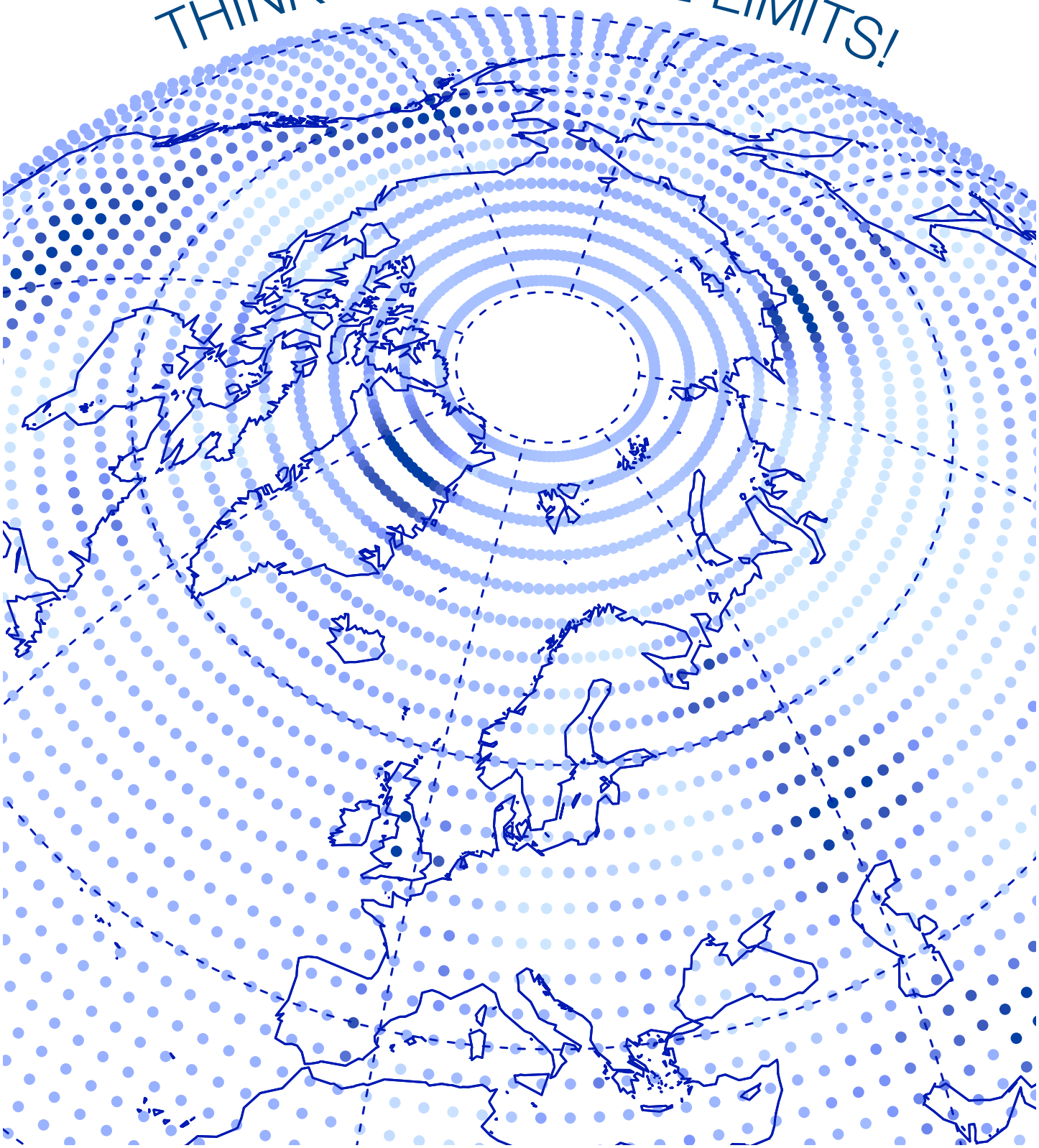
2017

Annual Report
Jahresbericht

Heidelberg Institute for
Theoretical Studies

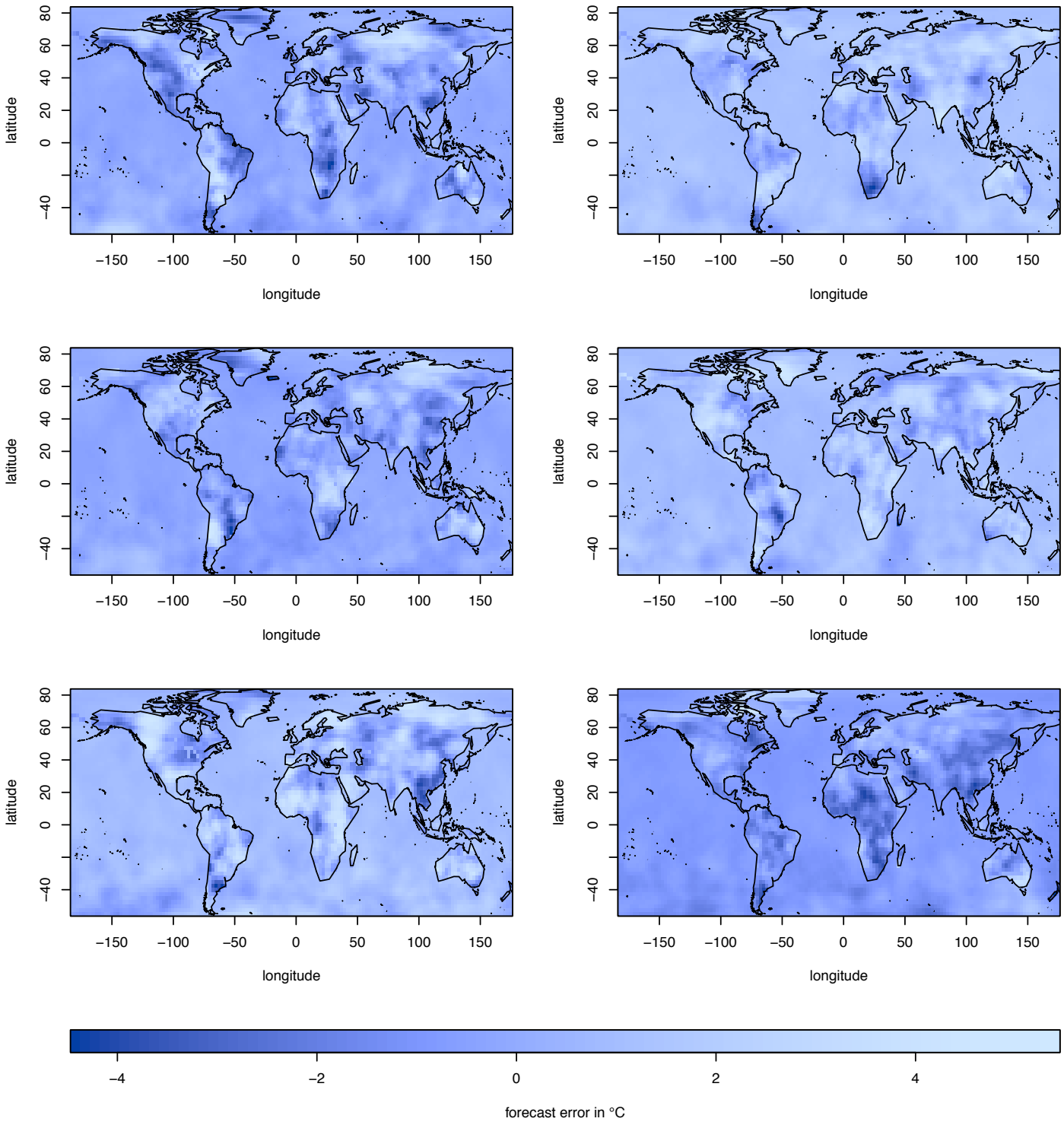
HITS

THINK BEYOND THE LIMITS!



Multiple random realizations of a spatial model for errors of temperature forecasts (cf. *Chapter 2.3*, pp. 26 ff).

Einige Simulationen von Zufallsfeldern zur räumlichen Modellierung der Fehler von statistisch nachbearbeiteten Temperaturvorhersagen (vgl. Kapitel 2.3, S. 26 ff).


latitude
80
60
40
20
0
−40
−150
−100
−50
0
50
100
150
longitude
−4
−2
0
2
4
forecast error in °C

Inhalt
Table of Contents

1
Think Beyond the Limits!

Prof. Dr. Michael Strube
(Scientific Director / Institutssprecher)

Dr. Gesa Schönberger
(Managing Director / Geschäftsführerin)

Basic research in the natural sciences, mathematics, and computer science were of fundamental importance to Klaus Tschira when he founded HITS gGmbH together with his foundation on January 1, 2010. The new research institute initially consisted of the four research groups incorporated from EML Research gGmbH: Molecular Biomechanics (MBM), Molecular and Cellular Modeling (MCM), Natural Language Processing (NLP), and Scientific Databases and Visualization (SDBV).

However, in keeping with Klaus Tschira's vision, HITS was intended to be larger and thematically broader in scope than its predecessor. In addition to its existing areas, new research groups focusing on astrophysics, statistical methods, and computer sciences were thus introduced, which HITS provided with a vast spatial and technical infrastructure. The goal then became to develop methods that enable computer-aided research – for example, with the help of simulations, visualizations, data analyses, or theoretical models. However, conducting research was merely one of several goals at HITS. According to its founder, the institute was also specifically designed to be a facility for scientific education. Therefore, HITS now closely collaborates with Heidelberg University and the KIT Karlsruhe, both of which are also affiliated with HITS as minority shareholders. Many group leaders at HITS are additionally university professors who accompany young individuals from across the globe throughout their scientific careers.

Scientists from about 20 nations currently work at HITS, including numerous guest researchers. These scientists are united by their academic curiosity and their computer-aided approach to investigating complex scientific issues. At HITS, these scientists can conduct research for their doctoral or post-doctoral programs at one of the two partner universities or at another university, and other scientists can complete work for their bachelor's

and master's thesis. HITS represents an important milestone in young individuals' scientific careers. In 2017, a total of 29 doctoral students worked at HITS, 6 of whom completed and defended their dissertations. In 2018, Rüdiger Pakmor from the group Theoretical Astrophysics (TAP) will become the second HITS scientist to successfully complete post-doctoral research at HITS, which was only possible through a close collaboration with both aforementioned universities (in addition to others). While young researchers at HITS mainly conduct research, they also gain experience at universities in teaching lectures, seminars, and other courses – an essential element of personal qualification.

HITS hosts many young scientists who – according to psychological research – hold great potential for innovation. Today's researchers generally work in groups, and due to the complexity of their tasks, they now work in increasingly close collaboration with other disciplines. This exchange is particularly encouraged by the vast scope of research topics available at HITS. If a HITS researcher needs to collaborate with a neighboring group, the result is often a discussion between an astrophysicist and a biologist or a linguist and a mathematician. Most individuals who participate in professional training are familiar with the feeling that there is never enough time to complete the necessary work and that the personnel frequently change. Although these changes are intentional, the skills needed in the groups must nevertheless be maintained. Therefore, each group can employ one permanent Staff Scientist in addition to the group leader. HITS allows promising young scientists who succeed in raising considerable funding to take an additional step forward in their career: They have the option of building their own group and can thereby begin to lay the foundation for a later position as a university professor. The most recent chapters of this success story belong to two astrophysicists: Christoph Pfrommer and Andreas Bauswein. In 2017, Pfrommer, who had earned an ERC Consolidator Grant and led his own group on the subject of High Energy Astrophysics and Cosmology (HAC) at HITS, accepted a professorship at the University of Potsdam as well as a group leader position at the Leibniz Institute for Astrophysics (see *Chapter 2.6*). Bauswein, a member of the Physics of Stellar Objects (PSO) group, received an ERC Starting Grant in October 2017 for his work on neutron star mergers (see *Chapter 9.5*).

However, it is not only junior scientists who come and go at HITS: Beginning on August 1, 2018, Volker Springel – head of the TAP group since 2010 – will leave the institute to assume a directorial position at the Max Planck Institute for Astrophysics in Garching, a position he had been offered in 2017. Moreover, in February 2018, the junior group Computational Biology (CBI) successfully completed its work after five years and was granted two publications in and a cover of the prestigious academic journal "Nature". The vision of HITS founder Klaus Tschira has been fulfilled: HITS has made a name for itself as a research institute internationally while simultaneously establishing itself as a sound scientific training institute. The necessary collaborations in both areas have had benefits for all parties involved, especially for the many young individuals who have found their way to HITS along their career paths, which has always been one of the institute's essential goals.

1
Think Beyond the Limits!

Die Grundlagenforschung in den Naturwissenschaften, der Mathematik und der Informatik steht im Vordergrund, als Klaus Tschira und seine Stiftung die HITS gGmbH zum 1. Januar 2010 gründen. In das neue Forschungsinstitut gehen die vier Forschungsgruppen der EML Research gGmbH ein: Molecular Biomechanics (MBM), Molecular and Cellular Modeling (MCM), Natural Language Processing (NLP) und Scientific Databases and Visualization (SDBV). Doch das HITS soll nach dem Willen des Stifters größer und thematisch breiter aufgestellt sein als sein Vorgänger. Neben den bestehenden Gebieten sollen sich neue Forschungsgruppen mit Astrophysik, statistischen Methoden und Computerwissenschaften befassen. Die räumliche und technische Infrastruktur dafür steht zur Verfügung. Der Auftrag lautet, Methoden zu erarbeiten, die rechnergestützte Forschung ermöglichen – zum Beispiel mit Hilfe von Simulationen, Visualisierungen, Datenanalysen oder theoretischen Modellen.

Doch Forschung ist nur einer von mehreren Aufträgen an das HITS. Nach dem Willen seines Gründers ist das Institut auch ganz gezielt als Einrichtung der wissenschaftlichen Ausbildung angelegt. Deshalb kooperiert HITS eng mit den Universitäten Heidelberg und KIT Karlsruhe, die dem Institut beide auch als Minderheits-Gesellschafter verbunden sind. Viele Gruppenleiter am HITS sind zugleich Professoren an einer Universität und begleiten junge Menschen aus aller Welt auf dem Weg in eine wissenschaftliche Karriere.

Aktuell arbeiten Wissenschaftler aus etwa 20 Nationen am Institut, darunter zahlreiche Gastwissenschaftler. Sie alle eint die Neugier der Forschenden und eine rechnergestützte Herangehensweise an komplexe Fragestellungen. Sie können hier forschen, um an einer der beiden Partneruniversitäten oder einer anderen Hochschule zu promovieren, sich zu habilitieren, Bachelor- und Masterarbeiten zu schreiben.

HITS will für junge Menschen eine wichtige Station in ihrer wissenschaftlichen Karriere sein. Im Jahr 2017 waren insgesamt 29 Doktoranden am HITS tätig, 6 haben ihre Dissertation abgeschlossen und verteidigt. Im Jahr 2018 wird mit Rüdiger Pakmor aus der Gruppe Theoretical Astrophysics (TAP) der zweite HITS-Wissenschaftler seine Habilitation erfolgreich abschließen. Dies ist nur möglich durch die enge Zusammenarbeit mit den beiden genannten sowie weiteren Universitäten. Während Nachwuchswissenschaftler am HITS ausschließlich forschen, halten sie an den Universitäten Vorträge, Seminare und Kurse – ein wesentliches Element der persönlichen Qualifizierung.

Am HITS arbeiten viele junge Wissenschaftlerinnen und Wissenschaftler. Das hat Vorteile, denn junge Leute haben nach Erkenntnissen der Psychologie das höchste Innovationspotenzial. In der Regel arbeiten Forscher heutzutage in Gruppen und aufgrund komplexer Aufgaben zunehmend in engem Austausch mit anderen Disziplinen. Das wird durch die Breite der vorhandenen Themen am HITS besonders gefördert. Wenn der

Dialog mit der Nachbargruppe gesucht wird, dann spricht die Astrophysikerin mit dem Biologen, der Linguist mit der Mathematikerin. Wer ausbildet, spürt immer auch den Nachteil, dass die zur Verfügung stehende Zeit begrenzt ist und der personelle Wechsel häufig. Obwohl diese Veränderungen gewollt sind, müssen die in den Gruppen benötigten Kompetenzen gleichwohl erhalten bleiben. Deshalb kann jede Gruppe zusätzlich zum Gruppenleiter einen festangestellten Staff Scientist beschäftigen.

Vielversprechenden Jungwissenschaftlern, denen es gelingt, nennenswerte Fördergelder einzuwerben, ermöglicht HITS einen weiteren Karriereschritt: Sie können hier eine eigene Gruppe aufbauen und damit den Grundstein für eine spätere Berufung auf eine Professur legen. Die jüngsten Kapitel dieser Erfolgsgeschichte schrieben zwei Astrophysiker: Christoph Pfrommer und Andreas Bauswein. Pfrommer, der einen ERC Consolidator Grant einwarb, startete seine eigene Gruppe High Energy Astrophysics and Cosmology (HAC) und nahm 2017 den Ruf auf eine Professur an der Universität Potsdam und eine damit verbundene Gruppenleiterstelle am dortigen Leibniz-Institut für Astrophysik an (siehe Kapitel 2.6). Andreas Bauswein, Mitarbeiter der Gruppe Physics of Stellar Objects (PSO), erhielt im Oktober 2017 für seine Arbeit zur Verschmelzung von Neutronensternen einen ERC Starting Grant (siehe Kapitel 9.5).

Doch nicht nur Nachwuchswissenschaftler kommen und gehen. Auch ein langjähriger HITS Gruppenleiter wird einem Ruf folgen, der ihm 2017 erteilt wurde. Zum 1. August 2018 wird Volker Springel, Leiter der TAP-Gruppe seit 2010, das Institut verlassen und eine Direktorenstelle am Max-Planck-Institut für Astrophysik in Garching annehmen. Und im Februar 2018 beendet die Juniorgruppe Computational Biology (CBI) nach fünf Jahren ihre Arbeit erfolgreich, mit zwei Publikationen und einem Titelbild im Fachjournal „Nature".

Die Vision seines Gründers ist aufgegangen: HITS hat sich als einzigartiges Forschungsinstitut international einen Namen gemacht und sich zugleich als wissenschaftlicher Ausbildungsbetrieb gut aufgestellt. Die dafür jeweils nötige Zusammenarbeit funktioniert gut. Sie schafft für alle Partner Vorteile – ganz besonders für die vielen jungen Menschen, die dem HITS auf ihrem Karriereweg zeitweise angehören. Das war und ist ein wesentliches Ziel.

2 Research

2.1 Astroinformatics (AIN)

2.1
Astroinformatics (AIN)

The Astroinformatics Junior 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, new, untapped wavelength regimes have yet to be investigated. Dedicated survey telescopes map the sky and constantly collect data. Our goal is to enable scientists to analyze this increasing amount of information in a less-biased manner.

The Astroinformatics Group is interested in the development of improved methods for time-series analysis and redshift models based on photometric measurements. These tools will be critical in the analysis of data in upcoming large survey projects, such as SKA, Gaia, LSST, and Euclid. Another scientific objective 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, a 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.

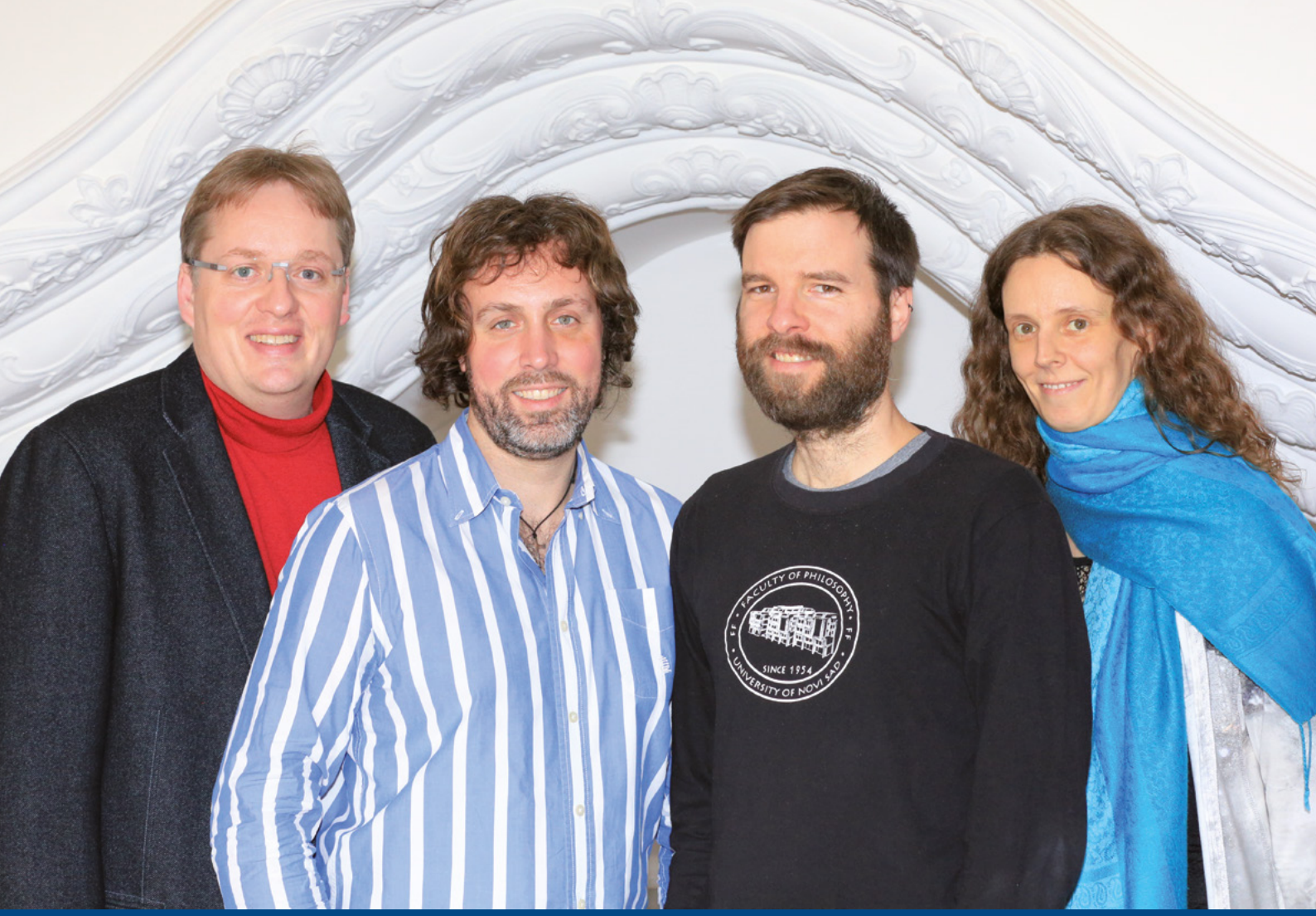
Moreover, we are engaged in multiple methods of conveying scientific topics to the public through initiatives such as participation in HITS events, public talks (see *Chapter 4*), and blog articles intended for a broader audience. For example, our scholarship holder Antonio D'Isanto wrote 32 articles for the Italian online magazine "Tom's Hardware Italia" in 2017.

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 gewonnenen 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.

Darüber hinaus engagieren wir uns in vielfältiger Weise, Wissenschaft in die Öffentlichkeit zu tragen, sei es durch die Teilnahme an Veranstaltungen des HITS, durch öffentliche Vorträge (siehe Kapitel 4) und Artikel für ein breites Publikum. So hat unser Doktorand Antonio D'Isanto im Jahr 2017 insgesamt 32 Artikel über astronomische Themen für das italienische Online-Magazin „Tom's Hardware Italia" geschrieben.



Group Leader

Dr. Kai Polsterer

Staff members

Dr. Nikos Gianniotis
Dr. Jennifer Wagner *(until December 2017)*

Scholarship holder

Antonio D'Isanto *(HITS Scholarship)*

Student assistant

Erica Hopkins *(since April 2017)*

Redshift estimation

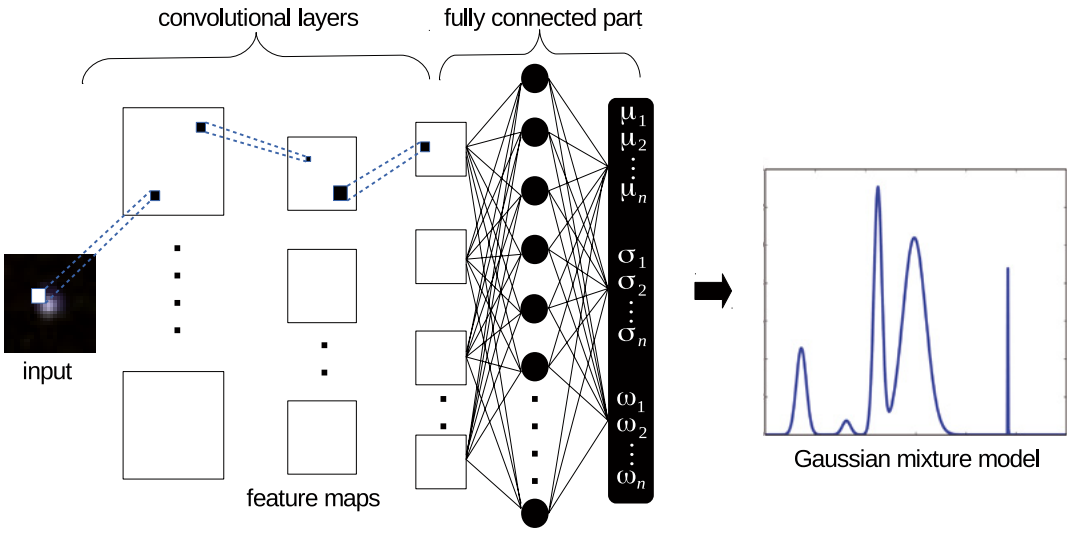


Figure 1: The architecture of the Deep Convolutional Mixture Density Network (DCMDN) used for the experiments. The convolutional and the fully connected (mixture density) network can be seen.

Deriving the distance of extragalactic objects via redshift analysis is an essential task in astronomy and is fundamental to cosmological research with a special interest in estimating the redshift of quasars. Quasars are super-massive black holes that accrete material and thereby create extreme luminosities. They are among the brightest objects in the Universe and can thus be observed over extreme distances, thereby providing insights into makeup of the early Universe. Measuring redshift via spectroscopic analysis is extremely demanding due to the necessity of long integration times and costly instrumentation requirements. In contrast to the expensive and time-consuming redshift determination via spectra, the use of photometry has become a popular and practical alternative. In order to process the exponentially increasing number of sources, the use of machine learning methods has become mandatory.

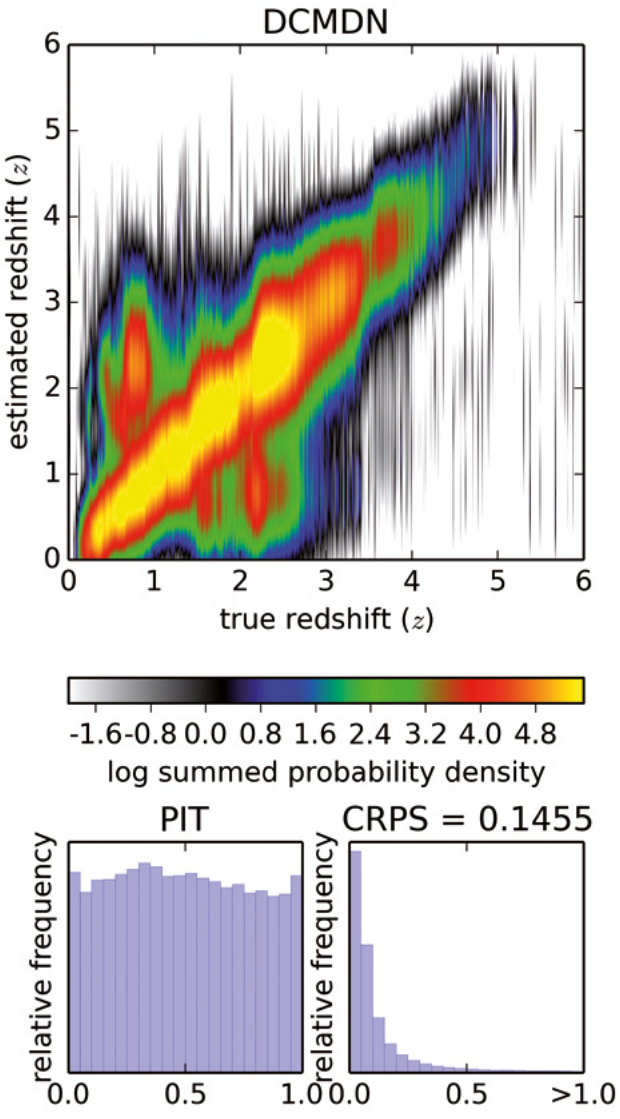
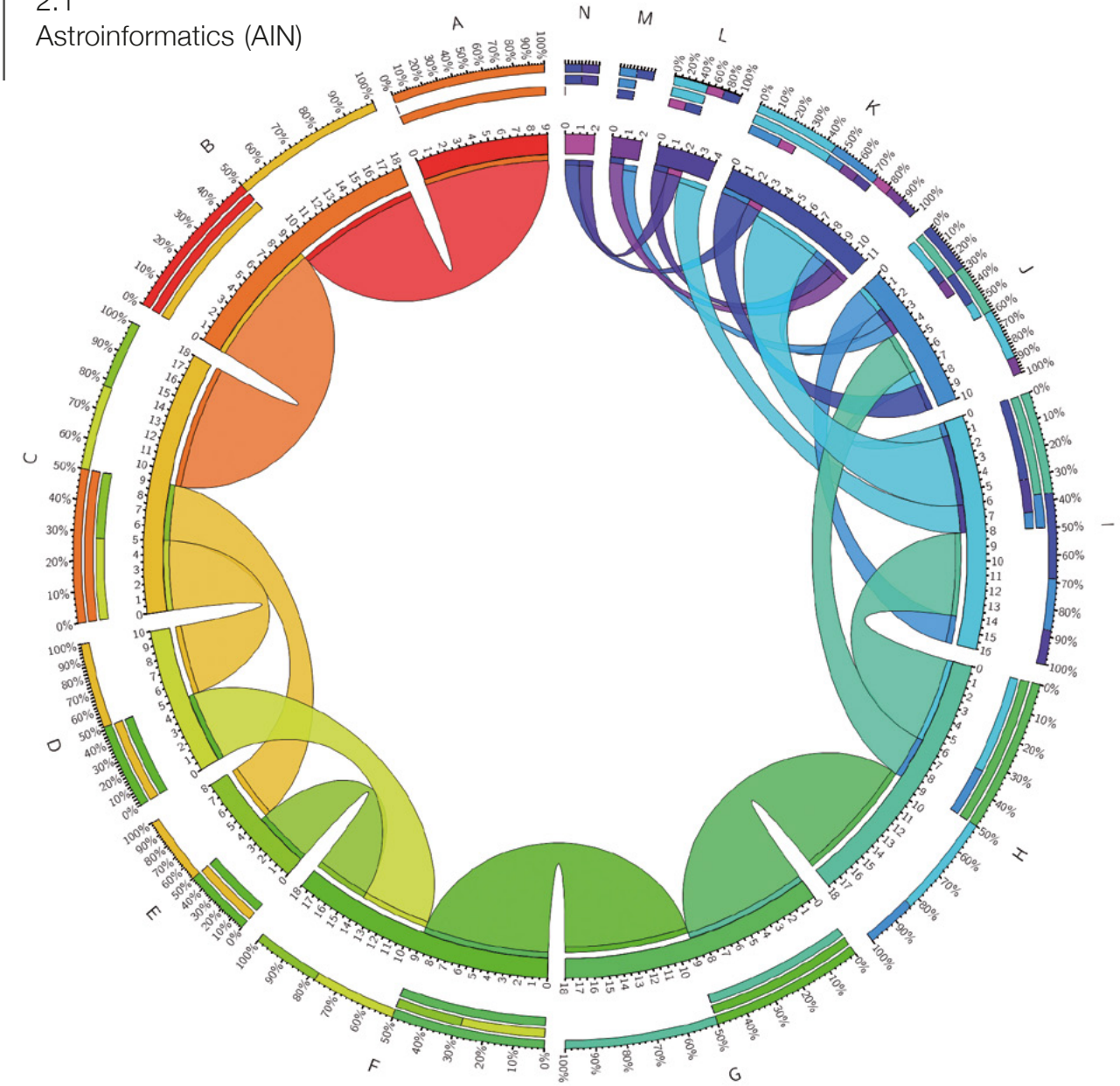


Figure 2: Density plot for the quasar experiment with the DCMDN. The plot displays the spectroscopic redshift against the summed distribution of the estimated probability density functions for the photometric redshift over every y axis. The PIT histogram and the CRPS are also shown.

Using manually selected features is the standard approach when training machine learning models; however, thanks to recent developments in GPU computing, it is now possible to explore new methods for an efficient photometric redshift estimation based on machine learning. In our recent works, two main methodologies have been analyzed. In the first case, we developed a model based on the application of deep learning technologies, to be more precise convolutional neural networks, which allow for the estimation of photometric redshifts directly from images in a fully probabilistic manner. The need for pre-classification and feature extraction is thereby rendered obsolete. In this framework, the process of feature extraction and selection is fully automatized. We developed a Deep Convolutional Mixture Density Network (DCMDN) and applied it to data from the Sloan Digital Sky Survey (SDSS). To be specific to catalogs with quasars, galaxies, and on a mixed one composed of the already-mentioned categories plus stars (used as contaminants). The DCMDN performed significantly better with respect to feature-based models in all cases.

On the other hand, feature-based methods remain important in gaining resources and computation time and because features are physically interpretable with respect to the feature maps generated by automatic methods. Moreover, the application of GPU computing also enables a massive feature selection to be performed that computes thousands of feature combinations and selects the best-performing ones through a greedy forward selection scheme. In our work, a framework was developed to perform an efficient selection from among the thousands of features obtained by combining all the photometric data for quasars in the Sloan Digital Sky Survey (SDSS). About 5,000 features were obtained by performing these combinations. A forward selection process was then applied to find the best-performing set of features using a k-Nearest Neighbor (kNN) model. In repeated experiments, a tree of the most probable features was generated. Finally, we performed experiments with a Random Forest (RF) model on the different branches of the tree in order to estimate the best set with an independent method.

The features used in nearly all publications in this field are plain magnitudes as well as simple colors. For SDSS, this process results in 10 features. The experiments have resulted superior performance with respect to the classical features adopted in the literature. Moreover, the selected features also perform better with respect to the DCMDN model when comparing the root mean square error (RMSE) and the continuous rank probability score (CRPS). The DCMDN remains better calibrated after the analysis of the PIT due to the loss function used to train the DCMDN (i.e., the CRPS itself) and thereby focuses the model to optimize calibration and sharpness. The structure of the feature tree reveals a comparable performance for all sets and that they are all equally better than the classic features. The selected features begin to outperform the classic ones from the 4th feature on, and the improvement becomes asymptotic from around the 9th to the 10th features. Furthermore, a very important role is given by ratios between different magnitudes, which are the dominant subgroup between the selected features.



color	name	feature
	A	i_{petro}/i_{psf}
	B	g_{psf}-u_{model}
	C	i_{exp}/r_{psf}
	D	$\sqrt{\sigma^2_{r_{model}} + \sigma^2_{r_{dev}}}$
	E	$\sqrt{\sigma^2_{g_{model}} + \sigma^2_{r_{dev}}}$
	F	r_{psf}/g_{exp}
	G	i_{psf}/z_{model}
	H	$i_{psf} - i_{dev}$
	I	r_{petro}/r_{psf}
	J	$z_{psf} - z_{model}$
	K	i_{psf}-r_{model}
	L	$r_{psf} - r_{petro}$
	M	z_{exp}/z_{psf}
	N	i_{psf}-i_{dev}

Figure 3: Chord diagram representing the structure of the feature tree. Every feature is associated with a specific color, and beginning with the first feature (A), it is possible to follow all the possible paths of the tree, which depict the different feature sets. The external arcs – from the inside to the outside – represent the occurrences of a particular feature, the number and type of connections exiting that feature, the number of connections entering that feature, and the total percentage of these two last categories.

In any case, the obtained results clearly indicate that it is not important to use a specific feature; instead, the global structure of the sets and the type of selected features play a major role in the improvement of the results. In order to inspect the structure of the tree and the relations between the different features, we adopted the chord diagram as a visualization tool. Using a color code and a structure based on ribbons connecting the features, this plot is able to visually represent the entire structure of all the trees in a single diagram. The role of the dominant features is evident upon inspecting the diagram, as are the relations among them.

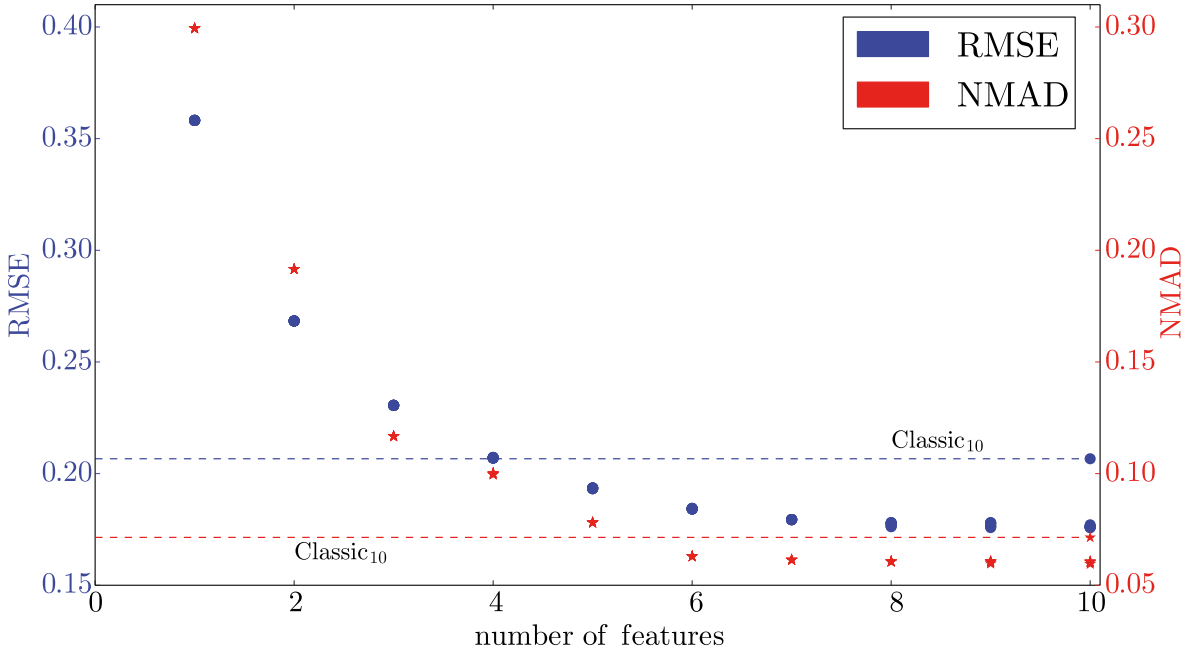


Figure 4: Results of the experiments with the RF for all sets of selected features with respect to the number of added features. The performance is expressed in terms of the RMSE and the NMAD. The performance of the classic features with respect to these indicators is shown for comparison.

After identifying the best feature set, the individual features were analyzed in order to understand their physical meaning. We found a clear correspondence between several of them and the emission lines of the Lyman-α series. This method was used on different quasar catalogs extracted from the SDSS's Data Release 7 and Data Release 9 in order to inspect the variation of the selected features in the presence of a different redshift distribution. From this point of view, the model displayed great stability and generated trees with a very similar structure and several common features. The final purpose of these works was to establish an efficient and affordable method of photometric redshift estimation based on features or automatized methods. In this sense, the two approaches are complementary. The tremendous effort in this field should serve as preparation for several upcoming projects, such as the EUCLID mission, LOFAR, PANSTARRS, and many others, whose success will mainly be based on the availability of highly affordable and precise photometric redshifts. The methods we have developed are vital to the success of these projects.

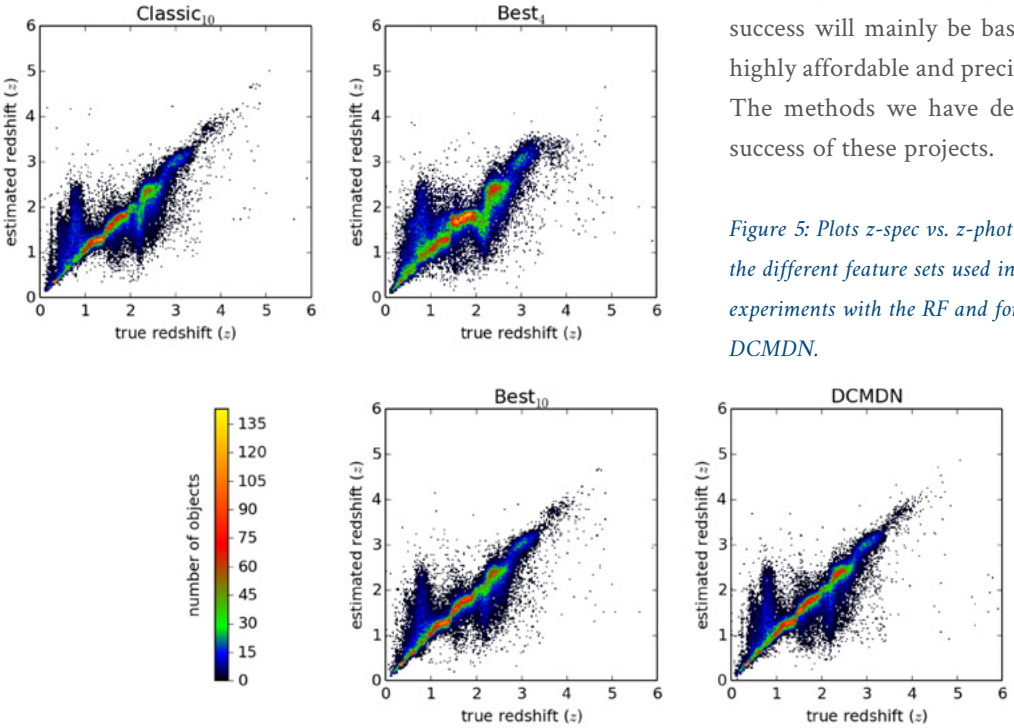


Figure 5: Plots z-spec vs. z-phot for the different feature sets used in the experiments with the RF and for the DCMDN.

Probabilistic principal component analysis for time series

The application of machine learning algorithms to time series requires that particular issues pertaining to the temporal nature of the time series be taken into account. For instance, the individual observed time series may be phase-shifted with respect to one another or have different lengths. Typically, phase-shift is of no consequence to us, and we would like to compare time series in a shift-invariant fashion. Fox example, when classifying time series, the observed length or phase shift should not have any influence on the assigned label.

While there are works that acknowledge the temporal nature of the time series in classification scenarios, little has been done for the unsupervised scenario. Generally speaking, it is quite common to avoid modeling the temporal nature of time series by transforming them into feature vectors (by extracting certain properties, such as moments, amplitudes, extrema, etc.) before feeding the time series as input into a machine learning algorithm. However, it is not entirely clear why such feature vectors are supposed to provide good characterizations of the underlying temporal dynamics. A further problem of the feature representation in an unsupervised context is that it is unclear what it means to subject the features to dimensionality reduction or clustering. In the case of principal component analysis (PCA) (which we concentrate on), the aim is to find a linear mapping to a lower-dimensional space that preserves the variance of the data. Preserving the variance of the feature vectors does not necessarily equate with preserving the dynamics that drive the time series. In the case of clustering (not dealt with by us, but similarly problematic), it is unclear what metric should drive the formation of the clusters as the inevitable distortion (i.e., quantization error) measure between prototypes and exemplars cannot be related back to the original dynamics.

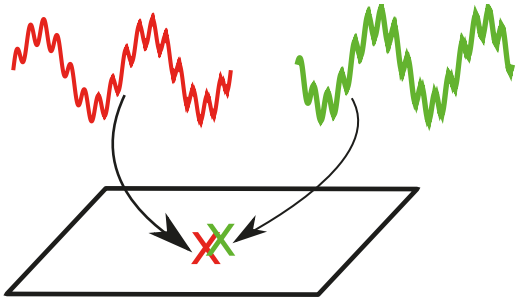


Figure 6: Stylized example: Two time series that differ only in phase. A phase-invariant dimensionality reduction should be invariant to this issue and project the two time series onto the same coordinates.

In this work, we have extended the ubiquitous Probabilistic Principal Component Analysis (PPCA) algorithm to the dimensionality reduction of time series. PPCA is a very important algorithm that has found applications in virtually all machine learning domains. PPCA formulates a noisy, affine map between low-dimensional latent variables and the observed data. We exploit the linear nature of PPCA (i.e., linear in parameters) in conjunction with the linear nature of Echo State Network (ESN). The ESN is a recurrent neural network that foregoes backpropagation through time by constructing its recurrent, hidden layer in a stochastic fashion. The stochastic construction is justified by empirical and theoretical findings that recurrent neural networks with small weights lead to the emergence of clusters in the recurrent activations

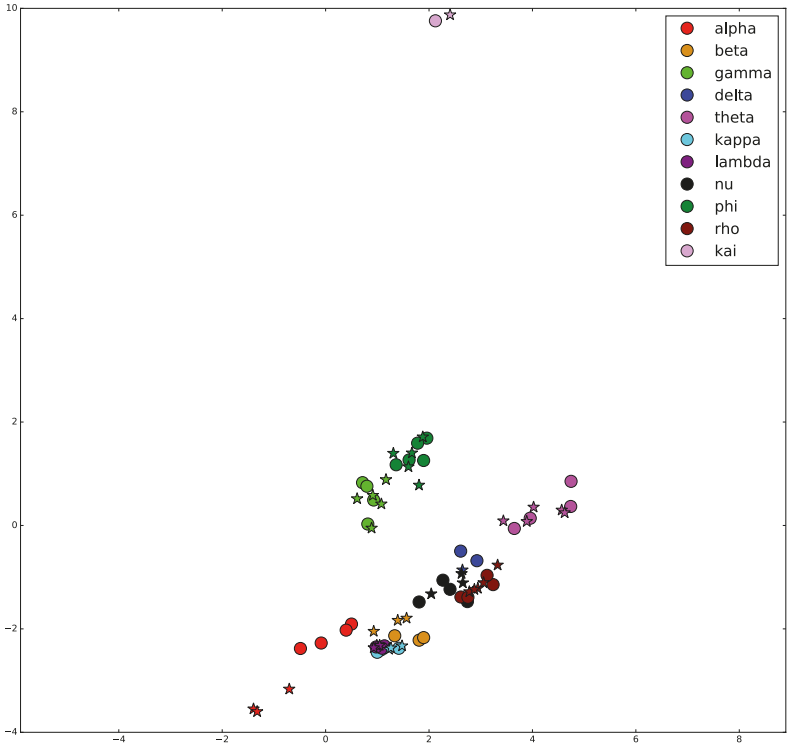


Figure 7: Example application on the X-ray binary time series. Dimensionality reduction by proposed scheme. We note that each projection corresponds to an entire light curve. The circles reveal projections of the training data, while the star markers display projections of testing data.

that correspond to Markov prediction contexts. The only free parameters in the ESN are its output weights (also referred to as readouts), which map the hidden activations to the predictions. Essentially, the operation of the ESN amounts to the recursive calculation of the hidden states followed by training the readouts so that they map the hidden states to the given predictions. This process also makes the ESN linear in its parameters and thus renders its combination with PPCA particularly appealing. However, as opposed to PPCA, which maps the latent variables directly to the data space, in the proposed model, we first affinely map the latent variables to ESN readouts. Hence, each latent variable addresses a recurrent network, and each recurrent network, in turn, is responsible for modeling a time series in the dataset. Thus, the mapping from latent variables to time series is indirect. The free parameters of the combined model include the latent low-dimensional variables as well as the parameters of the affine map that map the latent variables to readout weights. Model training can be formulated as an expectation-maximization algorithm in which the latent variables are marginalized in the expectation step to give rise to the marginal log-likelihood, which is optimized with respect to the parameters of the affine map.

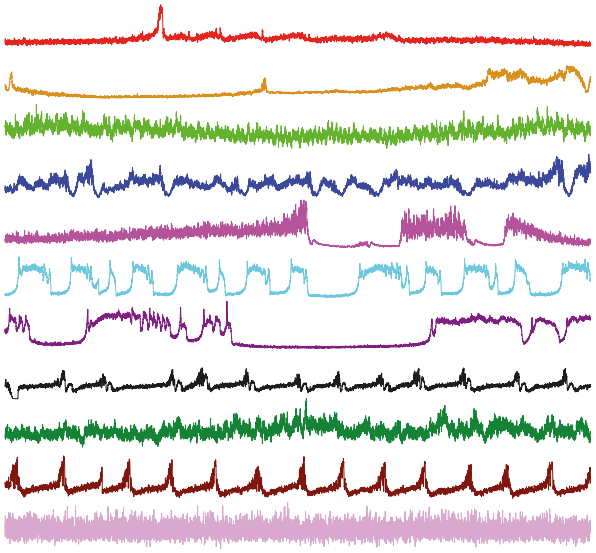
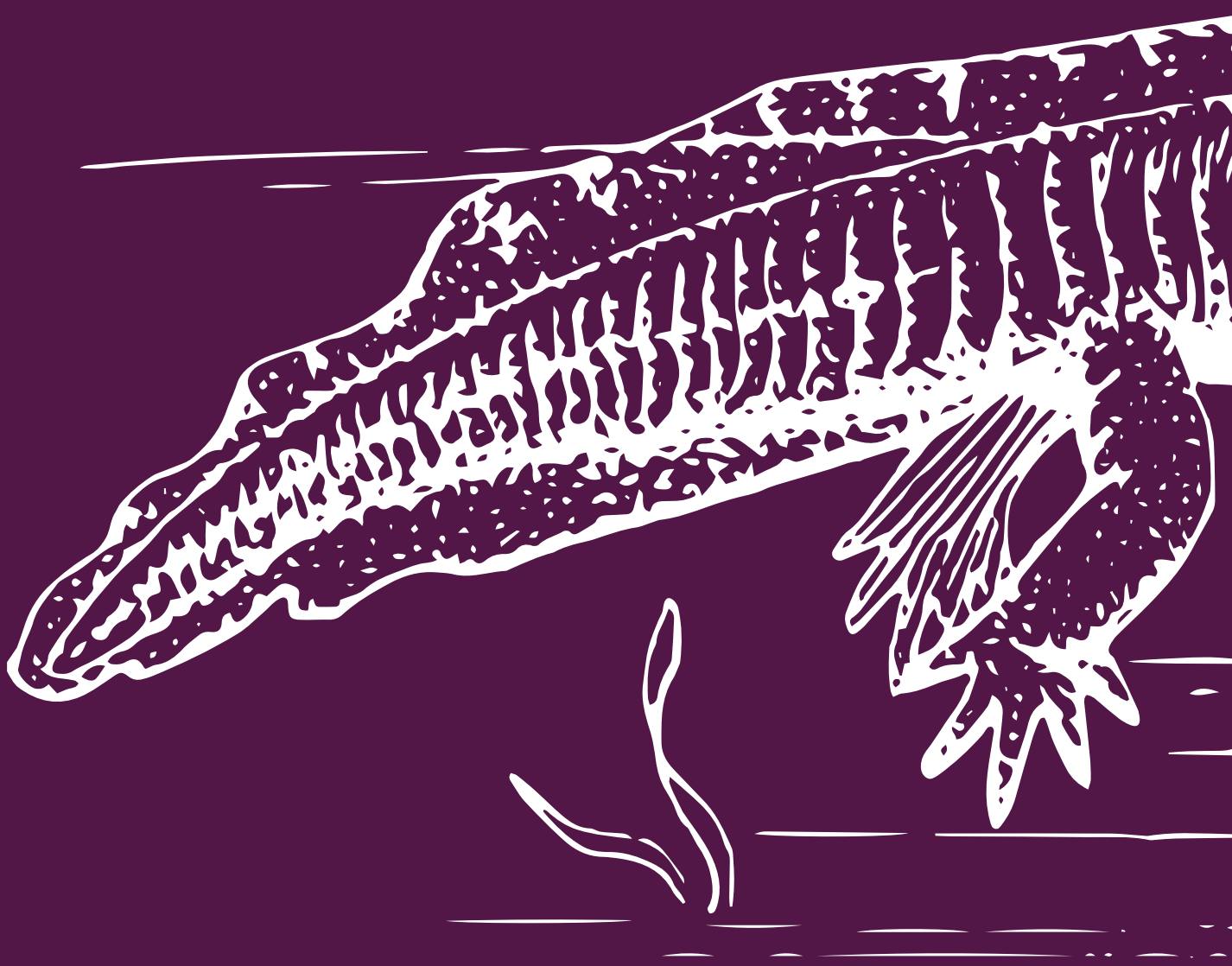
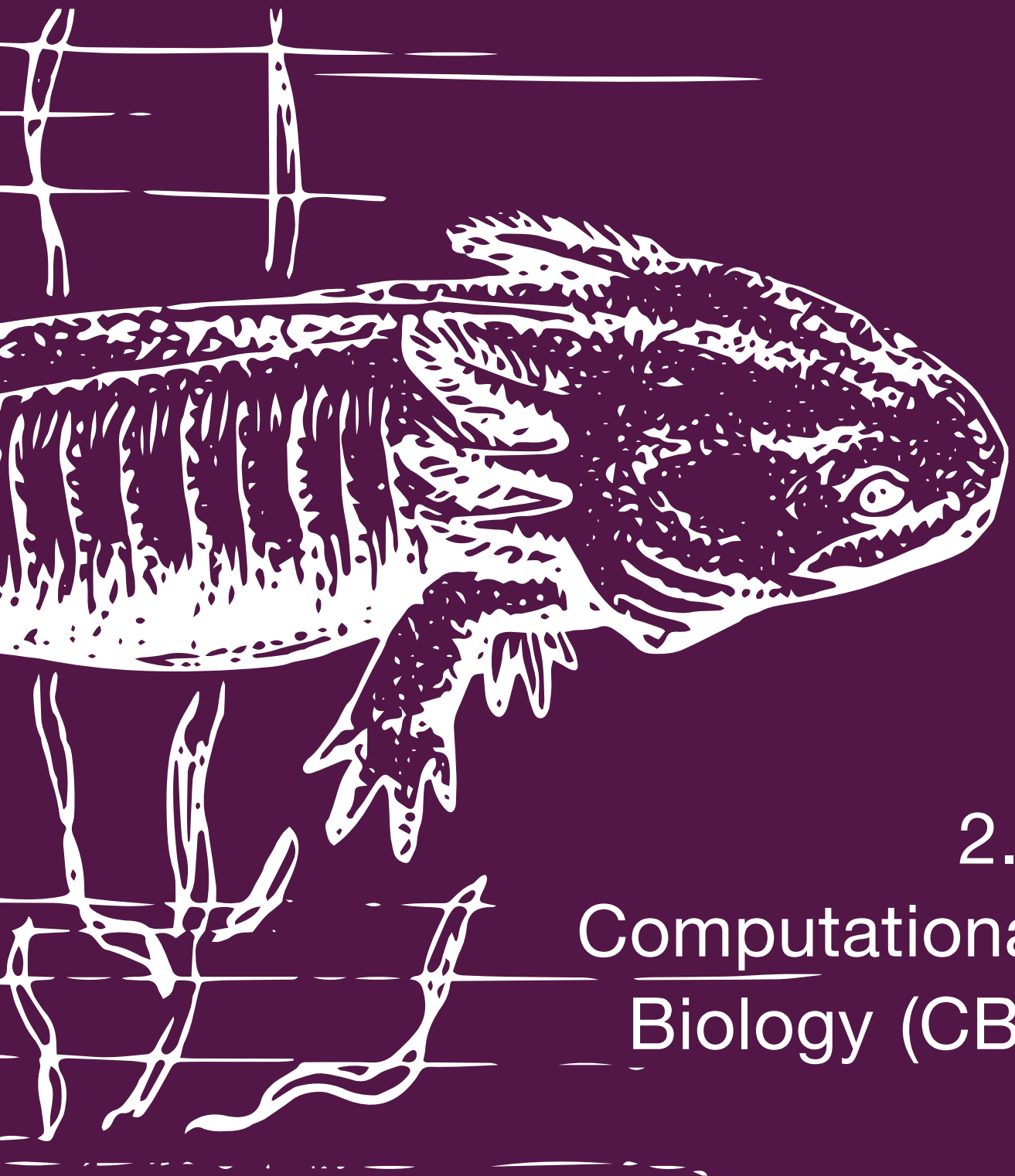


Figure 8: Instances of twelve distinct regimes of time series originating from the binary black hole system GRS1915+105.

As our proposed model is based on PPCA, it naturally inherits favorable properties of PPCA. For instance, it inherits the potential to be extended to mixtures if one ESN is not sufficient to capture the expressed dynamics in their entirety. The mixture approach enables the construction of hierarchical visualization by endowing the model with a tree structure in which parent nodes split the data space into (softly) disjoint parts and in which children nodes deliver a PPCA for each part. Furthermore, a Bayesian treatment may allow for the automatic determination of the number of latent dimensions as well as the automatic regularization of the weight matrix.

2 Research





2.2 Computational Biology (CBI)

2.2
Computational Biology (CBI)

Since 2013, the CBI group has been working at the interface(s) of 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 this field, we are especially interested in whole-genome assembly – that is, 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, albeit with high error rates, so 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 were the main goals of the group. After almost five years of intensive research, we managed to decipher the genome of the Mexican salamander Axolotl and the flatworm Schmidtea mediterranea together with colleagues from Dresden and Vienna. The results were published in "Nature". With this, our mission was completed, and the group's work came to an end in January 2018.

Seit 2013 arbeitete die CBI Gruppe an der Schnittstelle von Informatik, 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 der 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. Nach fast fünf Jahren intensiver Forschung gelang es uns, gemeinsam mit Kollegen aus Dresden und Wien, das Genom des mexikanischen Salamanders Axolotl und des Plattwurms Schmidtea mediterranea zu entschlüsseln. Die Ergebnisse konnten wir im Fachjournal „Nature" veröffentlichen. Damit haben wir unsere Mission erfüllt, und die Arbeit der Gruppe endet im Januar 2018.

Group Leader

Dr. Siegfried Schloissnig

Staff members

Martin Pippel
Thomas Sean Powell

Scholarship holders

Philipp Bongartz
Philipp Kämpfer

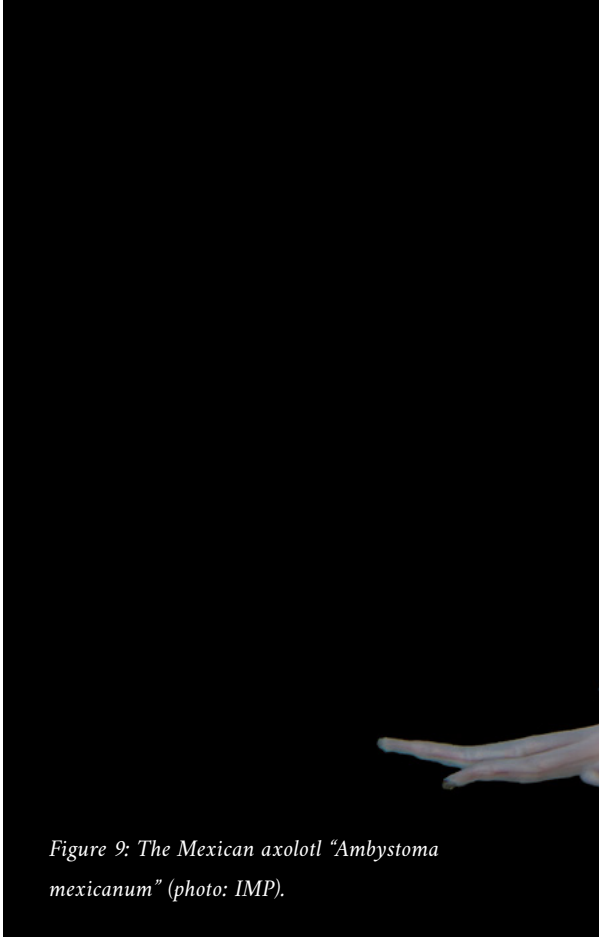


Figure 9: The Mexican axolotl "Ambystoma mexicanum" (photo: IMP).

The MARVEL assembler

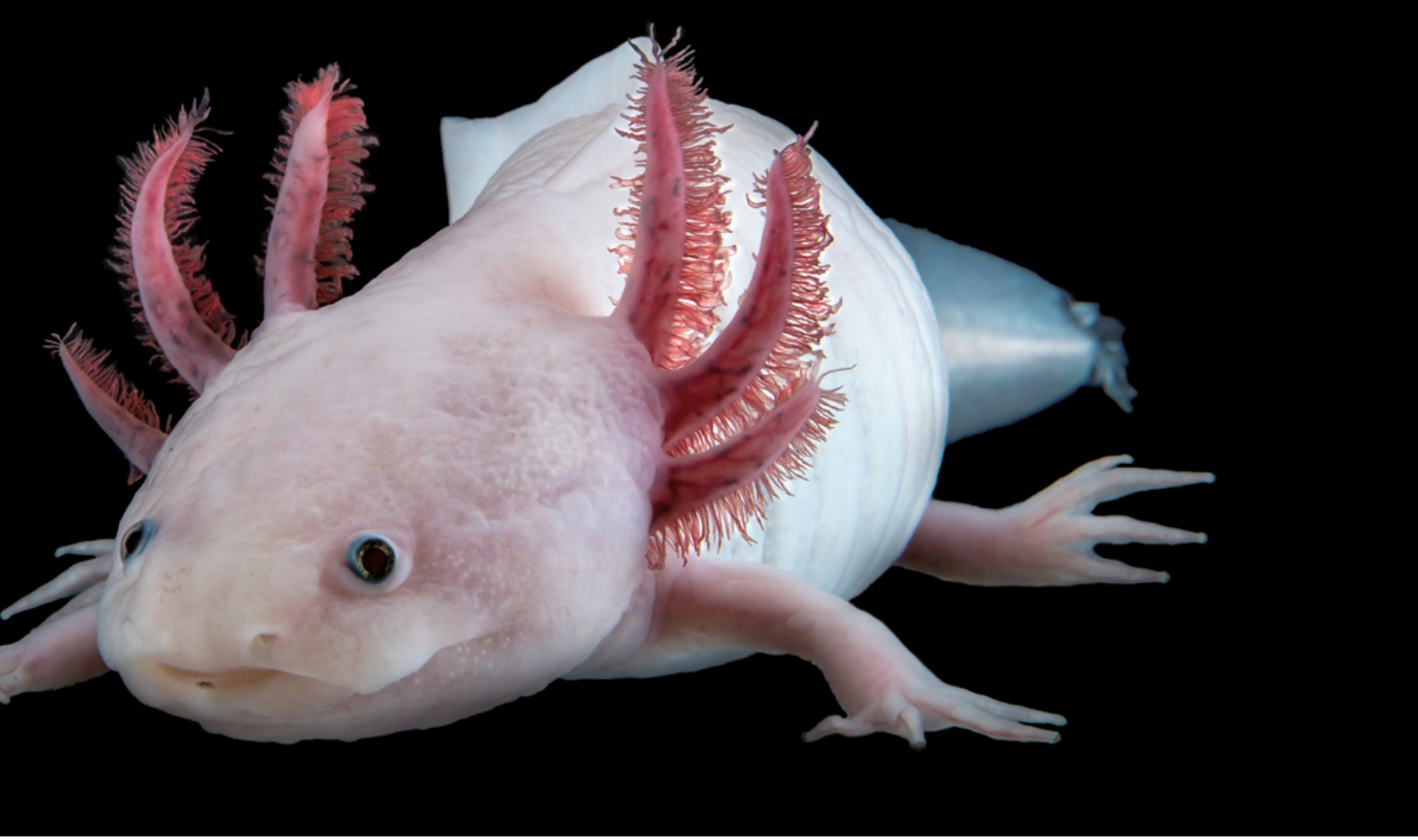
The long reads from single molecule sequencers (which can be upwards of 10k in length) are impressive, but with an error rate of 15 %, the resulting computational complexity of such reads is exuberant. However, truly random error positioning and near-Poisson single-molecule sampling imply that reference-quality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

We built an assembler we call MARVEL that has been used to assemble repetitive genomes of up to 32 Gb directly from a shotgun, long-read dataset. These genomes are currently only producible with the PacBio RS II sequencer. The assembler 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 developed a local alignment finder for noisy reads called Daligner.

One of the impediments to scaling the overlapper to large genomes lies in these genomes' repetitiveness. If every fragment of the genome were unique, then only true overlaps would be calculated; however, this is only partly the case, even for the simplest bacterial genomes. The repetitiveness that is ever-present to varying degrees induces partial matches (i.e., local alignments) of the reads that are responsible in large part 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 tagging occurs parallel to overlapping and results in the continuous exclusion of parts of a read from the overlapping process, which 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; rather, it gathers local alignment statistics from all overlaps jobs and distributes repeat interval annotations for these reads to all overlapper jobs.

Scrubbing detects and corrects read artifacts that would lead to partial alignments, which are subsequently 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 to a) detect and trim low-quality regions, b) repair regions in a read containing large numbers of insertions or add missing sequencing lost through polymerase jumps, and c) split reads at missed adapters and polymerase jumps.

The pipeline was developed with the common overlap-correct-overlap paradigm in mind, meaning we overlap the initial raw reads, scrub and correct them, and overlap the corrected reads again. However, we realized that the correction can actually be skipped and the assembly can be made directly from the raw reads after scrubbing, which necessitated a more thorough scrubbing phase that is good enough to guarantee not only a < 1 % error rate after correction but also a proper assembly. The move to uncorrected assembly was motived by the savings in computation time and by 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 others reads, the sequence used for the correction only needs to be good enough to allow the alignment to continue at the native error rate and doesn't actually have to



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

We have now successfully used the MARVEL assembler to assemble the Axolotl and Schmidtea mediterranea genomes. Both assemblies are now being used as the official reference genomes for both species. With the recent technological developments in DNA sequencing, we foresee a large demand for low-coverage assemblies that are subsequently scaffolded using Hi-C data. A proof-of-concept for this new approach to assembly is currently underway.

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 – 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.

An axolotl can reach a length of up to 30 cm (12 inches), lives up to 10 years, and is cannibalistically inclined. Due to its ability to regenerate, it is a model organism in regeneration and evolutionary research.

Its high regenerative capabilities enable the axolotl to regenerate entire body structures, such as its limbs, tail, and jaw, as an adult. Another characteristic of the axolotl is neoteny, meaning that it reaches sexual maturity without undergoing metamorphosis (i.e., it retains juvenile features during maturation). In addition to its regenerative abilities, the size of the genome, which is estimated at 32 gbp base and is distributed among 14 haploid chromosomes, makes this species an interesting as well as challenging one for de-novo genome assembly. Initial sequencing was based on BAC libraries, which aimed to construct an expressed sequence tag (EST) library. These approaches indicated that the genes of the axolotl are on average 10x larger than human genes 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

Figure 10: Regenerating planarian flatworm Schmidtea mediterranea. (Picture: Jochen Rink / MPI-CBG, Dresden).

consisted of a 28-fold random oversampling of this genome. We performed an assembly using MARVEL, which resulted in roughly 37 billion local alignment, took 350,000 CPU hours to compute, and revealed a repeat content of at least 60 %. The assembly contained 28.5 gbp, the longest contig was 3 mbp, and its N50 was 250 kbp. We then scaffolded the assembly using BioNano, which resulted in an N50 of 3 mbp in 125,000 contigs. Overall, those statistics represent an improvement of one order of magnitude in the basic assembly metrics over the next-largest sequenced genome (the white spruce) while and required significantly less CPU time to measure.

The assembly revealed that the size of the genome is largely due to a massive expansion of the LTR (long terminal repeat) family, which essentially "inflates" the genome and stretches coding and regulatory elements farther apart when compared with all other known genomes [Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AWC, Pippel M, Winkler S, Hastie AR, Young G, Roscito JG, Falcon F, Knapp D, Powell S, Cruz A, Cao H, Habermann B, Hiller M, Tanaka EM, Myers EW. The axolotl genome and the evolution of key tissue formation regulators. Nature volume 554, pages 50 – 55, February 1, 2018].

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, not only the chromosomes, but also the number of histone proteins has to be doubled to package the newly created DNA. This necessity 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 and even with the longer Sanger reads would have been unthinkable. However, with the advent of long-read sequencing machines, the assembly of this highly repetitive region may finally be possible. 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. We assembled the complete complex based

on further analyses of the (multi-)alignments, the detection of unique anchor points in the histone complex, clustering, and deep-learning approaches. We are now further investigating the possibility of performing the assembly of such complexes in a completely automated fashion, which has been proven rather difficult due to the noisy nature of the dataset and the often only subtle differences in the repeat copies. In order to effectively cope with these issues, we use deep-learning approaches to perform a targeted selection and correction of discriminating read features that are needed to automate the assembly process of repeat clusters.

