principles, Problems, Practices, and Prospects", Eds. Atmanspacher H, Maasen S. Wiley, Hoboken (2016) pp. 141–167.

Ehm W, Gneiting T, Jordan A, Krüger F. Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings (with discussion). J. Royal Stat. Soc. Ser. B: Stat. Methodol. (2016) 78:505–562.

7
Publications

Ehm W, Wackermann J. Geometric-optical illusions and Riemannian geometry. J. Math. Psychol. (2016) 71:28–38.

Elliott G, Ghanem D, Krüger F. Forecasting conditional probabilities of binary outcomes under misspecification. Rev. Econ. Stat. (2016) 98:742–755.

Fetzer A, Metzger J, Katic D, März K, Wagner M, Philipp P, Engelhardt S, Weller T, Zelzer S, Franz AM, Schoch N, Heuveline V, Maleshkova M, Rettinger A, Speidel S, Wolf I, Kengott H, Mehrabi A, Müller-Stich BP, Maier-Hein L, Meinzer HP, Nolden M. Towards an open-source semantic data infrastructure for integrating clinical and scientific data in cognitionguided surgery, in: Proc. SPIE 9789, Medical Imaging 2016: PACS and Imaging Informatics: Next Generation and Innovations, 97890O. pp. 97890O-1--97890O-8. (March 25, 2016); doi:10.1117/12.2217163 (2016).

Fissler T, Ziegel JF, Gneiting T. Expected shortfall is jointly elicitable with value-at-risk: Implications for backtesting. Risk Mag. (2016) January: 58–61.

Fujimoto Y, Bryan GL, Tasker EJ, Habe A, Simpson CM. GMC evolution in a barred spiral galaxy with star formation and thermal feedback. Mon. Notices Royal Astron. Soc. (2016) 461:1684–1700.

Gdynia G, Sauer SW, Kopitz J, Fuchs D, Duglova K, Ruppert T, Miller M, Pahl J, Cerwenka A, Enders M, Mairbäurl H, Kaminski MM, Penzel R, Zhang C, Fuller JC, Wade RC, Benner A, Chang-Claude J, Brenner H, Hoffmeister M, Zentgraf H, Schirmacher P, Roth W. The HMGB1 protein induces a metabolic type of tumour cell death by blocking aerobic respiration. Nat. Commun. (2016) 7:10764.

Gianniotis N, Kügler SD, Tino P, Polsterer KL. Model-coupled autoencoder for time series visualisation. Neurocomputing (2016) 192:139–146.

Gianniotis N, Schnörr C, Molkenthin C, Bora SS. Approximate variational inference based on a finite sample of gaussian latent variables. Pattern Analysis Appl. (2016) 19:475–485.

Girichidis P, Naab T, Walch S, Hanasz M, Mac Low MM, Ostriker JP, Gatto A, Peters T,Wünsch R, Glover SCO, Klessen RS, Clark PC, Baczynski C. Launching Cosmic-Ray-driven Outflows from the Magnetized Interstellar Medium. Astrophys. J. Lett. (2016a) 816:L19.

Girichidis P, Walch S, Naab T, Gatto A, Wünsch R, Glover SCO, Klessen RS, Clark PC, Peters T, Derigs D, Baczynski C. The SILCC (SImulating the LifeCycle of molecular Clouds) project – II. Dynamical evolution of the supernova-driven ISM and the launching of outflows. Mon. Notices Royal Astron. Soc. (2016b) 456:3432–3455.

Gómez FA, White SDM, Marinacci F, Slater CT, Grand RJJ, Springel V, Pakmor R. A fully cosmological model of a Monoceros-like ring. Mon. Notices Royal Astron. Soc. (2016) 456:2779–2793.

Goriely S, Bauswein A, Janka HT, Panebianco S, Sida JL, Lemaître JF, Hilaire S, Dubray N. The r-process nucleosynthesis during the decompression of neutron star crust material. J. Phys. Conf. Ser. (2016) 665:012052.

Grand RJJ, Kawata D. Radial migration in numerical simulations of Milky- Way sized galaxies. Astron. Nachrichten (2016a) 337:957.

Grand RJJ, Springel V, Gómez FA, Marinacci F, Pakmor R, Campbell DJR, Jenkins A. Vertical disc heating in Milky Way-sized galaxies in a cosmological context. Mon. Notices Royal Astron. Soc. (2016b) 459:199–219.

Grand RJJ, Springel V, Kawata D, Minchev I, Sánchez-Blázquez P, Gómez FA, Marinacci F, Pakmor R, Campbell DJR. Spiral-induced velocity and metallicity patterns in a cosmological zoom simulation of a Milky Way-sized galaxy. Mon. Notices Royal Astron. Soc. (2016c) 460:L94–L98.

Gundic M, Tomic A, Wade RC, Matovina M, Karacic Z, Kazazic S, Tomic S. Human DPP III – keap1 interactions: A combined experimental and computational study. Croat. Chem. Acta. (2016) 89:217–228.

Haider M, Steinhauser D, Vogelsberger M, Genel S, Springel V, Torrey P, Hernquist L. Large-scale mass distribution in the Illustris simulation. Mon. Notices Royal Astron. Soc. (2016) 457:3024–3035.

Hammer N, Jamitzky F, Satzger H, Allalen M, Block A, Karmakar A, Brehm M, Bader R, Iapichino L, Ragagnin A, Karakasis V, Kranzlmüller D, Bode A, Huber H, Kühn M, Machado R, Grünewald D, Edelmann PVF, Röpke FK, Wittmann M, Zeiser T, Wellein G, Mathias G, Schwörer M, Lorenzen K, Federrath C, Klessen R, Bamberg K, Ruhl H, Schornbaum F, Bauer M, Nikhil A, Qi J, Klimach H, Stüben H, Deshmukh A, Falkenstein T, Dolag K, Petkova M. Extreme scale-out supermuc phase 2 – lessons learned. In "Parallel Computing: On the Road to Exascale, Proceedings of the International Conference on Parallel Computing, ParCo 2015, 1–4 September 2015, Edinburgh, Scotland, UK", Eds. Joubert GR, Leather H, Parsons M, Peters FJ, Sawyer M. IOS Press (2016) vol. 27 of Advances in Parallel Computing, pp. 827–836.

Heinzerling B, Judea A, Strube M. HITS at TAC KBP 2015: Entity discovery and linking, and event nugget detection. In "Proceedings of the Text Analysis Conference, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 16 – 17 November 2015" (2016).

Hemri S, Haiden T, Pappenberger F. Discrete post-processing of total cloud cover ensemble forecasts. Mon. Weather. Rev. (2016) 144:2565–2577.

Henkel R, Hoehndorf R, Kacprowski T, Knüpfer C, Liebermeister W, Waltemath D. Notions of similarity for systems biology models. Briefings Bioinforma. (2016) pii: bbw090.

Hoang DT, Vinh LS, Flouri T, Stamatakis A, von Haeseler A, Minh BQ. A new phylogenetic tree sampling method for maximum parsimony bootstrapping and proof-of-concept implementation. In "Knowledge and Systems Engineering (KSE), 2016 Eighth International Conference on". IEEE (2016) pp. 1 – 6.

Hocq L, Sénéchal F, Lefebvre V, lehner a, Domon JM, Mollet JC, Dehors J, Pageau K, Marcelo P, Guerineau F, Kolsek K, Mercadante D, Pelloux J. Combined experimental and computational approaches reveal distinct phdependence of pectin methyl esterase inhibitors. Plant Physiol. (2016).

Hoff M, Orf S, Riehm B, Darriba D, Stamatakis A. Does the choice of nucleotide substitution models matter topologically? BMC Bioinforma. (2016) 17:143.

Jones S, Röpke FK, Pakmor R, Seitenzahl IR, Ohlmann ST, Edelmann PVF. Do electron-capture supernovae make neutron stars? First multidimensional hydrodynamic simulations of the oxygen deflagration. Astron. Astrophys. (2016a) 593:A72.

Jones S, Ritter C, Herwig F, Fryer C, Pignatari M, Bertolli MG, Paxton B. H ingestion into he-burning convection zones in super-agb stellar models as a potential site for intermediate neutrondensity nucleosynthesis. Mon. Notices Royal Astron. Soc. (2016b) 455:3848–3863.

Joubert GR, Leather H, Parsons M, Peters FJ, Sawyer M (Eds.). "Parallel Computing: On the Road to Exascale, Proceedings of the International Conference on Parallel Computing, ParCo 2015, 1 – 4 September 2015, Edinburgh, Scotland, UK", vol. 27 of Advances in Parallel Computing. IOS Press (2016).

Judea A, Strube M. Incremental global event extraction. In "Proceedings of the 26th International Conference on Computational Linguistics, Osaka, Japan, 11 – 16 December 2016" (2016) pp. 2279–2289.

Just O, Obergaulinger M, Janka HT, Bauswein A, Schwarz N. Neutron-star Merger Ejecta as Obstacles to Neutrino-powered Jets of Gamma-Ray Bursts. ApJ (2016) 816:L30.

Kannan R, Springel V, Pakmor R, Marinacci F, Vogelsberger M. Accurately simulating anisotropic thermal conduction on a moving mesh. Mon. Notices Royal Astron. Soc. (2016) 458:410–424.

Kannan R, Vogelsberger M, Stinson GS, Hennawi JF, Marinacci F, Springel V, Macciò AV. Galaxy formation with local photoionization feedback – II. Effect of X-ray emission from binaries and hot gas. Mon. Notices Royal Astron. Soc. (2016) 458:2516–2529.

7
Publications

Kent LM, Loo TS, Melton LD, Mercadante D, Williams MA, Jameson GB. Structure and properties of a non-processive, salt-requiring, and acidophilic pectin methylesterase from aspergillus niger provide insights into the key determinants of processivity control. J. Biol. Chem. (2016) 291:1289–1306.

Khan KM, Mesaik MA, Abdalla OM, Rahim F, Soomro S, Halim SA, Mustafa G, Ambreen N, Khalid AS, Taha M, Perveen S, Alam MT, Hameed A, Ul-Haq Z, Ullah H, Rehman ZU, Siddiqui RA, Voelter W. The immunomodulation potential of the synthetic derivatives of benzothiazoles: Implications in immune system disorders through in vitro and in silico studies. Bioorg. Chem. (2016) 64:21–28.

Kim Jh, Agertz O, Teyssier R, Butler MJ, Ceverino D, Choi JH, Feldmann R, Keller BW, Lupi A, Quinn T, Revaz Y, Wallace S, Gnedin NY, Leitner SN, Shen S, Smith BD, Thompson R, Turk MJ, Abel T, Arraki KS, Benincasa SM, Chakrabarti S, DeGraf C, Dekel A, Goldbaum NJ, Hopkins PF, Hummels CB, Klypin A, Li H, Madau P, Mandelker N, Mayer L, Nagamine K, Nickerson S, O'Shea BW, Primack JR, Roca-Fàbrega S, Semenov V, Shimizu I, Simpson CM, Todoroki K,Wadsley JW, Wise JH, AGORA Collaboration. The AGORA High-resolution Galaxy Simulations Comparison Project. II. Isolated Disk Test. Astrophys. J. (2016) 833:202.

Kokh DB, Czodrowski P, Rippmann F, Wade RC. Perturbation approaches for exploring protein binding site flexibility to predict transient binding pockets. J. Chem. Theory Comput. (2016) 12:4100–4113.

Konstandin L, Schmidt W, Girichidis P, Peters T, Shetty R, Klessen RS. Mach number study of supersonic turbulence: the properties of the density field. Mon. Notices Royal Astron. Soc. (2016) 460:4483–4491.

Kornilios P, Thanou E, Kapli P, Parmakelis A, Chatzaki M. Peeking through the trapdoor: Historical biogeography of the Aegean endemic spider Cyrtocarenum Ausserer, 1871 with an estimation of mtDNA substitution rates for Mygalomorphae. Mol. (2016) 98:300–313.

Kozlov AM, Zhang J, Yilmaz P, Glöckner FO, Stamatakis A. Phylogeny-aware identification and correction of taxonomically mislabeled sequences. Nucleic Acids Res. (2016) 44:5022–5033.

Kratzke J, Rengier F, Weis C, Beller CJ, Heuveline V. In vitro flow assessment: from pc-mri to computational fluid dynamics including fluid-structure interaction, in: Proc. SPIE 9783, Medical Imaging 2016: Physics of Medical Imaging,97835C (April 1, 2016); doi:10.1117/12.2217336 (2016).

Kromer M, Fremling C, Pakmor R, Taubenberger S, Amanullah R, Cenko SB, Fransson C, Goobar A, Leloudas G, Taddia F, Röpke FK, Seitenzahl IR, Sim SA, Sollerman J. The peculiar Type Ia supernova iPTF14atg: Chandrasekharmass explosion or violent merger? MNRAS (2016) 459:4428–4439.

Kromer M, Fremling C, Pakmor R, Taubenberger S, Amanullah R, Cenko SB, Fransson C, Goobar A, Leloudas G, Taddia F, Röpke FK, Seitenzahl IR, Sim SA, Sollerman J. The peculiar Type Ia supernova iPTF14atg: Chandrasekharmass explosion or violent merger? Mon. Notices Royal Astron. Soc. (2016) 459:4428–4439.

Krüger F, Nolte I. Disagreement versus uncertainty: Evidence from distribution forecasts. J. Bank. & Finance (2016) 72:S172–S186.

Kügler SD, Gianniotis N, Polsterer KL. An explorative approach for inspecting kepler data. Mon. Notices Royal Astron. Soc. (2016a) 455:4399–4405.

Kügler SD, Gianniotis N, Polsterer KL. A spectral model for multimodal redshift estimation. In "IEEE Symposium on Computational Intelligence and Data Mining (SSCI)". IEEE (2016b).

Latour M, Heber U, Irrgang A, Schaffenroth V, Geier S, Hillebrandt W, Röpke FK, Taubenberger S, Kromer M, Fink M. Quantitative spectral analysis of the sdB star HD 188112: A helium-core white dwarf progenitor. A&A (2016) 585:A115.

Lensink MF, Velankar S, Kryshtafovych A, Huang SY, Schneidman-Duhovny D, Sali A, Segura J, Fernandez-Fuentes N, Viswanath S, Elber R, Grudinin S, Popov P, Neveu E, Lee H, Baek M, Park S, Heo L, Rie Lee G, Seok C, Qin S, Zhou HX, Ritchie DW, Maigret B, Devignes MD, Ghoorah A, Torchala M, Chaleil RA, Bates PA, Ben-Zeev E, Eisenstein M, Negi SS, Weng Z, Vreven T, Pierce BG, Borrman TM, Yu J, Ochsenbein F, Guerois R, Vangone A, Rodrigues JP, van Zundert G, Nellen M, Xue L, Karaca E, Melquiond AS, Visscher K, Kastritis PL, Bonvin AM, Xu X, Qiu L, Yan C, Li J, Ma Z, Cheng J, Zou X, Shen Y, Peterson LX, Kim HR, Roy A, Han X, Esquivel-Rodriguez J, Kihara D, Yu X, Bruce NJ, Fuller JC, Wade RC, Anishchenko I, Kundrotas PJ, Vakser IA, Imai K, Yamada K, Oda T, Nakamura T, Tomii K, Pallara C, Romero-Durana M, Jiménez-García B, Moal IH, Férnandez-Recio J, Joung JY, Kim JY, Joo K, Lee J, Kozakov D, Vajda S, Mottarella S, Hall DR, Beglov D, Mamonov A, Xia B, Bohnuud T, Del Carpio CA, Ichiishi E, Marze N, Kuroda D, Roy Burman SS, Gray JJ, Chermak E, Cavallo L, Oliva R, Tovchigrechko A, Wodak SJ. Prediction of homo- and hetero-protein complexes by protein docking and template-based modeling: a CASP-CAPRI experiment. Proteins (2016) 84:323–348.

