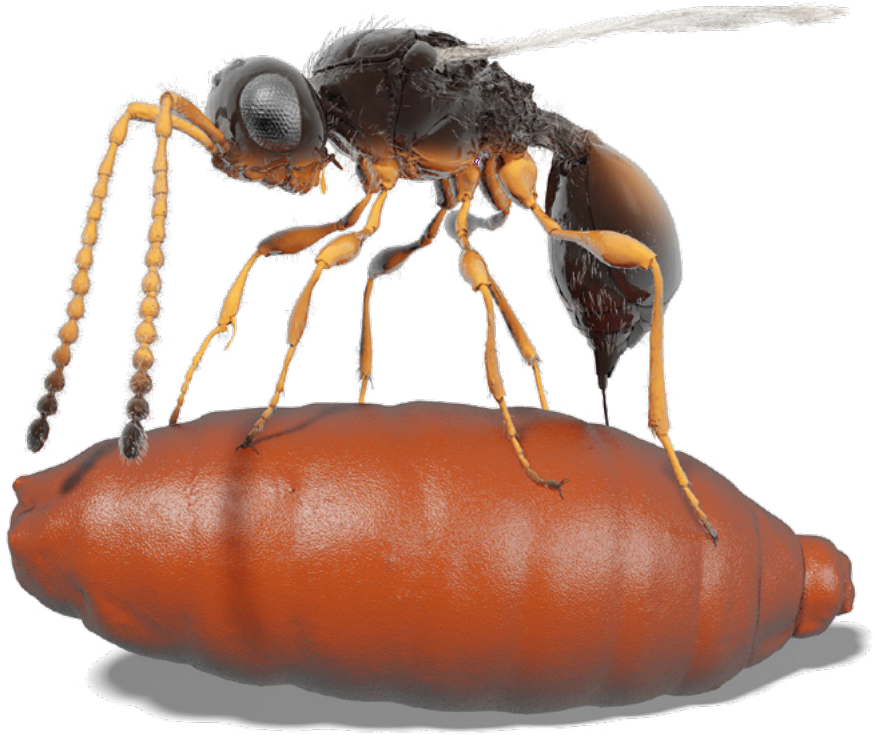
2018

Annual Report
Jahresbericht

Heidelberg Institute for
Theoretical Studies

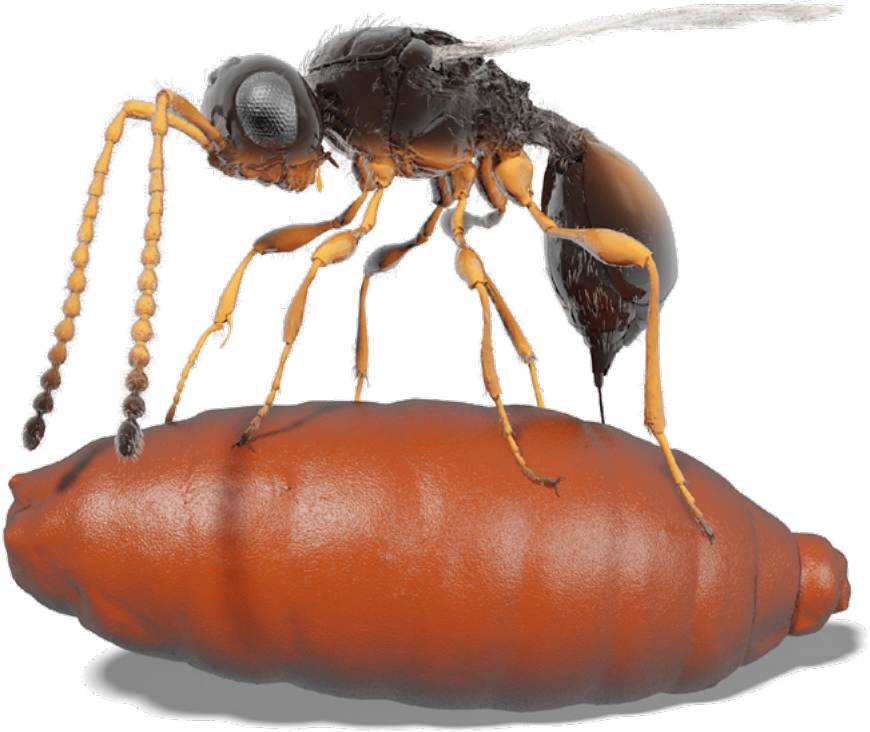
HITS



# THINK BEYOND THE LIMITS!

Digitally resurrected: The parasitic wasp Xenomorphia resurrecta deposits an egg in a fly pupa (cf. *Chapter 2.3*, p. 35. Figure: Thomas van de Kamp, KIT; Nature Communications).

*Digital wiederauferstanden: Die parasitische Wespe Xenomorphia resurrecta legt ein Ei in einer Fliegenpuppe ab. (vgl. Kapitel 2.3, S. 35. Bild: Thomas van de Kamp, KIT; Nature Communications).*



Inhalt
Table of Contents

# 1
# Think Beyond the Limits!

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

*Dr. Gesa Schönberger*
*(Managing Director / Geschäftsführerin)*

Have you ever wanted to scale down the Universe to 11.5 cm²? HITS makes it possible! Just take a look at the cover letter of this Annual Report – or place an online order with Deutsche Post at https://bit.ly/2SPiC5A. In late 2018, the German Postal Service released a postage stamp with a new simulation of the Universe generated by the Illustris TNG Simulation.

Earlier in the year, the HITS research group Theoretical Astrophysics – led by Volker Springel – published a series of reports on the new simulation. The researchers calculated how black holes influence the distribution of dark matter, how heavy elements are produced and distributed throughout the cosmos, and where magnetic fields originate. These calculations were enabled via the development and programming of a new simulation model for the Universe, Illustris TNG, the most complete simulation of its kind to date. Illustris TNG is only the latest piece of highly visible research from Volker Springel's Theoretical Astrophysics group *(see Chapter 2.11)*.

Hence, while we were sad when we learned that Volker would leave HITS in August 2018 to take up a new position as Director at the Max Planck Institute for Astrophysics in Garching, we were also happy to see that HITS enables researchers to grow and move on to the most prestigious positions in their respective fields. To honor Volker's contribution to HITS' reputation, he was appointed a HITS Fellow in July 2018 *(see Chapter 9.5)*.

The Computational Biology junior group brought its work to a close in 2018, which was according to plan. Group leader Siegfried Schloissnig accepted a position at the Institute for Molecular Pathology in Vienna, where he now heads the IT department for regeneration re-

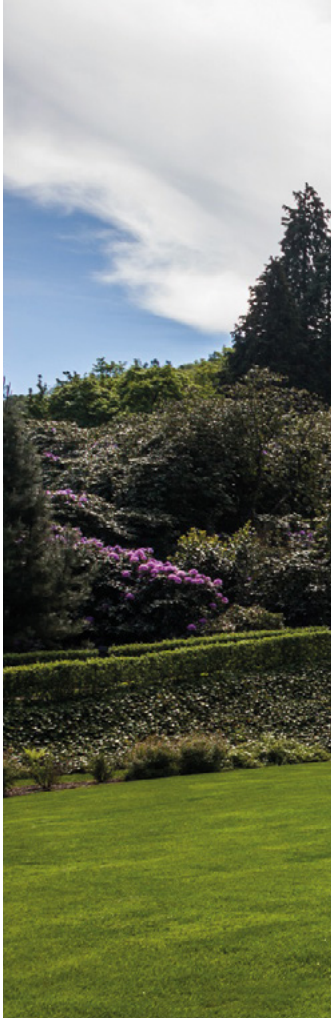
search. Just before leaving, the group co-authored two papers published in Nature that dealt with deciphering the Axolotl and the Schmidtea mediterranea genomes, two species characterized by high regeneration capabilities. Again, we were proud to see researchers at HITS publish papers in the most prestigious journal in the natural sciences *(see Chapter 4)*.

Change is a common phenomenon at HITS, we wish all the best to all those who left the institute in 2018 and hope to see them at our Alumni Meetings, the next of which will be held in July 2019.
However, there is no danger of HITS' gradual disappearance. With Ganna (Anya) Gryn'ova we hired a new group leader for a junior group on Computational Materials Science who will assume her position in April 2019 *(see Chapter 5.1.4)*. A second process of hiring a group leader for a full group on Computational Astrophysics together with the Heidelberg University is ongoing. Finally, we initiated the hiring process for a junior group leader on Machine Learning together with the Karlsruhe Institute of Technology.

Since its inception, HITS has emphasized interdisciplinary research. Group leaders constantly strive to strengthen existing formats – such as the lab meeting and the scientific seminar series – and to come up with new formats. Hence, to foster interdisciplinary collaboration between the HITS research groups, we organized the "HITS Fest" in July 2018 *(see Chapter 4)*. The event began with an invited talk by Matthias Scheffler (Fritz Haber Institute of the Max Planck Society in Berlin). We continued with a ten-station world café at which HITSters contributed their thoughts on interdisciplinary research, on HITS' future, on computing, on making science accessible to the public, and on social life at HITS. The results were presented during a barbecue, and a pub quiz concluded the event. Moreover, the group leaders are working on a strategy to intensify collaboration between different scientific disciplines for 2019. We will keep you informed of further developments.
In order to be proactive in these change processes, HITS launched an internal vision and mission process and is currently working on a basic concept for the development of the institute in the near future. This concept will also form the basis of the new corporate design.

Reaching out to the public is another regular part of the scientific work conducted at HITS. We feel that our highly specialized scientific research should be shared with as many people as possible. For example, in our efforts to make astronomy accessible to the public, we were very pleased to finalize a large joint project with the Klaus Tschira Stiftung: the ESO Supernova Planetarium and Visitor Centre. After years of planning and building, programming and implementing we celebrated the center opening in April 2018 *(see Chapter 6)*. If you are interested in astronomy, this extraordinary site located in Garching near Munich is worth visiting. There, you can explore the Universe yourself, which is scaled down not to the size of a stamp, but rather to 7,500 $m^2$.


# 1 Think Beyond the Limits!



*Wollten Sie schon immer mal das Universum auf 11,5 cm² sehen? Das HITS macht's möglich! Werfen Sie einfach einen Blick auf den Begleitbrief dieses Jahresberichts – oder bestellen Sie online unter https://bit.ly/2SPiC5A. Ende 2018 gab die Deutsche Post eine Briefmarke heraus, die eine neue Simulation des Universums zeigt, die mit Hilfe der Illustris TNG Simulation erstellt wurde.*

*Im Vorfeld hatte die Forschungsgruppe „Theoretical Astrophysics" am HITS unter der Leitung von Volker Springel dazu eine Reihe von Artikeln veröffentlicht, in denen die Wissenschaftlerinnen und Wissenschaftler berechneten, in welcher Weise Schwarze Löcher die Verteilung von dunkler Materie beeinflussen, wie Schwermetalle erzeugt und über den Kosmos verteilt werden und wie Magnetfelder entstehen. Ermöglicht wurden diese Berechnungen durch die Entwicklung und Programmierung eines neuen Simulationsmodels für das Universum, Illustris TNG, die bislang umfassendste Simulation dieser Art. Dabei reiht sich Illustris TNG nahtlos ein in die beeindruckende Serie von Forschungsergebnissen aus dieser Gruppe (siehe Kapitel 2.11).*

*Und so sehen wir den Weggang von Volker, der das HITS im August 2018 verließ, um Direktor am Max-Planck-Institut für Astrophysik in Garching zu werden, mit einem lachenden und einem weinenden Auge. Denn so traurig wir über den Verlust sind, so froh sind wir darüber, dass Wissenschaftler am HITS die Möglichkeit haben, sich zu entwickeln und sich für die angesehensten Positionen in ihrem Forschungsgebiet zu qualifizieren. Um Volkers Beitrag für den Ruf des Instituts entsprechend zu würdigen, wurde er im Juli 2018 zum HITS Fellow ernannt (siehe Kapitel 9.5).*

*Auch die Arbeit der Juniorgruppe „Computational Biology" endete planmäßig im Jahr 2018. Gruppenleiter Siegfried Schloissnig wechselte ans Forschungsinstitut für Molekulare Pathologie in Wien, wo er die Leitung der IT-Abteilung für Regenerationsforschung übernahm. Kurz zuvor war die Gruppe jeweils als Co-Autor an zwei „Nature"-Artikeln über die Entschlüsselung der Genome von Axolotl und Schmidtea mediterranea beteiligt. Beide Spezies verfügen über hervorragende Regenerationsfähigkeiten. Wieder einmal konnten wir mit Stolz zur Kenntnis nehmen, dass Wissenschaftlerinnen und Wissenschaftler am HITS in den renommiertesten naturwissenschaftlichen Fachzeitschriften publizieren (siehe Kapitel 4).*