Sequencing and assembly of Schmidtea mediterranea

Schmidtea mediterranea (S.med) is a free-living freshwater flatworm and member of the invertebrate Plathyhelminthes phylum. The full genome is estimated at 800 mbp to 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 costs of keeping the animals 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 is currently available. We are 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. Scrubbing and assembly was performed as previously explained and resulted in an initial assembly containing 726 mbp and with an N50 of 0.7 mbp [Grohme MA, Schloissnig S, Rozanski A, Pippel M, Young G, Winkler S, Brandl H, Henry I, Dahl A, Powell S, Hiller M, Myers EW, Rink JC. The genome of S. mediterranea and the evolution of cellular core mechanisms. Nature volume 554, pages 56 – 61, February 1, 2018].

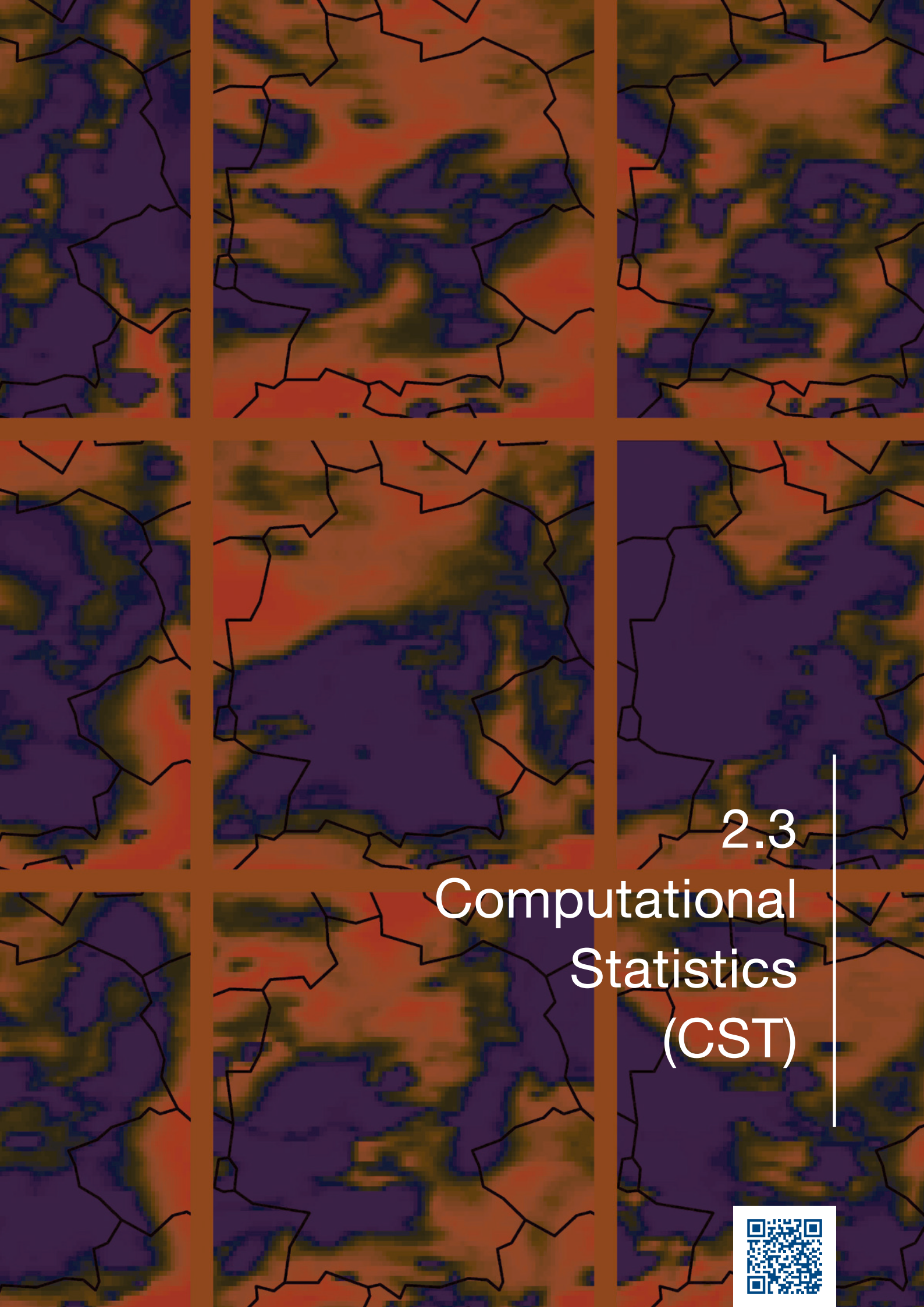
A hybrid-graph approach for short-read assembly

Currently, the reconstruction of genomic information using short-read data utilizes two distinct graph-based approaches, namely the overlap-layout-consensus concept (OLC) and de Bruijn graphs (dBG). The OLC concept is considered superior to the dBG in that the unit of assembly is a read as opposed to a small k-mer, which causes the graph and its path structure to be simpler and easier to disambiguate, thereby resulting in higher contig lengths. However, popular NGS assemblers rely solely on the de Bruijn graph due to its superior runtime efficiency compared with the quadratic nature of the OLC approach.

We developed a new hybrid graph approach that includes a variety of novel graph algorithms for fast and effective error correction. These algorithms eliminate more than 99% of all sequencing errors in linear time while taking advantage of both approaches. We combined the fast, linear-time construction of a dBG with the higher contig resolution of the OLC approach, which is accomplished by touring the dBG and collecting read information while simultaneously constructing an overlap graph directly from an expanded dBG that contains information on the original sequencing reads prior to their decomposition into k-mers. By further deriving a string graph from the transitively reduced overlap graph, it is possible to reconstruct large unique contigs of the genome. Finally, the paired-end read information is incorporated into the string graph to facilitate scaffolding-like contig ordering and the resolution of repetitive sequences. The hybrid-graph approach can be used with various insert sizes and sequencing technologies.

Initial tests on several bacterial short-read datasets have shown that the hybrid approach is comparable with the time and space complexity of classical dBG assemblers as well as with the high contig resolution of established OLC assemblers.

2 Research



2.3 Computational Statistics (CST)

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 is focused on the theory and practice of forecasting, and a secondary focus lies on spatial and spatio-temporal statistics and related areas.

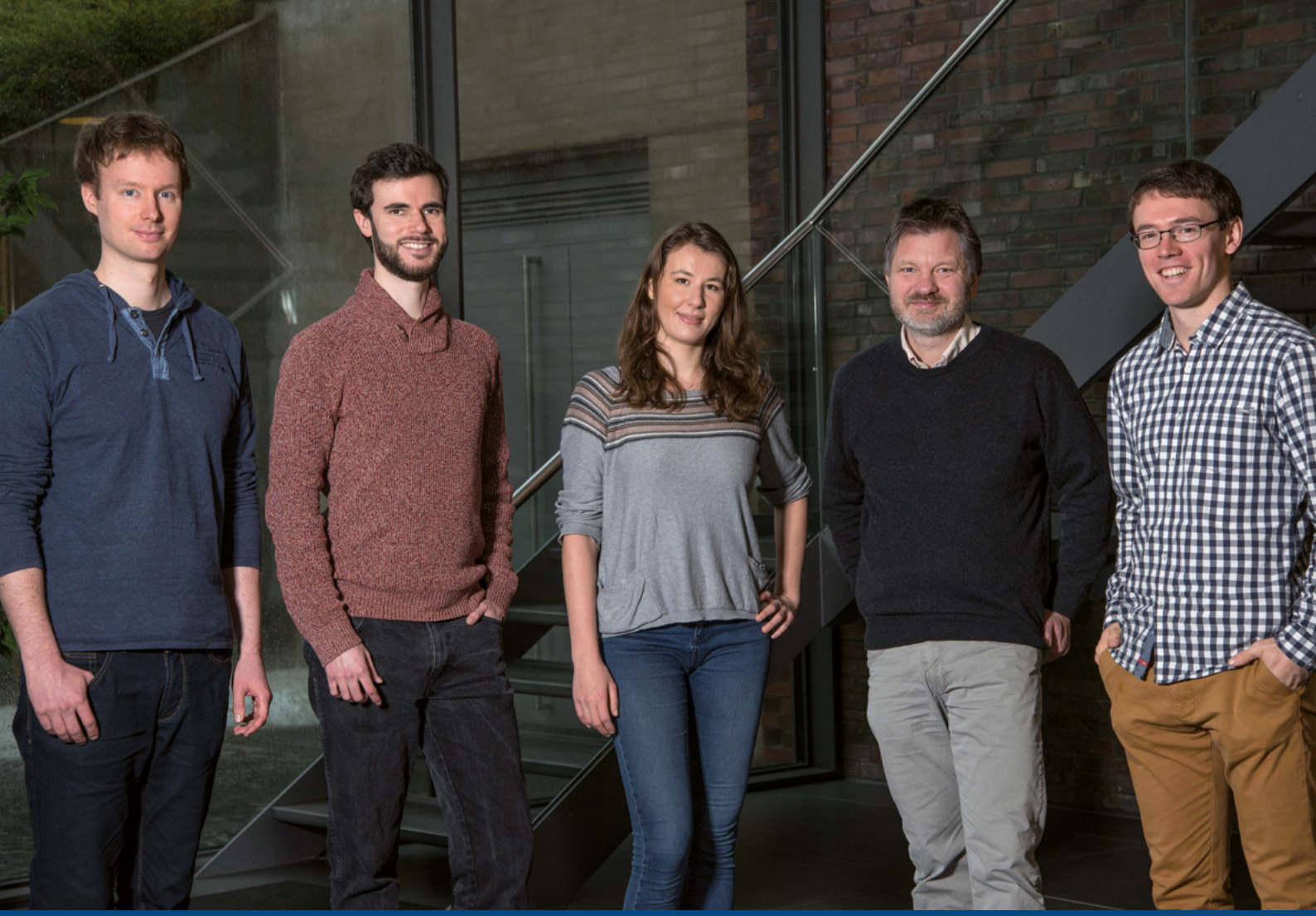
As the future is uncertain, forecasts should be probabilistic in nature, meaning they should take the form of probability distributions over future quantities or events. Accordingly, we are currently witnessing a trans-disciplinary shift of paradigms from deterministic or point forecasts to probabilistic forecasts. The CST group seeks to provide guidance and leadership in this transition by developing the theoretical foundations for the science of forecasting as well as 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 as well as other sources of external funding.

Weather forecasting represents a prime example of our work. In this context, the group maintains research contacts and collaborative relationships with national and international hydrological 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, d.h., 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 *(until September 2017)*
Dr. Alexander Jordan
Dr. Sebastian Lerch

Visiting scientists

Prof. Dr. Sándor Baran *(July 2017)*

Students

Maximiliane Graeter *(January 2017)*
Patrick Schmidt

General news

The focus of our research remains on the theory and practice of forecasting, with external funding provided by the European Research Council via the Advanced Grant ScienceFore and by the German Research Foundation within the Collaborative Research Group "Waves to Weather". Our interdisciplinary collaborations on forecasting problems include joint work with hydrologists at the German Federal Institute of Hydrology (BfG) in Koblenz and with meteorologists at our home university (the Karlsruhe Institute of Technology (KIT)) and at the European Centre for Medium-Range Weather Forecasts (ECMWF) in Reading in the United Kingdom.

In continuation of the cherished tradition, an integral aspect of our work is the intense scientific exchange with which we are involved on many occasions. A particular highlight in 2017 was the interdisciplinary "Summer School" on "The Science of Forecasting" held from October 3rd to 6th at the International Academic Forum in Heidelberg's Old Town. With funding provided by the ScienceFore project, we were happy to welcome about 40 young researchers from all over the world in disciplines ranging from mathematics and statistics to economics, meteorology, hydrology, and seismology. The invited speakers were Tom Hamill and Michael Scheuerer from the U.S. National Oceanic and Atmospheric Administration (NOAA, Boulder, Colorado), Barbara Rossi from Pompeu Fabra University (Barcelona, Spain), and Thordis Thorarinsdottir from the Norwegian Computing Center (Oslo, Norway). *Figure 11* shows the scientific organizing team, which includes not only our current team members but also CST alumnus Fabian Krüger, who is now an assistant professor in Heidelberg University's Department of Economics, as well as Peter Vogel from KIT. Further information about the summer school is provided in *Chapter 5.1.5* of this report.

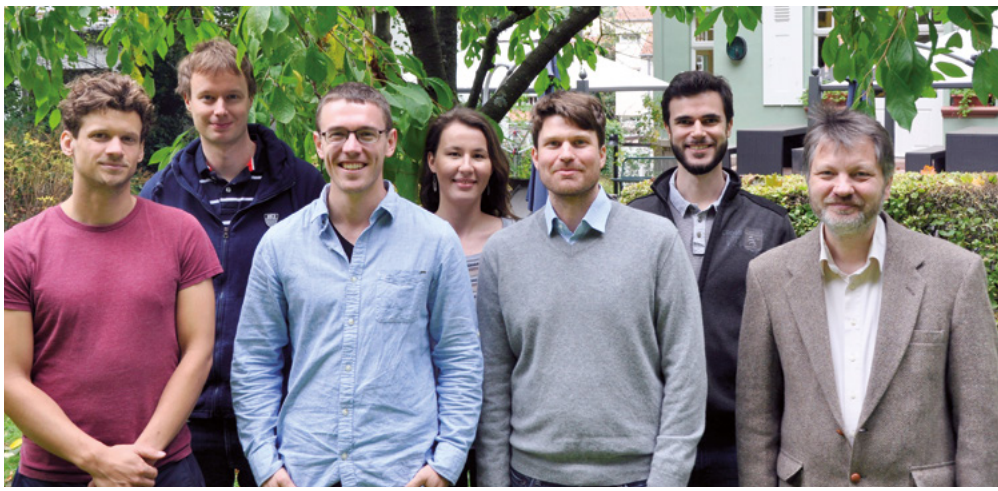


Figure 11: The scientific organizing team for the ScienceFore Summer School, including (from left to right) Patrick Schmidt, Sebastian Lerch, Peter Vogel, Kira Feldmann, Fabian Krüger, Alexander Jordan, and Tilmann Gneiting.

Throughout the year, we were happy to welcome guests from around the world who worked in a diverse range of scientific disciplines. Werner Stuetzle from the University of Washington in Seattle visited in January, and Sándor Baran from Debrecen University in Hungary spent the month of July in our group for joint research on probabilistic weather forecasting. In January, we held a mini-symposium on "Spatial Statistics" with students and colleagues at KIT, Heidelberg University, and Mannheim University.

The subsequent sections describe various facets of our research. We begin by reviewing the proper scoring paradigm in forecast evaluation and continue with our work on the statistical postprocessing of hydrologic ensemble forecasts. In the final section, we turn to research on geometric-optical illusions, with links to an ongoing exhibit at the Carl Bosch Museum in the immediate vicinity of the HITS premises.

Principles of forecast evaluation

Throughout history, humanity has shown a desire to predict the future, which is inherently uncertain and deterministically unpredictable. Therefore, forecasts ought to be probabilistic in nature and take the form of probability distributions over future quantities or events. To provide an example, *Figure 12* displays the projection of future inflation in the United Kingdom issued by the Bank of England in February 2018 as measured by the consumer price index (CPI). The forecasts take the form of probability distributions and are visualized by plotting prediction intervals. The most central band in the fan chart depicts a pointwise 30 % prediction interval. The pairs of the lighter red areas extend by 30 % each such that the entire fan corresponds to a 90 % interval, which is expected to cover the actual inflation rate 9 out of 10 times on average.

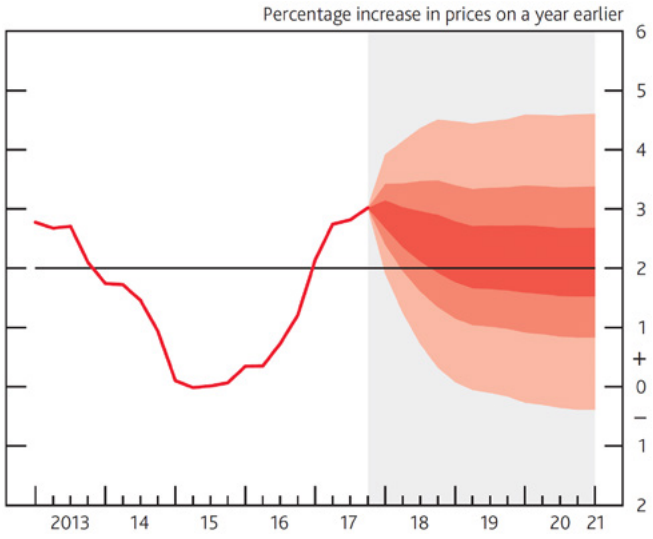


Figure 12: Bank of England projection of United Kingdom consumer price index (CPI) inflation in percent, as issued in February 2018. The shaded bands frame prediction intervals in increments of 30 %. Reproduced from the Bank of England's February 2018 Inflation Report with kind permission.

It is now generally recognized that the goal of probabilistic forecasting is to maximize the sharpness of the predictive distributions subject to calibration. The notion of calibration refers to the statistical compatibility between the probabilistic forecasts and the observations. Essentially, a forecast is calibrated if the realizing observations are statistically indistinguishable from random numbers drawn from the forecast distributions. Sharpness refers to the concentration of the predictive distributions and is thereby only a property of the forecasts.

In the context of prediction intervals, this goal can be phrased very simply: The shorter the intervals are, the sharper they are, and the sharper they are, the better they are (assuming that the nominal coverage is attained in practice). A schematic illustration is provided in *Figure 13*, where both the "blue" and the "red" forecaster issue prediction intervals on ten occasions. For both forecasters, the intervals capture the realizing observation eight times out of ten. In terms of calibration, they thus perform equally. However, the "blue" forecast is much sharper as it is updated at each new time step, whereas the "red" forecast is issued one single time without being updated again as new information becomes available.

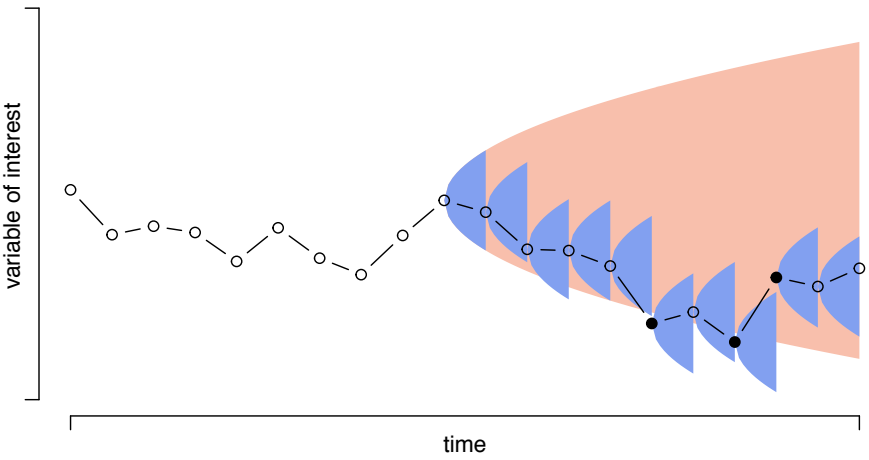


Figure 13: Schematic illustration of calibration and sharpness for interval forecasts. While the "blue" and the "red" forecasts perform equally in terms of calibration, the "blue" forecast is much sharper and therefore preferable.

Scoring rules are omnibus performance measures for probabilistic forecasts that simultaneously address calibration and sharpness. In a nutshell, a scoring rule assigns a numerical penalty based on the forecast and the event or value that materializes. A critically important requirement for a scoring rule is that it be "proper" in the technical sense that a forecaster should minimize the expected penalty by issuing a forecast that agrees with his or her best judgment [Gneiting T, Raftery AE. Strictly proper scoring rules, prediction and estimation. Journal of the American Statistical Association (2007) 102:359-378]. Proper scoring rules have therefore also been claimed to provide a "truth serum", and we speak of this guiding principle as the "proper scoring paradigm". Proper scoring tools serve as major work horses in the scientific study of forecasting problems, and the construction of proper scoring rules for a given purpose continues to pose challenges in our work [Hemri and Klein, 2017; Lerch et al., 2017].

Fortunately, in the above-mentioned context of prediction intervals, a very simple and intuitively appealing scoring rule turns out to be proper. Specifically, suppose that a central prediction interval with nominal coverage α is sought. In this case, the following scoring rule would be proper: Take the length of the prediction interval. If the realizing observation falls into the interval, the score or penalty incurred is simply this length; otherwise, the score or penalty is obtained as the sum of the length and $2/(1-\alpha)$ times the distance from the observation to the prediction interval.

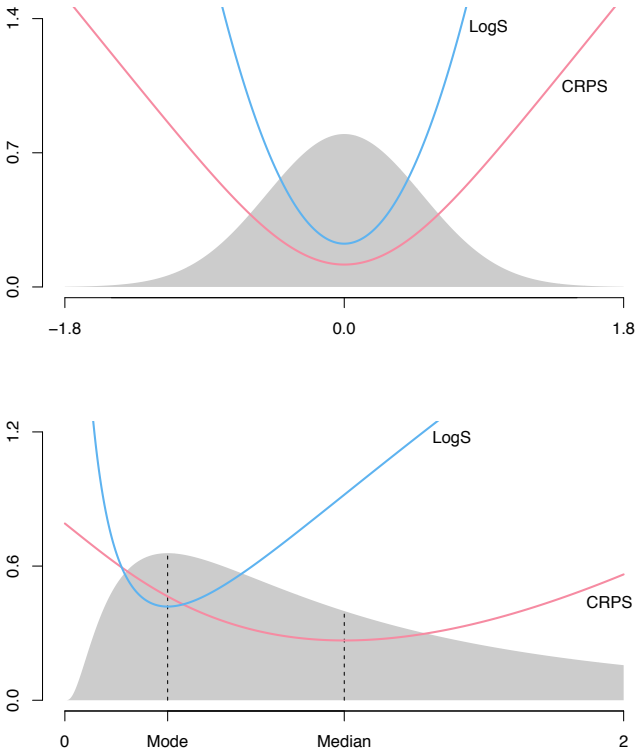


Figure 14: Continuous ranked probability score (CRPS) and logarithmic score (LogS) when the probabilistic forecast is Gaussian (top panel) or skewed (bottom panel).

Let us now return to probabilistic forecasts in the form of full probability distributions. *Figure 14* illustrates the two most popular proper scoring rules in this setting, namely the continuous ranked probability score (CRPS) and the logarithmic score (LogS). In the top panel, the probabilistic forecast is Gaussian, and we see that while both the CRPS and the LogS incur the smallest score or penalty at the center of the distribution, the LogS grows quadratically, whereas the CRPS only grows linearly as the realizing observation moves away from the center of the predictive distribution. In the bottom panel, the probabilistic forecast is skewed, and we see that the LogS incurs the smallest penalty at the mode of the predictive distribution, whereas the CRPS is lowest at its median.

A natural question is whether the proper scoring paradigm also applies to deterministic or point forecasts. Perhaps surprisingly, the answer is yes, namely via the notion of a "consistent scoring function". Specifically, if we are given a loss or scoring function for a point forecast, such as the absolute error (AE) or the squared error (SE), we can find the respective Bayes rule (i.e., the point forecast that minimizes the expected loss or score). For example, the Bayes rule under the AE is the median, and the Bayes rule under the SE is the mean or expectation of the predictive distribution. If we use the Bayes rule as point forecast, both the AE and the SE can be interpreted as proper scoring rules, as illustrated in *Figure 15*. This construction principle applies in full generality, and ongoing work in our group contributes to the theory and application of consistent soring functions. Expected and mean scores admit decompositions into calibration and sharpness components, and finite-sample biases in the respective terms can be efficiently corrected for [Ehm and Ovcharov (2017)].

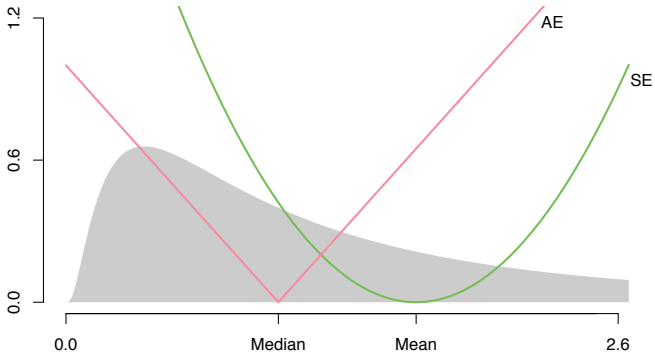


Figure 15: Absolute error (AE) and squared error (SE) for a point forecast in the form of the median and mean of the associated probabilistic forecast, respectively.

Statistical post-processing of hydrologic ensemble forecasts

In order to provide forecast information for water levels in major rivers, such as the Rhine, Danube, or Elbe in Germany, hydrologists draw on highly sophisticated numerical models that are run forward in real time. 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 100 – of the hydrologic model, with weather forecast information provided by output from distinct versions of numerical weather prediction models operated by national and international meteorological centers. The rationale underlying this approach is that by accounting for the uncertainty in the weather forecast information, the hydrologic ensemble output provides a means of assessing the uncertainty in the forecasts of water levels. Despite their undisputed successes, hydrologic 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). We have therefore joined forces with colleagues at the German Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde: BfG) in order to statistically post-process hydrologic ensemble forecasts. State-of-the-art techniques for this process 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 or normal family, for 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.

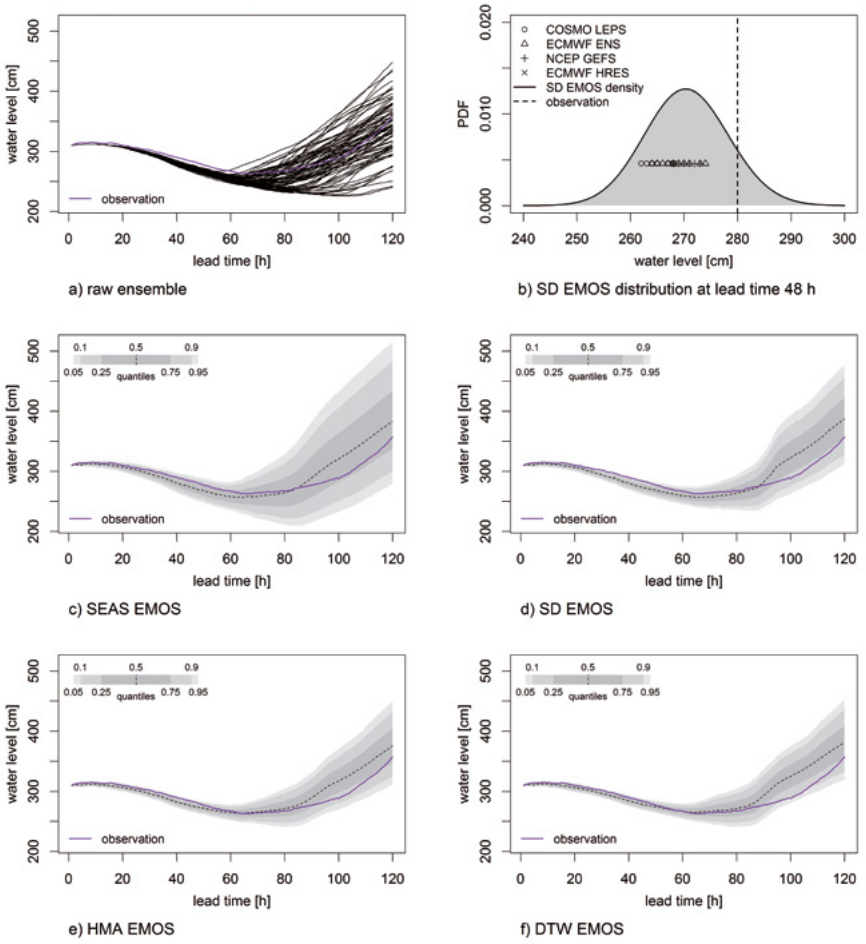


Figure 16: Raw ensemble and corresponding statistically post-processed EMOS forecast for water level at the Kaub station on the Rhine river, issued December 27th, 2009, at 6 GMT. Panel (a): Raw ensemble trajectories and the corresponding observed values. Panel (b): Raw ensemble and EMOS forecast density at a lead time of 48 h. Each dot represents a raw ensemble member. Panels (c) to (f): Quantiles of the predictive distributions obtained by standard, seasonally trained EMOS (SEAS EMOS) and the analog-based approaches SD EMOS, HMA EMOS, and DTW EMOS, respectively.

In recent work [Hemri and Klein, 2017], we have addressed the particular challenge faced by shipping companies when they need to determine the optimal loading of their ships. This optimal loading depends on future water levels: Loadings that are too high may force ships to stand still, whereas unnecessarily low loadings entail extra costs. To address this specific problem, we have developed EMOS techniques for the generation of probabilistic forecasts that are tailored to this setting and – importantly – we have constructed and applied proper scoring rules – variants of the aforementioned CRPS – that take the characteristics of the shipping companies' vessels into account. *Figure 16* illustrates the approach, with panel (a) showing the trajectories of the raw ensemble members along with the realizing, observed water levels at the Kaub station on the Rhine river. The forecast was issued on December 27th, 2009, and extends to a prediction horizon of 120 hours. Panel (b) examines the various forecasts at the particular prediction horizon of 48 hours, where the raw ensemble forecasts have labels that correspond to the specific weather forecast information they use. Finally, panels (c) through (f) display quantiles of the statistically post-processed forecasts under various innovative, analog-based variants of the standard EMOS approach – called SD EMOS, HMA EMOS, and DTW EMOS, respectively – which differ in the judicious, hydrologically motivated choice of the training data used to estimate statistical parameters.

Geometric-optical illustions

Located just a stone's throw away from HITS, the Carl Bosch Museum offers an exciting range of technical and historical exhibits. The permanent exhibits focus on the life and discoveries of Carl Bosch (1874 – 1940), who was awarded the Nobel Prize in Chemistry in 1931 and had his family residence on what are today the HITS premises. The topics covered range from Bosch's life to the development of chemical engineering to its political and economic effects as Bosch rose to become chairman of both BASF and IG Farben. In addition, the Carl Bosch Museum offers temporary exhibits on biannually changing themes. Since September 23rd, 2017, the temporary exhibits have featured optical illusions – the perfect motive for us to introduce recent research by Werner Ehm in this area [Ehm W, Wackermann J. Geometric-optical illusions and Riemannian geometry. Journal of Mathematical Psychology (2016) 71:28-38].

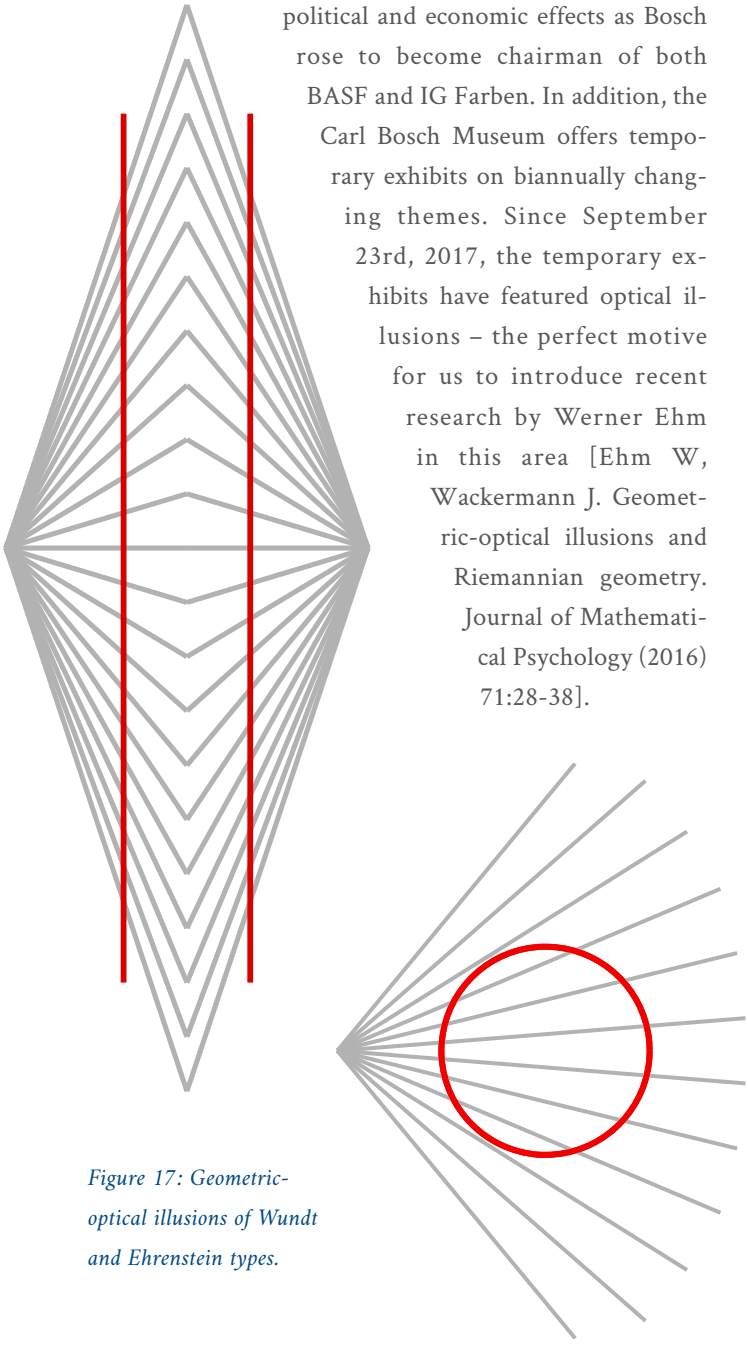


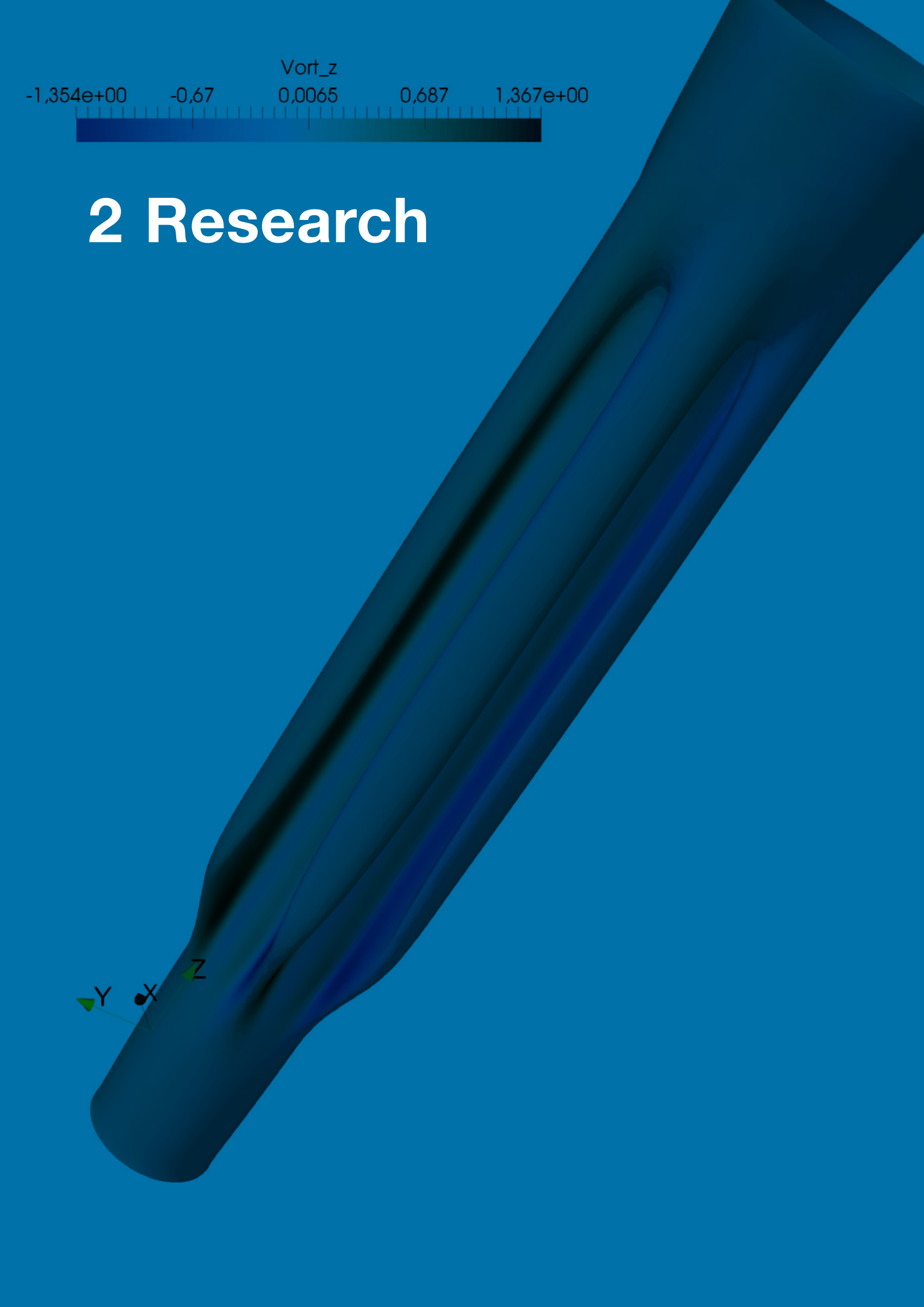
Figure 17: Geometric-optical illusions of Wundt and Ehrenstein types.

Simple geometric objects (called "targets" in this setting) appear distorted when presented along with suitable patterns of further lines (called "context"). To provide an example, *Figure 17* shows illusions of the Wundt and Ehrenstein types, for which the target is either a straight line or a circle. While attempts to explain such geometric-optical illusions abound in the literature, a generally accepted, unifying theory still is a distant prospect. In this light, it appears natural to develop descriptive tools for a quantitative appreciation of the visual phenomena in the first place.

The approach begins with two well-supported empirical observations in the extant literature. The first is the "local interaction" principle, meaning that the distortions depend on the context in the immediate vicinity of the target only and not on its global structure. The second observation is "regression to orthogonality" as small angles between context lines and the target appear larger than they are, thereby regressing toward a right angle. In the mathematical model, the context lines are interpreted as flow lines of a vector field, and local interactions between target and context are modeled as functions of the intersection angle.

This ansatz allows for the application of classical tools of differential geometry and variational calculus, and the solution to the associated variational problem serves as a "prediction" of the target as it is seen. Specifically, the true and the perceived target are represented as geodesics (i.e., shortest paths) in distinct geometries. For the true target, a baseline applies (Euclidean geometry for linear targets and hyperbolic geometry for circular targets), and for the perceived target, a context-induced perturbation of the baseline geometry is used. To a good approximation, the difference between the perceived and the true target factorizes as "shape of the distortion" times a parameter that measures effect size. While shape is computable from the target and context, the magnitude parameter varies between individuals and has to be determined empirically. Experiments with illusions of the Wundt and Ehrenstein types verify the adequacy of the predictions. The present approach may thus supply a template for variational principles applicable to a wider range of geometric-optical illusions.

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. Dr. Heuveline.

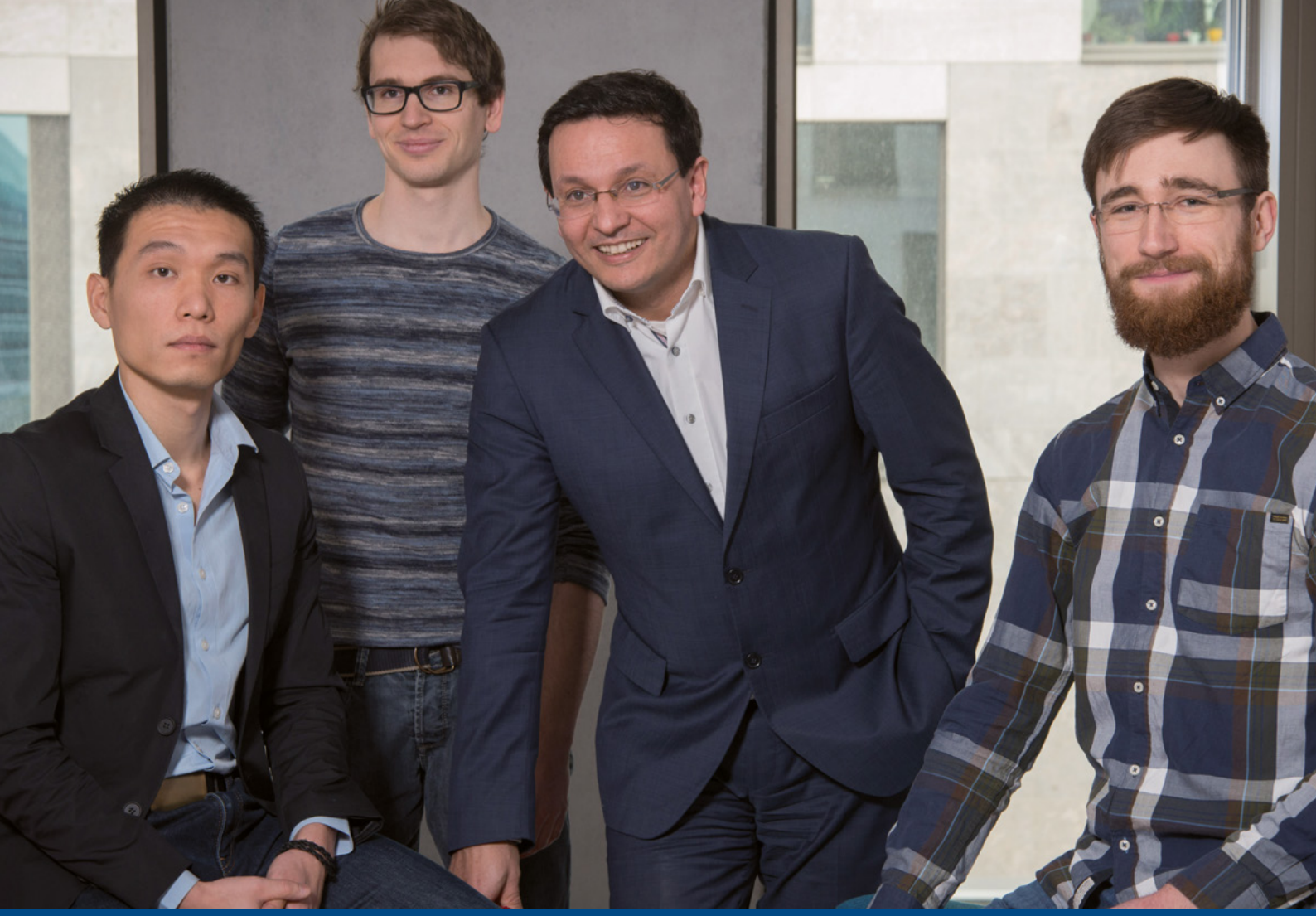
DMQ's research focus lies on gaining knowledge from extremely large and complex datasets through data mining technologies. Reliability considerations with respect to these datasets are addressed by uncertainty quantification methods. Both fields – data mining and uncertainty quantification – require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

In 2017, DMQ focused on research activities in the areas of uncertainty quantification for blood-flow simulations in the human aorta and in a blood pump device as well as for thermal electro-hydrodynamical systems. Moreover, the further development of our open-source finite element library HiFlow3, which incorporates new numerical schemes and data-driven computation, came under focus in 2017.

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 (IWR) der Universität Heidelberg zusammen, welches ebenfalls 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.

Im Jahr 2017 arbeitete die Gruppe dazu in folgenden Anwendungsbereichen: Uncertainty Quantification für Blutflusssimulationen in der menschlichen Aorta und in einer Blutpumpe, sowie für thermal-elektro-hydrodynamische Strömungen. Darüber hinaus stand die Weiterentwicklung der Finite-Element Software HiFlow3 im Fokus, wofür neue numerische Methoden und datengesteuerte Algorithmen implementiert wurden.



Group Leader

Prof. Dr. Vincent Heuveline

Scholarship holder

Sotirios Nikas *(since December 2017)*

Staff members

Chen Song
Philipp Gerstner
Dr. Peter Zaspel *(until February 2017)*
Fabian Kißler *(until April 2017)*
Suranita Kanjilal *(April – August 2017)*

Visiting scientist

Wei Zhang *(since November 2017)*

Moving towards uncertainty quantification for thermal electro-hydrodynamics (TEHD)

Containment in engineering often provides thermal insulation, as in heat exchanger systems, for which an improvement of heat transfer via efficient enhancement is of general interest due to its low operational costs and sustainable usage of energy. One possible heat transfer enhancement technique is given by applying electric fields, 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 fields and temperature gradients by combining numerical simulations with experimental data. We consider a multi-physics model based on the well-known Boussinesq equations for natural convection combined with an additional electrical body force.

In order to obtain an approximate solution for this set of partial differential equations, we use the so-called Finite Element Method. For this type of discretization, we have been able to derive a priori error estimates that show that the numerical solution approximately converges to the exact solution under certain conditions. Moreover, by comparing the numerical solution with experimental data, we could observe a good fit of qualitative solution characteristics for a range of predefined benchmark problems.

When conducting simulations for various temperature differences between the inner and outer wall, we observed the emergence of axially oriented vortices in both the experiment and the numerical simulation. Due to these vortices, radial heat transfer is enhanced, as is demonstrated by an increase of the corresponding Nusselt number. Moreover, since the number of vortices and therefore the hydrodynamical behavior varies for different temperature gradients and fluid parameters – which are not exactly known – our goal is to employ uncertainty quantification methods in order to obtain reliable results.

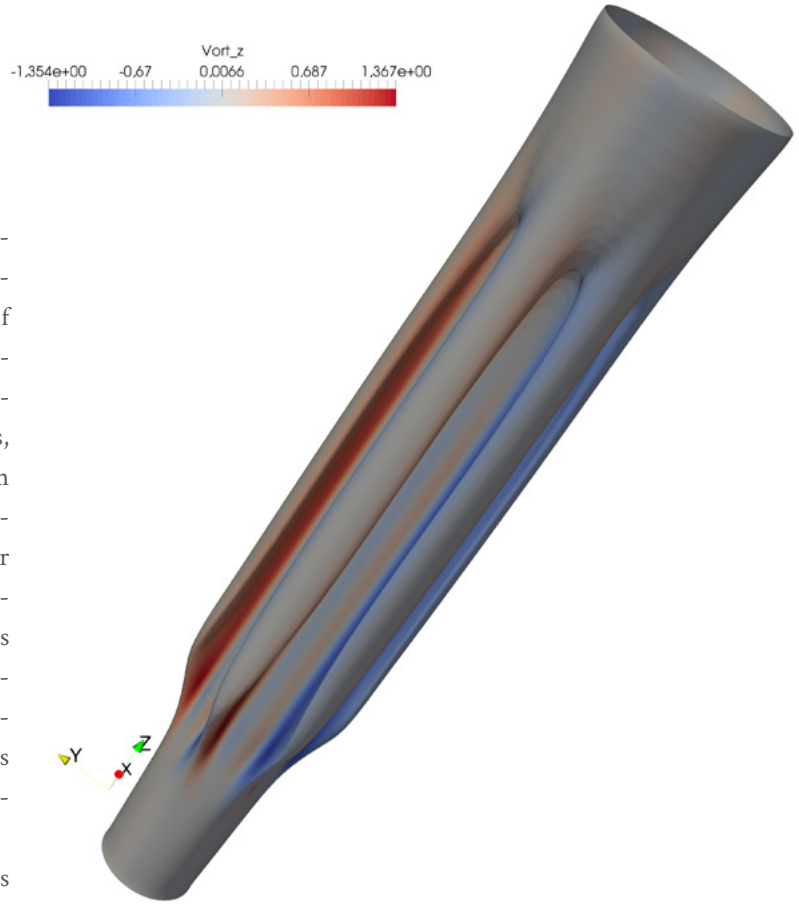


Figure 18 illustrates the considered geometry, for which the fluid is contained inside the gap of two concentric cylinders. The inner cylinder is heated, whereas the outer cylinder is cooled. In addition, an electric field is applied between the two of them. On the left, the isosurface of some specific temperature is shown; on the right, the temperature profile is illustrated on a horizontal cut in the middle of the cylinder. *Figure 19* reveals how the heat transfer to the inner wall increases due to arising vortices.

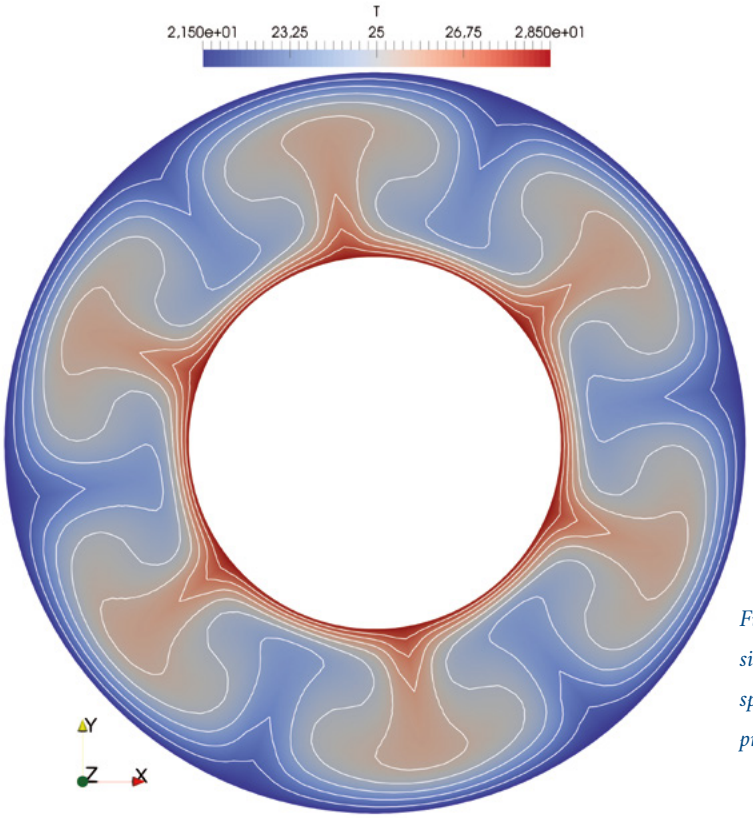


Figure 18: Geometry of TEHD simulation, together with isosurface of specific temperature and temperature profile on horizontal cut.

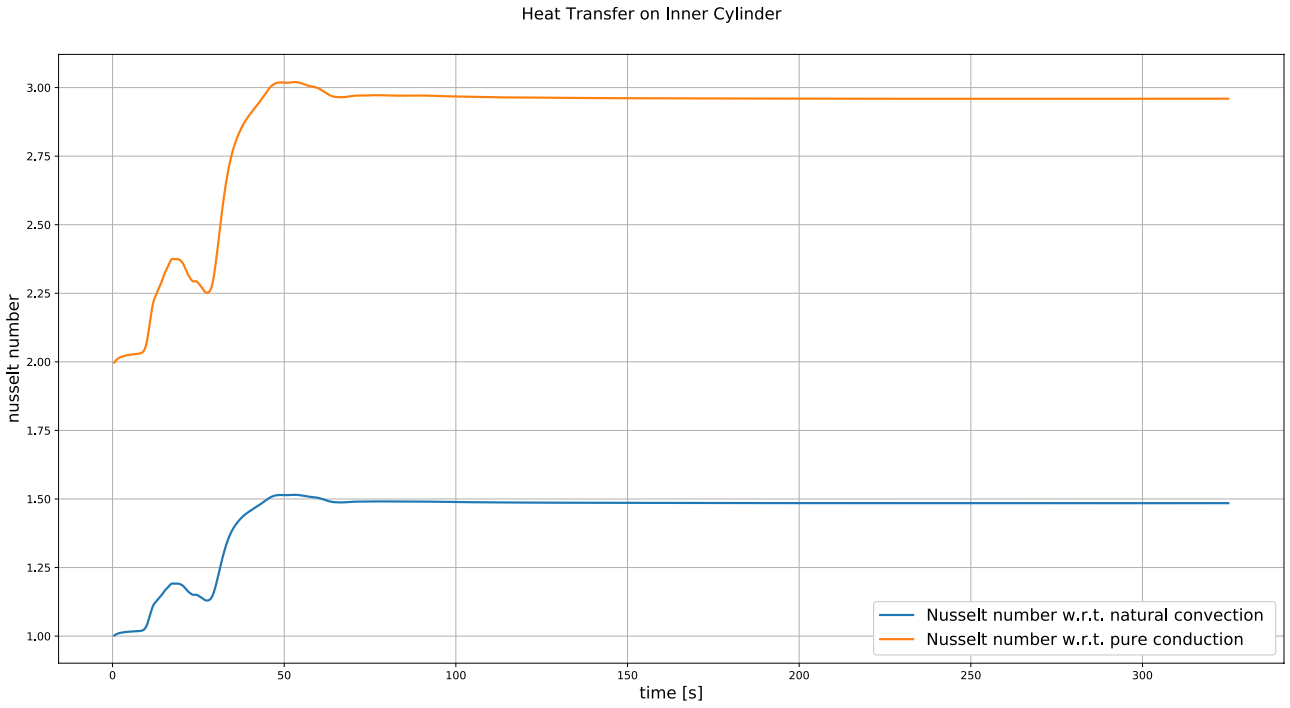


Figure 19: Nusselt number, i.e. heat transfer on inner wall, over time.

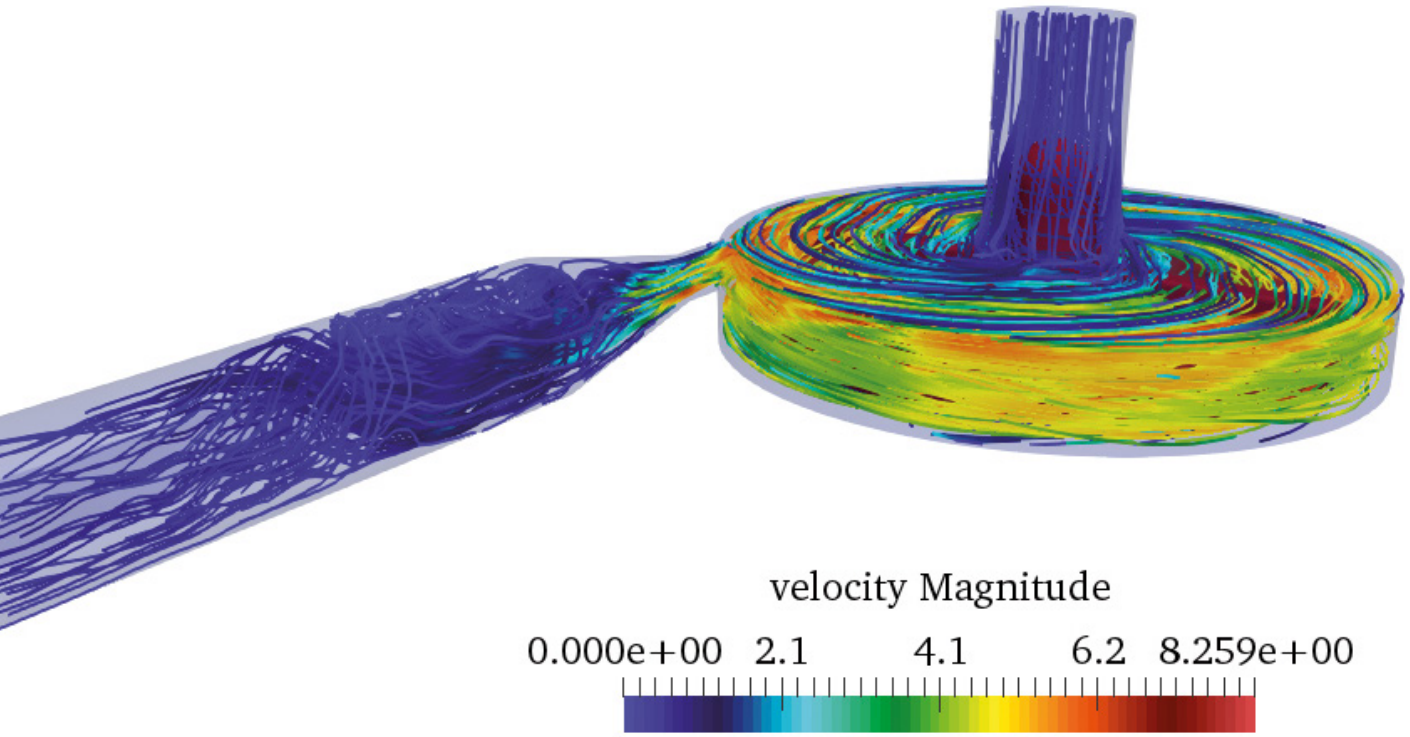


Figure 20: Streamline in the blood pump chamber.

Uncertainty quantification for a blood pump device

Heart failure (HF) – also known as congestive heart failure (CHF) – is a cardiovascular disease that renders the heart no longer capable of sustaining the blood flow that the human body requires to the circulatory system. There are more than 40 million people in the world who suffer from heart failure, which causes more than 17.5 million deaths per year worldwide. Although the best option at the moment for a heart-failure patient is to have a heart transplantation, the shortage of available heart donors in addition to the increasingly ageing population has put the healthcare system under enormous pressure.

As implantable supporting heart systems – or blood pumps – are an effective temporary solution for patients suffering from heart failure, the number of blood pump implants has now surpassed the number of heart transplants. However, the mortality rate of ventricular assist devices (VADs) is still very high (i.e., 20 % after the first year). In addition, the heart failure patients have a higher risk of stroke than the healthy people. Therefore, there is still room to improve the performance of blood pump devices.

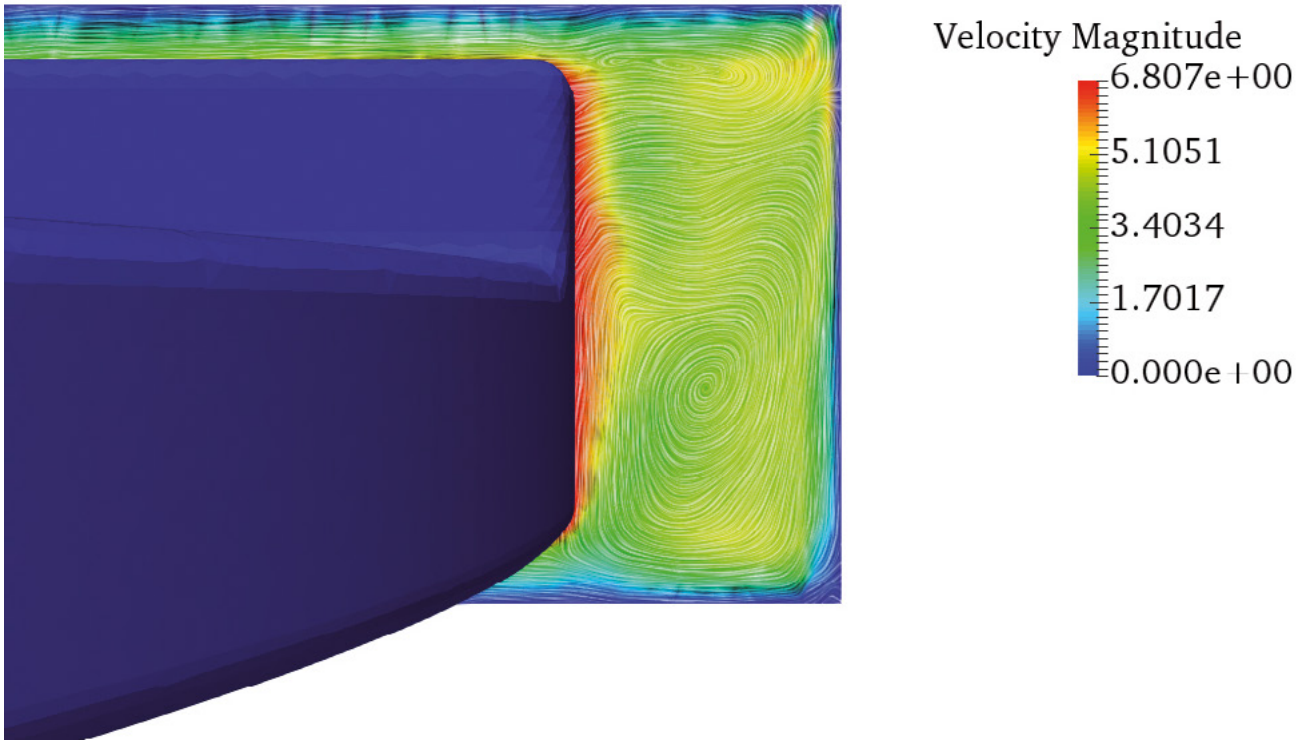


Figure 21: The flow velocity distribution in the casing.

The finite element method (FEM) is a sophisticated numerical method for studying complex mechanical devices. For our blood pump application, it provides information about the velocity pressure field, pressure and force loadings, hemolysis, etc. This information can help us to improve the performance of the ventricular assist devices. In addition to modeling the rotation process and blood flow into the pump, the actual applications of the pump are uncertain. As a result, Uncertainty Quantification (UQ) provides a framework that can be used to characterize uncertainties in simulations and is also an important tool in studying the variability and probabilistic behavior of physical systems.

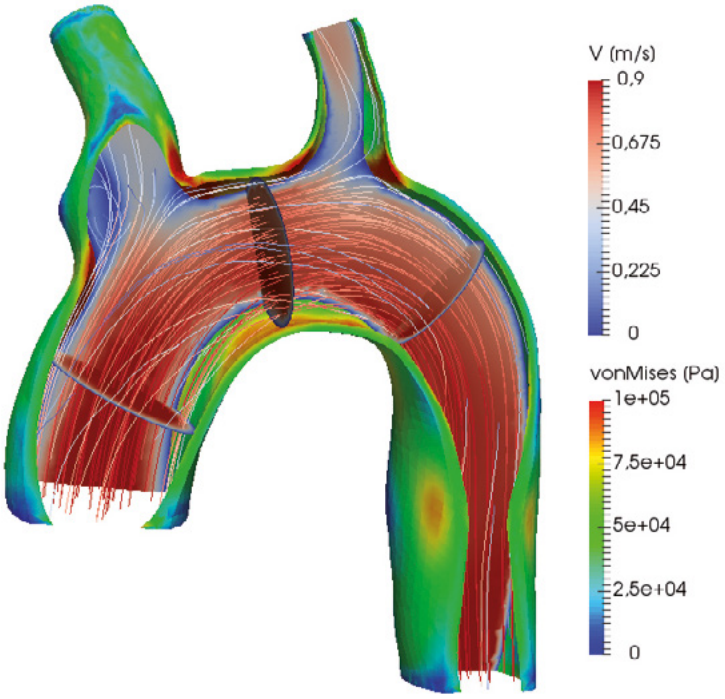
In our work, we incorporate the stochastic Galerkin projection technique and the variational multiscale formulation in a rotating system to simulate the propagation of uncertainties – which are influenced by uncertain input data – in a blood pump. The simulation is achieved via our open-source Finite Element library HiFlow³. In order to be able to efficiently solve the coupled stochastic system, we have additionally developed a multilevel preconditioner that is suitable for high-performance computing (HPC).

Will your aorta rupture? – Blood-flow simulation under uncertainties

In recent decades, biomedical studies with living probands (in vivo) and artificial experiments (in vitro) have been complemented more and more by computation and simulation (in silico). In silico techniques for medical engineering can provide information on the diagnosis and risk stratification of cardiovascular disease, one of the most commonly occurring causes of death in developed countries. Other uses for in silico methods are provided by virtual prototyping and the simulation of possible surgery outcomes. High reliability is a requirement for cardiovascular diagnosis and risk-stratification methods, especially with surgical decision-making. Given uncertainties in the input data of a simulation, this reliability requirement implies the need to quantify the uncertainties in simulation results.

For aneurysms of the aorta, the conditions for vessel dilation and for an increased risk of rupture are an ongoing subject of research. Numerical simulations can be used as a tool to assess aneurysm growth from a biomechanical point of view. For example, the stress load within the vessel wall can be a factor in aneurysm development. However, the stress load is not directly measurable, and modeling of aortic vessel walls and blood-flow dynamics leads to complex fluid-structure interaction (FSI) problems. Additionally, the model parameters can be highly uncertain for a given patient.

In our work, we develop a numerical framework that computes the probability that a threshold for the stress load within a vessel wall will be exceeded. This framework is based on an FSI model of blood vessels, and the model is configured patient-specifically by 4D flow magnetic resonance imaging (MRI). The propagation of uncertainties relies on a stochastic collocation method. Each collocation point is given by a finite element simulation using HiFlow³. As a result, the probability that a critical threshold of the stress load in the aortic vessel wall will be exceeded can be visualized over the vessel wall geometry (*cf. Figure 22*). The results open new perspectives for clinical studies.



HiFlow³

HiFlow³ is a multi-purpose finite element software that provides powerful tools for the efficient and accurate solution of a wide range of problems modeled by partial differential equations (PDEs). Most prominently, the considered problems arise from the areas of medical engineering, meteorology, environmental physics, and energy. In the field of medical engineering, patient-individual functional modeling and numerical simulation can provide (virtual) insights into operated tissues as well as into their behavior and functionality.

Computational fluid dynamics or fluid-structure interaction simulations can depict the flow behavior, for example, of blood in the aorta and hence enable a risk analysis in aortic aneurysms. In the field of meteorology, numerical simulation with different physical models (i.e., different sets of underlying PDE) allows for investigating the effects of specific physical phenomena on the numerical solution in a unified and mathematically sound framework.

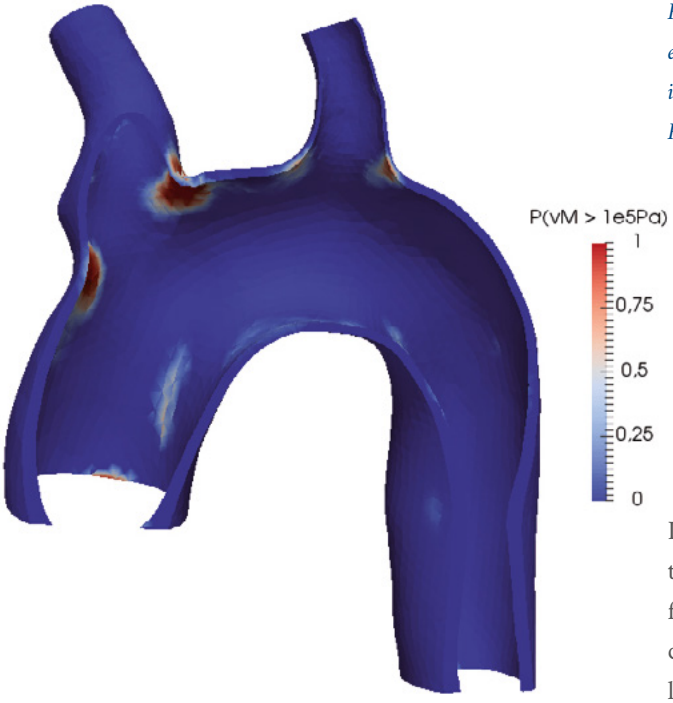


Figure 22: Simulation of the coupled blood flow and vessel wall elasticity of a human aortic bow. Left image: Mean value of the flow indicated by streamlines and the stress distribution within the wall. Right image: Probability of stress overlaod in the vessel wall.

HiFlow³ was designed as a library with high modularity and consequently with flexibility, thereby enabling it to address a broad range of application scenarios in which each scenario can be treated with a variety of methods in order to determine and choose the solution methodology that is best suited for the problem at hand. "Best suited" is not only meant in terms of approximation quality but also in terms of time-to-solution, which implies that each component of the software needs to excel in performance as well as in parallel scaling behavior. The importance of scalability arises from the fact that on the one hand, the challenging problems of current research in the considered fields of application require vast amounts of computing power, and on the other hand, scientists develop their codes on their local machines (e.g., laptops). Therefore, HiFlow³ is required to scale excellently from laptops to clusters.