Li H, Li R, Mao S, Xu D, Long RJ, Emsellem E. Assessing the Jeans Anisotropic Multi-Gaussian Expansion method with the Illustris simulation. Mon. Notices Royal Astron. Soc. (2016) 455:3680–3692.

Lippok S, Kolšek K, Löf A, Eggert D, Vanderlinden W, Müller JP, König G, Obser T, Röhrs K, Schneppenheim S, Budde U, Baldauf C, Aponte-Santamaría C, Gräter F, Schneppenheim R, Rädler JO, Brehm MA. von willebrand factor is dimerized by protein disulfide isomerase. Blood (2016) 127:1183–1191.

Loesel P, Heuveline V. Enhancing a diffusion algorithm for 4d image segmentation using local information (2016).

Magee MR, Kotak R, Sim SA, Kromer M, Rabinowitz D, Smartt SJ, Baltay C, Campbell HC, Chen TW, Fink M, Gal-Yam A, Galbany L, Hillebrandt W, Inserra C, Kankare E, Le Guillou L, Lyman JD, Maguire K, Pakmor R, Röpke FK, Ruiter AJ, Seitenzahl IR, Sullivan M, Valenti S, Young DR. The type Iax supernova, SN 2015H. A white dwarf deflagration candidate. A&A (2016a) 589:A89.

Magee MR, Kotak R, Sim SA, Kromer M, Rabinowitz D, Smartt SJ, Baltay C, Campbell HC, Chen TW, Fink M, Gal-Yam A, Galbany L, Hillebrandt W, Inserra C, Kankare E, Le Guillou L, Lyman JD, Maguire K, Pakmor R, Röpke FK, Ruiter AJ, Seitenzahl IR, Sullivan M, Valenti S, Young DR. The type Iax supernova, SN 2015H. A white dwarf deflagration candidate. Astron. Astrophys. (2016b) 589:A89.

Mahe F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, Singer D, Mayor J, Bunge J, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell E, Seppey C, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M. Soil Protists in Three Neotropical Rainforests are Hyperdiverse and Dominated by Parasites. bioRxiv 050997, Nature Ecology & Evolution (2016).

Mendoza-Temis JJ, Wu MR, Martínez-Pinedo G, Langanke K, Bauswein A, Janka HT, Frank A. On the robustness of the r-process in neutron-star mergers against variations of nuclear masses. J. Phys. Conf. Ser. (2016) 730:012018.

Mercadante D, Milles S, Fuertes G, Svergun D, Lemke EA, Gräter F. Rescuing the over-collapse of intrinsically disordered proteins using a force field derived by a new paradigm. Biophys. J. (2016) 110:556a.

Mesgar M, Strube M. Lexical coherence graph modeling using word embeddings. In "Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, Cal., 12 – 17 June 2016" (2016) pp. 1414 – 1423.

Mocz P, Pakmor R, Springel V, Vogelsberger M, Marinacci F, Hernquist L. A moving mesh unstaggered constrained transport scheme for magnetohydrodynamics. Mon. Notices Royal Astron. Soc. (2016) 463:477–488.

Monachesi A, Gómez FA, Grand RJJ, Kauffmann G, Marinacci F, Pakmor R, Springel V, Frenk CS. On the stellar halo metallicity profile of Milky Way-like galaxies in the Auriga simulations. Mon. Notices Royal Astron. Soc. (2016) 459:L46–L50.

7
Publications

Monari G, Famaey B, Siebert A, Grand RJJ, Kawata D, Boily C. The effects of bar-spiral coupling on stellar kinematics in the Galaxy. Mon. Notices Royal Astron. Soc. (2016) 461:3835–3846.

Moosavi NS, Strube M. Search space pruning: A simple solution for better coreference resolvers. In "Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, Cal., 12 – 17 June 2016" (2016a) pp. 1005 – 1011.

Moosavi NS, Strube M. Which coreference evaluation metric do you trust? A proposal for a link-based entity aware metric. In "Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7 – 12 August 2016" (2016b) pp. 632 – 642.

Nelson D, Genel S, Pillepich A, Vogelsberger M, Springel V, Hernquist L. Zooming in on accretion – I. The structure of halo gas. Mon. Notices Royal Astron. Soc. (2016) 460:2881–2904.

Nickerson D, Atalag K, de B, Geiger J, Goble C, Hollmann S, Lonien J, Müller W, Regierer B, Stanford N, Golebiewski M, Hunter P. The human physiome: how standards, software, and innovative service infrastructures are providing the building blocks to make it achievable. Interface Focus. (2016) 6: 20150103.

Öztürk MA, Pachov GV, Wade RC, Cojocaru V. Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. Nucleic Acids Res. (2016) 44:6599–6613.

Ohlmann ST, Röpke FK, Pakmor R, Springel V. Hydrodynamic Movingmesh Simulations of the Common Envelope Phase in Binary Stellar Systems. Astrophys. J. (2016a) 816:L9.

Ohlmann ST, Röpke FK, Pakmor R, Springel V, Müller E. Magnetic field amplification during the common envelope phase. Mon. Notices Royal Astron. Soc. (2016b) 462:L121–L125.

Ozboyaci M, Kokh DB, Wade RC. Three steps to gold: Mechanism of protein adsorption to a gold surface revealed by Brownian and molecular dynamics simulations. Phys. Chem. Chem. Phys. (2016a) 18:10191–10200.

Ozboyaci M, Kokh DB, Corni S, Wade RC. Modeling and simulation of protein–surface interactions: achievements and challenges. Q. Rev. Biophys. (2016b) 49.

Ozboyaci M, Martinez M, Wade RC. An efficient low storage and memory treatment of gridded interaction fields for simulations of macromolecular association. J. Chem. Theory Comput. (2016c) 12:4563–4577.

Pakmor R, Pfrommer C, Simpson CM, Kannan R, Springel V. Semi-implicit anisotropic cosmic ray transport on an unstructured moving mesh. Mon. Notices Royal Astron. Soc. (2016a) 462:2603–2616.

Pakmor R, Pfrommer C, Simpson CM, Springel V. Galactic Winds Driven by Isotropic and Anisotropic Cosmic-Ray Diffusion in Disk Galaxies. Astrophys. J. Lett. (2016b) 824:L30.

Pakmor R, Springel V, Bauer A, Mocz P, Munoz DJ, Ohlmann ST, Schaal K, Zhu C. Improving the convergence properties of the moving-mesh code AREPO. Mon. Notices Royal Astron. Soc. (2016c) 455:1134–1143.

Panov IV, Korneev IY, Blinnikov SI, Röpke F. Neutron excess number and nucleosynthesis of heavy elements in a type Ia supernova explosion. Sov. J. Exp. Theor. Phys. Lett. (2016) 103:431–434.

Parveen D, Mesgar M, Strube M. Generating coherent summaries of scientific articles using coherence patterns. In "Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Tex., 1 – 5 November 2016" (2016) pp. 772 – 783.

Pignatari M, Herwig F, Hirschi R, Bennett M, Rockefeller G, Fryer C, Timmes FX, Ritter C, Heger A, Jones S, Battino U, Dotter A, Trappitsch R, Diehl S, Frischknecht U, Hungerford A, Magkotsios G, Travaglio C, Young P. NuGrid Stellar Data Set. I.Stellar Yields from H to Bi for Stars with Metallicities Z = 0.02 and Z = 0.01. ApJS (2016) 225:24.

Polsterer KL. Dealing with Uncertain Multimodal Photometric Redshift Estimations. IAU Symp. Astroinformatics (2016a) 325.

Polsterer KL, Gieseke F. Probability Density Functions for Astronomy. In "Astronomical Data Analysis Software an Systems XXV (ADASS XXV)", Eds. Taylor AR, Rosolowsky E (2016b) vol. 496 of Astronomical Society of the Pacific Conference Series.

Polsterer KL, Gieseke F, Igel C, Doser B, Gianniotis N. Parallelized rotation and flipping invariant kohonen maps (pink) on gpus. In "European Symposium on Artificial Neural Networks (ESANN)" (2016c) pp. 405 – 410.

Rafiullah R, Aslamkhan M, Paramasivam N, Thiel C, Mustafa G, Wiemann S, Schlesner M, Wade RC, Rappold GA, Berkel S. Homozygous missense mutation in the LMAN2L gene segregates with intellectual disability in a large consanguineous Pakistani family. J. Med. Genet. (2016) 53:128–144.

Remse M, Mesgar M, Strube M. Feature-rich error detection in scientific writing using logistic regression. In "Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications, San Diego, Cal., 16 June 2016" (2016) pp. 162 – 171.

Renner SS, Grimm GW, Kapli P, Denk T. Species relationships and divergence times in beeches: new insights from the inclusion of 53 young and old fossils in a birth–death clock model. Phil. Trans. R. Soc. B (2016) 371:20150135.

Resch B, Summa A, Zeile P, Strube M. Citizen-centric urban planning through extracting emotion information from Twitter in an interdisciplinary spacetime- linguistics algorithm. Urban Plan. (2016) 1:114–127.

Rodriguez-Gomez V, Pillepich A, Sales LV, Genel S, Vogelsberger M, Zhu Q, Wellons S, Nelson D, Torrey P, Springel V, Ma CP, Hernquist L. The stellar mass assembly of galaxies in the Illustris simulation: growth by mergers and the spatial distribution of accreted stars. Mon. Notices Royal Astron. Soc. (2016) 458:2371–2390.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ (2016) 4:e2584.

Ruiz-Lapuente P, The LS, Hartmann DH, Ajello M, Canal R, Röpke FK, Ohlmann ST, Hillebrandt W. The Origin of the Cosmic Gamma-ray Background in the MeV Range. ApJ (2016) 820:142.

Sadiq SK. Reaction–diffusion basis of retroviral infectivity. Philos. Transactions Royal Soc. Lond. A: Math. Phys. Eng. Sci. (2016) 374:2080.

Sánchez-Menguiano L, Sánchez SF, Kawata D, Chemin L, Pérez I, Ruiz-Lara T, Sánchez-Blázquez P, Galbany L, Anderson JP, Grand RJJ, Minchev I, Gómez FA. Evidence of Ongoing Radial Migration in NGC 6754: Azimuthal Variations of the Gas Properties. Astrophys. J. (2016) 830:L40.

Schaal K, Springel V, Pakmor R, Pfrommer C, Nelson D, Vogelsberger M, Genel S, Pillepich A, Sijacki D, Hernquist L. Shock finding on a moving-mesh – II. Hydrodynamic shocks in the Illustris universe. Mon. Notices Royal Astron. Soc. (2016) 461:4441–4465.

Schefzik R. A similarity-based implementation of the Schaake shuffle. Mon. Weather Rev. (2016a) 144:1909–1921.

Schefzik R. Combining parametric low-dimensional ensemble postprocessing with reordering methods. Q. J. Royal Meteorol. Soc. (2016b) 142:2463–2477.

Schibich D, Gloge F, Pöhner I, Björkholm P, Wade RC, von Heijne G, Bukau B, Kramer G. Global profiling of SRP interaction with nascent polypeptides. Nature (2016) 536:219–223.

Schneider A, Teyssier R, Potter D, Stadel J, Onions J, Reed DS, Smith RE, Springel V, Pearce FR, Scoccimarro R. Matter power spectrum and the challenge of percent accuracy. J. Cosmol. Astropart. Phys. (2016) 4:047.

Schoch N, Engelhardt S, De Simone R, Wolf I, Heuveline V. High performance computing for cognition-guided cardiac surgery: Soft tissue simulation for mitral valve reconstruction in knowledge-based surgery assistance. Proc. 6th Int. Conf. on High-Performance Sci. Comput. 2015 (LNCS Springer) (2016) p. [in press].

Schoch N, Kißler F, Stoll M, Engelhardt S, de Simone R, Wolf I, Bendl R, Heuveline V. Comprehensive patient-specific information preprocessing for cardiac surgery simulations. Int. J. Comput. Assist. Radiol. Surg. (2016) 11:1051–1059.

Schoch N, Philipp P, Weller T, Engelhardt S, Volovyk M, Fetzer A, Nolden M, De Simone R, Wolf I, Maleshkova M, Rettinger A, Studer R, Heuveline V. Cognitive tools pipeline for assistance of mitral valve surgery, In: Proc. SPIE 9786, Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling, 978603. pp. 978603-1--978603-8.(March 18, 2016); doi:10.1117/12.2216059 (2016).

Schoch N, Speidel S, Sure-Vetter Y, Heuveline V. Towards semantic simulation for patient-specific surgery assistance. Proc. First Int. Work. on Surg. Data Sci. (2016) 1:58–63.

Scholl C, Kobert K, Flouri T, Stamatakis A. The divisible load balance problem with shared cost and its application to phylogenetic inference. In "Parallel and Distributed Processing Symposium Workshops, 2016 IEEE International". IEEE (2016) pp. 408 – 417.

Schreiber F, Bader G, Gleeson P, Golebiewski M, Hucka M, Le N, Myers C, Nickerson D, Sommer B, Waltemath D. Specifications of standards in systems and synthetic biology: Status and developments in 2016. J. Integr. Bioinforma. (2016) 13:289.

Seitenzahl IR, Kromer M, Ohlmann ST, Ciaraldi-Schoolmann F, Marquardt K, Fink M, Hillebrandt W, Pakmor R, Röpke FK, Ruiter AJ, Sim SA, Taubenberger S. Three-dimensional simulations of gravitationally confined detonations compared to observations of SN 1991T. Astron. Astrophys. (2016) 592:A57.

Simpson CM, Pakmor R, Marinacci F, Pfrommer C, Springel V, Glover SCO, Clark PC, Smith RJ. The Role of Cosmic-Ray Pressure in Accelerating Galactic Outflows. Astrophys. J. Lett. (2016) 827:L29.

Skourtanioti E, Kapli P, Ilgaz c, Kumlutas Y, Avcı A, Ahmadzadeh F, Crnobrnja-Isailovic J, Gherghel I, Lymberakis P, Poulakakis N. A reinvestigation of phylogeny and divergence times of the ablepharus kitaibelii species complex (sauria, scincidae) based on mtdna and nudna genes. Mol. (2016) 103:199–214.

Sparre M, Springel V. Zooming in on major mergers: dense, starbursting gas in cosmological simulations. Mon. Notices Royal Astron. Soc. (2016) 462:2418–2430.

Stank A, Kokh DB, Fuller JC, Wade RC. Protein binding pocket dynamics. Acc. Chem. Res. (2016a) 49:809–815.

Stank A, Richter S, Wade RC. ProSAT+: visualizing sequence annotations on 3D structure. Protein Eng. Des. Sel. (2016b) 29:281–284.

Steinhauser D, Schindler S, Springel V. Simulations of ram-pressure stripping in galaxy-cluster interactions. Astron. Astrophys. (2016) 591:A51.

Subramanian SR, Singam ERA, Berinski M, Subramanian V, Wade RC. Identification of an electrostatic ruler motif for sequence-specific binding of collagenase to collagen. J. Phys. Chem. B (2016) 120:8580–8589.