*Der Wandel ist ein wohlbekanntes Phänomen am HITS. Und so wünschen wir allen, die das Institut 2018 verlassen haben, alles Gute und freuen uns darauf, sie bei einem unserer Alumni Meetings begrüßen zu können, von denen das nächste im Juli 2019 stattfindet.*
*Von einer allmählichen Auflösung des HITS kann jedoch keine Rede sein: 2018 konnten wir mit Ganna (Anya) Gryn'ova die Leiterin für eine neue Juniorgruppe mit dem Forschungsschwerpunkt „Computergestützte Materialforschung" verpflichten, die ihre Arbeit im April 2019 aufnehmen wird (siehe Kapitel 5.1.4). Ein weiterer Rekrutierungsprozess mit der Universität Heidelberg für eine Gruppenleiterstelle im Bereich „Computergestützte Astro-*

*physik“ ist im Gange. Zusammen mit dem Karlsruher Institut für Technologie haben wir die Suche nach einem Junior-Gruppenleiter für maschinelles Lernen erfolgreich auf den Weg gebracht.*

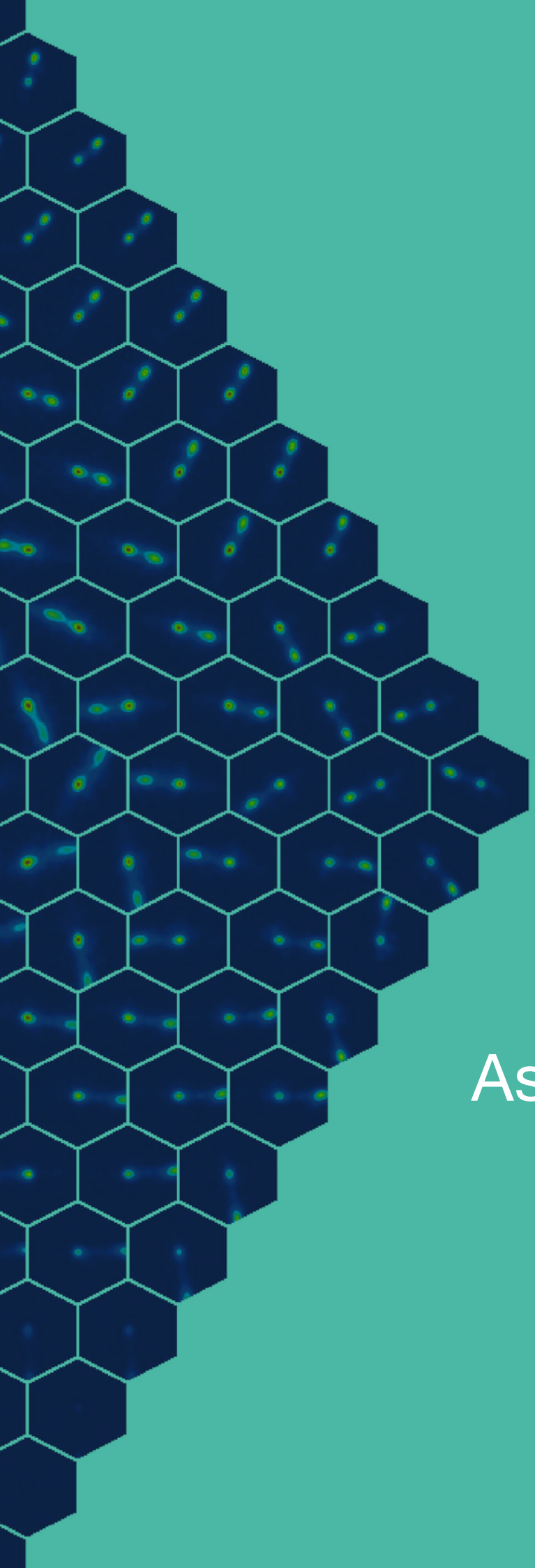
*Seit seiner Gründung legt das HITS großen Wert auf interdisziplinäre Forschung,. Die Gruppenleiter sind stets bestrebt, bewährte Formate wie zum Beispiel das Lab Meeting oder die „Scientific Seminar Series“ zu stärken und neue Formate zu entwickeln. So diente das “HITS-Fest” im Juli 2018 dazu, die interdisziplinäre Zusammenarbeit zwischen den einzelnen Forschungsgruppen zu fördern (siehe Kapitel 4). Die Veranstaltung begann mit einem Vortrag von Matthias Scheffler vom Fritz-Haber-Institut der Max-Planck-Gesellschaft in Berlin. Danach konnten die HITSter in einem World-Café an zehn verschiedenen Stationen ihre Ideen äußern. Die Themen reichten dabei von interdisziplinärer Forschung und der Zukunft des HITS, fachspezifischene Themen und Öffentlichkeitsarbeit bis hin zum sozialen Miteinander am Institut. Die Ergebnisse wurden während des anschließenden Grillfestes vorgestellt, ein Pub Quiz rundete die Veranstaltung ab.*

*Darüber hinaus arbeiten die Gruppenleiter an einer Strategie, um die Zusammenarbeit zwischen den verschiedenen wissenschaftlichen Disziplinen 2019 zu intensivieren. Wir werden Sie über die weitere Entwicklung auf dem Laufenden halten.*

*Um diesen Prozess aktiv mitzugestalten, entwickelt das HITS zurzeit interne Leitlinien, die das Institut in den nächsten Jahren prägen und die Grundlage für das neu zu entwickelnde Corporate Design bilden werden.*

*Eine engagierte Öffentlichkeitsarbeit ist ein weiterer wichtiger Bestandteil der wissenschaftlichen Arbeit am HITS, denn wir sind der Meinung, dass die Ergebnisse unserer Forschung ein möglichst breites Publikum erreichen sollten. Ein gelungenes Beispiel dafür ist das erst kürzlich eröffnete „ESO Supernova Planetarium and Visitor Centre“, ein gemeinsames Projekt mit der Klaus Tschira Stiftung, das astronomische Forschung der allgemeinen Öffentlichkeit zugänglich macht. Nach Jahren umfangreicher Planung und Bauarbeiten, des Programmierens und schließlich der Umsetzung der Inhalte konnten wir die Eröffnung des Zentrums im April 2018 in Garching bei München feiern (siehe Kapitel 6). Falls Sie sich für dieses Themengebiet interessieren, können wir Ihnen einen Besuch an diesem besonderen Ort nur wärmstens empfehlen. Dort haben Sie die Gelegenheit, das Universum auf immerhin 7500 m² selbst zu erkunden.*

# 2 Research



# 2.1 Astroinformatics (AIN)

## 2.1
## Astroinformatics (AIN)

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

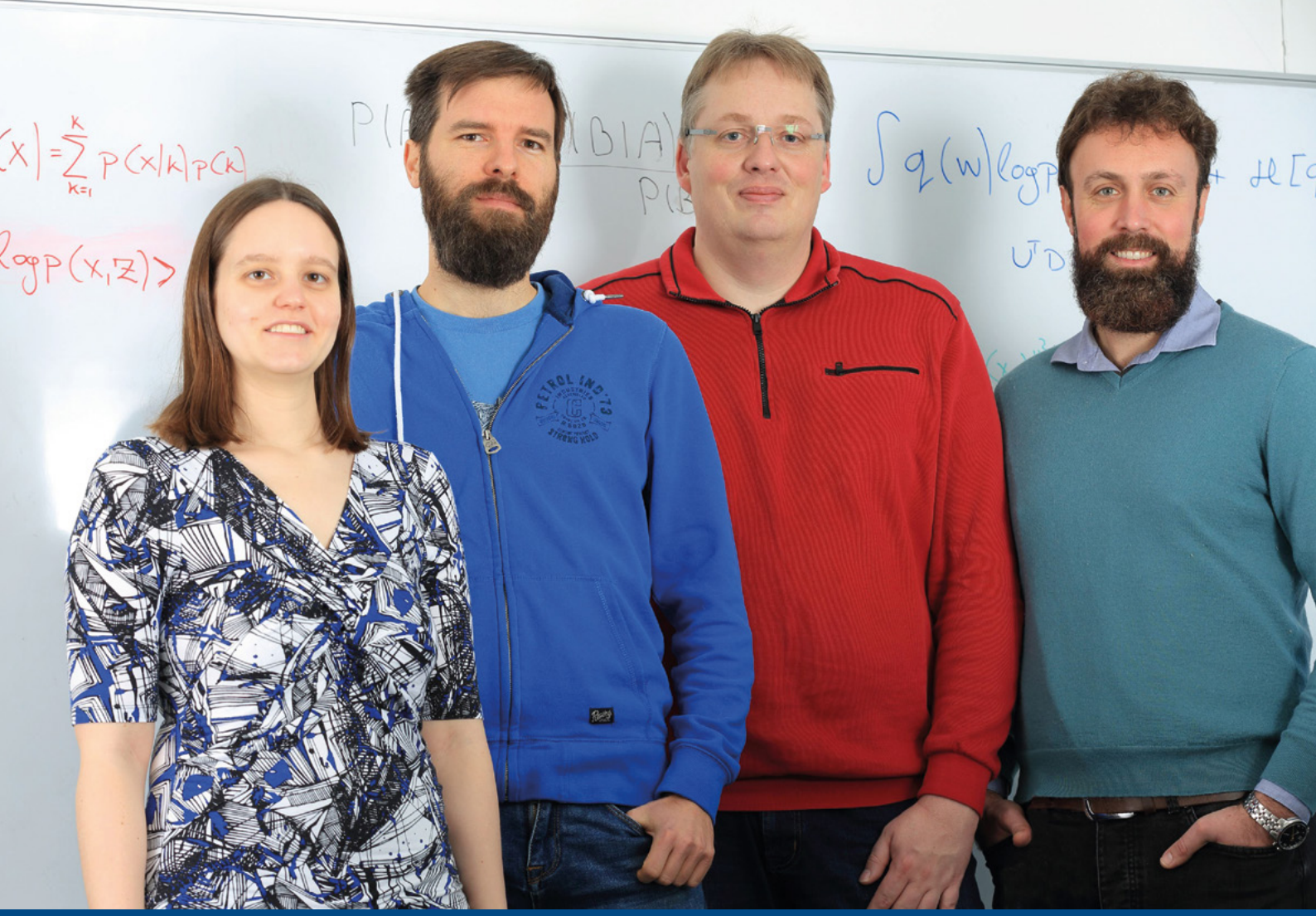
Over the past two decades, computers have revolutionized astronomy. Advances in technology have given rise to new detectors, complex instruments, and innovative telescope designs. These advances enable today's astronomers to observe more objects than ever before and with high spatial/spectral/temporal resolution. In addition, 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 developing 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 to extract and filter rare objects (outliers) for detailed follow-up analysis with 8-m class telescopes. With estimated occurrences of only a few objects per million, manually inspecting the existing catalogs is not an option. The Astroinformatics Group's other interests include morphologically classifying galaxies based on imaging data as well as measuring similarity in high-dimensional data spaces.

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

*In den letzten zwanzig Jahren hat der Einsatz von Computern die Astronomie stark beeinflusst. Durch technologische Fortschritte wurde es möglich, neue Detektoren sowie innovative Instrumente und Teleskopdesigns zu realisieren. Dadurch können Astronomen nun Objekte mit bisher unerreichtem Detailreichtum und in neuen Wellenlängenbereichen beobachten. Mit speziell dafür vorgesehenen Teleskopen wird der Himmel wiederholt durchmustert, und die so 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.*