In order to cope with the tasks and problems stated above, the HiFlow³ framework maintains several specialties. The first is a generic modular structure, which reflects a typical simulation cycle based on finite elements. This modular structure ensures the flexibility and maintainability of the software. Furthermore, the gained flexibility is a key prerequisite in adapting HiFlow³ to different mathematical and algorithmic needs as well as to different hardware- and computing architectures. The importance of the latter cannot be underestimated because high performance computing clusters feature highly heterogeneous hardware nowadays. This observation motivates the second specialty of the HiFlow³ framework, namely the ability of hardware-aware computing, which refers to the software's ability to provide and utilize specialized implementations of functions (especially performance-critical ones) that are adapted to the underlying hardware of the current computing system in order to unleash as much of the available computing power as possible. In the current state of the software, this process occurs extensively in the linear algebra module of HiFlow³. The third specialty is the HiFlow³ framework's awareness of the topic of energy consumption and energy efficiency, which plays an important role in current high-performance computing clusters. In the context of HiFlow³, this awareness is tightly connected to hardware awareness on the one hand and is currently respected in the design and implementation of a specific algorithm – namely the geometric multigrid method – on the other hand.

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 field of hyperbolic geometry, and projective geometry, which has its origins in the study of perspective in art and is based on incidence relations rather than on the measurement of distances. 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 (e.g., manifolds and metric spaces) and groups that act as symmetry groups on these spaces. We place a special focus 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, und bilden so eine Gruppe.

Im 19. Jahrhundert 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

Mareike Pfeil
Florian Stecker
Nicolaus Treib

Visiting scientists

Anna Schilling
Dr. Andreas Ott
Dr. Gye-Seon Lee
Dr. Sourav Ghosh
Dr. Daniele Alessandrini
Dr. Shinpei Baba
Evgenii Rogozinnikov
Johannes Horn
Dr. Federica Fanoni
Dr. Andy Sanders
Christoph Karg
Dr. Beatrice Pozzetti
Dr. Valentina Disarlo

Students

Sven Grützmacher
Tim Adler
Menelaos Zikidis

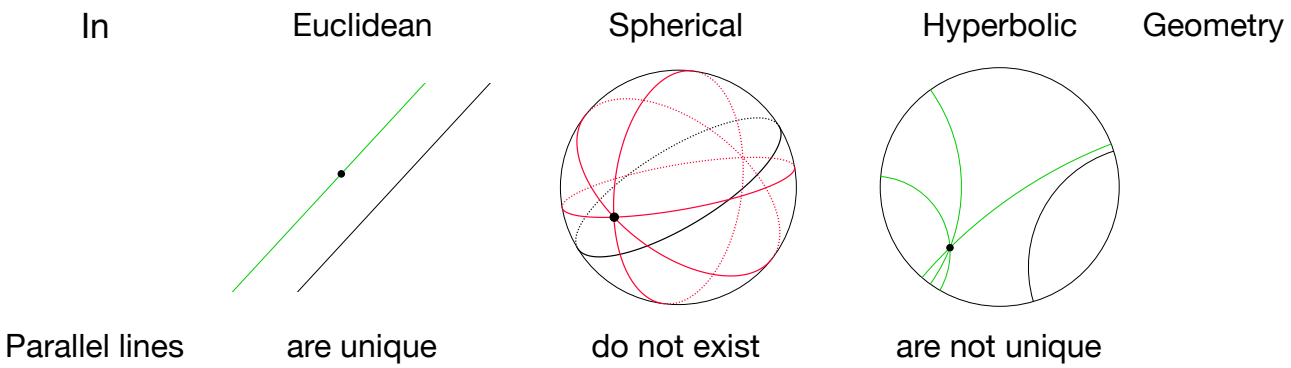


Figure 23: The three types of two-dimensional geometries: Euclidean, spherical, and hyperbolic.

Types of geometries

For most people, the word geometry – together with its various notions and laws – is strongly connected with 2- or 3-dimensional Euclidean geometry. Angles and distances come to mind, as do compass- and straightedge constructions. In Euclidean geometry, the sum of the interior angles of a triangle is 180 degrees, and the relation between its sides and angles is governed by the sine- and cosine theorems. Another important feature is the parallel postulate: Given a line and a point not on it, there is a unique second line through this point that does not intersect the first line and is parallel to it. This prevalent focus on Euclidean geometry stands to reason, of course, since it can be used to describe most aspects of our daily lives.

Taking a look at a map of flight routes, however, it is clear that there are also other important types of geometries. We live on the surface of the earth (a sphere), and the shortest connection between two points follows a great circle, which is why flight routes to the United States appear to curve towards the North Pole. When replacing straight lines with great circles, it becomes clear that the familiar rules of Euclidean geometry do not apply in this setting. In this context, sums of the interior angles of triangles are greater than 180 degrees and depend on the triangle in question. Moreover, there are no parallel great circles, as any two of them intersect. These properties are intimately connected to the fact that the sphere is positively curved at every point, which is not true in flat Euclidean space. As a consequence of the curved nature of the sphere, maps can never correctly depict all distances. While this effect is negligible on small scales (e.g., in city maps), it becomes very apparent when comparing the different projections people have used to draw maps of the whole surface of the earth.

There is a third type of geometry in dimension 2 that is called hyperbolic geometry and is in many ways the counterpart to spherical geometry. In this geometry, the sum of the interior angles of triangles is always smaller than 180 degrees, and given a line and a point not on it, there are many other lines through this point that do not intersect the first line. Hyperbolic geometry is a negatively curved geometry. At every point, it looks like the surface of a saddle. Unfortunately, this is more difficult to visualize. Any embedding of the hyperbolic plane in Euclidean space will not depict distances correctly.

Accepting this drawback, one possible model is the Poincaré disk model, which correctly depicts angles while heavily distorting distances towards the boundary of the disk. A traveler starting at the center of the disk and walking at constant speed towards the boundary would appear to the naïve observer to slow down more and more and never actually reach the boundary, which is in fact infinitely far away. In this model, the shortest path from one point to another follows a circular arc that is orthogonal to the boundary. These arcs are called geodesics. Like great circles on the sphere, they are the "straight lines" of hyperbolic geometry (see *Figure 23*).

Following Felix Klein's (1849 – 1925) fundamental ideas, what we consider to be geometry nowadays is a pair (X;G) of a space X and a group G of transformations or symmetries that act on this space. The term "group" expresses the idea that two transformations can be composed, thereby yielding a new transformation. Moreover, for every transformation, we also have its inverse, which moves every point back to where it was. The symmetries of a geometry should include sufficiently many transformations such that for any two points in the space X, there is a transformation that sends one to the other. Notably, interesting invariants of a geometry are not encoded in the space X. Instead, notions like angles or distances are only significant because they are preserved by the group G. For instance, X could be 2-dimensional space, and G the group of Euclidean isometries, maps preserving all Euclidean distances. The result is Euclidean 2-dimensional geometry, exactly as we are used to. On the other hand, X could be the sphere, and G the group of all rotations, which would yield spherical geometry.

In this way, the task of understanding a specific geometry becomes the task of understanding its group of symmetries. It also gives rise to many less familiar geometries, including affine geometry, in which we allow not only Euclidean isometries but also scalings and shearings in the group G. The notion of distance thus loses its meaning as it is not preserved by scalings, but straight lines and the parallelism of straight lines are. Another example is projective geometry, which has its origins in the study of perspective in art. In this geometry, the only invariant properties are incidences. For example, the only relation between a point and a line that is invariant under the group of projective transformations concerns whether or not the point lies on the line.

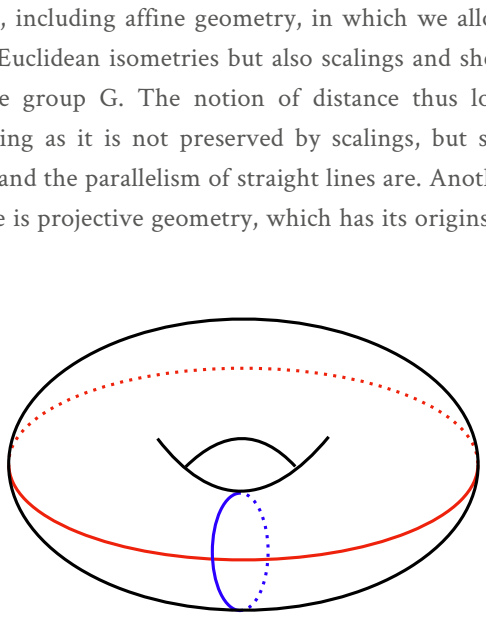


Figure 25: Noncontractible loops on the torus.

Loops and the fundamental group

In order to understand and distinguish more complicated geometric objects, it is often helpful to associate some algebraic invariants to them. One particularly important invariant is the fundamental group, which consists of all loops that can be drawn on the object that begin and end at the same fixed base point. Two loops are considered the same if one can be deformed into the other while keeping the base point fixed. The "neutral" loop is then the one that never leaves the starting point, and the inverse of a loop is the same but with the opposite direction. The fundamental group describes some topological properties of the shape we are considering. On the sphere, for example, every loop can be contracted to the neutral loop. For example, assume that we have a loop that begins and ends at the South Pole and that – for simplicity's sake – does not pass through the North Pole. Imagine pushing the whole loop away from the North Pole and towards the South Pole until it finally stays fixed at the South Pole (see *Figure 24*). On the torus, however, this is not always possible. Any loop that winds around the torus "horizontally" or "vertically" cannot be contracted (see *Figure 25*). A closer analysis reveals that the fundamental group of the torus is given by $\mathbb{Z}^2$, the group consisting of pairs of integer numbers, with composition given by adding the respective numbers. This means that any loop is characterized by the number of times it winds around the torus horizontally and vertically.

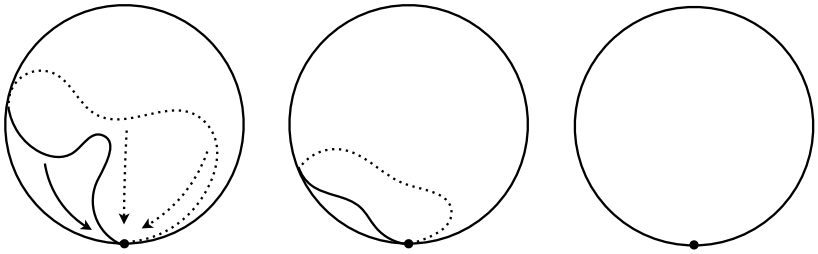
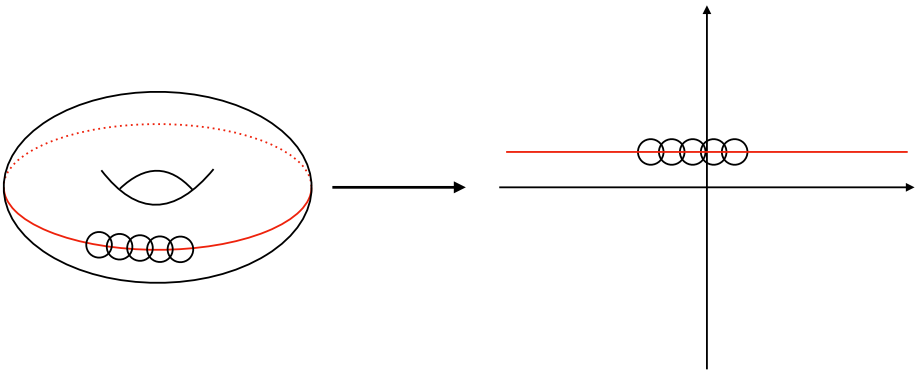


Figure 24: Contracting a loop on the sphere.

Geometric structures

In our research, we are rarely confronted with questions about these model geometries, which are already well-understood. Instead, a more common problem is deciding whether a given topological object M can be equipped with a geometric structure that locally looks like the model space X. For example, beginning with a square and gluing opposite edges together results in a torus. In a small neighborhood of any point, the square looks just like the Euclidean plane, so this construction induces a Euclidean structure on the torus. Charles Ehresmann (1905 – 1979) and William Thurston (1946 – 2012) developed an important link between geometric structures on a topological space and homomorphisms from its fundamental group into the group of symmetries G. The basic idea is that of "unrolling" the shape onto the model space: A small patch of M looks exactly like a small patch of the model space X. Towards the edge of this patch, we find another small patch that looks like a part of the model space, and we obtain a second, overlapping patch in X. An interesting phenomenon occurs as we repeat this procedure along a loop in M, covering it in model patches along the way: Its image in X does not necessarily close up. We can see this happen in the torus example from before: If we consider a closed curve that winds around the torus, its "unrolled" image in the plane is simply a straight line that never closes (see *Figure 26*). In this way, we associate each loop in M with a transformation of X that sends the starting point to the endpoint of the unrolled curve. In other words, this is a map from the fundamental group of M to the group of symmetries G that preserves the group structure – a homomorphism. It describes which points of the model space X we have to glue together if we want to get back to the topological object M and is called the holonomy representation of this (X;G) structure on M. Ehresmann observed that locally, the space of such group homomorphisms looks exactly the same as the space of (X;G) geometric structures on M.

Figure 26: "Unrolling" the shape: a map of the torus.



Both directions of this equivalence have interesting applications: On the one hand, they provide a means of tackling the question raised above: Namely, in order to find out whether a given topological object M admits geometric structures of type (X;G), we can examine the possible group homomorphisms from the fundamental group of M into the group G of symmetries. Classifying such homomorphisms even reveals all the possible ways of endowing M with an (X;G) structure. On the other hand, we might stumble upon an interesting class of group homomorphisms in a number of ways by generalizing a known construction or because mathematicians or physicists from a different field are interested in them. If the homomorphisms fit the framework of Ehresmann and are holonomy representations of an (X;G) geometric structure on M, then they have a geometric meaning.

Schottky groups

To illustrate the point of finding a geometric interpretation for a "nice" class of group homomorphisms, we examine Schottky groups, which are named after German mathematician Friedrich Schottky (1851 – 1935). These groups of isometries of the hyperbolic plane are constructed as follows: Begin with an even number of disjoint halfspaces (i.e., regions in the hyperbolic plane bounded by a geodesic). Then, pair these halfspaces using isometries of the hyperbolic plane (see *Figure 27*). Each of these isometries will push everything that lies outside of the first of its halfspaces into the second one. We therefore refer to them as the repelling and attracting halfspaces. All possible compositions of these transformations and their inverses form the group in which we are interested.

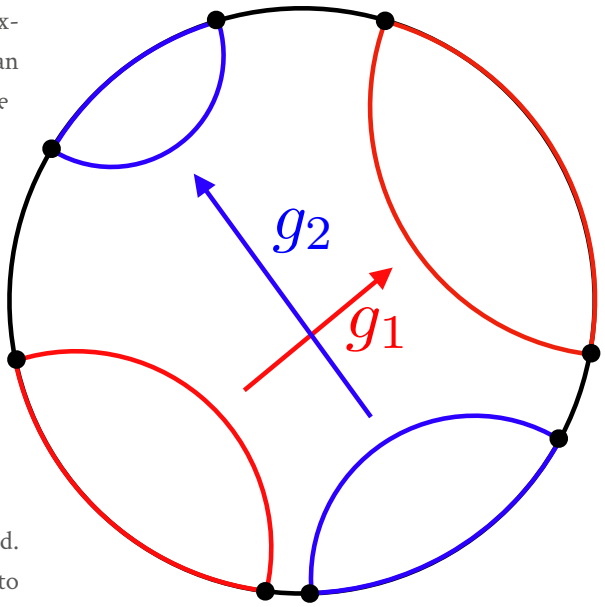


Figure 27: A Schottky pairing in the hyperbolic plane.

These groups are fairly easy to understand due to what is called the "ping-pong lemma". The idea behind it is easy to visualize: Consider a point in the center of the disk and imagine where it ends up as we apply a series of pairing transformations to it. Since the point is not contained in any repelling halfspace, the first transformation will send it into its attracting halfspace. Applying the second transformation afterwards, the point is now pushed into the attracting halfspace of this transformation, and the process repeats itself. The point jumps back and forth between the different halfspaces, like in a game of ping pong – albeit on an unconventional table (see *Figure 28*). Most importantly, the point never returns to its starting position, so the composition of transformations can never be the identity transformation. In mathematical terms, this is called a free group, which emphasizes the idea that the generating transformations are not related in any way.

Figure 28: A point "bouncing" between halfspaces.

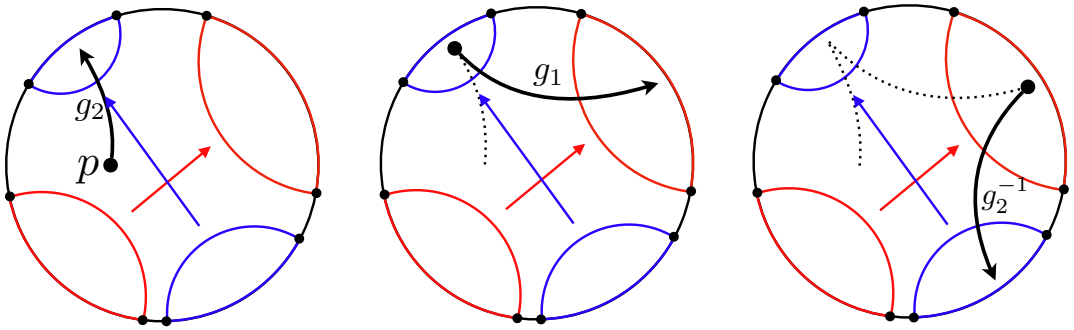


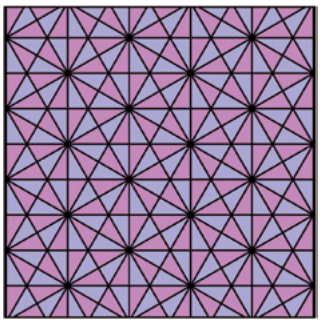
Figure 29: Visualizing Schottky groups.

The previous description of geometric structures opens up a geometric point of view on these groups: We can interpret them as a rule that tells us which points of the hyperbolic plane we should identify. If we used two pairs of halfspaces that were paired crosswise, for example, we would end up with a torus that had a hole cut into its surface. The Schottky group thus endows this surface with a hyperbolic structure. Tim Adler and Sven Grützmacher have used our 3D printer to create tangible models associated with Schottky groups for use at outreach events (see *Figure 29*).

A popular approach in mathematics is to begin from a well-understood concept and apply it to a different setting, modifying it along the way as necessary. This works well with Schottky groups. The observation at the heart of the approach is that there is a unique geodesic connecting any two points in the boundary of the hyperbolic plane. Consequently, we may as well forget about the interior of the disk and consider only the boundary, which is simply a circle: Halfspaces become intervals in the circle, and the hyperbolic isometries used for the pairing have similar attracting-repelling dynamics on the circle. Adding yet another level of abstraction, intervals can be seen as being determined by the cyclic order on the circle. The three points x, y, and z are in positive order if – beginning at x and going around the circle counter-clockwise – we first encounter y and then z. The interval between two points is simply the set of all points that lie between them in this order. The benefit of this formulation is that such structures can be found in a variety of different settings, and we can therefore construct generalized Schottky groups there. For example, the Einstein Universe – a compactification of 3-dimensional Minkowski space – carries a partial cyclic order related to its causal structure. A point lies between two other points if it is in the future of one and in the past of the other. In the more general setting of Hermitian Lie groups, generalized Schottky groups turn out to be equivalent to so-called maximal representations. These are special homomorphisms from the fundamental group of a surface into a Hermitian Lie group that are characterized by the fact that an elaborate algebraic invariant takes on its maximal possible value. The interpretation as generalized Schottky groups provides a very explicit description of maxi-

mal representations. Generalized Schottky groups also provide examples of Anosov representations, which is yet another class of homomorphisms that has received a great deal of attention in recent years and that is connected to geometric structures on higher dimensional manifolds.

However, many interesting examples do not fit into the convenient framework of free groups. Indeed, we do not need to go far to leave this realm. Earlier, we saw that a free group can describe a torus with a hole cut into its surface. From a geometric point of view, it appears more natural to consider surfaces without boundaries. Fundamental groups of such surfaces are not free. For example, a "double torus" can be glued together out of an octagon. Its fundamental group has four generators with one relation between them (see *Figure 30*). The presence of relations between the generators makes it much more difficult to analyze such groups. In general, it is difficult to decide whether a given sequence in the generators is in fact trivial, which is in stark contrast to the case of free groups. This difficulty also makes the geometric objects with which these groups are connected more difficult to understand, and it is thus worthwhile to look for examples with a manageable amount of complexity. Such examples include reflection groups, which can be considered in various geometries, such as Euclidean, spherical, or hyperbolic geometry.



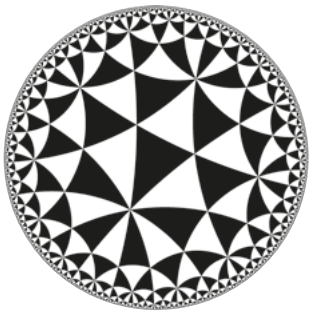


Figure 31: Triangular tilings in Euclidean, spherical, and hyperbolic geometry.

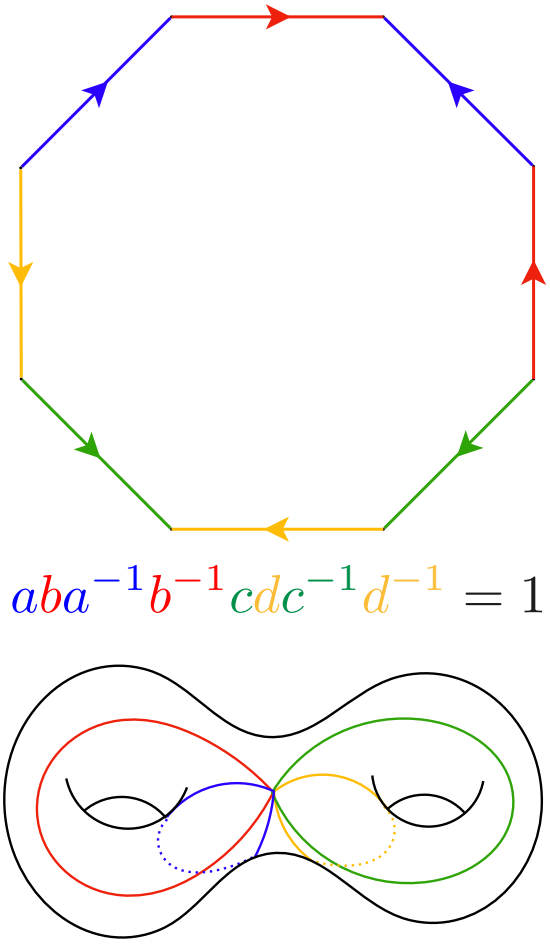
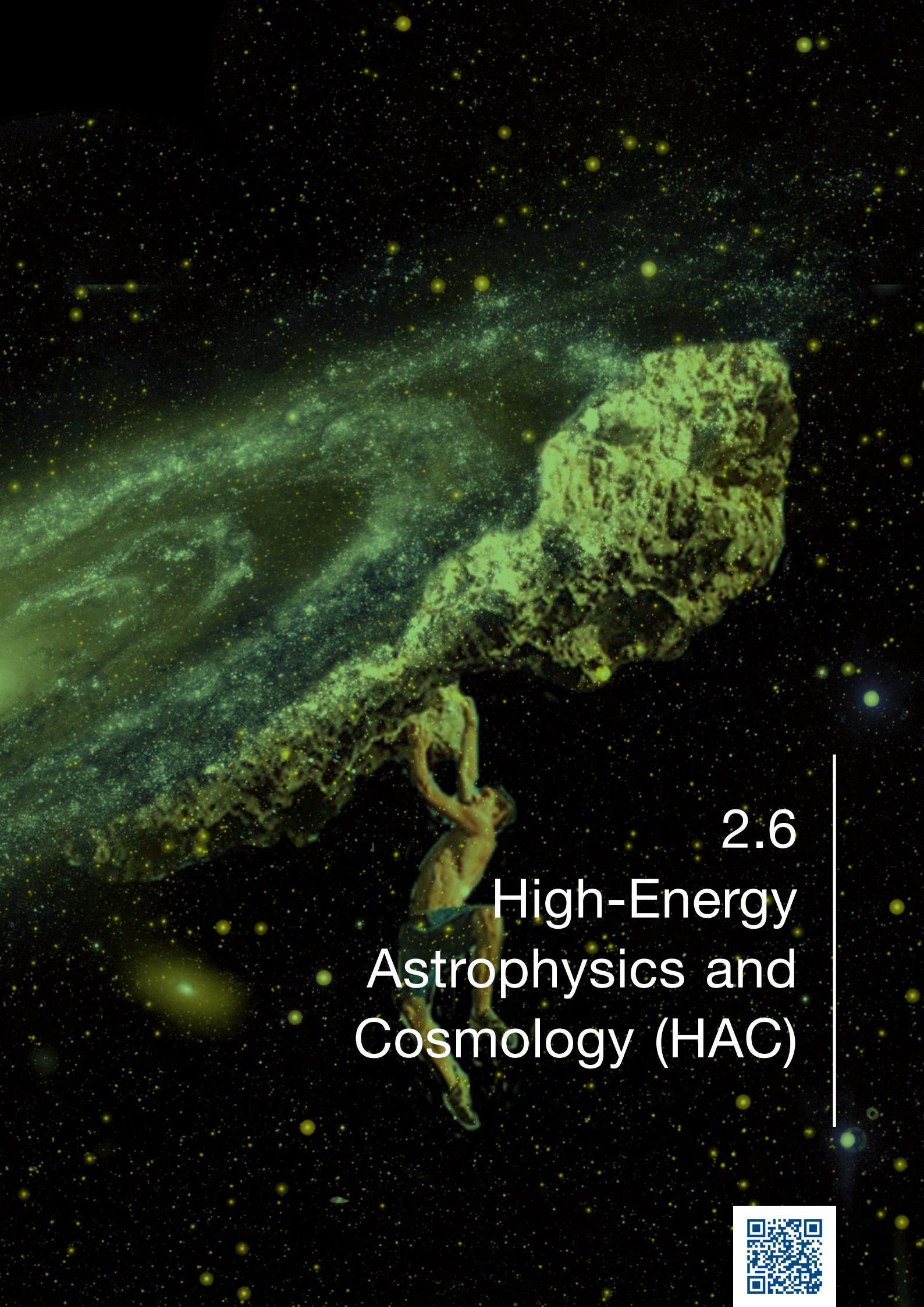


Figure 30: Gluing a double torus out of an octagon.

These groups are closely connected to tilings of the Euclidean plane, the sphere, or the hyperbolic plane and are thus well-suited for visualization purposes (see *Figure 31*). Their eponymous property is the fact that all generators are reflections along hyperplanes, which also determines the kind of relations that occur between generators: First, applying any reflection twice is the identity. Moreover, if two hyperplanes intersect at an angle α, composing them yields a rotation by the angle 2α. To obtain a tiling, this angle should divide 360°, so the corresponding power of the rotation is the identity.

As it turns out, this determines all relations between the generators, making reflection groups reasonably easy to understand. There are special hyperbolic surfaces, named Hurwitz surfaces after Adolf Hurwitz (1859 – 1919), whose fundamental group is mapped to a group of hyperbolic isometries that is almost a reflection group under the holonomy representation. Such connections between seemingly unrelated topics are a key component in our research.

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 – namely 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 program of focused observations of 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 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 in den Zentren dieser Galaxienhaufen befinden, das kühlende Gas zu heizen. Aber wie funktioniert dieser Mechanismus? 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

Prof. Dr. Christoph Pfrommer

Postdoc

Dr. Philipp Girichidis

PhD students

Matteo Pais
Georg Winner *(HGSFP Scholarship)*

Master student

Kristian Ehlert

The HAC group moved in August 2017 to the Leibniz-Institute for Astrophysics Potsdam (AIP) where Prof. Dr. Christoph Pfrommer has become a jointly appointed professor with the University of Potsdam as of April 2017.

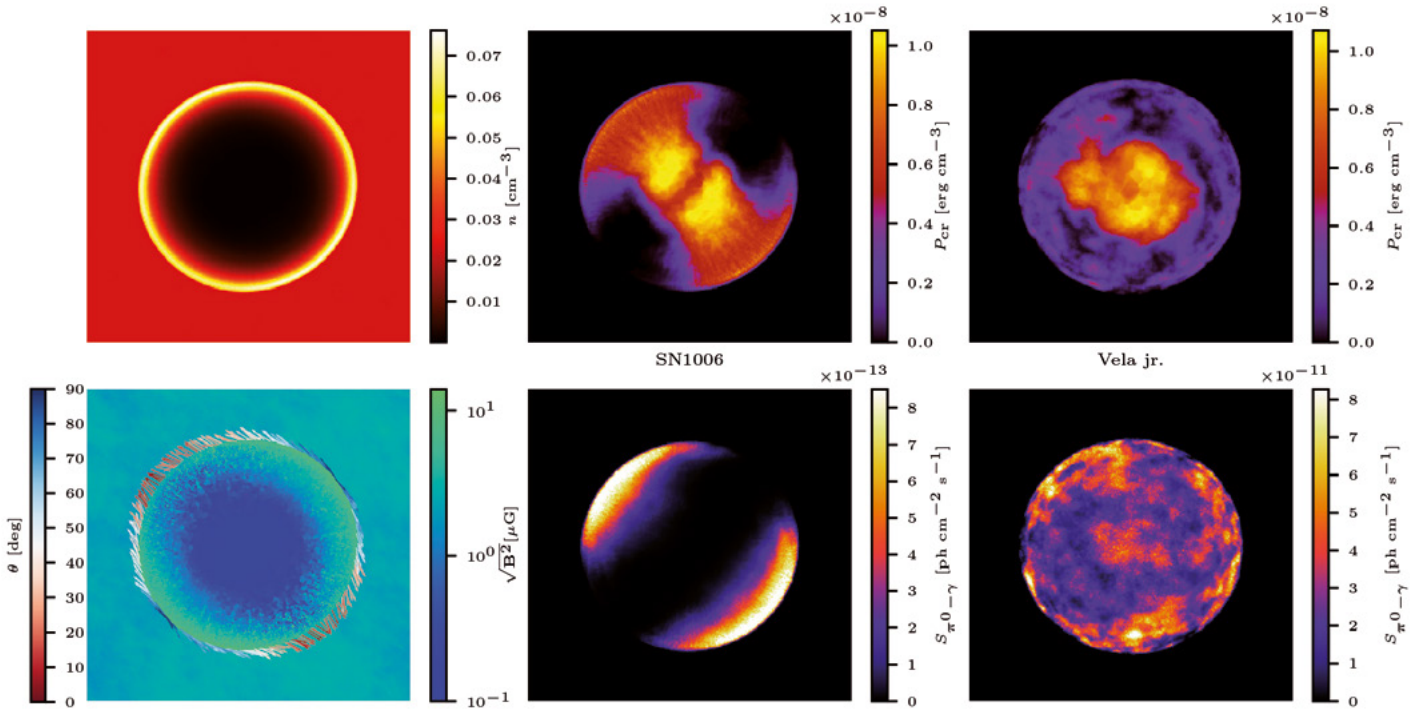


Figure 32: Expanding supernova shock with cosmic ray acceleration. We display cuts through the 3D simulation with a homogeneous magnetic field (initially oriented from top left to bottom right) of the density (top left) and magnetic field strength (bottom left), supplemented by magnetic field vectors at the shock. The resulting cosmic ray pressure (top center) and gamma-ray emission (bottom center) show the hourglass- and polar cap morphologies, respectively, which resemble those of supernova SN1006. Switching to a turbulent magnetic field yields a patchy cosmic ray distribution (top right) that is edge-brightened in the gamma-ray emission map (bottom right) and resembles observations of the supernova remnant Vela Jr.

Cosmic-ray acceleration at supernova remnants

When a massive star ends its life, it explodes as a supernova, driving a powerful shock wave into the surrounding interstellar medium. Most of the kinetic energy of the expanding shock is thermalized and heats up the initially cold interstellar medium to X-ray-emitting temperatures. However, recent plasma simulations have demonstrated that a small fraction of protons and ions that impinge on the shock are accelerated to extreme energies, thereby giving rise to a population of cosmic rays.

However, the efficiency of this acceleration depends on the orientation between the magnetic field vector and the direction of shock propagation – the so-called magnetic obliquity. If the shock propagates along the magnetic field (or moves at a narrow angle to it), particles that are reflected at the electrostatic shock potential can easily diffuse back into the region ahead of the shock, scatter there at electromagnetic fluctuations, and again be advected past the shock. At every shock crossing, each particle acquires a small amount of energy. It turns out that a particle can undergo many of these shock-crossing cycles until it eventually reaches extreme energies. Thus, such a quasi-parallel shock geometry maximizes the acceleration efficiency and converts around 15 percent of the available energy to cosmic rays. In contrast, a shock that propagates perpendicular to the magnetic field (or at a large angle to it) is inefficient in accelerating cosmic rays because any charged elementary particle is bound to gyrate around

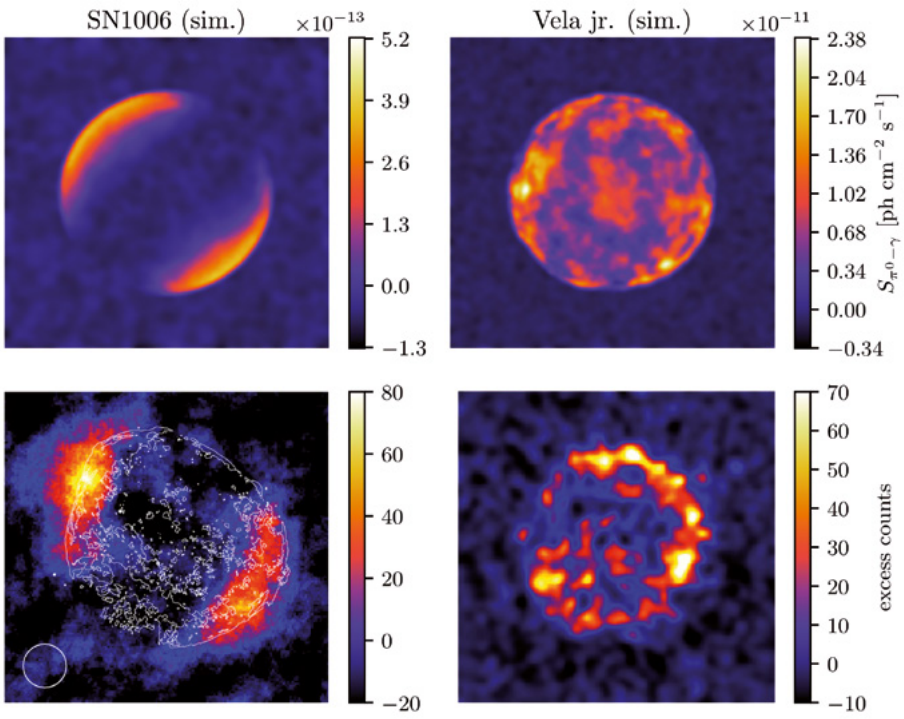


Figure 33: Comparison between simulated (top) and observed (bottom) gamma-ray emission maps in which we account for magnetic obliquity-dependent shock acceleration. Our model for SN1006 (top left) uses a superposition of a homogeneous magnetic field (initially oriented from top left to bottom right) and a mildly turbulent magnetic field with a coherence length equal to the box size. Our simulations match the TeV observations of SN 1006 (bottom left) in morphology and flux. In contrast, our model for Vela Jr. (top right) employs a purely turbulent magnetic field with a correlation scale equal to half the box size. The patchy gamma-ray morphology statistically resembles the TeV observations of SN Vela Jr. In the bottom panels, we display the acceptance-corrected smoother excess maps of HESS TeV gamma-ray observations.

a magnetic field, which itself is frozen in the plasma. As the magnetized plasma is swept past the shock, so is the particle, which has no chance of returning to the region ahead of the shock. Hence, this quasi-perpendicular configuration is unlikely to energize cosmic rays.

As the spherical supernova shock expands into the magnetized interstellar medium, it sweeps over patches with quasi-parallel and with quasi-perpendicular magnetic geometries. Consequently, the number of generated cosmic rays varies widely along the shock. We have devised a novel algorithm that identifies the shock in our simulation, measures its strength and the magnetic orientation, and accelerates cosmic rays according to the magnetic obliquity. The resulting cosmic ray distribution critically depends on the global magnetic geometry: A homogeneous magnetic field generates abundant cosmic rays at the quasi-parallel polar cap regions, while a turbulent magnetic field generates a patchy cosmic ray distribution (see *Figure 32*, from Pais, Pfrommer, Ehlert et al. in prep.).

These freshly accelerated cosmic rays collide inelastically with the ambient gas and produce gamma rays that can be detected on Earth. The gamma-ray emission resembles the cosmic ray morphology convolved with the gas distribution (which peaks at the shock). Thus, resolved gamma-ray observations of supernova remnants enable us to infer the otherwise unobservable magnetic field distribution with which the shock wave collided. Surprisingly, our simulations agree nicely with different TeV observations of shell-type supernova remnants (see *Figure 33*), which enables us to draw conclusions about the magnetic coherence scale in the local vicinity of these supernova remnants.

Cosmic rays in the interstellar medium

The space between stars is not empty; rather, it is filled with a diffuse medium out of which new stars are born and into which the products of exploding stars (such as heavy metals) are delivered once the stars have ended their lives. In this interstellar gas, there is only one cosmic ray particle for one billion low-energy atoms and molecules. However, as each cosmic ray particle is highly energetic, the total energy contained in cosmic rays is comparable to other forms of energies, be they thermal or magnetic. This observation opens the door to the exciting possibility that cosmic rays may be able to actively participate in the dynamical evolution of a galaxy.

Most of the cosmic rays are accelerated when massive stars explode as supernovae. Approximately 10 percent of the supernova energy can be converted into cosmic rays. The remaining 90 percent is available for the powerful explosion that directly stirs the interstellar medium and distributes the ejecta into the galactic disk. Hence, the generation of cosmic rays in supernovae shocks is closely related to the formation of (massive) stars.

Once cosmic rays have been produced, they can stream through the interstellar medium. An important transport feature of cosmic rays is their fast relative motion with respect to gas, which allows them to quickly escape from their production sites and spread throughout the galaxy on time scales that are shorter than dynamical time. Frequent scattering with the diffuse medium allows these cosmic rays to exert a force on the gas in the galaxy. Since they lose their energy very slowly – only 5 – 10 percent of the cosmic ray energy is cooled down – they can maintain this force, continuously accelerate the gas, and efficiently drive some fraction of it out of the galaxy. These outflows are observed ubiquitously but have only recently been successfully modeled in numerical simulations.

In our group, we investigate the impact of cosmic rays from their production at supernovae to their long-term effect on the structure of the galaxy and the gas distribution therein. We use hydrodynamical simulations in which we dynamically couple the cosmic ray pressure with the gas. By following heating and cooling processes in the interstellar medium, we aim to precisely investigate the evolution of different energy components. We are particularly interested in the relative dynamical impact of the thermal fraction of the supernova explosion and the cosmic ray contribution in shaping the galactic disk and driving outflows.

The direct explosion of the supernovae drives fountain flows of the gas and expels it temporarily from the galactic disk. However, by only employing the thermal supernova explosion, the gas would not leave the galaxy; rather, it would eventually return. The long-lived cosmic ray fluid plays a minor role in the dynamics within the disk but acts as the main driving force that accelerates the previously ejected gas. *Figure 34* displays the time evolution of the vertical

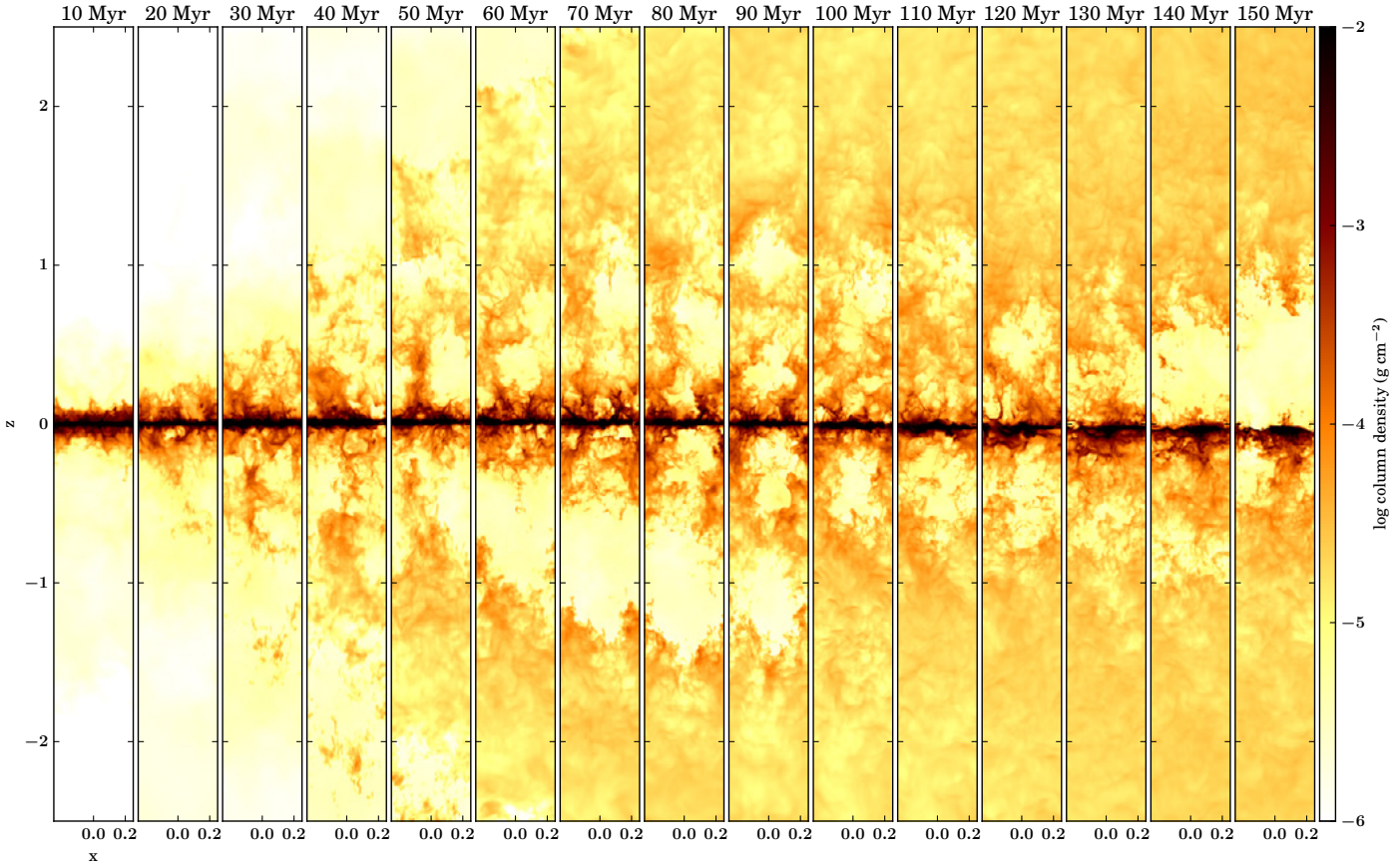


Figure 34: Time evolution of the vertical gas distribution. After approximately 40 Myrs, the supernovae begin to launch outflows with strong density contrasts up to a height of around 1 kpc. Above this height, cosmic rays dominate the energy content and further accelerate the gas. The outflowing gas is smooth (Girichidis et al. submitted).

gas structure in the disk. Most supernovae explode close to the midplane with altitudes of less than 0.1 kpc. An extended distribution of supernovae is distributed up to a height of approximately 1 kpc. The supernovae are the main driver that lifts the gas to heights of 1 – 2 kpc. This region of direct supernova impact displays strong density contrasts. Above a height of 1 – 2 kpc, cosmic rays continue to lift the gas. Due to their fast diffusion and smooth distribution, their force on the gas is uniform, which leads to a smooth outflow. The combined effect of direct supernovae and cosmic rays drives outflows with rates that exceed the star formation rate; that is, more than one solar mass of gas is driven out of the galaxy for one solar mass of formed stars. We find that the gaseous outflows begin with small velocities from the midplane of the galactic disk and gain speeds at altitudes of a few kilo-parsecs in height due to the continuous acceleration by cosmic rays. With terminal velocities of a few hundred km/s, the gas is able to escape the gravitational attraction of the galaxy. Overall, we find that cosmic rays are a dynamically important driver in the interstellar medium. The energy structure reveals three different regimes. In the disk, at heights below z = 0.1 kpc, the cosmic ray pressure dominates the thermal pressure due to the efficient cooling of the supernova energy and the negligible cooling of cosmic rays. We observe approximate equilibrium between thermal and cosmic ray pressure for the strongly structured gas up to a height of 1 kpc. Altitudes above 1 kpc lack the thermal energy input by supernovae and are again dominated by cosmic ray pressure.

Jets from super-massive black holes in galaxy clusters

Galaxy clusters consist of hundreds to thousands of galaxies. The most massive of these galaxies are giant elliptical galaxies, which lie at the very center of galaxy clusters. Each giant elliptical galaxy harbors a supermassive black hole at its center with billions to tens of billions of solar masses. The evolution of black holes is intimately and symbiotically connected to galaxy clusters, and both influence each other.

The gas at the cluster center is very dense and thus quickly loses energy, which implies the formation of stars and gas accretion onto the central super-massive black hole. As the cooling gas sinks to the center, it conserves angular momentum and spins up such that it joins the accretion disk around the black hole. Because the spin of a black hole is limited in general relativity, the accretion disk-black hole system is forced to either transport some of the angular momentum outwards in the accretion disk or eject a fraction of the fastest spinning gas in the form of jets along the rotation axis. Jets are launched back to back through magnetic torques in the immediate vicinity of a black hole and consist of magnetic fields and cosmic rays, which are accelerated at internal shocks in the jets.

Eventually, the ram pressure of the ambient cluster gas slows jets down and inflates lobes of relativistic plasma. As jets terminate, the lobes detach from the central region. Because these lobes are injected at the bottom of the gravitational cluster potential and are filled with relativistic particles, they are lighter than the ambient cluster gas on top of them. Thus, the lobes begin to rise buoyantly in the stratified cluster atmosphere. While the total available energy in these lobes is more than enough to offset the radiative cooling of the cluster gas, it is far from clear how exactly this energy heats the cooling cluster gas or whether it does so at all. This uncertainty constitutes the "cooling flow problem" in galaxy clusters. Moreover, the black hole horizon is more than one hundred million times smaller than the cluster radius, where the gas begins to cool and precipitates. How is the black hole accretion rate (which dictates the jet launching) tuned to the heating rate of the gas at large radii? Does this process yield a stable heating-cooling balance?

To answer these important questions, we simulate jets that interact with the magnetized, turbulent cluster gas. We account for a realistic density ratio of the jets to the cluster gas to predict cluster X-ray- and radio emissions (Ehlert, Weinberger, Pfrommer, et al. in prep.). *Figure 35* displays a snapshot of an upwards-rising underdense lobe filled with cosmic rays (central panel) and magnetic fields (right panel). While the magnetized cluster gas streams around the lobe, the magnetic field lines get stuck at the interface and drape around the lobe. This process stabilizes the lobe against hydrodynamical instabilities that would otherwise quickly disrupt it. As the flow of cluster gas converges behind the lobe, the magnetic field is compressed to form strongly magnetized filaments that morphologically resemble observed H-alpha filaments.

Over time, cosmic rays escape from the lobe and are conducted along the magnetic filaments towards the cluster center. They isotropize in the center and heat the cluster gas at a rate that exactly offsets cooling (see *Figure 36*). The upwards-moving lobe induces vortical gas motions that decay into smaller vortices, thereby injecting a small amount of turbulence. The kinetic motions in these vortices eventually dissipate and also heat the cluster gas.

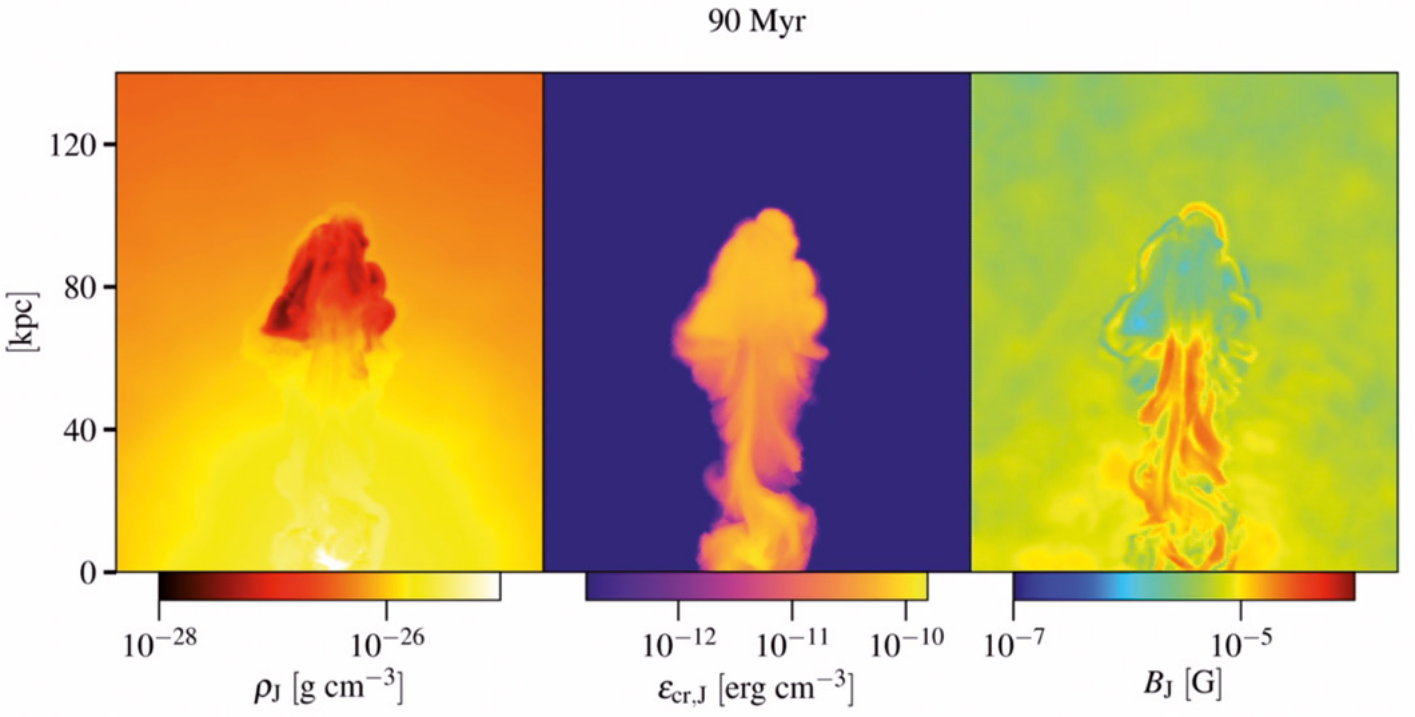


Figure 35: Rising lobe of relativistic plasma in a turbulent cluster atmosphere. We show cross-sections through the center of the 3D simulation of density (left), cosmic ray energy density (middle), and magnetic field strength (right).

However, turbulent heating is strictly confined to the wake because these swirling motions quickly decay long before they could be advected to the entire cooling region in the cluster core.

In fact, magnetic and kinetic turbulence are merely two sides of the same coin: Both processes necessarily accompany each other. Ubiquitous cluster phenomena such as orbiting galaxies or merging smaller clusters generate kinetic and magnetic turbulence, which fills the volume of the entire core region of the cluster and is not necessarily connected to the rising lobes. Moreover, the level of magnetic turbulence as inferred from observations of polarized radio emissions is consistent with the observed small motions that only cause a moderate broadening of X-ray lines.

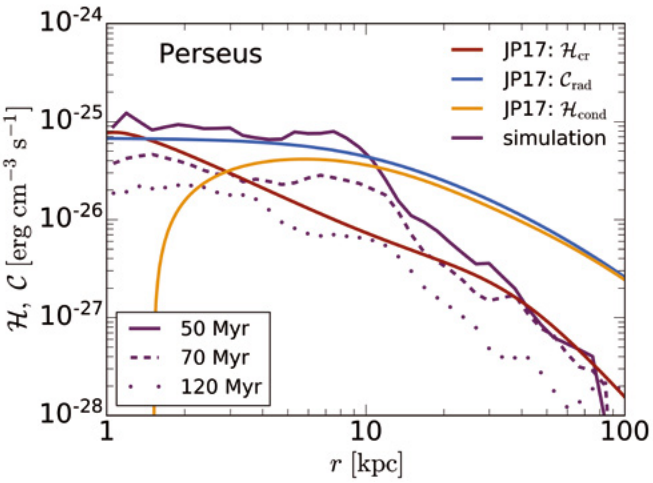
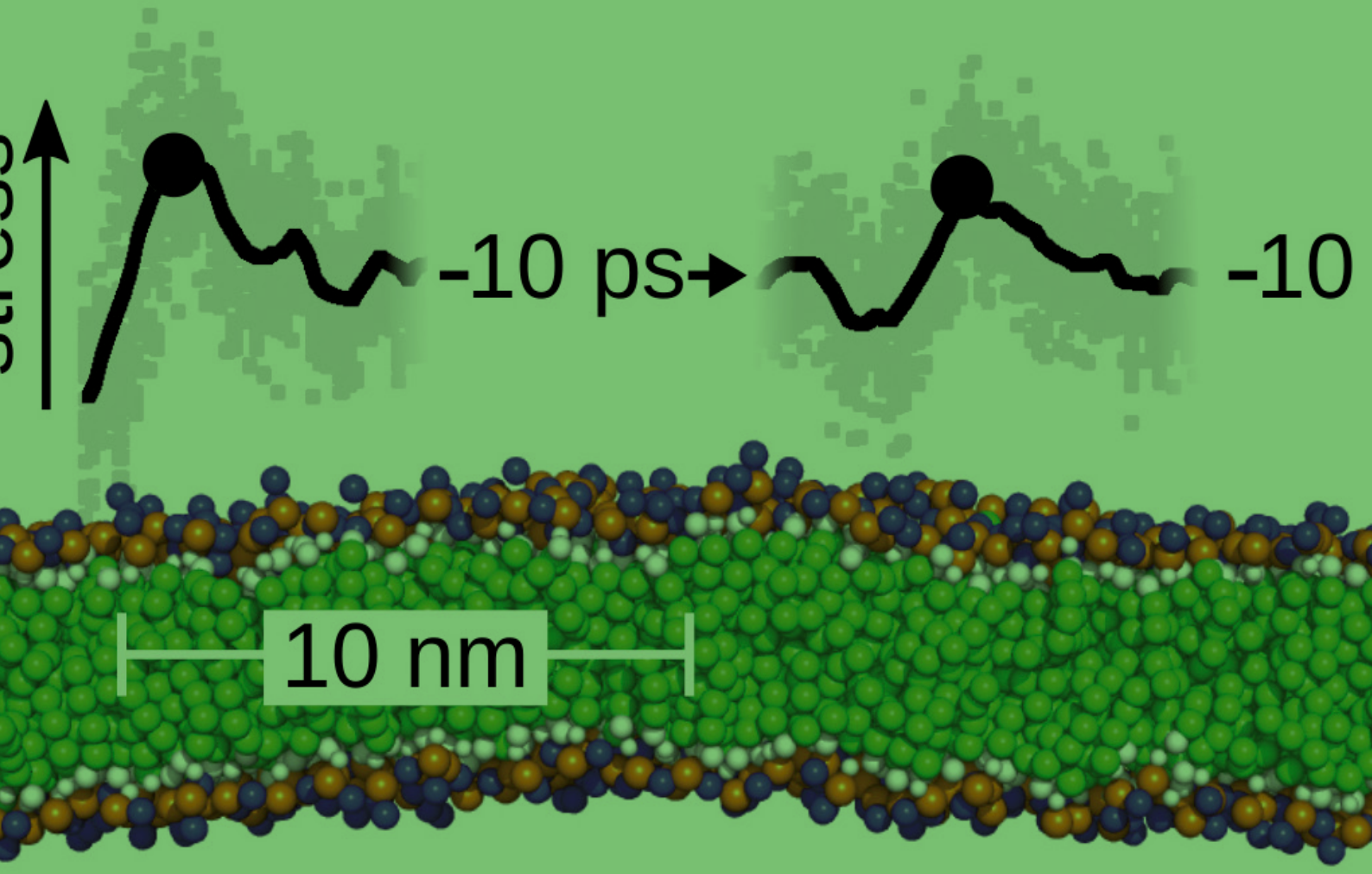
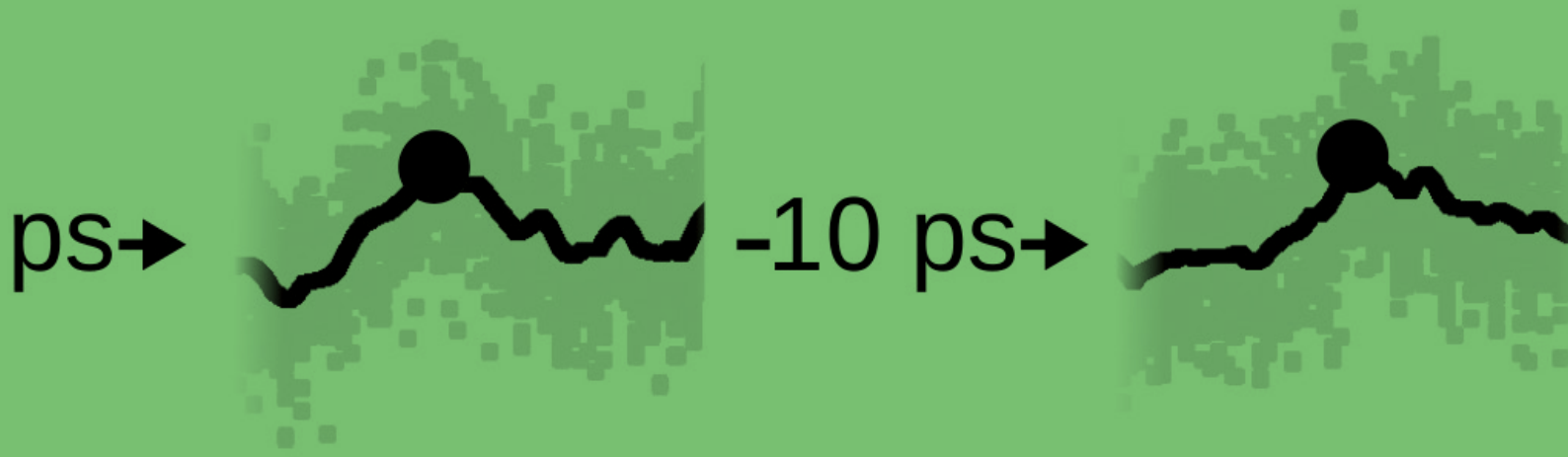
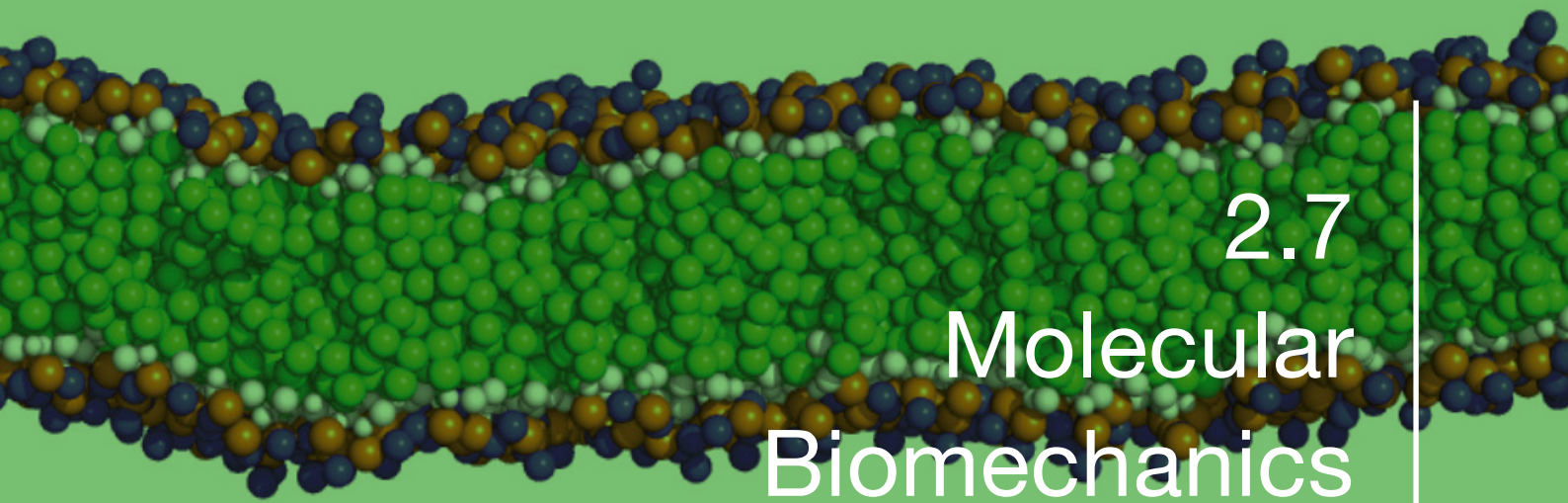


Figure 36: Radial profiles of the rate of radiative cooling vs. cosmic ray heating. We contrast our simulated averaged cosmic ray-heating profiles at different times (violet) to those of a radial, steady state solution for the Perseus cluster (red, Jacob & Pfrommer, 2017). At larger radii, thermal conduction heating (orange) balances radiative cooling (blue).

2 Research







2.7 Molecular Biomechanics (MBM)

2.7
Molecular Biomechanics (MBM)

Living organisms are incessantly under mechanical stress. Pushing, pulling, or shearing forces drive processes on all scales, from embryogenesis to locomotion. The MBM groups aims to uncover how mechanical forces are coupled with biological processes on the molecular scale. The molecules involved in such "mechano-sensation" are proteins, lipids, and DNA. To this end, we use quantum chemical calculations, Molecular Dynamics simulations on an atomistic and coarse-grained scale, and continuum mechanics methods to examine the behavior of these molecules under various types of mechanical perturbations.

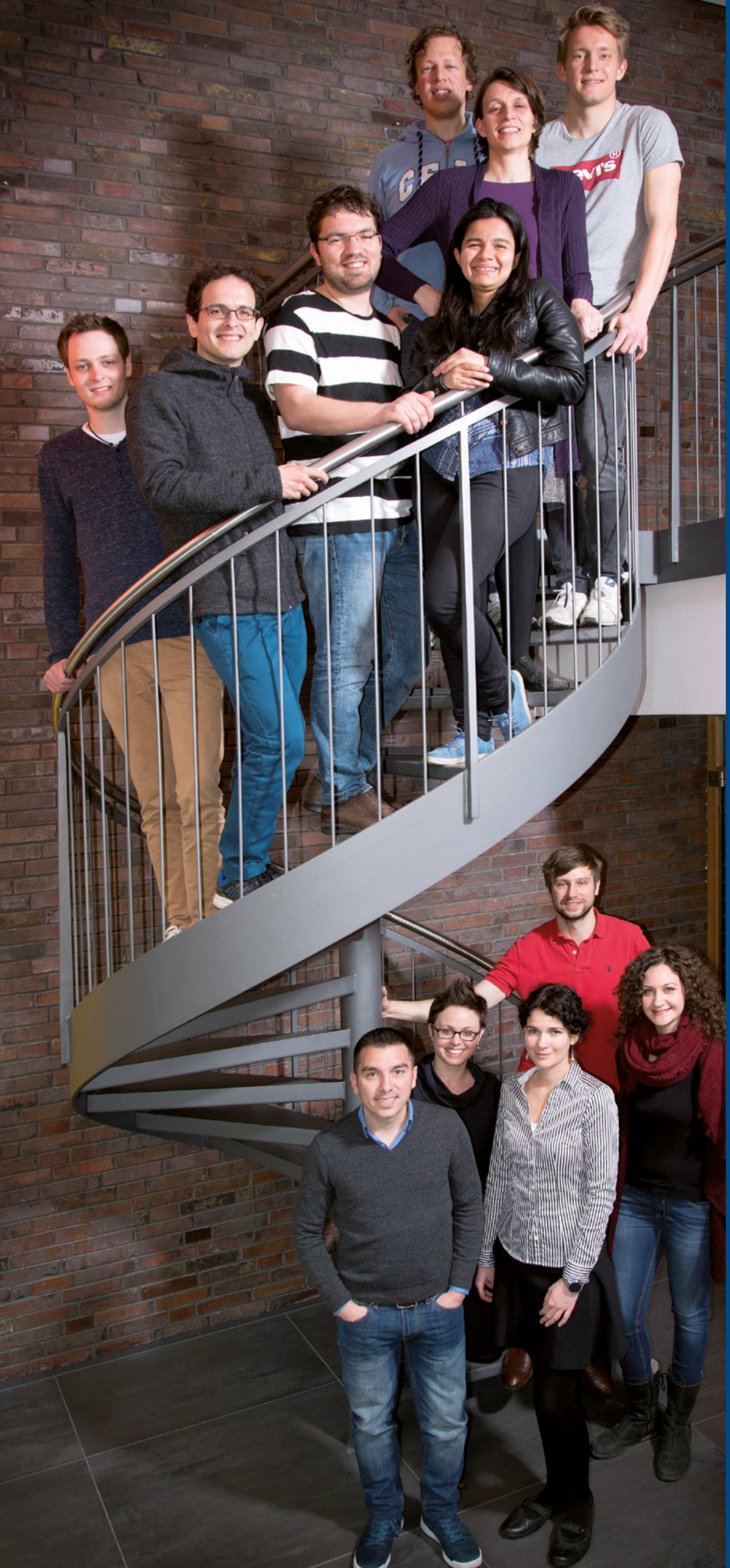
Lebende Organismen stehen ständig unter mechanischer Belastung. Druck-, Zieh- oder Scherkräfte treiben diese Prozesse in allen Größenordnungen an, von der Embryogenese bis zur Fortbewegung. Die MBM Gruppe will entdecken, wie mechanische Kräfte mit biologischen Prozessen auf der molekularen Ebene gekoppelt sind. Die Akteure dieser „Kraftmessung" sind Proteine, Lipide und DNA. Wir arbeiten mit quantenchemischen Berechnungen, Molekulardynamik-Simulationen auf atomaren und grobauflösenden Skalen sowie mit Kontinuumsmechanik, um das Verhalten dieser Moleküle unter verschiedenen Formen mechanischer Störungen zu untersuchen.

In 2017, we contributed to the understanding of mechano-sensation at cell-cell junctions as they occur in skin by studying the force-response of an important linker protein at these junctions – desmoplakin – within a German-wide priority program on this topic.

Im Jahr 2017 trugen wir zum Verständnis der Kraftmessung an Schnittstellen zwischen Zellen bei, wie sie in der Haut vorkommen: Wir untersuchten die Kraft-Antwort eines wichtigen Verbindungsproteins an diesen Stellen – Desmoplakin – im Rahmen eines deutschlandweiten Schwerpunktprogramms zu diesem Thema.

Any such linkers are ultimately also connected – directly or indirectly – to the cell membrane, which is thus also involved in force transduction both within and between cells. We were able to analyze how tiny force pulses can travel through the membrane very efficiently with surprisingly little attenuation. This and other related studies emphasize the theory that membranes might play a more important role in mechano-sensing than previously thought. Finally, we also made further progress in other related fields, including mechanochemistry and intrinsically disordered proteins, as described below.

Alle diese Verbindungsproteine sind letztlich – ob direkt oder indirekt – mit der Zellmembran verknüpft, die dadurch auch in die Kraftübertragung sowohl innerhalb als auch zwischen Zellen eingebunden ist. Uns gelang es zu analysieren, wie kleine Kraftimpulse wie Schallwellen sehr effizient durch die Membrane „reisen" können und sich dabei überraschend wenig abschwächen. Diese und andere damit verbundene Studien bestärken die Theorie, dass Membranen eine wichtigere Rolle in der Mechanosensorik spielen, als bisher angenommen wurde. Schließlich konnten wir auch in anderen verwandten Feldern Fortschritte verzeichnen, einschließlich Mechanochemie und ungeordneten Proteinen, wie weiter unten berichtet wird.



Group Leader

Prof. Dr. Frauke Gräter

Staff members

Dr. Camilo Aponte-Santamaría *(until March 2017)*
Dr. Csaba Daday
Florian Franz *(from August 2017)*
Ana María Herrera-Rodríguez
Dr. Katra Kolšek *(until January 2017)*
Fabian Kutzki *(from April 2017)*
Dr. Davide Mercadante *(until April 2017)*
Dr. Vedran Miletić
Dr. Fan Jin *(from July 2017)*
Dr. Agnieszka Obarska-Kosinska
Johannes Wagner *(until July 2017)*
Christopher Zapp

Students

Sara Becker *(until February 2017)*
Florian Franz *(until July 2017)*
Tobias Jäger *(from June 2017)*
Christopher Lance *(June – July and again from October 2017)*
Salvador Ramirez Ciarreta *(from March 2017)*
Steven Sheridan *(from October 2017)*
Max Simon *(April – August 2017)*
Jana Tünnermann *(April – June 2017)*

Guest scientist

Prof. Dr. Fei Xia *(January – February 2017)*

Swapping (sulfur) partners – Who, where and when?