Summa A, Resch B, Strube M. Microblog emotion classification by computing similarity in text, time, and space. In "Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, Osaka, Japan, 12 December 2016" (2016) pp. 153 – 162.

Terada Y, Maeda K, Fukazawa Y, Bamba A, Ueda Y, Katsuda S, Enoto T, Takahashi T, Tamagawa T, Röpke FK, Summa A, Diehl R. Measurements of the Soft Gamma-Ray Emission from SN2014J with Suzaku. ApJ (2016) 823:43.

Tong R, Wade RC, Bruce NJ. Comparative electrostatic analysis of adenylyl cyclase for isoform dependent regulation properties. Proteins (2016) 84:1844–1858.

Touré V, Mazein A,Waltemath D, Balaur I, Saqi M, Henkel R, Pellet J, Auffray C. Ston: exploring biological pathways using the sbgn standard and graph databases. BMC Bioinforma. (2016) 17:494.

Tursch A, Mercadante D, Tennigkeit J, Gräter F, Özbek S. Minicollagen cysteine-rich domains encode distinct modes of polymerization to form stable nematocyst capsules. Sci. reports (2016) 6.

Vogelsberger M, Zavala J, Cyr-Racine FY, Pfrommer C, Bringmann T, Sigurdson K. ETHOS – an effective theory of structure formation: dark matter physics as a possible explanation of the small-scale CDM problems. Mon. Notices Royal Astron. Soc. (2016) 460:1399–1416.

Wild K, Bange G, Motiejunas D, Kribelbauer J, Hendricks A, Segnitz B, Wade RC, Sinning I. Structural basis for conserved regulation and adaptation of the signal recognition particle targeting complex. J. Mol. Biol. (2016) 428:2880–2897.

Wlotzka M, Heuveline V, Klatt S, Kraus D, Haas E, Kiese K, Butterbach- Bahl K, Kraft P, Breuer L. Parallel multiphysics simulations using openpalm with application to hydro-biogeochemistry coupling, in: Hans Georg Bock, Hoang Xuan Phu, Rolf Rannacher, Johannes Schloeder (Eds.): Modeling, Simulation and Optimization of Complex Processes - HPSC 2015. Proceedings of the Sixth International Conference on High Performance Scientific Computing, March 16-20, 2015, Hanoi, Vietnam. Springer, in press (2016).

Wolstencroft K, Krebs O, Snoep J, Stanford N, Bacall F, Golebiewski M, Kuzyakiv R, Nguyen Q, Owen S, Soiland-Reyes S, Straszewski J, van N, Williams A, Malmström L, Rinn B, Müller W, Goble C. Fairdomhub: a repository and collaboration environment for sharing systems biology research. Nucleic Acids Res. (2016) 45(D1):D404–D407.

Xu D, Sluse D, Schneider P, Springel V, Vogelsberger M, Nelson D, Hernquist L. Lens galaxies in the Illustris simulation: power-law models and the bias of the Hubble constant from time delays. Mon. Notices Royal Astron. Soc. (2016) 456:739–755.

Yu X, Kokh DB, Nandekar P, Mustafa G, Richter S, Wade RC. Dynathor: Dynamics of the complex of cytochrome p450 and cytochrome p450 reductase in a phospholipid bilayer. In "High Performance Computing in Science and Engineering'15: Transactions of the High Performance Computing Center, Stuttgart (HLRS) 2015". Springer International Publishing, Cham, Switzerland (2016a) pp. 255 – 264.

Yu X, Nandekar P, Mustafa G, Cojocaru V, Lepesheva GI, Wade RC. Ligand tunnels in T. brucei and human CYP51: Insights for parasite-specific drug design. Biochim. Biophys. Acta (BBA) – Gen. Subj. (2016b) 1860:67–78.

Zakhartsev M, Medvedeva I, Orlov Y, Akberdin I, Krebs O, Schulze W. Metabolic model of central carbon and energy metabolisms of growing arabidopsis thaliana in relation to sucrose translocation. BMC Plant Biol. (2016) 16:262.

Zhou X, Frandsen PB, Holzenthal RW, Beet CR, Bennett KR, Blahnik RJ, Bonada N, Cartwright D, Chuluunbat S, Cocks GV, et al. The Trichoptera barcode initiative: a strategy for generating a species-level Tree of Life. Phil. Trans. R. Soc. B (2016) 371:20160025.

Zhu Q, Marinacci F, Maji M, Li Y, Springel V, Hernquist L. Baryonic impact on the dark matter distribution in Milky-Way-sized galaxies and their satellites. Mon. Notices Royal Astron. Soc. (2016) 458:1559–1580.

Degrees

Barbera, Pierre:
"Efficient and Massively Parallel Implementation of the Evolutionary Placement Algorithm", Master thesis, Computer Science, Karlsruhe Institute of Technology, and HITS: Alexandros Stamatakis, 2016.

Bolaños-Rosales, Alejandro:
"Low Mach Number Simulations of Convective Boundary Mixing in Classical Novae", Ph.D. thesis, Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, and HITS: Friedrich Röpke, 2016.

Cejka, Fabian:
"Jørgensen Lemma", Bachelor thesis, Mathematics, Heidelberg University: Anna Wienhard, 2016.

Claus, Patrick:
"Pronomenresolution über Abhängigkeitspfade und Semantic Web Queries innerhalb eines Koreferenzresolutionssystems", Master thesis, Neuphilologische Fakultät, Heidelberg University, and HITS: Michael Strube, 2016.

Darriba, Diego:
"Selection of models of genomic evolution in HPC environments", Ph.D. thesis, Computer Science, University of A Coruna: Guillermo López Taboada, and HITS: Alexandros Stamatakis, 2016.

Dilek, Yalim:
"On Complex Schottky groups in higher Dimensions", Master thesis, Mathematics, Heidelberg University: Anna Wienhard, 2016.

Gräter, Maximiliane:
"Simulation Study of Dual Ensemble Copula Coupling", Master thesis, Faculty of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting, 2016.

Hemri, Stephan:
"Probabilistic Forecasting Based on Hydrometeorological Ensembles", Ph.D. thesis, Faculty of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting, 2016.

Hoecker, Maximilian:
"Clustering von großen hochdimensionalen und unsicheren Datensätzen in der Astronomie", Ph.D. thesis, Computer Science, Physics, Heidelberg University, and HITS: Vincent Heuveline, 2016.

Horst, Leonhard:
"On the path to multidimensional simulations of shear instabilities in stellar evolution: mapping 1D initial data into 3D space", Master thesis, Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, and HITS: Friedrich Röpke, 2016.

Hou, Yufang:
"Unrestricted Bridging Resolution", Ph.D. thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2016.

Jordan, Alexander:
"Facets of Forecast Evaluation", Ph.D. thesis, Faculty of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting, 2016.

Kobert, Kassian:
"Mathematical Problems in Molecular Evolution and Next Generation Sequencing", Ph.D. thesis, Computer Science, Karlsruhe Institute of Technology, and HITS: Alexandros Stamatakis, 2016.

Kissler, Fabian:
"A family of representations for the modular group", Master thesis, Mathematics, Heidelberg University: Anna Wienhard, 2016.

Lach, Florian:
"Two- and three-dimensional simulations of classical novae", Master thesis, Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, and HITS: Friedrich Röpke, 2016.

Lambusch, Fabienne:
"Towards a biological modelling tool recommending proper subnetworks", Master thesis, Informatik, Universität Rostock, and HITS: Ron Henkel.

Lerch, Sebastian:
"Probabilistic Forecasting and Comparative Model Assessment, With Focus on Extreme Events", Ph.D. thesis, Faculty of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting, 2016.

Li, Liang:
"Assessing Point Forecasts – Economic and Statistical Measures", Masters thesis, Faculty of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting, 2016.

Ohlmann, Sebastian T.:
"Hydrodynamics of the Common Envelope Phase in Binary Stellar Evolution", Dissertation, Heidelberg University, and HITS: Friedrich Röpke, 2016.

Orth, Joshua:
"Simulations of Type Ia Supernovae in Pre-expanded White Dwarf Stars", Master thesis, Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, and HITS: Friedrich Röpke, 2016.

Ozboyaci, Musa:
"Modeling and Simulation of Protein-Surface Interactions", Ph.D. thesis, Combined Faculty for the Natural Sciences and Mathematics, Heidelberg University: Peter Bastian, and HITS: Rebecca C. Wade, 2016.

Psonis, Nikolaos:
"Molecular Phylogeny and Phylogeography of the Podarcis tauricus (Sauria: Lacertidae) Species Subgroup", Ph.D. thesis, Department of Biology, University of Crete: Nikolaos Poulakakis, and HITS: Alexandros Stamatakis, 2016.

Remse, Madeline Christin:
"Automated Evaluation of Scientific Writing, Sentence-level Error Detection Using Machine Learning", Master thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2016.

Rogozinnikov, Evgenii:
"Spektrale Netze und Fock-Goncharov Koordinaten", Master thesis, Mathematics, Heidelberg University: Anna Wienhard, 2016.

Schaal, Kevin:
"Shocks in the Illustris Universe and Novel Numerical Methods", Ph.D. thesis, Physics, Department of Physics and Astronomy, Heidelberg University and HITS: Volker Springel, 2016.

Schmitt, Stephan:
"Harmonic maps and Teichmüller spaces", Bachelor thesis, Mathematics, Heidelberg University: Anna Wienhard, 2016.

Stubbusch, Astrid:
"Case Study of Human Growth Hormone Interactions by Brownian Dynamics Simulation", Bachelor thesis, Molecular Biotechnology, Faculty of Biosciences, Ruprecht-Karls University Heidelberg, and HITS: Rebecca C. Wade and Neil Bruce, 2016.

Lectures, Courses and Seminars

Andreas Bauswein:
"Gravitational waves", lecture course, Heidelberg University, Heidelberg, summer semester 2016.

Steffen Brinkmann:
"Programmieren I", Institut für Computerlinguistik, Universität Heidelberg, October 2016 – February 2017.

Neil Bruce:
Practical course on Brownian dynamics simulations, School on Molecular Modeling for Life Sciences, Pula, Sardinia, Italy, June 6 – 10, 2016.

Csaba Daday:
Course on *"Bioinformatics"* for Bachelor students in Biology, Heidelberg University, Germany, January, 2016. Course on *"Fundamentals of Simulation Methods"* for Master's students in Physics, Heidelberg University, Germany, October 2015 – January 2016.

Simon Gawlok:
Lecture on *"Mathematik und Informatik B für molekulare Biotechnologie"*, Heidelberg University, winter term 2015 / 2016.

8 Teaching

Philipp Girichidis:

"Practical course on hydrodynamical simulations", Lecture at summer school *"The ISM-SPP Olympian School of Astrophysics 2016"*, Paralia Katerini, Mount Olympus, Greece. October 3 – 7, 2016.

Tilmann Gneiting:

Lecture course on *"Time series analysis"*, Karlsruhe Institute of Technology, April 2016 – July 2016.
Lecture course on *"Forecasting: Theory and practice I"*, Karlsruhe Institute of Technology, October 2016 – February 2017.

Tilmann Gneiting, Sebastian Lerch:

Seminar on *"Statistical forecasting"*, Karlsruhe Institute of Technology, October 2015 – February 2016.

Frauke Gräter:

Notre Dame University summer school course on *"molecular simulations Molecular Dynamics simulations"*, Notre Dame University, USA, July 4 – 6, 2016.

Frauke Gräter, Rebecca Wade:

Course on *"Bioinformatics"* for Bachelor students in Biology, Heidelberg University, Germany, January 25 – 29, 2016. Lecture and practical's on *"Computational molecular biophysics"* for Masters students in Physics and Biology, Heidelberg University, Germany, October 2015 – February 2016. Lecture and tutorials on *"Fundamental methods of Computational Physics"* for Masters students in Physics, Heidelberg University, Germany, October 2015 – February 2016. Contribution to lecture on *"Computational biochemistry"* for Master's students in Biochemistry, Heidelberg University, Germany, November, 2016.

Frauke Gräter, Katra Kolsek, Csaba Daday (MBM) Rebecca C. Wade, Mehmet Ali Öztürk, Joanna Panecka, Neil J. Bruce (MCM):

Master's course *"Computational Molecular Biophysics"*, Heidelberg University, January 13 – February 3, 2016.

Vincent Heuveline:

Lecture on Uncertainty Quantification, Heidelberg University, winter term 2015 / 2016. Lecture on High performance computing and systems, NiPS Summer School 2016, Aalborg University, Denmark, August 13 – 16, 2016.

Vincent Heuveline, Simon Gawlok:

Proseminar Einführung in die Numerik, Heidelberg University, winter term 2016 / 2017.

Vincent Heuveline, Marcel Kunze:

Seminar IT-Cybersecurity, Heidelberg University, winter term 2016 / 2017.

Vincent Heuveline, Marcel Kunze, Wolf Schünemann:

Seminar IT-Cybersecurity, Heidelberg University, winter term 2015 / 2016. Seminar IT-Cybersecurity, Heidelberg University, summer term 2016.

Vincent Heuveline, Wolf Schünemann:

Lecture on *"Cybersicherheit und Internet Governance"*, Heidelberg University, summer term 2016.

Paschalia Kapli:

Practical course on *"Species Concepts and Species Delimitation"*, ForBio, research school in biosystematics, Trondheim, Norway, October 24 – 28, 2016.

Alexey Kozlov:

Computational Molecular Evolution summer school, HCMR, Heraklion, Greece, May 8 – 19, 2016.

Antonia Stank, Rebecca C. Wade:

Practical course on *"Using protein structures to understand protein function"*, EMBO Course on *"Computational Biology: Genomes to Systems"*, EMBL, Heidelberg, June 23, 2016.

Michael Strube:

PhD Colloquium, Department of Computational Linguistics, Heidelberg University (October 2015 – February 2016). Seminar: *"Discourse Processing and Applications"*, Department of Computational Linguistics, Heidelberg University (October 2015 – February 2016). PhD Colloquium, Department of Computational Linguistics, Heidelberg University (April 2015 – July 2016).

Rüdiger Pakmor:

Block Course *"Python for Scientists"*, Department of Physics and Astronomy, Heidelberg University, August 2016.

Rüdiger Pakmor, Frauke Gräter:
Lecture *"Fundamentals of Simulation Methods"*, Department of Physics and Astronomy, Heidelberg University, October 2016 – February 2017.

Christoph Pfrommer:
"The Physics of Star Formation" Lecture at Potsdam University, January 13 – 14, 2016. *"Galaxy Formation and High-Energy Astrophysics"*, Master seminar at Heidelberg University, summer term 2016. *"Cosmology"*, Master seminar at Heidelberg University, winter term 2016 / 2017.

Friedrich Röpke:
"Computational Astrophysics", lecture course, Heidelberg University, Heidelberg, summer semester 2016. *"Stellar structure, evolution, and explosions"*, lecture course, Heidelberg University, Heidelberg, winter semester 2016 / 2017.

Friedrich Röpke, Andreas Bauswein:
"White Dwarfs, Neutron Stars, and Black Holes – Compact Objects in Astrophysics", lecture course, Heidelberg University, Heidelberg, winter semester 2015 / 2016.

Volker Springel:
Study group *"Experimental Physics III: Atom- and Quantumphysics"*, Department of Physics and Astronomy, Heidelberg University, October 2016 – February 2017.

Volker Springel, Rüdiger Pakmor:
Lecture *"Fundamentals of Simulation Methods"*, Department of Physics and Astronomy, Heidelberg University, October 2015 – February 2016.