## Group Leader

Dr. Kai Polsterer

## Staff Member

Dr. Nikos Gianniotis

## Scholarship Holder

Antonio D'Isanto *(HITS Scholarship)*

## Student Assistant

Erica Hopkins

## Return of the features

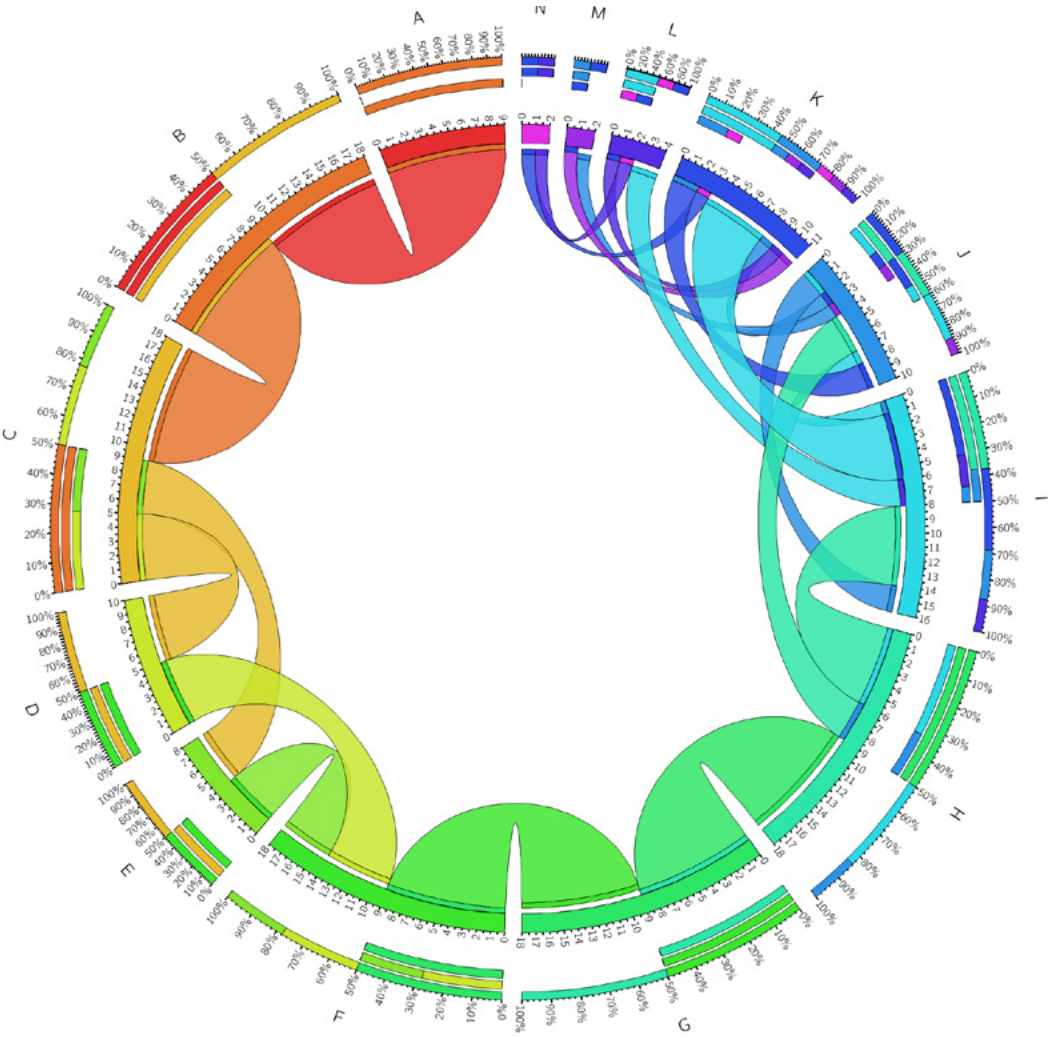
The explosion in data volume that astronomy has experienced in the recent years has increased the demand for innovative analysis techniques capable of dealing with this level of complexity. In this sense, machine learning is the key to providing practical solutions that perform well. In particular, the application of deep learning models brings the advantage of fully automating tasks (e.g., features extraction and selection), which is of fundamental importance to both regression and classification tasks. Features constitute a representation of the original information given by astronomical data (e.g., images or time series). An accurate selection of the best-performing features is of extreme importance in order to guarantee the reliability of predictive models. Automating such a task constitutes an improvement in terms of speed, efficiency, and the amount of information used. On the other hand, deep learning models have the disadvantage of being black boxes that lack physical interpretability in terms of the features they automatically extract. Motivated by this lack of interpretability, we developed a model that combines the advantage of the massive use of information – which characterizes deep learning – with the interpretability offered by the use of features.

The proposed model has been applied to the scientific case of photometric redshift estimation. Redshift is a fundamental measure in astronomy and is measured via spectroscopy. However, due to the difficulty and time requirements in obtaining spectroscopic redshifts for all required sources, obtaining estimates of redshift photometrically has become increasingly popular. The availability of photometric redshifts for a large number of sources is of high priority and is key in the successful deployment of a number of forthcoming projects and missions, such as Euclid. In this work, we synthetically built 4,520 features by adopting a controlled combination strategy of magnitudes, errors, radii, and ellipticities for spectroscopically confirmed quasars from the Sloan Digital Sky Survey (SDSS). Among these features, we selected several best-set candidates by means of a forward selection scheme. This feature-selection strategy was based on recursive experiments performed through a k-Nearest Neighbors (kNN) algorithm in which the best-performing feature was selected and fixed at each step. In this manner, a tree of feature sets was built in which each branch of the tree constituted a best-set candidate. As expected, we noted in the results that no shape-related features were selected as the objects were quasar point-like sources. Preliminary tests performed on extended sources (e.g., galaxies), however, revealed an increasing importance of features that capture shape variations.

The feature sets were finally used as inputs to a random forest – and not the previously used kNN algorithm – in order to reduce the potential risk of overtuning the discovered features to the particular use of the kNN. The experiments demonstrate that the discovered set of features leads to an improvement in the predictive performance obtained when using the standard 10 classic features often employed in the relevant literature (i.e., magnitudes and colors). In actual fact, we noticed in the experiments that it was possible to achieve superior predictive performance with as few as four selected features. Furthermore, the best sets of features were able to outperform even a fully automatic model based on a convolutional neural network. The performance was quantified and verified using the Continuous Ranked Probability Score (CRPS) and additional standard measures, such as the root mean square error (RMSE).

The structure of the feature tree was visually represented using a chord diagram (commonly adopted in the biologic field) (see *Figure 1*) in order to visualize and represent the correlations and interconnections between the different features. The use of known parameters taken from the SDSS database allowed us to physically interpret a subset of the discovered features. In fact, we found a clear correlation between a subset of the features and certain quasar emission lines, which are representative of the physical processes that occur in the selected sources. The correlation was found by comparing the importance of the selected features calculated by means of the random forest (i.e., the Gini index) per redshift bin with the position of the emission lines in the redshift space. From the plots (*Figure 2*), a striking correspondence between features and spectral lines is evident yet is only partially true in the case of the classic features. These considerations lead us to the conclusion that the developed model is capable of predicting affordable photometric redshift estimates by selecting only features that perform well and are physically meaningful.

*Figure 1: Example of chord diagram of the derived features. Every feature corresponds to a color, and beginning with the first feature (A), all possible paths of the tree can be followed, and different feature subsets can be defined. Ordered from outside to inside, the arc-segments represent the occurrences of a particular feature, namely the total percentage of the individual connections, numbers, and sources of entering connections as well as the numbers and targets of exiting connections.*



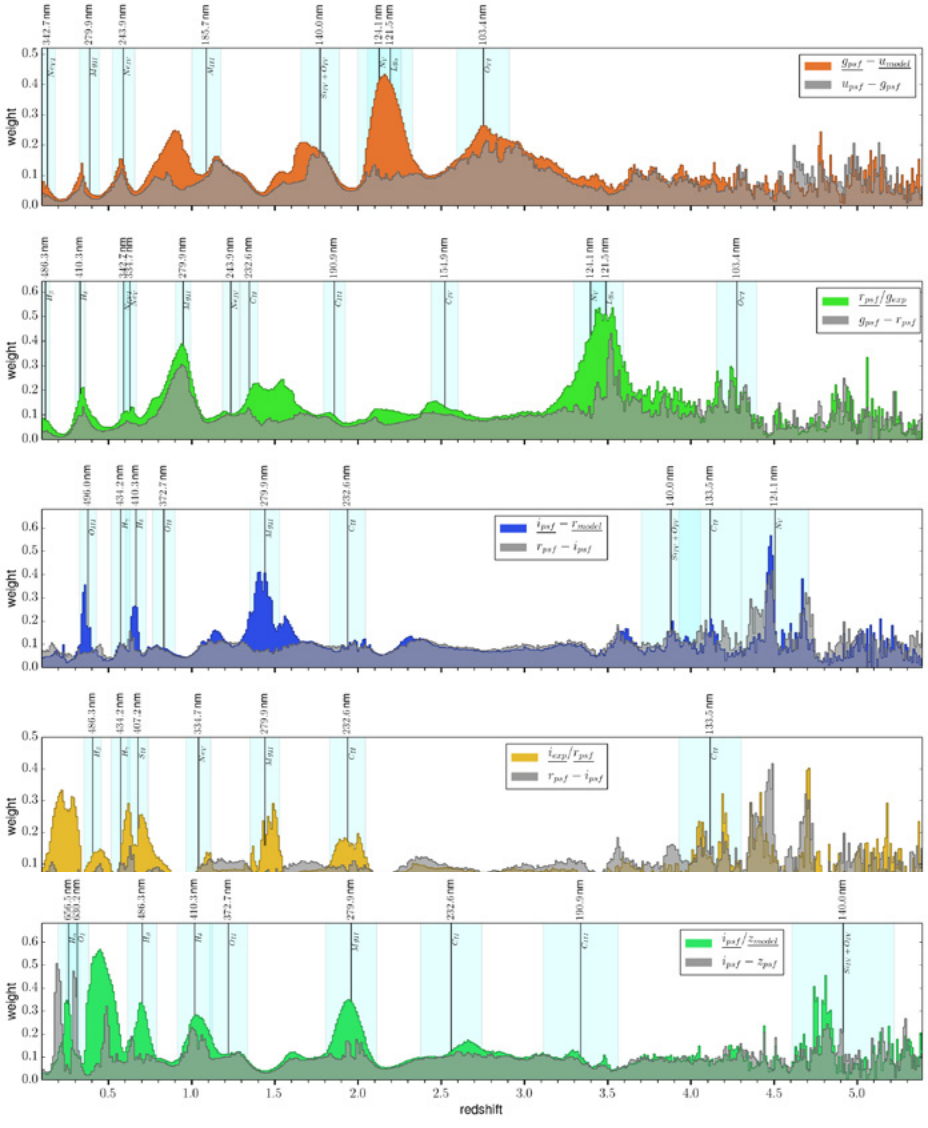




*Figure 2: Feature importance of some features from the obtained best set composed by magnitudes from neighboring bands. The importance of every feature was calculated for a sliding redshift bin of Δ z = 0.2. The results were compared with the classic features of the same bands. Based on the characteristics of the SDSS photometric filters system, the wavelengths indicating the beginning, center, and end of the overlapping regions were used to overplot the positions of particular quasar emission lines. The same operation was performed for the classic features (always shown in grey).*

## Mixed Variational Inference

In this line of research, we are concerned with obtaining the set of all plausible parameter values of a model (i.e., the set of possible solutions) that explain the observed dataset. This procedure contrasts with the more usual setting, in which the model is fitted to the data at hand and a single, optimal parameter vector (a point estimate) is obtained. The set of all plausible parameter values can be succinctly described using a probability density and is called a posterior density in this particular case. Using densities allows us to express our (subjective) uncertainty that governs the model parameters. Before observing the data, we may express our assumptions and beliefs regarding how the model parameters are distributed in terms of another density, called the prior density. Thus, given the observed data and our a priori beliefs as a prior density, we want to infer the posterior density. Inferring the posterior density is often a difficult task except in cases in which mathematically convenient functional forms are adopted for specifying the model (i.e., likelihood) and the prior density. Deviation from such convenience results in intractable calculations that call for approximations.

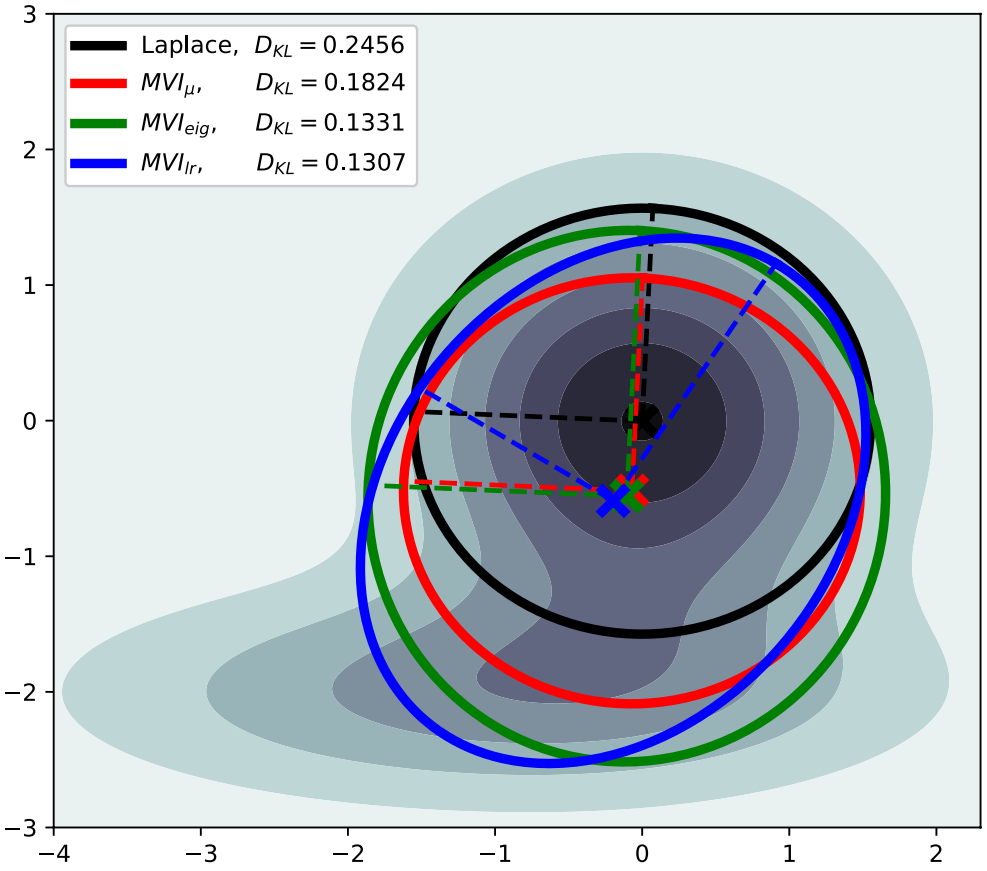



*Figure 3: Contour plot of a synthetic 2D posterior p(w). We plot the Gaussian Laplace posterior and the proposed MVI posteriors. The means of the Gaussian posteriors are shown with a cross. We also show an ellipse for the 70% confidence interval of each covariance matrix. The dashed lines are the axes of each ellipse (eigenvectors).*

An important tool for inferring the posterior density is the Laplace approximation, which has had a tremendous impact on advancing the use of Bayesian methodology in the field of machine learning. The Laplace approximation approximates the sought posterior density as a Gaussian density. However, in so doing, it does not necessarily take into account where the actual density lies and instead employs a quadratic approximation at the posterior mode (i.e., the most likely parameter value). Despite this weakness, the Laplace approximation enjoys widespread use and often produces good approximations in a variety of problem settings. Variational inference (VI) provides an alternative approximation that does take into account where the density lies by explicitly minimizing an objective (Kullback-Leiber divergence) that measures the "distance" of the approximation to the sought posterior density. However, such approaches also depend on the tractability of certain mathematical terms and are therefore not always applicable. As a response to this issue, the so-called stochastic VI approach has been put forward. Typically, stochastic VI approaches deliver an approximating posterior density that is factorized – that is, it discards correlations in the parameters. Stochastic VI typically adopts a factorized approximating density because it is economical in the number of free parameters and hence reduces the optimization effort.