K. Kolšek, C. Aponte-Santamaría, and F. Gräter

Chemical bonds within a protein are very stable. When a protein changes its struc-ture, the only labile chemical bonds are disulfide bonds formed by one type of amino acid, cysteine. Disulfide bonds are important for folding and can be prone to spontaneous rearrangements via the thiol-disulfide reaction. This reaction is very simple: Two sulfurs are bound to one another through a bond, and a third sulfur (thiol) attacks the disulfide bond and forms a new bond with one of first two, thereby kicking the other original sulfur out of the previous bond. This process results in a rearrangement by swapping the bond partners and leads to a state with the same energy as the original state for simple and symmetric systems. The situation become complex with more than one thiol and/or disulfide. In this case, many thiols could attack a number of disulfides in various ways, resulting in many possible situations that can change dynamically over time.

Thiol / disulfide exchange turns out to be an important process in various biological tissues. A prominent example is blood coagulation: One critical factor, the von Willebrand factor, is stretched by flowing blood and only then aggregates by interacting with itself and blood platelets. This aggregation also involves thiol/disulfide exchange reactions, resulting in tethering to other von Willebrand factor molecules instead of a tethering within the same molecule. Despite this importance, which has also only recently begun to be discovered, the molecular determinants of the disulfide bond shuffling remain largely unknown. As a proof-of-concept, we used an immunoglobin domain that undergoes unfolding when exposed to force (*Figure 37a*). Previous experiments have shown that once the immunoglobulin domain contains a number of free cysteines and a disulfide bond, it can show disulfide bond swapping during unfolding. After partial unfolding, one of the cysteines can attack the disulfide bond and form a new bond, resulting in further unfolding of the freed peptide chain (Alegre-Cebollada J, Kosuri P, Rivas-Pardo JA, Fernández JM (2011) Direct observation of disulfide isomerization in a single protein. Nat Chem 3(11):882–887). Single-molecule experiments have been able to measure the speed for all possible cases of thiol and disulfide exchange, and we asked if simulations can help explain the experimental data and thereby shed light on the underlying principles.

By using a novel hybrid of the Monte Carlo- and Molecular Dynamics methods, we were able to reproduce the experimental results, namely the chance of 32Cys to react with 55Cys, which is approximately 3 times higher when compared with 24Cys. We found that the most important determinant of the observed selectivity of this thiol-disulfide reaction was the accessibility of the disulfide bond by the attacking thiol (*Figure 37b*), which was determined by the intrinsic motions of the protein and not the nucleophilicity of the three reactive sulfurs.

As the most commonly used simulation method for proteins, Molecular Dynamics simulations cannot take covalent bond rupture into account and instead use simple unbreakable springs to describe a bond. We introduced a Monte Carlo step to the simulations described above in order to change the location of the bond during the Molecular Dynamics simulations. This recipe is much more efficient than quantum mechanical calculations, which can accurately describe the chemistry of the reaction, as has been shown by others and us in previous studies. Our new procedure represents a road towards simulating chemical events somewhere in a dynamic biomolecule during Molecular Dynamics simulations on the fly. *More information can be found in: [Kolšek, 2017].*

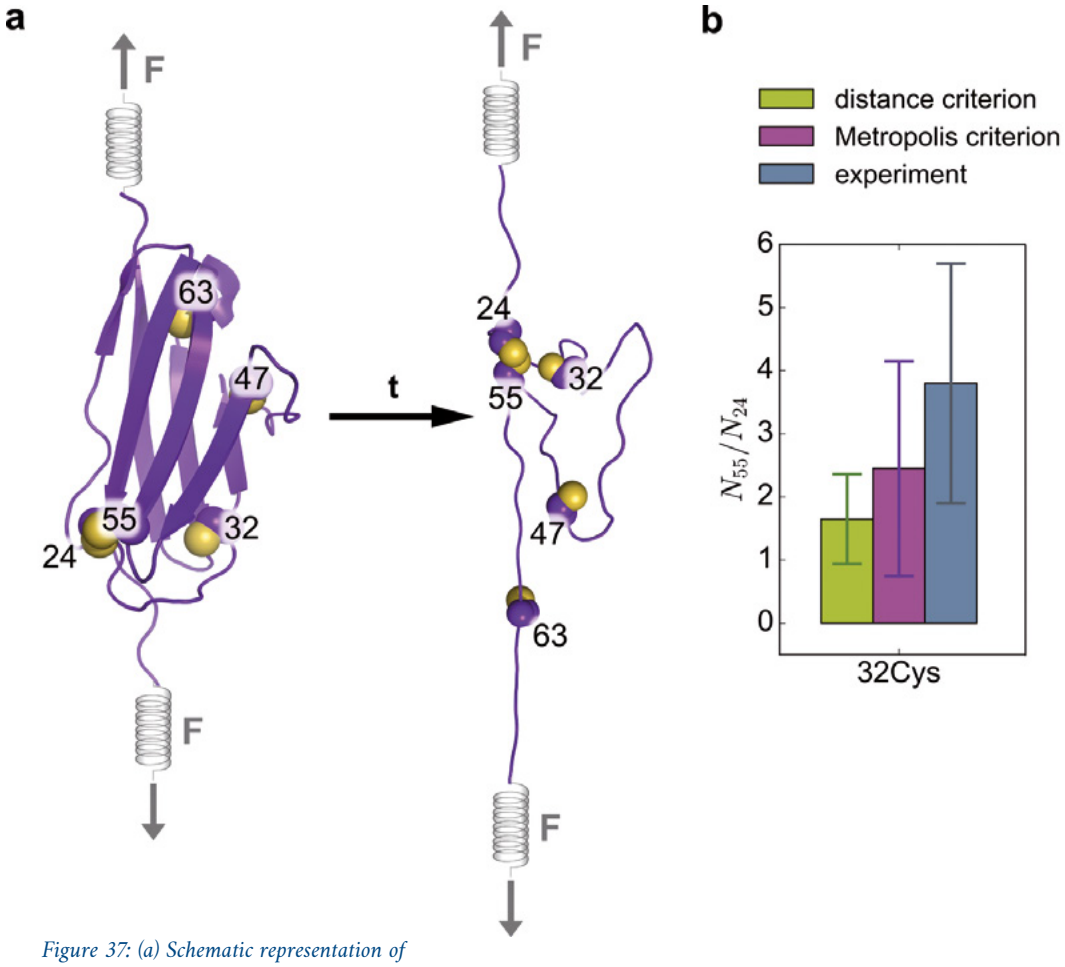


Figure 37: (a) Schematic representation of unfolding of mutated immunoglobulin domain under force with five cysteine residues (presented in purple spheres) and sulfur atom (in yellow sphere). (b) Ratio of 55Cys (N55) and 24Cys (N24) swap counts, calculated with a distance criterion (green) and with energy-based Metropolis criterion (magenta). The experimental ratios (gray) were taken from Alegre-Cebollada J, Kosuri P, Rivas-Pardo JA, Fernández JM (2011) Direct observation of disulfide isomerization in a single protein. Nat Chem 3(11):882–887. The qualitative agreement with experiments is evident. We can also observe that including energy criteria does not significantly changes the ratio, which implies that accessibility is the major determinant of reaction.

Disordered proteins: a challenge for the computer

Davide Mercadante and Frauke Gräter

One of the long-standing paradigms of biology is the relation between macromolecular structure and function. As an example, immediately after the resolution of the DNA structure, it appeared to be clear that the double helix of DNA served the purpose of both storing and transferring the genetic material of a cell from one generation to the next. For proteins, such a structure-to-function paradigm has been equally embraced by the scientific community and has proven valid for many years, as confirmed by the characterization of numerous structured proteins. Nevertheless, some proteins escape such a rule, for despite being functional, they lack well-defined structural elements but still mediate the functional recognition of other partners and are hence still functional. These proteins – dubbed intrinsically disordered proteins – are able to perform a variety of complex cellular functions despite their lack of a stable secondary structure, and they are involved in crucial cellular processes ranging from signaling to transport and transcription and co-translation, which makes them particularly important to investigate.

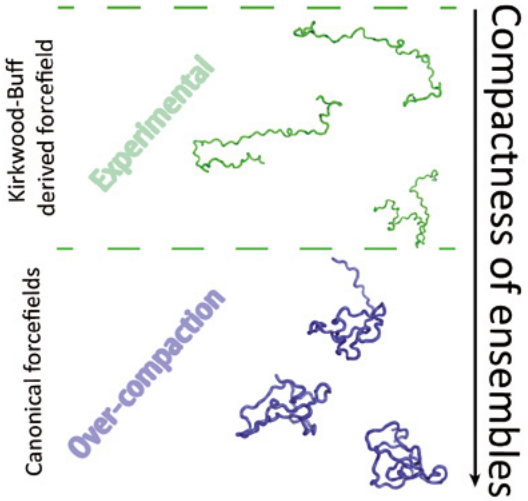
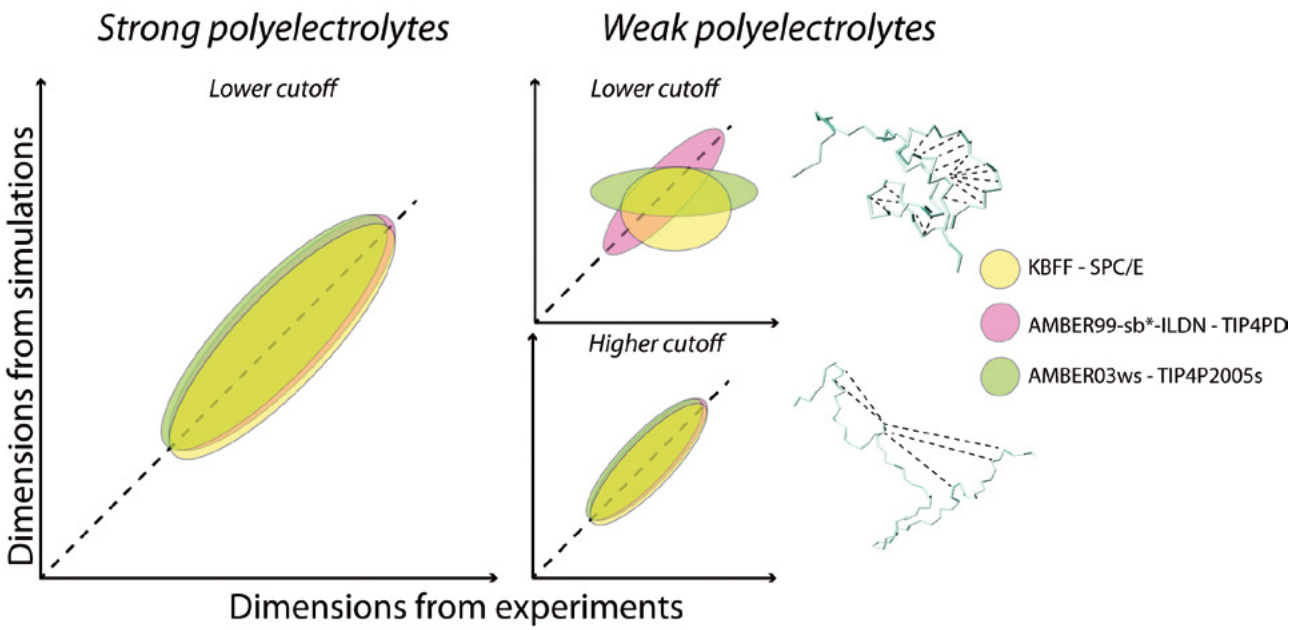


Figure 38: Schematic representation of the ability of the proposed KBFF to rescue intrinsically disordered proteins from collapse, which is not reported by experimental investigations. The Kirkwood-Buff force field is able to increase the dimensions of simulated, intrinsically disordered peptides and yield ensemble dimensions in line with experimental studies.

However, the lack of structure in IDPs makes such entities difficult to characterize because of their extreme conformational versatility and high dynamics in short timescales. The free-energy landscape of IDPs is different from that of structured proteins, for it has many shallow energy minima that describe a series of highly different conformations. Due to these properties, computational approaches are particularly suited to couple experimental techniques and to ultimately define IDPs' behavior. Nevertheless, molecular dynamics and Monte Carlo simulations aimed at sampling the conformational dynamics of disordered proteins have encountered enormous difficulties in sampling the correct conformational dynamics of IDPs and report considerable inconsistencies when comparing simulated and experimentally derived dimensions. Simulations reveal overly compacted conformers that can be ascribed to the impossibility of the force fields to correctly reproduce the dimensions of disordered chains and compact IDPs into globular conformations that do not comply with their functional conformational dynamics.

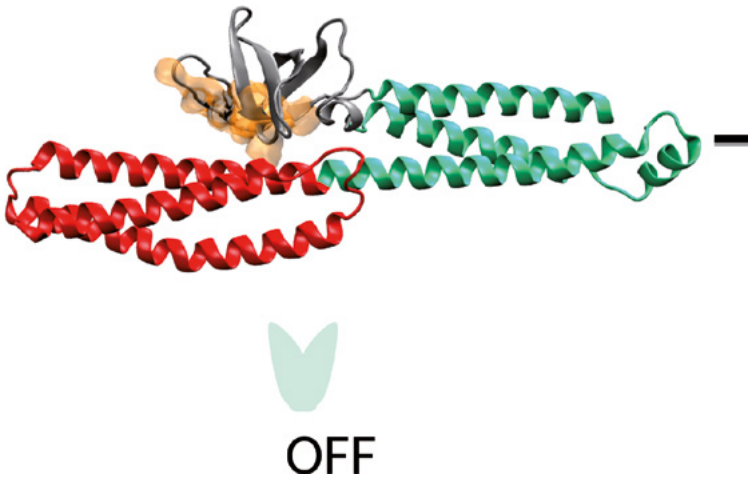
In previous studies, we have tackled the over-compaction of IDPs and proposed the sampling of these entities via an innovative approach: a force field named KBFF and developed on the basis of the Kirkwood-Buff force theory of solutions (*Figure 38*). With the advantage of describing the interactions between amino acids by using parameters obtained from a comparative match of thermodynamic quantities derived from experiments, KBFF is able to rebalance the equilibrium between protein-protein and protein-water interactions, hence sampling the correct overall dimensions of IDPs. In parallel to KBFF, other approaches aimed at scaling – either empirically or from a complete re-parameterization of explicit water – have also been developed and favor consistent improvements in MD simulations in sampling the dimensions of IDP ensembles.

Although KBFF (and the water-rescaling methods suggested to resolve the problem of IDPs over compaction) yield overall dimensions of the polypeptide chains that better comply with experimental findings, it is still unknown if intra-chain dynamics are correctly reproduced. We have recently contributed to the understanding of this design and performance of MD simulations in two distinct systems: 1) a strong polyelectrolyte composed of 30 amino acids and featuring either blocky or randomly distributed charged residues and 2) a fragment of nucleoporin 153 (Nup153), the dimensions of which were shown to be correctly reproduced using KBFF and that is representative, on the other hand, of a weak polyelectrolyte. We therefore rationally designed variants of this fragment characterized by mutations with an expected effect on the overall dimensions of the chains. We mutated all the phenylalanine residues in the Nup153 to proline and alanine, which resulted in reducing or expanding the dimensions of the polypeptide's ensemble. While all the tested force fields correctly reproduced the trends expected for strong polyelectrolites, the results from sampling weak polyelectrolites (such as the tested Nup153 fragment) were problematic.



Sampling the dynamics of such systems revealed that the force fields and water models that best solve the problem of over-compaction fail at reproducing the expected effects of the designed mutations on the ensemble's dimensions, which is also shown from smFRET experiments. Importantly, we discovered that the limitations of force fields can be bypassed if simulations are performed by choosing longer interaction cutoffs to sample intra-chain interactions. By increasing the coulomb and van der Waals cutoffs from 1.0 to 1.5 nm, simulations reproduce the trends described by the experiments with the ensemble of the mutated Nup153 and expand or contract if phenylalanine residues are mutated to alanine or proline, respectively (*Figure 39*). This simple modification reveals the importance of long-range interactions for the definition of intrachain dynamics and the overall dimensions of IDPs and paves the way for further computational studies in which longer interaction distances need to be considered in order to correctly sample dynamics. *More information can be found in [Mercadante, 2017].*

Figure 39: Schematic representation of the importance of long-range interactions in defining the dimensions of intrinsically disordered proteins for strong and weak polyelectrolites. The dimensions of highly charged proteins can be reproduced by all the force fields and /or water models suggested to solve the over-collapse of IDPs in simulations (left panel). Simulations of weak, intrinsically disordered polyelectrolites for which the effect of charges is negligible need longer interaction cutoffs to reproduce experimental dimensions.



How do cell-cell junctions sense force? Molecular dynamics clarified

Csaba Daday, Katra Kolšek, and Frauke Gräter

Cell junctions at the edges of tissues (e.g., focal ad-hesions) have been exten-sively studied, and the molecular mechanisms behind them are relatively well un-derstood. On the other hand, cell-cell junctions inside tissues are much more poorly documented despite their being more common in the body and being crucial to tissue stability. This state is mainly due to additional complexities, for measurements are much more difficult to set up and proteins often are hard to purify and solvate. Furthermore, creating model systems of these cell-cell junctions is much more challenging in vitro than are cell-extracellular matrix junctions. One of the classes of cell-cell jun-ctions is desmosomes, which are very common in skin tissue and heart muscle. Given the very strong mechanical stresses involved in the daily use of skin and in the heartbeat, it stands to reason that each cell in these tissues needs a mechanism to sense and react to force. However, due to the aforementioned difficulties in studying these systems, not much is known about how this force sensing happens.

We simulated the force response of an important structural protein in desmosomes, namely desmoplakin. The central plakin domain of desmoplakin is largely constituted by α-helical domains called spectrin repeats (SRs), but one of them is interrupted by a very different structure: a barrel-shaped SH3 domain. SH3 domains are known to be signaling actors (actors that send chemical messages by binding to other proteins and activating them), but the barrel interacts with the protein itself in the known structure (i.e., it is autoinhibited).

In our simulations, we found that the SH3 domain is freed under force, and we therefore hypothesized that it acts as a force sensor (*Figure 40*): When the protein – and hence, the entire cell-cell junction – is under stress, it becomes activated and is ready to send chemical signals to downstream partners. We repeated our simulations on several modified systems, including the functionally similar "cousin protein" plectin, which has an analogous role in the cell-extracellular junctions called hemidesmosomes. We found that the presence of the SH3 domain also stabilizes the inter-SR junction and that desmoplakin is activated under larger loads than plectin. Overall, our simulations provided very strong indications that this SH3 domain insertion has a signaling role and acts as an "emergency system" that is activated when the junction is under severe stress and when large parts of the protein have therefore been lost. Further clarification from experiments can help us, for example, by measuring the typical forces through these junctions (in general) and in desmoplakin (in particular). Comparing the loads on desmoplakin with those on plectin would also prove interesting. Our results predict that desmoplakin is tuned to higher forces, and this is therefore probably also the hierarchy in vivo. *More information can be found in: [Daday, 2017].*

Stress propagation through biological membranes in silico

Camilo Aponte-Santamaría and Frauke Gräter

Biological membranes are essential for any form of life and provide the casing for the precious molecules of life inside the cell.Membranes also contain many important molecules themselves, such as lipids (arranged as a bilayer) and proteins (embedded in such bilayers). Such diverse composition allows for tightly controlled information exchange with the out-

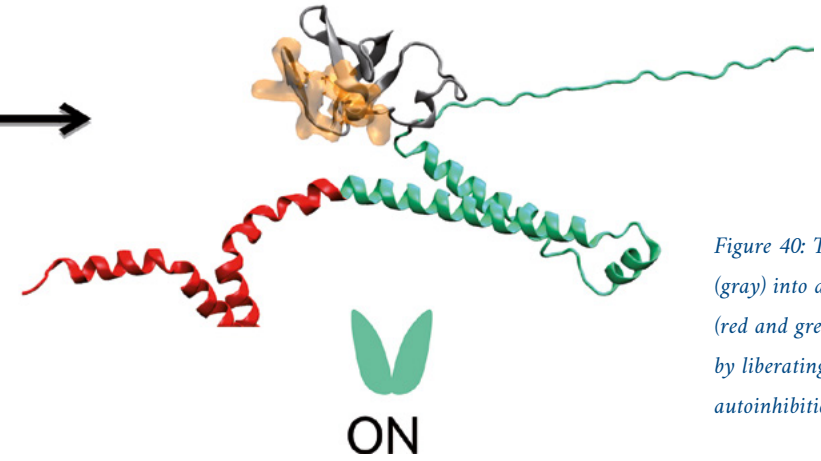


Figure 40: The insertion of an SH3 domain (gray) into a network of spectrin repeats (red and green). Force activates the protein by liberating the binding site (orange) from autoinhibition.

side. Cells are constantly being pushed and pulled by their neighboring cells while they divide, communicate, move, and die. Accordingly, biological membranes are subjected to mechanical stress while these processes occur.

Mechanical stress not only passively perturbs biological membranes but also critically influences their shape, elasticity, dynamics, and fusogenicity. Furthermore, mechanical stress has been found to trigger the opening and closing of mechanosensitive membrane-embedded ion- and water channels. In the latter case, it remained unclear whether the lipid bilayer alone provides a pathway to dynamically transmit the forces necessary to activate these channels in the absence of auxiliary scaffold proteins and – if so – how fast and how far membrane-mediated mechanical stress propagates.

We used molecular dynamics simulations to monitor the dynamics of lipid bilayers at quasi-atomistic coarse-grained levels of resolution and employed our in-house force distribution analysis method to extract the mechanical stress at every point of the bilayer and at every time of the simulation. We demonstrated that nanometer-wide pulses of mechanical stress efficiently propagate longitudinally through biological lipid bilayers (*Figure 41*). They travel at very high speeds (in the order of kilometers per second) in very good agreement with the experimental speed of sound observed for these biomaterials. Remarkably, the pulses traveled up to several tens of picoseconds, implying travelled distances of up to tens of nanometers before damping. These observations are highly robust regarding the simulation parameters and are consistent with the continuum visco-elastic models proposed for the propagation of acoustic waves through interfaces.

Our work supports the idea that the lipid bilayer can couple mechanosensitive elements in crowded biological membranes by itself by quickly transmitting localized pulses of stress. Rapid information transfer across cells is vital for the inner workings of our tissues, from the brain to muscles. We speculate that the bilayer confers an advantage by providing a narrow 2-dimensional medium for directed-stress propagation. The actual nature and role of such ultrafast information transfer, however, remains to be tested in future experimental studies. *Publication: [Aponte, 2017a].*

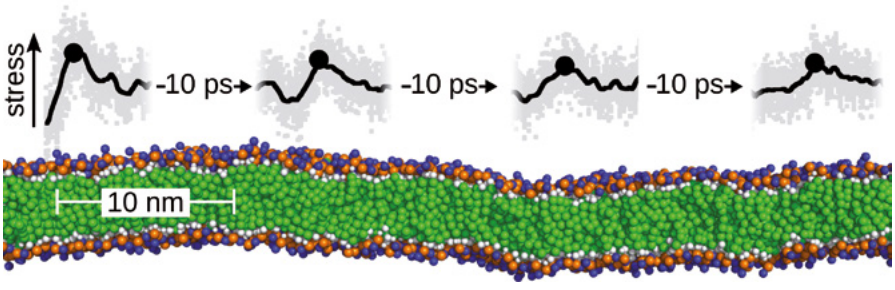
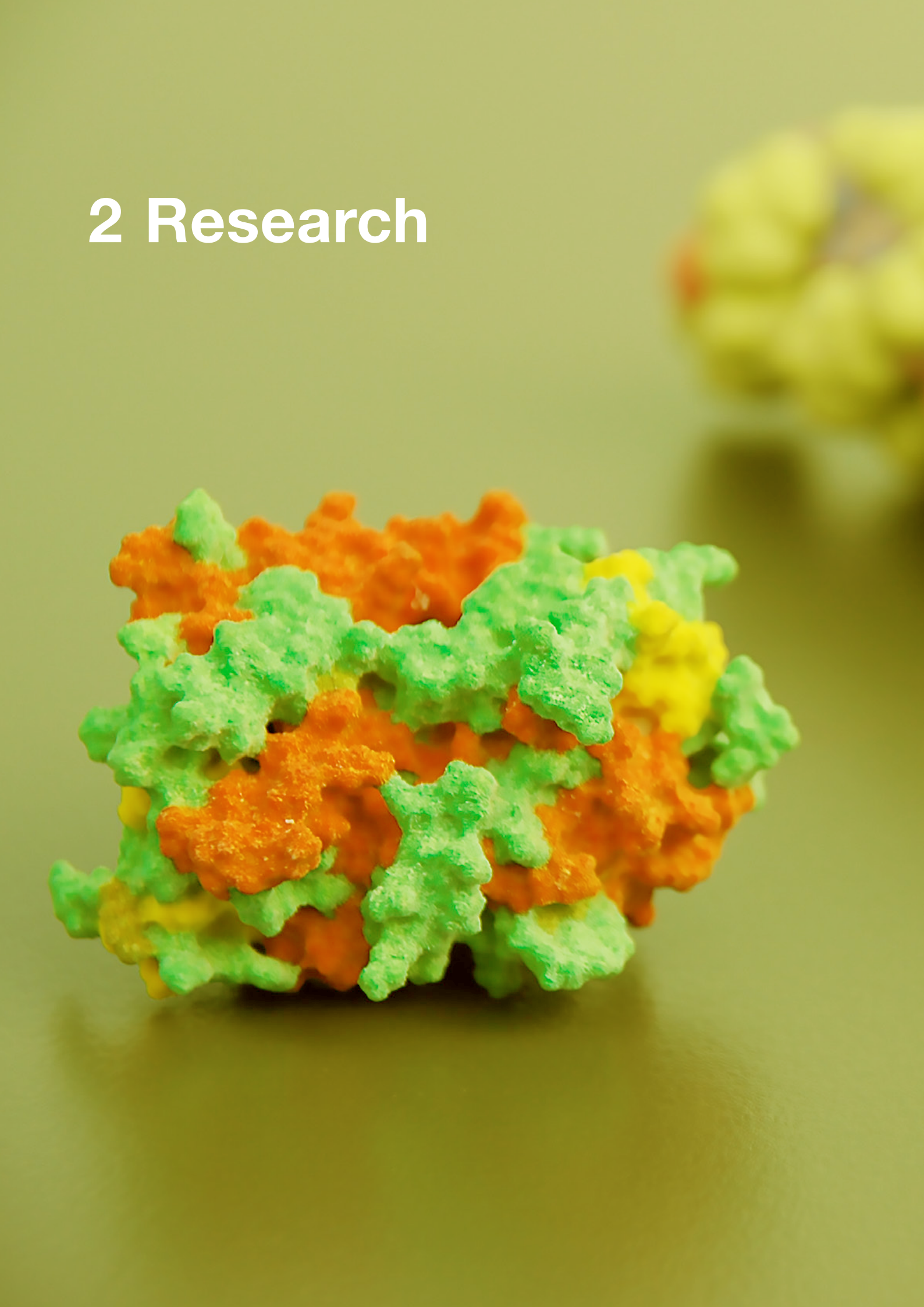
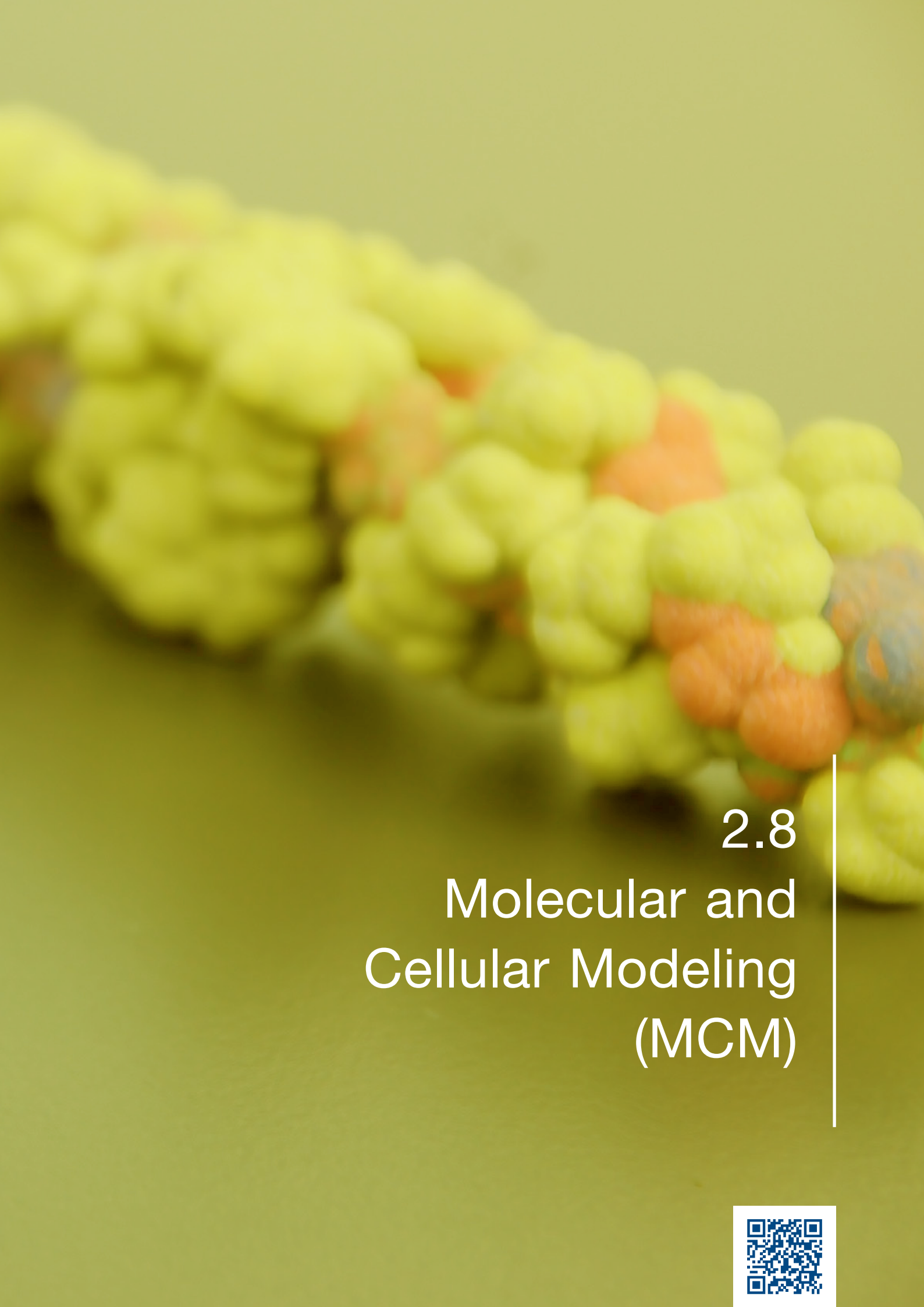


Figure 41: Propagation of mechanical stress through lipid bilayers recovered from coarse-grained molecular dynamics simulations. Localized pulses of stress (nanometers in width) longitudinally propagate through the bilayer (spheres). They efficiently move along the membrane at speeds in the order of nanometers/picosecond (km/s) during several tens of ps, implying several tens of nm before attenuation.

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 this 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 problems 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 collaboration 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 outline 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 focus on projects on (i) drug-target binding kinetics and protein binding pocket dynamics, (ii) structure-based drug design for neglected tropical diseases caused by trypanosomatid parasites, and (iii) electrostatic analysis of protein-protein interactions.

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 web-basierten 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 mit (i) der Wirkstoff-Protein-Bindungskinetik und der Dynamik von Protein-Bindungs-Stellen, (ii) strukturbasiertem Wirkstoffdesign für vernachlässigte tropische Krankheiten, verursacht von Trypanosomatidparasiten, und (iii) der elektrostatischen Analyse von Protein-Protein Wechselwirkungen.

Group Leader

Prof. Dr. Rebecca Wade

Staff members

Dr. Neil Bruce
Dr. Daria Kokh
Mehmet Öztürk *(since July 2017)*
Dr. Joanna Panecka *(until January 2017)*
Ina Pöhner
Dr. Stefan Richter
Dr. Kashif Sadiq

Scholarship holders

Gaurav Ganotra
Ghulam Mustafa
Mehmet Öztürk *(until June 2017)*

Visiting scientists

Dr. Goutam Mukherjee *(BIOMS fellow, Heidelberg University, since May 2017)*
Dr. Prajwal Nandekar *(Heidelberg University)*
Marcus Mendes *(PDSE-Capes Fellowship, Federal University of Rio Grande do Sul, Brazil, August – November 2017)*

Students

Thore Burgel *(May – June 2017)*
Carl Bürkel *(April – August 2017)*
Patrick Friedrich *(since December 2017)*
Max Horn *(until June 2017)*
Kai Horny *(until April 2017)*
Anika Liu *(March–April 2017)*
Rebecca Neil *(until June 2017)*
Martin Reinhardt *(until April 2017)*
Imme Roggenbach *(October – December 2017)*
Jui-Hung Yuan *(since July 2017)*

2.8 Molecular and Cellular Modeling (MCM)

Figure 42: MCM alumni meeting participants.

General news

In 2017, Dr. Goutam Mukherjee joined the group as a BIOMS postdoctoral fellow. Marcus Mendes visited for four months from the Federal University of Rio Grande do Sul, Brazil, with the support of a PDSE-Capes fellowship to learn how to apply our PIPSA methodology to his studies of molecular recognition in the immune system. Dr. Joanna Panecka returned to Warsaw to establish her own research project but continues to collaborate with us and has made several short visits to Heidelberg supported by BIOMS. Antonia Stank defended her doctoral thesis and began working in the pharmaceutical industry. Martin Reinhardt completed his master's thesis in Physics and has gone on to doctoral studies at the Max Planck Institute for Biophysical Chemistry in Göttingen. Rebecca Neil finished her Erasmus year with us and returned to Imperial College, London, to complete her B.Sc. in Biochemistry. Carl Bürkel successfully completed his bachelor's thesis in Molecular Biotechnology in the summer, and Ghulam Mustafa defended his doctoral thesis at the end of the year. Several master's students have done internships in the group throughout the year: Thore Burgel, Anika Liu, Imme Roggenbach, Jui-Hung Yuan (Molecular Biotechnology), and Kai Horny (Biochemistry). Two major projects involving large consortia came to an end this year: the EU FP7-supported NMTrypI (New Medicines for Trypanosomatidic Infections) consortium and the K4DD (Kinetics for Drug Discovery) project, which was a European Innovative Medicines Initiative (IMI). Both projects proved most successful, and some of the results are discussed below. With the support of the Heidelberg University Innovation Fund Frontier program, we also completed the funding phase of our project on structure-function relationships of actin filament stability in collaboration with Dr. Ross Douglas and Prof. Dr. Freddy Frischknecht at the Center for Infectious Diseases at Heidelberg University Medical School. We gained new insights into profilin-actin interactions in Plasmodium [Moreau, 2017] and identified key determinants of differences in the stability of mammalian and Plasmodium actin, which may provide avenues for the discovery of new anti-malarial agents.

We held an MCM alumni meeting on July 1st, 2017, in conjunction with the HITS alumni meeting. About 20 MCM alumni attended, and together with the current group members, we had a morning of diverse and stimulating scientific talks and a lively poster session followed by lunch, beer, and lots of catching up and reminiscing about old times (going back 25 years to when Rebecca established her research group at EMBL in 1992) (see *Figure 42*).

Drug-target binding kinetics and the dynamics of protein binding sites

During the K4DD (Kinetics for Drug Discovery; www.k4dd.eu) project, we focused on the development of computational methods to compute drug-target binding kinetic parameters as well as on joint experimental and computational studies aimed at gaining an understanding of the determinants of drug-target binding kinetics [Schuetz, 2017]. The computation of binding kinetic pa-

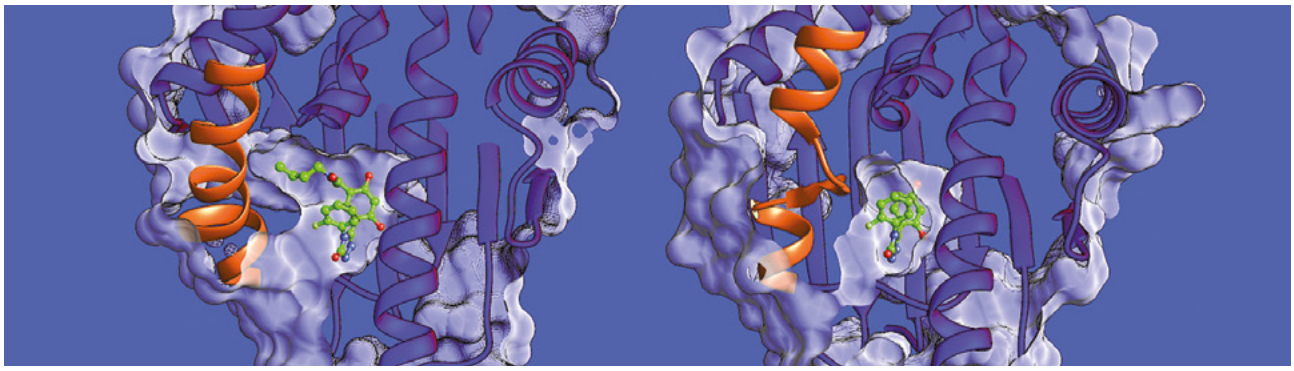
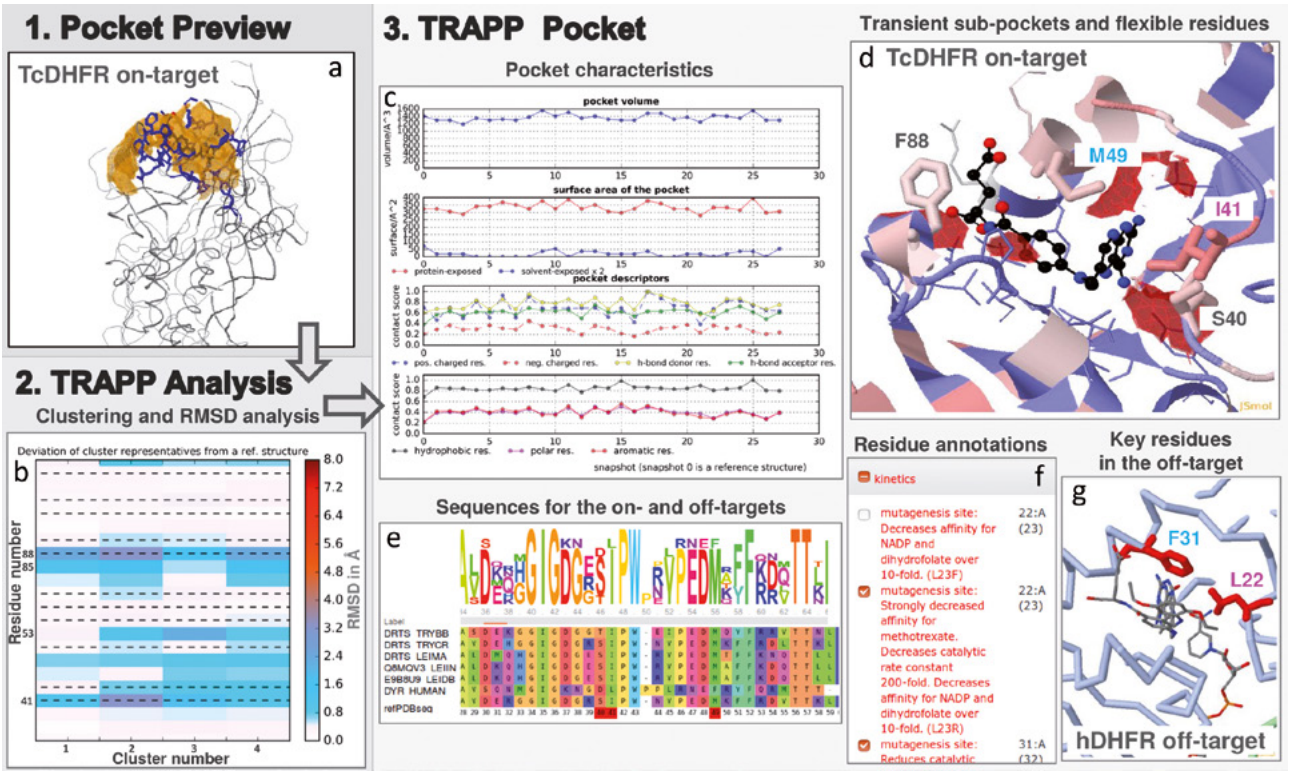


Figure 43: Greater binding site mobility leads to longer residence times of drug-like compounds binding to the cancer target: heat shock protein 90 [Amaral, 2017]. Differences in the structure and dynamics of the helical region (orange) of the protein when two different inhibitor molecules (green) are bound affect the shape of the binding pocket (shown by light-blue surfaces) and the rate of formation and dissociation of the complexes. The protein-inhibitor complex on the left has a longer lifetime and – surprisingly – its binding is entropically driven, with an important contributor being the greater mobility of the helical region in the bound complex.

rameters is generally more challenging than the more established computation of binding affinity because it requires the characterization of the transition state for binding or unbinding as well as the bound and unbound states of the ligand-receptor system. Moreover, drugs can have long residence times (extending to hours), and the computation of drug residence times is thus beyond the scope of conventional molecular dynamics simulations on microsecond timescales. Therefore, many novel methods of computing rate constants for receptor-ligand binding processes that seek to overcome these challenges have been developed over the last few years. We have recently reviewed and compared these methods [Bruce, 2018] . We have also released KBbox (kbbox.h-its.org), a webserver that provides a toolbox of computational methods for studying the kinetics of molecular binding. KBbox includes descriptions of methods and tutorials as well as guidelines to aid researchers in the choice of appropriate methods for their studies. KBbox includes some of the methods developed in the K4DD project, and more will be added as they are published. We are currently using the systematically measured datasets of drug-target binding kinetics measured by K4DD partners as a basis for developing and validating several new computational approaches to estimating binding kinetic parameters. The following computational methods are under development: τ-random acceleration molecular dynamics (τRAMD) for relative residence times, Brownian dynamics simulations for on-rates, and the chemometric method, Comparative Binding Energy (COMBINE) analysis, for receptor-specific models for off-rates.

One of the factors influencing drug-target binding kinetics is the dynamics of the protein binding pocket. Together with partners at Merck KGaA (Darmstadt) and the Instituto de Biologia Experimental e Tecnológica (iBET) (Lisbon), we carried out a detailed experimental and theoretical study of the binding of inhibitors to the N-terminal nucleotide-binding domain of heat shock protein 90 (HSP90) [Amaral, 2017]. The binding pocket is lined by a region that can take the form of a helix or a loop when bound to different inhibitors (see *Figure 43*). We found that compounds that bind when a helix is present do so for a longer time. Surprisingly, an important contributor to the long residence times was the greater mobility of the helical region of the binding pocket when the inhibitor was bound. This study thus suggests a new way of finding more effective drug candidates with optimal kinetic and thermodynamic properties, namely by considering less rigid protein targets and identifying molecules that stabilize more mobile forms of the protein upon binding.



To facilitate the study of protein binding pocket dynamics, we developed the TRAPP (transient pockets in proteins) webserver (http://trapp.h-its.org) [Stank, 2017], which we designed to aid the prediction of binding site flexibility and the detection of transient binding pockets. The webserver incorporates tools for sequence and structure analysis and for molecular simulation. An example of its application to a target for the design of anti-parasitic agents is shown in *Figure 44*.

Figure 44: Screenshot images from an example application of the TRAPP webserver (trapp.h-its.org) to the analysis of the binding site of a parasite drug target to identify transient subpockets that are selective with respect to the human homologue. Structures of Trypanosoma cruzi dihydrofolate reductase (TcDHFR) are analyzed to explore variation in the binding pocket (orange isocontours, a) with clustering of the structures into four clusters (b), computation of pocket characteristics (c), detection of transient subpockets (red isocontours, d) and residues that are not conserved between the on- and off-targets (TcDHFR and human DHFR (hDHFR) (d, e)), and annotations of the effects of mutating the corresponding residues in hDHFR on inhibitor binding and catalytic activity from our ProSAT+ webserver (f, g) (image from [Stank, 2017]).

Protein structure-based approaches to discovering new medicines for neglected trypanosomatidic infections

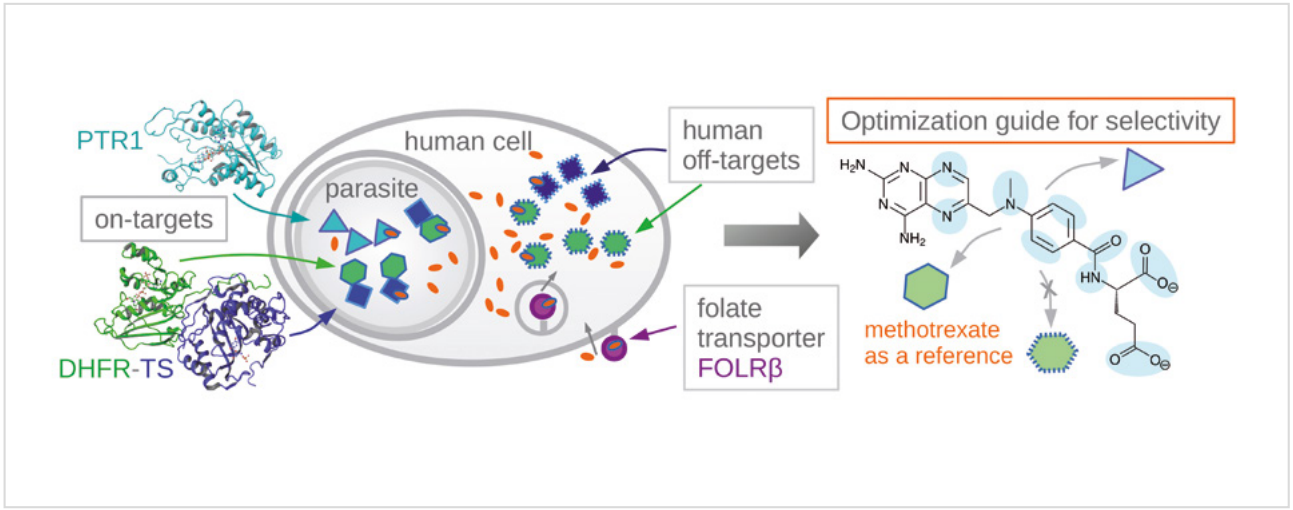


Figure 45: Computational mapping of on- and off-targets of the folate pathway provides a guide for the optimization of compounds for selectivity. Reprinted from [Panecka-Hofman et al., 2017], with permission from Elsevier.

In the NMTrypI (New Medicines for Neglected Trypanosomatidic Infections, http://fp7-nmtrypi.eu/) project, an international, multidisciplinary consortium employed phenotypic screening, ligand-based optimization, and structure-based design approaches to discover anti-leishmanial and anti-trypanosomal agents. Our contribution has primarily been to pursue protein structure-based approaches aimed at discovering compounds that selectively target the folate pathway in the parasites.

To provide a basis for structure-based design and optimization of enzyme inhibitors as anti-parasitic agents, we carried out a systematic comparative mapping of on-targets and off-targets in the folate pathway [Panecka-Hofman et al., 2017]. Our ligand-based comparative computational mapping methodology involved comparing the sequence, structural, dynamic, and physico-chemical properties of the target protein binding sites. The TRAPP webserver (*Figure 44*) and other computational tools were used for the analysis, which provided insights into which parts of protein binding sites can be used to design compounds to bind to multiple parasitic targets or to gain selectivity for the parasitic target relative to a similar human target. The webserver thereby provides a basis for the re-purposing of anti-folates and the design of new anti-trypanosomal agents (see *Figure 45*).

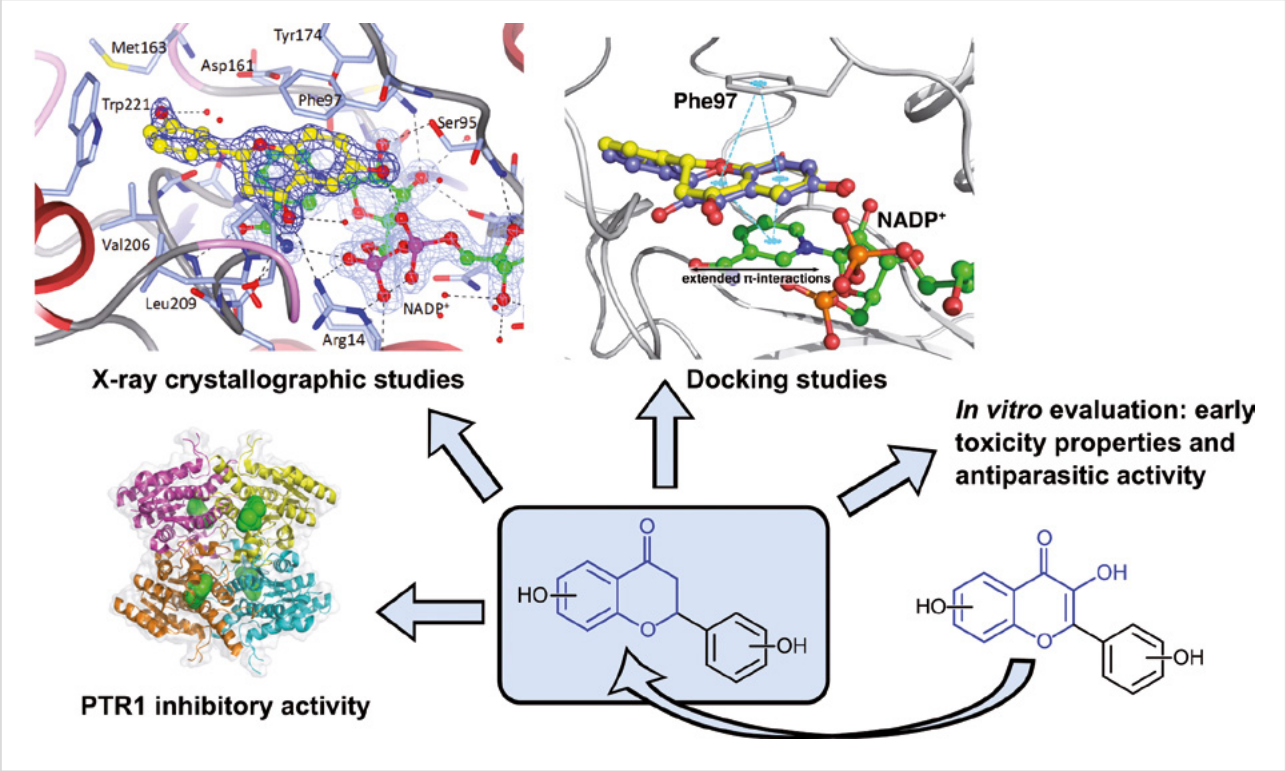


Figure 46: A concerted structure-based approach to optimize compounds identified from a screen of a library of natural products yielded compounds that bind the enzyme – pteridine reductase 1 – and display anti-parasitic activity. Adapted from [Di Pisa et al, 2017].

In the NMTrypI project, we used computational docking approaches to aid the design of compounds with different structural scaffolds that inhibit parasitic pteridine reductase 1 enzymes [Di Pisa, 2017, LInciano, 2017]. Compounds with new scaffolds were identified, and iterative rounds of optimization led to improved selectivity for established scaffolds. The computational work was performed in close collaboration with partners who determined crystallographic structures and measured the activities of compounds on the target enzymes and parasites (see *Figure 46*).

We also applied fragment-based drug-design approaches to the design of inhibitors of pteridine reductase 1 with the goal of identifying fragments that individually bind weakly to the protein target but yield a compound that binds tightly when connected. Computationally, it is difficult to correctly score the fragments, and standard empirical methods are often inadequate; therefore, we investigated a simple quantum chemical approach to ligand fragment scoring for Trypanosoma brucei pteridine reductase 1 inhibition [Jedwabny, 2017] and found that this approach performed well compared to ab initio quantum mechanical calculations and a commonly used empirical scoring method. We demonstrated the approach's potential for general application in fragment-based drug design approaches.

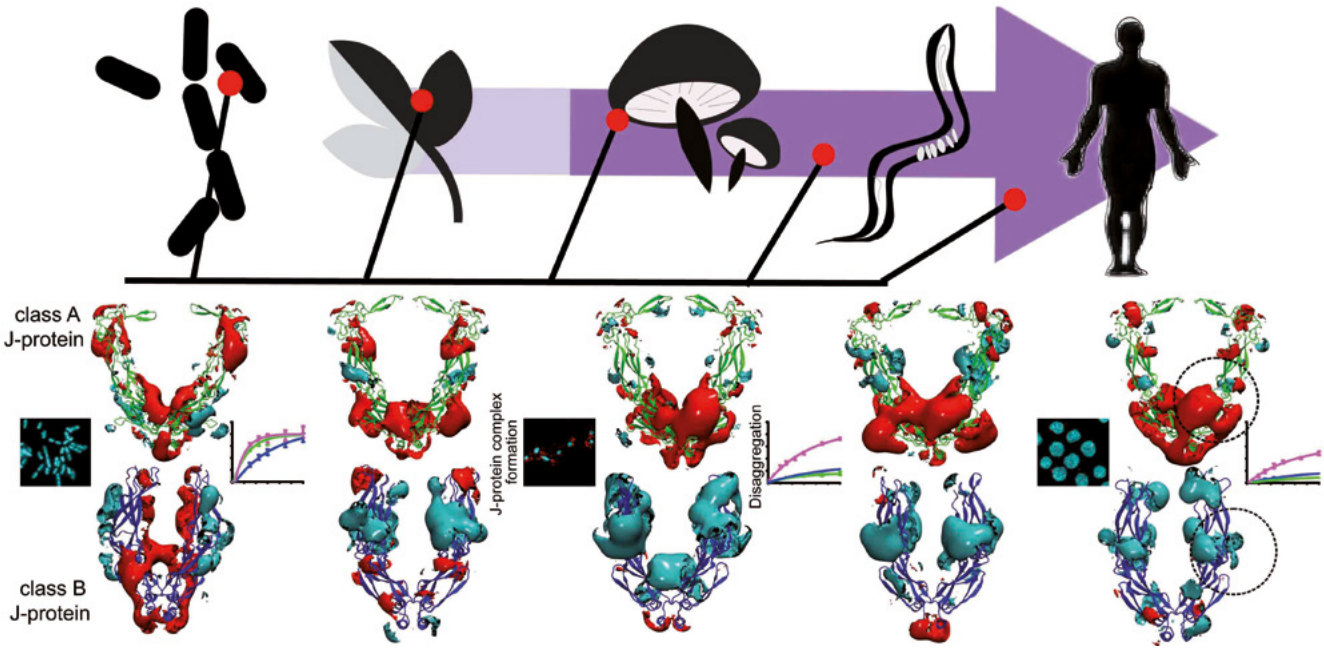


Figure 47: Evolution of the J-domain co-chaperone network for efficient protein disaggregation. Experiments and bioinformatics analysis reveal synergetic disaggregation by class A and class B J-proteins in eukaryotes but not in bacteria. Corresponding differences have evolved in the electrostatic potentials of the C-terminal domains of the J-proteins (green cartoon representation with positive (cyan) and negative (red) isopotential contours) at the proposed J-domain binding sites (dotted circles) [Nillegoda et al., 2017].

Learning from an electrostatics-centric view of protein-protein interactions

Electrostatic interactions play a key role in determining how proteins interact. In addition to positive and negative charges, the elements of molecular shape, flexibility, and polarizability all contribute to intermolecular electrostatic interactions. We apply the computation of molecular electrostatic potentials in bioinformatics analyses and in molecular simulations to investigate protein interactions. In a range of biological applications designed to study binding mechanisms, we identified specific and non-specific protein-protein interactions and a spectrum of ways by which electrostatic interactions contribute to biomolecular recognition, including domain-domain interactions, electrostatic steering [Öztürk, 2017], electrostatic patch complementarity [Moreau, 2017], charged binding motifs [Rafiullah, 2017], and phosphorylation-dependent allosteric binding.

In one project in which we studied the co-chaperone J-protein (DnaJ, HSP40) network with Bernd Bukau's group (ZMBH, Heidelberg University), we developed computational procedures to allow for the systematic analysis of weak transient interactions between flexible proteins with multiple binding sites. Experiments demonstrated a role for electrostatic interactions, and we found that the comparison of sequences, molecular electrostatic potentials, and Brownian dynamics docking allowed classes of J-protein to be distinguished and evolutionary differences to be detected. Our previous work had revealed that metazoan mixed-class J-protein complexes are crucial for efficient protein disaggregation activity in human and nematode HSP70 systems and that they are transient and involve complementary, charged regions. Complex formation allows J-proteins to initiate transient higher-order chaperone structures involving HSP70 and interacting nucleotide exchange factors. Extending the analysis to investigate the evolution of J-protein networks, we found that interclass J-protein networks can be formed by canonical class A and class B J-proteins in eukaryotes but not in bacteria (see *Figure 47*) [Nillegoda et al., 2017].

2 Research

2.9 Natural Language Processing (NLP)

2.9 Natural Language Processing (NLP)

Natural Language Processing (NLP) is an interdisciplinary research area 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.

In 2017, Sebastian Martschat successfully defended his thesis, entitled "Latent Structures for Coreference Resolution," at which point he had already begun a new job as a PostDoc at the University of Heidelberg. Moreover, our former PhD student Daraksha Parveen now works at Microsoft Research Hyderabad.

The other members of the NLP group continued their high-quality research, which led us to publish papers at all important conferences in the field (ACL, EACL, EMNLP, and IJCNLP). PhD. students in the NLP group are also encouraged to complete internships in the industry. Mohsen Mesgar concluded a six-month internship at SAP in Walldorf, Germany, in early 2017, and in the summer of 2017, he completed a three-month internship at Apple in Cupertino, California, USA.

After three years of contemplating the risks of NLP and discussing them with different audiences, group leader Michael Strube co-organized the "First ACL Workshop on Ethics in NLP," which was held in Valencia, Spain, during the EACL 2017 conference. The workshop hosted invited speakers and panelists from academia and industry and proved highly interdisciplinary, with speakers from different scientific fields, including NLP, artificial intelligence, philosophy, and law. The speakers were in different stages of their careers and ranged from well-established senior scientists to graduate students (see *Chapter 5.1.1*).

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

Im Jahr 2017 verteidigte Sebastian Martschat seine Dissertation zum Thema „Latent Structures for Coreference Resolution" erfolgreich. Zu diesem Zeitpunkt arbeitete er bereits als PostDoc an der Universität Heidelberg. Daraksha Parveen, vormals Doktorandin in der NLP Gruppe, arbeitet inzwischen für Microsoft Research Hyderabad.
Die anderen Mitglieder der NLP Gruppe setzten ihre hochqualitativen Forschungsarbeiten fort, was zu Publikationen auf allen wichtigen Konferenzen des Gebiets (ACL, EACL, EMNLP und IJCNLP) führte.
Die NLP Gruppe ermutigt grundsätzlich ihre Doktorandinnen und Doktoranden, Praktika in der Industrie zu absolvieren. Mohsen Mesgar verbrachte im Frühjahr 2017 ein halbjähriges Praktikum bei SAP SE in Walldorf (Deutschland), und im Sommer arbeitete er drei Monate als Praktikant bei Apple Inc. in Cupertino (Kalifornien, USA).

Nach drei Jahren intensiven Nachdenkens und nach Diskussionen mit unterschiedlichen Kreisen initiierte NLP Gruppenleiter Michael Strube 2017 als Mitorganisator den „First ACL Workshop on Ethics in NLP", der während der EACL-Konferenz in Valencia (Spanien) stattfand. Am Workshop nahmen Referenten und Podiumsgäste aus dem akademischen Umfeld und der Industrie teil. Das Treffen hatte stark interdisziplinären Charakter, mit Rednern verschiedener Wissenschaftsfelder, darunter NLP, künstliche Intelligenz, Philosophie und Recht. Die Palette der Referenten reichte von etablierten Wissenschaftlern in leitenden Funktionen bis hin zu Masterstudierenden (siehe Kapitel 5.1.1).

Group Leader

Prof. Dr. Michael Strube

Staff members

Alex Judea *(since October 2017)*
Dr. Mark-Christoph Müller

Scholarship holders

Alex Judea *(until September 2017)*
Mohsen Mesgar
Nafise Moosavi
Feifei Peng *(until August 2017)*

Visiting scientist

Benjamin Heinzerling *(Research Training Group AIPHES)*

Students

Leo Born
Raphael Schumann
Julia Suter
Isabell Wolter

What is wrong with state-of-the-art coreference resolvers, and how can they be addressed?

Coreference resolution refers to the task of finding different expressions in a text that refer to the same real-world entity. For instance, there is a coreference relation between "it" and "Ethernet cable" in (1).

(1) I want to order a long Ethernet cable.
It shouldn't be shorter than 4m.

There have been impressive improvements in the performance of coreference resolvers in recent years. For instance, from 2011 to 2017, performance has improved by more than ten percent. The problem is that all this improvement is based on evaluations of a single standard test set. If we apply recent coreference resolvers to new test data, the large gap between the performance of a 2017 neural network-based system compared with a 2011 rule-based system disappears.

For instance, Example (2) reveals the output of the state-of-the-art system in a commonly used evaluation set in which the detected coreferring relations are marked in blue. The system correctly recognizes that "the guerrillas" refers to "the country's leftist rebels."

(2) El Salvador's government opened a new round of talks with the country's leftist rebels. A spokesman said the guerrillas would present a cease-fire proposal.

Example (3), on the other hand, reveals how this same coreference resolver performs poorly in Example (1). Example (1) is considerably simpler than Example (2) but does not appear in the above-mentioned dataset.

(3) I want to order a long Ethernet cable.
It should n't be shorter than 4m.

We investigate the reasons behind this substantial difference between the performance of coreference resolvers in the standard evaluation set and in new domains and have found (a) that there is great overlap between the training and test datasets that are used for standard coreference evaluations and (b) that more complex coreference resolvers are biased towards resolving expressions that are seen during training. Therefore, these systems do not generalize well to unseen expressions.

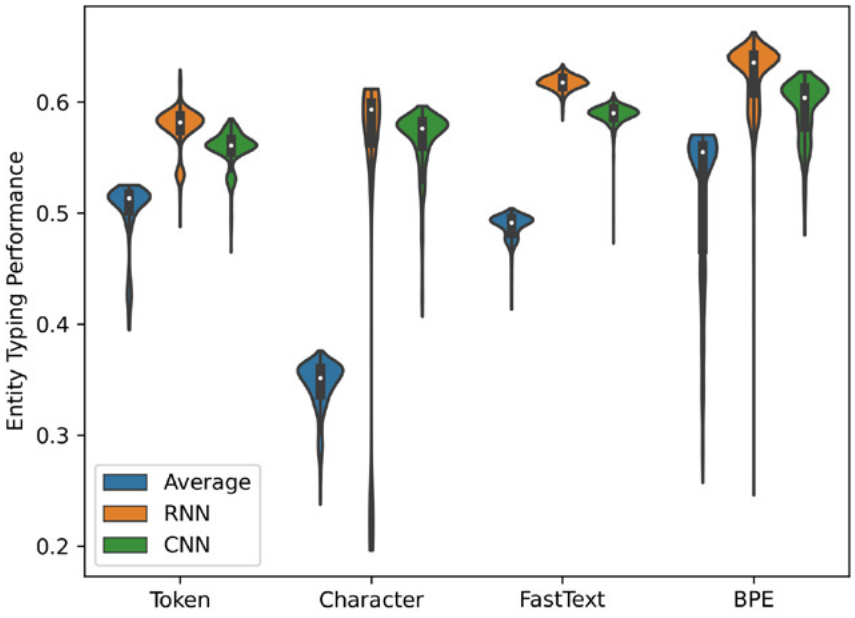
In earlier systems, coreference relations were represented by several heuristic features that were designed based on linguistic intuitions. However, recent research has abandoned all linguistic features. Instead, the systems automatically learn the representation of coreference relations solely based on strings. According to our experiments, the incorporation of linguistic features into state-of-the-art coreference resolvers is a promising method for improving generalizations across various domains.

Subword embeddings in 275 languages

Learning good representations of rare words or words not seen during training at all is a difficult challenge in natural language processing. As a makeshift solution, systems have typically replaced such words with a generic UNKNOWN token. Recently, based on the assumption that a word's meaning can be reconstructed from its parts, several subword-based methods have been proposed for dealing with the unknown word problem: character-based recurrent neural networks (RNN), character-based convolutional neural networks (CNN), word embeddings enriched with subword information (FastText), and Byte-Pair Encoding (BPE), among others. While pre-trained FastText embeddings are publicly available, embeddings for BPE units are commonly trained on a per-task basis (e.g., a specific language pair for machine translation) and

not published for general use. To address this problem, we published BPEmb, a collection of pre-trained subword embeddings in 275 languages based on Byte-Pair Encoding.

Byte Pair Encoding is a variable-length encoding that views text as a sequence of symbols and iteratively merges the most frequent symbol pair into a new symbol. For example, encoding an English text might first consist of merging the most frequent symbol pair "t h" into a new symbol "th" and then merging the symbol pair "th e" into "the" in the next iteration, and so on. The number of merge operations determines whether the resulting encoding creates mostly short character sequences (e.g., 1,000 merge operations) or whether it includes symbols for many frequently occurring words (e.g., 30,000 merge operations). Since the BPE algorithm works with any sequence of symbols, it requires no preprocessing and can be applied to untokenized text. We apply BPE to all Wikipedia editions of sufficient size and pre-train embeddings for the resulting BPE symbol by using the word embedding algorithm GloVe, which results in byte-pair embeddings for 275 languages.



An evaluation using entity typing as a testbed reveals that BPEmb compares favorably with FastText (a state-of-the-art approach that combines embeddings of tokens and subword units) as well as with character embeddings (*Figure 48*). Entity typing involves assigning a semantic category to mentions of entities (e.g., /location to Melfordshire or /other/health/malady to Myxomitosis). This task is well suited for evaluating subword models since subwords often allow for inferring a word's semantic category, even if the whole word has never been seen before: The suffix -shire hints at a location, the suffix -osis at an abnormal state or sickness.

Figure 48: English entity typing performance of subword embeddings across different architectures. The violin plot shows smoothed distributions of the scores obtained during a hyper-parameter search. White points represent medians, and boxes represent quartiles. Distributions are cut to reflect the highest and lowest scores.

Event extraction

Automatic event extraction is a challenging information extraction task. Systems need to find not only what kind of events are mentioned in a sentence but also the entities, places, and times that play a role in these events. Consider the following example: "Bush returned from a summit in Northern Ireland." Here, the word "returned" indicates the mention of a transport event. Furthermore, "Bush" is the artifact being transported, and "Northern Ireland" is the origin of the movement. Note that neither the vehicle used for transportation nor the destination of the movement is specified here. An event extractor needs to automatically recognize all of this information. While this example is rather simple, event structure can become quite complex. It is possible to mention multiple entities, places, or times that fill one role (e.g., multiple artifacts being transported), or one entity could fill different roles in different events (e.g., the target of an attack may also be the victim of a death).

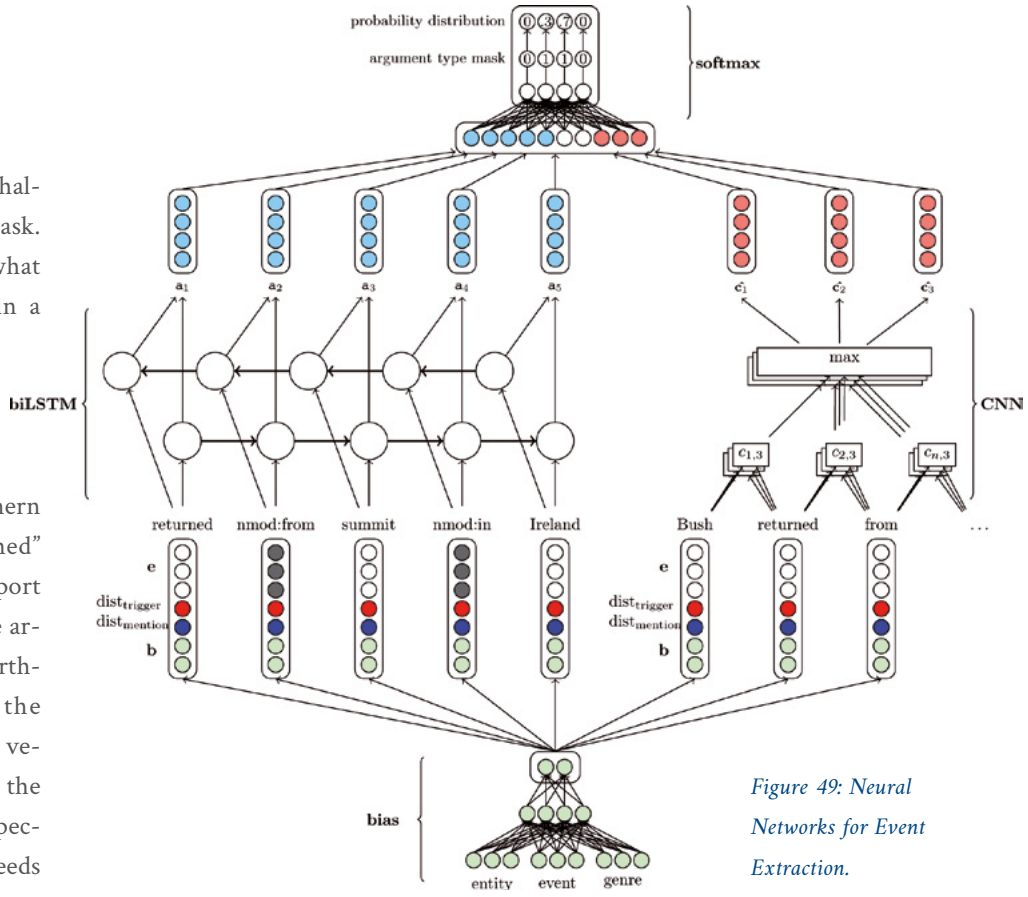


Figure 49: Neural Networks for Event Extraction.