Volker Springel, Karl-Heinz Brenner, Joachim Wambsganß:
Seminar on *"Computational Methods in Optics and Astrophysics"*, Department of Physics and Astronomy, Heidelberg University, April 2016 – July 2016.

Alexandros Stamatakis, Alexey Kozlov, Tomas Flouri, Lucas Czech, Diego Darriba:
Lecture *"Introduction to Bioinformatics for Computer Scientists"*, computer science Master's program at Karlsruhe Institute of Technology, winter term, 2016.

Alexandros Stamatakis, Alexey Kozlov, Tomas Flouri, Lucas Czech, Diego Darriba, Paschalia Kapli:
Seminar *"Hot Topics in Bioinformatics"*, computer science Master's program at Karlsruhe Institute of Technology, summer term, 2016.

Rebecca Wade:
Ringvorlesung *"Electrostatics and Solvation for Biomolecules"*, M. Sc. Biochemistry, Heidelberg University, November 11, 2016. MoBi4all: *"Biomolecular interactions: a computational approach"*, M. Sc. Molecular Biotechnology, Heidelberg University, January 18, 2016.Module 4, *"Protein Dynamics and Biomolecular Recognition: Insights from Simulations"*, M. Sc. Molecular Cell Biology, Heidelberg University, March 9, 2016. Module 3, *"Protein Modeling"*, M. Sc. Molecular Cell Biology, Heidelberg University, May 4 & 9, 2016. Ringvorlesung *"Structure and Dynamics of Biological Macromolecules"*, *"Electrostatics, solvation and protein interactions"*, B. Sc. Biosciences, Heidelberg University, June 23, 2016. Ringvorlesung *"Biophysik"*, *"Receptor-Ligand Interactions: Structure and Dynamics"*, B. Sc. Molecular Biotechnology, Heidelberg University, December 1, 2016.

Rebecca Wade, Neil Bruce, Anna Feldman-Salit, Gaurav Ganotra, Daria Kokh, Prajwal Nandekar, Mehmet Ali Öztürk, Joanna Panecka, Ina Pöhner, Stefan Richter, Antonia Stank (MCM), Frauke Gräter, Camilo Aponte-Santamaria, Eduardo Cruz-Chú, Jing Zhou (MBM):
B.Sc. Biosciences practical course *"Grundkurs Bioinformatik"*, Heidelberg University, January 25 – 29, 2016.

Anna Wienhard:
"Differentialgeometrie I", Heidelberg University, Summer Term 2016. "*Differentialgeometrie II"*, Heidelberg University, Winter Term 2017.

Peter Zaspel:
Lecture on *"Mathematik und Informatik A für molekulare Biotechnologie"*, Heidelberg University, winter term 2015 / 2016. Lecture on *„Einführung in die Numerik"*, Heidelberg University, summer term 2016.

9.1 Guest Speaker Activities

Camilo Aponte-Santamaría:

"Disentangling biomolecular interaction networks in hemostasis through multiscale simulations", Selection Symposium, Max Planck Tandem Groups, University of Andes, Bogotá, Colombia, February 15 – 16, 2016.

Christian Arnold:

"Zoomed cosmological simulations in f(R) modifed gravity", 8th Bethe Center Workshop *"Particle Physics meets Cosmology"*, Bad Honnef, Germany, October 2016.

Andreas Bauswein:

"Perspectives in modeling stellar mergers", NuPECC Town Meeting, GSI Darmstadt, Germany, February 17, 2016. *"Colliding neutron stars as sources of gravitational waves"*, Physics Colloquium, University of Mainz, Germany, May 17, 2016. *"Revealing the high-density equation of state through binary neutron star mergers"*, MITP workshop: Neutron Skins of Nuclei, Mainz, Germany, May 18, 2016. *"Neutron-star merger ejecta"*, The r-process: connecting FRIB with the cosmos, East Lansing, USA, June 14 2016. *"Neutron Star Mergers"*, NAVI-JINA-CEE 3rd Astrophysical Nuclear Reaction Network School, Schmitten, Germany, August 22, 2016; Zakopane Conference on Nuclear Physics, Zakopane, Poland, September 3, 2016. *"Dynamics and gravitational-wave emission of neutron-star mergers"*, Neutron Stars: A Cosmic Laboratory for Matter under Extreme Conditions, Bad Honnef, Germany, October 28, 2016. *"Equation of state constraints from postmerger gravitational-wave emission"*, Workshop on Numerical Relativity in Matter Space Times for Gravitational Wave Astronomy, Valencia, Spain, December 13, 2016.

Neil Bruce:

"Simulating Macromolecular Recognition through Brownian Dynamics Simulations", School on Molecular Modeling for Life Sciences, Pula, Sardinia, Italy, June 6 – 10, 2016.

Lucas Czech and Alexey Kozlov:

"Phylogenetic Placement of Short Reads: Visualization, Species Delimitation and Taxonomic Annotation", 1st UniEuks Implementation Meeting, Paris, France, May 2016.

Philipp Edelmann:

"Modeling Dynamic Phases in Stellar Evolution using Multidimensional Hydrodynamics Simulations", APEC seminar, Kavli IPMU, Japan, January 21, 2016. *"Low-Mach Number Hydrodynamics in Astrophysics"*, Astro Lunch Seminar, Newcastle University, UK, October 19, 2016. *"Improving Stellar Models with Low-Mach Number Hydrodynamics"*, RIKEN, Japan, November 10, 2016. *"Hydrodynamics of Rotating Stars"*, APEC seminar, Kavli IPMU, Japan, November 17, 2016.

Philipp Girichidis:

"Dynamical impact of cosmic rays on the ISM and galactic outflows", conference *"Computational Galaxy Formation"*, Ringberg Castle, Germany, May 9, 2016; conference "ASTRONUM", Monterey, California, USA, June 7, 2016. *"The Physics of the interstellar medium in cluding cosmic rays"*, SFB seminar, University of Cologne, Cologne, Germany, June 20, 2016. *"How CRs help driving outflows"*, MPA-UWC meeting, Stellenbosch, South Africa, November 22, 2016. *"Cosmic ray-driven outflows from the ISM"*, workshop *"Cosmidyn"*, Montpellier, France, December 1, 2016. *"Importance of the initial conditions for star and star cluster formation"*, conference *"Stellar aggregates over mass and spatial scales"*, Bad Honnef, Germany, December 6, 2016.

Tilmann Gneiting:

"Probabilistic weather forecasts, with a view towards seismology", INGV – Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy, February 18, 2016. *"Extreme events and verification"*, ECMWF - European Centre for Medium-Range Weather Forecasts, Reading, UK, March 2, 2016. *"Strictly and non-strictly positive definite functions on spheres"*, International Conference on Multivariate Approximation, Schloss Rauischholzhausen, Germany, April 1, 2016. *"Uncertainty quantification in weather forecasting"*, SIAM Conference on Uncertainty Quantification, Lausanne, Switzerland, April 7, 2016. *"Positive definite functions on spheres"*, Festkolloquium on the occasion of the 60th birthday of Professor Zoltán Sasvári, Technical University Dresden, Dresden, Germany, April 28, 2016. *"Probabilistische Wettervorhersage"*, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany, June 2, 2016. *"Statistical post-processing of ensemble weather forecasts"*, Université de Paris – Diderot, Paris, France, November 21, 2016. *"Probabilistische Wettervorhersage"*, Inaugural lecture, Karlsruhe Institute of Technology, Karlsruhe, Germany, November 25, 2016. *"Of quantiles and expectiles: Consistent scoring functions, mixture representations, and forecast rankings"*, Conference on Computational and Financial Econometrics – CM Statistics, Seville, Spain, December 9, 2016.

Martin Golebiewski:

"Standardising Data in Biotechnology: From Community to ISO Standards", Keynote lecture at the symposium *"The 50th Anniversary of World Data Center for Microorganisms"*, Beijing, China, September 6 – 8, 2016. *"COMBINE and ISO/TC 276 Biotechnology: From grassroots community to official ISO standards"*, COMBINE 2016, Newcastle upon Tyne, UK, September 19 – 24, 2016. Panel discussion *"Harmonising standardisation strategies to increase efficiency and competitiveness of European life sciences research"*, BioDataWorld Congress 2016, Wellcome Genome Campus, Hinxton, UK, October 26 – 27, 2016.

Frauke Gräter:

"Mechanosensing of Focal Adhesion Kinase", Seminar, Applied Physics Department, Ludwig Maximilian University of Munich (LMU), Germany, February 16, 2016. *"Molecular mechanisms of primary hemostasis and links to sepsis"*, Kick-off Workshop SCIDATOS, Heidelberg University, Germany, February 23, 2016. *"Visualizing molecular stresses and dynamics: proteins in the crash test"*, VIZBI Conference, The European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, March 9 – 11, 2016. *"Mechanochemistry from computer simulations"*, Department Seminar – Theoretical Chemistry, University of Würzburg, Germany, March 22, 2016. *"Proteins under stress – how force ruptures molecules and fibers"*, Conference "Across Scales", East China Normal University, Shanghai, China, May 26 – 30, 2016. *"Lack of structures as a challenge for protein simulations"*, SFB Seminar, University of Stuttgart, Germany, June 9, 2016. *"Molecular high-five: ultrafast binding of IDPs to structured proteins"*, Gordon Research Conference on Intrinsically disordered proteins, Les Diablerets, Switzerland, June 27 – 29, 2016. *"Molecular sensors: translating force into biochemical signals"*, Symposium Doktorandenkolleg, University of Linz, St Wolfgang, Austria, July 1 – 3, 2016. *"Disulfide bond mechanochemistry from simulations"*, FASEB Meeting, Steamboat Springs, USA, August 15 – 18, 2016. *"Mechanosensing of proteins and membranes from simulations"*, GTBio Annual Meeting of French Biophysical Society, Obernai, France, December 14 – 16, 2016.

Samuel Jones:

"Stars on Death Row", APEC Seminar, Kavli Institute for the Physics and Mathematics of the Universe (WPI), Tokyo, Japan, January 16, 2016. *"Perspectives in modelling stellar evolution"*, NuPECC Town Meeting, GSI Darmstadt, February 17, 2016. *"Simulating the lives and deaths of 8 – 10 solar-mass stars"*, Theorieseminar, TU Darmstadt, Germany, April 14, 2016; Astrophysics Seminar, Los Alamos National Laboratory, New Mexico, USA, May 24, 2016; *"Supernovae, Hypernovae and Binary Driven Hypernovae An Adriatic Workshop"*, Pescara, Italy June 20, 2016. *"Convection and deflagrations in the progenitor systems of supernovae"*, APEC Seminar, Kavli Institute for the Physics and Mathematics of the Universe (WPI), Tokyo, Japan, November 10, 2016. *"Supernovae and their progenitor systems"*, University of Hull, UK, December 15, 2016.

Alexey Kozlov:

"Identification of taxonomically mislabeled sequences in large databases", 2nd EukREF workshop, Barcelona, Spain, Nov. 28 – Dec. 2 2016.

9 Miscellaneous

Sebastian Lerch:

"Probabilistic forecasting and comparative model assessment based on MCMC output", 36th International Symposium of Forecasters, Santander (Spain), June 20, 2016.

Davide Mercadante:

"The basis and advantages of extreme plasticity: nucleoporins as a paradigm", 50th Biophysical Society Meeting, Los Angeles, USA, February 27 – March 2, 2016. *"Exploring the advantages of lacking a structure: the functional space of the dark proteome"*, Seminar at Newcastle University, Newcastle upon Tyne, Newcastle, United Kingdom, February, 2016. *"Innovating the sampling of intrinsically disordered proteins and exploring functional advantages of structural absence"*, Seminar at the Leibniz Institute for Molecular Pharmacology, Berlin, Germany, June, 2016. *"New paradigms of protein behaviour: innovating the sampling of intrinsically disordered proteins and exploring functional advantages of structural absence"*, Seminar at The University Federico II, Naples, Italy, July, 2016. *"Exploring advantages of intrinsic disorder and solving the over-compaction of IDPs"*, Seminar at the Technical University of Darmstadt, Darmstadt, Germany, July, 2016. *"Computational investigation of protein dynamics: exploring the nanoworld through high-performance computing"*, Bw4 supercomputing symposium, Heidelberg, Germany, August, 2016.

Sebastian Ohlmann:

"Hydrodynamics of the common envelope phase", Max Planck Institute for Astrophysics, Garching, Germany, January 13, 2016. *"Hydrodynamics of the common envelope phase"*, Computational Astrophysics Seminar, Mount Stromlo Obervatory, Australian National University, Canberra, Australia, March 30, 2016. *"Hydrodynamics of the Common Envelope Phase in High Resolution"*, Argelander Institute for Astronomy, Bonn, Germany, October 20, 2016.

Rüdiger Pakmor:

"SNe Ia from WD mergers", Astrophysical Colloquium, Australian National University, Canberra, Australia, March 2016.

Christoph Pfrommer:

"30,000 foot view of blazar heating", Conference on *"Feedback over 44 Orders of Magnitude: From Gamma-rays to the Universe"*, Perimeter Institute, Waterloo, Canada, Mar 14 – 16, 2016. *"Unveiling the physics of feedback in galaxy formation"*, Astrophysics Seminar, University of Waterloo, Canada, Mar 23, 2016. *"Cosmic ray physics in galaxy formation"*, Workshop on *"Computational Galaxy Formation"*, Ringberg, Tegernsee, Germany, May 9 – 13, 2016; Workshop on "*Magnetic Fields in Galaxies"*, Bochum, Germany, June 15 – 17, 2016. *"Cosmic ray feedback in galaxies and AGN"*, 6th International Symposium on Gamma-Ray Astronomy 2016, Heidelberg, Germany, July 11 – 15, 2016. *"Is cosmic ray heating relevant in cool core clusters?"*, Workshop on Physics of the Intracluster Medium: Theory and Computation, Minneapolis, USA, Aug 22 – 24, 2016. *"High-energy astrophysics – status and perspectives"*, plenary talk, Annual Meeting of the German Astronomical Society: *"The Many Facets of Astrophysics – Photons, Particles, and Spacetime"*, Bochum University, Germany, Sep 12 – 16, 2016. *"How cosmic rays shape the interstellar medium and galaxies"*, Annual Meeting of the German Astronomical Society: *"The Many Facets of Astrophysics – Photons, Particles, and Spacetime"*, Bochum University, Germany, Sep 12 – 16, 2016. *"The physics of propagating TeV gamma-rays: from plasma instabilities to cosmological structure formation"*, Conference on *"Dynamical processes in space plasmas"*, The Dead Sea, Israel, Apr 3 – 11, 2016; Ruhr Astroparticle-Plasma Physics (RAPP) Center Inauguration Meeting, Bochum University, Germany, Sep 22 – 23, 2016. *"How cosmic rays shape galaxies"*, Heidelberg Joint Astronomical Colloquium, Heidelberg University, Germany, July 26, 2016; Physics Colloquium, Innsbruck University, Austria, Nov 21 / 22, 2016. *"Feedback by Active Galactic Nuclei in Galaxy Clusters"*, Joint Astronomical Seminar, Innsbruck University, Austria, Nov 21 / 22, 2016. *"Cosmic ray transport in galaxy formation"*, Workshop on Dynamical effects of cosmic rays, Montpellier, France, Dec 1 – 2, 2016. *"Cosmic ray heating in cool core clusters"*, Conference *"Galaxy clusters: physics laboratories and cosmological probes"*, Cambridge, UK, Dec 5 – 9, 2016.