In this work, we devised a method that builds on the Laplace approximation and variational inference, which we term "Mixed Variational Inference (MVI)." The method captures correlations in the posterior density, is applicable to certain non-tractable situations (non-conjugate models), and maintains a low number of free parameters. The main idea is to take the parameters of the Gaussian posterior-, mean-, and covariance matrix produced by the Laplace approximation and adapt them only partially. In this manner, we maintain some of the correlation structure given by Laplace but also adapt it by optimizing the same objective function that variational inference also uses.

In this work, we propose three different methods of partially adapting the Laplace posterior – that is, we produced three different types of approximations to the true posterior, which are illustrated in *Figure 3*. The figure displays a contour plot of a synthetic 2D posterior p(w), which we would like to approximate with a Gaussian. The legend reports the Kullback-Leibler divergence ($D_{KL}$) between each approximation and the true posterior. The Laplace approximation is plotted in black. The first proposed method (red) – $MVI_{\mu}$ – adapts only the mean and keeps the covariance matrix equal to that of Laplace. The second approximation (green) – $MVI_{eig}$ – adapts the mean and also allows for scaling the Laplace covariance matrix along its axes. The third approximation (blue) – $MVI_{lr}$ – adapts the mean and the Laplace covariance matrix with a low-rank update. We observe that Laplace (black), by design, places its mean on the mode of p(w). All other approximations place their means on alternative locations, albeit fairly close to one another. We note that even the modest flexibility of $MVI_{\mu}$ (red) – whose covariance is constrained to be equal to that of Laplace – can achieve a better $D_{KL}$ (lower is better). $MVI_{eig}$ (green) and $MVI_{lr}$ (blue) achieve an even lower $D_{KL}$ due to the partial flexibility allowed in adapting their covariance matrices.

Furthermore, we tested the proposed MVI in the setting of logistic and softmax regression. In numerical experiments, we found that the proposed scheme compares favorably with the Laplace approximation and VI and employs a factorized posterior on a number of benchmark datasets. Future work will focus on designing a more elaborate approximating posterior density capable of capturing higher statistical moments than the currently employed Gaussian density.

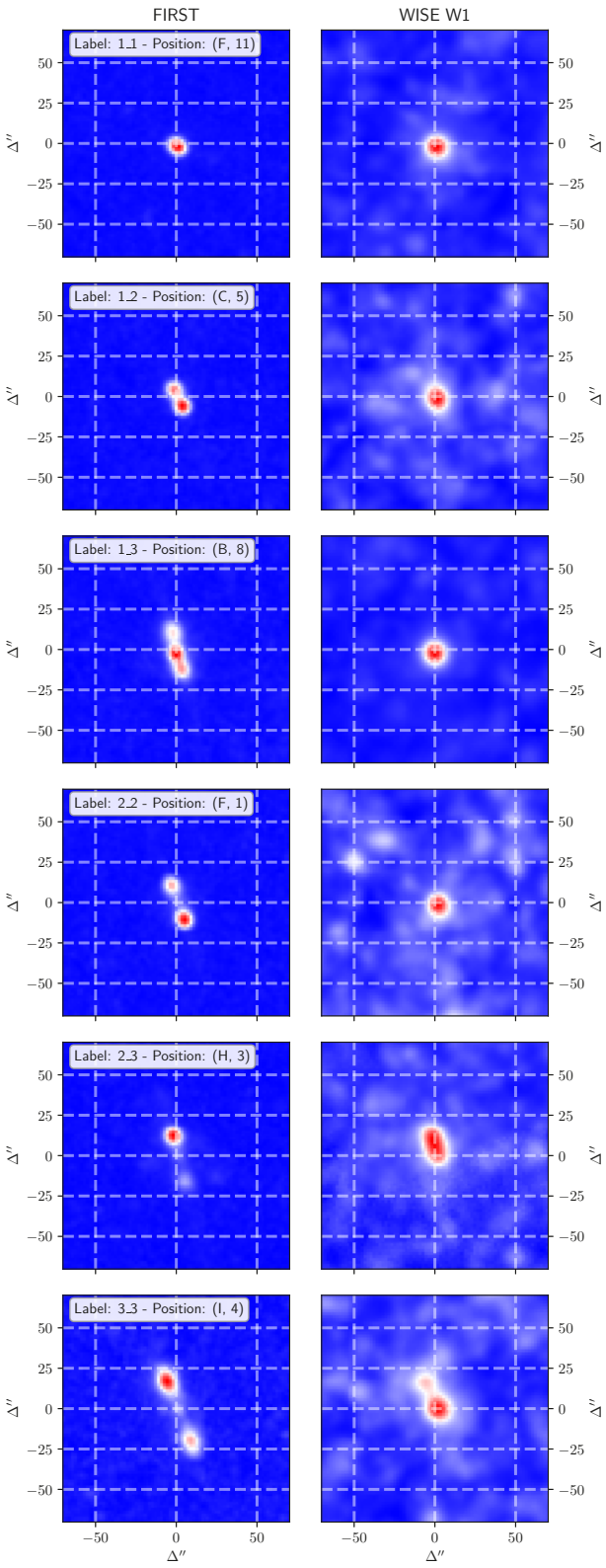
## Galaxy classification via dimensionality reduction

When galaxies emit radio emissions, this emission commonly arises from jets from the supermassive black hole (SMBH) in the center. However, based on the radio data alone, it is impossible to determine if two nearby sources are part of the same galaxy or if they just happen to appear close to each other in the sky. IR data are necessary to make this determination. In the IR, we can detect the center of the host galaxy. If there is an IR source between the two radio sources and not at the same point as the radio sources, they are more likely part of the same galaxy. We hope to achieve this classification by using a self-organizing map (SOM) code written by Kai Polsterer and Bernd Doser called Parallelized rotation and flipping INvariant Kohonen maps (PINK).

A SOM is a dimensionality reduction technique that reduces the high-dimensional data to a set of two-dimensional prototypes. Determining which prototype is most similar to an image enables us to learn about this image without needing to analyze it independently. PINK, in particular, has the capability of learning these prototypes in a rotation-invariant fashion (i.e., the rotation of the images is immaterial).

We collaborated with Tim Galvin at CSIRO on the use of PINK with data from the Radio GalaxyZoo (RGZ) to combine radio- and IR data. RGZ is a citizen-scientist project in which volunteers visually inspect galaxies and click on the interesting components. We used the labels obtained from RGZ and introduced a new method of transferring the label information. The labels obtained from RGZ divide the radio objects into classes based on the number of peaks and components that volunteers are able to identify. Finally, we compared images with the prototypes in the SOM and were able to correctly identify the number of components and peaks with 85.7% and 80.7% accuracy, respectively.

We also ran a similar experiment on the IR images in which we were able to find a correlation between what was observed in the radio and IR, as expected. This approach is important in astronomy because it allows us to identify whether two nearby radio sources are part of the same galaxy or whether they are unrelated and are simply visually close to each other (see *Figure 4*).



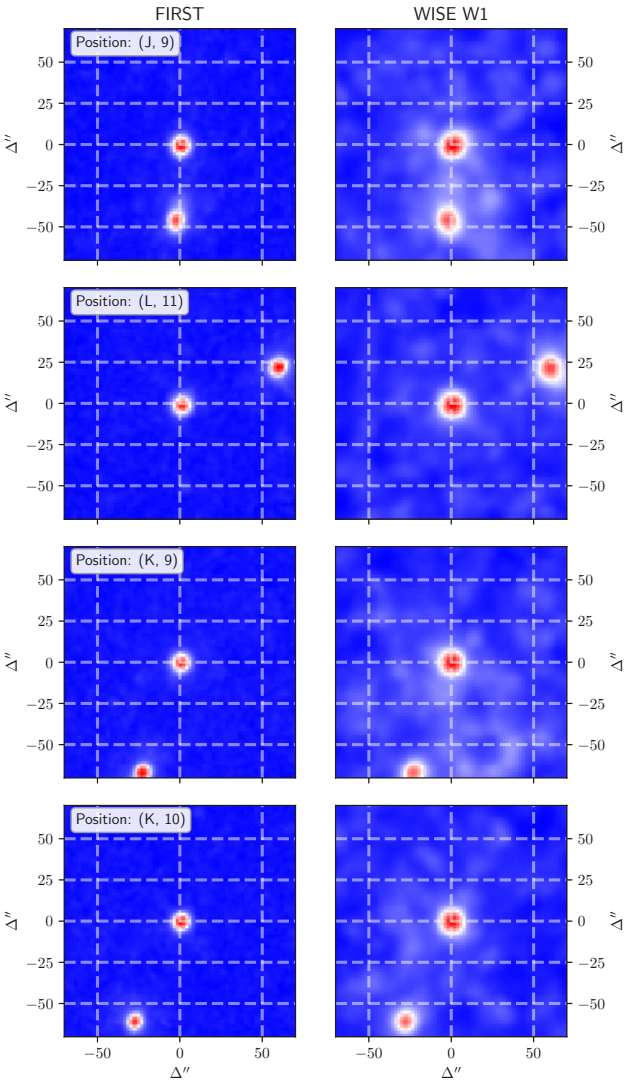
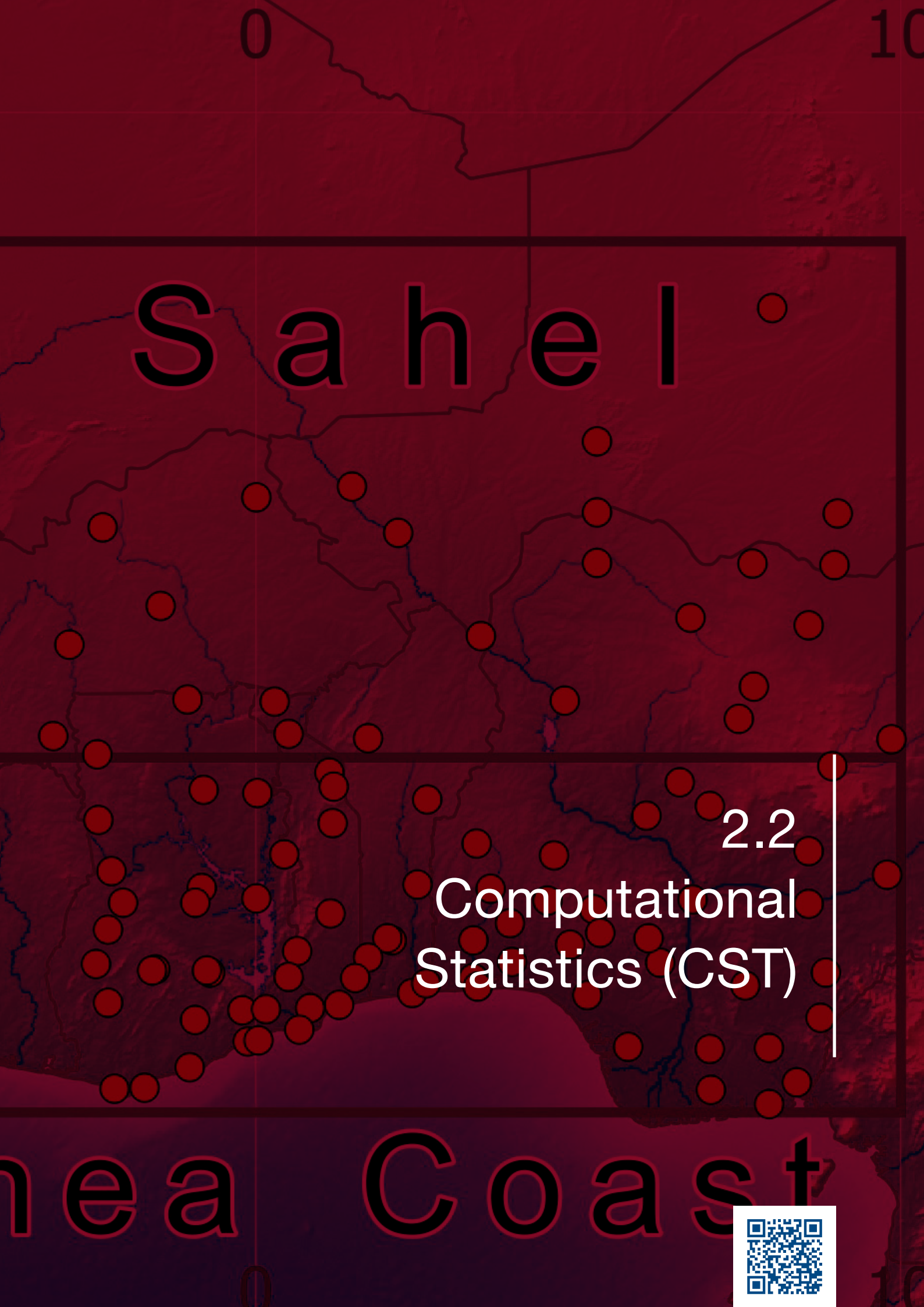