Within event extraction, the extraction of arguments – the entities, places, and times that play a role – turns out to be very difficult. We examined the involved phenomena and found, inter alia, syntactic complexity to be of crucial importance. Argument extraction is relatively easy for arguments that are close to the words that most clearly express the event within a sentence (the triggers) but difficult when arguments are far away from the triggers. The problem is diversity: If arguments are close, the syntactic structure and the words that connect them to their triggers are short and low in number, and they can be learned from a set of event-annotated documents. If arguments are distant, however, the syntactic structures involved – and even the words that stand between the argument and the trigger – are very diverse, and most of them are never seen during training.

We decided to use methods introduced in the renaissance of neural networks to better handle unknown and diverse structures for argument extraction. Our system [Judea 2017] relies on a special kind of neural network, namely a Recurrent Neural Network (RNN) that directly operates on lexicalized syntactic paths (i.e., both the paths of syntactic relations between words and the words themselves). These paths connect triggers and potential arguments. The RNN learns what such paths look like in the training data. They can also learn which parts are most likely irrelevant for argument extraction. RNNs ultimately produce representations of syntactic paths in a very high-dimensional space in which it is easy to correctly classify them.

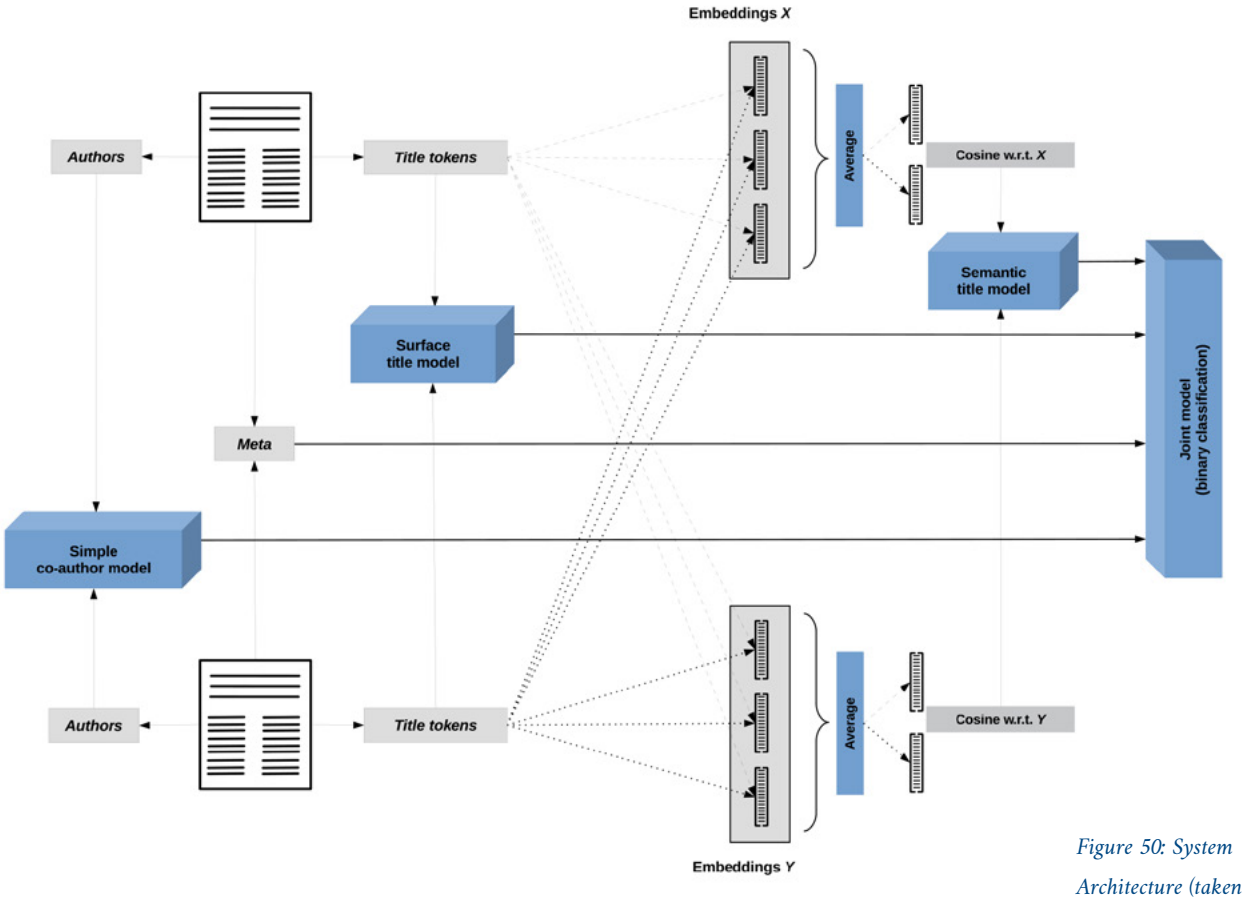


Figure 50: System Architecture (taken from [Müller 2017b]).

Author disambiguation

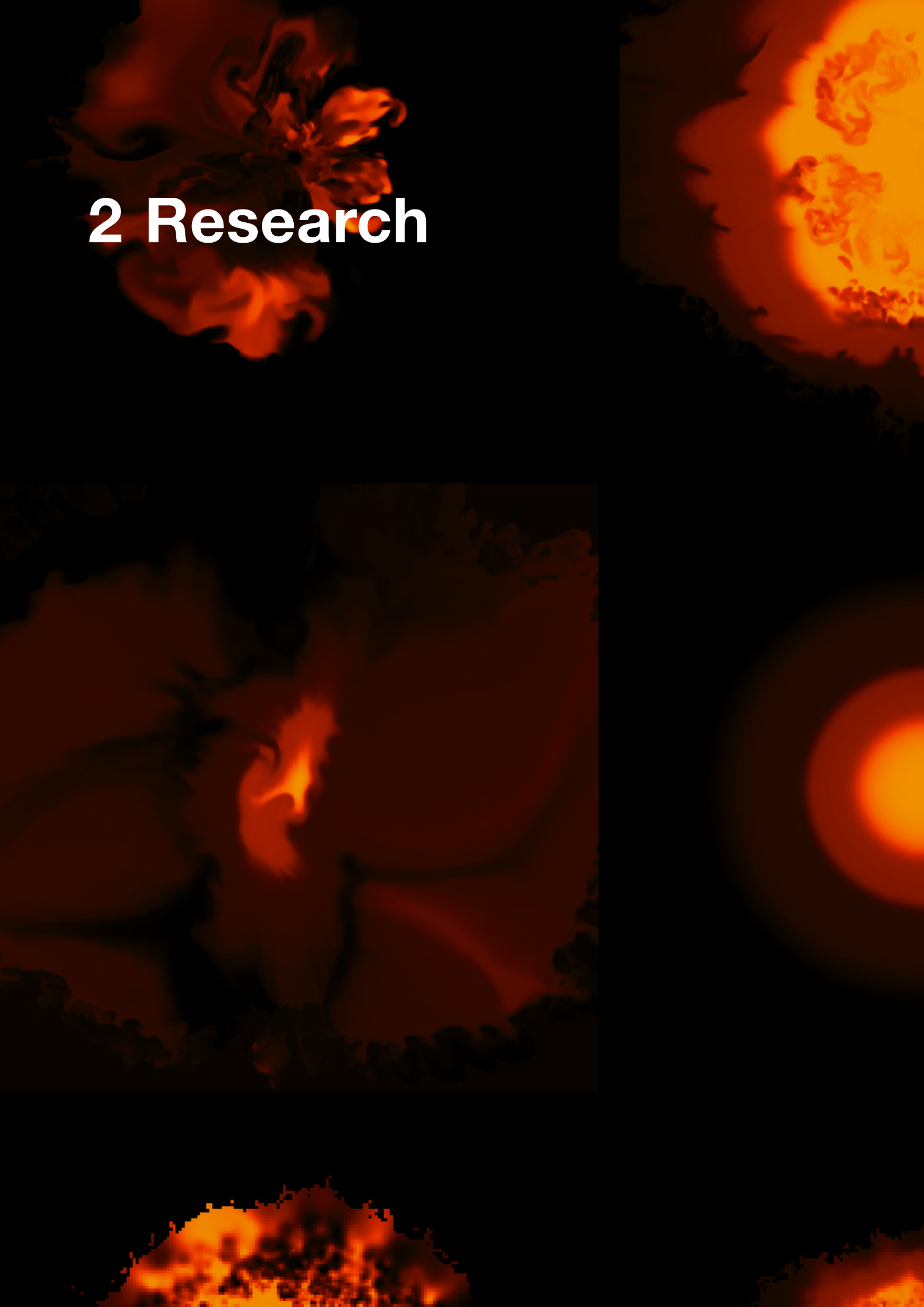
Research on author disambiguation at HITS was conducted as part of the project "Scalable Author Disambiguation for Bibliographic Databases" (SCAD) in cooperation with DBLP and zbMATH and was funded by the Leibniz-Gemeinschaft. 2017 was the second year of the project.

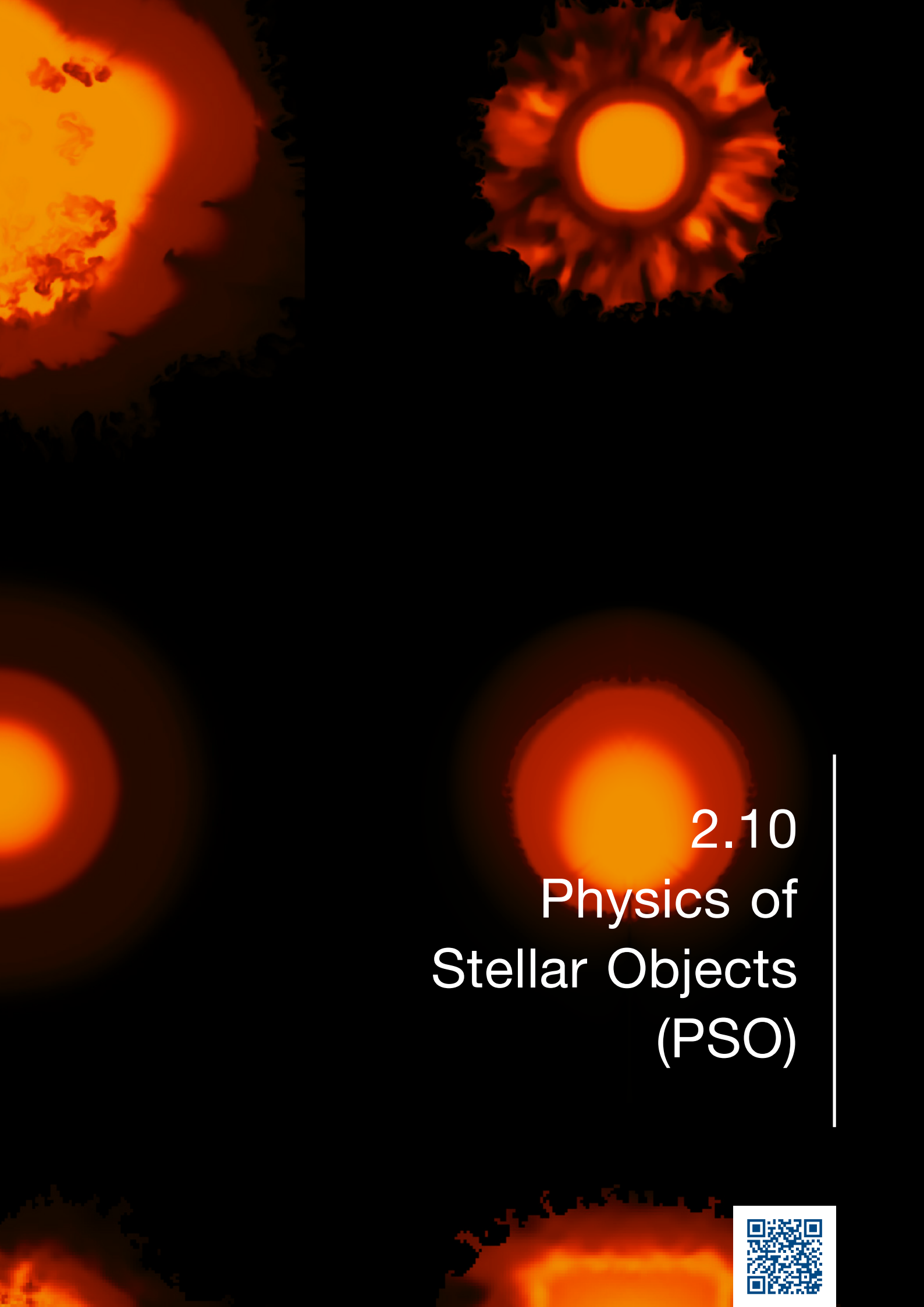
After more theoretically inclined work in the previous year, which led to the publication of Müller (2017a), we focused on implementation and evaluation. In terms of methodology, using word embeddings was an obvious choice to try to capture semantic similarity between publications by authors with ambiguous names that could – or could not – be the same real-life person. In author disambiguation, considering the semantic similarity of two publications is not yet the standard, and co-author similarity is the single most widely used feature. The success of co-author-based features is due to the fact that author disambiguation has mainly been attempted for publications from domains with a high proportion of multi-author papers. In the context of the SCAD project, in contrast, we also target publications from the domain of mathematics, which is characterized by less-rich co-author information.

In 2017, we published the first research paper in the field of author disambiguation that applied word embeddings for this task (Müller 2017b).

While the paper was a significant success for the project in that it demonstrated the utility of Word Embedding-based semantics for author disambiguation, it also indicated some directions for future work, which, in turn, have defined much of our project schedule for 2018. In doing all this, we focus on methods that are lightweight in the sense that they should not require specialized hardware. This self-limitation is both a necessity for building systems that are practically applicable in real-life situations (e.g., at digital libraries) as well as justified due to the nature of the semantic relations observed in author disambiguation.

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 was formed and how it was distributed throughout 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 optical and other parts of the electromagnetic spectrum. They are 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 take place in stars and stellar explosions. Newly developed numerical techniques and the ever-increasing power of supercomputers facilitate the modeling of stellar objects in unprecedented detail and with unparalleled precision.

One of our group's primary goals is to model the thermonuclear explosions of white dwarf stars that lead to the astronomical phenomenon known as Type Ia supernovae. These supernovae are the main source 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 in the PSO group to other kinds of cosmic explosions.

Classical astrophysical theory describes stars as one-dimensional objects in hydrostatic equilibrium, an approach that has proven extremely successful and explains why stars are observed in different configurations while also providing a qualitative understanding of stellar evolution. However, simplifying assumptions limit 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 that take 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. Fundamentale Fragen sind, wie dieses Material gebildet wurde und wie es sich im Universum verteilt hat. 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 aufwendigen 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. Sebastian Ohlmann

Scholarship holders

Dr. Samuel Jones (*Humboldt Research Fellowship until April 2017*)
Leonhard Horst

Visiting scientists

Dr. Markus Kromer (*Gliese Fellowship*)
Aron Michel (*until October 2017*)
Florian Lach (*since March 2017*)
Sabrina Gronow (*since April 2017*)

Students

Christian Sand
Alberto Botto Poala (*until October 2017*)

Neutron stars on the brink of collapse

When a very massive star dies, its core contracts. In a supernova explosion, the star's outer layers are expelled, leaving behind an ultra-compact neutron star. Massive stars often occur in binaries. The neutron stars that are formed from them remain bound after the supernova explosions and may be close enough to eventually merge. For such an event to occur, the system must lose angular momentum and energy. According to the predictions of Einstein's theory of general relativity, this occurs as a consequence of the emission of gravitational waves, which have indeed been detected very recently with the extremely sensitive gravitational-wave detectors Advanced LIGO in the United States. The first measured events were associated with mergers of two black holes, and the 2017 Nobel Prize in Physics was awarded for this detection. On August 17, 2017, for the first time, the Advanced LIGO and Virgo (Italy) Observatories were able to observe the merger of two neutron stars, which caused great excitement in the scientific community.

While neutron stars have a slightly larger mass than our Sun, their diameter is a mere 10 km. These stars thus contain a large mass in a very small spatial volume, which leads their interior to the most extreme conditions under which matter is found in the Universe, resembling those encountered in atomic nuclei. Theorists have been exploring these conditions for several decades and have developed models to describe matter under these conditions – a challenging task that involves many aspects of nuclear and particle theory and gives rise to uncertainties. Therefore, an experimental test or observation is urgently needed to pin down the properties of matter in neutron stars and to constrain the theoretical models. Researchers are particularly interested in determining the radius of neutron stars as their size depends on the unknown properties of matter.

Based on the observational data from the LIGO and Virgo instruments, an international team of scientists from Germany, Greece, and Japan (including HITS astrophysicist Dr. Andreas Bauswein) has managed to narrow down the size of neutron stars with the aid of computer simulations. The calculations suggest that the neutron star radius must be at least 10.7 km.

The collapse as evidence

In neutron star collisions, two neutron stars orbit around each other, eventually merging to form a star with approximately twice the mass of the individual stars. In this cosmic event, gravitational waves – oscillations of space-time – are emitted. This event resembles what happens when a stone is thrown into water and waves form on the water's surface. The heavier the stone, the higher the waves. Thus, the signal characteristics reveal properties of the source. In the case of merging binary objects, the particular shape of the gravitational-wave signal reveals the masses of the merging stars. The first detected neutron star merger event was dubbed GW170817, and the gravitational-wave signal was interpreted as being produced by a binary system with a total mass of 2.74 solar masses.

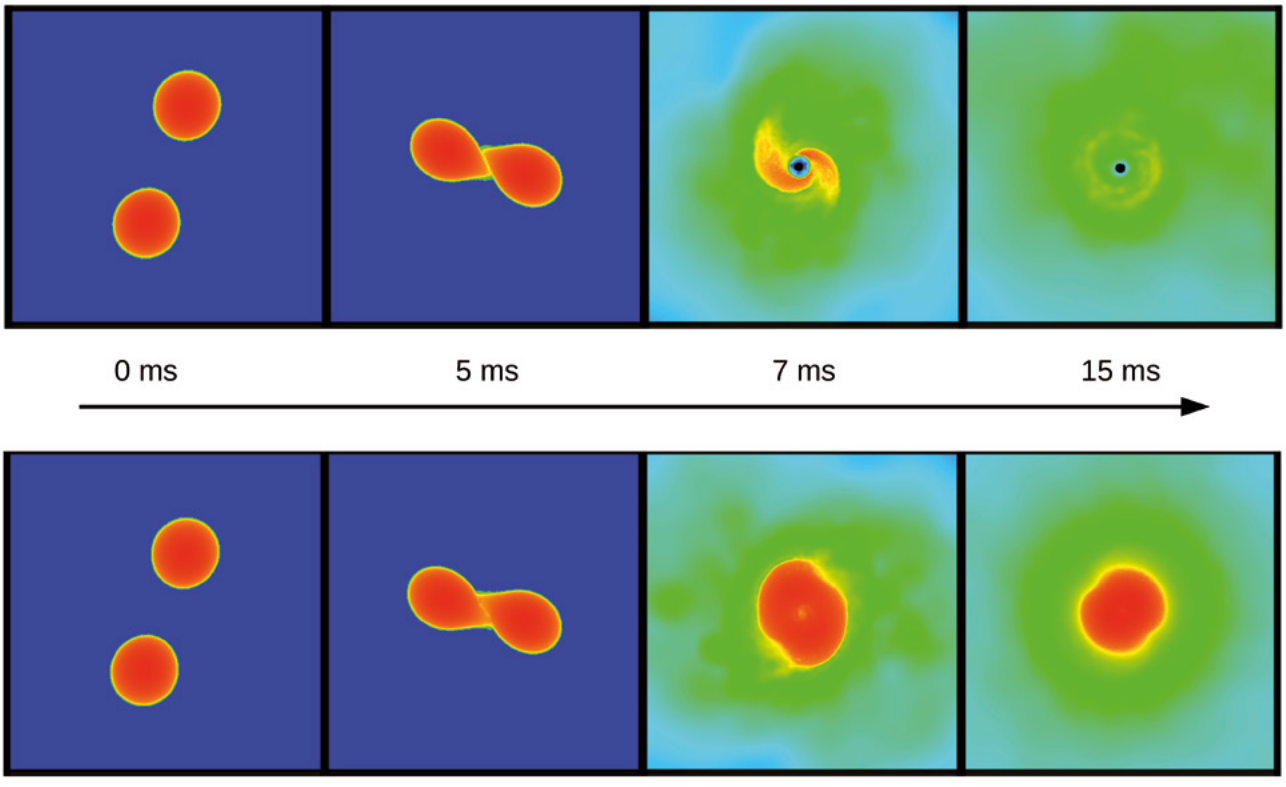
Researchers from the international team around PSO group member Andreas Bauswein simulated different merger scenarios for the recently measured masses to determine the radius of the neutron stars. In so doing, they relied on different models and equations of state describing the exact structure of neutron stars. Then, they checked whether the calculated merger scenarios were consistent with the observations. For the merger of two neutron stars, two outcomes are conceivable (see *Figure 51*). As the product is unstable, it may directly collapse into a black hole; however, this process may be delayed, and a massive remnant that is formed from the merged neutron stars may exist for several milliseconds until it also finally collapses into a black hole. Based on their computer simulations, the team concluded that all models leading to a direct collapse of the merger remnant could be ruled out because in this case, relatively little light is emitted during the collision; however, different telescopes have observed a bright light source at the location of the stars' colli-

sion, which provides clear evidence against the hypothesis of direct collapse. The new simulation results thereby exclude a number of models of neutron star matter, namely all models that predict a neutron star radius smaller than 10.7 kilometers.

Neutron stars reveal fundamental properties of matter

Such new measurements and calculations help theorists to better understand the properties of high-density matter in our Universe. The recently published study by [Bauswein et al., 2017d] marks of scientific progress as it ruled out some theoretical models, but there are still several possible models with neutron star radii greater than 10.7 km. However, the authors of this publication were able to demonstrate that further observations of neutron star mergers will continue to tighten the constraints. The LIGO and Virgo Observatories have just begun taking measurements, and the sensitivity of the instruments will continue to increase over the next few years and provide even more precise observational data. Therefore, it is very likely that more neutron star mergers will soon be observed and that the observational data from these events will reveal more about the internal structure of matter.

Figure 51: The upper and lower panels each visualize a simulation of a neutron star merger. The density in the equatorial plane is color-coded. In the scenario shown in the upper panels, the star collapses after the merger and forms a black hole within less than a millisecond, whereas the scenario displayed in the lower row leads to an at least temporarily stable star, which survives for many tens of milliseconds (picture: Andreas Bauswein, HITS).



Where gold is made in the Universe

In addition to providing clues regarding the properties of high-density matter, the observations of the neutron star merger event GW170817 solve an old mystery about the origins of heavy elements such as gold, platinum, and uranium, which had not been precisely known before even though neutron star collisions had been discussed as promising candidates. Optical observations of the "kilonova" associated with GW170817, in which Markus Kromer (Gliese Fellow at ZAH, Heidelberg University, and member of the PSO group) was involved [Smartt et al., 2017], now demonstrate that the explosion produced the precise conditions necessary to forge gold and other heavy elements. In fact, some astronomers believe to have identified traces of these heavy elements in their observations.

Introducing HESMA – the HEidelberg Supernova Model Archive

Type Ia supernovae are spectacular cosmic explosions that impact on many aspects of astronomy and astrophysics. They are believed to be the main source of iron and similar elements in today's Universe. Their extreme brightness and the remarkable homogeneity of their observable properties render them "cosmic lighthouses" that have been used to measure cosmic distances. These measurements reveal fundamental properties of the Universe, and its expansion has been found to be accelerating – a discovery that was awarded the 2011 Nobel Prize in Physics.

The need for Type Ia supernova models

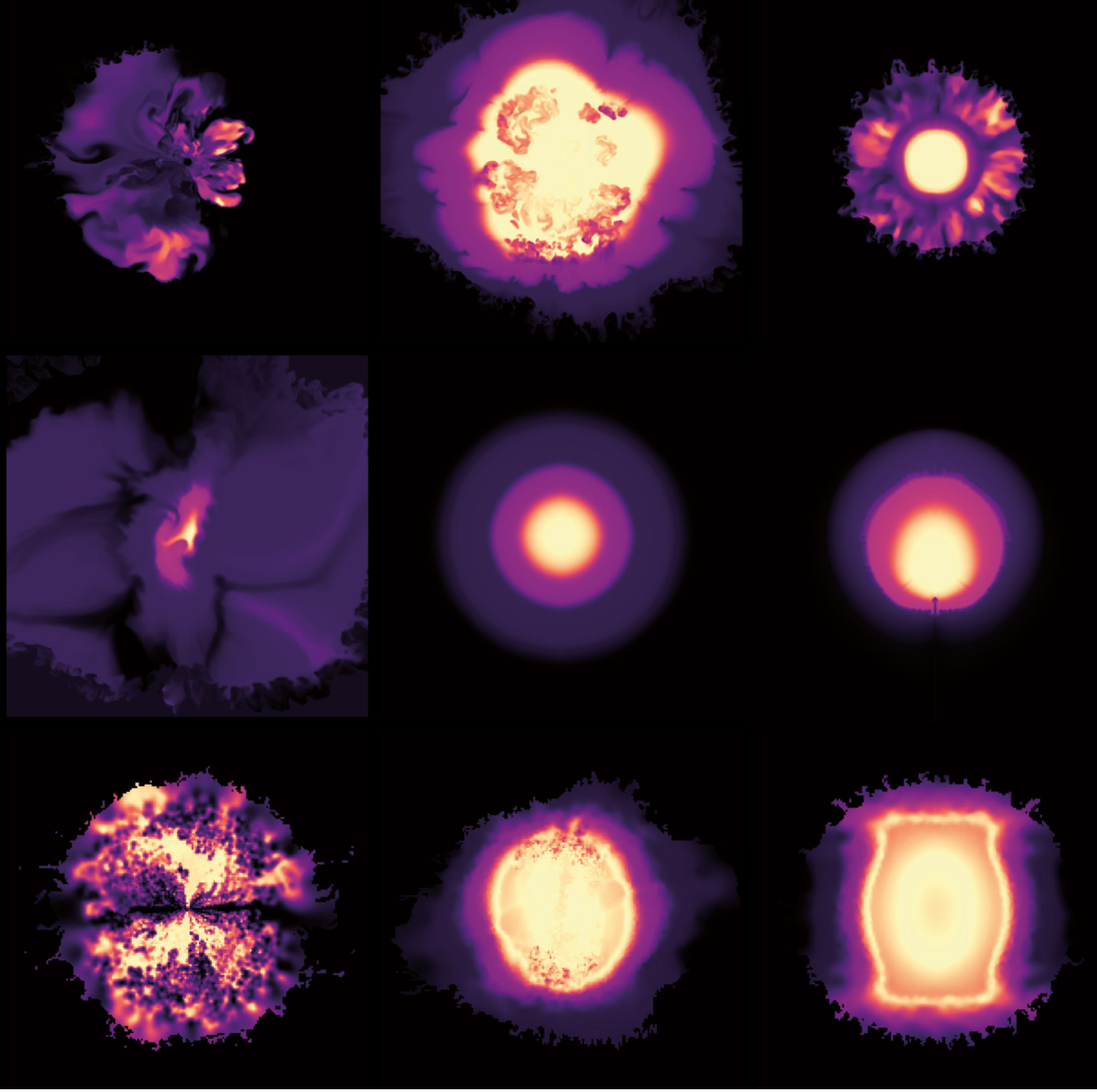
While Type Ia supernovae are believed to result from thermonuclear explosions of white dwarf stars with roughly the mass of the Sun and the volume of the Earth, details on the physical mechanism are not fully understood. Therefore, in collaboration with an international team involving researchers from the UK, Sweden, and Australia, the PSO group performs extensive simulations on some of the world's fastest supercomputers to model such events in three spatial dimensions. These models provide a detailed picture of the explosion physics, the production of chemical species in nuclear reactions, the structure of the ejected material, and the observational properties expected from different explosion scenarios.

This effort has been underway for over a decade and has produced numerous results that have been published in the astrophysical literature. Astrophysicists around the world use this wealth of data for follow-up studies and comparisons with observational findings. As a result, the need for a central repository as a service to the community has emerged.

The HESMA database

Following repeated requests by the community for nucleosynthetic yields, ejecta profiles, and synthetic observables of our explosion models of Type Ia supernovae, the PSO group at HITS decided to make the data of all their published models available in electronic form. To this end, we have established an online model archive, the HEidelberg Supernova Model Archive (HESMA), at https://hesma.h-its.org, see also [Kromer et al., 2017].

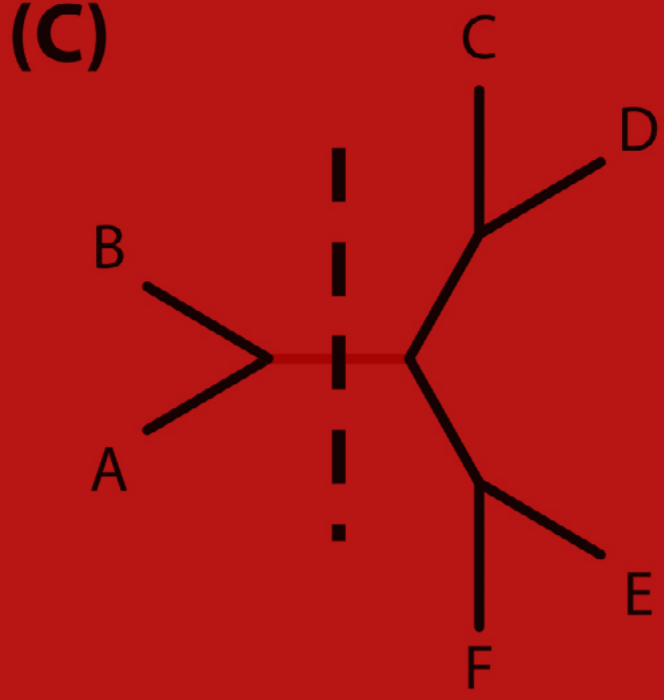
HESMA currently includes data on about 70 Type Ia supernova explosion models. The models cover a wide range of progenitor scenarios and explosion mechanisms of Type Ia supernovae and were produced by the PSO group at HITS and its collaborators at the Max Planck Institute for Astrophysics in Garching, Queen's University Belfast (UK), and UNSW Canberra (Australia). A graphical representation of a subset of the models is displayed in *Figure 52*. HESMA provides integrated isotopic abundances as well as radially averaged density and isotopic abundance profiles of the explosion ejecta after they have achieved homologous expansion. It also offers angle-averaged UVOIR bolometric light curves as well as gamma-ray and optical spectral time series. Bolometric light curves are tabulated from ~ 6 to 80 days past explosion. Gamma-ray spectra for selected models are available from about 6 to 100 days past explosion. Optical ($\lambda \in [3500,9500]$ Å) spectral time series are provided in our database from about 6 to 40 days past explosion. Additionally, early synthetic light curves for the epoch from 1.1 to 10 days past explosion are included for some models and selected filter bands.



Since its launch in June 2017, HESMA has been very well received in the astrophysical community. At the time of writing this report, HESMA had more than 440 registered users. We hope that the HESMA data will be useful for a wide range of applications, from supernova modeling over the interpretation of observed supernova spectra to galactic and cosmic chemical evolution. Future models will be added to HESMA after publication.

Figure 52: This image displays slices of the mean atomic number at the end of the hydrodynamics simulation for a variety of models from HESMA. The top row shows Chandrasekhar-mass explosions (from left to right: deflagration, delayed detonation, and gravitationally confined detonation), the middle row shows explosions of a sub-Chandrasekhar-mass primary WD (from left to right: violent merger, pure detonation, and double detonation), and the bottom row displays super-Chandrasekhar-mass explosions (from left to right: deflagration, delayed detonation, and detonation). Figure taken from [Kromer et al., 2017].

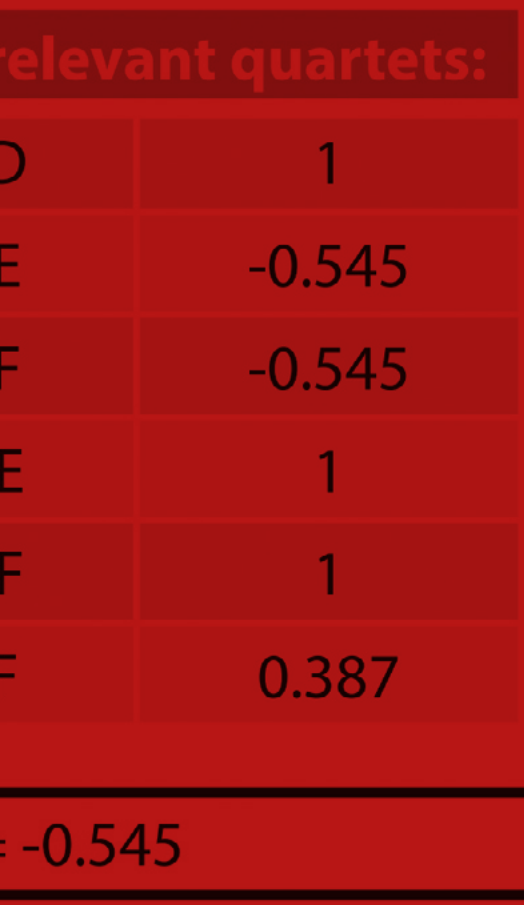
2 Research



Bip: AB|CDEF

Q-IC of
AB|CD
AB|CE
AB|CF
AB|DE
AB|DF
AB|EF

LQ-IC = min Q–IC =



2.11 Scientific Computing (SCO)

2.11
Scientific Computing (SCO)

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

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

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

In the following section, we outline our current research activities, which are situated at the interface(s) between computer science, biology, and bioinformatics.

The overall goal is to devise new methods, algorithms, computer architectures, and freely available/accessible tools for molecular data analysis and to make them available to evolutionary biologists.

In other words, we strive to support 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 that can be analyzed, observed, and tracked at the DNA level.

As evolutionary biologist Theodosius Dobzhansky's famous and widely quoted dictum states, "Nothing in biology makes sense except in the light of evolution".

Note: In 2018, the name of the group will be changed into "Computational Molecular Evolution" (CME).

Die Gruppe Wissenschaftliches Rechnen (SCO) beschäftigt sich mit Algorithmen, Modellen 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, Forschung zu unterstützen. 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".

Hinweis: Im Jahr 2018 wird die Gruppe umbenannt in „Rechnerbasierte Molekulare Evolution" (CME).

Group Leader

Prof. Dr. Alexandros Stamatakis

Staff members

Dr. Diego Darriba *(until October 2017)*
Dr. Tomáš Flouri *(until April 2017)*
Alexey Kozlov

Scholarship holders

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

Visiting scientists

Khouloud Madbouh *(January / February 2017)*
Umberto Perron *(December 2017)*
Laura Rubinat *(June 2017)*
Theodora Serdari *(February – December 2017)*

Students

Rudolf Biczok
Sebastian Giesse
Fernando Ramirez
Tobias Ribizel
Axel Trefzer

What happened at the lab in 2017?

In the winter of 2016 / 2017, Alexis, Tomas, Alexey, and Pierre taught the "Introduction to Bioinformatics for Computer Scientists" class at the Karlsruhe Institute of Technology (KIT). As in previous years, we received very 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/Winter16_17.pdf). Enrollment in our course at KIT further increased over the winter semester of 2017 / 2018.

During the summer semester of 2017, we were involved in a plethora of teaching activities. We again taught our main seminar "Hot Topics in Bioinformatics" and a new joint interdisciplinary main seminar entitled "Algorithmic Methods in the Humanities" with colleagues from the KIT Departments of Computer Science, Philosophy, and Language. In collaboration with Prof. Wagner at the Institute for Theoretical Informatics, we also offered a bioinformatics programming practical whose results are outlined in more detail below. The practical also received a learning quality index rating of 100 based on student evaluation. 2017 was very important for our current PhD student Sarah Lutteropp, who successfully completed her master's thesis at the Department of Computer Science at KIT. Nico Schwaiger also completed his master's thesis – which was co-supervised by Tilmann Gneiting from HITS – at the KIT's Department of Mathematics.

A total of five KIT master's students joined the lab as research assistants in 2017 to work on various programming projects. Moreover, Sebastian Giesse, who – like Sarah – also won the national "Bundeswettbewerb für Informatik" ("National Computer Science Competition"), joined the lab to compose his master's thesis.

In 2017, we hosted several PhD students via our Visiting PhD Student Program. Khouloud Madbouh, a computer scientist from the University of Tunis, joined us for a second visit until mid-February 2017. Laura Rubinat (CNRS, France) visited us in June 2017 to work on novel methods of assessing microbial diversity. Theodora Serdari worked with us as an intern from February to December 2017 on assessing whether computational species delimitation methods can be applied in classifying viruses. Finally, Umberto Perron (EBI Hinxton, UK) spent some time with us in December 2017 to implement novel models of protein evolution that take the three-dimensional structure of proteins into account.

Another highlight of 2017 was the course Computational Molecular Evolution, for which Alexis again served as co-organizer. The course took place for the 9th time at the European Bioinformatics Institute in Hinxton, UK. Former lab member Paschalia Kapli and current PhD student Lucas Czech contributed significantly to the success of the course as teaching assistants (see *Chapter 5.1.2*).

During the year, we had to say goodbye to Tomas, who went on to begin a second postdoc at University College London, and to Diego, who was offered an assistant professor position in Spain.

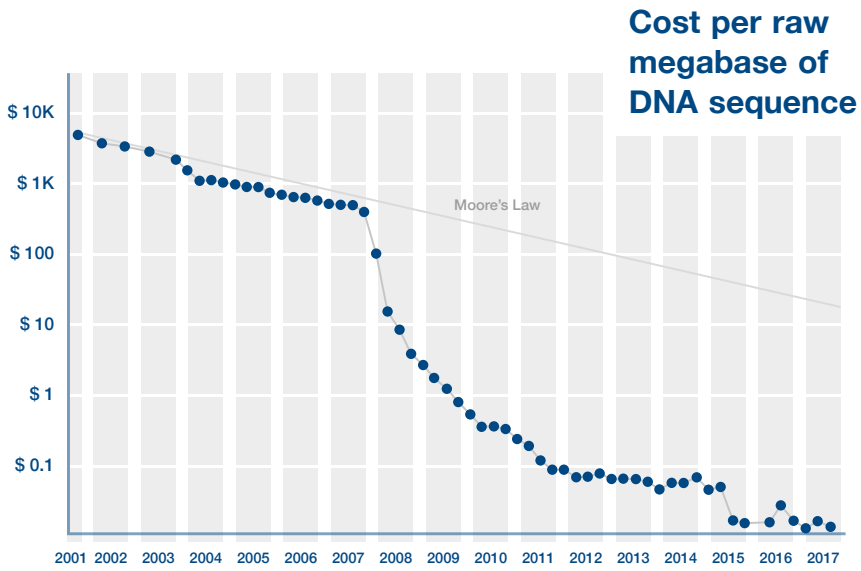
Finally, the year was characterized by the creation of the preproposal and full proposal entitled "Algorithm Engineering for the Scalability Challenge" for the German Excellence Initiative by colleagues who mostly came from the KIT's Department of Computer Science. Alexis is currently writing and coordinating the chapter on the genome analysis application challenge.

Introduction

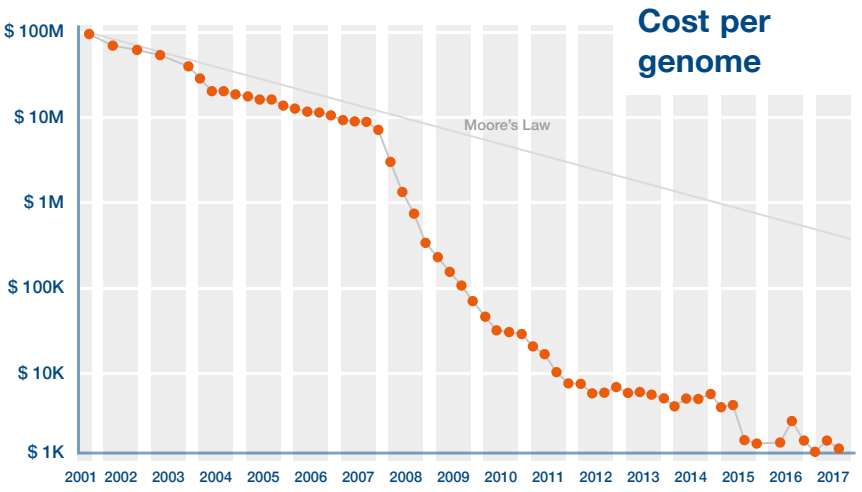
The term "computational molecular evolution" refers to computer-based methods of reconstructing evolutionary trees from DNA or, for example, from protein or morphological data.

The term also refers to the design of programs that estimate 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.



Source: National Human Genome Research Institute (NHGRI).



Source: National Human Genome Research Institute (NHGRI).

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 addressed is the fact that the number of molecular data available in public databases is growing at a significantly faster rate than the computers that are capable of analyzing the data can keep up with.

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 2017 (see https://www.genome.gov/sequencingcosts/).

Thus, we are facing a scalability challenge – that is, we are constantly trying to catch up with the data avalanche and to 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 compute-intensive models of evolution.

Another difficulty of computational molecular evolution 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 output requires the continuous development of new algorithms and tools to filter, puzzle together, and analyze these molecular data. For instance, our student programmer Fernando is working on a smartphone application that can rapidly identify unknown sequences. Small sequencing devices that can be connected to smartphones and allow for sequencing and identifying viruses during viral outbreaks on spot already exist. 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, for instance, to disentangle the origin of bacterial strains in hospitals, to determine the correlation between the frequency of speciation events (species diversity) and climatic changes in the past, and to analyze microbial diversity in the human gut.

As already mentioned, one key challenge for computer science is scaling existing analytic methods to the enormous new datasets produced by next-generation sequencing methods. We applied for and received 70 million CPU hours on the Munich supercomputer SuperMUC to conduct a new insect transcriptome analysis project. The dataset is already one order of magnitude larger than that published in Science in late 2014. Thus, our codes need to be capable of leveraging the computational resources provided by supercomputers. In this context, we made several important software releases in 2017. Alexey released RAxML-Next Generation (RAxML-NG), a completely re-designed version of our highly cited flagship tool RAxML, which is used for phylogenetic inference.

Pierre released EPA-NG, the next-generation implementation of our evolutionary placement algorithm that places millions of anonymous short sequences into reference trees of known sequences. The new code displays high parallel efficiency on up to thousands of cores.

Diego released Modeltest-NG, a next-generation tool for determining the best-fit statistical model of evolution for a given dataset. The new tool is at least one order of magnitude faster than its predecessor.

Benoit released the second version of our low-level library (libpll2) for computing likelihood scores on phylogenetic trees. This new version now implements novel algorithmic techniques that accelerate likelihood calculations, which account for 90 – 95 % of the overall run-time in all the aforementioned tools. Finally, Diego and Alexey collaborated on implementing the pll-modules library, which offers a vast number of frequently used functions that build on the core likelihood library. *Figure 53* provides an overview of the modular software stack that we developed for implementing evolutionary analysis tools. However, it remains unclear if we really need to deploy such compute-intensive statistical models of evolution to infer phylogenetic relationships given the mass of data at hand. Thus, we are investigating whether less-expensive discrete methods could be used or combined with the application of statistical methods to a subset of the input data in order to obtain equally accurate results while utilizing significantly fewer computational resources. Our PhD student Sarah is currently working on such unconventional inference methods. Moreover, our master's student Sebastian is investigating a novel discrete criterion for reconstructing phylogenies. The criterion aims to minimize the conflict in the evolutionary signal inherent in the data. There has also been substantial progress in our new research line on cancer-cell evolution. Together with David Posada at the University of Vigo, Spain, we devised and implemented novel models of cancer evolution. Our prototype implementation displays better accuracy than current competing tools for this purpose.

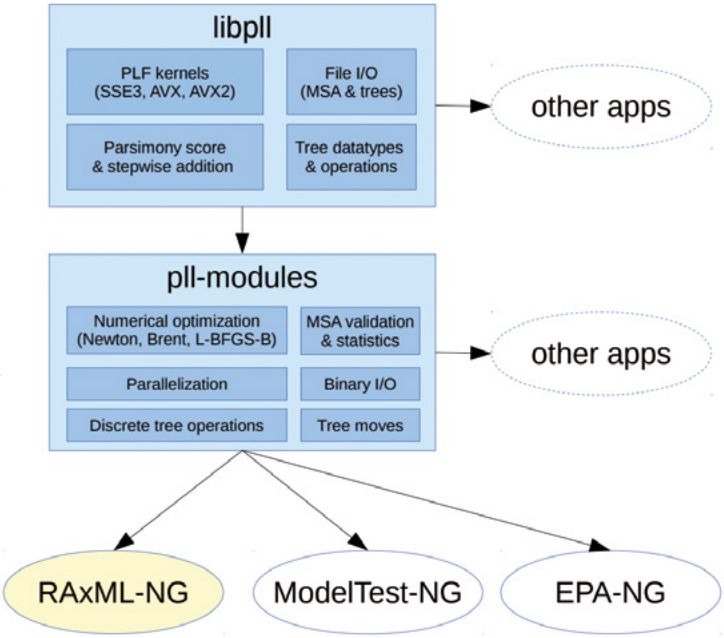


Figure 53: overview of the modular software stack for implementing evolutionary analysis tools.

RAxML-NG

A major milestone for the lab was reached in March 2017 when Alexey released a completely re-designed version of RAxML, called RAxML-NG. As RAxML had been developed over 15 years and had become unmaintainable, we decided to re-implement the tool from scratch by relying on the aforementioned new libraries (llpll and pll-modules). Our goal was to develop a new, more user-friendly tool that is also more flexible with respect to the statistical models of evolution that the user can specify and that is easier to maintain as well as to extend. In addition, our aim was to design a single code that scales from a single core up to a supercomputer and takes into account all the lessons learned during the RAxML program development.

We had previously developed two separate yet highly similar tools for laptops / servers (RAxML) and for supercomputers (ExaML) that were hard to keep synchronized. Thus, RAxML-NG now implements all features and algorithms from RAxML and ExaML, such as checkpointing, optimal data-distribution algorithms, binary input data-file formats for concurrent I/O, etc. While we relied on the efficient parallelization scheme of ExaML, we managed to further improve parallel efficiency by adopting a hybrid Pthreads / MPI-based parallelization approach. In *Figure 54* below, we compare the scalability of ExaML with RAxML-NG on up to 2,048 cores on two large empirical DNA- (left) and protein- (right) datasets. It is worth noting that RAxML-NG attains more than 100 % parallel efficiency (known as super-linear speedups in computer science) because of increased cache efficiency (i.e., the availability of the increased accumulated -over all cores- capacity of short and fast-to-access small on-chip memory). While such super-linear speedups are not uncommon for data bandwidth-bound scientific applications, a parallel efficiency of 140 % as achieved on the protein dataset is rather spectacular. Apart from improved parallel efficiency, RAxML-NG is also about a factor of 2 faster than RAxML because it uses the more efficient likelihood implementation in libpll2. Moreover, it is more accurate because of slight modifications to the tree-search algorithm. RAxML-NG also outperforms the main competitor (IQ-Tree) with respect to speed as well as accuracy.

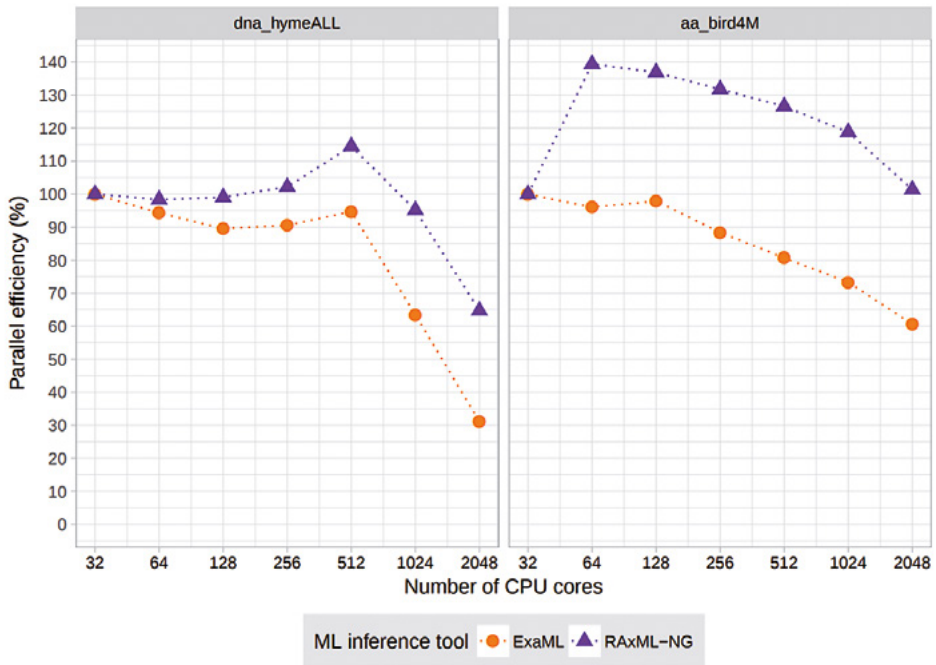


Figure 54: Comparison of the scalability of ExaML with RAxML-NG on up to 2,048 cores on two large empirical DNA- (left) and protein- (right) datasets.

Quantifying uncertainty in phylogenetic trees

In 2017, Sarah continued our successful work with our collaborator Antonis Rokas (Vanderbilt University, Nashville, USA) on quantifying the uncertainty of phylogenetic trees [Zhou et. al., 2017]. Our new and substantially more flexible method relies on sampling all quartets (all induced subtrees with 4 species) from a given set of phylogenetic trees to infer the degree of uncertainty for an inner branch of the corresponding species tree. The extremely large number of induced quartets gave rise to an abundance of interesting discrete algorithmic problems. For instance, together with Ulrich Meyer in Frankfurt am Main, Germany, we are co-supervising a master's thesis on developing an external memory algorithm (i.e., an algorithm that intelligently transfers data between disk and RAM for problems that require more memory than is available in RAM) to calculate the frequency of occurrence of quartets. Our master's student Sebastian is currently implementing search algorithms to infer a species tree that minimizes the degree of uncertainty. A schematic outline of the uncertainty score computation is provided in *Figure 55* below.

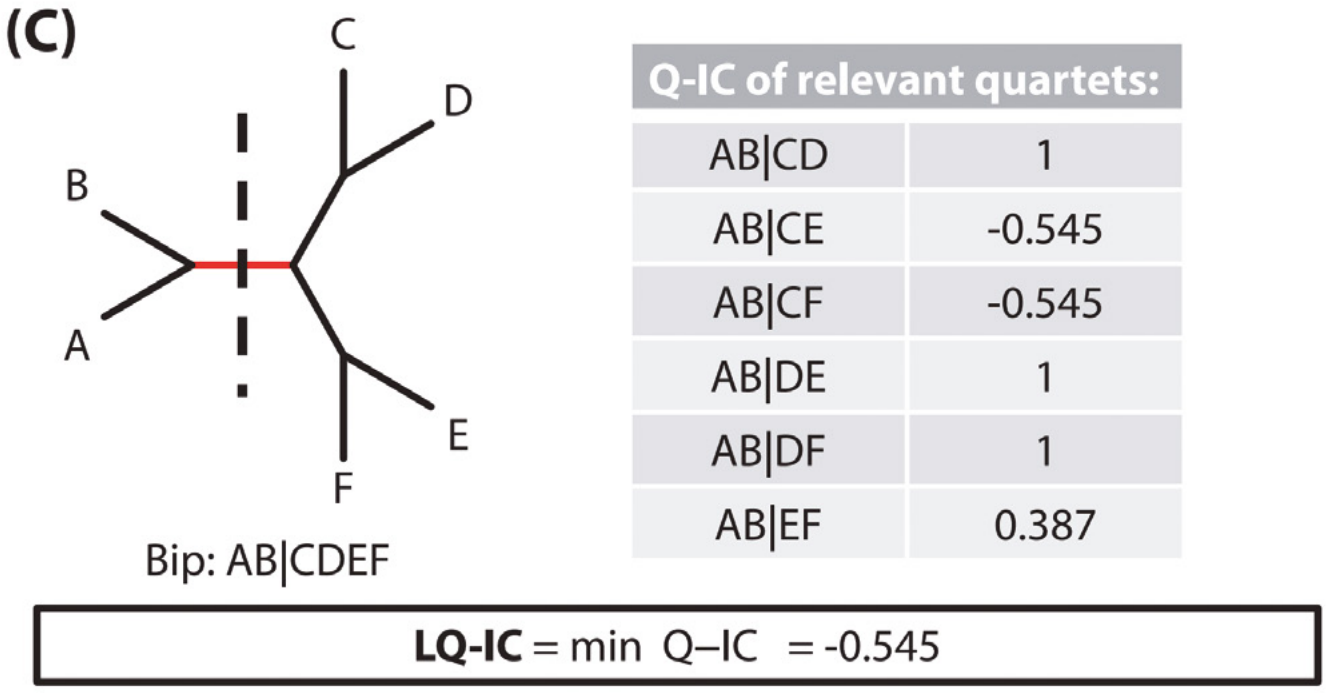


Figure 55: Schematic outline of the uncertainty score computation to infer a species tree that minimizes the degree of uncertainty.

Terraces in tree space

So-called terraces in phylogenetic tree space occur under specific evolutionary model settings and for specific input-data patterns. When a phylogenetic tree resides on such a terrace, this means that several other distinct tree topologies with the exact same analytical likelihood score also exist.

In statistical terms, this implies that the model is not identifiable – that is, several different parameter values exist (in this case, different tree topologies, which are a parameter of the model) that have the same likelihood score and can thus not be distinguished.

Based on previous work conducted by us and other research groups, it is already known that terraces can have a detrimental impact on the post-analysis and biological interpretation of phylogenetic trees if they are not detected. For instance, if biologists do not notice that a final tree that they publish resides on a terrace with potentially millions of other trees, the evolutionary assessment of that single tree will be of little value. While this phenomenon had already been described in 2011, none of the standard tools for phylogenetic inference investigate whether the final output tree resides on a terrace. We believe that this failure is mainly due to the lack of an efficient C/C++ library for executing the known (albeit somewhat complicated) discrete algorithm for counting and enumerating trees on a phylogenetic terrace.

To this end, we set forth to implement such an efficient library with the students of our programming practical at KIT. In fact, two teams of 4 students each worked on two independent implementations, which allowed for better verification of the correctness of the implementations by comparing the results and for generating competition in implementing the fastest code.

The so-called terraphast-I implementation by the first student team won and was 28 times faster than a previously available python implementation of the same algorithm while, also requiring one order of magnitude less memory. The library has already been integrated into RAxML-NG and will also be integrated into the IQ-Tree tool for phylogenetic inference. For further details, please see [Biczok et al., 2017].

2 Research


D-3-HYDROXYACYL-ENZYME
D-3-HYDROXY-BUTYRYL-ENZYME
trans-2,3-DEHYDRO-ACYL-ENZYME
CROTONYL-ENZYME
(PALMITOYL-CoA-ENZYME PALMITOYL TRANSFERASE)
for chain elongation (mitochondrial)
FATTY AC

2.12 Scientific Databases and Visualization (SDBV)

2.12
Scientific Databases and Visualization (SDBV)

Our mission is to make data accessible to researchers and to everyone. The challenge of fulfilling users' needs is multi-faceted, and we are interested in bringing them scientific data that can be more easily used. A key part of our work involves the preparation of data, which enables users to more quickly find what they need, namely by performing biocuration – that is, by enriching, structuring, and interrelating data. What exactly does curation mean, and what is sufficient data quality? Finding pragmatic answers to these questions proves challenging.

Biocuration is inherently limited, at least if there are few individuals who curate data for the many. This limitation leads to another field of our work: How can we help to put some of the curation load onto other people's shoulders? How can we enable people to curate their own data? How can we simplify and incentivize self-curation to the point that it is taken up? How can we turn doing something that is perceived as being rather altruistic and simultaneously having unsure long-term benefits into an immediately gratifying experience? There are numerous open-ended challenges in the field. We tackle these challenges in both concrete projects with real people and real data needs as well as in tool design and development. Since its inception, the group has benefitted from the fact that it brings together computer scientists and scientists from other disciplines in the creation of tools adapted to our users' needs and that are part of national, European, and international data infrastructures.

Wir wollen Daten für Forscher leichter zugänglich machen. Das Ziel, die Bedürfnisse der Nutzer zu befriedigen, stellt uns vor vielschichtige Herausforderungen. Wir konzentrieren uns darauf, wissenschaftliche Daten zu den Nutzern zu bringen, und wir wollen sie befähigen, mehr Nutzen aus ihnen zu ziehen. Ein wesentlicher Punkt unserer Arbeit ist „Biocuration", also Daten durch Anreicherung, Strukturierung und Erstellung von Verbindungen zu verbessern. Was ist Kuratierung, was ist hinreichende Datenqualität? Für diese Fragen pragmatische Lösungen zu finden, fordert uns heraus.

„Biocuration" ist limitiert, zumindest wenn wenige für viele Daten kuratieren. Dies eröffnet ein anderes Arbeitsfeld: Wie können wir die Last des Kuratierens auf mehr Schultern verteilen? Wie können wir Menschen dazu bewegen, ihre eigenen Daten zu kuratieren? Wie können wir vereinfachen und Anreize setzen? Wie machen wir aus einer altruistischen Übung mit unsicherer Langzeit-Perspektive eine nutzbringende Erfahrung, die sofort wirkt? Wir stellen uns dieser Vielzahl von Fragen, einerseits in Projekten, in denen echte Menschen echte Fragen stellen, und in der Konzeption und Erstellung von Werkzeugen. Seit ihrem Beginn profitiert die Gruppe von ihrem Mix aus Informatikern und Fachwissenschaftlern. Zusammen bauen wir Werkzeuge, die an den Bedarf der Nutzer angepasst sind. Diese Werkzeuge sind Teil von nationalen, europäischen und internationalen Forschungsinfrastrukturen.

Group Leader

Priv.-Doz. Dr. Wolfgang Müller

Staff members

Dr. Steffen Brinkmann *(until December 2017)*
Martin Golebiewski
Ron Henkel
Renate Kania *(until February 2017)*
Dr. Olga Krebs
Dr. Hadas Leonov
Quyen Nguyen *(until August 2017)*
Dr. Maja Rey
Dr. Andreas Weidemann
Dr. Ulrike Wittig

Students

Marcel Petrov
Jumana Abu-Khader *(since November 2017)*

The Liver Systems Medicine network (LiSyM)

The SDBV group is responsible for the central data-management support of the German Liver Systems Medicine network (LiSyM). LiSyM represents a research network of more than 20 institutions that have been brought together for 5 years by a 20-million-euro funding program of the German Federal Ministry of Education and Research (BMBF). The 39 contributing research groups work together as coordinated, geographically distributed teams. In LiSyM, experimental laboratory researchers work together closely with mathematicians, bioinformaticians, modelers, clinical investigators, and data managers to develop a systems medicine approach to studying liver diseases. The aim is to acquire and use experimental data together with existing data from previous projects and literature to build computational models that address various functional aspects of the human liver and its crucial role in development and progression of many diseases. These models will facilitate decision-making at the patient's bedside and will help to predict the effects of new medicines in the treatment of diseases.

The liver supports almost every organ in the human body and is necessary for its survival. It is both the largest internal organ and the largest gland in body and maintains over 500 functions, many of them vital. The liver plays a major role in metabolism (i.e., the synthesis and breakdown of small and complex molecules). All the blood leaving the stomach and intestines passes through the liver, which processes it and breaks down the nutrients and drugs into forms that are easier for the body to use. Because of its vital functions, the liver is prone to many diseases, that in turn lead to serious problems worldwide.

More than 5 million people in Germany suffer from diseases of the liver. Non-alcoholic fatty liver disease (NAFLD) is rapidly becoming the most prevalent form liver disease. In Europe, NAFLD is estimated to affect about 20 % (and in the US, even 30 %) of the population. As a central metabolic organ, the liver plays a dominant role in metabolic changes that underlie widespread diseases, such as atherosclerosis, chronic heart failure, dementia, and type 2 diabetes. One of the major goals of LiSyM is thus to unravel the mechanisms leading to NAFLD and non-alcoholic steatohepatitis (NASH).

The LiSyM network builds on the Virtual Liver Network (VLN), which ran from 2010 to 2015 and in which SDBV was responsible for the central data management (just as it is now in LiSyM). Scientifically, LiSyM is based on four discrete (albeit related) pillars, each of which focuses on a particular phase of disease. These phases are described by specific clinical phenomena and underpinned by experimental and theoretical studies:

Pillar I: Early Metabolic Injury

Pillar I examines the transition from steatosis – the process describing the abnormal retention of lipids within liver cells – to NASH as the disease-defining moment in NAFLD. Through a systematic analysis, the molecular mechanisms that trigger the establishment of NAFLD in affected patients are elucidated.

Pillar II: Chronic Liver Disease Progression

Chronic liver diseases (CLD) progression leads to cirrhosis, often to cancer, and ultimately to organ failure and death. A systems medicine approach was chosen to develop strategies for better characterizing liver fibrosis (the formation of excess fibrous connective tissue). The aim is to define key molecular mechanisms and structural changes in tissue architecture during the progression of CLD at all biological scales (cellular, tissue and organ level).

Pillar III: Regeneration and Repair in Acute-on-Chronic Liver Failure

Acute-on-chronic liver failure (ACLF) is a recently recognized syndrome in cirrhosis that is characterized by acute decompensation, organ failure, and high short-term mortality. Early detection and the discovery of a cure are urgent clinical needs for the increasing number of patients affected by this often-fatal disease. A systems medicine approach helps to identify the critical mechanisms of ACLF and to foster liver regeneration and repair.

Pillar IV: Liver Function Diagnostics

Data (breath test scores, imaging data from MRE, histopathological characterization, and proteomics data of liver biopsies) are acquired from two different groups of patients and used in computer-assisted diagnostic tools for early detection and a comprehensive evaluation of altered liver functions.

In addition, four junior research groups funded by LiSyM are working in close collaboration with the pillars. There is also some degree of collaboration between the pillars, and the scientific project management coordinates these actions, which is supported and complemented by the central data management for the whole network – the task of SDBV in LiSyM, in which Hadas Leonov and Steffen Brinkmann (software developers) and Martin Golebiewski (requirements engineer and user contact) contributed during 2017.

Data management for LiSyM

The purpose of LiSyM data management is to support research activities inside the network. We aim to provide one location at which all data can be stored or registered (if stored elsewhere) and to provide one location at which the relations of data items to other data items as well as to people and publications are visible. These data are stored in as FAIR a manner as possible (meaning Findable, Accessible, Interoperable, and Reusable).

An added complication within LiSyM is data of varying protection levels:

(I) Named patient data that can only be shared within an organization.

(II) Pseudonymized or anonymized data that can be shared across organizational boundaries.

(III) Information that would be legal to share but that may or may not be restricted in sharing, e.g., by IP considerations. To this end, we run a multi-pronged, federated data-management approach: We provide (1) Service, (2) Training, and (3) Development; we run (4) a facility; and we are a key contributor to (5) LiSyM policy-building and implementation.

Facility and development

The tools we provide are part of an integrated yet distributed federated infrastructure: In the center lies a LiSyM-specific instance (i.e. tailored installation) of the SEEK system (https://seek.lisym.org) that was released at the beginning of 2017. SEEK is the centerpiece of the FAIRDOM software platform and is also used by many other research networks and projects in the field, e.g., for the publicly available FAIRDOM-Hub (https://fairdomhub.org) [Wolstencroft et al., 2017]. LiSyM SEEK provides a space for the whole network to structure, store, and share data, models, SOPs, and publications, as well as to provide information and metadata on assays, used biological samples, events, and people/institutions involved in the project. The diverse types of data and information in SEEK can be associated to each other so as to depict relations between them.

In addition to LiSyM SEEK as a central exchange platform, local systems are used within LiSyM in some labs, and we connect SEEK to some of them. We are currently establishing an application programming interface (API) that reads and writes data, information, and metadata from/to SEEK. The API can be used to connect to other systems and to transfer data between two SEEK instances (e.g., from Virtual Liver SEEK to LiSyM SEEK) as well as for automated data retrieval from SEEK. Publication lists are exported to the public LiSyM website (lisym.org). The JSON Read API was released in 2017 and allows for searching, listing, and reading assets described in SEEK, as well as their attributes.

Other achievements in the past year include progress on the SEEK write API, the integration of iPython visualizations into SEEK, and containerized (Docker) installation of SEEK.

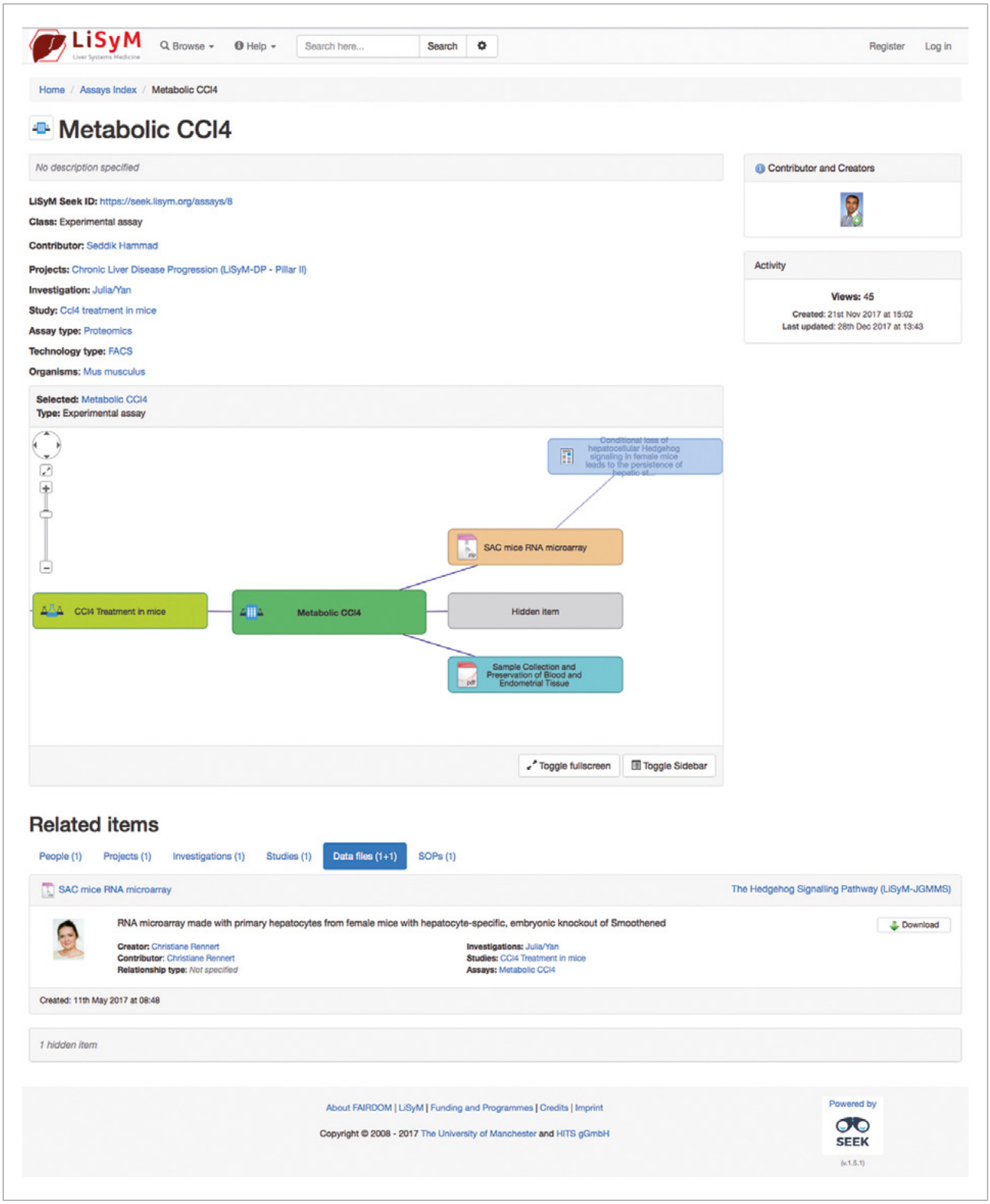


Figure 56: Screenshot of an experimental assay described in LiSyM SEEK, the central data-management platform of the LiSyM network.