Ina Pöhner:

"Computational drug design for Trypanosomatidic infections", BioMac Seminar Series, University of Bayreuth, Bayreuth (Germany), January 14, 2016.

Kai Polsterer:

"KDDIG the future of machine learning in the IVOA", IVOA Interop, Cape Town, South Africa, May 13, 2016. *"Machine Learning in Astronomy: lessons learned from learning machines"*, MPIA Colloquium, Heidelberg. Germany, April 15, 2016; SKA-South Africa / IDIA Colloquium, Cape Town, May 15, 2016. *"Knowledge Discovery and Astronomical Databases"*, ZAH Astronomers' Convention 2016, University of Heidelberg, Germany, June 15, 2016. *"Datenimpulse aus der Wissenschaft: Astroinformatik / Maschinelles Lernen in der Astronomie"*, Netzwerk Recherche, Hamburg, Germany, July 9, 2016. *"Uncertain Photometric Redshifts: proper use of probability density functions"*, Annual Meeting of the German Astronomical Society, Splinter Meeting 2016: E-Science & Virtual Observatory, Ruhr-Universität Bochum, Germany, September 14, 2016. *"Multimodal Redshift Estimation"*, IAU Symposium 325, Astroinformatics 2016, Sorrento, Italy, October 24, 2016.

Friedrich Röpke:

"Typ Ia Supernovae – wie Sternenexplosionen im Computer simuliert werden", 61. Rhein-Neckar Gesprächskreis, SAP, Walldorf, Germany, January 28, 2016. *"Perspectives in modeling stellar explosions"*, NuPECC Town Meeting, GSI Darmstadt, Germany, February 17, 2016 *"Models for Type Ia supernovae – status and challenges"*, *"The Next Decade of European SN Cosmology"* Meeting, Humboldt University Berlin, Germany, April 15, 2016. *"Supernovae and nuclear astrophysics"*, Nuclear Astrophysics in Germany – A Community Meeting in Darmstadt, Germany, November 16, 2016.

Kashif Sadiq:

"Towards multiscale modeling of retroviral self-assembly and maturation", Solvay Workshop: Bridging the gaps at the PCB interface – Multiscale modelling in physics chemistry and biology, International Solvay Institutes, Brussels, Belgium, April 19 – 21, 2016. *"Predictive synthetic retrovirology: A new paradigm for gene delivery"*, European Health Science Match, Life Sciences Cluster, Heidelberg, Germany, September 26, 2016.

Roman Schefzik:

"Physically coherent probabilistic weather forecasts via discrete copula-based ensemble postprocessing methods", Computational Genome Biology Group, German Cancer Research Center, Heidelberg, Germany, January 13, 2016. *"Physically coherent probabilistic weather forecasts via discrete copula-based ensemble postprocessing methods"*, Royal Meteorological Institute of Belgium, Brussels, Belgium, March 2, 2016.

Martin Sparre:

"The Illustris zoom simulation project: Zooming in on major merger", Center for Computational Astrophysics, New York City, USA, September 19, 2016; Max Planck Institute for Astronomy, Heidelberg, September 22, 2016; Geneva Observatory, Versoix, Switzerland, November 29, 2016. *"The Illustris zoom simulation project: Merger-induced starbursts and the morphology of merger remnants"*, Max Planck Institute for Astrophysics, Garching, September 13, 2016; European Southern Observatory, Garching, September 14, 2016.

Volker Springel:

"Exascale simulations of the magnetic universe", SPPEXA Symposium, Leibniz Supercomputing Centre, Garching, January 2016. *"Supercomputer Simulationen von Galaxien, Dunkler Materie und Dunkler Energie"*, Alfried Krupp Wissenschaftskolleg Greifswald, Vortragsreihe *"Welt.Raum"*, Greifswald, January 2016. *"Simulating Cosmic Structure Formation"*, Workshop on High Performance Computing, International School for Advanced Studies, SISSA, Trieste, Italy, February 2016. *"Cosmic magnetism in simulations of disk galaxy formation"*, Astrophysical Colloquium, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK, February 2016; Astrophysical Colloquium, University of Sussex, UK, April 2016. *"Properties of galaxies reproduced in the Illustris simulation"*, ZAH Scientific Advisory Council, Heidelberg, February 2016. *"Magnetic fields, cosmic rays, and outflows in galaxies"*, Simons Symposium on Galactic Outflows, Schloss Elmau, April 2016. *"Cosmic structure formation in non-standard gravity scenarios"*, Results Workshop John von Neumann-Institut für Computing, Forschungszentrum Jülich, April 2016. *"Circumventing road blocks towards multi-scale, multi-physics simulations of galaxy formation"*, Computational Galaxy Formation Conference, Ringberg Castle, Tegernsee, May 2016. *"Exploring the physics of galaxy formation with supercomputers"*, Physical Colloquium, Rheinische Friedrich-Wilhelms-Universität Bonn, June

2016. *"Cosmology on a moving mesh"*, ZAH Astronomer's Convention, Heidelberg, June 2016. *"Simulating cosmic structure formation"*, 4th Heidelberg Laureate Forum, Wine & Dine Event, Speyer, September 2016. *"Hydrodynamical simulations of galaxy formation"*, Joint Astrophysics Colloquium, Institute for Advanced Study and Princeton University, Princeton, USA, September 2016; 11th Sino-German Conference on Cosmology and Galaxy Formation, Guangzhou, China, December 2016. *"Hydrodynamical simulations of dark universes"*, Particle Physics meets Cosmology, Physics Centre Bad Honnef, October 2016. *"The Illustris++ Project: Next generation galaxy formation simulations"*, Results & Review Workshop, High-Performance Computing Center Stuttgart, University of Stuttgart, October 2016. *"Das Universum im Supercomputer"*, Vortragsreihe Faszination Astronomie, Haus der Astronomie, Heidelberg, November 2016. *"Star formation physics in simulations on galaxy scales"*, Heidelberg-Harvard Star Formation Conference, Heidelberg, November 2016. *"Galaxy Formation and Fundamental Physics"*, Jülich Supercomputer Center, JUQUEEN Nachfolge Workshop, Jülich, November 2016. *"Dunkle Materie, Galaxien, und das kosmische Netz"*, 29. Raumfahrt Kolloquium, FH Aachen, November 2016.

Alexandros Stamatakis:

"Challenges & problems in phylogenetic inference and bioinformatics", JOBIM 2016, Lyon, France, June 2016; Max Planck Institute & University of Saarland, Saarbrücken, Germany, November 2016; Hellenic Bioinformatics Conference (HBio) 2016, Thessaloniki, Greece, November 2016.

Michael Strube:

"The Non-Utility of Semantics for Coreference Resolution (Corbon Remix)", Keynote Speech at the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2016) at NAACL 2016, San Diego, CA, USA, June 2016. *"Big Brother liest mit: Informationstechnologie zwischen Überwachungsstaat und gesellschaftlichem Nutzen"*, Lecture Series *"Zukunftsfragen kontrovers"*, TU Braunschweig, Germany, November 2016. *"Fiktion und Realität in Dave Eggers' Roman 'The Circle'"*, Lecture Series *"Gegenwartsliteratur und Zeitkritik"*, Heidelberg University, Germany, December 2016.

Freeke van de Voort:

"Cosmic gas flows: the key to understanding galaxy formation", University of Surrey, Guildford, UK, November 2016.

Rebecca Wade:

"Exploring protein binding pocket dynamics for ligand design", In silico tools in drug design and target discovery, FIB: Barcelona School of Informatics, Barcelona, Spain, Feb 29, 2016. *"Prediction of the effects of protein mutations on binding properties"*, Protein Engineering Symposium, LOEWE Center for Synthetic Microbiology, Marburg, Germany, May 30, 2016. *"Computational exploration of protein dynamics for ligand design"*, 3rd NovAliX Conference *"Biophysics in Drug Discovery 2016: Developing the Synergy between Biophysics and Medicinal Chemistry to Deliver Better Drugs"*, Strasbourg, France, June 7 – 10, 2016. *"In silico prediction of biomolecular recognition"*, ISQBP President's Meeting 2016, Bergen, Norway, June 19 – 22, 2016. *"Simulation of protein recognition"*, EMBO Course on *"Computational Biology: Genomes to Systems"*, Heidelberg, Germany, June 23, 2016. *"Simulation of protein adsorption on inorganic surfaces"*, VII-th conference on *"Modeling and Design of Molecular Materials 2016 (MDMM 2016)"*, Trzebnica, Wroclaw, Poland, June 26 – 30, 2016. *"The charged affairs of proteins: simulating the weak and the strong"*, Gordon Research Conference on Computational Chemistry on *"Theory and simulation across scales in molecular science"*, Girona, Spain, July 24 – 29, 2016. *"Exploring protein dynamics for ligand design"*, 21st Euro-QSAR Symposium, Verona, Italy, Sep 4 – 8, 2016. *"In silico prediction of biomolecular recognition"*, 6th EuCheMS Chemistry Congress, Sevilla, Spain, Sep 11 – 15, 2016.

Johannes Wagner:

"Efficient molecular dynamics computing setups and multiscale simulations of spider silk fibers", Seminar at The University Federico II, Naples, Italy, July, 2016.

Anna Wienhard:

"How to order surface groups?", Distinguished Women in Mathematics Lecture Series, UT Austin, January 26, 2016. *"On rigidity and flexibility of (some) discrete subgroups of Lie groups"*, Conference on Dynamics, Geometry and Number Theory, Institut Henri Poincaré, Paris, France, June 14 2016. *"Maximal representations, deformation spaces, and geometric structures."*, Geometries, Surfaces and Representations of Fundamental Groups, University of Maryland, College Park, Maryland, USA, June 22, 2016. *"Rotation numbers – from elements to groups"*, Kolloquium, Universität Hamburg, Germany, July 5, 2016. *"Positivity and higher Teichmüller theory"*, Hitchin 70: Celebrating 30 years of Higgs bun-

dles and 15 years of generalized geometry, ICMAT, Madrid, Spain, September 12, 2016. *"Generalizing Lusztig's total positivity"*, Conference on Lie Theory and Cluster Algebras, Notre Dame University and La Sapienzia, Rome, Italy, October 17, 2016. *"Rigidity and flexibility of discrete subgroups of Lie groups"*, Mathematisches Kolloquium, Universität Paderborn, Germany, December, 12, 2016.

Dandan Xu:

"Lensing flux ratio anomalies and the projected subhalo mass fraction – from a simulation perspective", Workshop on *"Dark Matter on the Smallest Scales"*, Lorentz Center, Leiden, Netherlands, April 2016. *"The inner structures of early-type galaxies in the Illustris simulation"*, Argelander Institute for Astronomy, University of Bonn, Bonn, Germany, November 2016. *"Learning about early-type galaxies and their observational methods using cosmological hydrodynamic simulations"*, The fine structure of the local Universe, Chinese Academy of Sciences, Zhuhai, China, December, 2016.

9.2 Presentations

Talks

Irene Adrian-Kalchhauser, Martin Pippel, Jean-Claude Walser, Siegfried Schloissnig:

"Transgenerational adaptation mechanisms in wild populations of Neogobuis melanostomus (the round goby)", in EMBO|EMBL Symposium: New Model Systems for Linking Evolution and Ecology, EMBL Heidelberg, Germany, May 8 – 11, 2016.

Christian Arnold:

"Zoomed cosmological simulations in f(R) modifed gravity", Harvard-Smithsonian Center for Astrophysics, Cambridge, USA, October 2016; Massachusetts Institute of Technology, Cambridge, USA, October 2016. *"Cosmological hydrodynamical simulations in f(R) modifed gravity"*, Max Planck Institute for Astrophysics, Munich, Germany, November 2016; Institute of Astronomy, University of Cambridge, Cambridge, UK, July 2016; Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK, July 2016. *"Cosmological simulations in f(R) modifed gravity"*, Institute for Theoretical Physics, Heidelberg University, Germany, November 2016.

Andreas Bauswein:

"Gravitational wave emission from neutron star mergers", NAVI Physics Days, Darmstadt, Germany, January 19, 2016. *"Detection of Gravitational Waves"*, special seminar , HITS, Heidelberg, Germany, February 22, 2016. *"Oscillations of Neutron-Star Merger Remnants"*, NewCompStar meeting on oscillations and instabilities in neutron stars, Southampton, United Kingdom, September 14, 2016.

Lucas Czech:

"Visualizing Evolutionary Placement of Short Reads on Phylogenetic Trees", Heidelberg Unseminars in Bioinformatics (HUB) 22 – Visual Perspectives in Science, Heidelberg, Germany, March 8, 2016. *"Mathematik auf evolutionären Stammbäumen"*, Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg, Germany, June 20, 2016. *"Do Phylogenetic Tree Viewers correctly display Support Values?"*, Hellenic Bioinformatics Conference (HBio) 2016, Thessaloniki, Greece, November 2016.

Csaba Daday:

"Simulating force response at cell junctions: desmoplakin as a molecular force sensor?", Annual Meeting of the German Biophysical Society, Erlangen, Germany, September 27, 2016.

Antonio D'Isanto:

"Uncertain photometric redshifts with machine learning methods", IMPRS Seminar, Heidelberg (Germany), June 16, 2016.

Philipp Edelmann:

"Dynamical Shear Instabilities in Massive Stars", 18th Workshop on Nuclear Astrophysics, Ringberg Castle, Tegernsee, Germany, March 14, 2016. *"A 3D hydrodynamical perspective on shear instabilities in massive stars"*, Bridging the gap: from massive stars to supernovae, Chicheley Hall, UK, June 1, 2016. *"Sind Sterne rund? Mathematik in der stellaren Astrophysik"*, Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg, Germany, June 28, 2016.

9 Miscellaneous

Kira Feldmann:

"Die Mathematik der Wettervorhersage", Sprechen Sie Mathematik? Akademische Mittagspause, Heidelberg University, Heidelberg (Germany), July 8, 2016.

Tomas Flouri:

"Efficient parallelization schemes for molecular phylogenetics applications", Big Data Institute, University College London, UK, Nov. 28 2016.

Nikos Gianniotis:

"Autoencoding Astronomical Time Series for Visualisation", ITA Blackboard Colloquium, Heidelberg University (Germany), January 25, 2016; ZAH Astronomers' Convention 2016, University of Heidelberg (Germany), June 15, 2016. *"Visualisation of time series via neural networks"*, Summer school on Astrostatistics & Data Mining, Heidelberg (Germany), September 16, 2016; Annual Meeting of the German Astronomical Society, Splinter Meeting 2016: E-Science & Virtual Observatory, Ruhr-Universität Bochum (Germany), September 14, 2016. *"A Spectral Model for Multimodal Redshift Estimation"*, IEEE Symposium on Computational Intelligence and Data Mining, Athens (Greece), December 8, 2016.

Philipp Girichidis:

"Shock structure of the multiphase ISM in cosmic-ray-MHD simulations", conference *"SHOCKS2016"*, Torun, Poland, September 16, 2016.