*Figure 4: Both figures display examples of SOM neurons. The larger figure shows examples in which the IR source is in between the two radio sources, meaning they are all likely part of the same galaxy. The smaller figure shows examples of IR sources aligning with radio sources, indicating they are likely unrelated galaxies.*

# 2 Research


Sahel
2.2
Computational
Statistics (CST)

## 2.2
## 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.

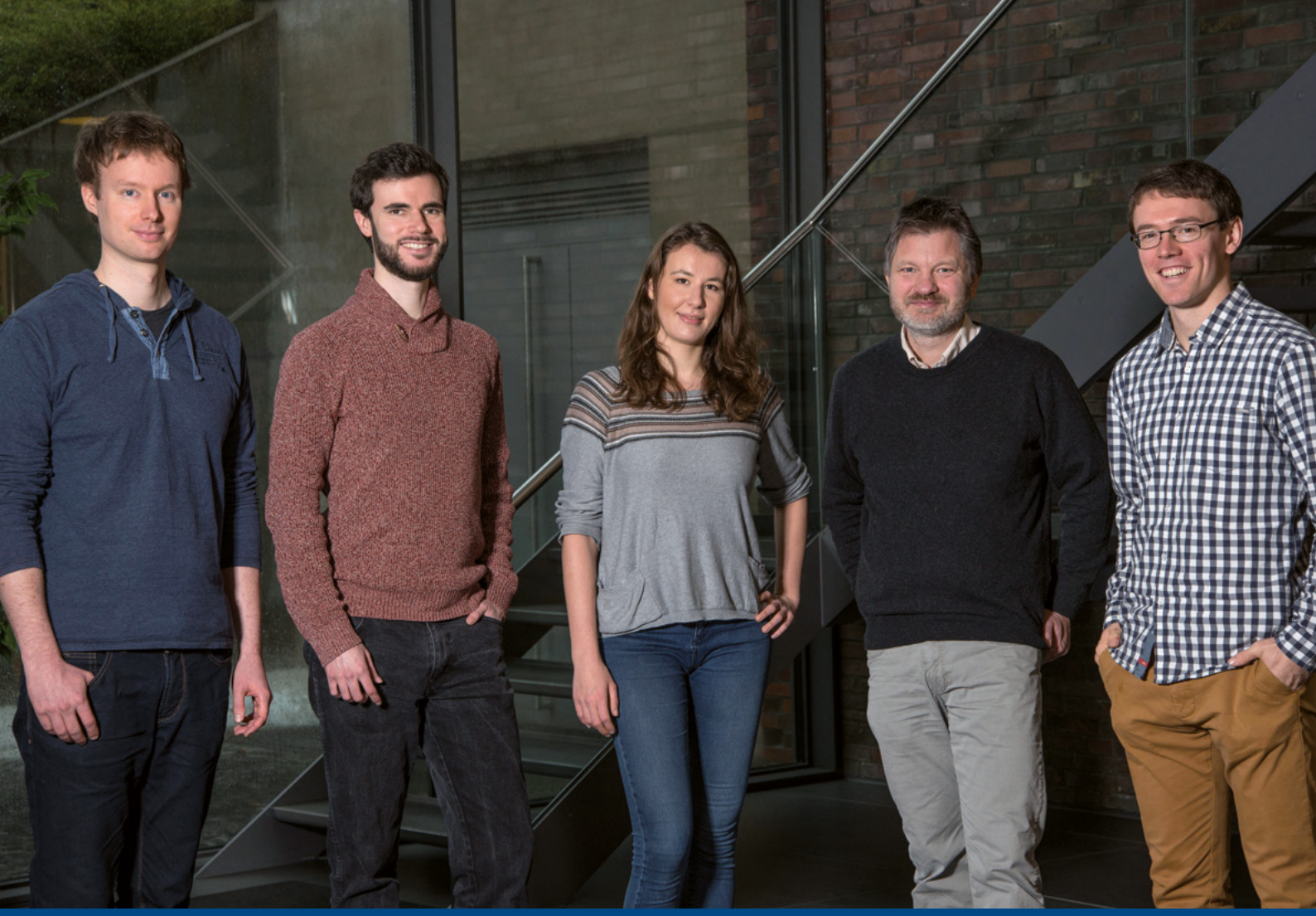
As the future is uncertain, forecasts should be probabilistic in nature, meaning they should take the form of probability distributions over future quantities or events. Accordingly, we are currently witnessing a trans-disciplinary shift of paradigms from deterministic or point forecasts to probabilistic forecasts. The CST group seeks to provide guidance and leadership in this transition by developing both the theoretical foundations for the science of forecasting and cutting-edge statistical methodology, notably in connection with applications.

Weather forecasting represents a prime example of our work. In this context, the group maintains research contacts and collaborative relations with meteorologists at KIT and at 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 Gruppe liegt in der Theorie und Praxis der Vorhersage.*

*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 in der Wettervorhersage, zum Einsatz bringt.*

*In diesem Zusammenhang pflegen wir Kontakte und Kooperationen mit Meteorolog(inn)en am KIT und am Europäischen Zentrum für mittelfristige Wettervorhersage.*



## Group Leader

Prof. Dr. Tilmann Gneiting

## Staff Members

Dr. Werner Ehm *(until February 2018)*
Kira Feldmann
Alexander Jordan *(until February 2018)*
Sebastian Lerch
Johannes Resin *(since June 2018)*

## Scholarship Holder

Patrick Schmidt *(HITS Scholarship)*

## Visiting Scientists

Sándor Baran *(July 2018)*
Peter Vogel

## General news

The focus of our research remains on the theory and practice of forecasting, with external funding support provided by the European Research Council via the Advanced Grant ScienceFore (which expired in 2018) and by the German Research Foundation within the collaborative research center "Waves to Weather." Our interdisciplinary collaborations include joint work 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, United Kingdom.

In continuation of a precious tradition, an integral and very enjoyable aspect of our work is the intense disciplinary and interdisciplinary scientific exchange in which we take part on many occasions. A highlight in 2018 was the mini-symposium "Proper scoring rules and consistent scoring functions" held on May 16 on the HITS premises, which drew an international audience, with colleagues joining from as far as New Zealand. *Figure 5* shows the participants – who discussed the latest advances in theory and methodology for forecast evaluation – at the main entrance to the HITS building. We were again happy to welcome guests from all over the world throughout the entire year.

*Figure 5: Participants in the mini-symposium "Proper scoring rules and consistent scoring functions" held on May 16, 2018.*

The next three sections describe facets of our research that have been developed within the "Waves to Weather" consortium, which addresses fundamental questions concerning the predictability of weather. We begin by reviewing the role of ensemble forecasts in numerical weather prediction and elucidating the need for statistical postprocessing. Afterward, we report on joint research with meteorologists at KIT, where we seek to improve weather forecasts across Africa and report humbling results that are generating new challenges of scientific importance and societal relevance. In the final section, we describe novel ways in which the use of neural networks improves methods for the postprocessing of weather forecasts. Perhaps surprisingly, these methods relate to techniques developed in the Astroinformatics (AIN)group (*see Chapter 2.1*), which serves to illustrate the fruitful interdisciplinary research environment at HITS.

### STATISTICAL POSTPROCESSING OF ENSEMBLE WEATHER FORECASTS

In order to provide weather forecast information, weather centers worldwide draw on highly sophisticated numerical models that are run in real time on supercomputers and produce point forecasts of future atmospheric states. In a strong move toward probabilistic forecasts, these efforts have been transformed through the operational implementation of so-called ensemble systems since the 1990s. An ensemble forecast consists of multiple, simultaneous runs – typically between 10 and 50 – of numerical weather prediction (NWP) models, which differ from one another in terms of the two major sources of uncertainty, namely the initial state of the atmosphere and the mathematical representation of the respective physical processes.

Ensemble systems are now run at all major weather centers, leading to the creation of The International Grand Global Ensemble (TIGGE) multi-model ensemble database. TIGGE contains forecasts from up to ten ensemble systems worldwide, the ensemble of the European Centre for Medium-Range Weather Forecasts (ECMWF) being the most skillful and the most widely used.

Despite their undisputed successes, ensemble systems continue to exhibit systematic errors, such as biases (predictions are systematically too high or too low) and dispersion errors (the output is systematically either too spread out, or not enough). It is therefore common practice to statistically postprocess the NWP output. State-of-the-art methods for this postprocessing include the distributional 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 family, where the parameters depend on the ensemble forecast at hand in physically meaningful ways. In contrast, the BMA method uses a mixture distribution in which the mixture components depend on the ensemble member forecasts in suitable ways, with the mixture weights reflecting the corresponding member's skill. The choice of the parametric family used in the EMOS approach and of the mixture components employed by the BMA technique depends on the weather variable of interest.

The simplest case arises in the EMOS model for temperature, which uses the Gaussian family. In contrast, EMOS and BMA models for nonnegative weather quantities – such as precipitation accumulation – require more complex formulations involving the use of gamma and generalized extreme value (GEV) distributions, among others. The respective statistical parameters need to be estimated from training data based on forecast-observation pairs from the recent past.

## BETTER WEATHER FORECASTS FOR AFRICA

When will it rain, where, and how much? Reliable weather forecasts are taken for granted in developed countries. However, the state of affairs is much worse for many parts of Africa. In a recently published joint study with meteorologists at KIT *[Vogel et al., 2018]*, we demonstrated that progress made in weather forecasting in Europe and North America does not directly translate to Africa. Moreover, we revealed what science can do to tackle this problem. Accumulated precipitation forecasts are of great socioeconomic importance for agriculturally dominated societies in northern tropical Africa, such as the West Sahel, East Sahel, and Guinea Coast regions, as illustrated in *Figure 6*.