Figure 57: Shooting the TV documentary "Operieren und Kassieren" ("Paying for Surgery") at HITS.

Service

To provide consulting support and service to LiSyM users, we answered about 200 emails from at least 43 users and addressed their data-management needs. We complemented this process via phone calls and participated in several LiSyM meetings. An important part of the service is consulting users on how to apply domain-specific standards for data formats as well as on metadata and semantic content descriptions. Consequently, we are active in several standardization initiatives [Schreiber et al., 2017] and organizations, such as the ISO (International Organization for Standardization), COMBINE (Computational Modeling in Biology Network [Myers et al., 2017]), and others, and we thereby provide a link between these standard-developing bodies and our SEEK users. We are also involved in leadership and coordination in the ISO committee ISO/TC 276 Biotechnology (https://www.iso.org/committee/4514241.html) and in COMBINE (http://co.mbine.org).

Training

To help gather requirements from users and prioritize them adequately, we have a team of about 12-15 "PALs" (Project Area Liaisons), who work within LiSyM as experimentalists, modelers, or clinicians and feed us with the users' specific data-management needs in addition to helping us disseminate newly implemented developments and features. In 2017, we ran one face-to-face meeting and several online PAL meetings, during which the PALs were able to demonstrate how they work with data and spawned a discussion among the participants and us on how given data-management problems can be addressed. A discussion in the Hünfeld PALs meeting led to the idea of organizing a data-structuring meeting (see Events section). In addition, we organized a hands-on data-management training as part of a LiSyM student- and postdoc retreat in November 2017 and a modelling tutorial as satellite to the 18th International Conference on Systems Biology (ICSB) in August 2017.

Policy-building and implementation

In addition to our development, facility, training, and service work, we have completed policy work. Together with the LiSyM leadership team, we have created a data sharing policy. After incorporating changes and running a discussion with the LiSyM SAB, the sharing policy was finally signed at beginning of 2018 by all PIs in LiSyM. The policy contains rules on how to handle the end of the project and how to access the data, SOPs, and models.

Operation Explorer

The Operation Explorer is a tool developed by the SDBV group in collaboration with science journalists with the purpose of visualizing statistical data about (mainly) medical diagnoses and procedures, such as surgeries. This tool simplifies the process of generating maps by enabling these data to be queried via selection lists. Results are presented as color-binned maps that contain data related to age-standardized populations in regions within Germany. The work on the Operation Explorer began when Volker Stollorz (now with the Science Media Center Germany in Cologne) was Journalist in Residence at HITS. During this time, Meike Hemschemeier's documentary "Operieren und Kassieren – Ein Klinik-Daten-Krimi" ("Paying for Surgery – A Clinical Data Thriller") went live on German public TV and attracted a large number of viewers. Over the past year, we have made plans to flexibilize the architecture of the operation explorer and have begun designing new projects that will be useful both for health-services research and journalism (see also *Chapter 6*).

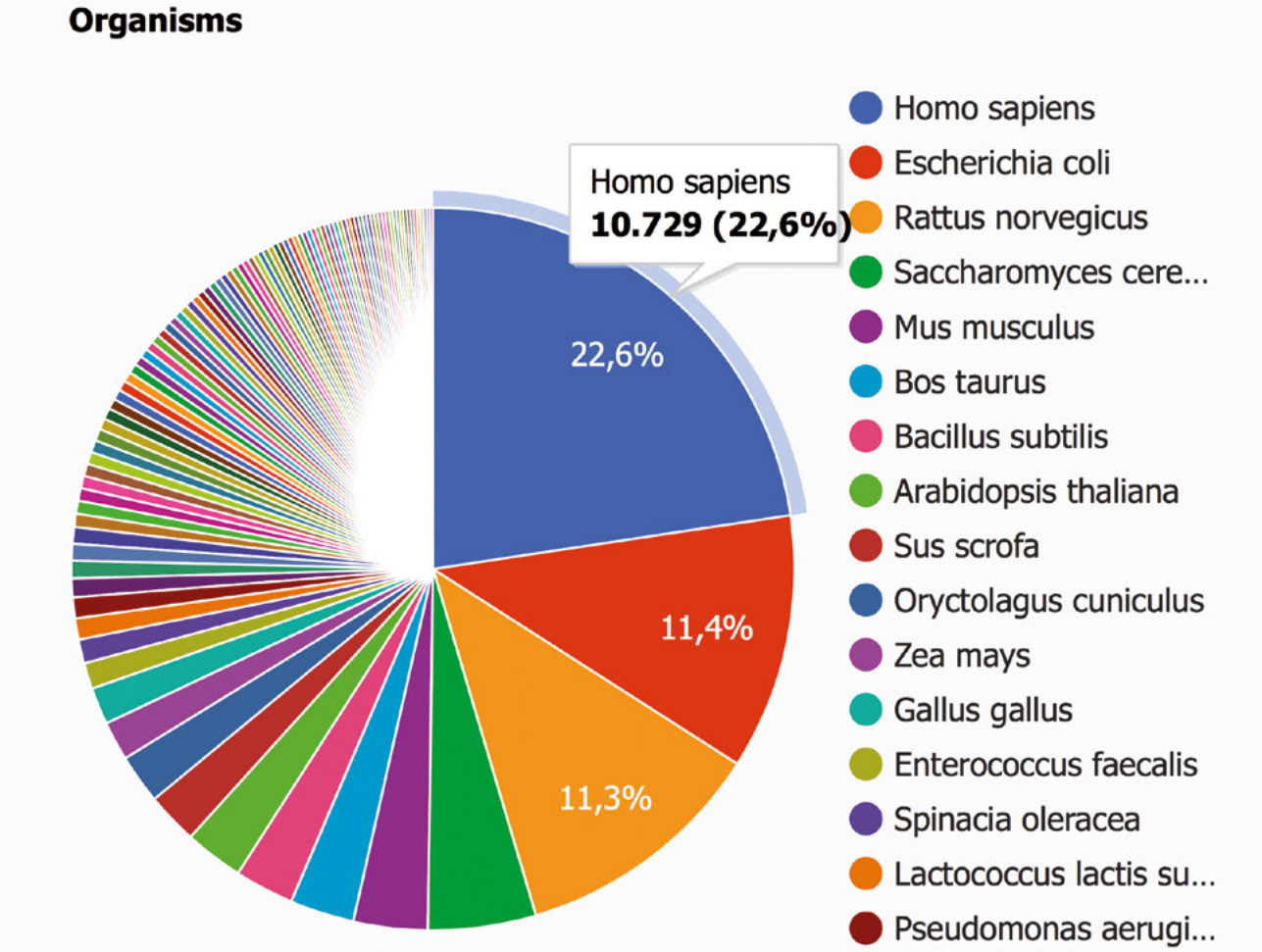


Figure 58: Taxonomical distribution of organisms in SABIO-RK.

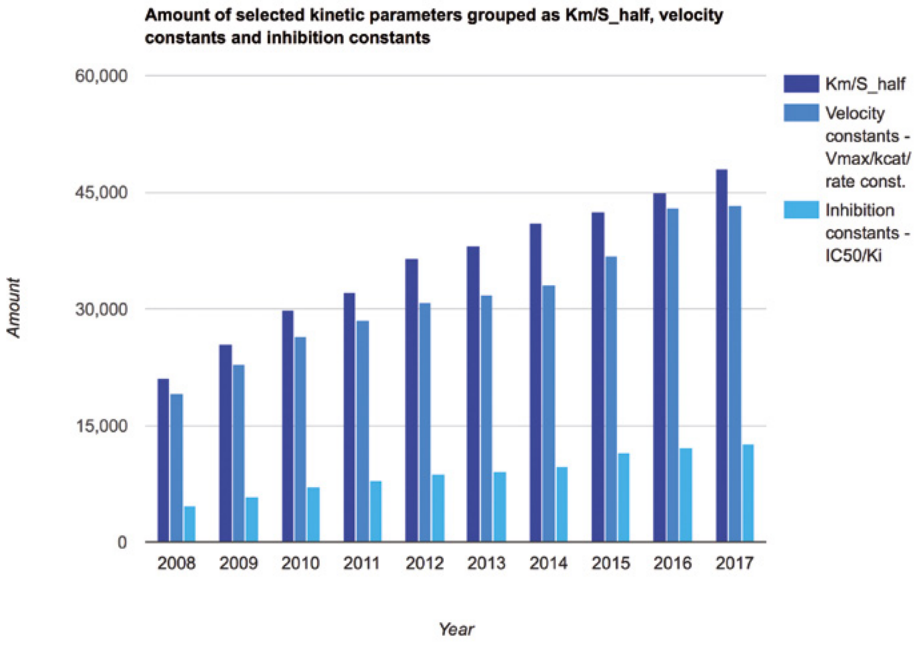
SABIO-RK

The SABIO-RK database (http://sabiork.h-its.org/) supports modelers of biochemical reactions and complex networks [Müller et al., 2017 and Wittig et al., 2018]. SABIO-RK represents a repository for structured, curated, and annotated data on reactions and their kinetics. The data are manually extracted from the scientific literature and stored in a relational database. The content comprises both naturally occurring and alternatively measured biochemical reactions, and the data are made available to the public via a web-based search interface and web services for programmatic access. Data are highly interlinked to external databases, ontologies, and controlled vocabularies. Links are implemented to KEGG for reactions; to UniProtKB for proteins; to NCBI taxonomy for organisms; to Brenda TissueOntology (BTO) for tissues; to PubMed for publications; to ChEBI, KEGG, and PubChem for compounds; to Gene Ontology (GO) for cell locations and signaling events; to Systems Biology Ontology (SBO) for kinetic laws and parameters; and to ExPASy, KEGG, BRENDA, IntEnz, IUBMB, Reactome, and MetaCrop for enzymes. Kinetics data and related information can be exported in different formats, including SBML, XML, and as spreadsheets.

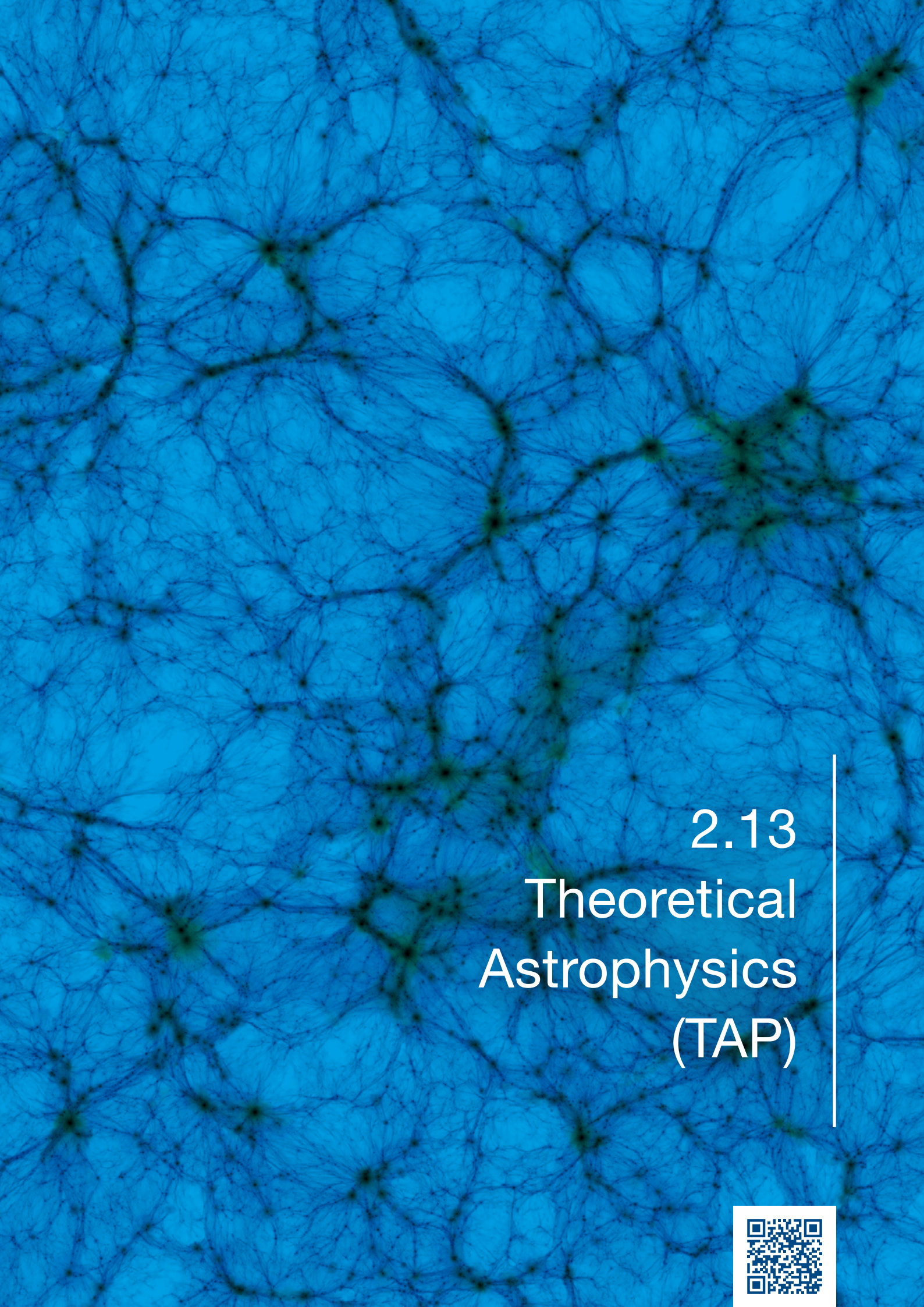
SABIO-RK web services can be used to automatically access the database, which is also used for the retrieval of kinetics data by third-party software tools and data workflows. These tools include CellDesigner, VirtualCell, Sycamore, SBMLsqueezer, cy3sabiork, Path2Models, The MCM group's LigDig, and FAIRDOMHub. Currently, SABIO-RK is accessed mostly (ca. 90 %) via web services, which underlines the importance of its integration into modelling and visualization tools.

In 2017, SABIO-RK was further improved functionally, and new features were implemented, as users had requested. I) Aside from SBML (Systems Biology Markup Language), SBBioPax, and spreadsheet export, SABIO-RK now supports the export of kinetics data in a variety of additional formats, including Matlab, BioPAX level 2 and level 3, DOT, Octave, XPP, and SBGNML. II) MetaNetX integrated SABIO-RK and implemented cross-references to compounds and reactions in SABIO-RK. MetaNetX is an online platform for accessing, analyzing, and manipulating genome-scale metabolic networks as well as biochemical pathways. III) SABIO-RK reactions were added to the biochemical reaction database BKMS-react, which is an integrated and nonredundant biochemical reaction database that contains known enzyme-catalyzed and spontaneous reactions. IV) The appearance of SABIO-RK on the web was improved (e.g., outdated Flash animations were replaced by modern JavaScript functionality). V) The data publication workflow in SABIO-RK was updated. In SABIO-RK, a complex procedure is required from the initial state of the entered 'raw' data until the final state, when it is ready to be searched through by users (e.g., the manually entered data have to be supplemented and annotated by additional unique identifiers to other databases). This process requires expertise and thus needs manual control. The full transfer procedure requires more than 30 individual steps, which have been performed semi-automatically in the past. To increase the speed of this routine and to lower the burden of manual work, the procedure was implemented as a mostly automated process, thereby allowing for faster and more frequent releases of new datasets. VI) Webservices and manual search functionalities were extended, and VII) the information provided in the data export was extended (e.g., the tissue definition is now provided as additional annotation). Kinetics data were extracted by the SABIO-RK students and curators from 304 publications and stored in the database for public retrieval. Both the curation and improvements were enabled by de.NBI, the German Network for Bioinformatics Infrastructure. As a shared de.NBI / COMBINE training event, we also organized a workshop on "Modelling and Simulation Tools in Systems Biology" at the ICSB conference in Blacksburg (Virginia), USA, which covered SABIO-RK in addition to other tools.

Figure 59: Distribution of selected kinetics parameters in SABIO-RK over the past 10 years.



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 more narrowly constrain the properties of dark matter and dark energy, the two enigmatic matter and energy components that dominate today's Universe and lie at the root of some of the most fundamental problems in modern physics.

Numerical simulations of both the collisionless and the hydrodynamic types play a prominent role in our work on a variety of scales. To this end, we develop novel numerical schemes that can be efficiently used on very large supercomputers. Our goal is to exploit these supercomputers at their full capacity to link the initial conditions of the Universe with its complex and evolved state today. The simulation models are indispensable for the interpretation of observational data and its comparison with theoretical models. Using simulations, we are able to study how diverse physical processes that are relevant in structure formation interact in a complex and highly non-linear fashion.

A current priority in our group is to incorporate physics into our models that is thought to be important but has thus far been often neglected, such as supermassive black hole formation, cosmic rays, or radiative transfer. In this report, we highlight 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. Rüdiger Pakmor
Dr. Christine Simpson
Dr. Dandan Xu
Dr. Martin Sparre *(until July 2017)*
Dr. Freeke van de Voort
Dr. Thomas Guillet
Dr. Felipe Goicovic *(since September 2017)*

Students

Rainer Weinberger
Christian Arnold *(until July 2017)*
Jolanta Zjupa
Svenja Jacob
Sebastian Bustamante

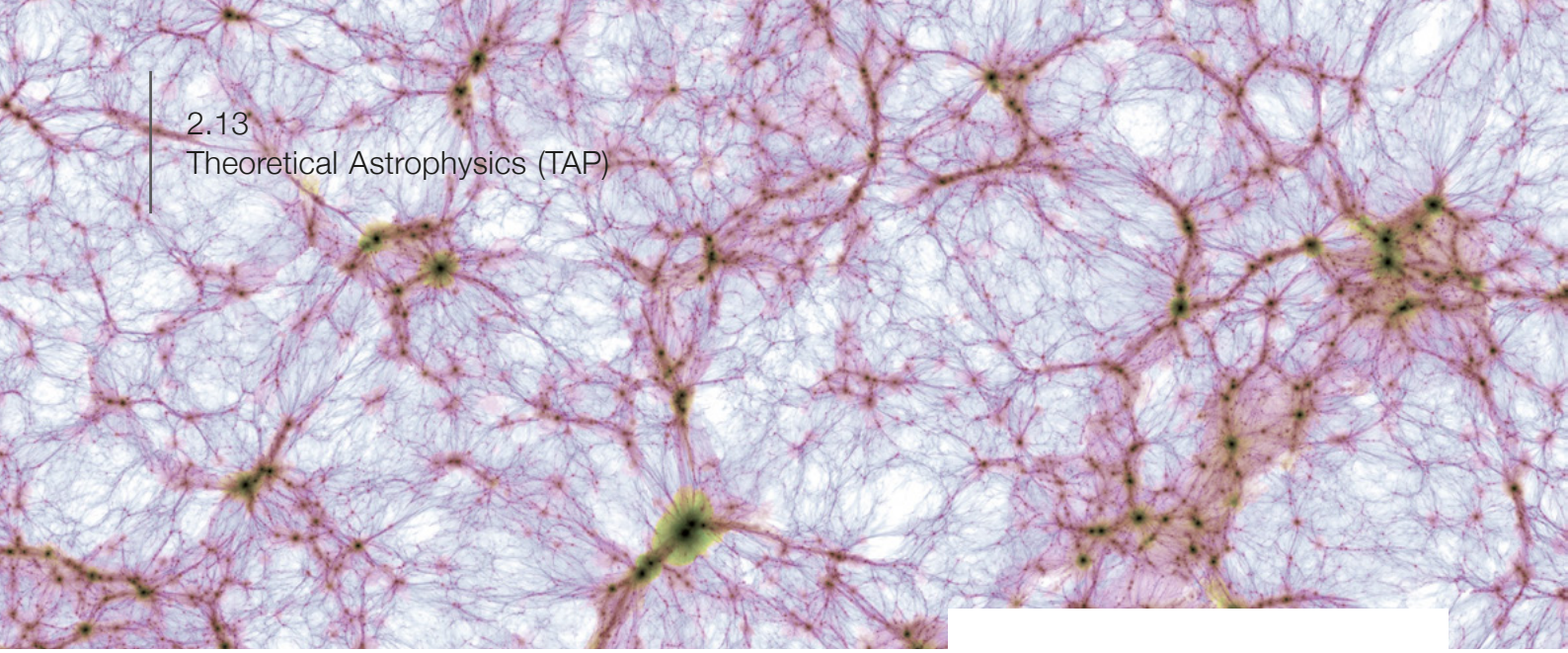


Figure 60: Thin slice through the cosmic large-scale structure in the largest simulation of the Illustris-TNG project, which contains more than 30 billion resolution elements. A projection of the baryonic density field is shown in which the image brightness indicates the projected mass density and the color hue depicts the mean projected gas temperature. The displayed region extends by about 1.2 billion lightyears from left to right.

Cosmic large-scale structure in the Next-Generation Illustris Simulations

Observed galaxies range in mass from a few thousand to a few trillion times the mass of the Sun, encompass physical sizes from a fraction to tens of kilo-parsecs, and span a variety of morphologies. Galaxies can reside in diverse environments – in isolation or as members of rich groups and clusters. They are self-gravitating systems of stars and gas embedded in a halo of dark matter, and their distribution throughout space outlines a cosmic web defined by filaments, nodes, sheets, and voids of matter. The highly clustered large-scale structure of the Universe today (at mega-parsec and giga-parsec scales) arose from more than 13 billion years of evolution, beginning with a nearly homogeneous distribution of matter in the early Universe.

In order to gain a theoretical, ab-initio understanding of the structure formation process and the role that different physical processes play in shaping galaxies, simulations that account for the multi-scale physics involved are the tool of choice. As a starting point, such calculations rely on initial conditions that are known and well constrained by observational data from the cosmic microwave background radiation, which can now be measured to exquisite precision by the Planck satellite. The calculations then need to accurately compute the dominant physical force (i.e., gravity) that acts upon all matter within the accelerating expansion of the Universe. Furthermore, (magneto) hydrodynamic processes for modeling the evolution of the gaseous component of the Universe need to be followed. Finally, it is necessary to account for all other relevant astrophysical processes, from atomic-level interactions that govern the radiative cooling of a metal-enriched gas to the formation of stars and supermassive black holes, with their subsequent expulsion of mass, metals, and "feedback" energy, which can have an impact on scales even larger than entire galactic halos. In 2014, we and a group of international collaborators created one of the first high-resolution, large-scale hydrodynamic simulations of galaxy formation. This simulation, dubbed "Illustris", used a novel and innovative numerical approach that employed a moving, unstructured mesh as realized in our AREPO code. One of the major achievements of Illustris was its ability to follow the small-scale evolution of gas and stars within a representative portion of the Universe, thereby yielding a population of thousands of well-resolved elliptical and spiral galaxies. For the first time, the simulation reproduced the observed morphological mix of galaxies and its dependence on stellar mass.

However, some aspects of Illustris were found to contradict a number of observable constraints. These important discrepancies could be traced to the Illustris feedback model from active galactic nuclei, thereby pointing to deficits in our physics model for supermassive black hole accretion and its associated energy release. In the last three years, we have therefore undertaken a campaign of new model development in order to address this key deficiency and also improve upon other lingering problems of the original Illustris simulation. This effort has led to a new kinetic wind model for black hole feedback physics

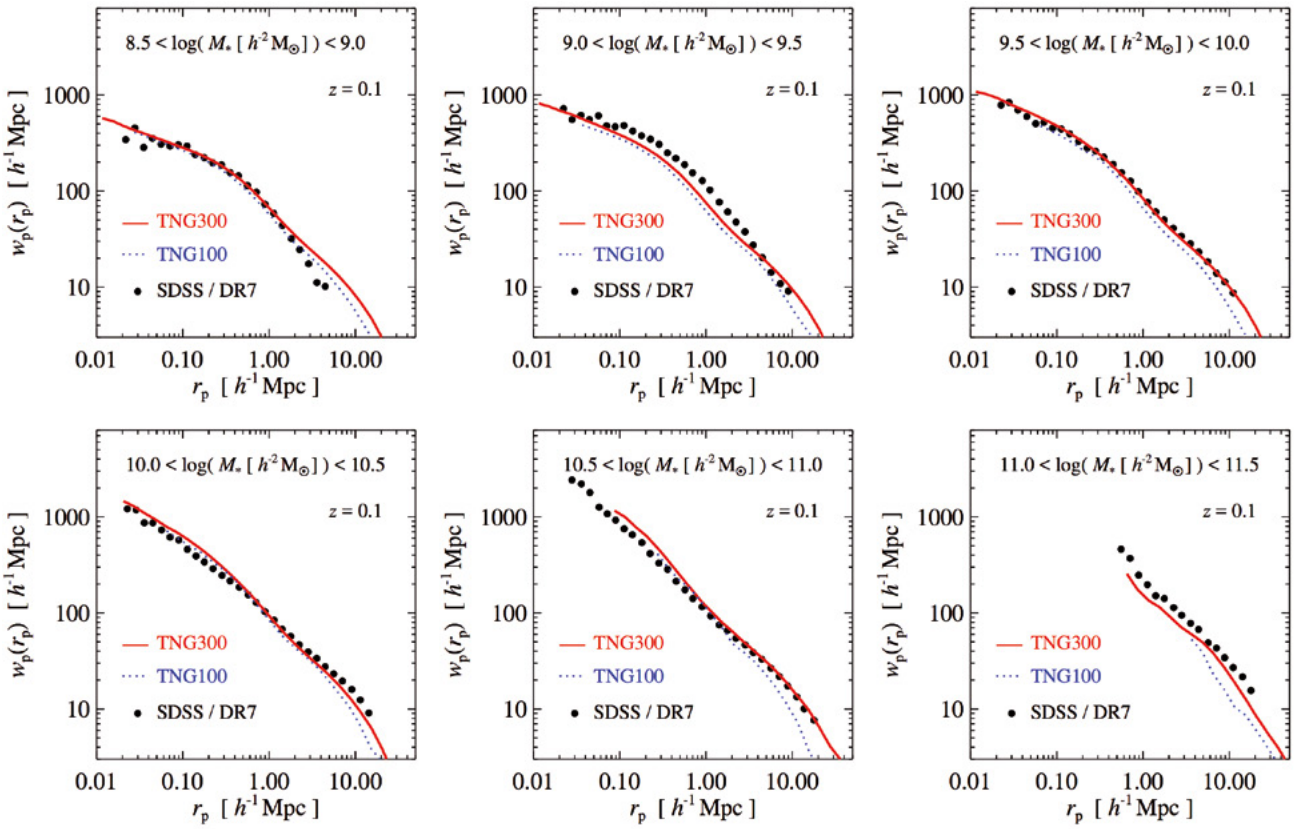


Figure 61: Comparison of the projected two-point galaxy correlation functions of TNG300 (solid) and TNG100 (dotted) at redshift z = 0.1 with data from the Sloan Digital Sky Survey (symbols) in six different stellar mass ranges.

[Weinberger et al. 2017a] and the realization of "The Next-Generation Illustris" (IllustrisTNG) simulations. This ambitious simulation program actually consists of three complementary cosmological simulations, with the runs TNG300 (large volume, box-size 300 Mpc), TNG100 (intermediate volume, box-size 110 Mpc), and TNG50 (small volume, box-size 50 Mpc). The calculations were carried out on the Hazel Hen supercomputer at the High Performance Computing Centre Stuttgart, with CPU-time awarded by the Gauss Centre for Supercomputing.

Figure 60 provides an overview of the cosmic large-scale structure as seen in the TNG300 simulation. Interestingly, hydrodynamic simulations of galaxy formation have now reached sufficient volume to make precise predictions for clustering on cosmologically relevant scales. For example, we have used our new IllustrisTNG simulations to study the non-linear correlation functions and power spectra of baryons, dark matter, galaxies, and haloes over an exceptionally large range of scales [Springel V, Pakmor R, Pillepich A, Weinberger R, Nelson D, Hernquist L, Vogelsberger M, Genel S, Torrey P, Marinacci F, Naiman J. First results from the IllustrisTNG simulations: matter and galaxy clustering. Mon. Notices Royal Astron. Soc. (2018) 475:676–698, 2018].

We have found that baryonic effects increase the clustering of dark matter on small scales and diminish the total matter power spectrum on scales of up to 3 Mpc. We have also found that the two-point correlation function of the simulated galaxies agrees well with observational data from the Sloan Digital Sky Survey (see *Figure 61*) both as a function of stellar mass and when split according to galaxy color, a remarkable result that has been achieved for the first time with cosmological hydrodynamic simulations.

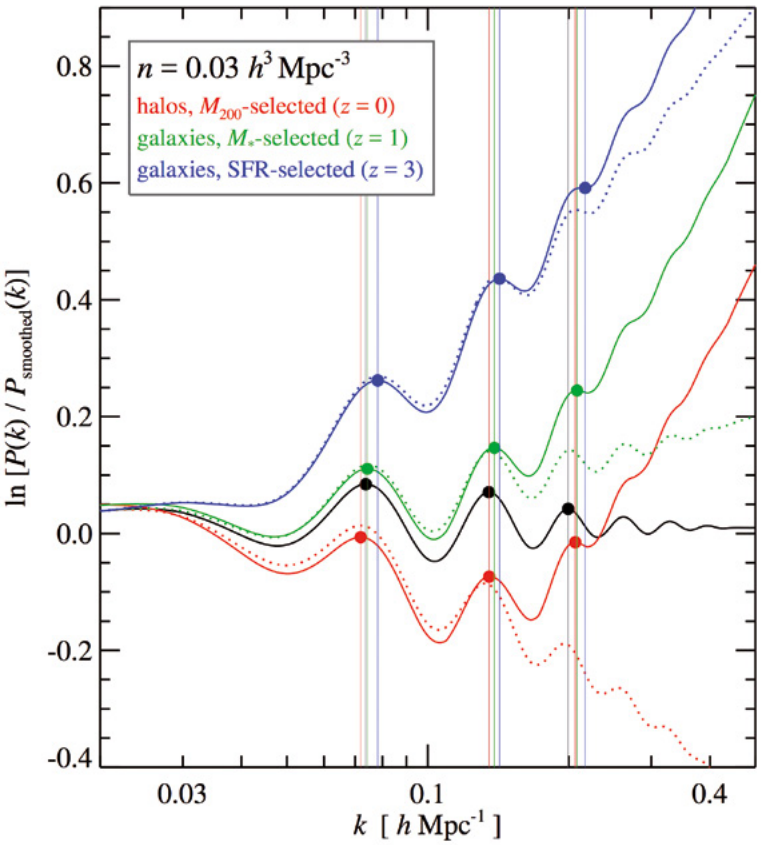


Figure 62: Power spectra estimates based on IllustrisTNG relative to the linearly evolved, initial power spectrum at different times. Three combinations of tracer type and redshift are shown with different colors (as labelled), with the relative bias on the largest scales renormalized to unity. The unperturbed linear theory baryonic acoustic oscillations are shown as a black line. The positions of the local maxima in the first three wiggles have been located and marked with circles and thin vertical lines. The scale-dependent bias leads to sizable shifts of up to 6% in the k-values of these peak positions.

Given this agreement, the TNG simulations can make valuable theoretical predictions for the so-called clustering bias of different galaxy samples, which has important cosmological implications. For example, we have detected significant scale-dependencies in the bias of different observational tracers of large-scale structure, extending well into the range of the baryonic acoustic oscillations that can be used to reconstruct the expansion history of the Universe and thus constrain dark energy. The resulting nominal shifts of the acoustic peaks can be around 5 % in spatial scale (see *Figure 62*) and correspond to a substantial mistuning of this cosmic ruler. Fortunately, it appears that this distortion is correctable with sufficient amounts of galaxy survey data thanks to the smooth variation of the bias with spatial scale.

These results merely scratch the surface of the rich scientific fields that can be addressed with the IllustrisTNG simulations. Indeed, already more than a dozen articles have been published on different aspects of the calculations, and a great deal of additional work is currently in preparation.

The interaction of jets with the intracluster medium

The short radiative cooling times of the gas in galaxy clusters – combined with the paucity of cold gas and star formation – suggest the presence of a central heating source. The energy from jets driven by the central supermassive black hole is widely considered to be a promising candidate for balancing the cooling losses. This hypothesis is observationally supported by the fact that most galaxy clusters with short cooling times show signatures of jet activity, and their jet power correlates with the cooling rate. However, how the highly collimated jets distribute energy in a volume-filling fashion to the cluster gas remains a topic of intense debate. Suggested mechanisms include heating by weak shocks and sound waves; mixing the lobe with surrounding material, cosmic rays, and turbulent dissipation; and mixing by turbulence, which may be promoted by anisotropic thermal conduction. Jets from supermassive black holes in the centers of galaxy clusters are also a prominent potential candidate for moderating gas cooling and subsequent star formation through depositing energy in the intra-cluster gas. In a recent project at HITS [Weinberger et al., 2017b], we therefore revisited this topic through an analysis of a new set of high-resolution magneto-hydrodynamic simulations of jets from supermassive black holes and their interaction with the surrounding medium.

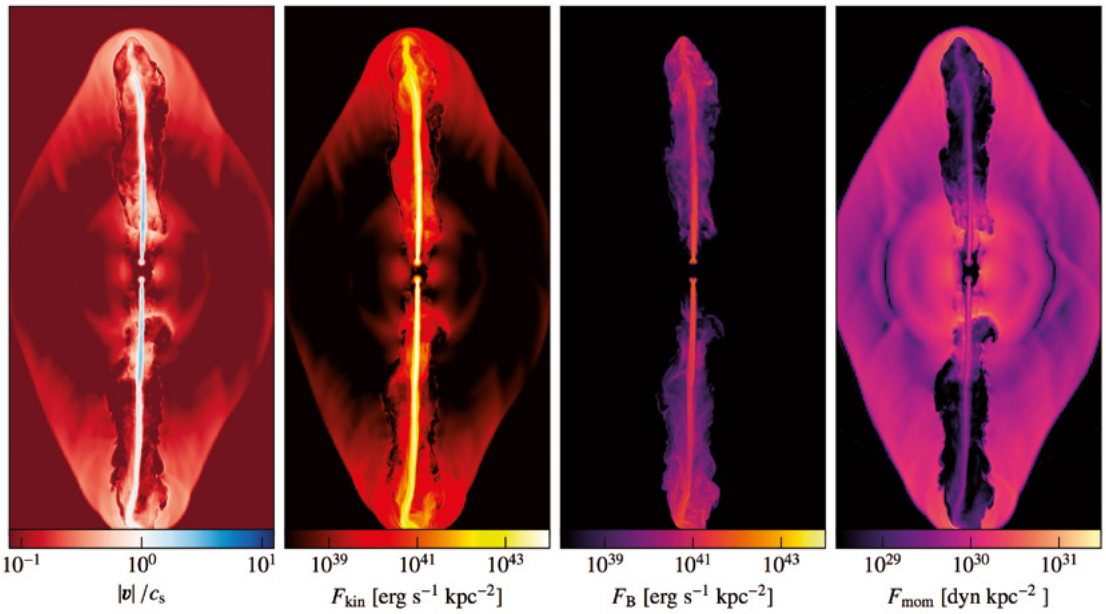


Figure 63: Left to right: Jet velocity, kinetic energy flux, magnetic energy flux, and momentum flux of a powerful jet after 42 mega-years, all measured in the black hole rest frame. Each panel is 200 kpc in the vertical direction and 100 kpc in the horizontal direction and displays averaged quantities over 10 kpc in depth.

We used idealized magnetohydrodynamic simulations that conserve the total energy of the gas – apart from the energy injection from the jet – in a stationary, spherically symmetric gravitational potential. This simple setup allows us to simulate the inflation of a low-density cavity in the surrounding intracluster medium and the subsequent lobe evolution and disruption after a few hundred mega-years. Specifically, we used the moving-mesh magnetohydrodynamics code AREPO to inject supersonic, low-density, collimated, and magnetized outflows into cluster centers (*Figure 63*), which are then stopped by the surrounding gas, thermalize, and produce low-density cavities filled with cosmic rays. In this manner, we could carry out high-resolution, non-radiative simulations of lobe creation, expansion, and disruption. We found that the dynamic evolution of the jet-inflated lobes (*Figure 64*) is in qualitative agreement with simulations of idealized low-density cavities that are dominated by large-scale Rayleigh-Taylor instabilities. The buoyant rise of the lobe does not create energetically significant small-scale chaotic motion in a volume-filling fashion, but rather a systematic upward motion in the wake of the lobe and a corresponding back-flow perpendicular to it. Overall, 50 % of the injected energy ends up in material that is not part of the lobe, and about 25 % remains in the inner 100 kpc. Thus, jet-inflated, buoyantly rising cavities drive systematic gas motions, which play an important role in heating the central regions of galaxy clusters; meanwhile, the mixing of lobe material is sub-dominant. Encouragingly, the main mechanisms responsible for this energy deposition can be modeled at resolutions that are within reach in future, high-resolution cosmological zoom simulations of galaxy clusters – something that we will pursue in forthcoming work.

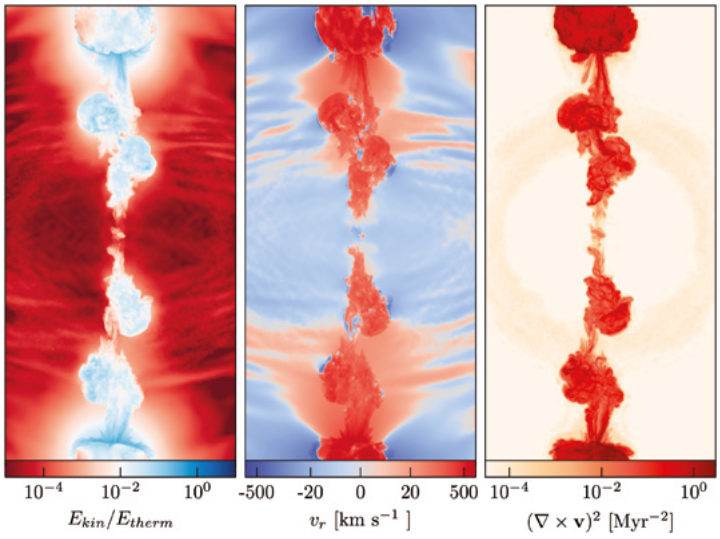


Figure 64: Left to right: Entropy, kinetic over thermal energy ratio, radial velocity, and vorticity squared of the intra-cluster medium after passage of a magnetized radio-lobe (after 168 mega-years). Each projection is 150 kpc wide, 300 kpc high, and 75 kpc deep.

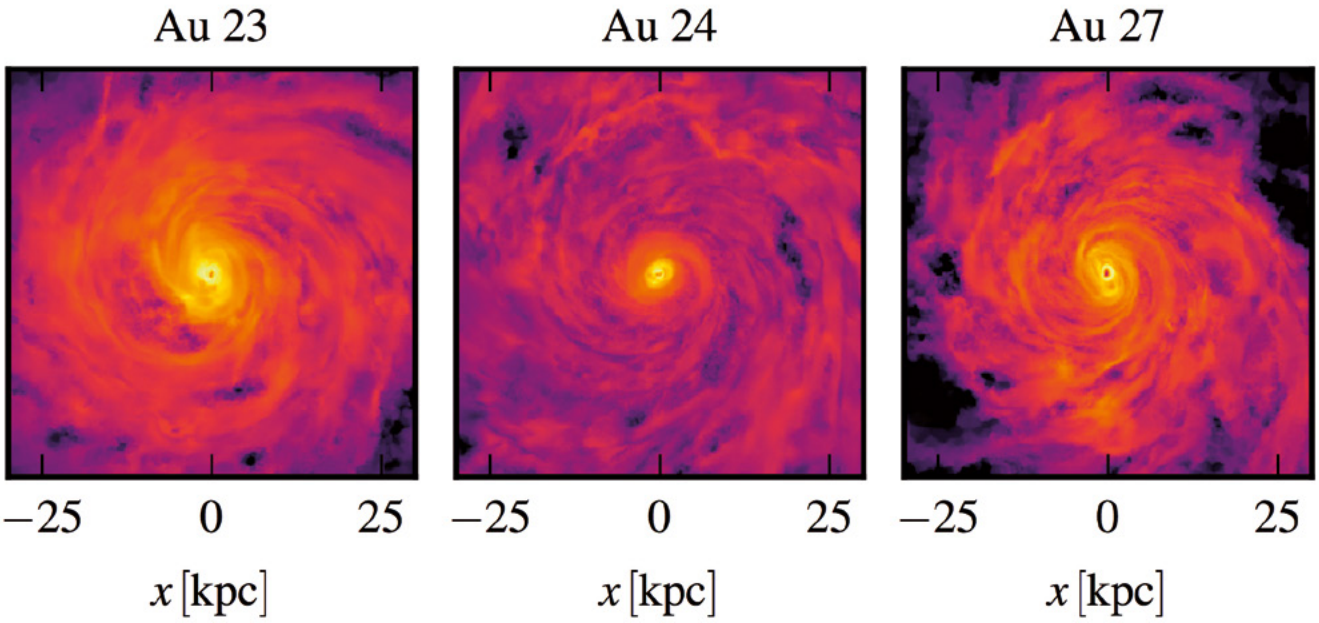


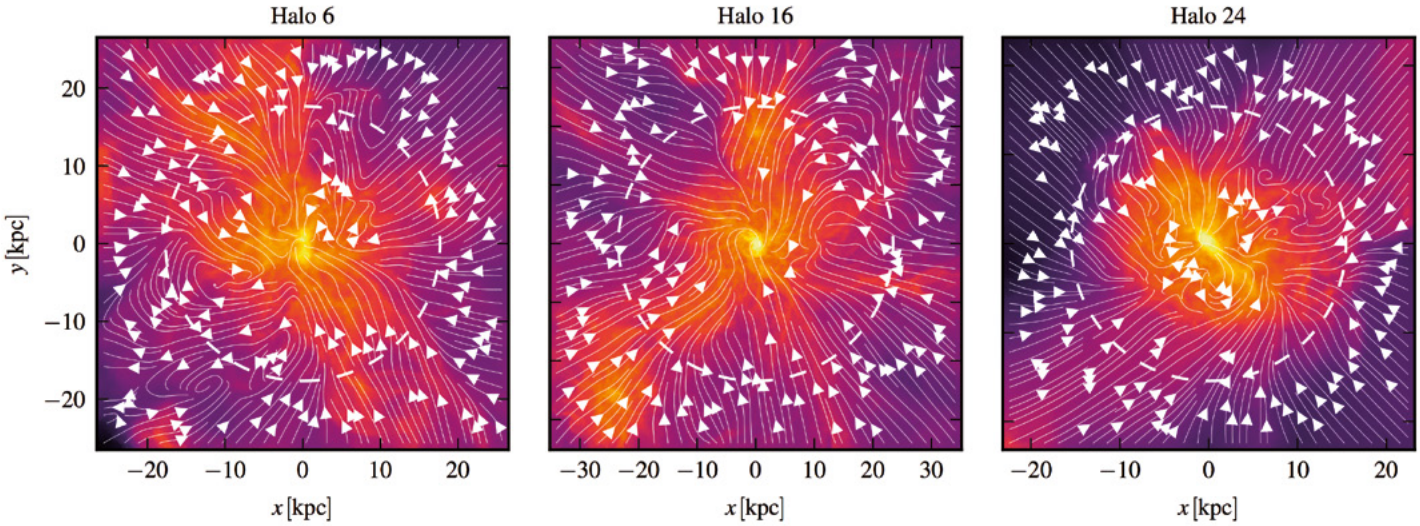
Figure 65: Magnetic field strength in the disks of two selected halos at redshift z = 0. The images show face-on projections with a depth of 2 kpc centered on the plane of the disk.

Magnetic field formation in the Auriga project

Magnetic fields in the interstellar medium have been observed for a large number of galaxies. One of the most important findings is that the magnetic and turbulent energy densities in nearby galaxies are roughly in equipartition. Therefore, at least today, magnetic fields are expected to be dynamically important for the evolution of galaxies. Moreover, strong magnetic fields have also been measured in galaxies at much earlier times, up to redshift z = 2. In addition to their direct influence on gas dynamics, magnetic fields also play an important role in controlling the direction of anisotropic transport processes of charged particles, which is of immediate relevance for thermal conduction and the motion of cosmic rays.

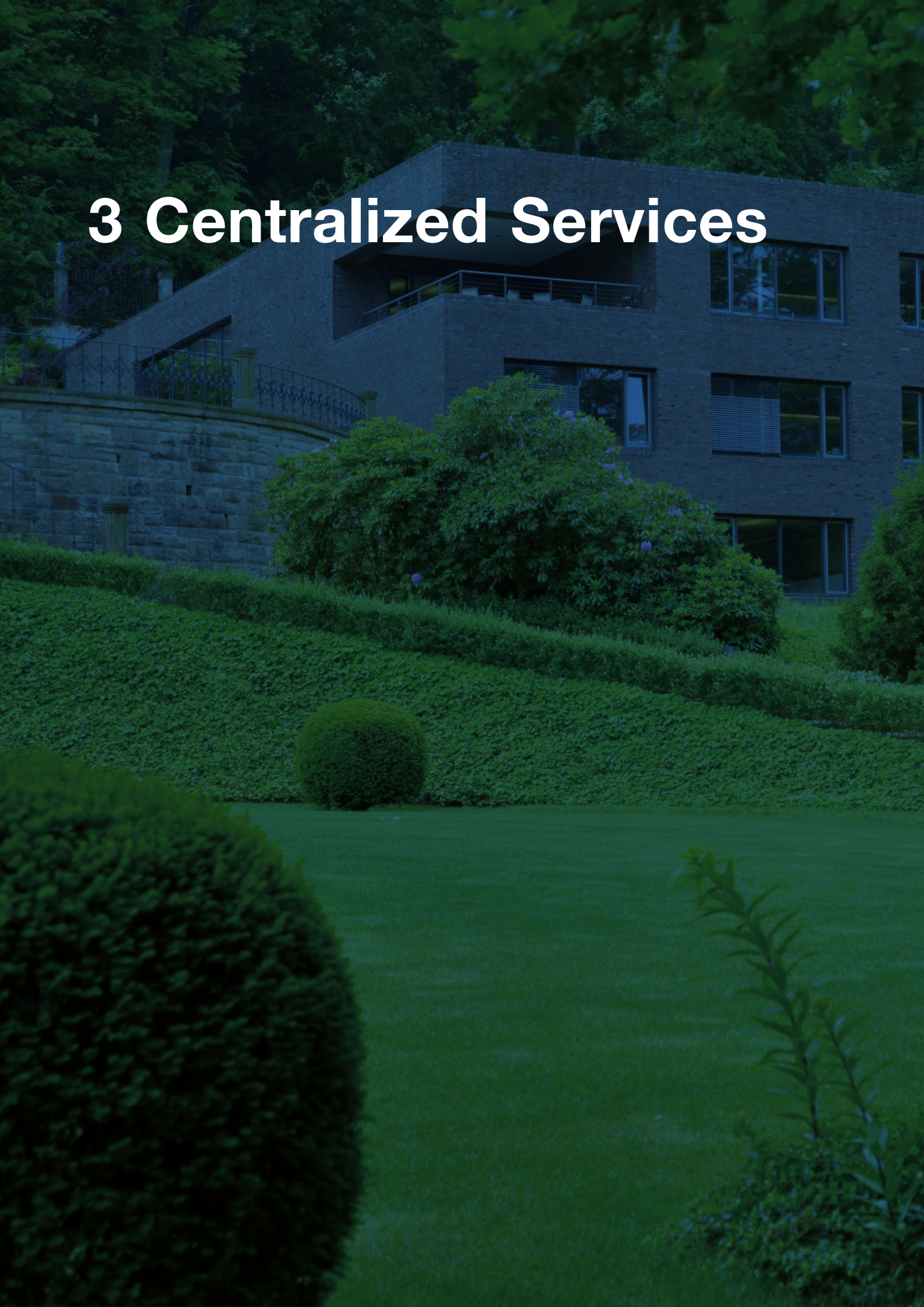
Despite their importance, the origin and evolution of galactic magnetic fields is still not fully understood. To reach their observed strength, magnetic fields have to be amplified by several orders of magnitude from cosmological seed fields or weak magnetic fields generated in stars and supernovae. The amplification process is highly non-linear and is inherently coupled with the complex baryonic physics that regulates star formation and the formation and evolution of galaxies. This process is thus best studied by means of detailed numerical simulations that attempt to take all relevant processes into account. Utilizing some of our newly developed modeling capabilities, we recently carried out the first set of 30 high-resolution cosmological zoom-simulations of the formation of Milky Way-like galaxies, including magnetic fields and a sophisticated modeling of galaxy formation physics, entitled the "Auriga" project [Grand et al., 2017]. We also extensively studied the overall evolution of the magnetic field strength and how it correlates with other properties of galaxies [Pakmor et al., 2017]. In *Figure 65*, we display examples of the final magnetic field strength in some of the formed disk galaxies.

We found that the magnetic fields grow exponentially at early times owing to a small-scale dynamo with an e-folding time of roughly 100 mega-years in the center of halos. This dynamo is driven by turbulent gas flows that originate in the tumultuous rapid growth phases of the



earliest galaxies. Saturation of the dynamo amplification occurs around redshift z = 2 – 3, when the magnetic energy density reaches about 10% of the turbulent energy density with a typical strength of 10 – 50 μG. In the galactic centers, the ratio between magnetic and turbulent energy subsequently remains nearly constant until z = 0. At larger radii, differential rotation in the disks leads to a further modest amplification that is linear in time. Interestingly, the final radial and vertical variations of the magnetic field strength can be well described by two joint exponential profiles and are in good agreement with observational constraints. Overall, however, the magnetic fields have only little influence on the global evolution of the star formation rate of the galaxies as it takes comparatively long to reach equipartition, at which point most of the stars of the galaxies have already formed.

Figure 66: Gas density projections of three high-resolution halos at redshift z = 5. The color-coding depicts the logarithmic column density of the gas. The streamlines reveal the direction of the projected velocity field. The dashed line shows the virial radius of the halos at this cosmic epoch.



3 Centralized Services

3.1 Administrative Services

3.2 IT Infrastructure and Network (ITS)

3.1 Administrative Services

Group Leader

Dr. Gesa Schönberger

Staff members

Christina Blach *(office)*
Christina Bölk-Krosta *(controlling)*
Benedicta Frech *(office)*
Ingrid Kräling *(controlling)*
Kerstin Nicolai *(controlling; until June 2017)*
Thomas Rasem *(controlling; since June 2017)*
Rebekka Riehl *(human resources and assistant to the managing director)*
Stefanie Szymorek *(human resources)*
Thu Phuong *(student assistant; until September 2017)*

The HITS administration takes care of almost all administrative tasks for the scientific groups. It looks after the day-to-day running of the two HITS sites, oversees personnel and accounting, deals with legal issues, and organizes events.

Administration is always characterized by attention to detail. Individual questions have to be clarified, and concrete problems need to be solved. It would thus be impossible to provide a comprehensive list of all of the administration's work. Instead, we present a few selected topics from 2017 that kept us busy in the administration. Every change of personnel brings about new workflows. When Michael Strube (NLP) took over Rebecca Wade's (MCM) position as Scientific Director at the beginning of 2017, a good and effective collaboration was quickly achieved. The year also saw personnel changes in the area of controlling, which gave us an incentive to redistribute tasks.

Beginning in the spring, we prepared for Christoph Pfrommer's (HAC) move to the Leibniz Institute for Astrophysics in Potsdam. In terms of the administration, this move required controlling and the human resources team to ensure a smooth transfer of the group along with the existing ERC grant (see *Chapter 2.6*). Later that year, the departure of Siegfried Schloissnig and his CBI group was scheduled for January 2018 (see *Chapter 2.2*). In the summer, the two-day alumni meeting with 54 alumni from HITS and its predecessor institutions (EML and EML Research) as well as 50 current HITS employees took place. In the process, contacts were strengthened and new connections were made (see *Chapter 5.4*).

Finally, a mixed team consisting of members from the communications and the administration groups worked hard to improve internal communications. In so doing, the intranet was entirely restructured with the goal of increasing user-friendliness through various search options, namely a so-called form finder. In the autumn of 2017, the new intranet went online. It not only provides interesting information from the administration but also keeps users informed about current developments at the institute. We close with a few figures about the institute:

In HITS, a total of 12 research groups worked in the autumn and winter of 2017, including two junior groups and two associated groups. Six of the twelve group leaders were jointly appointed at KIT Karlsruhe and the University of Heidelberg and were assigned to HITS. At the end of the year, a total of 90 employees were active at HITS. Including scholarship holders, visiting scholars, delegates, and students, more than 130 people worked at HITS in 2017.

HITS oversaw a total of 43 research projects in 2017, 31 of which were financed via third-party funds. With the support of the European Union alone, HITS participated in 13 projects, including three large grants from the European Research Council (ERC), seven projects under the EU Research Framework Program Horizon 2020, and two in the so-called 7th Framework Programme for Research and Technological Development (FP7). Five additional projects funded by various third-party sponsors were approved in 2017 and will begin in 2018, including an ERC Starting Grant.

Group Leader

Dr. Ion Bogdan Costescu

Staff members

Dr. Bernd Doser *(Software Developer)*
Dr. Christian Goll *(System Administrator, until June 2017)*
Cristian Huza *(System Administrator)*
Norbert Rabes *(System Administrator)*
Andreas Ulrich *(System Administrator)*
Dr. Nils Wötzel *(Software Developer, until July 2017)*

Students

Stevan Bratic *(until May 2017)*
Alexander Rausch *(until October 2017)*

3.2 IT Infrastructure and Network (ITS)

One of the things that marked the IT world in 2017 was an acute crisis of GPU cards. We contributed somewhat to this crisis by acquiring 48 NVidia P40 cards and 4 NVidia P100 cards and adding them to our HPC cluster environment. However, unlike most of the world, which uses GPUs for the so-called mining of cryptocoins, the HITS scientists perform machine learning, molecular dynamics simulations, or linear algebra calculations. Our "harvest" consists of scientific results in areas such as the classification of images from space telescopes, analyzing natural language, and studying interactions between biological molecules or their mechanical properties. Our GPUs are paired with powerful Intel CPUs, allowing scientific applications that can distribute their computation over both CPUs and GPUs (like the molecular dynamics code GROMACS) to run optimally. The input and output data are stored on our large storage systems and accessed at high speed via an FDR Infiniband network used together with the other cluster nodes. This setup provides the HITS scientists with access to the same data from both CPU-only and GPU-equipped cluster nodes and thus allows them to choose the optimal architecture for a calculation and to mix calculations of different types in a workflow.

As our HPC cluster environment is now split between the computer rooms at HITS and at the University of Heidelberg (see Annual Report 2016), we have looked into the possibility of making a storage system from one computer room directly accessible from the compute nodes in the other room. This process would involve connecting the Infiniband networks from both sides over a distance of around 11km. Although long-range Infiniband networking solutions do exist, the calculated total cost of ownership over a longer time has proven unreasonably high when compared with the advantages that it would bring, namely access to data without the need to copy it first and a reduction in duplicated data, which could also be achieved through better data management. Furthermore, such a solution would only make direct access to a remote storage system possible, but would not increase the total available capacity. Indeed, in the first part of the year, the need for such an increase quickly became apparent, in order to cope with the amounts of data generated in our HPC cluster environment as well as data produced by HITS projects at large HPC centers and transferred back to us for further processing or analysis. Finally, we decided to extend our storage systems based on BeeGFS by another petabyte of fast storage, thereby raising the total capacity to more than 4.5 petabytes.

Hardware replacements in our server and network infrastructure also kept us busy in 2017. The storage system used for our central backup facilities had begun showing its age, so its function was assumed by a faster and more extensible device. The new system has a larger amount of cache and uses it more aggressively such that metadata operations involving many small files are significantly sped up. After renewing the core of our network infrastructure at the end of the last year, we replaced the network switches connecting our workstations and servers in 2017. In addition, we extended and improved the Wi-Fi coverage on the HITS campus, which involved replacing old access points and installing new ones, connecting them to a new network switch that offers Power-over-Ethernet, and setting them up to be centrally managed. Finally, we started the process of replacing the institute firewall towards the end of the year, but this had to be postponed until the beginning of 2018 for technical reasons.

4 Communication and Outreach


Heidelberg Institute for
Theoretical Studies
HITS
Heidelberg Institute for
Theoretical Studies
HITS
Heidelberg Institute for
Theoretical Studies
HITS
Heidelberg Institute for
Theoretical Studies
HITS

4 Communication and Outreach

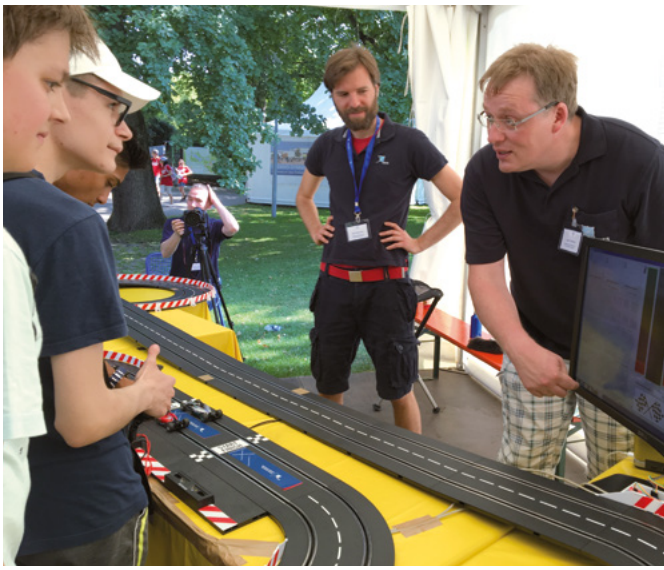


Figure 67: Ready, Set, Go: Kai Polsterer (right) and Nikos Gianniotis (middle, both AIN) explaining school students how to save energy in a racing car game.

The HITS communications team works to establish "HITS" as a brand name for a small but excellent interdisciplinary research institute that is perfectly suited for young, aspiring researchers. At the same time, we want to draw the media's attention to the excellence of our scientists in their respective fields. Last but not least, we help in building up the HITS alumni network as a hub for strengthening the ties between current and former HITSters.

In 2017, these goals were supported by a plethora of events and several scientific achievements that our researchers made. Some highlights are presented below.

Frauke Gräter, head of the Molecular Biomechanics group, won the PRACE Ada Lovelace Award for High Performance Computing (HPC). This was only the second time that the "Partnership for Advanced Computing in Europe" (PRACE) association had awarded a female scientist for her outstanding impact on HPC research at a global level. The award was presented at the PRACEdays conference in Barcelona, Spain, in May 2017. In July, "Groups and Geometry" (GRG) group leader Anna Wienhard was invited to give a presentation at the next International Congress of Mathematicians (ICM) in Rio de Janeiro, Brazil, in 2018. Being invited to give a lecture at the ICM is among the most prestigious recognitions of research accomplishment in mathematics.

In November, we were happy to announce that the European Research Council had awarded HITS astrophysicist Andreas Bauswein (Physics of Stellar Objects group) an ERC Starting Grant worth approximately 1.5 million euros. The aim of Andreas's project is to better understand the collisions between neutron stars – the potential candidates for the formation of heavy elements, such as gold and silver – by using computer simulations.

In the same month, we released the news that three HITSters again ranked among the group of most highly cited researchers worldwide. The "Highly Cited Researchers" report by Clarivate Analytics stated that the publications of Tilmann Gneiting (Computational Statistics (CST)), Volker Springel (Theoretical Astrophysics (TAP)), and Alexandros Stamatakis (Scientific Computing (SCO)) rank in the top one percent by citations in their field and publication year in the Web of Science.

With their current research, Gneiting, Springel, and Stamatakis also made an impact on the science media with their studies on the "Forecaster's dilemma" (in "Statistical Science") and the "Auriga" galaxy simulations (in "Royal Astronomical Society") as well as with a paper on protists in the rainforest (in "Nature Ecology and Evolution").

Figure 68: The "Girls' Day" 2017 at HITS.

Media relations: Journalist in Residence program

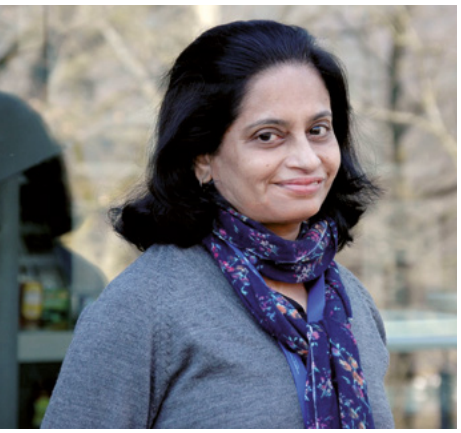


Figure 69: T.V. Padma, HITS Journalist in Residence 2017.

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 pressures of the "daily grind." In February 2017, HITS welcomed its sixth Journalist in Residence, T.V. Padma from Delhi (India), for a five-month stay. Padma is a freelance science journalist who writes for Nature, New Scientist, Physics World, and other media outlets. During her stay, she held an internal seminar with the HITSters on social media and gave a public speech on science and science journalism in India. Moreover, she got to know Heidelberg better. "I discovered that this city is full of science, an ideal place for a science journalist", she stated. Padma returned to Delhi with some new projects that had begun at HITS. In the summer, HITS began its next call for applications. Peter Saueressig promoted the program at the World Conference of Science Journalists in San Francisco in October, where over 1,300 international science journalists were in attendance. The campaign – amplified by our social media channels – proved successful: We received a great number of applications by excellent candidates from six continents. The jury, which consisted of science journalists and scientists from universities, Max Planck institutes and HITS, chose award-winning German TV journalist Kerstin Hoppenhaus as the "HITS Journalist in Residence" for 2018. She will join us in June and spend her time investigating new audiovisual formats in science journalism.

Regular and special outreach activities: From STEM girls to Nobel laureates

On April 27, HITS once again participated in the national "Girls' Day". The goal of this yearly event is to expand the minds of young girls and to get them interested in, say, a STEM subject, such as research at HITS. Twenty-five girls between the ages of eleven and fifteen visited us to participate in our workshops and find out what it's like to be a HITS researcher. Scientists from three research groups (CST, MCM, and TAP) offered small-scale, hands-on workshops to show the girls what the daily work

Figure 70: In Lindau: Agnieszka Obarska-Kosinska (right) and Christopher Zapp (left) explaining their research to Ulrich Steinbach (middle), Ministry of Science (image: Staatsministerium Baden-Württemberg).

and life of a researcher look like in fields ranging from molecular biology and mathematics to astrophysics. After the workshops, the girls had lunch with the scientists and enjoyed a guided tour through the HITS server room.

For the second time, HITS researchers were invited to the "Netzwerk Recherche" ("Network Research") conference in Hamburg, one of the biggest annual events in German journalism. On June 10, group leaders Kai Polsterer (AIN) and Michael Strube (NLP) were speakers on a panel about data science and its implications for data journalism.

As in previous years, HITS took part in the Explore Science event, which once again took place in Mannheim's Luisenpark from June 21-25. 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 2017, Explore Science's motto was "Energy", and the event attracted more than 45,000 visitors from all over the region. HITS researchers from seven groups presented the topic in three elaborated interactive hands-on stations: "Gravity Pong", a touch screen computer game in which the gravity of black holes is used to move the ball (i.e., the Earth) into the opponent's field; "Ready, Set, Go!" an energy-saving game that uses a Carrera slot-car racing set; and a multiple-choice "Energy Quiz" with questions about energy sources and consumption.

On June 30, two HITS researchers took part in a special outreach activity: For the first time, the institute participated in the science exhibition at the Nobel Laureates' Meeting in Lindau, Germany. Agnieszka Obarska-Kosinska and Christopher Zapp from the MBM group presented the institute and their research on "robust protein materials in the stress test" to the laureates and to journalists and politicians. In the summer, the institute again participated in the International Summer Science School Heidelberg (ISH). The SDBV group hosted two school students (one each from Israel and the U.S.), who learned more about the "Operation Explorer" (see *Chapter 6*) and programmed a 3D map presentation. In September, another event conceived by the Klaus Tschira Stiftung took place: the 5th installment of the Heidelberg Laureate Forum (HLF). This year, we again hosted a group of young researchers from all over the world (see *Chapter 6*).

Figure 71: Rebecca Wade talking to the students from Delft, the Netherlands.

Head of Communications

Dr. Peter Saueressig

Staff member

Isabel Lacurie

Students

Richard Seidemann *(from February to July 2017)*
Julia Klawitter *(since August 2017)*

Guests from academia and politics

This year, we received visitors from both academia and politics at the institute. In May, we saw a group of 30 students from the Study Association Nanobiology at the Technical University of Delft, the Netherlands. Michael Strube gave an introduction to HITS and its research, which was followed by speeches from Rebecca Wade and Csaba Daday on the institute's most salient molecular biology topics. The visit also included a poster session and a discussion with researchers from the MCM and MBM groups.

In July, we hosted Theresia Bauer (Minister of Science and the Arts, Baden-Württemberg), and members of the Deutsche Bundestag Franziska Brantner and Konstantin von Notz for a conversation on big data, data security, and data protection (see *Chapter 5.3*). Finally, in October, a group from the Heidelberg Club International (HCI) visited HITS for a guided tour through the institute and the gardens as well as an overview of HITS scientific work by Michael Strube and a talk on galaxy formation and supercomputers by Volker Springel.

Speaking up for science

Scientists usually speak about the details of their work or the methods they apply. This year, however, our scientists – as well as many others across the world – felt compelled to raise their voices in support of science itself. On April 22, HITSters joined the "March for Science" in Heidelberg – a demonstration for freedom, democracy, and reason. Our Scientific Director Michael Strube addressed a crowd of 1,800 people at the University Square and also appeared at several other public events related to the "science march" theme (see *Chapter 5.3*).

A network for the future

HITS is 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. In summer 2017, we organized a two-day alumni meeting that focused on interactions between alumni and current HITSters and featured a public talk as well as career talks by alumni (see *Chapter 5.4*).

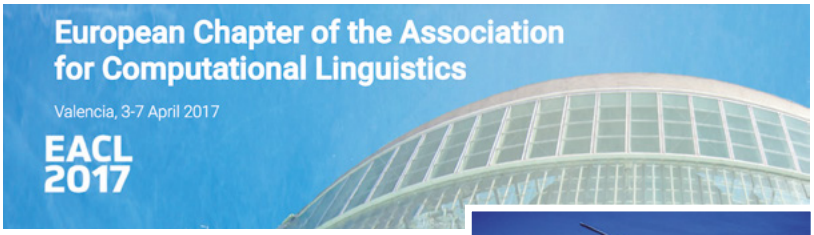
5 Events

5.1 Conferences, Workshop & Courses

5.2 HITS Colloquia

5.3 Science March and more

5.4 HITS Alumni Meeting



5.1.1 First ACL Workshop on Ethics in Natural Language Processing

April 4, 2017, Valencia, Spain

For a long time, Natural Language Processing (NLP) was thought to involve doing linguistics with computers. As such, it didn't have any applications in industry, it was not of interest to governments, and it did not have an impact on people's lives; however, this has changed dramatically in the last couple years. Search engines process and analyze queries and reply not only by retrieving webpages of interest but also via condensed information. Moreover, dialogue systems guide customers through call centers and help desks, machine translation is real and works for very many languages, students' essays can be scored automatically, customers' product reviews can be analyzed for sentiment, and people even have devices in their homes that perform tasks when provided with speech input. While all of this technology may sound useful, there is also a downside to these technical achievements: The corporations that provide search and translation services collect, process, and analyze data about their users. Furthermore, phone calls, emails, and chat communication is intercepted by governments and automatically analyzed for suspicious behavior, and the collected data are used to train machine-learning algorithms that accurately reproduce any prejudice contained in the training data. To make the NLP community aware of cases of dual use and of bias in machine learning applied to language, NLP group leader Prof. Michael Strube co-organized the Workshop on Ethics in Natural Language Processing that was held in Valencia in April 2017 at the largest European conference in the field, the EACL 2017. The workshop featured guest researchers as speakers in the fields of NLP, artificial intelligence, philosophy, and law. Papers submitted to the workshop dealt, inter alia, with gender bias in language data, with the problem of the reproducibility of research results, and with best practices in NLP research both in general as well as when applied to medical data. The workshop concluded with a panel discussion that quickly turned into an open discussion with the audience. The timeliness of the workshop topic were confirmed by the fact it drew the largest audience of all workshops at the EACL with about 80 registered participants and sometimes more than 100 people in the room. The workshop topic was picked up by mainstream papers in the field. There will also be a second event in June 2018 at the North American conference in our field, the NAACL 2018.