Martin Golebiewski:

"NormSys – Standardization in the Systems Biology Research", Round table meeting about standardisation in the life sciences, 6th Conference on Systems Biology of Mammalian Cells (SBMC 2016), Munich, Germany, April 6 – 8, 2016. *"COMBINE: Harmonizing the Standardization Processes for Modeling in Systems Biology"*, Conference *"The CHARME of standardisation in life sciences"*, Warsaw, Poland, June 21 – 22, 2016. *"Standardising Data in Systems Biology: From Community to ISO Standards"*, FAIRDOM User Meeting, International Conference on Systems Biology (ICSB 2016), Barcelona, Spain, September 15, 2016. *"Standards and Tools for Systems Biology Modelling"* and *"SABIO-RK – Reaction Kinetics Database"*, COMBINE & de.NBI Tutorial *"Modelling and Simulation Tools in Systems Biology"*, International Conference on Systems Biology (ICSB 2016), Barcelona, Spain, September 16, 2016. *"The NormSys registry for modeling standards in systems and synthetic biology"*, COMBINE 2016, Newcastle upon Tyne, UK, September 19 – 24, 2016. *"ISO/ TC 276 Biotechnology"*, ISO /TC215 & CEN / TC251 Health Informatics Joint Meeting, Lillehammer, Norway, November 13 – 17, 2016. *"NORMSYS: Building a bridge between grass-roots initiatives and standardization bodies"*, NormSys Conference *"Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences"*, Potsdam, Germany, December 5 – 6, 2016.

Robert Grand:

"The Auriga Simulation Project", SuperMUC status and results workshop, Leibniz Supercomputing Centre, Garching, Germany, April 2016. *"The Giga-galaxy project"*, Computational Galaxy Formation Conference, Ringberg Castle, Tegernsee, Germany, May 2016. *"Formation and evolution of Milky Way sized galaxies in high resolution cosmological zoom simulations"*, Conference: Discs in galaxies, European Southern Observatory, Garching, Germany, July 2016; Conference: The Milky Way and its Environment, Institut d'Astrophysique de Paris, France, September 2016. *"Formation of disc galaxies"*, Seminar, Max-Planck Institute for Astrophysics, Garching, Germany, July 2016. *"Disc dynamics in the Auriga cosmological simulations"*, Dark matter distribution in the era of Gaia, Nordita, Stockholm, Sweden, October 2016. *"Auriga Simulations"*, Galaxy Coffee Seminar, Max-Planck Institute for Astronomy, Heidelberg, Germany, October 2016.

Stephan Hemri:

"Multivariate postprocessing techniques for probabilistic hydrological forecasting", European Geosciences Union General Assembly, Vienna (Austria), April 20, 2016. *"Unsupervised postprocessing based on the Nadler approach"*, European Centre for Medium-Range Weather Forecasts, Reading (UK), November 9, 2016.

Ron Henkel:

"Model Management – Hands On", Kinetics on the move – Workshop on Data for Computational Modeling, Heidelberg, Germany, May 30 – 31, 2016. *"Combine archive and OMEX Format"*, Combine Tutorial, International Conference on Systems Biology (ICSB 2016), Barcelona, Spain, September 16, 2016. *"German Network for Bioinformatics Infrastructure (de.NBI)"*, NormSys Conference *"Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences"*, Potsdam, Germany, December 5 – 6, 2016.

Ana Maria Herrera-Rodriguez:

"Towards Modeling Self-Assembly of Silk Proteins under Flow", Heidelberg Graduate School of Mathmatical and Computational Methods for the Sciences 8th Annual Colloquium, Speyer, Germany, December 1 – 2, 2016.

Vincent Heuveline:

"Uncertainty Quantification and High Performance Computing: Challenges and Perspectives", Dagstuhl Seminar on Uncertainty Quantification and High Performance Computing, Dagstuhl, Germany, September 11 – 16, 2016. HPC in Heidelberg, ZKI Supercomputing, Heidelberg, Germany, October 13 – 14, 2016. *„Wie si-cher ist die numerische Simulation? Uncertainty Quantification als Paradigmenwechsel"*, Robert Bosch GmbH, Renningen, Germany, November 21, 2016.

Samuel Jones:

"The 'Bermuda Triangle': evolution and fate of 8 – 12 solar-mass stars", 18th Workshop on Nuclear Astrophysics, Ringberg Castle, Tegernsee, Germany, March 14, 2016. *"ONe Deflagrations"*, 11th Bi-Annual Würzburg Workshop, Universität Würzburg, Germany, December 8, 2016.

Katra Kolšek:

"Modelling Cysteine Disulfide Exchange Reactions by Molecular Dynamics Simulations", 60th Annual Meeting Biophysical Society, Los Angeles, USA, February 27 – March 2, 2016.

Jonas Kratzke:

In vitro flow assessment: From PC-MRI to computational fluid dynamics including fluid-structure interaction, SPIE – Medical Imaging 2016, San Diego, CA, USA, April 1, 2016.

Olga Krebs:

The 8th Young Scientists School *"Systems Biology and Bioinformatics"*, SBB-2016, Novosibirsk, Russia, August 23 – 25, 2016. *"FAIRDOM: Data and Model Management for Systems Biology Projects"*, International Conference on Bioinformatics of Genome Regulation and Structure / Systems Biology (BGRS 2016), Novosibirsk, Russia, August 29, 2016. *"FAIRDOM approach for semantic interoperability of systems biology data and models"*, Ontologies and Data in Life-Sciences (ODLS 2016), Halle, Germany, 29th – September 30, 2016. *"The FAIR Data Principles"*, NormSys Conference *"Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences"*, Potsdam, Germany, December 5 – 6, 2016.

Fabian Krüger:

"Probabilistic forecasting and predictive model assessment based on MCMC output", Conference of the International Society for Bayesian Analysis, Cagliari (Italy), June 17, 2016. *"Combining density forecasts under various scoring rules: An analysis of UK inflation"*, Econometric Society European Meeting, Geneva (Switzerland), August 23, 2016. *"Murphy diagrams for evaluating forecasts of value-at-risk and expected shortfall"*, Statistische Woche, Augsburg (Germany), September 15, 2016.

Sebastian Lerch:

"Probabilistic forecasting and comparative model assessment based on MCMC output", 12th German Probability and Statistics Days, Bochum (Germany), March 1, 2016; Workshop on Advances in Economic Forecasting, Heidelberg (Germany), June 6, 2016. *"Similarity-based semi-local estimation of EMOS models"*, European Geosciences Union General Assembly, Vienna (Austria), April 20, 2016. *"Forecaster's dilemma: Extreme events and forecast evaluation"*, 9th European Central Bank Workshop on Forecasting Techniques, Frankfurt (Germany), June 4, 2016; Volkswagen-Stiftung Symposium Extremes 2016 – Modeling, Analysis, and Prediction of Extreme Events, Hannover (Germany), September 14, 2016.

Philipp Lösel:

"Enhancing a diffusion algorithm for 4D image segmentation using local information", SPIE Medical Imaging 2016, San Diego, USA, February 27 – March 3, 2016. "*The NOVA project and a GPU based diffusion method for image segmentation"*, Workshop der Deutschen Zoologischen Gesellschaft Fachgruppe Morphologie, Functional Morphology and Biomechanics, University of Kiel, Kiel, Germany, September 9, 2016. *"The Biomedical Image Segmentation App"*, 9. Graduiertenforum der DZG-Fachgruppe Morphologie, ANKA, KIT, Karlsruhe, Germany, October 14, 2016. *"A GPU based Diffusion Method for Whole-Heart and Great Vessel Segmentation"*, MICCAI 2016: 19th International Conference on Medical Image Computing and Computer Assisted Intervention, Athen, Greece, October 17 – 21, 2016. *"The Biomedical Image Segmentation App"*, 12. Modellierungstag Rhein-Neckar, IWR, University Heidelberg, Heidelberg, Germany, December 8, 2016.

9 Miscellaneous

Aron Michel:
"Gravitational settling of 22Ne in sub-Chandrasekhar models of Type Ia supernovae", 18th Workshop on Nuclear Astrophysics, Ringberg Castle, Tegernsee, Germany, March 17, 2016.

Vedran Miletić:
"Towards fully open source GPU accelerated molecular dynamics simulation", Lightning talk, 2016 European LLVM Developers' Meeting, Barcelona, Spain, March 17 – 18, 2016. *"Reactive MD"*, GROMACS Developer Workshop 2016, Göttingen, Germany, May 19 – 20, 2016.

Wolfgang Müller:
"The LiSyM Data Management", LiSyM Kickoff Meeting, Munich, Germany, April 5, 2016.

Mehmet Ali Öztürk:
"Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome", Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, May 27 – 28, 2016; 3rd DKFZ / ZMBH Alliance Retreat, Kloster Schöntal, Germany, June 19 – 21, 2016.

Sebastian Ohlmann:
"Hydrodynamics of the common envelope phase", 18th Workshop on Nuclear Astrophysics, Ringberg Castle, Tegernsee, Germany, March 14 – 19, 2016; *"Binary Stars in Cambridge"* conference, Cambridge, UK, July 24 – 30, 2016.

Andreas Ott:
"Symmetrie", Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg, Germany, July 18, 2016.

Rüdiger Pakmor:
"Future computational challenges in explosion modelling", The Physics of Supernovae, Garching, September 2016. *"Subluminous violent mergers"*, Supernovae: The Outliers, Garching, September 2016. *"Cosmic ray driven winds in disk galaxies"*, Computational Galaxy Formation Conference, Ringberg Castle, Tegernsee, May 2016. *"SNe Ia from WD mergers"*, 18th workshop on Nuclear Astrophysics, Ringberg Castle, Tegernsee, March 2016.

Martin Pippel, Siegfried Schloissnig:
"De novo assembly of highly repetitive genomes", Sequencing, Finishing, and Analysis in the Future (SFAF), Santa Fe, New Mexico, USA, May 27 – 29, 2016. *"MARVEL – An efficient implementation regarding repeat saturated genomes"*, Big Data in Biology and Health, EMBL Heidelberg, Germany, September 25 – 27, 2016.

Kai Polsterer:
"Parallelized rotation and flipping invariant Kohonen maps (PINK) on GPUs", European Symposium on Artificial Neural Networks 2016, Brugge (Belgium), April 28, 2016. *"Wie lernende Maschinen Astronomen bei der Arbeit unterstützen"*, Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg (Germany), June 7, 2016.

Ina Pöhner:
"Insight for anti-parasitic drug design – From comparative pocketome analysis to computational modeling of a parasitic folate & biopterin pathway", Meeting on *"Computational Modelling with COPASI"*, Manchester Institute of Biotechnology, Manchester (UK), May 12 – 13, 2016. *"Targeting the parasitic folate pathway by a structure-based drug design approach"*, Synergy meeting for FP7-Health Drug Development for Neglected Parasitic Diseases, University of Modena and Reggio Emilia, Modena (Italy), June 15 – 16, 2016. *"Wie Mathematik helfen kann, Parasiten zu bekämpfen"*, Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg (Germany), June 27, 2016.

Friedrich Röpke:
"Modeling stars and stellar explosions", ZAH Astronomers' convention, Heidelberg, Germany, June 15, 2016. *"SN Ia explosion models: approaches – achievements – challenges"*, The Physics of Supernovae Workshop, Munich Institute for Astro- and Particle Physics, Germany, August 29, 2016. *"Thermonuclear supernovae as sources of iron group elements in the Milky Way"*, SFB 881 Meeting, Heidelberg University, Germany, October 9, 2016.

Patrick Schmidt:

"Optimal forecasting strategy for value-at-risk forecasts under the Basel Accord", Conference on Computational and Financial Econometrics – CMStatistics, Seville (Spain), December 10, 2016.

Nicolai Schoch:

"Cognitive tools pipeline for assistance of mitral valve surgery", SPIE 2016 Medical Imaging Conference (Robotics and Modeling Track), San Diego, California, USA, February 28, 2016. *"Towards Semantic Simulation for Patient-Specific Surgery Assistance"*, International Workshop on Surgical Data Science SDS 2016, Heidelberg, Germany, June 20, 2016. *"Comprehensive patient-specific information preprocessing for cardiac surgery simulations"*, Information Processing in Computer-Assisted Interventions IPCAI 2016, Heidelberg, Germany, June 21 – 22, 2016. *"Matching Presentation: Towards cognition-guided patient-specific numerical simulation for cardiac surgery assistance"*, European Health Science Match 2016, Heidelberg, Germany, September 26, 2016.

Constantin Scholl:

"The divisible load balance problem with shared cost and its application to phylogenetic inference", 15th IEEE International Workshop on High Performance Computational Biology, in conjunction with IPDPS 2016, Chicago, Illinois, USA, May 2016.

Chen Song:

Uncertainty Quantification for Biomedical Bloodpump Application, SIAM Conference on UQ 16, Lausanne , Switzerland, April 5 – 8, 2016.

Martin Sparre:

"High-resolution cosmological simulations of major mergers: some merger remnants are star-forming and have stellar discs", Conference: Mapping the Pathways of Galaxy Transformation Across Time and Space, Catalina Island (California, US), August 2016. *"Major mergers in cosmological zoom-in simulations: starbursting gas, metallicity distribution and morphology of merger remnants"*, Conference: European Week of Astronomy and Space Science 2016, Athens, Greece, July 2016.

Volker Springel:

"Galaxien, Dunkle Materie und Dunkle Energie", HITS Tag der offenen Tür, Heidelberg, Germany, July 2016. *"Wie die größten Zahlenfresser das Universum enträtseln"*, Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg, Germany, July 2016. *"Supercomputer-Simulationen von Galaxien und Dunkler Materie"*, Physik am Samstagmorgen, Studio Villa Bosch, Heidelberg, Germany, November 2016.

Alexandros Stamatakis:

"Maximizing the Number of Bugs in Bioinformatics Software", CLIMB 2016, retirement colloquium for Bernard Moret, EPFL, Lausanne, Switzerland, November 2016.

Antonia Stank:

"A combinatorial puzzle: Modelling disaggregase co-chaperone complexes", DKFZ-ZMBH Alliance Seminar, Heidelberg, Jan. 13, 2016.

Florian Stecker, Nicolaus Treib und Anna Schilling:

"Der Zauberwürfel entzaubert", Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg, Germany, May 19, 2016.

Michael Strube:

"Textkohärenz, Textqualität, Automatisches Textverstehen", Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg, Germany, July 4, 2016.

Jenny Wagner:

"Model-independent characterisation of strong gravitational lenses", Annual Meeting of the German Astronomical Society, Galaxy Cluster Session, Bochum (Germany), September 15, 2016; Sino-Germany Workshop on Galaxy and Cosmology, Guangzhou (China), December 12, 2016.

Rainer Weinberger:

"Modeling AGN feedback in the next generation cosmological volume simulations", The Changing Face of Galaxies: Uncovering Transformational Physics, Hobart, Tasmania, Australia, September 2016. *"Modeling feedback from AGN jets"*, Institute of Astronomy, Cambridge University, UK, December 2016.

Anna Wienhard:

"Die Geometrie des Raumes: Vom Leben in verschiedenen Welten", Akademische Mittagspause 2016 – Sprechen Sie Mathematik?, Peterskirche, Heidelberg, Germany, April 29, 2016.

Ulrike Wittig:

"SABIO-RK: Finding, curating, and publishing data", Kinetics on the move – Workshop on Data for Computational Modeling, Heidelberg, Germany, May 30 – 31, 2016.

Martin Wlotzka:

"An Energy-Efficient Parallel Multigrid Method. 5th European Seminar on Computing", Pilsen, Czech Republic, June 5 – 10, 2016. "*Recent Results from the Exa2Green Project"*, ICT-Energy Community Workshop, 11th HiPEAC Conference, Prague, Czech Republic, Jan 18 – 20, 2016. *"An Energy-Efficient Parallel Multigrid Method"*, Workshop on *"Minimizing Energy Consumption of Computing to the Limit"*, 11th HiPEAC Conference, Prague, Czech Republic, Jan 18 – 20, 2016.