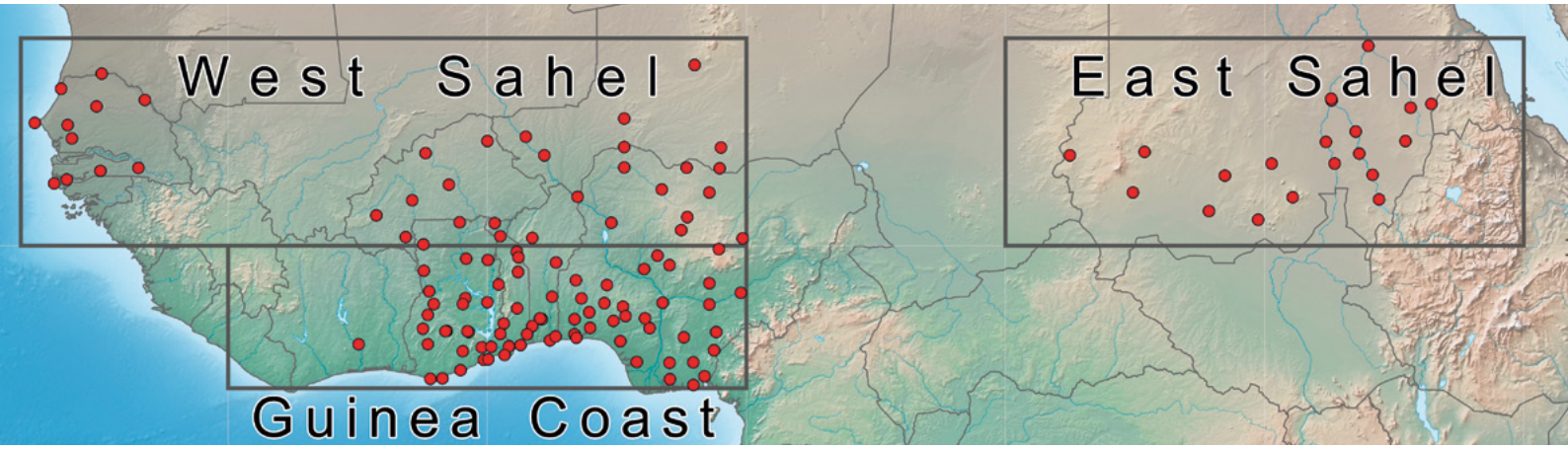



*Figure 6: Geographical overview of the study domain, with the locations of the observation stations (dots) within the three considered regions in northern central Africa.*

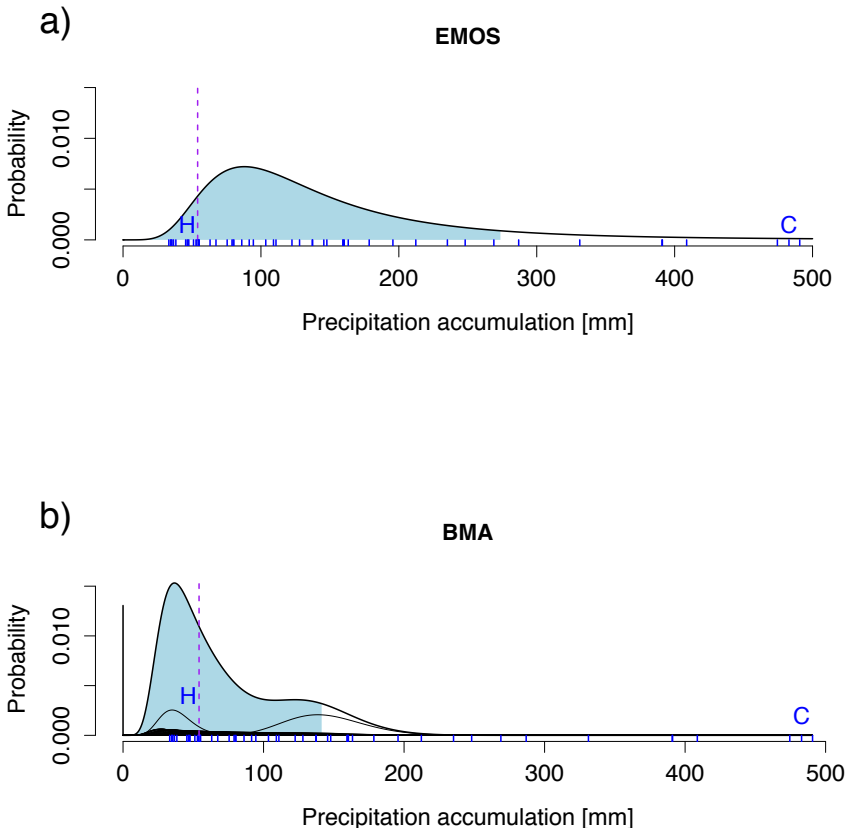



*Figure 7: Postprocessed ensemble forecasts for 5-day accumulated precipitation at Ouagadougou, Burkina Faso, valid 3–8 August 2007. The blue marks at the bottom represent the raw ECMWF ensemble members, including the high-resolution (H) and control (C) runs. (a) The EMOS postprocessed forecast distribution includes a tiny point mass at zero and a censored GEV density for positive accumulations. (b) The BMA postprocessed forecast includes a point mass at zero, which is represented by the solid bar, and a mixture of power-transformed gamma densities for positive accumulations. The component densities are represented by the thin black curves. The lower 90% prediction interval is indicated in light blue, and the dashed bar represents the verifying precipitation accumulation.*

In our study, we analyzed the performance of NWP ensemble forecasts from the TIGGE database relative to simplistic, climatology-based forecasts. To assess the full potential of the ensemble forecasts, we applied state-of-the-art statistical postprocessing methods in the form of BMA and EMOS. For illustration, Panel a) in *Figure 7* displays an EMOS postprocessed forecast distribution for 5-day accumulated precipitation at Ouagadougou, Burkina Faso. The members of the raw ECMWF ensemble include several values in excess of 200 mm, the highest member being close to 500 mm. These members inform the statistical parameters of the EMOS postprocessed forecast distribution, which includes a tiny point mass at zero and a GEV density for positive precipitation accumulations, with the 90th percentile being at 174 mm. Panel b) shows the respective BMA postprocessed forecast distribution, which involves a point mass of about 0.01 at zero and a mixture of power-transformed gamma densities for positive accumulations, with the 90th percentile being at 141 mm. In this example, the BMA and EMOS postprocessed distributions are sharper than the raw ECMWF ensemble, yet the verifying accumulation is nevertheless well captured.

The goal of the study was to provide an exhaustive assessment of our current ability to predict rainfall over northern tropical Africa by considering the skill of raw and postprocessed forecasts from TIGGE. We examined accumulation periods of 1 to 5 days for monsoon seasons from 2007 to 2014 and verified the forecasts against rainfall observations from the Karlsruhe African Surface Station Database (KASS-D) as well as satellite-based, gridded products. As a benchmark, we introduced the concept of a probabilistic climatology that consists of the observations during the 30 years prior to the considered year at the considered day of the year and location. This method represents the static, climatological distribution of rainfall at a given location and date but does not incorporate dynamic information about the state of the atmosphere. We extended the probabilistic climatology by including observations in a 2-day window around the considered day and refer to this technique as Extended Probabilistic Climatology (EPC).

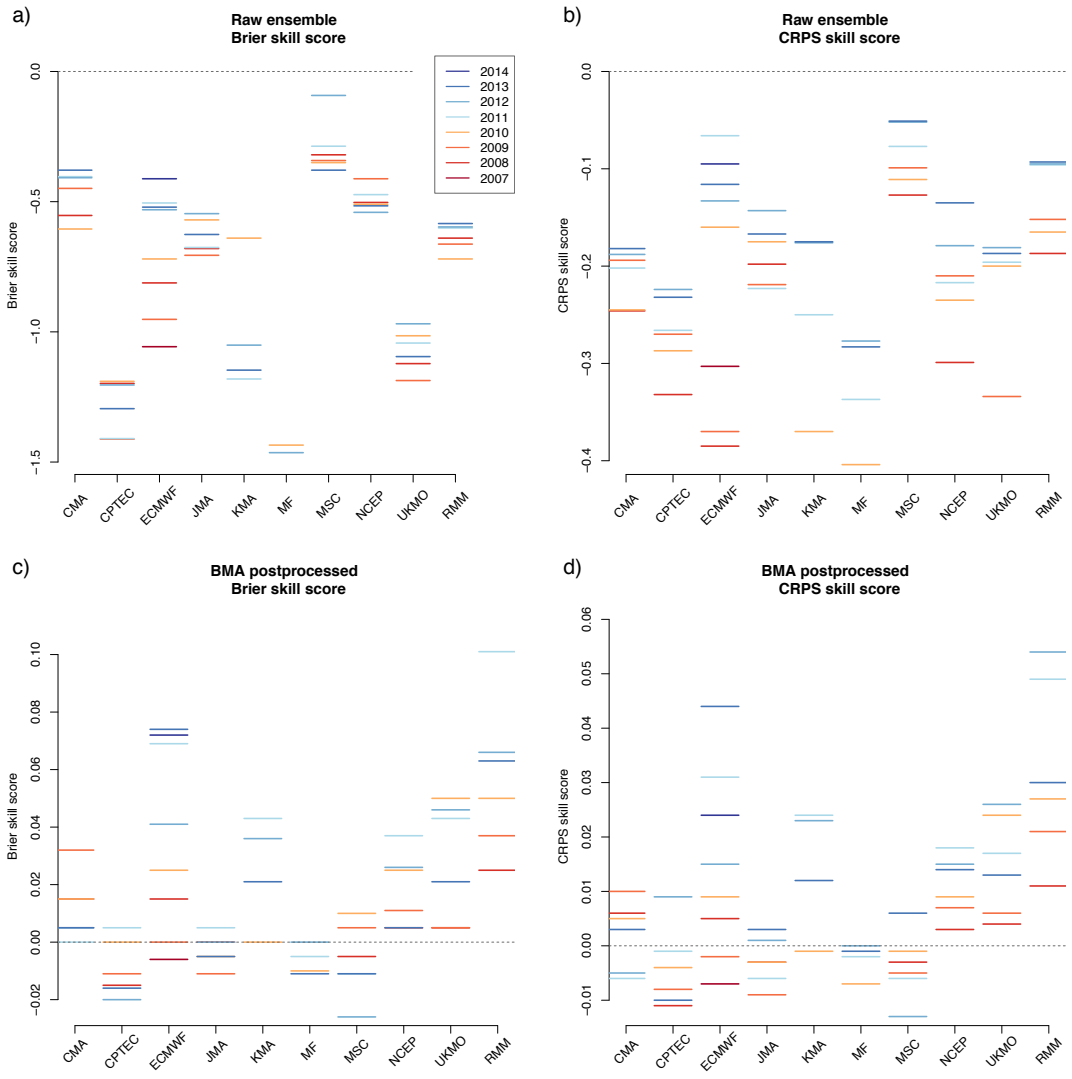



*Figure 8: Brier and CRPS skill scores for raw and BMA postprocessed TIGGE sub-ensemble forecasts of one-day accumulated precipitation over West Sahel in monsoon seasons from 2007 to 2014, verified against station observations. Skill equal to EPC is indicated by the dashed line.*

The ensemble forecasts from the TIGGE database were clearly expected to outperform EPC by a large margin. In putting this expectation to the test, *Figure 8* displays Brier and CRPS skill scores relative to EPC for raw and BMA postprocessed TIGGE sub-ensemble forecasts from 2007 to 2014, verified against station observations in the West Sahel region. The skill scores are listed in terms of widely used, theoretically principled loss functions, namely the Brier score and the continuous ranked probability score (CRPS). A negative (positive) skill score corresponds to a forecast that is inferior (superior) to EPC. Surprisingly, the raw ensemble forecast underperforms relative to EPC, a striking finding that remains true independently of region, accumulation time, monsoon season, and ensemble. BMA and EMOS postprocessed forecasts strongly improve on the raw ensembles but – rather disappointingly – typically do not outperform EPC. Overall, raw ensemble forecasts are of no use for the prediction of precipitation over northern tropical Africa, and even EMOS and BMA postprocessed forecasts have little added value compared with EPC.

The fact that EPS precipitation forecasts are so poor over northern tropical Africa is a strong demonstration of both the complexity of the underlying forecast problem and the need for vigorous scientific development. Given the growing socio-economic impact of rainfall in northern tropical Africa – with its rain-fed agriculture – statistical and statistical-dynamical approaches should be fostered in parallel in order to improve the predictability of rainfall in this region.

## USING NEURAL NETWORKS TO POSTPROCESS ENSEMBLE FORECASTS

As noted, distributional regression or ensemble model output statistics (EMOS) approaches model the conditional probability distribution of future weather quantities in terms of parametric forecast distributions, such as Gaussian distributions. To this end, the parameters of the forecast distribution are connected to summary statistics from the NWP ensemble forecast through suitable link functions. The use of additional predictor variables, such as forecasts of related albeit distinct weather quantities, can improve the predictive performance. However, the specification of suitable link functions for arbitrary predictor variables is not straightforward, and elaborate techniques are required to avoid overfitting.

In a recent paper *[Rasp and Lerch, 2018]*, we proposed a new approach based on modern machine learning techniques. Motivated by discussions with Antonio D'Isanto and Kai Polsterer (AIN) and by adopting machine-learning approaches to astrophysical image analysis developed by the AIN group, we demonstrated how neural networks can be employed successfully in the postprocessing of ensemble weather forecasts. The network architecture allows for incorporating various features that are relevant for correcting systematic deficiencies of ensemble forecasts as well as for estimating network parameters by optimizing the aforementioned CRPS.