5.1.2 Wellcome Trust Advanced Course on Computational Molecular Evolution

May 8–19, 2017, Hinxton, Cambridge, UK

This practical course took place for the 9th time at the Wellcome Genome Campus in Hinxton near Cambridge, UK. It 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's foremost scientists and authors of famous analysis tools in evolutionary bioinformatics, including Tanja Stadler, Nick Goldman, Bruce Rannala, and Ziheng Yang. Alexandros Stamatakis (SCO group leader) has been a co-organizer of this event since 2011. Former SCO lab member Paschalia Kapli and current SCO member Lucas Czech made substantial contributions as teaching assistants. The 10th iteration of this course will take place in Crete in 2018 and will again be co-sponsored by HITS.

Figure 72: Participants of the workshop "Bridging Nuclear and Gravitational Physics: The Dense Matter Equation of State" in Trento.

5.1.3 Bridging nuclear and gravitational physics: The dense matter equation of state

June 5 – 9, 2017, European Center for Theoretical Studies in Nuclear Physics and Related Areas (ECT*), Trento, Italy

The Physics of Stellar Objects (PSO) group co-organized the workshop "Bridging Nuclear and Gravitational Physics: The Dense Matter Equation of State" at the European Center for Theoretical Studies in Nuclear Physics and Related Areas (ECT*) in Trento, Italy (http://www.ectstar.eu/). The meeting aimed at gathering international experts in the fields of nuclear physics and gravitational-wave astronomy. Both fields are tightly connected because the detailed signal properties of gravitational waves can reveal the currently unknown nuclear equation of state and thus describe the fundamental properties of high-density matter. The workshop fostered communication between the different communities, which is necessary to maximize the scientific value of current and future detections. The format of the workshop – with its numerous informal discussions – was well received by the participants.

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

August 6, 2017, Blacksburg, Virginia, USA

Systems biologists from all around the world gathered at the 18th International Conference on Systems Biology (ICSB) in Blacksburg, Virginia (USA), from August 6th to 12th, 2017 (http://www.cpe.vt.edu/icsb2017/). HITSter Martin Golebiewski (SDBV) organized and coordinated the one-day COMBINE & de.NBI Tutorial Workshop "Modeling and Simulation Tools in Systems Biology" as a conference satellite (http://co.mbine.org/events/tutorial2017). The tutorial showed the mainly young scientists from many countries how to set up computer models of biological networks and simulate these models in different software platforms. Lectures, software demonstrations, and hands-on sessions provided the attendees with the necessary skills to create, simulate, and handle such models.

As an important prerequisite for the information and data exchange among scientists, domain-specific community standard formats were demonstrated to be crucial to the data and model exchange. Such modeling standards for the life sciences are defined by the COMBINE network (http://co.mbine.org), which co-hosted the tutorial. Martin Golebiewski, as a member of the COMBINE coordination board, provided an overview of the network's standards. An international team of tutors then instructed the scientists how to use modeling- 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, and Martin Golebiewski) presented their SABIO-RK database for reaction kinetics data and the SEEK system (FAIRDOMHub) for integrated FAIR data and model management. Other software tools were also introduced, such as COPASI for the simulation and analysis of biochemical networks and their dynamics, which was demonstrated by HITS alumni Ursula Kummer and Sven Sahle (Heidelberg University) as well as by Pedro Mendes (UConn Health Center, USA) and their teams. Moreover, the software CellDesigner (Akira Funahashi, Japan) and VirtualCell (VCell, Leslie M. Loew, UConn Health Center, USA) – two widespread modeling and simulation platforms – were also presented.

Figure 73: Group photo in the garden of the International Academic Forum, Heidelberg.

5.1.5 The science of forecasting: Models and data instead of crystal balls

"Summer School" in Heidelberg as part of the "ScienceFore" Project, led by Tilmann Gneiting (HITS). The project is funded by an Advanced Grant from the European Union.

Figure 74: Sebastian Lerch (CST group) talking at a tutorial session.

"Que sera, sera": As the future is uncertain, forecasts need to be probabilistic in nature and take the form of probability distributions over future quantities or events. Accordingly, a transdisciplinary transition from point forecasts to probabilistic forecasts is well underway. The ScienceFore project, led by HITS group leader Tilmann Gneiting (CST), seeks to provide guidance and leadership in this transition by developing the theoretical foundations of the science of forecasting, cutting-edge statistical methodology, and applications in meteorology and economics. ScienceFore is supported by an Advanced ERC grant from the European Research Council (ERC). From October 3 – 6, 2017, the CST Group together with HITS alumnus Fabian Krüger (Heidelberg University) organized the "ScienceFore Summer School" in the International Academic Forum, Heidelberg.

Figure 75: Scientific session.

About 40 participants from four continents discussed topics including forecast evaluation, forecasting across disciplines, and statistical postprocessing. The invited speakers were Barabara Rossi from Pompeu Fabra University (Barcelona, Spain), Thordis Thorarinsdottir from the Norwegian Computer Center (Oslo, Norway) as well as Tom Hamill and Michael Scheuerer from the U.S. National Oceanic and Atmospheric Administration (NOAA, Boulder / Colorado).

This project receives funding from the European Union's 7th Framework Programme for Research, Technological Development, and Demonstration under Grant Agreement No 290976.

5.1.6 LiSyM/FAIRDOM/ de.NBI Data Structuring Course

November 22 – 23, 2017, Hünfeld, Germany

The SDBV group organized a Data Structuring Course for scientists in the life sciences (https://seek.lisym.org/events/16) with a special focus on systems biology and systems medicine. Martin Golebiewski and Olga Krebs coordinated this two-day training event in Hünfeld (near Fulda), a small German town where famous computer pioneer Konrad Zuse is buried. The course was targeted at wet-lab scientists and people who use experimental and/or clinical data for research and was organized and funded by the Liver Systems Biology Network LiSyM (http://www.lisym.org) (in which the SDBV group is responsible for the central data management) in conjunction with the initiatives FAIRDOM (https://fair-dom.org), ERASysAPP (European ERA-Net for Systems Biology Applications), and the German Network for Bioinformatics Infrastructure de.NBI (https://www.denbi.de).

The purpose of this highly interactive training course was to provide users with an impression and hands-on training regarding how they could use a variety of specific tools to prepare and structure their research data and corresponding metadata. Introductory lectures, software demonstrations, and hands-on sessions provided the attendees with the necessary skills to structure and describe their research data in a meaningful and efficient manner. Topics included best practice data structuring for experimental and (clinical) sample data; standards for data formats, metadata, and data annotations; using SEEK & FAIRDOMhub for integrated data structuring and data management; and many others. The tutors were Olga Krebs, Martin Golebiewski, and Wolfgang Müller from SDBV, supported by Maja Rey and some experienced users of the systems developed by the group.

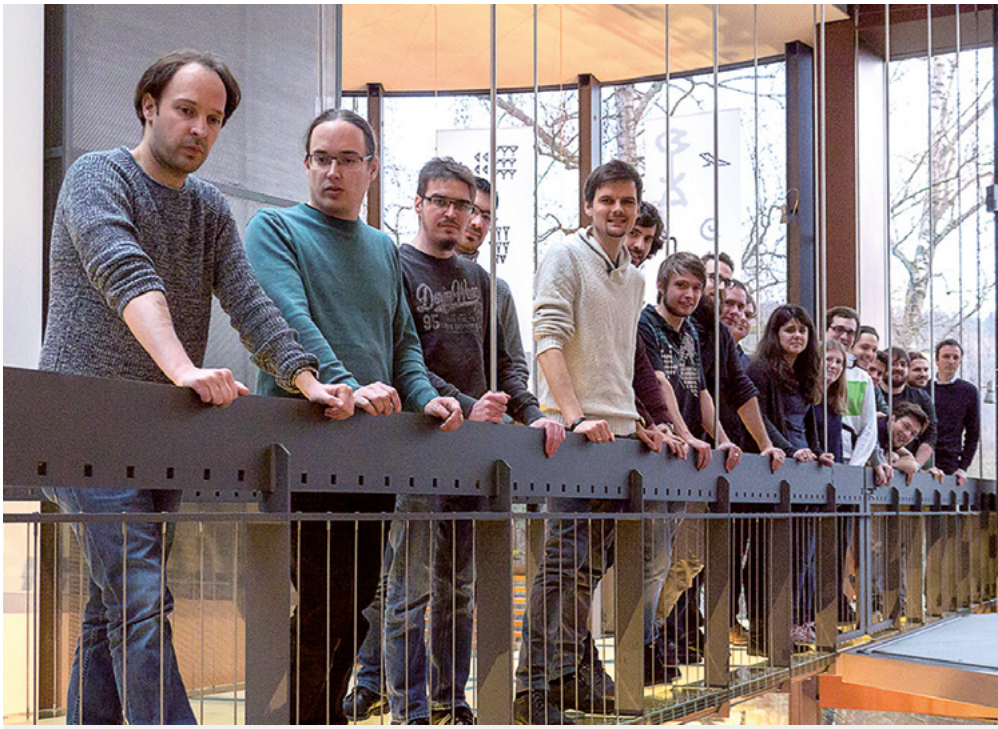


Figure 76: The workshop participants during a break in the Studio Villa Bosch.

5.1.7 Workshop on Supernovae and Stellar Hydrodynamics,

December 14 – 15, 2017, Heidelberg, Germany

On December 14 and 15, 2017, the PSO group organized the 12th Würzburg Workshop on Supernovae and Stellar Hydrodynamics in Heidelberg at HITS and Studio Villa Bosch. More than 20 researchers from international universities and institutions gathered to discuss new developments in the field and to collaborate on new and existing projects.

5.2
HITS Colloquia

Prof. Dr. Bert de Groot
Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
January 16, 2017: The Molecular Dynamics of Channel Permeation and Molecular Recognition

Prof. Dr. Victoria Stodden
School of Information Sciences, University of Illinois at Urbana-Champaign, USA
February 20, 2017: Implementing Reproducibility in Computational Science

Prof. Dr. Konrad Polthier
Mathematical Geometry Processing, Freie Universität Berlin, Germany
April 24, 2017: New Shapes, New Materials and New Processes

T.V. Padma
HITS Journalist in Residence 2017
May 10, 2017: Against the Odds: Indian Science and Science Journalism

Prof. Dr. J. Andrew McCammon
University of California, San Diego, USA
May 15, 2017: Computer-aided Drug Discovery

Prof. Dr. Antonis Rokas
Dept. of Biological Sciences, Vanderbilt University, Nashville, USA
June 19, 2017: Resolving Challenging Branches of the Tree of Life

Prof. Dr. Stefan Wuchty
Dept. of Computer Science, University of Miami, USA
June 30, 2017: Quantifying Group Dynamics in Online Social Media: ISIS and Beyond

Prof. Dr. Niels Grabe
National Center for Tumor Diseases, BIOQUANT Center, University of Heidelberg, Germany
July 17, 2017: Machine Learning Based Image Analysis for Supporting Microscopic Tissue Analysis in Cancer

Dr. Ata Kaban
School of Computer Science, University of Birmingham, UK
September 18, 2017: From Random Projections to Learning Theory to Algorithms and Back

Prof. Dr. Gerhard Jäger
Institute for Linguistics, University of Tübingen, Germany
October 16, 2017: From Words to Features to Trees: Computing a World Tree of Languages from Word Lists

Prof. Dr. Peter V. Coveney
Computational Chemistry Section, University College London, UK
November 20, 2017: The Virtual Human: In Silico Methods for Personalized Medicine


Prof. Dr. Bert de Groot


Prof. Dr. Victoria Stodden


Prof. Dr. Konrad Polthier


T.V. Padma


Prof. Dr. J. Andrew McCammon


Prof. Dr. Antonis Rokas


Prof. Dr. Stefan Wuchty


Prof. Dr. Niels Grabe


Dr. Ata Kaban


Prof. Dr. Gerhard Jäger


Prof. Dr. Peter V. Coveney

Figure 77: "March for Science" in Heidelberg: the closing event at University Square.

On Earth Day (April 22, 2017), one million people all over the world participated in the "March for Science", a non-partisan movement to celebrate science and the role it plays in everyday lives.

According to the organizers, the goal was to emphasize the notion that science upholds the common good and to call for evidence-based policies, which are in the public's best interest. In Heidelberg, 1,800 people attended the "March for Science", during which they walked through the historic center and gathered at University Square for a closing event with speeches and music. HITS Scientific Director Michael Strube was among the speakers. "The freedom of science also comes with responsibility," he said. "The filter bubble, for instance, is not an invention of evil politicians; rather, it is based on concepts from computer science". Strube reminded scientists of the importance of openly informing society not only about their research, but also about opportunities and risks. The event was widely publicized by the press and social media channels and led to a series of related public events.

On May 15, Michael Strube was invited by the city of Heidelberg to give a talk at the German-American Institute (Deutsch-Amerikanisches Institut (DAI)) on the "dark side" of computational linguistics.

Three days later, Michael was a speaker at a panel discussion at the German Cancer Research Center (DKFZ) in the aftermath of the "March for Science." Together with Theresia Bauer (Minister of Science and the Arts, Baden-Württemberg), Michael Baumann (Chairman and Scientific Director, DKFZ), Matthias Henze (Co-Director, EMBL), and Amardeo Sarma (GWUP, society for the scientific investigation of parasciences), he discussed the public image of scientists.

Figure 78: A conversation on data and civil rights (f.l.t.r.): Minister Theresia Bauer, Franziska Brantner MdB, Konstantin von Notz MdB, Michael Strube, Wolfgang Müller.

In July, Michael Strube again met with Theresia Bauer, this time at HITS. Bauer was visiting the institute together Konstantin von Notz and Franziska Brantner, both members of the Green Party faction in the German Bundestag. HITS Managing Director Gesa Schönberger and Deputy Scientific Director Wolfgang Müller attended the meeting, as well. Together, they all had a conversation on big data, data security, and data protection. "Digital politics should be viewed as educational politics and has to eventually lead to individuals' digital 'maturity'", Strube emphasized. "In the 21st century, computational thinking is a fundamental skill, like reading, writing, and arithmetic," he said, quoting Jeannette Wing, former member of the HITS Scientific Advisory Board. Finally, Strube reminded the politicians of their duty to ensure that "civil rights not be harmed by digitalization". ■

Figure 79: Michael Strube delivering his speech at the "March for Science".

5.4
HITS Alumni Meeting

Figure 80: HITS Alumnus Stefan Wuchty (University of Miami) gave a colloquium talk on social network groups.

Ten years after the first alumni meeting took place at Villa Bosch, HITS management invited alumni from HITS, its predecessor (EML Research), and EML for a two-day event during the final weekend of June.

Based on feedback from previous meetings, the format of the event was changed such that the emphasis was placed on interactions between alumni and current HITSters. The meeting was therefore open to alumni and to all current HITSters. The response proved impressive as about 100 people registered for the meeting.

The event commenced on Friday, with a public scientific colloquium in the Studio Villa Bosch given by Stefan Wuchty, professor at the University of Miami, Florida, USA, and EML alumnus (1999 – 2002). During his talk "Quantifying Group Dynamics in Online Social Media: ISIS and Beyond," Stefan spoke about his research on social networks groups. This topic attracted several journalists who had interviewed Stefan before the event.

After a break, HITS Managing Director Gesa Schönberger and Scientific Director Michael Strube welcomed the participants to the event. Their welcome address was followed by a panel discussion on careers in which three alumni talked about their career paths and life after leaving EML / HITS. These discussions were held by Katja Filippova (now with Google Research, Zürich, Switzerland), who had worked in the NLP group; Jon Fuller (now at the KNIME company, Zürich, Switzerland), a former postdoc in the MCM group; and Fernando Izquierdo-Carrasco (now at Oxford Nanopore Technologies, Oxford,

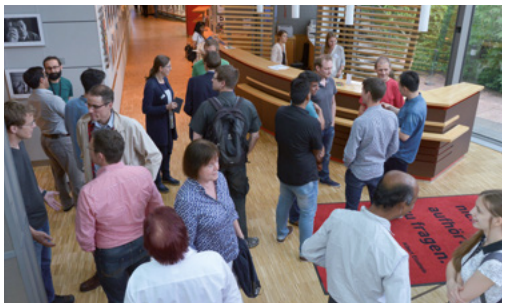


Figure 81: HITS and what next? HITS Alumni Fernando Izquierdo-Carrasco (top left), John Fuller (center left), and Katja Filippova (center) talked about their careers in the panel discussion and shared their experiences with the audience.

United Kingdom), who had been a doctoral student in the CME (former SCO) group. The vivid discussion was moderated by Rebecca Wade (MCM). In the evening, alumni and HITSters met for an informal dinner at the Essighaus restaurant in Heidelberg's historic center.

On Saturday morning, some HITS groups had special meetings with their alumni. The MCM group, for instance, organized a half-day workshop with short talks and discussions (see *Chapter 2.8*). In the afternoon, a "grill & chill" party was held in the HITS cafeteria that lasted until sundown. HITS chef Ralf Westermann and his team treated the participants to delicious food and beverages. Alumni and HITSters came with their families, so there was plenty of room for friendly discussions and laughs.

Figure 82: Gesa Schönberger and Michael Strube during their welcome address to the Alumni and HITSters.

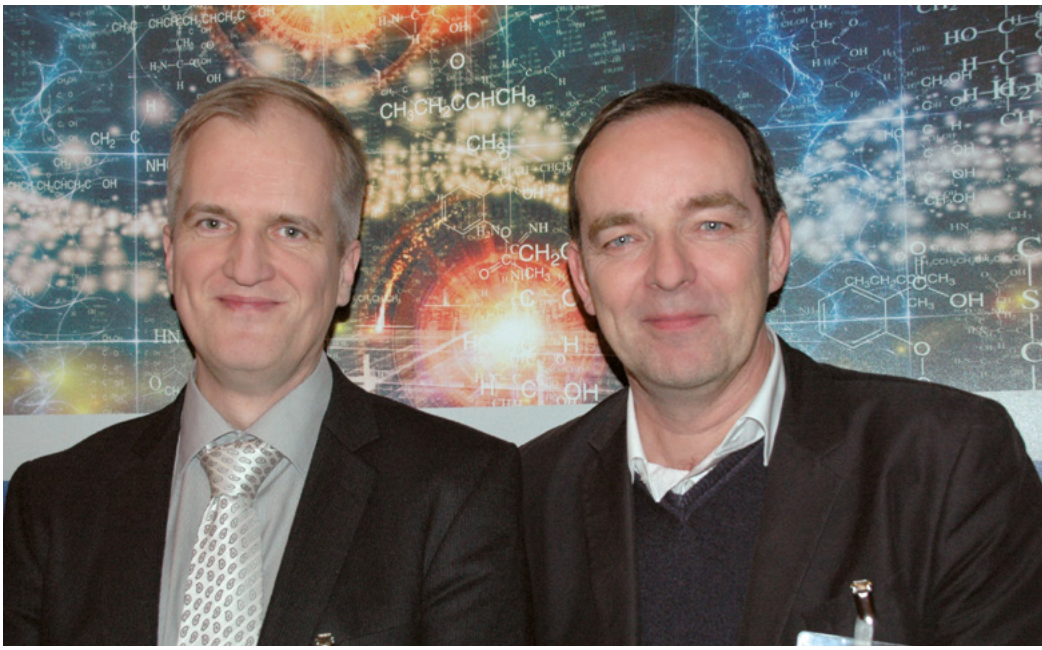


Figure 83: Wolfgang Müller (left) and Volker Stollorz (right).

Operation Explorer

In 2012, science journalist Volker Stollorz was the first "journalist in residence" at HITS. During his stay, he met group leader Priv.-Doz. Dr. Wolfgang Müller and the Scientific Databases and Visualization (SDBV) group, which specializes in the search for life science data and makes storage, search, and processing simple to use for domain experts who are not computer scientists. Together, they discussed Volker's idea of a research tool for science journalists who work on issues involving the situation of medical care in Germany. Over the course of these discussions, "Operation Explorer" was conceived.

At the beginning: A journalist in residence

The SDBV group and Volker Stollorz began a joint project funded by the Robert Bosch Foundation's program "New Pathways in Science Journalism" ("Neue Wege im Wissenschaftsjournalismus"). Together with SDBV members Dr. Meik Bittkowski and Dr. Andreas Weidemann, Wolfgang Müller developed the first version of "Operation Explorer," which enables journalists to quickly mine the sizable datasets of all diagnoses made, all operations performed, and the remedial procedures resorted to in a visual and exploratory way (see more in *Chapter 2.12*).

Figure 84: "Paying for Surgery": Science journalists used the "Operation Explorer" to conduct research on a story that was broadcast on German TV.

Figure 85: The HLF young researchers at HITS.

In the end: An exciting story

The project was introduced at the German science journalists' conference "Wissenswerte" ("Valuable Information") in 2013 and made a big splash. Over the next two years, the SDBV group and Stollorz proceeded with "Operation Explorer" and collaborated with a team at the German television station WDR. The TV journalists used the research tool and produced a movie that was broadcast on German TV in June 2017. The issue at hand was overwhelmingly reflected by the German media, and the expertise of HITS researchers was highlighted. Since the establishment of the German "Science Media Center" (SMC) in Cologne (with Volker Stollorz as Chief Editor), "Operation Explorer" has been established as a research tool for accredited science journalists.

Heidelberg Laureate Forum

Since 2013, the Heidelberg Laureate Forum (HLF) has taken place annually in Heidelberg. Award-winning scientists from mathematics and computer science come together at this networking event to exchange ideas with each other as well as with selected young scientists and students. The HLF is organized by the Heidelberg Laureate Forum Foundation (HLFF). Since 2016, HITS and Heidelberg University have served as the foundation's scientific partners, and both partners contribute their scientific expertise.

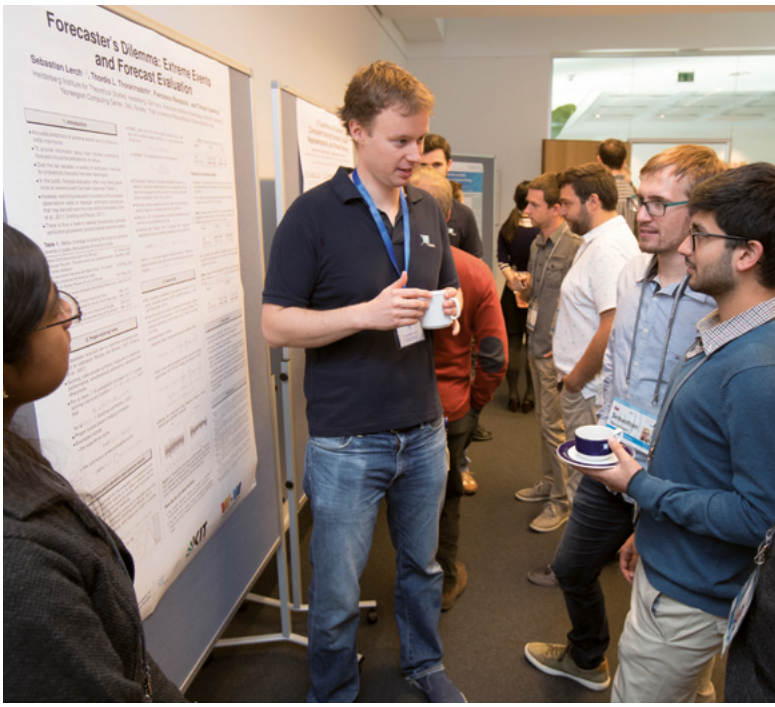


Figure 86: Poster session: Sebastian Lerch (CST) explaining his research topic to the HLF young researchers (photo: Kreutzer/HLFF).

Figure 87: ESO Supernova, artistic impression. Architekten Bernhardt + Partner.

HITS: The forum's scientific partner

The HLF resulted from a joint initiative of the Klaus Tschira Foundation, which supports both the HLF and HITS. Prof. Dr. Andreas Reuter, founding Managing Director and longtime manager of HITS, has been involved since 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 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 was intended by its founder, Klaus Tschira.

Meet the young researchers

As in the previous four years, HITS hosted a group of young researchers from HLF 2017 ranging from undergraduate students to postdoctoral scientists. These researchers enjoyed group presentations and a poster session in which members of several HITS groups presented their current research topics and publications.

ESO Supernova

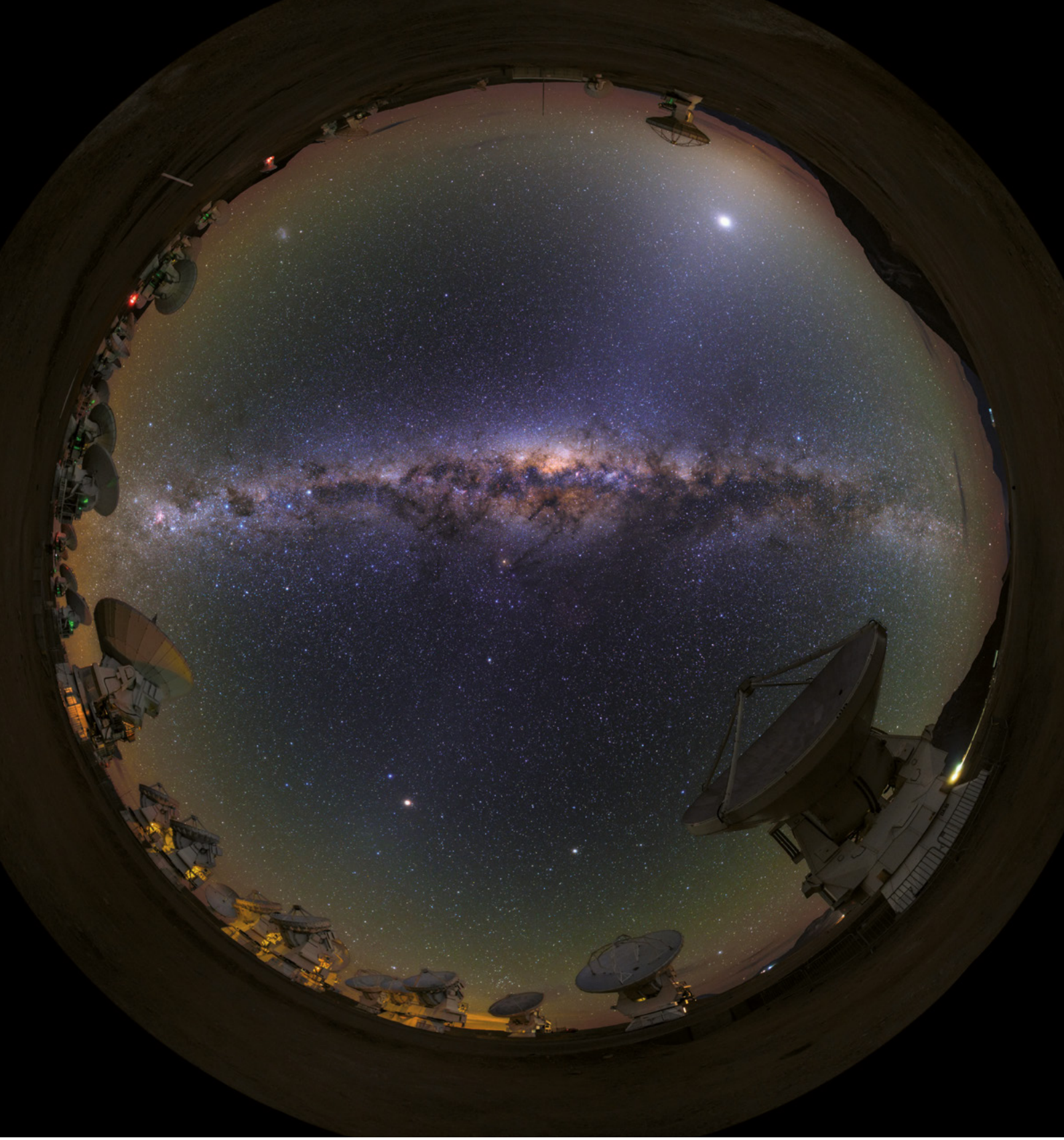
Stars, space, and the Universe hold an almost magical attraction for many people. While astronomy is a vivid and rapidly evolving science, transferring current and new knowledge from researchers to the public is often difficult, and the requirements of daily work limits this transfer to rare occasions 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, thereby allowing visitors to learn about updates to the status of astronomy and recent discoveries as well as to become and remain inspired by the fascinating Universe in which we live.

Grand opening in 2018

The center is the fruit in 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 Foundation will fund the construction of the premises, and ESO will run the facility. The opening ceremony for the ESO Supernova will be on April 26, 2018, at which point the center will open its doors to the public.

Virtual reality and interactive simulations

The ESO 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 in to a broad range of topics by means of interactive computer simulations, virtual reality, and state-of-the-art comput-



er graphics, which will enable them to discover astronomy at an individual and personal level, deepen their understanding of it, 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.

Figure 88: A UHD fish-eye/fulldome view of the Atacama Large Millimeter/submillimeter Array (ALMA). Overhead, the Milky Way shines brightly. Taken during the ESO Ultra HD Expedition. ESO/B. Tafreshi (twanight.org).

7 Publications

Abbott BP et al. (incl. Kromer M). Multi-messenger Observations of a Binary Neutron Star Merger. ApJ Lett. (2017) 848:L12.

Ahmadzadeh F, Lymberakis P, Pirouz RS, Kapli P. The evolutionary history of two lizards (Squamata: Lacertidae) is linked to the geological development of Iran. Zool. Anzeiger – A J. Comp. Zool. (2017) 270:49–56.

Amaral M, Kokh D, Bomke J, Wegener A, Buchstaller H, Eggenweiler H, Matias P, Sirrenberg C, Wade R, Frech M. Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. Nature Communications (2017) 8:2276.

Andrews S, Fryer C, Even WP, Jones S, Pignatari M, NuGrid Collaboration. Uncertainty in Explosive Yields of Core-Collapse Supernovae. In "American Astronomical Society Meeting Abstracts" (2017) vol. 229 of American Astronomical Society Meeting Abstracts, p. 341.21.

Aponte-Santamaría C, Brunken J, Gräter F. Stress propagation through biological lipid bilayers in silico. JACS Commun. (2017) 139:13588–13591.

Aponte-Santamaría C, Fischer G, Båth P, Neutze R, de Groot B. Temperature dependence of protein-water interactions in a gated yeast aquaporin. Sci. Reports (2017) 7:4016.

Aponte-Santamaría C, Lippok S, Mittag J, Obser T, Schneppenheim R, Baldauf C, Gräter F, Buddee U, Rädler J. Mutation g1629e increases von willebrand factor cleavage via a cooperative destabilization mechanism. Biophys. journal (2017) 112:57–65.

Bank S, Sann M, Mayer C, Meusemann K, Donath A, Podsiadlowski L, Kozlov A, Petersen M, Krogmann L, Meier R, et al. Transcriptome and target DNA enrichment sequence data provide new insights into the phylogeny of vespid wasps (Hymenoptera: Aculeata: Vespidae). Mol. Phylogenetics Evol. (2017) 116:213–226.

Barna B, Szalai T, Kromer M, Kerzendorf WE, Vinkó J, Silverman JM, Marion GH, Wheeler JC. Abundance tomography of Type Iax SN 2011ay with tardis. MNRAS (2017) 471:4865–4877.

Barsukow W, Edelmann PVF, Klingenberg C, Miczek F, Röpke FK. A numerical scheme for the compressible low-mach number regime of ideal fluid dynamics. J. Sci. Comput. (2017) 72:623–646.

Barsukow W, Edelmann PVF, Klingenberg C, Röpke FK. A low-mach roe-type solver for the euler equations allowing for gravity source terms. In "Workshop on low velocity flows, Paris, 5 – 6 November 2015", Eds. Dellacherie S, et al (2017) vol. 58 of ESAIM: Proceedings and Surveys, pp. 27 – 39.

Baumann M, Effinger M, Eller D, Heuveline V, Kempf C, Loos L, Maylein L, Peltzer J, Pfeiffer M, Probst V, et al. heimap: Virtueller Kartentisch. EScience-Tage 2017: Forschungsdaten managen (2017).

Baumann M, Gehbart F, Mattes O, Nikas S, Heuveline V. Development and implementation of a temperature monitoring system for hpc systems EMCL Preprint series (2017) 07.

Bauswein A. Dynamics of compact object mergers. In "APS Meeting Abstracts" (2017) p. B15.002.

Bauswein A, Ardevol Pulpillo R, Clark JA, Just O, Goriely S, Janka HT, Stergioulas N. Neutron-star Mergers and Nuclear Physics. Acta Phys. Polonica B (2017) 48:651.

Bauswein A, Just O, Janka HT, Stergioulas N. Neutron-star Radius Constraints from GW170817 and Future Detections. ApJ Lett. (2017) 850:L34.

Bauswein A, Stergioulas N. Semi-analytic derivation of the threshold mass for prompt collapse in binary neutron-star mergers. MNRAS (2017) 471:4956–4965.

Berney C, Ciuprina A, Bender S, Brodie J, Edgcomb V, Kim E, Rajan J, Parfrey LW, Adl S, Audic S, et al. UniEuk: Time to Speak a Common Language in Protistology! J. Eukaryot. Microbiol. (2017) 64:407–411.

Bertsch V, Fichtner W, Heuveline V, Leibfried T. Advances in energy system optimization: Proceedings of the first International Symposium on Energy System Optimization (2017).

Costescu BI, Gräter F, Sturm S. Dynamic disorder can explain non-exponential kinetics of fast protein mechanical unfolding. J. Struct. Biol. (2017) 197:43–49.

Biczok R, Bozsoky P, Eisenmann P, Ernst J, Ribizel T, Scholz F, Trefzer A, Weber F, Hamann M, Stamatakis A. Two C++ Libraries for Counting Trees on a Phylogenetic Terrace. bioRxiv (2017) p. 211276.

Bogdán Á, Bourdin H, Forman WR, Kraft RP, Vogelsberger M, Hernquist L, Springel V. Probing the Hot X-Ray Corona around the Massive Spiral Galaxy, NGC 6753, Using Deep XMM-Newton Observations. Astrophys. J. (2017) 850:98.

Born L, Mesgar M, Strube M. Using a graph-based coherence model in document-level machine translation. In "Proceedings of the 3rd Workshop on Discourse in Machine Translation, Copenhagen, Denmark, 8 September 2017" (2017) pp. 26 – 35.

Briones B, Aponte-Santamaría C, de Groot B. Localization and ordering of lipids around aquaporin-0: Protein and lipid mobility effects. Front. Physiol. (2017) 8:124.

Bruce NJ, Ganotra GK, Kokh DB, Sadiq SK, Wade RC. New approaches for computing ligand-receptor binding kinetics. Curr. Opin. Struct. Biol. (2018) 49:1–10.

Burelle-JP, Trieb N. Schottky groups and maximal representations (2017). To appear in Geometriae Dedicata.

Chan C, Müller B, Heger A, Pakmor R, Springel V. Black Hole Formation and Fallback during the Supernova Explosion of a 40 Msun Star. Astrophys. J. (2018) 852:L19.

Chatziioannou K, Clark JA, Bauswein A, Millhouse M, Littenberg TB, Cornish N. Inferring the post-merger gravitational wave emission from binary neutron star coalescences. Phys. Rev. D (2017) 96:124035.

Cruz-Chú E, Xiao S, Patil S, Gkagkas K, Gräter F. Organic filling mitigates flaw-sensitivity of nanoscale aragonite. ACS Biomater. Sci. & Eng. (2017) 3:260–268.

Daday C, Kolšek K, Gräter F. The mechano-sensing role of the unique sh3 insertion in plakin domains revealed by molecular dynamics simulations. Sci. Reports (2017) 7:11669.

Deason AJ, Belokurov V, Koposov SE, Gómez FA, Grand RJ, Marinacci F, Pakmor R. The slight spin of the old stellar halo. Mon. Notices Royal Astron. Soc. (2017) 470:1259–1273.

Denissenkov PA, Herwig F, Battino U, Ritter C, Pignatari M, Jones S, Paxton B. i-process Nucleosynthesis and Mass Retention Efficiency in He-shell Flash Evolution of Rapidly Accreting White Dwarfs. ApJ Lett. (2017) 834:L10.

Di Pisa F, Landi G, Dello Iacono L, Pozzi C, Borsari C, Ferrari S, Santucci M, Santarem N, Cordeiro-da Silva A, Moraes CB, Alcantara LM, Fontana V, Freitas-Junior LH, Gul S, Kuzikov M, Behrens B, Pöhner I, Wade RC, Costi MP, Mangani S. Chroman-4-one derivatives targeting pteridine reductase 1 and showing anti-parasitic activity. Mol. (2017) 22:426.

Diemer B, Sparre M, Abramson LE, Torrey P. Log-normal Star Formation Histories in Simulated and Observed Galaxies. Astrophys. J. (2017) 839:26.

D'Isanto A, Polsterer KL. Photometric redshift estimation via deep learning. Generalized and pre-classification-less, image based, fully probabilistic redshifts. Astron. Astrophys. (2017) 609:A111.

D'Isanto A, Polsterer KL. Uncertain photometric redshifts via combining deep convolutional and mixture density networks. In "25th European Symposium on Artificial Neural Networks", ESANN 2015, Bruges, Belgium, April 26 – 28, 2017" (2017).

Edelmann PVF, Röpke FK, Hirschi R, Georgy C, Jones S. Testing a onedimensional prescription of dynamical shear mixing with a two-dimensional hydrodynamic simulation. A&A (2017) 604:A25.

Ehm W, Ovcharov E. Bias-corrected score decomposition for generalized quantiles. Biometrika (2017) 104:473–480.

7
Publications

Farnham A, Kurz C, Öztürk MA, Solbiati M, Myllyntaus O, Meekes J, Pham TM, Paz C, Langiewicz M, Andrews S, Kanninen L, Agbemabiese C, Guler AT, Durieux J, Jasim S, Viessmann O, Frattini S, Yembergenova D, Benito CM, Porte M, Grangeray-Vilmint A, Curiel RP, Rehncrona C, Malas T, Esposito F, Hettne K. Early career researchers want Open Science. Genome Biol. (2017) 18:221.

Flouri T, Zhang J, Czech L, Kobert K, Stamatakis A. An Efficient Approach to Merging Paired-End Reads and Incorporation of Uncertainties. Chapter in "Algorithms for Next- Generation Sequencing Data", pp. 299 – 325, Springer (2017).

Fuertes G, Banterle N, Ruff K, Chowdhury A, Mercadante D, Koehler C, Kachala M, Girona GE, Milles S, Mishra A, Onck PR, Gräter F, Esteban-Martín S, Pappu RV, Svergun DI, Lemke EA. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in saxs vs. fret measurements. Proc. Natl. Acad. Sci. United States Am. (2017) 114:E6342–E6351.

Gatto A, Walch S, Naab T, Girichidis P, Wünsch R, Glover SCO, Klessen RS, Clark PC, Peters T, Derigs D, Baczynski C, Puls J. The SILCC project – III. Regulation of star formation and outflows by stellar winds and supernovae. Mon. Notices Royal Astron. Soc. (2017) 466:1903–1924.

Gawlok S, Gerstner P, Haupt S, Heuveline V, Kratzke J, Lösel P, Mang K, Schmidtobreick M, Schoch N, Schween N, Schwegler J, Song C, Wlotzka M. Hiflow3 – technical report on release 2.0. Prepr. Ser. Eng. Math. Comput. Lab (2017).

Ghosh S, Trieb N. Affine Anosov representations and proper actions (2017). Preprint, arXiv:1711.09712.

Gianniotis N. Linear dimensionality reduction for time series. In "Neural Information Processing", Eds. Liu D, Xie S, Li Y, Zhao D, El-Alfy ESM. Cham (2017) pp. 375 – 383.

Gieseke F, Polsterer KL, Mahabal A, Igel C, Heskes T. Massively-parallel best subset selection for ordinary least-squares regression. In "2017 IEEE Symposium Series on Computational Intelligence (SSCI)" (2017) pp. 1 – 8.

Gneiting T. When is the mode functional the Bayes classifier? Stat (2017) 6:204–206.

Goicovic FG, Sesana A, Cuadra J, Stasyszyn F. Infalling clouds on to supermassive black hole binaries – II. Binary evolution and the final parsec problem. Mon. Notices Royal Astron. Soc. (2017) 472:514–531.

Gómez FA, Grand RJJ, Monachesi A, White SDM, Bustamante S, Marinacci F, Pakmor R, Simpson CM, Springel V, Frenk CS. Lessons from the Auriga discs: the hunt for the Milky Way's ex situ disc is not yet over. Mon. Notices Royal Astron. Soc. (2017) 472:3722–3733.

Gómez FA, White SDM, Grand RJJ, Marinacci F, Springel V, Pakmor R. Warps and waves in the stellar discs of the Auriga cosmological simulations. Mon. Notices Royal Astron. Soc. (2017) 465:3446–3460.

Grand RJJ, Gómez FA, Marinacci F, Pakmor R, Springel V, Campbell DJR, Frenk CS, Jenkins A, White SDM. The Auriga Project: the properties and formation mechanisms of disc galaxies across cosmic time. Mon. Notices Royal Astron. Soc. (2017) 467:179–207.

Grand RJJ, Bustamante S, Gómez FA, Kawata D, Marinacci F, Pakmor R, Rix HW, Simpson CM, Sparre M, Springel V. Origin of chemically distinct discs in the Auriga cosmological simulations. Mon. Notices Royal Astron. Soc. (2018) 474:3629–3639.

Grímsson F, Kapli P, Hofmann CC, Zetter R, Grimm G. Eocene Loranthaceae pollen pushes back divergence ages for major splits in the family. PeerJ (2017) 5:e3373.

Guilet J, Bauswein A, Just O, Janka HT. Magnetorotational instability in neutron star mergers: impact of neutrinos. MNRAS (2017) 471:1879–1887.

Hachinger S, Röpke FK, Mazzali PA, Gal-Yam A, Maguire K, Sullivan M, Taubenberger S, Ashall C, Campbell H, Elias-Rosa N, Feindt U, Greggio L, Inserra C, Miluzio M, Smartt SJ, Young D. Type Ia supernovae with and without blueshifted narrow Na I D lines – how different is their structure? MNRAS (2017) 471:491–506.

Haettel T, Schilling A, Wienhard A. Horofunction compactifications of symmetric spaces (2017). Preprint, arXiv:1705.05026.

Heinzerling B, Moosavi NS, Strube M. Revisiting selectional preferences for coreference resolution. In "Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7 – 11 September 2017" (2017) pp. 1343 – 1350.

Heinzerling B, Strube M, Lin CY. Trust, but verify! Better entity linking through automatic verification. In "Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, 3 – 7 April 2017" (2017) pp. 828–838.

Hemri S, Klein B. Analog-based postprocessing of navigation-related hydrological ensemble forecasts. Water Resour. Res. (2017) 53:9059–9077.

Heuveline V, Schick M, Webster C, Zaspel P. Uncertainty quantification and high performance computing (Dagstuhl Seminar 16372). In "Dagstuhl Reports". Schloss Dagstuhl-Leibniz-Zentrum für Informatik (2017).

Hilbert S, Xu D, Schneider P, Springel V, Vogelsberger M, Hernquist L. Intrinsic alignments of galaxies in the Illustris simulation. Mon. Notices Royal Astron. Soc. (2017) 468:790–823.

Hovy D, Spruit S, Mitchell M, Bender EM, Strube M, Wallach H (Eds.). "Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, Valencia, Spain, 4 April 2017" (2017).

Hull CLH, Mocz P, Burkhart B, Goodman AA, Girart JM, Cortés PC, Hernquist L, Springel V, Li ZY, Lai SP. Unveiling the Role of the Magnetic Field at the Smallest Scales of Star Formation. Astrophys. J. (2017) 842:L9.

Hunt JAS, Kawata D, Monari G, Grand RJJ, Famaey B, Siebert A. Stars with fast Galactic rotation observed in Gaia TGAS: a signature driven by the Perseus arm? Mon. Notices Royal Astron. Soc. (2017) 467:L21–L25.

Jacob S, Pfrommer C. Cosmic ray heating in cool core clusters – I. Diversity of steady state solutions. Mon. Notices Royal Astron. Soc. (2017) 467:1449–1477.

Jacob S, Pfrommer C. Cosmic ray heating in cool core clusters – II. Selfregulation cycle and non-thermal emission. Mon. Notices Royal Astron. Soc. (2017) 467:1478–1495.

Jacob S, Pakmor R, Simpson CM, Springel V, Pfrommer C. The dependence of cosmic ray-driven galactic winds on halo mass. Mon. Notices Royal Astron. Soc. (2018) 475:570–584.

Jedwabny W, Panecka-Hofman J, Dyguda-Kazimierowicz E, Wade RC, Sokalski WA. Application of a simple quantum chemical approach to ligand fragment scoring for trypanosoma brucei pteridine reductase 1 inhibition. J. Comput. Aided Mol. Des. (2017) 31:715–728.

Johansson J, Goobar A, Kasliwal MM, Helou G, Masci F, Tinyanont S, Jencson J, Cao Y, Fox OD, Kromer M, Amanullah R, Banerjee DPK, Joshi V, Jerkstrand A, Kankare E, Prince TA. Spitzer observations of SN 2014J and properties of mid-IR emission in Type Ia supernovae. MNRAS (2017) 466:3442– 3449.

Jones S, Andrassy R, Sandalski S, Davis A, Woodward P, Herwig F. Idealized hydrodynamic simulations of turbulent oxygen-burning shell convection in 4_ geometry. MNRAS (2017) 465:2991–3010.

Judea A, Strube M. Event argument identification on dependency graphs with bidirectional LSTMs. In "Proceedings of the 8th International Joint Conference on Natural Language Processing, Taipei, Taiwan, 27 November – 1 December 2017" (2017) pp. 822 – 831.

7 Publications

Just O, Goriely S, Bauswein A, Obergaulinger M, Ardevol Pulpillo R, Janka HT. Impact of Neutrino Interactions on Outflows of Neutron-Star Mergers. In "14th International Symposium on Nuclei in the Cosmos (NIC2016)", Eds. Kubono S, Kajino T, Nishimura S, Isobe T, Nagataki S, Shima T, Takeda Y (2017) p. 010704.

Kannan R, Vogelsberger M, Pfrommer C, Weinberger R, Springel V, Hernquist L, Puchwein E, Pakmor R. Increasing Black Hole Feedback-induced Quenching with Anisotropic Thermal Conduction. Astrophys. J. Lett. (2017) 837:L18.

Kawata D, Grand RJJ, Gibson BK, Casagrande L, Hunt JAS, Brook CB. Impacts of a flaring star-forming disc and stellar radial mixing on the vertical metallicity gradient. Mon. Notices Royal Astron. Soc. (2017) 464:702–712.

Klatt S, Kraus D, Kraft P, Breuer L, Wlotzka M, Heuveline V, Haas E, Kiese R, Butterbach-Bahl K. Exploring impacts of vegetated buffer strips on nitrogen cycling using a spatially explicit hydro-biogeochemical modeling approach. Environ. Model. & Softw. (2017) 90:55–67.

Knowles K, Baker A, Basu K, Bharadwaj V, Deane R, Devlin M, Dicker S, de Gasperin F, Ferrari C, Hilton M, Hughes JP, Intema HT, Makhathini S, Moodley K, Oozeer N, Pfrommer C, Sievers J, Sikhosana SP, Smirnov O, Sommer MW, Stanchfield S, van der Heyden K, Zwart JTL. MERGHERS: An SZ-selected cluster survey with MeerKAT. ArXiv e-prints (2017).

Kolšek K, Aponte-Santamaría C, Gräter F. Accessibility explains preferred thiol-disulfide isomerization in a protein domain. Sci. Reports (2017) 7:9858.

Kromer M, Ohlmann S, Röpke FK. Simulating the observed diversity of Type Ia supernovae . Introducing a model data base. Mem. Soc. Astron. Italiana (2017) 88:312.

Kurohashi S, Strube M (Eds.). "Proceedings of the IJCNLP 2017, Tutorial Abstracts, Taipei, Taiwan, 27 November 2017" (2017).

Lerch S, Baran S. Similarity-based semilocal estimation of post-processing models. J. Royal Stat. Soc. Ser. C: Appl. Stat. (2017) 66:29–51.

Lerch S, Thorarinsdottir TL, Ravazzolo F, Gneiting T. Forecaster's dilemma: Extreme events and forecast evaluation. Stat. Sci. (2017) 32:106–127.

Li H, Mao S, Emsellem E, Xu D, Springel V, Krajnovic D. The origin and properties of massive prolate galaxies in the Illustris simulation. Mon. Notices Royal Astron. Soc. (2018) 473:1489–1511.

Linciano P, Dawson A, Pöhner I, Costa DM, Sá MS, Cordeiro-da Silva A, Luciani R, Gul S, Witt G, Ellinger B, Kuzikov M, Gribbon P, Reinshagen J, Wolf M, Behrens B, Hannaert V, Michels PAM, Nerini E, Pozzi C, di Pisa F, Landi G, Santarem N, Ferrari S, Saxena P, Lazzari S, Cannazza G, Freitas- Junior LH, Moraes CB, Pascoalino BS, Alcântara LM, Bertolacini CP, Fontana V, Wittig U, Müller W, Wade RC, Hunter WN, Mangani S, Costantino L, Costi MP. Exploiting the 2-amino-1,3,4-thiadiazole scaffold to inhibit trypanosoma brucei pteridine reductase in support of early-stage drug discovery. ACS Omega (2017) 2:5666–5683.

Loderer T, Heuveline V. New sparsing approach for real-time simulations of stiff models on electronic control units. Prepr. Ser. Eng. Math. Comput. Lab (2017).

Lösel P, Heuveline V. A GPU-based diffusion method for whole-heart and great vessel segmentation. In "Reconstruction, Segmentation, and Analysis of Medical Images, RAMBO 2016 and HVSMR 2016, Athens, Greece, October 17, 2016", Eds. Zuluaga MA, Bhatia K, Kainz B, Moghari MH, Pace DF. Lecture Notes in Computer Science Springer International Publishing (2017) pp. 121 – 128.

Maier-Hein L, Vedula S, Speidel S, Navab N, Kikinis R, Park A, Eisenmann M, Feussner H, Forestier G, Giannarou S, Hashizume M, Katic D, Kenngott H, Kranzfelder M, Malpani A, März K, Neumuth T, Padoy N, Pugh C, Schoch N, Stoyanov D, Wagner M, Taylor RH, Hager GD, Jannin P. Surgical data science: Enabling next-generation surgery. Nat. Biomed. Eng. (2017).

Marinacci F, Grand RJJ, Pakmor R, Springel V, Gómez FA, Frenk CS, White SDM. Properties of H I discs in the Auriga cosmological simulations. Mon. Notices Royal Astron. Soc. (2017) 466:3859–3875.

Mercadante D, Wagner J, Aramburu I, Lemke E, Gräter F. Ampling long versus short-range interactions defines the ability of force fields to reproduce the dynamics of intrinsically disordered proteins. J. Chem. Theory Comput. (2017) 13:3964–3974.

Mexner W, Bonn M, Kopmann A, Mauch V, Ressmann D, Chilingaryan SA, Jerome NT, Kamp TVD, Heuveline V, Lösel P, et al. Open GL API-based analysis of large datasets in a cloud environment. Adv. Comput. Electr. Eng. Des. Use Virtualization Technol. Cloud Comput. (2017) p. 161 – 181.

Meyer-Hübner N, Suriyah M, Leibfried T, Slednev V, Bertsch V, Fichtner W, Gerstner P, Schick M, Heuveline V. Optimal storage operation with model predictive control in the German transmission grid. In "Advances in Energy System Optimization". Birkhäuser, Cham (2017) pp. 31 – 45.

Miletic V, Odorcic I, Nikolic P, Svedružic Ž. In silico design of the first DNA-independent mechanism-based inhibitor of mammalian dna methyltransferase dnmt1. PLoS ONE (2017) 12:e0174410.

Mocz P, Burkhart B, Hernquist L, McKee CF, Springel V. Moving-mesh Simulations of Star-forming Cores in Magneto-gravo-turbulence. Astrophys. J. (2017) 838:40.

Moosavi NS, Strube M. Lexical features in coreference resolution: To be used with caution. In "Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Vancouver, B.C., Canada, 30 July – 4 August 2017" (2017).

Moosavi NS, Strube M. Use generalized representations, but do not forget surface features. In "Proceedings of the 2nd Workshop on Coreference Resolution Beyond OntoNotes, Valencia, Spain, 4 April 2017" (2017) pp. 1 – 7.

Moreau CA, Bhargav SP, Kumar H, Quadt KA, Piirainen H, Strauss L, Kehrer J, Streichfuss M, Spatz JP, Wade RC, Kursula I, Frischknecht F. A unique profilin-actin interface is important for malaria parasite motility. PLOS Pathog. (2017) 13:e1006412.

Morel B, Flouri T, Stamatakis A. A novel heuristic for data distribution in massively parallel phylogenetic inference using site repeats. In "Proceedings of 19th Conference on High Performance Computing and Communications (HPCC), Bangkok, Thailand". IEEE (2017) To be published.

Müller MC. Semantic author name disambiguation with word embeddings. In "Research and Advanced Technology for Digital Libraries: 21st International Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, 18 – 21 September 2017", Eds. Kamps J, Tsakonas G, Manolopoulos Y, Iliadis L, Karydis I. Springer International Publishing, Cham, Germany (2017) pp. 300 – 311.

Müller MC, Reitz F, Roy N. Data sets for author name disambiguation: An empirical analysis and a new resource. Sci. (2017) 111:1467–1500.

Müller W, Bittkowski M, Golebiewski M, Kania R, Rey M, Weidemann A, Wittig U. SABIO-RK, von Daten in der Publikation zur Suchlösung für Spezialisten. Datenbank Spektrum (2017) 17, pp. 21 – 28.

Myers CJ , Bader G, Gleeson P, Golebiewski M, Hucka M, Le Novère N, Nickerson DP, Schreiber F, and Waltemath D. A brief history of COMBINE. 2017 Winter Simulation Conference (WSC), Las Vegas, NV, USA (2017), pp. 884 – 895.

Nillegoda NB, Stank A, Malinverni D, Alberts N, Szlachcic A, Barducci A, De Los Rios P, Wade RC, Bukau B. Evolution of an intricate j-protein network driving protein disaggregation in eukaryotes. eLife (2017) 6:e24560.

Noebauer UM, Kromer M, Taubenberger S, Baklanov P, Blinnikov S, Sorokina E, Hillebrandt W. Early light curves for Type Ia supernova explosion models. MNRAS (2017) 472:2787–2799.

Ohlmann ST, Röpke FK, Pakmor R, Springel V. Constructing stable 3D hydrodynamical models of giant stars. Astron. Astrophys. (2017) 599:A5.

7 Publications

Öztürk MA, Cojocaru V, Wade RC. Drosophila melanogaster linker histone (dh1) binding to the nucleosome. In "Proceedings of the 3rd bwHPCSymposium: Heidelberg 2016". Heidelberg University Library, heiBOOKS, Heidelberg (2017) pp. 100 – 102.

Pakmor R, Gómez FA, Grand RJJ, Marinacci F, Simpson CM, Springel V, Campbell DJR, Frenk CS, Guillet T, Pfrommer C, White SDM. Magnetic field formation in the Milky Way like disc galaxies of the Auriga project. Mon. Notices Royal Astron. Soc. (2017) 469:3185–3199.

Panecka-Hofman J, Pöhner I, Spyrakis F, Zeppelin T, Di Pisa F, Dello Iacono L, Bonucci A, Quotadamo A, Venturelli A, Mangani S, Costi MP, Wade RC. Comparative mapping of on-targets and off-targets for the discovery of antitrypanosomatid folate pathway inhibitors. Biochimica et Biophys. Acta – Gen. Subj. (2017) 1861:3215–3230.

Pardi A, Girichidis P, Naab T, Walch S, Peters T, Heitsch F, Glover SCO, Klessen RS, Wünsch R, Gatto A. The impact of magnetic fields on the chemical evolution of the supernova-driven ISM. Mon. Notices Royal Astron. Soc. (2017) 465:4611–4633.

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, et al. Evolutionary history of the Hymenoptera. Curr. Biol. (2017) 27:1013–1018.

Peters T, Naab T, Walch S, Glover SCO, Girichidis P, Pellegrini E, Klessen RS, Wünsch R, Gatto A, Baczynski C. The SILCC project – IV. Impact of dissociating and ionizing radiation on the interstellar medium and H_ emission as a tracer of the star formation rate. Mon. Notices Royal Astron. Soc. (2017) 466:3293–3308.

Peters T, Zhukovska S, Naab T, Girichidis P, Walch S, Glover SCO, Klessen RS, Clark PC, Seifried D. The turbulent life of dust grains in the supernova-driven, multiphase interstellar medium. Mon. Notices Royal Astron. Soc. (2017) 467:4322–4342.

Petrushevska T, Amanullah R, Bulla M, Kromer M, Ferretti R, Goobar A, Papadogiannakis S. Testing for redshift evolution of Type Ia supernovae using the strongly lensed PS1-10afx at z = 1.4. A&A (2017) 603:A136.

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.

Pfrommer C, Pakmor R, Simpson CM, Springel V. Simulating Gamma-Ray Emission in Star-forming Galaxies. Astrophys. J. Lett. (2017) 847:L13.

Pineda JCB, Hayward CC, Springel V, Mendes de Oliveira C. Rotation curve fitting and its fatal attraction to cores in realistically simulated galaxy observations. Mon. Notices Royal Astron. Soc. (2017) 466:63–87.

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.

Polsterer KL. Astroinformatics; a new discipline or business as usual? In "Astronomical Data Analysis Software and Systems XXVII (ADASS XXVII)", Eds. Taylor AR, Rosolowsky E (2017) vol. 496 of Astronomical Society of the Pacific Conference Series.

Polsterer KL. Dealing with Uncertain Multimodal Photometric Redshift Estimations. In "Astroinformatics", Eds. Brescia M, Djorgovski SG, Feigelson ED, Longo G, Cavuoti S (2017) vol. 325 of IAU Symposium, pp. 156 – 165.

Posch S, Aponte-Santamaría C, Schwarz R, Karner A, Radtke M, Gräter F, Obser T, Brehm GK, Gruber H, Netz R, Baldauf C, Schneppenheim R, Tampé R, Hinterdorfer P. Mutual A domain interactions in the force sensing protein von Willebrand factor. J. Struct. Biol. (2017) 197:57–64.

Rafiullah R, Long AB, Ivanova AA, Ali H, Berkel S, Mustafa G, Paramasivam N, Schlesner M, Wiemann S, Wade RC, Bolthauser E, Blum M, Kahn RA, Caspary T, Rappold GA. A novel homozygous ARL13B variant in patients with Joubert syndrome impairs its guanine nucleotide-exchange factor activity. Eur. J. Hum. Genet. (2017) 25:1324–1334.

Ratmann O, Wymant C, Colijn C, Danaviah S, Essex M, Frost SD, Gall A, Saiseitsiwe S, Grabowski M, Gray R, et al. HIV-1 full-genome phylogenetics of generalized epidemics in sub-Saharan Africa: impact of missing nucleotidecharacters in next-generation sequences. AIDS Res. Hum. Retroviruses (2017) 33:1083–1098.

Rodriguez-Gomez V, Sales LV, Genel S, Pillepich A, Zjupa J, Nelson D, Griffen B, Torrey P, Snyder GF, Vogelsberger M, Springel V, Ma CP, Hernquist L. The role of mergers and halo spin in shaping galaxy morphology. Mon. Notices Royal Astron. Soc. (2017) 467:3083–3098.

Röpke FK. Combustion in thermonuclear supernova explosions. In "Handbook of Supernovae", Eds. Alsabti A, Murdin P. Springer (2017) pp. 1185–1209.

Sasdelli M, Hillebrandt W, Kromer M, Ishida EEO, Röpke FK, Sim SA, Pakmor R, Seitenzahl IR, Fink M. A metric space for Type Ia supernova spectra: a new method to assess explosion scenarios. Mon. Notices Royal Astron. Soc. (2017) 466:3784–3809.

Schmelzle S, Heethoff M, Heuveline V, Lösel P, Becker J, Beckmann F, Schluenzen F, Hammel JU, Kopmann A, Mexner W, Vogelgesang M, Jerome NT, Betz O, Beutel R, Wipfler B, Blanke A, Harzsch S, Hörnig M, Baumbach T, van de Kamp T. The nova project: maximizing beam time efficiency through synergistic analyses of SRmCT data. In Proc. SPIE 10391, Developments in X-Ray Tomography XI, 103910P, doi:1 0.1117/12.2275959.5581167496001, 2017. (2017) vol. 17.

Schmidt P. Discussion of "Elicitability and backtesting: Perspectives for banking regulation". Annals Appl. Stat. (2017) 11:1883–1885.

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) (2017).

Schoch N, Heuveline V. Towards an intelligent framework for personalized simulation-enhanced surgery assistance: Linking a simulation ontology to a reinforcement learning algorithm for calibration of numerical simulations. Prepr. Ser. Eng. Math. Comput. Lab (2017).

Schoch N, Heuveline V. Towards cognition-guided patient-specific FEM-based cardiac surgery simulation. Lect. Notes Comput. Sci. Proc. "Functional Imaging Model". Hear. (FIMH) 2017 (2017) .

Schreiber F, Bader GD, Golebiewski M, Hucka M, Kormeier B, Le Novère N, Myers CJ, Nickerson DP, Sommer B, Waltemath D, Weise S. Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2016. Journal of Integrative Bioinformatics (2017) 13(3), pp. 1 – 7.

Schuetz DA, de Witte WEA, Wong YC, Knasmueller B, Richter L, Kokh DB, Sadiq SK, Bosma R, Nederpelt I, Heitman LH, Segala E, Amaral M, Guo D, Andres D, Georgi V, Stoddart LA, Hill S, Cooke RM, Graaf CD, Leurs R, Frech M, Wade RC, de Lange ECM, IJzerman AP, Müller-Fahrnow A, Ecker GF. Kinetics for drug discovery: an industry-driven effort to target drug residence time. Drug Discov. Today (2017) 22:896–911.

Seifried D, Walch S, Girichidis P, Naab T, Wünsch R, Klessen RS, Glover SCO, Peters T, Clark P. SILCC-Zoom: the dynamic and chemical evolution of molecular clouds. Mon. Notices Royal Astron. Soc. (2017) 472:4797–4818.

Shalaby M, Broderick AE, Chang P, Pfrommer C, Lamberts A, Puchwein E. Importance of Resolving the Spectral Support of Beam-plasma Instabilities in Simulations. Astrophys. J. (2017) 848:81.

Shalaby M, Broderick AE, Chang P, Pfrommer C, Lamberts A, Puchwein E. SHARP: A Spatially Higher-order, Relativistic Particle-in-cell Code. Astrophys. J. (2017) 841:52.

7
Publications

Slednev V, Ruppert M, Bertsch V, Fichtner W, Meyer-Hübner N, Suriyah M, Leibfried T, Gerstner P, Schick M, Heuveline V. Regionalizing input data for generation and transmission expansion planning models. In "Advances in Energy System Optimization". Birkhäuser, Cham (2017) pp. 205 – 217.

Smartt SJ, et al. (incl. Kromer M). A kilonova as the electromagnetic counterpart to a gravitational-wave source. Nature (2017) 551:75–79.

Song C, Heuveline V. Multilevel preconditioner of polynomial chaos method for quantifying uncertainties in a blood pump. In "International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP), Greece". Scopus, Elsevier (2017).

Song C, Stoll M, Giske K, Bendl R, Heuveline V. Sparse grids for quantifying motion uncertainties in biomechanical models of radiotherapy patients. Prepr. Ser. Eng. Math. Comput. Lab (2017).

Sparre M, Hayward CC, Feldmann R, Faucher-Giguère CA, Muratov AL, Kereš D, Hopkins PF. (Star)bursts of FIRE: observational signatures of bursty star formation in galaxies. Mon. Notices Royal Astron. Soc. (2017) 466:88–104.

Sparre M, Springel V. The unorthodox evolution of major merger remnants into star-forming spiral galaxies. Mon. Notices Royal Astron. Soc. (2017) 470:3946– 3958.

Stank A, Kokh DB, Horn M, Sizikova E, Neil R, Panecka J, Richter S, Wade RC. Trapp webserver: predicting protein binding site flexibility and detecting transient binding pockets. Nucleic Acids Res. (2017) 45:W325–W330.

Sun Z, Wienhard A, Zhang T. Flows on the psl(v)-hitchin component (2017). Preprint, arXiv:1709.03580.

Šmíd J, Moravec J, Gvoždík V, Štundl J, Frynta D, Lymberakis P, Kapli P, Wilms T, Schmitz A, Shobrak M, Yousefkhani S, Rastegar-Pouyani E, Castilla A, Els J, Mayer W. Cutting the Gordian Knot: Phylogenetic and ecological diversification of the Mesalina brevirostris species complex (Squamata, Lacertidae). Zool. Scripta (2017) 46:649–664.

Terada Y, Maeda K, Fukazawa Y, Bamba A, Ueda Y, Katsuda S, Enoto T, Takahashi T, Tamagawa T, Röpke FK, Summa A, Diehl R. Soft gamma-ray observation of SN2014J with suzaku. In "14th International Symposium on Nuclei in the Cosmos (NIC2016)", Eds. Kubono S, Kajino T, Nishimura S, Isobe T, Nagataki S, Shima T, Takeda Y (2017) id. 010306.

Tiede P, Broderick AE, Shalaby M, Pfrommer C, Puchwein E, Chang P, Lamberts A. Bow Ties in the Sky. II. Searching for Gamma-Ray Halos in the Fermi Sky Using Anisotropy. Astrophys. J. (2017) 850:157.

Van de Voort F, Bahé YM, Bower RG, Correa CA, Crain RA, Schaye J, Theuns T. The environmental dependence of gas accretion on to galaxies: quenching satellites through starvation. Mon. Notices Royal Astron. Soc. (2017) 466:3460– 3471.

Wagner J. Generalised model-independent characterisation of strong gravitational lenses. I. Theoretical foundations. Astron. Astrophys. (2017) 601:A131.

Wagner J, Patil S, Greving I, Lämmel M, Gkagkas K, Seydel T, Müller M, Markert B, Gräter F. Stress-induced long-range ordering in spider silk. Sci. Reports (2017) 7:15273.

Weinberger R, Ehlert K, Pfrommer C, Pakmor R, Springel V. Simulating the interaction of jets with the intracluster medium. Mon. Notices Royal Astron. Soc. (2017) 470:4530–4546.

Weinberger R, Springel V, Hernquist L, Pillepich A, Marinacci F, Pakmor R, Nelson D, Genel S, Vogelsberger M, Naiman J, Torrey P. Simulating galaxy formation with black hole driven thermal and kinetic feedback. Mon. Notices Royal Astron. Soc. (2017) 465:3291–3308.

Wiener J, Pfrommer C, Oh SP. Cosmic ray-driven galactic winds: streaming or diffusion? Mon. Notices Royal Astron. Soc. (2017) 467:906–921.

Wienhard A, Zhang T. Deforming convex real projective structures. Geom. Dedicata (2018) 192:327–360.

Wittig U, Rey M, Weidemann A, Müller W. Data management and data enrichment for systems biology projects. J Biotechnol (2017) 261: 229-237.

Wittig U, Rey M, Weidemann A, Kania R, Müller W. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. Nucleic Acids Research (2018) 46(D1): D656-D660.

Wlotzka M, Heuveline V. Energy-efficient multigrid smoothers and grid transfer operators on multi-core and GPU clusters. J. Parallel Distributed Comput. (2017) 100:181–192.

Wlotzka M, Heuveline V. An energy-efficient parallel multigrid methods for multi-core CPU platforms and HPC clusters. EMCL Prepr. Ser. (2017) 03.