Dandan Xu:

"The inner structures of early-type galaxies in the Illustris simulation", Sino-Germany workshop on galaxy formation, GuangZhou, China, December 2016.

Jolanta Zjupa:

"Understanding galaxy formation on a moving mesh", International HPC Summer School, University of Ljubljana, Ljubljana, Slovenia, June 26 – July 1, 2016.

Posters

Camilo Aponte-Santamaría:

"Stress propagation through biological lipid bilayers", Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, May, 2016.

Lucas Czech, Simon Berger, Denis Krompaß, Jiajie Zhang, Paschalia Kapli, Pavlos Pavlidis, and Alexandros Stamatakis:

"Evolutionary Placement of Short Reads – Methods, Applications, and Visualization", Poster, EMBO EMBL Symposium: A New Age of Discovery for Aquatic Microeukaryotes, Heidelberg, Germany, Jan. 26, 2016.

Csaba Daday:

"Simulating force response at cell junctions: desmoplakin as a molecular force sensor?", Annual Meeting of the German Biophysical Society, Erlangen, Germany, September 25 – 28, 2016; poster, Gordon Research Conference, *"Single Molecule Approaches to Biology"*, Hong Kong, China, July 3 – 8, 2016. *"Molecular mechanism of force sensing in desmoplakin"*, Workshop on *"Computer Simulation and Theory of Macromolecules"*, Göttingen, Germany, May 27 – 28, 2016.

Antonio D'Isanto, Kai Polsterer:

"Uncertain Photometric Redshifts with Deep Learning Methods", Astroinformatics 2016, Sorrento, Italy, October 20 – 24, 2016.

Kira Feldmann:

"Spatial post-processing for temperature forecasts", Poster, European Geosciences Union General Assembly, Vienna (Austria), April 20, 2016.

Gaurav Kumar Ganotra, Daria B. Kokh, Rebecca C. Wade:

"Computational methods to predict drug-binding kinetics", HGS MathComp Annual Colloquium 2016, Speyer, December 1 – 2, 2016; Helmholtz Drug Research Conference on *"Drug Innovation in Academia"*, DKFZ Heidelberg, Germany, December 8 – 9, 2016.

Martin Golebiewski, Olga Krebs, Quyen Nguyen, Stuart Owen, Natalie Stanford, Dawie Van Niekerk, Katy Wolstencroft, Finn Bacall, Renate Kania, Maja Rey, Andreas Weidemann, Ulrike Wittig, Jacky L. Snoep, Wolfgang Mueller, Carole Goble:

"Data Needs Structure: Data and Model Management for Distributed Systems Biology Projects", 6th Conference on Systems Biology of Mammalian Cells (SBMC 2016), Munich, Germany, April 6 – 8, 2016.

Martin Golebiewski, Alexander Nikolaew, Nils Woetzel, Jill Zander:
"The NormSys registry for modeling standards in systems and synthetic biology", 6th Conference on Systems Biology of Mammalian Cells (SBMC 2016), Munich, Germany, April 6 – 8, 2016.

Martin Golebiewski, Olga Krebs, Rostyk Kuzyakiv, Katy Wolstencroft, Natalie J Stanford, Stuart Owen, Quyen Nguyen, Finn Bacall, Norman Morrison, Jakub Straszewski, Caterina Barillari, Lars Malmstroem, Bernd Rinn, Jacky Snoep, Wolfgang Müller, Carole Goble:
"FAIRDOM approach for Data and Model Management in Systems Biology", 17th International Conference on Systems Biology 2016 (ICSB 2016), Barcelona, Spain, September 16 – 20, 2016.

Martin Golebiewski, Alexander Nikolaew, Susanne Hollmann, Babette Regierer, Nils Wötzel, Bernd Müller-Röber:
"Harmonizing Standardization Processes for Model and Data Exchange in Systems Biology", 17th International Conference on Systems Biology 2016 (ICSB 2016), Barcelona, Spain, September 16 – 20, 2016.

Ana Maria Herrera-Rodriguez:
"Stretching and self-assembly of silk proteins under flow", Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, May 27 – 28, 2016.

Susanne Hollmann, Martin Golebiewski, Babette Regierer:
NormSys & CHARME Two initiatives that aim at Harmonizing the Standardization Processes for Data Exchange in Systems Biology, 6th Conference on Systems Biology of Mammalian Cells (SBMC 2016), Munich, Germany, April 6 – 8, 2016.

Wiktoria Jedwabny, Joanna Panecka, Edyta Dyguda-Kazimierowicz, Rebecca C. Wade, W. Andrzej Sokalski:
"Application of theoretical ab initio models for scoring inhibitory activities in fragment-based drug design against Trypanosoma brucei pteridine reductase 1.", 12th German Conference on Chemoinformatics, Fulda, Germany, November 6 – 8, 2016; Helmholtz Drug Research Conference on *"Drug Innovation in Academia"*, Heidelberg, Germany, December 8 – 9, 2016.

Philipp Kämpfer:
"Assembly of DNA Sequences", Genome Informatics 2016, Hinxton, UK, September 19 – 22, 2016.

Daria Kokh, Marta Amaral, Joerg Bomke, Matthias Dreyer, Matthias Frech, Maryse Lowinski, Alexey Rak, Rebecca C. Wade:
"Simulation of protein-small molecule unbinding kinetics using enhanced sampling molecular dynamics simulations", EuroQSAR21, Verona, Italy, Sept. 4 – 8, 2016.

Daria Kokh, Marta Amaral, Joerg Bomke, Matthias Dreyer, Matthias Frech, Maryse Lowinski, Alexey Rak, Rebecca C. Wade:
"In silico prediction of relative drug-protein binding kinetics", Helmholtz Drug Research conference on *"Drug innovation in Academia"*, Heidelberg, Germany, Dec. 8, 2016.

Katra Kolšek:
"Mechanistic determination of covalent ligand binding", 21st European Symposium on Quantitative Structure-Activity Relationship, Verona, Italy, September 4 – 8, 2016.

Hadas Leonov, Lihua An, Nils Wötzel, Ulrike Wittig, Wolfgang Müller:
"Drug Discovery Enhancing Systems", International de.NBI-Symposium *"Bioinformatics for Human Health and Disease"*, Heidelberg, Germany, November 7 – 9, 2016.

Sebastian Lerch:
"scoringRules – A software package for probabilistic model evaluation", Poster, European Geosciences Union General Assembly, Vienna (Austria), April 22, 2016.

Davide Mercadante:
"Forecasting the effects of increasing water-dispersion forces on intrinsically disordered proteins dynamics", Gordon Research Conference on Intrinsically disordered proteins, Les Diablerets, Switzerland, July, 2016.

9 Miscellaneous

Prajwal P. Nandekar, Ghulam Mustafa, Rebecca C. Wade: *"Dynamics of a Complex of Cytochrome P450 and CYP450 Reductase in a Phospholipid Bilayer"*, Helmholtz Drug Research conference on *"Drug Innovation in Academia"*. DKFZ Heidelberg, Germany, December 8 – 9, 2016.

Daniele Narzi, Neil Bruce, Francesco Colizzi, Rodrigo Perera Casasnovas, Elisa Frezza, Siri Camee van Keulen, Juliette Martin, Richard Lavery, Modesto Orozco, Ursula Roethlisberger, Rebecca Wade, Jeanette Hellgren Kotaleski, and Paolo Carloni:
"Molecular Simulation Methods for HBP", HBP Annual Summit 2016, Madrid, Spain, October 12 – 15, 2016.

Mehmet Ali Öztürk, Rebecca C. Wade:
"Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome", 7th EMBO Meeting, Mannheim, Germany, September 10 – 13, 2016.

Joanna Panecka, Ina Pöhner, Talia Zeppelin, Francesca Spyrakis, Maria Paola Costi, Rebecca C. Wade:
"Comparative mapping of on-targets and off-targets for the discovery of anti-trypanosomatid folate pathway inhibitors", 21st European Symposium on Quantitative Structure-Activity Relationship, Verona, Italy, September 4 – 8, 2016.

Martin Pippel:
"MARVEL – An efficient implementation regarding repeat saturated genomes", Genome Informatics 2016, Hinxton, UK, September 19 – 22, 2016.

Ina Pöhner, Joanna Panecka, Francesca Spyrakis, Talia Zeppelin, Maria Paola Costi and Rebecca C. Wade:
"Insights for anti-parasitic drug design – From comparative pocketome analysis to computational modeling of a parasitic folate & biopterin pathway", Meeting on Computational Modelling with COPASI, Manchester Institute of Biotechnology, Manchester (UK), May 12 – 13, 2016.

Ina Pöhner, Joanna Panecka, Talia Zeppelin, Francesca Spyrakis, Maria Paola Costi, Rebecca C. Wade:
"Comparative mapping of on-targets and off-targets for the discovery of anti-trypanosomatid folate pathway inhibitors", Helmholtz Drug Research conference on *"Drug Innovation in Academia"*, DKFZ Heidelberg, Germany, December 8 – 9, 2016.

Kai Polsterer, Fabian Gieseke:
"Probability Density Functions for Astronomy", ADASS 2016, Triest, Italy, October 16 – 20, 2016.

Maja Rey, Ulrike Wittig, Renate Kania, Andreas Weidemann, Wolfgang Müller:
"SABIO-RK database meets user requests", Biocuration 2016, Geneva, Switzerland, April 10 – 14, 2016.

Maja Rey, Andreas Weidemann, Renate Kania, Ron Henkel, Dagmar Waltemath, Steffen Klamt, Ursula Kummer, Wolfgang Müller:
"NBI-SysBio: de.NBI-node serving the Systems Biology Cycle", International de.NBI-Symposium *"Bioinformatics for Human Health & Disease"*, Heidelberg, Germany, November 7 – 9, 2016.

Maja Rey, Renate Kania, Andreas Weidemann, Wolfgang Müller:
"Kinetic data and data & models management for systems biology", de.NBI-SAB-Meeting, Berlin, Germany, November 28 – 29, 2016.

Thomas Sean Powell:
"MARVEL – A front-to-end assembler for noisy long-read sequencing", ISMB 2016, Orlando, FL, USA, July 8 – 12, 2016; Genome Informatics 2016; Hinxton, UK, September 19 – 22, 2016.

S. Kashif Sadiq:
"Computation of protein-ligand binding kinetics by combining Brownian dynamics, random accelerated molecular dynamics and unbiased MD simulations", Solvay Workshop: Bridging the gaps at the PCB interface – Multiscale modelling in physics chemistry and biology, International Solvay Institutes, Brussels, Belgium, April 19 – 21, 2016; Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, May 27 – 28, 2016; 3rd DKFZ / ZMBH Alliance Retreat, Kloster Schöntal, Germany, June 19 – 21, 2016.

Andreas Weidemann, Ulrike Wittig, Maja Rey, Renate Kania, Martin Golebiewski, Wolfgang Müller:
"SABIO-RK: kinetics on the move", 17th International Conference on Systems Biology 2016 (ICSB 2016), Barcelona, Spain, 16th-20th September 2016.

Andreas Weidemann, Martin Golebiewski, Olga Krebs, Stuart Owen, Quyen Nguyen, Natalie Stanford, Ulrike Wittig, Katy Wolstencroft, Rostyk Kuzyakiv, Finn Bacall, Norman Morrison, Jakub Straszewski, Caterina Barillari, Lars Malmstroem, Bernd Rinn, Jacky Snoep, Carole Goble, Wolfgang Müller:
"Data and Model Management for Distributed Systems Biology Projects", Big Data in Biology and Health – EMBL–Wellcome Genome Campus Conference, Heidelberg, Germany, September 25 – 27, 2016.

Rainer Weinberger:
Poster on *"Simulating galaxy formation – the role of supermassive black holes"*, ZAH's Astronomers Convention, Heidelberg, June 2016. Poster on *"Simulating galaxy formation with black hole driven thermal and kinetic feedback"*, Conference: Crossing the Rubicon: the fate of gas flows in galaxies, Santarcangelo di Romagna, Italy, September 2016.

Ulrike Wittig, Hadas Leonov, Lihua An, Nils Wötzel, Wolfgang Müller:
"Management and Analysis of Drug Discovery Data", Drug Innovation in Academia, Heidelberg, Germany, December 8 – 9, 2016.

Christopher Zapp:
"Steric Interaction of Molecular Motors and Passive Crosslinkers on Microtubules", Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, May 2 – 28, 2016.

Jolanta Zjupa:
Poster on *"The impact of outflows on the Ly-alpha forest in the Illustris simulations"*, From Wall to Web, Harnack-Haus conference venue of the Max Planck Society, Berlin, Germany, July 24 – 29, 2016.

Demos

Ron Henkel, Olga Krebs:
"Data and Model Management with FAIRDOM, Seek, and other tools", NormSys Conference & Tutorial *"Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences"*, Potsdam, Germany, December 5 – 6, 2016.

Stefan Richter, Antonia Stank, Daria Kokh, Mehmet Ali Öztürk:
"Sammelt Nährstoffe – das Brettspiel", Explore Science (www.explore-science.info), Luisenpark Mannheim, June 15 – 19, 2016.

Andreas Weidemann:
"SABIO-RK – Reaction Kinetics Database", COMBINE & de.NBI Tutorial *"Modelling and Simulation Tools in Systems Biology"*, Barcelona, Spain, September 16, 2016.

9.3 Memberships

Tilmann Gneiting:

Fellow, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading (UK); Affiliate Professor, Department of Statistics, University of Washington, Seattle (USA); Guest faculty member, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg (Germany); Faculty member, Research Training Group 1653, Spatial / Temporal *"Probabilistic Graphical Models and Applications in Image Analysis"*, Heidelberg University, Heidelberg (Germany); Faculty member, Research Training Group 1953, *"Statistical Modeling of Complex Systems and Processes: Advanced Nonparametric Methods"*, Heidelberg University and Mannheim University, Heidelberg and Mannheim (Germany); Institute of Mathematical Statistics (IMS) Council, ex officio member; Institute of Mathematical Statistics (IMS) representative, Committee of Presidents of Statistical Societies (COPSS) Awards Committee.

Martin Golebiewski:

Convenor (chair) of ISO / TC 276 Biotechnology working group 5 *"Data Processing and Integration"*, International Organization for Standardization (ISO); German delegate at the ISO technical committee 276 Biotechnology (ISO/TC 276), International Organization for Standardization (ISO); Chair of the national German working group *"Data Processing and Integration in Biotechnology"*, (NA 057-06-02-05 AK), German Institute for Standardization (DIN); Member of the national German standardization committee *("Nationaler Arbeitsausschuss")* NA 057–06 – 02 AA Biotechnology, German Institute for Standardization (DIN); Member of the board of coordinators of the COMBINE network (Computational Modeling in Biology network); Member of the Richtlinienausschuss (German committee for engineering standards) VDI 6320 *"Datenmanagement im Bereich Life Sciences"*, Association of German Engineers (VDI); Co-chair Joint Ad-hoc Group on Standardization of Genomic Information Compression and Storage between ISO/IEC JTC 1/SC 29/ WG 11 (MPEG) and ISO/TC 276/WG 5; Member of the management committee and co-leader Working Group 1 (Community / platform-building) of the European COST action CHARME (Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research).