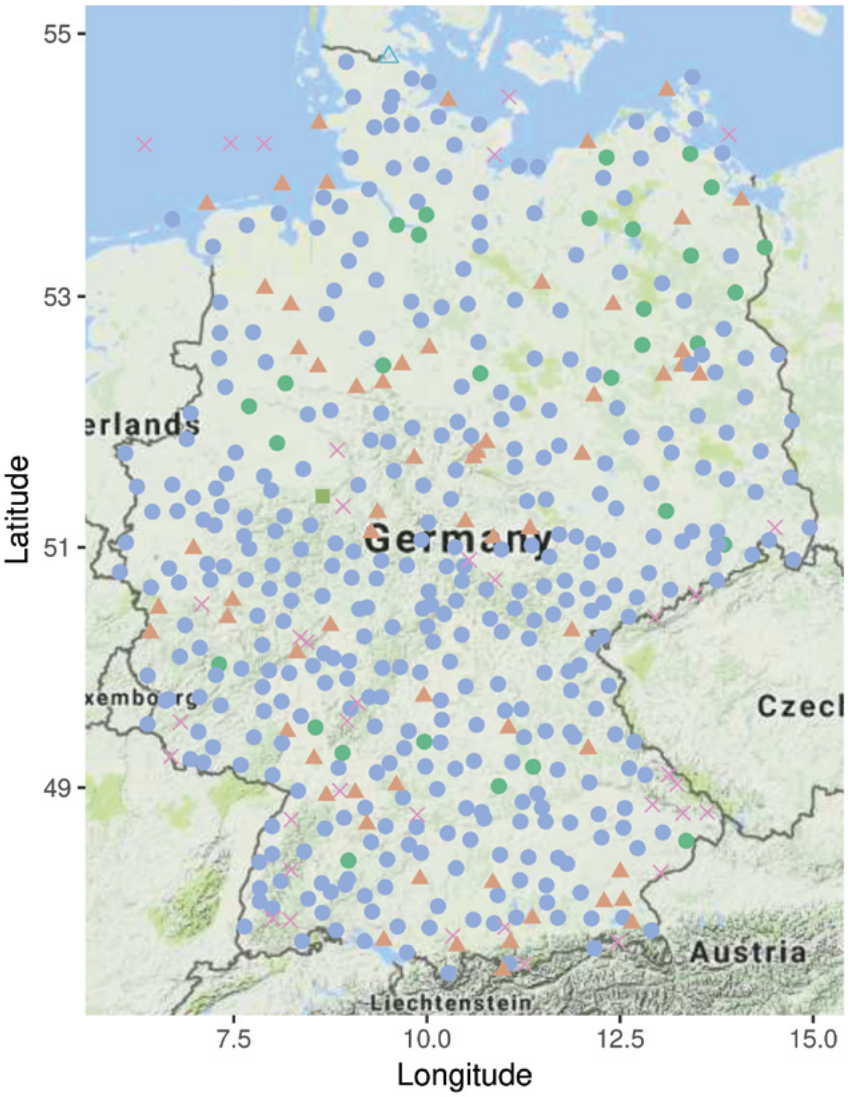



*Figure 9: Locations of observation stations for surface temperature, color coded by the best performing approach in terms of the CRPS in 2016. Neural network-based approaches are indicated by circles, benchmarks by triangles and crosses.*

Specifically, the output layer of the neural network represents the mean and variance parameter of the Gaussian forecast distribution. The corresponding probabilistic forecast describes the conditional distribution of surface temperature in terms of input features – namely summary statistics from the NWP ensemble forecast – and characteristics of observation stations. The use of a hidden layer allows the network model to learn arbitrary nonlinear relations between predictor variables and distribution parameters in a data-driven manner without requiring the prescription of a specific link function. To enable

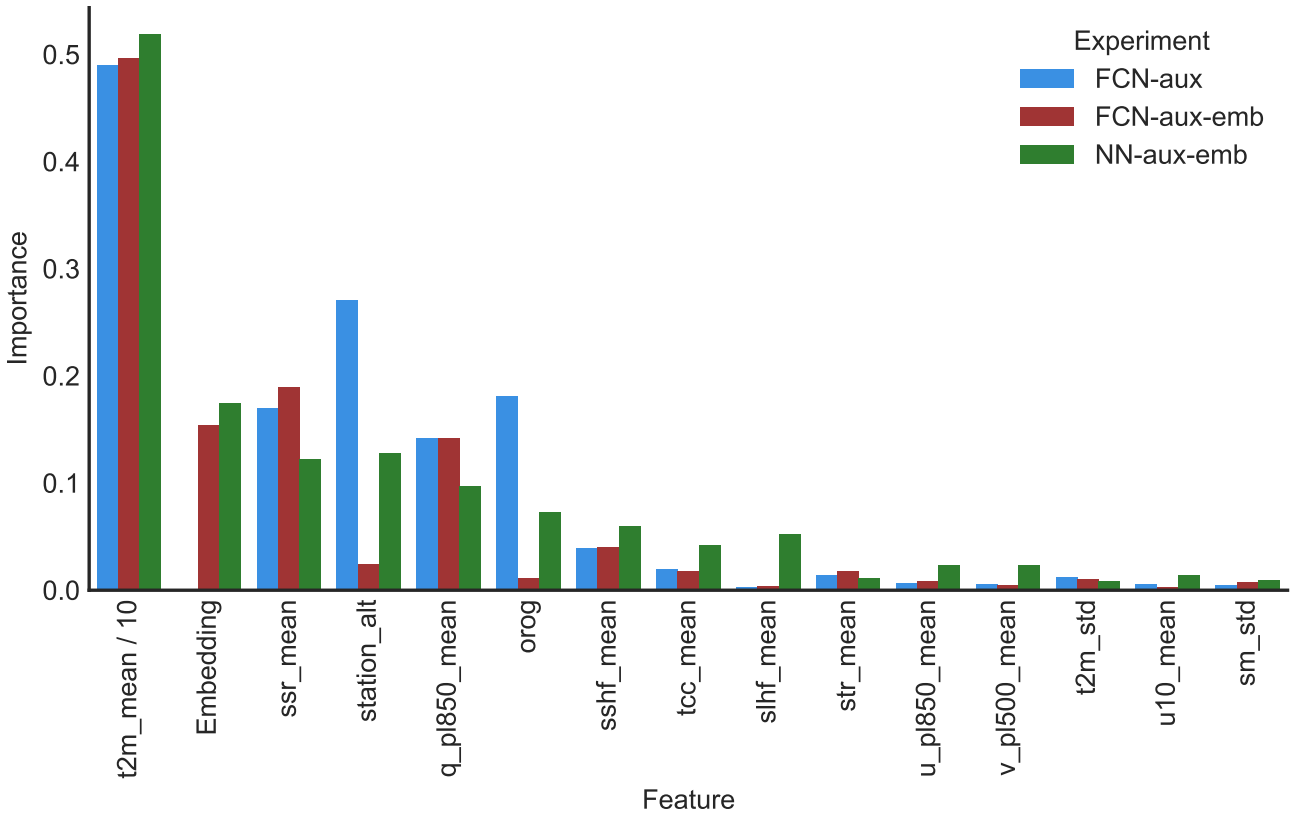


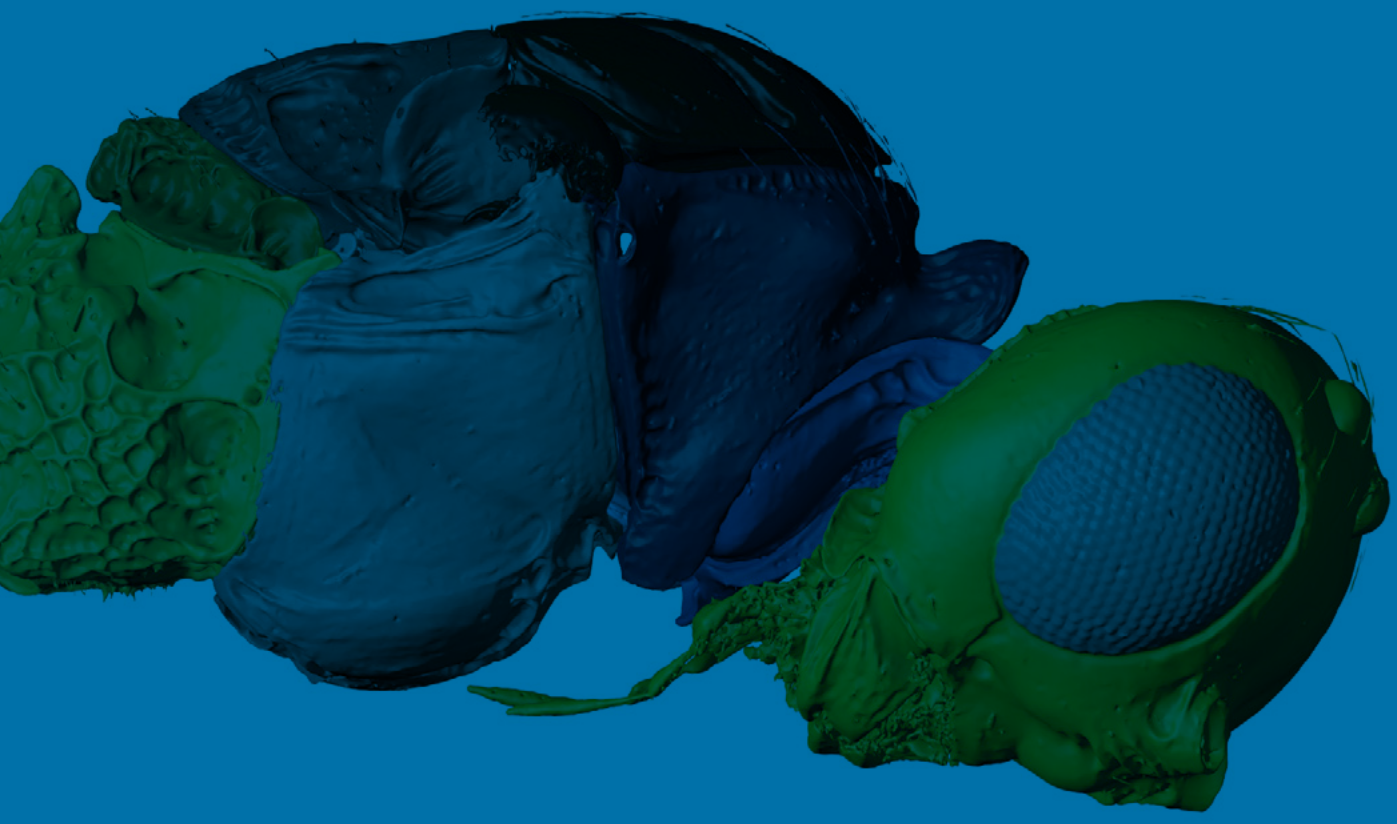

*Figure 10: Feature importance for the 15 leading predictor variables in three neural network architectures.*

the learning of station-specific information, we used embeddings, a common technique in natural-language processing and recommender systems. This method allows for estimating a single model based on data from all stations but is made locally adaptive by encoding station-specific information.

In a case study on ECMWF ensemble forecasts of surface temperature over Germany from 2007 to 2016, we compared the novel neural network-based approach with state-of-the-art techniques. While computationally much less demanding, the neural network-based approach provided the best forecasts at about 80% of the observation stations, as illustrated in *Figure 9*.

The trained machine learning model can be used to gain meteorological insight by identifying the most relevant predictor variables, thereby challenging the common view of neural networks as pure black boxes. For illustration, *Figure 10* displays the 15 most important predictors (or features) for three different network models, with the importance of a predictor variable measured in terms of the deterioration of the predictive ability when its values are subject to random permutations. The mean ensemble forecast for temperature (t2m_mean) is the most important predictor throughout, followed by features that account for local characteristics, such as embeddings, station altitude, and orography.

# 2 Research



# 2.3 Data Mining and Uncertainty Quantification (DMQ)

## 2.3
## 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 collaboration with the Engineering Mathematics and Computing Lab (EMCL) at the Interdisciplinary Center for Scientific Computing (IWR) at Heidelberg University, which is also headed by Vincent Heuveline.