Wlotzka M, Heuveline V, Dolz MF, Heidari MR, Ludwig T, Malossi ACI, Quintana-Orti ES. Energy-aware high performance computing. In "ICTEnergy Concepts for Energy Efficiency and Sustainability". InTech (2017)

Wlotzka M, Heuveline V, Klatt S, Kraus D, Haas E, Kiese R, Butterbach-Bahl K, Kraft P, Breuer L. Parallel multiphysics simulations using openpalm with application to hydrology biogeochemistry coupling. In "Modeling, Simulation and Optimization of Complex Processes – HPSC 2015", Eds. Bock H, Phu H, Rannacher R, Schlöder J. Springer (2017) .

Wlotzka M, Morel T, Piacentini A, Heuveline V. New features for advanced dynamic parallel communication routines in openpalm: Algorithms and documentation. EMCL Prepr. Ser. (2017) 04.

Wolstencroft K, Krebs O, Snoep JL, Stanford NJ, Bacall F, Golebiewski M, Kuzyakiv R, Nguyen Q, Owen S, Soiland-Reyes S, Straszewski J, van Niekerk DD, Williams AR, Malmström L, Rinn B, Müller W, Goble C. FAIRDOM-Hub: a repository and collaboration environment for sharing systems biology research. Nucleic Acids Research (2017) 45(D1): D404-D407.

Wright WP, Kneller JP, Ohlmann ST, Röpke FK, Scholberg K, Seitenzahl IR. Neutrinos from type la supernovae: The gravitationally confined detonation scenario. Phys. Rev. D (2017) 95:043006.

Xu D, Springel V, Sluse D, Schneider P, Sonnenfeld A, Nelson D, Vogelsberger M, Hernquist L. The inner structure of early-type galaxies in the Illustris simulation. Mon. Notices Royal Astron. Soc. (2017) 469:1824–1848.

Zhou B, Hogg P, Gräter F. One-way allosteric communication between the two disulfide bonds in tissue factor. Biophys. journal (2017) 112:78–86.

Zhou X, Lutteropp S, Czech L, Stamatakis A, von Looz M, Rokas A. Quartetbased computations of internode certainty provide accurate and robust measures of phylogenetic incongruence. bioRxiv (2017).

Zhu Q, Hernquist L, Marinacci F, Springel V, Li Y. Baryonic impact on the dark matter orbital properties of Milky Way-sized haloes. Mon. Notices Royal Astron. Soc. (2017) 466:3876–3886.

Zjupa J, Springel V. Angular momentum properties of haloes and their baryon content in the Illustris simulation. Mon. Notices Royal Astron. Soc. (2017) 466:1625–1647.

8 Teaching

Degrees

Adler, Tim:
"A boundary map to the Roller boundary of a CAT(0) cube complex", Master thesis, Heidelberg University: Anna Wienhard, 2017.

Arnold, Christian:
"Hydrodynamical cosmological simulations in f(R) modified gravity", Ph.D. thesis, Physics, Department of Physics and Astronomy, Heidelberg University and HITS: Volker Springel, 2017.

Born, Leo:
"Using Graph-Based Local Coherence for Document-Level MT Decoding", Bachelor thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2017.

Botto Poala, Alberto:
"Hydrodynamic Simulations of Convective Boundary Mixing", Master thesis, Fakultät für Physik und Astronomie, Heidelberg University: Friedrich Röpke, 2017.

Buerkel, Carl:
"Towards development of Random Acceleration Molecular Dynamics simulation analysis software for predicting drug residence time", Bachelor thesis, Molecular Biotechnology, Faculty of Biosciences, Heidelberg University and HITS: Rebecca Wade, Kashif Sadiq and Daria Kokh, 2017.

Ehlert, Kristian:
"Simulating the interaction of extragalactic jets with turbulent magnetic fields in galaxy clusters", Master thesis, Physics, Heidelberg University and HITS: Christoph Pfrommer, 2017.

Eppinger, Mirjam:
"Analyse von Koreferenz-Evaluationsmetriken anhand des Vergleichs mit menschlichen Urteilen", Master thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2017.

Ganster, Kevin:
"Deterministic Forecasts of Binary Events: Comparison of Performance Measures", Bachelor thesis, Faculty of Mathematics, Karlsruhe Institute of Technology: Tilmann Gneiting, 2017.

Gawlok, Simon:
"Numerical methods for compressible flow with meteorological applications", Ph.D. thesis, Mathematics, Heidelberg University: Vincent Heuveline, 2017.

Goicovic, Felipe:
"Infalling clouds onto supermassive black hole binaries", Ph.D. thesis, Physics, Department of Physics and Astronomy, Heidelberg University and Pontificia Universidad Catolica de Chile, Instituto de Astrofisica – Facultad de Fisica and HITS: Volker Springel, 2017.

Gronow, Sabrina:
"Sub-Chandrasekhar mass white dwarfs as Type Ia supernova progenitors", Master thesis, Fakultät für Physik und Astronomie, Heidelberg University: Friedrich Röpke, 2017.

Haupt, Saskia:
"Kalman Filter for the Heat Equation", Master thesis, Mathematics, Heidelberg University: Vincent Heuveline, 2017.

Klar, Manuel:
"Statistical Forecasts of Rain Occurrence over West Africa", Master thesis, Faculty of Mathematics, Karlsruhe Institute of Technology: Tilmann Gneiting, 2017.

Lutteropp, Sarah:
"Error-Profile-Aware Correction of Next Generation Sequencing Read", Master thesis, Computer Science, Karlsruhe Institute of Technology: Alexandros Stamatakis, 2017.

Martschat, Sebastian:
"Structured Representations for Coreference Resolution", Ph.D. thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2017.

Mustafa, Ghulam:
"Modeling and Simulation of Membrane Proteins to Understand their Structure, Dynamics and Function", Ph.D. thesis, Combined Faculty for the Natural Sciences and Mathematics, Heidelberg University and HITS: Rebecca Wade, 2017.

Nikas, Sotirios:
"Characterization Studies of PACIFIC readout ASIC for the LHCb SciFi Tracker", Master thesis, Physics, Heidelberg University: Ulrich Uwer, 2017.

Pardi, Anabele-Linda:
"The impact of supernova feedback on the evolution of the ISM", Ph.D. thesis, Faculty of Physics, LMU München: Simon D. M. White, and HITS: Philipp Girichidis, 2017.

Pfeil, Mareike Katharina:
"Earthquakes in the hyperbolic plane", Master thesis, Heidelberg University: Anna Wienhard, 2017.

Reinhardt, Martin:
"Development of Hydrodynamic Coarse-Grain Models for Brownian Dynamics Simulations of Crowded Protein Systems", Master thesis, Faculty of Physics, Heidelberg University and HITS: Rebecca Wade and Neil Bruce, 2017.

Schmidtobreick, Mareike:
"Parallel asynchronous matrix multiplication for a distributed pipelined neural network", Ph.D. thesis, Mathematics, Computer Science, Heidelberg University: Vincent Heuveline, 2017.

Schoch, Nicolai:
"Towards Cognition-Guided Patient-Specific Numerical Simulation for Cardiac Surgery Assistance", Ph.D. thesis, Mathematics, Medicine, Heidelberg University: Vincent Heuveline, 2017.

Schwaiger, Nico:
"MC3 für phylogenetische Bäume und molekulare Datierung", Master thesis, Mathematics, Karlsruhe Institute of Technology: Alexandros Stamatakis and Tilmann Gneiting, 2017.

Speranskaya, Marina:
"Lexical Coherence Graph: An Application to Authorship Identification", Bachelor thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2017.

Stank, Antonia:
"Computational Studies on the Relation Between Macromolecular Dynamics and Protein Binding and Function", Ph.D. thesis, Combined Faculty for the Natural Sciences and Mathematics, Heidelberg University and HITS: Rebecca Wade, 2017.

Wagner, Johannes:
"Stress-induced molecular processes on multiple scales from simulations", Ph.D. thesis, Faculty for Physics and Astronomy, Heidelberg University, and HITS: Frauke Gräter, 2017.

Wlotzka, Martin:
"Parallel numerical methods for model coupling in nutrient cycle simulations", Ph.D. thesis, Mathematics, Meteorology, Heidelberg University: Vincent Heuveline, 2017.

Yilmaz, Ozan:
"Eine Analyse von militärisch und zivil geförderter US-amerikanischer Forschung im Bereich NLP", Bachelor thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2017.

Lectures, Courses and Seminars

Andreas Bauswein:
"White Dwarfs, Neutron Stars, and Black Holes – Compact Objects in Astrophysics", lecture course, Heidelberg University, Germany, summer semester 2017. *"Neutron star mergers"*, lecture, Heidelberg Summer School 2017. *"Compact Objects & Gravitational Waves"*, Heidelberg University, Germany, September 12, 2017. *"Neutron star mergers"*, lecture, Helmholtz International Summer School, Dubna, Russia, August – September 2017. *"Nuclear Theory and Astrophysical Applications"*, Dubna, Russia, July 19, 2017. *"Tutorial Experimental Physics"*, tutorial, Heidelberg University, Germany, winter semester 2016 / 2017.

Csaba Daday:
"Introduction to equilibrium and force-probe molecular dynamics", IGIB, New Delhi, India, October 30 – November 11, 2017.

Florian Franz, Frauke Gräter:
"Fundamentals of Simulation Methods", Heidelberg University, Germany, October – December 2017.

Tilmann Gneiting:
Lecture course on *"Forecasting: Theory and practice I"*, Karlsruhe Institute of Technology, Germany, October 2016 - February 2017. Lecture course on *"Forecasting: Theory and practice II"*, Karlsruhe Institute of Technology, Germany, April - July 2017.
Seminar on *"Statistical forecasting and classification"*, Karlsruhe Institute of Technology, Germany, October 2017 - February 2018.

Martin Golebiewski, Olga Krebs, Wolfgang Müller:
Coordinating Action Systems Medicine (CASyM) & European Association of Systems Medicine (EASyM) winter school: *"3rd SysBioMED hands-on tutorial – Systems Medicine Approaches in Personalized Medicine"*, Ljubljana, Slovenia, March 29 – April 1, 2017.

Martin Golebiewski, Ron Henkel, Andreas Weidemann:
18th International Conference on Systems Biology (ICSB2018): *"COMBINE & de.NBI Tutorial – Modelling and Simulation Tools in Systems Biology"*, VirginiaTech, Blacksburg, Virginia, USA, August 6, 2017.

8 Teaching

Martin Golebiewski:
1st Liver Systems Medicine (LiSyM) Retreat Meeting: *"LiSyM data management training"*, Hünfeld, Germany, November 20 – 22, 2017.

Martin Golebiewski, Olga Krebs, Wolfgang Müller, Maja Rey:
LiSyM/FAIRDOM/de.NBI Data Structuring Course for Systems Biologists, Hünfeld, Germany, November 22 – 23, 2017.

Frauke Gräter:
Heidelberg Workshop "Multiscale challenges": *"Nacre mechanics: Which laws from the big world hold on the small world of molecules?"* Heidelberg, November 2017. *"Computational biochemistry"*, Heidelberg University, Germany, November 2017.

Frauke Gräter, Katra Kolsek, Csaba Daday (MBM), Rebecca Wade, Gaurav Ganotra, Neil Bruce (MCM):
M.Sc. course *"Computational Molecular Biophysics"*, Heidelberg University, Germany, June – July 2017.

Sabrina Gronow:
"Theoretical Astrophysics", exercises and tutorial accompanying the lecture course, Heidelberg University, Germany, winter semester 2017 / 2018.

Vincent Heuveline, Marcel Kunze:
"IT-Cybersecurity", seminar, Heidelberg University, Germany, winter semester 2016 / 2017 and summer semester 2017.

Vincent Heuveline, Simon Gawlok:
"Einführung in die Numerik", proseminar, Heidelberg University, Germany, winter semester 2016 / 2017.

Vincent Heuveline, Simon Gawlok, Saskia Haupt:
"Uncertainty Quantification 1", lecture, Heidelberg University, Germany, winter semester 2017 / 2018.

Vincent Heuveline:
"IT-Sicherheit", lecture, Heidelberg University, Germany, summer semester 2017.

Leonhard Horst:
"Theoretical Astrophysics", exercises and tutorial accompanying the lecture course, Heidelberg University, winter semester 2017 / 2018. *"Stellar structures, evolution, and explosions"*, exercises and tutorial accompanying the lecture course, Heidelberg University, Germany, winter semester 2016 / 2017.

Olga Krebs:
"2nd FAIRDOM training for for EmPowerPutida", Berlin, Germany, April 18 – 19, 2017. The 9th Young Scientists School *"Systems Biology and Bioinformatics"* SBB-2017: Data Management Training, Yalta, Russia, July 25 – 30, 2017. Data Management Training at Modelling Workshop, Rostock, Germany, 2017.

Markus Kromer:
"Python programming for scientists", lecture course, Heidelberg University, Germany, summer semester 2017. *"White Dwarfs, Neutron Stars, and Black Holes – Compact Objects in Astrophysics"*, seminar accompanying the lecture course, Heidelberg University, Germany, summer semester 2017.

Florian Lach:
"Theoretical Astrophysics", exercises and tutorial accompanying the lecture course, Heidelberg University, Germany, winter semester 2017 / 2018.

Wolfgang Müller:
"Moderne Web-Techniken", Bachelor / Master Seminar Medieninformatik, Universität Bamberg, Germany, summer semester 2017.

Rüdiger Pakmor, Frauke Gräter:
"Fundamentals of Simulation Methods", lecture, Department of Physics and Astronomy, Heidelberg University, Germany, October 2017 – February 2018.

Joanna Panecka-Hofman, Neil Bruce, Daria Kokh:
HGS MathComp Graduate School lecture course *"Computational Methods and Strategies in Structure-based Drug Design"*, Heidelberg University, Germany, December 2017.

Christoph Pfrommer:
"Cosmology", Master seminar, Heidelberg University, Germany, winter semester 2016 / 2017.

Friedrich Röpke:
"Stellar structure, evolution, and explosions", lecture course, Heidel-

berg University, Germany, winter semester 2016 / 2017. *"Thermonuclear Supernovae"*, lecture course, 47th Saas Fee Advanced Course of the Swiss Society for Astrophysics and Astronomy, Villarssur-Ollon, Switzerland, March 12 – 18, 2017.

Friedrich Röpke, Markus Kromer:
"Theoretical Astrophysics", lecture course, Heidelberg University, Germany, winter semester 2017 / 2018.

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

Alexandros Stamatakis, Alexey Kozlov, Tomas Flouri, Pierre Barbera:
"Introduction to Bioinformatics for Computer Scientists", lecture, computer science Master's program at Karlsruhe Institute of Technology, Germany, winter semester 2016 / 2017.

Alexandros Stamatakis, Benoit Morel, Sarah Lutteropp, Pierre Barbera:
"Hot Topics in Bioinformatics", seminar, computer science Master's program at Karlsruhe Institute of Technology, Germany, summer semester 2017.

Alexandros Stamatakis, Lucas Czech:
"Algorithmic Methods in the Humanities", seminar, computer science Master's program at Karlsruhe Institute of Technology, Germany, summer semester 2017. *"Computational Molecular Evolution"*, summer school, Hinxton, Cambridge, United Kingdom, May 2017.

Alexandros Stamatakis:
"Large-scale phylogenetic inference", tutotial, Hellenic Bioinformatics Conference 2017, Heraklion, Greece, September 2017. *"Hands on Bioinformatics Practical"*, practical, computer science Master's program at Karlsruhe Institute of Technology, Germany, summer semester 2017.

Michael Strube:
PhD Colloquium, Department of Computational Linguistics, Heidelberg University, Germany, October 2016 – February 2017. Seminar: *"The Dark Side of NLP: Gefahren automatischer Sprachverarbeitung"*, Department of Computational Linguistics, Heidelberg University, Germany, October 2016 – February 2017. PhD Colloquium, Department of Computational Linguistics, Heidelberg University, Germany, April – July 2017.

Rebecca Wade:
MoBi4all: *"Biomolecular interactions: a computational approach"*, M. Sc. Molecular Biotechnology, Heidelberg University, Germany, January and December 2017. Module 4, *"Biomolecular Recognition: Modeling and Simulation"*, M.Sc. Molecular Cell Biology, Heidelberg University, Germany, March 2017. Module 3, *"Protein Modeling"*, M.Sc. Molecular Cell Biology, Heidelberg University, Germany, May 2017. Ringvorlesung *"Structure and Dynamics of Biological Macromolecules"*, *"Electrostatics, solvation and protein interactions"*, B.Sc. Biosciences, Heidelberg University, Germany, June 2017. Ringvorlesung „Biophysik", *"Receptor-Ligand Interactions: Structure and Dynamics"*, B.Sc. Molecular Biotechnology, Heidelberg University, Germany, November 2017. Ringvorlesung *"Computational Biochemistry"*, *"Electrostatics and Solvation for Biomolecules"*, M. Sc. Biochemistry, Heidelberg University, Germany, December 2017.

Rebecca Wade, Neil Bruce, Anna Feldman-Salit, Gaurav Ganotra, Prajwal Nandekar, Mehmet Ali Öztürk, Ina Pöhner, Stefan Richter (MCM), Frauke Gräter, Davide Mercadente, Camilo Aponte-Santamaria (MBM):
B.Sc. Biosciences practical course *"Grundkurs Bioinformatik"*, Heidelberg University, Germany, January – February 2017.

Rebecca Wade, Daria Kokh, Kashif Sadiq, Stefan Richter, Neil Bruce:
HBIGS Graduate School practical course *"Computational analysis of protein binding properties"*, Heidelberg University, Germany, April 2017.

Anna Wienhard:
"Differentialgeometrie II", Heidelberg University, Germany, winter semester 2017.

Christopher Zapp, Csaba Daday, Frauke Gräter:
"Fundamentals of Simulation Methods", Heidelberg University, Germany, January – February 2017.

9.1 Guest Speaker Activities

Andreas Bauswein:

"Equation of state constraints from postmerger gravitational-wave emission", Neutron star mergers: From gravitational waves to nucleosynthesis, Hirschegg, Austria, January 16, 2017. *"Gravitational waves and heavy elements from neutron-star mergers"*, Niels Bohr Institute, Copenhagen, Denmark, February 6, 2017. *"Dynamics of compact object mergers"*, APS April meeting, Washington, DC, USA, February 28, 2017. *"Neutron-star mergers and highlights from 'Supernovae, Neutron Star Mergers and the Ignition of Dense Matter'"*, Arbeitstreffen Kernphysik 2017, Schleching, Germany, March 9, 2017. *"Neutron-star properties from the gravitational-wave signal of binary mergers"*, New Frontiers in Gravitational-Wave Astrophysics, Rome, Italy, June 6 ,2017. *"Unified picture of postmerger dynamics gravitational-wave emission"*, ECT* workshop *"Bridging Nuclear and Gravitational Physics: the Dense Matter Equation of State"*, Trento, Italy, June 8, 2017. *"Neutron-star mergers and the high-density equation of state"*, 2nd Workshop of the SFB 1245 *"Nuclei: From Fundamental Interactions to Structure and Stars"*, Budenheim, Germany, October 5, 2017. JINA-CEE LIVESTREAM *"The Impact of the LIGO/VIRGO Neutron Star Merger Discovery on Research in Nuclear Science and Nuclear Astrophysics"*, Panelist in online panel discussion, December 1, 2017. *"Gravitational waves from neutron-star mergers"*, ESNT workshop – nucleosynthesis and equation of state of nuclear matter, Saclay, France, December 8, 2017. *"Neutron star mergers"*, Seminar University of Tübingen, Germany, December 18, 2017.

Neil Bruce:

"SGA1 Progress at HITS". SP6 Planning Meeting, Geneva, Switzerland, February 2017. *"Computational Toolbox of Methods for Modelling Drug-Receptor Binding Kinetics"*. K4DD Project Meeting, Frankfurt, Germany, March 2017. *"Computational Approaches to Ligand-Binding"*. HBP CDP6 Inaugural Meeting, Institut Pasteur, Paris, July 2017.

Lucas Czech:

"Trees in the Exelixis Lab", Tree Vizualization Workshop, Chicago, USA, May 2017.

Antonio D'Isanto:

"The two worlds of photometric redshift estimation:fully automatic vs feature based models", Astroinformatics Seminar, Naples (Italy), December 20, 2017.

Philipp Girichidis:

"Dynamical impact of CRs on the structure of the ISM". Lorentz Center workshop on "Bayesian modeling of the Galactic magnetic field: From Faraday Rotation to Ultra-High Energy Cosmic Ray Deflections", Leiden University, The Netherlands, March 23, 2017.

Tilmann Gneiting:

"Statistical post-processing of ensemble weather forecasts", Conference on Predictability and Multi-Scale Prediction of High Impact Weather, Landshut, Germany, October 10, 2017.

Martin Golebiewski:

"Don't be lost in the jungle of standards: Cataloguing, harmonizing and application of modelling standards", Invited talk at COMBINE 2016, Milano, Italy, October 9 – 13, 2017.

Frauke Gräter:

"Support to HPC code: The gap between scientific code development and exascale technology". Podium discussion, PRACEdays, Barcelona, Spain, May 16 – 18, 2017. *"Molecular High-five: fast binding of intrinsically disordered proteins"*. Berlin, FU, International Symposium SFB 765 *"Mulitvalency"*, October 5, 2017; "Viruses and Cells – Computational Challenges and Approaches" Conference, Heidelberg, October 6 – 7, 2017.

Robert Grand:

"Formation of simulated Milky Way-sized galaxies", University of Zuerich, Zuerich, Switzerland, November 2017. *"Cosmo zooms of Milky Way-like galaxies and Galactic archeology"*, Leibniz Institute for Astrophysics, Potsdam, Germany, December 2017. *"News from the dark"* workshop, Montpellier, France, May 2017.

Vincent Heuveline:

Research and and innovation at EMCL (Engineering Mathematics and Computing Lab), Institute of Applied Informatics and Formal Description Methods (AIFB), Karlsruhe Institute of Technology, March 31, 2017.

Svenja Jacob:

"Cosmic-ray heating and mini-halos", Workshop on Diffuse Synchrotron Emission in Clusters of Galaxies, Leiden, Netherlands, October 2017.

Daria Kokh:

"In Silico Prediction of Relative Drug-Protein Residence Times", American Chemical Society meeting, Washington, USA, August 20 – 24, 2017.

Alexey Kozlov:

"Tools for large-scale phylogenetic inference", Global Plant Phylogeny Synthesys Workshop 2017, London, UK, October 2017.

Markus Kromer:

"Towards an understanding of Type Ia supernovae from a synthesis of theory and observations", The AGB-Supernova Mass Transition, International Conference, Observatory of Rome, Monteporzio Catone, Italy, March 28, 2017. *"Deciphering the progenitors and explosion mechanisms of Type Ia supernovae"*, ARI Colloquium, Astronomisches Recheninstitut, Heidelberg, Germany, May 18, 2017. *"Kosmischen Leuchtfeuern auf der Spur"*, Studium Generale, Universität Stuttgart, Germany, May 24, 2017.

Sebastian Lerch:

"Statistical post-processing of ensemble forecasts", International Workshop on Meteorology and Air Traffic Management, Seville, Spain, May 24, 2017.

Wolfgang Müller:

"SABIO-RK from Experiment to Database to Analysis", Beilstein Enzymology Symposium 2017, Rüdesheim, Germany, September 19 – 21, 2017.

Rüdiger Pakmor:

"Die Milchstraße im Computer", Haus der Astronomie, Heidelberg, Germany, October 2017.

Christoph Pfrommer:

"How cosmic rays shape galaxies", Erlangen Center for Astroparticle Physics Seminar, Erlangen University, Germany, Jan 25, 2017; Colloquium, ASTRON, Dwingeloo, The Netherlands, Feb 2, 2017; Colloquium, Center for Computational Astrophysics, New York, USA, May 19, 2017; Astroparticle Seminar, DESY, Zeuthen, Germany, June 16, 2017. *"Cosmic rays and magnetic fields in galaxies"*, Lorentz Center workshop on "Bayesian modeling of the Galactic magnetic field", Leiden University, The Netherlands, March 20 – 24, 2017. *"The impact of cosmic rays on galaxy formation"*, Joint Particle- and Astroparticle Seminar, Karlsruhe Institute of Technology, Germany, July 4, 2017. *"Illuminating cosmological formation shocks"*, Workshop on Diffuse Synchrotron Emission in Clusters of Galaxies, Leiden, Netherlands, October 2017.

Kai Polsterer:

"Machine Learning in Astronomy: lessons learned from learning machines", Colloquium Royal Observatory, Edinburgh, UK,May 31, 2017. *"Astroinformatik / Maschinelles Lernen in der Astronomie; Unsicherheiten machen Aussagen sicherer"*, Netzwerk Recherche Jahreskonferenz, Hamburg (Germany), June 10, 2017. *"Reproducibility in the Era of Data Driven Science; software, data, publications"*, EWASS 2017, Prague, Czech Republic, June 28, 2017; Astroinformatics 2017, Cape Town, South Africa, November 9, 2017. *"Uncertain Photometric Redshifts: deep learning meets probability density functions"*, EWASS 2017, Prague (Czech Republic), June 29, 2017.

Friedrich Röpke:

"The Seven-League Hydro code and adventures in stellar physics" talks at 1st ISSI Meeting "Towards a New Generation of Massive Star Models", ISSI Bern, Switzerland, February 3, 2017. *"Stars on fire – simulating astrophysical burning in stellar evolution and thermonuclear supernovae"*, Innsbruck Physics Colloquium, University of Innsbruck, Austria, May 9, 2017. *"Typ Ia Supernovae – wie explodierende Sterne unser kosmologisches Weltbild erschütterten"*, talk at Haus der Astronomie, Heidelberg, Germany, May 11, 2017. *"Low Mach number fluid dynamics for stellar applications with the Seven-League Hydro code"*, talk at "Stellar Hydro Days IV", University of Victoria, Canada, May 31, 2017. *"Type Ia supernovae and nucleosynthesis"*, talk at "Felsenkeller Workshop", HZDR Dresden, Germany, June 27, 2017. *"Simulating thermonuclear supernovae"*, seminar talk at RIKEN, Tokyo, Japan, September 22, 2017. *"Chandrasekhar-mass SN Ia explosion models"*, talk at conference "Stellar Evolution, Supernova and Nucleosynthesis Across Cosmic Time", Kavli IPMU, Tokyo, Japan, September 27, 2017.

9 Miscellaneous

Kashif Sadiq:

"Heidelberg Institute for Theoretical Studies – Associate Partner", CompBioMed meeting, University College London, London, UK, September 28, 2017. *"Towards multiscale spatiotemporal modeling of retroviral maturation"*, Meeting on "Viruses and Cells – Computational Challenges and Approaches", Heidelberg University, Germany, October 6 – 7, 2017. *"Towards Multiscale Spatiotemporal Modelling of Retroviral Maturation"*, Institute of Theoretical Physics (ITP) Seminar, Heidelberg University, Germany, November 16, 2017.

Anna Schilling, Florian Stecker, Nicolaus Treib:

"Wie Mathematiker Billard spielen – von Dreieckstischen und unendlichen Bahnen", Life Science Lab, Heidelberg, Germany, January 13, 2017; "Tag der Mathematik", Heidelberg University, Germany, October 5, 2017.

Christine Simpson:

"Quenching and ram pressure stripping of simulated Milky Way satellite galaxies", Durham University, Durham, UK, June 2017.

Chen Song:

"Application of Uncertainty Quantification for a rotating device", Northeast Electric Power University (NEEPU), Jilin City, China, July 9 – 11, 2017.

Alexandros Stamatakis:

"Designing next generation RAxML and Tales of Bugs in Bioinformatics Software & Algorithms", Max Planck Institute for Plant Breeding Research, Cologne, Germany, March 2017. *"Supercomputing Challenges in Evolutionary Biology"*, International Supercomputing Conference (ISC2017), Frankfurt, Germany, June 2017.

Volker Springel:

"Hydrodynamical simulations of galaxy formation", Carving Through the Codes Conference, Davos, Switzerland, February 2017. *"Cosmic Structure Formation"*, 28. Edgar Lüscher Seminar on Solid State Physics, Klosters, Switzerland, February 2017. *"Simulierte Universen: Ursprung und Schicksal unserer Milchstraße"*, Leopoldina Vorlesung, Halle, Germany, March 2017; Abendvortrag, Alfried Krupp Wissenschaftskolleg, Greifswald, Germany, September 2017; Blick über den Tellerrand Vortragsreihe, Fraunhofer IWTM, Kaiserslautern, Germany, October 2017. *"Making the Universe and its Contents"*, Wallenberg Foundation Symposium on Big Questions in Astrophysics, Lund, Sweden, April 2017. *"Dark matter on large scales in the Universe"*, Wallenberg Foundation Centenial Celebration Workshop on Dark Matter, Lund, Sweden, April 2017. *"Supercomputer explorations of galaxy formation"*, Physics Colloquium, Massachusetts Institute of Technology, Cambridge, USA, April 2017. *"Supercomputer simulations of the dark and luminous matter in the Universe"*, Physics Colloquium, DESY Hamburg, Germany, May 2017; Physics Colloquium, DESY Zeuthen, Germany, May 2017; Physics Colloquim, Max-Planck Institute for Plasma Physics, Garching, Germany, November 2017; Physics Colloquium, Johannes Gutenberg University Mainz, Germany, November 2017. *"Magnetic fields and cosmic rays in cosmic structure formation"*, Max-Planck Princeton Center for Plasmaphysics Conference (MPPC), Greifswald, Germany, September 2017. *"Cosmic large-scale structure in the IllustrisTNG simulations"*, Results & Review Workshop, High-Performance Computing Center Stuttgart, University of Stuttgart, Germany, October 2017. *"Das Universum: Simulation als Forschungsinstrument"*, Zwischen Theorie und Experiment, Einstein Forum, Potsdam, Germany, November 2017. *"Simulating the dark and luminous Universe"*, Public Talk, University of Iceland, Reykjavik, Iceland, November 2017. *"Simulations of cosmic structure formation"*, Universe Cluster Workshop, Kloster Seeon, Seeon, Germany, November 2017.

Michael Strube:

"Algorithmen und Verantwortung", Public Speech at the March for Science, Heidelberg, Germany, April 23, 2017. *"Big Brother liest, denkt und wirkt mit. Informationstechnologie zwischen gesellschaftlichem Nutzen und Schaden"*, TU Ilmenau, Ilmenau Germany, May 2017. *"Big Brother liest, denkt und wirkt mit. Künstliche Intelligenz zwischen gesellschaftlichem Nutzen und Schaden"*, Vortragsreihe "Heidelberg Digital", Deutsch-Amerikanisches Institut, Heidelberg, Germany, May 15, 2017. *"The Dark Side of NLP: Computerlinguistik zwischen gesellschaftlichem Nutzen und Schaden"*, Netzwerk Recherche 2017, Hamburg, Germany, June 10, 2017. *"Thinking in Graphs. Graph-based Coherence Modeling"*, Keynote Speech at TextGraphs-11: The Workshop on Graph-based Methods for Natural Language Processing, Vancouver, B.C., Canada, August 2017. *"Text Mining and Natural Language Processing"*, SciCAR: Where Science Meets Computer Assisted Reporting, Dortmund, Germany, September 6, 2017. *"Chancen und Gefahren automatischer Sprachverarbeitung"*, BMJV Webdays 2017: Deine Daten, Deine Sicherheit, Deine Meinung, Berlin, Germany, November 2017.

Nicolaus Treib:

"Schottky groups in SL(n,R)", August 6 – 12, 2017, 3rd GEAR Junior Retreat, Stanford / CA, USA.

Freeke van de Voort:

"Gas accretion onto galaxies (or the lack thereof)", Workshop on the physics of quenching massive galaxies at high redshift, Leiden, Netherlands, November 2017. *"Gas accretion onto galaxies and its dependence on stellar mass, redshift, and environment"*, Workshop: A decade of the star-forming main sequence, Leiden, Netherlands, September 2017.

Rebecca Wade:

"In silico prediction of biomolecular recognition", Modelling Chemical and Biological (Re-)activity: MCBR-5, Chennai, India, February 18 – 21, 2017. *"Exploring protein dynamics for ligand design"*, Eleventh European Workshop in Drug Design (XI EWDD), Siena, Italy, May 21 – 26, 2017. *"Computational Approaches to Protein Target Dynamics for Drug Discovery"*, 17th Hellenic Symposium on Medicinal Chemistry (17th HSMC), Thessaloniki, Greece, June 1 – 3, 2017. *"Current State of the Art and Results from Computational Methods for Rates of Binding and Dissociation"*, GSK Webinar, June 23, 2017. *"Exploring protein dynamics for ligand design"*, 11th International Medicinal Chemistry Symposium, AIMECS 2017, and Royal Australian Chemical Institute (RACI) Centenary Congress, Melbourne, Australia, July 23 – 28, 2017. *"Insights into the Membrane and Protein Interactions of Cytochrome P450 Enzymes from Molecular Simulations"*, 20th International Conference on Cytochrome P450: Biochemistry, Biophysics and Biotechnology (ICCP450), Düsseldorf, Germany, August 27 – 31, 2017. *"Computational Mapping of Protein Binding Sites for Drug Discovery"*, SBDD2017, Lausanne, Switzerland, September 5 – 8, 2017. *"Computing Protein Binding Properties"*, DKFZ-ZMBH Alliance Group leaders meeting, Heidelberg, Germany, September 18, 2017. *"Computational Mapping of Protein Binding Sites for Drug Discovery"*, Heinrich Heine University, Düsseldorf, Germany, Nov 6, 2017. *"Learning from an electrostatics-centric view of protein-protein interactions"*, IRB Barcelona, Spain, Nov 10, 2017. *"Modern simulation approaches to study biomolecular interactions"*, iNEXT workshop on Macromolecular Interactions, Eilat, Israel, Dec 1 – 3, 2017. *"Learning from an electrostatics-centric view of protein-protein interactions"*, Conference on "Molecular Perspectives on Protein-Protein Interactions", 2017, Eilat, Israel, Dec 3 – 7, 2017. *"Brownian dynamics simulation of the diffusional association of macromolecules"*, CNR-NANO, Modena, Italy, December 20, 2017.

Jennifer Wagner:

"Linsen & Schätzle – wie reiche Galaxienhaufen auf die Geschichte des Universums blicken", Studium Generale, University of Stuttgart, Germany, July 12, 2017. *"A model-independent approach to gravitational lensing"*, Astrophysics Talk, INAF Bologna Italy, November 14, 2017; Astrophysics seminar, Ben-Gurion University, Be'er Sheva Israel, December 13, 2017.

Rainer Weinberger:

"Feedback from supermassive black holes in cosmological simulations", Leibniz Institute for Astrophysics, Potsdam, Germany, November 2017.

Anna Wienhard:

"A tale of rigidity and flexibility: discrete subgroups of higher rank Lie groups", Berlin Mathematical School, Germany, December 15, 2017.

Dandan Xu:

"The inner structure of early-type galaxies since z=1 from a simulation perspective", Tsinghua University, Beijing, China, April 2017; Beijing University, Beijing, China, April 2017; Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing, China, May 2017; Nanjing University, Nanjing, China, May 2017; Shanghai Jiaotong University, Shanghai, China, May 2017; Shanghai Observatory, Chinese Academy of Sciences, Shanghai, China, May 2017.

Jolanta Zjupa:

"How galactic feedback regulates the angular momentum content of galaxies", ITA Blackboard Colloquium, Institute for Theoretical Astrophysics, Heidelberg University, Germany, July 2017.

9.2 Presentations

Demos

Martin Golebiewski, Olga Krebs, Hadas Leonov, Wolfgang Müller:
"LiSyM SEEK/FAIRDOMhub Demo & Hands-On", Joint LiSyM & ERASysAPP Data Management PALs Meeting, Hünfeld, Germany, May 2 – 3, 2017.

Martin Golebiewski:
"LiSyM SEEK Demo", 1st LiSyM Jamboree & SAB Meeting, Dresden, Germany, May 16 – 17, 2017.

Martin Golebiewski, Ron Henkel, Andreas Weidemann:
"Integrated Data and Model Management: FAIRDOM & SEEK", 18th International Conference on Systems Biology (ICSB2018): "COMBINE & de.NBI Tutorial – Modelling and Simulation Tools in Systems Biology", VirginiaTech, Blacksburg, Virginia, USA, August 6, 2017.

Ron Henkel, Olga Krebs:
"Data and Model Management with FAIRDOM, Seek, and other tools", DigitalLife Bergen Norway, May 2017.

Olga Krebs:
"FAIRDOM Data Management", Sevastopol, Russia, August 2 – 3, 2017.

Andreas Weidemann, Martin Golebiewski:
"SABIO-RK – Reaction Kinetics Database", 18th International Conference on Systems Biology (ICSB2018): "COMBINE & de.NBI Tutorial – Modelling and Simulation Tools in Systems Biology", VirginiaTech, Blacksburg, Virginia, USA, August 6, 2017.

Posters

Camilo Aponte-Santamaría:
"Mechano-induced unfolding of von Willebrand factor: a clinical example of protein destabilization", 61th Annual Meeting of the Biophysical Society. New Orleans, USA, February 11 – 15, 2017.

Pierre Barbera:
"Massively Parallel Evolutionary Placement of Genetic Sequences", Poster, International Supercomputing Conference (ISC2017) PhD Symposium, Frankfurt am Main, Germany, June 2017.

Neil Bruce, Martin Reinhardt, Rebecca Wade, Daniele Narzi, Siri Camee van Keulen, Ursula Roethlisberger, Elisa Frezza, Juliette Martin, Richard Lavery, Modesto Orozco, Francesco Colizzi, Rodrigo Perera Casasnovas, Giulia Rossetti and Paolo Carloni:
"Simulating Neuronal Processes at the Molecular Level". HBP Annual Summit, Glasgow, UK, October 2017.

Csaba Daday:
"Simulating force response at cell junctions: Desmoplakin as a molecular force sensor?", 61th Annual Meeting of the Biophysical Society. New Orleans, USA, February 11 – 15, 2017; EMBO|EMBL Symposium: Mechanical Forces in Biology. Heidelberg, Germany, July 12 – 15, 2017; Multiscale Mechanochemistry & Mechanobiology. Berlin, Germany, October 16 – 18, 2017. *"Coarse-grained Simulations of Conformational Changes in Focal Adhesion Kinase upon Oligomerization"*, Biophysical Society Thematic Meeting on "Conformational Ensembles from Experimental Data and Computer Simulations", Berlin, August 25 – 29, 2017. *"ConAn: Understanding Equilibrium and Force-Probe MD Simulations through Dynamic Contact Maps"*, Workshop on computer simulation and theory of macromolecules. Hünfeld, Germany, March 24 – 25, 2017.

Antonio D'Isanto, Kai Polsterer:
"Uncertain photometric redshifts via combining deep convolutional and mixture density networks", European Symposium on Artificial Neural Networks, ESANN 2017, Bruges, Belgum, April 26 – 28, 2017.

Florian Franz:

"Mechanical indentation of biological membranes", Workshop on computer simulations and theory of macromolecules. Hünfeld, Germany. March 24 – 25, 2017. *"Stress propagation through biological lipid bilayers revealed through atomistic simulations"*, 61th Annual Meeting of the Biophysical Society. New Orleans, USA, February 11 – 15, 2017.

Gaurav Kumar Ganotra, Daria B. Kokh, S. Kashif Sadiq, Rebecca C. Wade:

"COMparative BINding Energy (COMBINE) analysis to predict drug-binding kinetics", K4DD Meeting: "Binding kinetics: Time is of the essence", Berlin, Germany, October 16 – 18, 2017.

Nikos Gianniotis:

"Probabilistic PCA for Time Series", Astroinformatics 2017, Cape Town, South Africa, November 7 – 10, 2017.

Martin Golebiewski, Maja Rey, Andreas Weidemann, Ulrike Wittig, Ron Henkel, Steffen Klamt, Ursula Kummer, Wolfgang Müller:

"de.NBI SysBio: German ELIXIR node serving the Systems Biology Cycle", ELIXIR All Hands Meeting 2017, Rome, Italy, March 21 – 23, 2017.

Robert Grand:

Poster on *"The formation and evolution of Milky Way sized galaxies in high-resolution cosmological zoom simulations"*, Rediscovering our Galaxy Conference, Leibniz Institute for Astrophysics, Potsdam, Germany, July 2017.

Stephan Hemri:

"Discrete post-processing of total cloud cover ensemble forecasts", Poster, European Geosciences Union General Assembly, Vienna, Austria, April 27, 2017.

Ana María Herrera-Rodríguez:

"Towards simulating large-scale self-assembly of proteins under flow", 61th Annual Meeting of the Biophysical Society. New Orleans, USA, February 11 – 15, 2017; workshop on computer simulations and theory of macromolecules, Hünfeld, Germany. March 24 – 25, 2017.

Daria B. Kokh, Marta Amaral, Joerg Bomke, Matthias Dreyer, Matthias Frech, Maryse Lowinski, Alexey Rak, Rebecca C. Wade:

"In Silico Prediction of Relative Drug-Protein Residence Times", Gordon Research Conference on "Next Generation Computer-Aided Medicine Design: Going Beyond Traditional Targets, Pathways, Modalities, Agents and Techniques", Mount Snow, West Dover, Vermont, USA, July 16 – 21, 2017.

Antonia Stank, Daria B. Kokh, Max Horn, Elena Sizikova, Rebecca Neil, Joanna Panecka, Stefan Richter, Rebecca C. Wade:

"Exploring Target Flexibility for Drug Design", American Chemical Society meeting, Washington, USA, August 20 – 24, 2017.

Daria B. 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", K4DD meeting: "Binding kinetics: Time is of the essence", Berlin, Germany, October 16 – 18, 2017.

Sebastian Lerch:

"Comparison of multivariate post-processing approaches", Poster, European Geosciences Union General Assembly, Vienna, Austria, April 27, 2017. *"Evaluation of probabilistic forecasts with the scoring Rules package"*, Poster, European Geosciences Union General Assembly, Vienna, Austria, April 27, 2017.

Benoit Morel:

"A novel heuristic for data distribution in massively parallel phylogenetic inference using site repeats", Poster, International Supercomputing Conference (ISC2017) PhD Symposium, Frankfurt am Main, Germany, June 2017.

Ghulam Mustafa, Prajwal P. Nandekar, Rebecca C. Wade:

"Simulation of Human P450-membrane Interactions", 20th International Conference on Cytochrome P450: Biochemistry, Biophysics and Biotechnology. Düsseldorf, Germany, August 27 – 31, 2017.

9 Miscellaneous

Prajwal P. Nandekar, Goutam Mukherjee, Ghulam Mustafa, Rebecca C. Wade:
"Simulations of Cytochrome P450 and CYP450 Reductase in a Phospholipid Bilayer", 20th International Conference on Cytochrome P450: Biochemistry, Biophysics and Biotechnology. Düsseldorf, Germany, August 27 – 31, 2017.

Prajwal P. Nandekar, Ghulam Mustafa, Rebecca C. Wade:
"Dynamics of a Complex of Cytochrome P450 and CYP450 Reductase in a Phospholipid Bilayer", Workshop on "Computer Simulation and Theory of Macromolecules", Hünfeld, Germany, March 24 – 25, 2017.

Prajwal P. Nandekar, Ross G. Douglas, Julia Aktories, Rebekka Weber, S. Kashif Sadiq, Hirdesh Kumar, Friedrich Frischknecht, Rebecca C. Wade:
"Understanding the Differences in Rabbit and Plasmodium Actin Filament Structure Using Molecular Dynamics Simulations", eCheminfo Euro 2017 Training and Innovation Course in Drug Design, Milan, Italy, July 17 – 21, 2017.

Mehmet Ali Öztürk, Vlad Cojocaru, Rebecca C. Wade:
"Molecular simulations to investigate the sequence dependence of linker histone – nucleosome binding", EMBO Conference: "The Nucleosome: From Atoms to Genome", Heidelberg, Germany, September 1, 2017.

Agnieszka Obarska-Kosinska:
"Atomistic Structural Modeling and Molecular Dynamics Simulation of Collagen Fibril", Workshop on computer simulations and theory of macromolecules. Hünfeld, Germany, March 24 – 25, 2017; EMBO|EMBL Symposium: Mechanical Forces in Biology. Heidelberg, Germany. July 12 – 15, 2017: Multiscale Mechanochemistry & Mechanobiology. Berlin, Germany, October 16 – 18, 2017.

Joanna Panecka-Hofman, Rebecca C. Wade:
"Exploring the flexibility of homotetrameric pteridine reductase 1, a folate pathway enzyme from trypanosomatid human parasites", CECAM Workshop: "Computational approaches to investigating allostery", Lausanne, Switzerland, October 30 – November 1, 2017.

Thomas Sean Powell:
Assembly optimization in both space and time of the largest genome to date Genome Informatics 2017, Cold Spring Harbor Laboratory, NY, USA November 1 – 4, 2017.

Martin Reinhardt, Neil J. Bruce, Rebecca C. Wade:
"Hydrodynamic Coarse-Grain Models for Brownian Dynamics Simulations of Crowded Protein Systems", Workshop on "Computer Simulation and Theory of Macromolecules", Hünfeld, Germany, March 24 – 25, 2017.

Maja Rey, Ulrike Wittig, Andreas Weidemann, Wolfgang Müller:
"SABIO-RK and SEEK: de.NBI Tools for Systems Biology", International Conference on Systems Biology of Human Disease (SBHD), Heidelberg, July 5 – 7, 2017.

Maja Rey, Ulrike Wittig, Ron Henkel, Andreas Weidemann, Wolfgang Müller:
"SEEK: Data management support for systems biology projects", 2nd International de.NBI Symposium "The Future Development of Bioinformatics in Germany and Europe", Bielefeld, Germany, October 23 – 25, 2017.

S. Kashif Sadiq, Daria B. Kokh, Gaurav K. Ganotra, Rebecca C. Wade: *"Computing protein-ligand association kinetics by combining Brownian dynamics and conventional molecular dynamics simulations"*, K4DD meeting: "Binding kinetics: Time is of the essence", Berlin, Germany, October 16 – 18, 2017.

Andreas Weidemann, Ulrike Wittig, Maja Rey, Martin Golebiewski, Wolfgang Müller:
"Kinetics data information retrieval from the literature as public service", 18th International Conference on Systems Biology (ICSB2018), VirginiaTech, Blacksburg, Virginia, USA, August 6 – 12, 2017.

Ulrike Wittig, Maja Rey, Andreas Weidemann, Wolfgang Müller:
"SABIO-RK database: Kinetics data from literature and experiments", 2nd International de.NBI Symposium "The Future Development of Bioinformatics in Germany and Europe", Bielefeld, Germany, October 23 – 25, 2017.

Christopher Zapp:
"Bond Rupture in Biomolecules from QM Calculations", Workshop on computer simulation and theory of macromolecules. Hünfeld, Germany, March 24 – 25, 2017.

Jolanta Zjupa:

Poster on *"The impact of galactic outflows on the IGM in the IllustrisTNG and IllustrisQLA simulations"*, The galaxy ecosystem – Flow of baryons through galaxies, ESO, Garching, Germany, July 2017.

Talks

Camilo Aponte-Santamaría:

"Stress propagation through biological lipid bilayers and the role of mechanosensitive proteins during blood coagulation, elucidated through atomistic simulations", seminar to the group of Prof. Alfredo Alexander-Katz at MIT, Cambridge/MA, USA, February 17, 2017.

Andreas Bauswein:

"Neutron star mergers", Scientific Advisory Board meeting of the Heidelberg Institute for Theoretical Studies, Heidelberg, Germany, December 4, 2017.

Csaba Daday:

"Simulating Force Response at Cell Junctions: Desmoplakin as a Molecular Force Sensor?", Workshop on computer simulation and theory of macromolecules. Hünfeld, Germany, March 24 – 25, 2017. *"Modelling mechanosensing at cellular junctions in health and disease: desmoplakin and plectin"*, Aachen meeting of SPP 1782 of the German Research Foundation, October 12 – 13, 2017; Seminar at IGIB, New Delhi, India, November 9, 2017.

Antonio D'Isanto:

"Deep Learning in Astronomy", astrostatistics seminar, MPIA, Heidelberg, Germany, April 4, 2017. *"Probabilistic photometric redshift derivation from multiband imaging data"*, Annual Meeting of the German Astronomical Society, Splinter Meeting 2017: E-Science & Virtual Observatory, Universität Göttingen (Germany), September 19, 2017. *"Probabilistic photometric redshifts via deep learning"*, Astroinformatics 2017, Cape Town, South Africa, November 9, 2017.

Nikos Gianniotis:

"Linear Dimensionality Reduction for Time Series", International Conference on Neural Information Processing, ICONIP, Guangzhou (China), November 17, 2017. *"Model basedphotometric redshift estimation"*, astrostatistics seminar, MPIA, Heidelberg, Germany, July 4, 2017.

Felipe Goicovic:

"Evolution of massive black hole binaries from incoherent accretion events", Workshop Young Astronomers on Galactic Nuclei, CEFCA, Teruel, Spain, October 2017.

Martin Golebiewski:

"Community Building Activities of CHARME", Combined CHARME working group meetings, Rome, Italy, March 20, 2017. *"Convenor Report of ISO/TC 276/WG5 Data Processing and Integration"*, Plenary Meeting of ISO/TC 276 Biotechnology, Seoul , Republic of Korea), May 13, 2017. *"COMBINE and its Standards: BioPax, SBGN, SBML, SED-ML, CellML, SBOL, NeuroML"*, 18th International Conference on Systems Biology , ICSB2018): "COMBINE & de.NBI Tutorial – Modelling and Simulation Tools in Systems Biology", VirginiaTech, Blacksburg, Virginia , USA, August 6, 2017. *"Data Needs Structure: Data and Model Management for Distributed Systems Biology Projects"*, 18th International Conference on Systems Biology (ICSB2018), VirginiaTech, Blacksburg, Virginia, USA, August 6 – 12 , 2017. *"Standardising Data for Biobanking"*, Global Biobank Week 2017, Stockholm, Sweden, September 13 – 15, 2017. *"COMBINE: A community of communities"*, COMBINE 2016: 8th Computational Modeling in Biology Network Meeting, Milano, Italy, October 9 – 13, 2017. *"FAIR Data Management"*, Workshop: "With CHARME into a transdisciplinary dialogue on Standardisation in life-sciences", Mediterranean institute for life sciences , MedILS), Split, Croatia, October 23 – 25, 2017. *"LiSyM Data Management"*, 1st Liver Systems Medicine (LiSyM) Retreat Meeting, Hünfeld, Germany, November 20 – 22, 2017.

Frauke Gräter:

"Multi-scale silk mechanics", Toyota, Brussels, Multiscale workshop, February 6 – 7, 2017. *"Mechanisms of protein-mechanosensing"*, CNRS IGIB institute, Paris, department seminar, March 29, 2017, Berlin, "Multiscale Mechanochemistry & Mechanobiology conference", October 16 – 18, 2017.

Robert Grand:

"The formation and evolution of Milky Way sized galaxies in high-resolution cosmological zoom simulations", CLUES meeting, UAM, Madrid, June 2017. Thin, thick and dark discs workshop, ETH, Ascona, Switzerland, July 2017. *"How gas shapes Milky Way-sized galaxies in cosmological simulations: mergers, gas-flows and feedback"*, Conference: The Dynamics of Gas in Galaxies, University of Malta, La Valetta, Malta, October 2017.

9 Miscellaneous

Sabrina Gronow:

"A powerful local shear instability", IMPRS Seminar Heidelberg, Germany, November 30, 2017. *"Sub-M_Ch white dwarfs as SN Ia progenitors"*, International Conference of Physics Students, Turin, Italy, August 13, 2017.

Thomas Guillet:

"Discontinuous Galerkin Magnetohydrodynamics in the EXAMAG project", SPPEXA plenary meeting 2017, Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, Garching, Germany, March 2017; SPPEXA Workshop, Heidelberg University, Heidelberg, Germany, July 2017. *"Discontinuous Galerkin methods for astrophysical MHD"*, Astronomische Gesellschaft Annual Meeting 2017, Universität Göttingen, Göttingen, Germany, September 2017; Virgo Consortium Meeting, Max-Planck Institute for Astrophysics, Garching, Germany, December 2017.

Stephan Hemri:

"Spatio-temporal post-processing of hydrological ensemble forecasts", Talk, TIES-GRASPA Conference on Climate and Environment, Bergamo, Italy, July 26, 2017. *"Statistische Post-Prozessierung von hydrologischen Ensemble Vorhersagen"*, Talk, Kolloquium Statistische Methoden in der hydrologischen Vorhersagepraxis und deren Nutzen, Koblenz , Germany, 14 September 2017.

Ana María Herrera-Rodríguez:

"Spider Silk Proteins under Flow by Combining Optical Tweezers and Microfluidic Devices", seminar to the Biological Physics group, University of Gothenburg, Sweden, October 18 – 27, 2017.

Svenja Jacob:

"Cosmic ray driven galactic winds and their dependence on halo mass", ITA Blackboard Colloquium, Heidelberg University, Germany, November 2017.

Alexander Jordan:

"Murphy diagrams", Talk, 7th International Verification Methods Workshop, Berlin, Germany, May 9, 2017. *"Of quantiles and expectiles: Consistent scoring functions, mixture representations, and forecast rankings"*, Talk, TIES-GRASPA Conference on Climate and Environment, Bergamo, Italy, July 26, 2017.

Olga Krebs:

"Guidelines on FAIR Data Management in Horizon 2020", 2nd FAIRDOM training for EmPowerPutida, Berlin, Germany, April 18 – 19, 2017. *"FAIRDOM: Data and Model Management for Systems Biology Projects"*, 9th Young Scientists School "Systems Biology and Bioinformatics", SBB-2017, Yalta, Autonomous Republic of Crimea/Russia, July 25 – 30, 2017. *"FAIRDOM"*, Sevastopol, Russia, August 1, 2017.

Sebastian Lerch:

"Forecaster's dilemma: Extreme events and forecast evaluation", Talk, ACINN Graduate Seminar, University of Innsbruck, Austria, January 11, 2017; Talk, 7th International Verification Methods Workshop, Berlin , Germany, May 8, 2017; Talk, Conference on Predictability and Multi-Scale Prediction of High Impact Weather, Landshut, Germany, October 10, 2017. *"Probabilistic forecasting and comparative model assessment based on Markov chain Monte Carlo output"*, Talk, TIES-GRASPA Conference on Climate and Environment, Bergamo, Italy, July 26, 2017; Talk, Deutsche Bundesbank Workshop on Forecasting, Frankfurt, Germany, September 8, 2017.

Benoit Morel:

"A novel heuristic for data distribution in massively parallel phylogenetic inference using site repeats", Talk, 19th International Conference on High Performance Computing and Communications (HPCC), Bangkok, Thailand, December 2017.

Wolfgang Müller, Martin Golebiewski:

"LiSyM Data Management", 1st LiSyM Jamboree & SAB Meeting, Dresden, Germany, May 16 – 17, 2017.

Rüdiger Pakmor:

"Magnetic fields in disk galaxies", SFB 881 Retreat, Kloster Schöntal, Schöntal, Germany, April 2017. *"MHD and CRs in Arepo"*, Virgo Consortium Annual Meeting, Garching, Germany, December 2017.

Kai Polsterer:

"Reproducibility in the Era of Data Driven Science; software, data, publications", Annual Meeting of the German Astronomical Society, Splinter Meeting 2017: E-Science & Virtual Observatory, Universität Göttingen, Germany, September 19, 2017. *"Astroinformatics; a new discipline or business as usual?"*, ADASS 2017, Santiago, Chile, October 25, 2017. *"Code to the Data: a hard, deep learning example"*, IVOA Interop 2017, Santiago, Chile, October 28, 2017.

Friedrich Röpke:

"Typ Ia Supernovae – Computersimulationen von Sternenexplosionen", HITS Parade, Heidelberg, Germany, February 14, 2017.

Christine Simpson:

"Quenching and ram pressure stripping of simulated Milky Way satellite galaxies", The Role of Gas in Galaxy Dynamics, University of Malta, Valletta, Malta, October 2017. *"Satellites in Auriga"*, Virgo Consortium Workshop, Max-Planck Institute for Astrophysics, Garching, December 2017. *"Cosmic Rays in the Interstellar Medium"*, Virgo Consortium Workshop, Max-Planck Institute for Astrophysics, Garching, December 2017.

Volker Springel:

"Simulations of the Milky Way's Formation in a Cosmological Context", SFB 881 Seminar, Heidelberg, Germany, January 2017. *"Making the Universe and its Contents"*, visit of the Heidelberg Club International at HITS, Heidelberg, Germany, October 2017.

Freeke van de Voort:

"Inflow, outflow, and enrichment", Conference: The circle of life: Connecting the intergalactic, circumgalactic, and interstellar media, Kruger Park, South Africa, August 2017. *"The environmental dependence of gas accretion onto galaxies"*, Conference: The galaxy ecosystem. Flow of baryons through galaxies, ESO, Garching, Germany, July 2017. *"How galactic outflows change the hot haloes around galaxies"*, Conference: On the origin (and evolution) of baryonic galaxy halos, Galapagos Islands, Ecuador, March 2017.

Rebecca Wade:

"Molecular and Cellular Modeling at HITS", HITS/MCM Alumni Meeting, Studio Villa Bosch, HITS, July 1, 2017.

Jennifer Wagner:

"Model-independent characterisation of strong gravitational lenses", Galaxy Clusters 2017, Santander (Spain, , July 4, 2017. *"Applications of Minkowski Tensors in Gravitational Lensing"*, Geometry and Physics of Spatial Random Systems, Bad Herrenalb , Germany, September 12, 2017.

Rainer Weinberger:

"Supermassive black hole feedback in the next generation Illustris simulations", From black hole to environment, Canberra, Australia, August 2017. *"Quenching of central galaxies in the next generation Illustris simulations"*, The role of gas in galaxy dynamics, Valetta, Malta, October 2017. *"Modelling active galactic nucleus feedback in galaxy scale simulations"*, Young astronomers on galactic nuclei, Teruel, Spain, October 2017. *"Modeling active galactic nucleus driven jets in simulations"*, Virgo Consortium Meeting, Max-Planck Institute for Astrophysics, Garching, Germany, December 2017.

Anna Wienhard:

"Die Geometrie des Raumes – vom Leben in verschiedenen Welten", HITS Parade, Heidelberg, Germany, February 14, 2017.

Dandan Xu:

"Combining simulation and observations – a study of early-type galaxies", Conference on Shedding Light on the Dark Universe with Extremely Large Telescopes, Lanzhou, China, Aug/Sep 2017. *"The inner structure of early-type galaxies since z=1 from a simulation perspective"*, Strong Gravitational Lensing Workshop, Cogne, Italy, June 2017.

Jolanta Zjupa:

"The impact of outflows on the CGM and IGM - Ly-alpha forest in the IllustrisTNG/QLA simulations", What Matter(s) Around Galaxies – Resolving the physics of the CGM, Durham University, Durham, UK, June 2017.

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). 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.

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). 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. 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. Associated faculty member, HGS MathComp Graduate School, University of Heidelberg. Faculty member, Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology (HBIGS), University of Heidelberg.

Vincent Heuveline:
Scientific Advisory Board, Potsdam Institute for Climate Impact. Research Scientific Advisory Counsel, Institute for Political Sciences, Heidelberg University. 2013 – 2017: Scientific Advisory Board, German-French University.

Markus Kromer:
Member of the Organizing Committee for the Heidelberg Joint Astronomical Colloquium.

Wolfgang Müller:
Member of the Scientific Advisory Board of the BioModels Database; Deputy Chairman of SIG 4 (Infrastructure & data management), German Network for Bioinformatics Infrastructure (de.NBI); Board Member and Treasurer of FAIRDOM e.V.

Kai Polsterer:
Member of the Section "Knowledge Discovery in Databases" of the Virtual Observatory Alliance, International Astronomical Union (IAU); 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; Arbeitskreis Physik, moderne Informationstechnologie und Künstliche Intelligenz der DPG.

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).

Christine Simpson:
Virgo Consortium Diversity Working Group.

Volker Springel:
Member of the National Academy of Sciences, Leopoldina.

Member of the Interdisciplinary Center for Scientific Computing (IWR), Heidelberg Heidelberg. External Scientific Member of the Max-Planck-Institute for Astronomy, Heidelberg. Member of the Cosmological Simulation Working Group (CSWG) of the EUCLID mission of ESA. Member of the Research Council of the Field of Focus "Structure and pattern formation in the material world" at Heidelberg University. Member of the Steering Committee of the Virgo Consortium. 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.

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. Member Heidelberger Akademie der Wissenschaften. Co-Speaker DFG Research Training Group 2229 "Asymptotic invariants and limits of groups and spaces", Heidelberg – Karlsruhe. Advisory Board of Springer Lecture Notes in Mathematics. Member Scientificc 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 Advidory Board "Wissenschaftskommunikation.de".

Ulrike Wittig:
Member of the STRENDA Commission (Standards for Reporting Enzymology Data).

9.4 Contributions to the Scientific Community

Program Committee Memberships

Wolfgang Müller:
de.NBI Symposium 2017: The Future Development of Bioinformatics in Germany and Europe, October 23 – 25, 2017, Bielefeld, Germany. Third International Workshop on Data Management and Analytics for Medicine and Healthcare, Workshop on the 43rd International Conference on Very Large Data Bases, September 1, 2017, Munich, Germany.

Michael Strube:
Program Co-Chair of the First ACL Workshop on Ethics in Natural Language Processing. Valencia, Spain, April 4, 2017. Organization Committee Memberships (Chair).

Andreas Bauswein:
Member of Organizing Committee for the ECT* workshop "Bridging nuclear and gravitational physics: the dense matter equation of state", Trento, Italy, June 5 – 9, 2017.

Martin Golebiewski:
Combined CHARME working group meetings, Rome, Italy, March 20, 2017. CHARME Management Committee Meeting, Riga, Latvia, May 28 – 29, 2017. COMBINE 2016: 8th Computational Modeling in Biology Network Meeting, Milano, ItalyOctober 9 – 13, 2017. Workshop: "With CHARME into a transdisciplinary dialogue on Standardisation in life-sciences", Mediterranean institute for life sciences (MedILS), Split, Croatia, October 23 – 25, 2017.

Christoph Pfrommer:
Chair of the Organizing Committee for the AIP Thinkshop 2018 on "The role of feedback in galaxy formation: from small-scale winds to large-scale outflows", Potsdam, Germany. Member of the Organizing Committee for the 2018 TeV Particle Astrophysics conference (TeVPA 2018), Berlin, Germany.

Friedrich Röpke, Andreas Bauswein:
Members of Organizing Committee for the 12th Heidelberg Summer School 2017 "Compact Objects & Gravitational Waves", Heidelberg, Germany, September 11 – 15, 2017.

Friedrich Röpke, Markus Krombер, Sebastian Ohlmann:
Members of Organizing Committee for the "12th Würzburg Workshop in Heidelberg", Heidelberg, Germany, December 14 /15, 2017.

Michael Strube:
Program Co-Chair of the First ACL Workshop on Ethics in Natural Language Processing. Valencia, Spain, April 4, 2017.

Rebecca Wade:
iNEXT workshop on Macromolecular Interactions, Eilat, Israel, Dec 1 – 3, 2017. Conference on "Molecular Perspectives on Protein-Protein Interactions", 2017, Eilat, Israel, Dec 3 – 7, 2017.

Workshop Organization

Daniele Alessandrini, Beatrice Pozzetti, Roman Sauer, Petra Schwer, Anna Wienhard:
Asymptotic geometry of groups and spaces, Heidelberg University, Germany, February 20 – 23, 2017.

Kira Feldmann, Svenja Jacob, Alexander Jordan, Daria Kokh, Sarah Lutteropp, Ina Pöhner, Jolanta Zjupa:
Girls' Day 2017 at HITS, Heidelberg, Germany, April 27, 2017.

Simon Gawlok, Vincent Heuveline:
Workshop on Energy-Aware High Performance Computing held in conjunction with ISC-High Performance June 22, 2017, Frankfurt, Germany.

Martin Golebiewski:
Joint LiSyM & ERASysAPP Data Management PALs Meeting, Hünfeld, Germany, May 2 – 3, 2017. COMBINE & de.NBI Tutorial – Modelling and Simulation Tools in Systems Biology, 18th International Conference on Systems Biology (ICSB2018), VirginiaTech, Blacksburg, Virginia, USA, August 6, 2017. CHARME WG 1 and 3 meeting, Milano, Italy, October 12 – 13, 2017. LiSyM/FAIRDOM/de.NBI Data Structuring Course for Systems Biologists, Hünfeld, Germany, November 22 – 23, 2017. Committee Meetings of ISO/TC 276 Biotechnology working group WG5 "Data Processing and Integration": Seoul, Republic of Korea, May 8 – 12, 2017. Rome, Italy, November 27 – December 1, 2017.

Vincent Heuveline, Jonas Kratzke:
E-Science Tage 2017, March 16 – 17, 2017, Heidelberg, Germany.

Vincent Heuveline, Chen Song:
Sino-German cooperation with Northeast Electric Power University (NEEPU), July 9 – 11, 2017, Jilin, China.

Olga Krebs:
Joint LiSyM & ERASysAPP Data Management PALs Meeting, Hünfeld (Germany), May 2 – 3, 2017.

Kai Polsterer:
IEEE SSCI special session on "Mining the Sky: Knowledge Discovery in Big and Complex Astronomical Data Sets and Data Streams" in Hawaii, USA Session of the KDDIG at the International Virtual Observatory Alliance Interoperability Meeting in Shanghai, China Session of the KDDIG at the International Virtual Observatory Alliance Interoperability Meeting in Santiago, Chile BoF on Machine Learning in Astronomy at ADASS 2017 in Santiago, Chile Co-organizer, splinter on E-Science & Virtual Observatory at the Annual Meeting of the Astronomische Gesellschaft 2017 Coorganizer, the "Astroinformatics 2017" in Cape Town, South Africa.

Anna Schilling, Petra Schwer, Anna Wienhard:
Compactifications of Buildings and Symmetric Spaces, Heidelberg, May 16 – 17, 2017.

Alexandros Stamatakis:
Co-organizer of 2017 Computational Molecular Evolution Summer School, Hinxton, UK.

Rebecca Wade:
MCM Group alumni meeting, Studio Villa Bosch, July 1, 2017.

Anna Wienhard:
Heidelberg-Karlsruhe-Strasbourg Geometry Day X, Heidelberg, December 1, 2017.

Jolanta Zjupa:
Dark Matter Day at Zentrum Astronomie Heidelberg (ZAH), Heidelberg University, Germany, October 31, 2017.

Editorial Work

Tilmann Gneiting:
Editor-In-Chief, Annals of Applied Statistics.

Frauke Gräter:
Biophysical Journal; Journal of the American Chemical Society PLoS Journals; Nature Journals; Proceedings of the National Academy of Sciences; German Research Society (DFG).

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.

Anna Wienhard:
Editor of Annales Scientifique de l'École normale supérieure. Editor of Geometriae Dedicata. Editor Forum Mathematicum. Editor Geometry and Topology.

9.5 Awards

Andreas Bauswein:
Starting Grant from the European Research Council (ERC). Research project: "GreatMoves: General Relativistic Moving-Mesh Simulations of Neutron Star Mergers".

Tilmann Gneiting:
Highly Cited Researcher in Mathematics for 2005 – 2015, Clarivate Analytics (formerly Thomson Reuters), 2017. Fellow, American Statistical Association, 2017.

Frauke Gräter:
PRACE Ada Lovelace Award for HPC, awarded by the Partnership for Advanced Computing in Europe (PRACE).

Prajwal Nandekar:
First prize, with Dr. Subhash Mohan Agarwal, BioSolveIT Spring 2016 Scientific Challenge.

Alexandros Stamatakis:
Highly Cited Researcher in Computer Science for 2005 – 2015, Clarivate Analytics (formerly Thomson Reuters), 2017.

Volker Springel:
Golden Spike Award of the High-Performance Computing Center Stuttgart, Germany, October 2017. Highly Cited Researcher in Space Science for 2005 – 2015, Clarivate Analytics (formerly Thomson Reuters), 2017.

Figure 89: The HITS Scientific Advisory Board at HITS 2017. From left to right: Wolfgang Müller (HITS Deputy Scientific Director), Michael Strube (HITS Scientific Director), Tony Hey, Heribert Knorr, Gert-Martin Greuel, Gesa Schönberger (HITS Managing Director), Alex Szalay, Thomas Lengauer, Adele Goldberg, Dieter Kranzlmüller.

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 2017, the board consisted of the following 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

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 Director:

Dr. Gesa Schönberger
(since January 2016)

Scientific Director:

Prof. Dr. Michael Strube
(2017 – 2018)

Deputy Scientific Director:

Priv.-Doz. Dr. Wolfgang Müller
(2017 – 2018)

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 | © 2018 HITS gGmbH

Twitter: @HITStudies
Facebook: /HITStudies
Youtube: /TheHITSters
LinkedIn: company/the-heidelberg-institute-for-theoretical-studies
Instagram: the_hitsters