Frauke Gräter:

Member of BIOMS (Heidelberg Center for Modeling and Simulation in the Biosciences) Steering Committee; Faculty member, Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, Germany; Associated faculty member, HGS MathComp Graduate School, University of Heidelberg, Germany; Faculty member, Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology (HBIGS), University of Heidelberg, Germany; Advisory Board member, Graduiertenkolleg NanoCell, University of Linz, Austria.

Wolfgang Müller:

Member of the Scientific Advisory Board of the BioModels Database; Member of the Richtlinienausschuss (German committee for engineering standards) VDI 6320 *"Datenmanagement im Bereich Life Sciences"*, Association of German Engineers (VDI): Deputy Chairman of SIG 4 (Infrastructure & data management), German Network for Bioinformatics Infrastructure (de.NBI); Board Member and Treasurer of FAIRDOM e.V.

Christoph Pfrommer:

Member of the Organizing Committee for the Heidelberg Joint Astronomical Colloquium.

Kai Polsterer:

Member of the IEEE Task Force on Mining Complex Astronomical Data; Member of the Standing Committee on Science Priorities of the International Virtual Observatory Alliance; Chair of the Knowledge Discovery in Databases Interest Group of the International Virtual Observatory Alliance.

Friedrich Röpke:

Advisory board, *"Sterne und Weltraum"*.

Kashif Sadiq:

Member of expert panel for life sciences for the European Extreme Data & Computing Initiative (EXDCI Project).

Volker Springel:

Member of the Interdisciplinary Center for Scientific Computing (IWR), Heidelberg, Germany; External Scientific Member of the Max-Planck-Institute for Astronomy, Heidelberg; Member of the Cosmological Simulation Working Group (CSWG) of the EUCLID mission of ESAM; Member of the Research Council of the Field of Focus *"Structure and pattern formation in the material world"* at Heidelberg University; Member of the Board of SFB 881 *"The Milky Way System"*; Member of the Scientific Advisory Board of the Gauss Centre for Supercomputing (GCS); Member of the International Advisory Board of the Institute for Computational Cosmology, Durham University, UK.

Alexandros Stamatakis:

Member of the steering committee of the Munich Supercomputing System HLRB at LRZ; Member of the Scientific Advisory Board of Elixir Greece; Member of the scientific advisory board of the Computational Biology Institute in Montpellier, France.

Michael Strube:

Research Training Group 1994, *"Adaptive Preparation of Information from Heterogeneous Sources (AIPHES)"*, TU Darmstadt / Heidelberg University / HITS; Scientific Advisory Board: Special Interest Group on Discourse and Dialogue (SIGdial).

Rebecca Wade:

Member of Scientific Advisory Council of the Leibniz-Institut für Molekulare Pharmakologie (FMP), Berlin-Buch; Member of Scientific Advisory Board of the Max Planck Institute of Biophysics, Frankfurt; Member: BIOMS Steering Committee, Heidelberg; Member at Heidelberg University of: CellNetworks Cluster of Excellence, HBIGS (Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology) faculty, HGS MathComp Graduate School faculty, Interdisciplinary Center for Scientific Computing (IWR), DKFZ-ZMBH Alliance of the German Cancer Research Center and the Center for Molecular Biology at Heidelberg University; Mentor, BioMedX, Heidelberg, *"Selective Kinase Inhibitors"* Team.

Anna Wienhard:

Fellow of the American Mathematical Society; Co-Speaker DFG Research Training Group 2229 *"Asymptotic invariants and limits of groups and spaces"*, Heidelberg – Karlsruhe; Editor of Annales Scientifique de l'École normale supérieure; 2011 Editor of Geometriae Dedicata; Editor Forum Mathematicum; Advisory Board of Springer Lecture Notes in Mathematics; Member Scientific Advisory Board of the Heidelberg Laureate Forum; Coordinator of the Upstream Mentoring Program at Heidelberg University (with M. Winckler); Member of the Kuratorium of the Internationales Wissenschaftsforum Heidelberg; Coordinator of Erasmus program in Mathematics, Heidelberg University; PI and member of the Network Excecutive Committee of the NSF funded Research network in the mathematical sciences *"Geometric structures and representation varieties"* (GEAR); Member Ahlfors-Bers Colloquium Organizing Board; Member of the Beirat *"Wissenschaftskommunikation.de"*.

9.4 Contributions to the Scientific Community

Program Committee Memberships

Martin Golebiewski:

CHARME Conference *"The CHARME of standardisation in life sciences"*, Warsaw (Poland), June 1 – 2, 2016; COMBINE 2016: 6th Computational Modeling in Biology Network Meeting, Newcastle upon Tyne (UK), September 19 – 24, 2016; NormSys Conference *"Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences"*, Potsdam (Germany), December 5 – 6, 2016.

Michael Strube:

Area Chair for the area Summarization, Generation, Discourse, and Dialogue at the Conference for Empirical Methods in Natural Language Processing, Austin, Texas, USA, November 2 – 6, 2016.

Editorial work

Tilmann Gneiting:

Editor-In-Chief, Annals of Applied Statistics.

Michael Strube:

Editorial Board: Dialogue & Discourse Journal.

Rebecca Wade:

Associate Editor: Journal of Molecular Recognition, PLOS Computational Biology; Section Editor: BMC Biophysics; Editorial Board: BBA General Subjects; Journal of Computer-aided Molecular Design; Biopolymers; Current Chemical Biology; Protein Engineering, Design and Selection; Computational Biology and Chemistry: Advances and Applications; Open Access Bioinformatics.

Organization Committee Membership (Chair)

Martin Golebiewski:

CHARME Conference *"The CHARME of standardisation in life sciences"*, Chair of the session *"Standardization Communities in the Life-Sciences"*, Warsaw, Poland, June 1 – 2, 2016; COMBINE 2016: 6th Computational Modeling in Biology Network Meeting, Chair of the session *"Emerging Standards"*, Newcastle upon Tyne, UK, September 19 – 24, 2016; NormSys Conference *"Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences"*, Chair of the session *"Industrial Needs"*, Potsdam, Germany, December 5 – 6, 2016.

Conference and Workshop Organization

Martin Golebiewski:

NormSys Round table talk about standardisation in the life sciences, Munich, Germany, April 6, 2016; COMBINE & de.NBI Tutorial *"Modelling and Simulation Tools in Systems Biology"*, International Conference on Systems Biology (ICSB 2016), Barcelona, Spain, September 16, 2016; NormSys Conference *"Future Needs – Today's Requirements: Building Bridges with Standards in the Life Sciences"*, Potsdam, Germany, December 5 – 6, 2016; Committee Meetings of ISO / TC 276 Biotechnology working group WG5 *"Data Processing and Integration"*: Paris, France, February 18 / 19, Washington DC, USA, May 9 – 13, Dublin, Ireland, October 24 – 28, 2016; Committee Meeting of the national German working group *"Data Processing and Integration in Biotechnology"* (NA 057-06-02-05 AK), German Institute for Standardization (DIN) Berlin, Germany, January 27, 2016.

Fabian Krüger:

Workshop on Advances in Economic Forecasting, Heidelberg (Germany), June 6, 2016.

Kai Polsterer:

IEEE SSCI special session on *"Mining the Sky: Knowledge Discovery in Big and Complex Astronomical Data Sets and Data Streams"* in Athens, Greece; Session of the KDDIG and a Focus Session at the International Virtual Observatory Alliance in Cape Town, South Africa; Mini-Symposium on: *"Improved Short Time Weather Prediction for Astronomical Observatories"*; The splinter on E-Science & Virtual Observatory at the Annual Meeting of the Astronomische Gesellschaft 2016 (co-organizer).

Alexandros Stamatakis:
Main organizer of Computational Molecular Evolution Summer School, HCMR, Heraklion, Greece, May 8 – 19 2016.

Anna Wienhard, Qiongling Li, Tengren Zhang:
Workshop on Surface Group Representations, Caltech, CA / USA, March 17 – 20, 2016.

Anna Wienhard, Andreas Ott, Jan Swoboda, Hartmut Weiss, Frederik Witt:
Workshop Higgs Bundles in Geometry and Physics, IWH Heidelberg, Germany, February 29 – March 3rd, 2016.

Ulrike Wittig, Renate Kania, Maja Rey, Martin Golebiewski, Ron Henkel, Olga Krebs, Andreas Weidemann, Wolfgang Müller:
Kinetics on the move: practical workshop on data for computational modeling, HITS Heidelberg, Germany), May 30 – 31, 2016.

9.5 Awards

Tilmann Gneiting:
"Highly Cited Researcher" in Mathematics for 2004 – 2014, Clarivate Analytics (formerly Thomson Reuters), 2016.

Michael Hoff, Stefan Orf, and Benedikt Riehm:
KIT prize from KIT for research work conducted by students in the framework of the winter term 2014 / 15 practical, Campus day of KIT, October 14, 2016.

Mehmet Ali Öztürk:
2016 Research Travel Award for the project *"Structural analysis of DNA binding mechanisms of FoxA1 – AR heterodimer complex"* from the Dr. Alexander und Dr. Rosemarie Bauer-Stiftung, Heidelberg University. Selected to represent Heidelberg University at the League of European Research Universities (LERU) 7th Doctoral Summer School *"Data Stewardship for Scientific Discovery and Innovation"* July 10 – 15, 2016, Leiden University, Netherlands.

Volker Springel:
Elected into the National Academy of Sciences Leopoldina, March 2016. Golden Spike Award of the High-Performance Computing Center Stuttgart, October 2016. *"Highly Cited Researcher"* in Space Science for 2004 – 2014, Clarivate Analytics (formerly Thomson Reuters), 2016.

Alexandros Stamatakis:
"Highly Cited Researcher" in Computer Science for 2004 – 2014, Clarivate Analytics (formerly Thomson Reuters), 2016.

Alexandros Stamatakis, Tomas Flouri:
Award for *"Teaching excellence"* from the dean of the computer science faculty at the Karlsruhe Institute of Technology for the programming practical *"Hands-on Bioinformatics Practical"* taught in summer 2015 based on the practical evaluation by the students, Karlsruhe, July 2016.

Rebecca Wade:
2016 International Society of Quantum Biology and Pharmacology (ISQBP) Award in Computational Biology, awarded June 20, 2016, Bergen, Norway.

The HITS Scientific Advisory Board.

Shareholders Board

HITS Stiftung

Prof. Dr.-Ing. Dr. h.c. Andreas Reuter
Member of the Board of Directors

Prof. Dr. Wilfried Juling
Member of the Board of Directors

Heidelberg University

Prof. Dr. Stephen Hashmi
Vice-President of Research and Structure

Karlsruhe Institute of Technology (KIT)

Dr. Hanns-Günther Mayer
Director of Shareholdings
("Leitung Beteiligungen")

Scientific Advisory Board

The HITS Scientific Advisory Board (SAB) is a group of internationally renowned scientists that supports the management of HITS in various aspects of running, planning, and directing the institute. The SAB is responsible for orchestrating the periodic evaluation of all the research groups of HITS. It presents the results to the HITS management and makes recommendations regarding how to further improve the institute's research performance. In 2016, the board consisted of ten members:

- Dr. Adele Goldberg, former President of the Association for Computing Machinery (ACM), USA (Vice Chair, SAB)
- Prof. Dr. Gert-Martin Greuel, University of Kaiserslautern, former Director of the "Mathematisches Forschungszentrum Oberwolfach", Germany
- Prof. Dr. Stefan Hell, Director at the Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
- Prof. Dr. Tony Hey, Chief Data Scientist Science and Technology Facilities Council, UK
- Prof. Dr. Masaru Kitsuregawa, University of Tokyo, Director General of the National Institute of Informatics, Japan
- Dr. Heribert Knorr, Head of Department at Ministry of Science, Research and the Arts Baden-Württemberg (retired), Germany
- Prof. Dr. Dieter Kranzlmüller, Ludwig Maximilians University, Munich, Director of the Leibniz Super Computing Center, Germany (Chair, SAB)
- Prof. Dr. Thomas Lengauer, Max-Planck-Institute for Computer Science, Saarbrücken, Germany
- Prof. Dr. Alex Szalay, Johns Hopkins University, USA
- Prof. Dr. Jeannette Wing, Carnegie-Mellon University, Corporate VP of Microsoft Research, USA

HITS Management

The HITS Management consists of the Managing Director and the Scientific Director ("Institutssprecher"). The latter is one of the group leaders appointed by the HITS shareholders for a period of two years. The scientific director represents the institute in all scientific matters vis-à-vis cooperation partners and the public.

Managing Directors:

Dr. Gesa Schönberger
(since January 2016)

Prof. Dr.-Ing. Dr. h.c. Andreas Reuter
(until April 2016)

Scientific Director:

Prof. Dr. Rebecca Wade
(2015 – 2016)

Deputy Scientific Director:

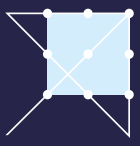
Prof. Dr. Michael Strube
(2015 – 2016)

Prof. Dr.-Ing. Dr. h.c. Andreas Reuter

On April 15, 2016, Andreas Reuter (photo: C. Flemming / HLF) stepped down as Managing Director of HITS. 18 years before, the computer scientist had given up his post as prorector of Stuttgart University to establish a private English-speaking university, at the same time accepting Klaus Tschira's offer to mastermind the development of the EML European Media Laboratory GmbH as its scientific and managing director. In 2003 he went on to found EML Research gGmbH, a basic research institute from which HITS originated in 2010. Andreas Reuter's extensive experience as a science manager has proven invaluable – particularly in the cooperation with KIT and Heidelberg University on various joint appointments and the decision by both universities to become shareholders of HITS in 2014. When Klaus Tschira passed away in March 2015, Andreas Reuter remained as the sole managing director for another year. Despite retirement as Managing Director, Andreas Reuter continues to be active in other institutions founded by Klaus Tschira, such as the HITS Stiftung, the Heidelberg Laureate Forum Foundation, and the EML European Media Laboratory GmbH. He is also a senior professor at Heidelberg University.

HITS

The Heidelberg Institute for Theoretical Studies (HITS) was established in 2010 by the physicist and SAP co-founder Klaus Tschira (1940 – 2015) and the Klaus Tschira Foundation as a private, non-profit research institute. HITS conducts basic research in the natural sciences, mathematics, and computer science, with a focus on the processing, structuring, and analyzing large amounts of complex data and the development of computational methods and software. The research fields range from molecular biology to astrophysics. The shareholders of HITS are the HITS Stiftung, Heidelberg University, and the Karlsruhe Institute of Technology (KIT). HITS also cooperates with other universities and research institutes and with industrial partners. The base funding of HITS is provided by the HITS Stiftung with funds received from the Klaus Tschira Foundation. The primary external funding agencies are the Federal Ministry of Education and Research (BMBF), the German Research Foundation (DFG), and the European Union.



Edited by
HITS gGmbH
Schloss-Wolfsbrunnenweg 35
D-69118 Heidelberg

Editor
Dr. Peter Saueressig
Head of Communications

Contact
info@h-its.org
Phone: +49 6221-533 262
Fax: +49 6221-533 298
www.h-its.org

Our e-mail addresses have the following structure:
Firstname.lastname@h-its.org

Pictures
HITS gGmbH *(unless otherwise indicated)*

Layout and Design
FEUERWASSER | grafik . web . design
www.feuerwasser.de

ISSN 1438-4159 | © 2017 HITS gGmbH

Twitter: @HITStudies
Facebook: /HITStudies
Youtube: /TheHITSters
LinkedIn: company/the-heidelberg-institute-for-theoretical-studies
Instagram: the_hitsters