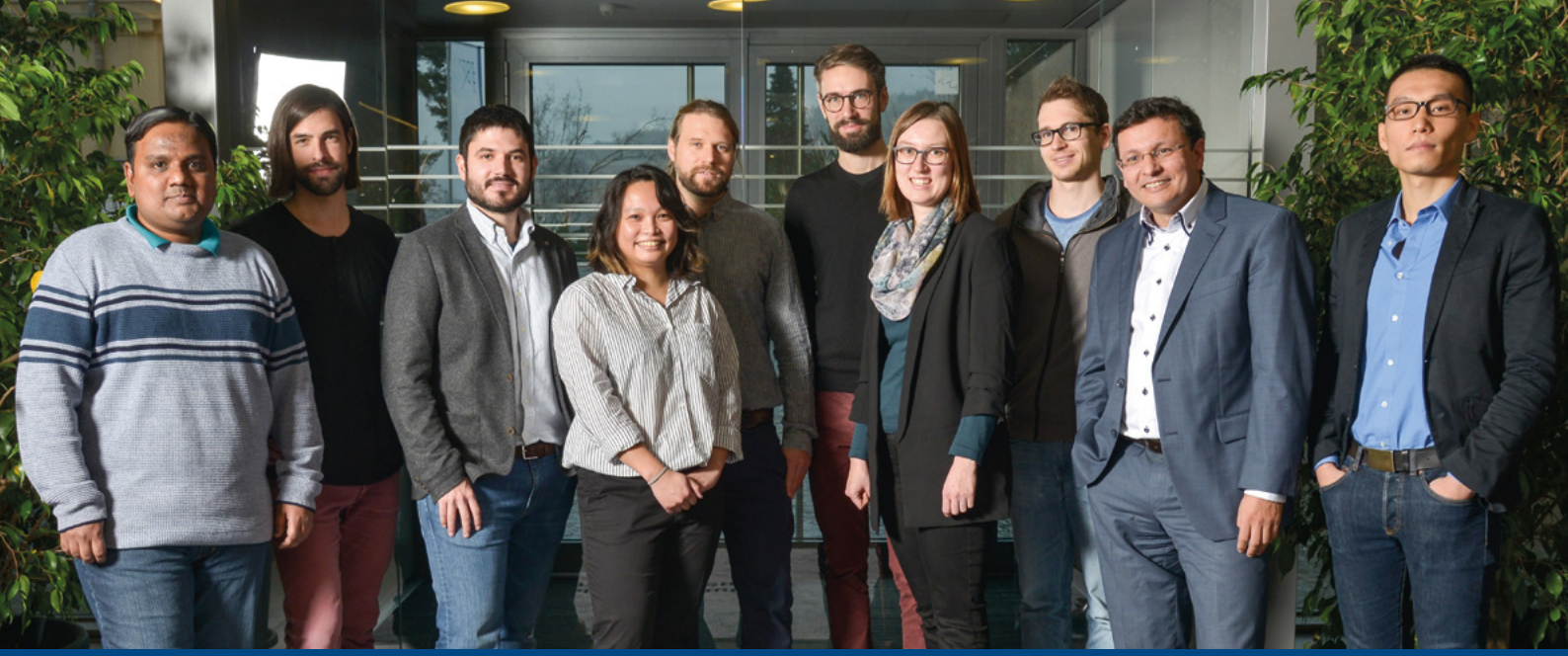
DMQ's research focus lies in gaining knowledge from extremely large and complex datasets through data-mining technologies. Reliability considerations with respect to these datasets are addressed via uncertainty quantification methods. Both fields – data mining and uncertainty quantification – require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

In 2018, DMQ focused on research activities in the following areas: Uncertainty quantification for medical engineering, multi-physics flow simulations and 3D image segmentation.

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

*2018 wurde dazu in der Gruppe in folgenden Anwendungsbereichen gearbeitet: Uncertainty Quantification für medizinische Anwendungen, Multiphysik-Strömungssimulationen und 3D Bildsegmentierung.*



## Group Leader

Prof. Dr. Vincent Heuveline

## Scholarship Holders

Sotirios Nikas *(until October 2018)*
Vijayasarathi Janardhanam *(since November 2018)*
Alejandra Jayme *(since December 2018)*

## Staff Members

Dr. Chen Song
Philipp Gerstner

## Visiting Scientists

Prof. Dr. Jiawei Zhang *(July 2018)*
Wei Zhang *(until February 2018)*

## Uncertainty quantification for a blood pump device

Heart failure (HF) – also known as congestive heart failure – is a cardiovascular disease in which the heart can no longer maintain the blood flow that the circulatory system in the human body needs. More than 40 million people worldwide suffer from heart failure, which causes more than 17.5 million deaths annually.

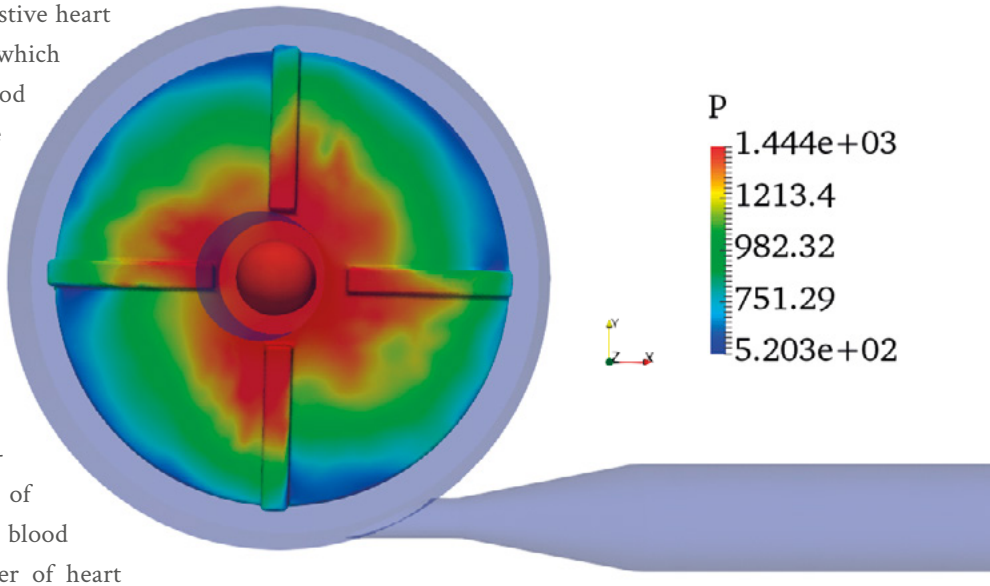



*Figure 11: Mean value of pressure in blood pump.*

Implantable cardiac assist systems – blood pumps – are an effective temporary solution for patients with heart failure. Due to the increasing number of heart-failure patients, the number of blood pump implants has exceeded the number of heart transplants in most developed countries. However, the mortality rate of individuals with ventricular assist devices (VAD) remains high (i.e., 20% after the first year), and improving the reliability of such blood pump devices will remain important in the future.

In our work, we incorporate the variational multiscale formulation to cope with turbulent flow under the finite element method (FEM) framework. We do not only build a numerical model for the blood pump device, but also include the uncertainty analysis in the modeling. We consider three different uncertain sources: input boundary condition, dynamic viscosity, and angular speed. We therefore also developed a sophisticated preconditioner – Polynomial Chaos Expansion multilevel preconditioner – in order to overcome this issue. The simulation is achieved via our open-source Finite Element library HiFlow[3]. To be able to efficiently deal with the coupled stochastic system, we additionally developed a multilevel preconditioner that is suitable for high-performance computing (HPC). Therefore, we can provide a reliability assessment – in terms of a probability – of the blood pump device based on several important quantities of interest, e.g. the velocity pressure field, pressure and force loadings, hemolysis.

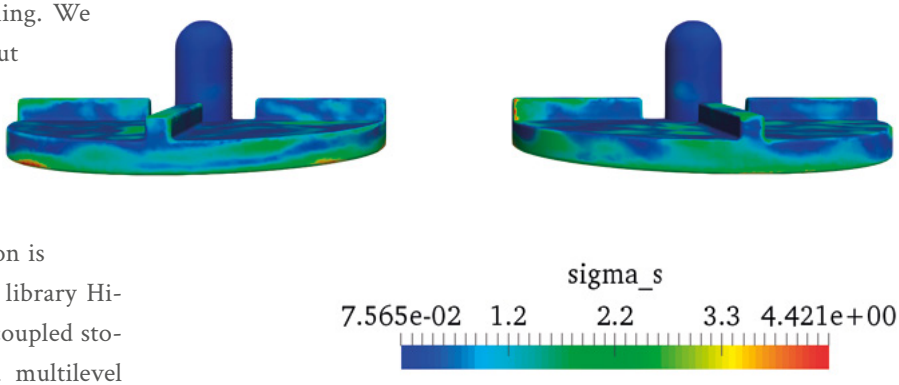



*Figure 12: Standard deviation of pressure in blood pump.*

## The influence of uncertainties in the biomechanical simulation of the aorta

In recent decades, biomedical studies with living subjects (in vivo) and artificial experiments (in vitro) have been supplemented more and more by calculation and simulation *(in silico). In silico* techniques for medical technology can provide information on the diagnosis and risk stratification of cardiovascular disease, which is one of the leading causes of death in industrialized countries. Further application examples for *in silico* methods include the virtual prototyping and simulation of possible operation results. High reliability is a prerequisite for cardiovascular diagnostics and risk stratification methods, especially in surgical decisions. Given the uncertainties in the input data of a simulation, this means that the uncertainties in the simulation results must be quantified.

For aneurysms of the aorta, the conditions for the dilation of the vessel and for an increased risk of rupture are a common research topic. 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 may be a factor in the development of the aneurysm. However, the stress load is not directly measurable, and the modeling of the aortic vessel wall and blood flow dynamics leads to complex problems of fluid-structure interaction (FSI). In addition, the model parameters may be highly uncertain for a particular patient.

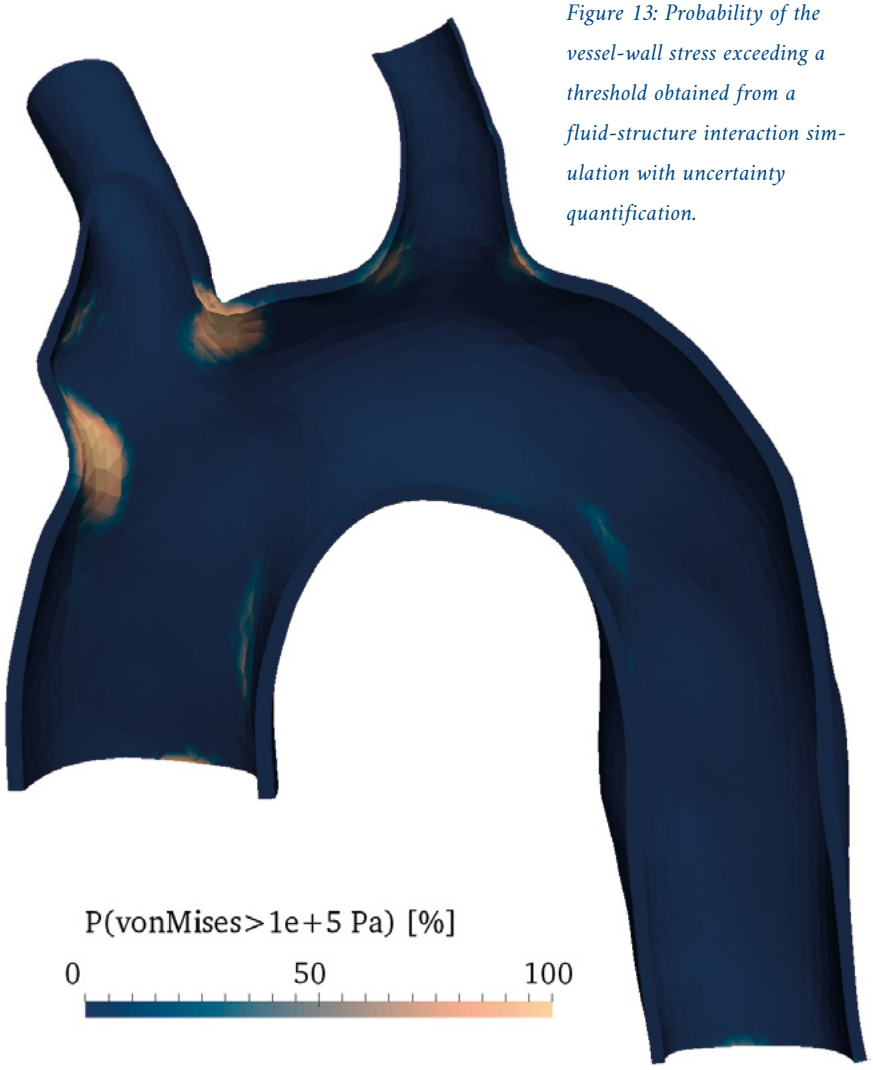



*Figure 13: Probability of the vessel-wall stress exceeding a threshold obtained from a fluid-structure interaction simulation with uncertainty quantification.*

In our work, we develop a numerical framework to calculate the likelihood that a threshold for the stress load in a vessel's wall will be exceeded. This framework is based on an FSI model of a blood vessel. The model is patient-specifically configured by 4D Flow Magnetic Resonance Imaging (MRI). The spread of uncertainties is based on a stochastic collocation method. Each collocation point is indicated by a finite element simulation with HiFlow³, which makes it possible to visualize the probability that a critical threshold of the load burden in the aortic vessel wall is exceeded above the vessel wall geometry (see *Figure 13*). The results open up new perspectives for clinical trials.

## Uncertainty quantification for multi-physic flow simulation

Containment in engineering often provides thermal insulation (e.g. heat exchanger systems). An improvement of heat transfer via efficient enhancement is of general interest due both to its benefits of low operational costs as well as its sustainable energy usage. One possible heat-transfer enhancement technique is given by applying electric fields, which is known as thermal electro-hydrodynamic- (TEHD) driven heat-transfer augmentation. In collaboration with our project partners at BTU Cottbus, we aim to gain further insight into the hydrodynamic behavior of dielectric fluids in a cylindrical enclosure with an applied electric field and temperature gradient by combining numerical simulations with experimental data. In particular, 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 could derive a priori error estimates that reveal 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 experiments and numerical simulations. Due to these vortices, radial heat transfer is enhanced, as shown by an increase of the corresponding Nusselt numbers.

*Figure 14* illustrates the considered geometry, in 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.

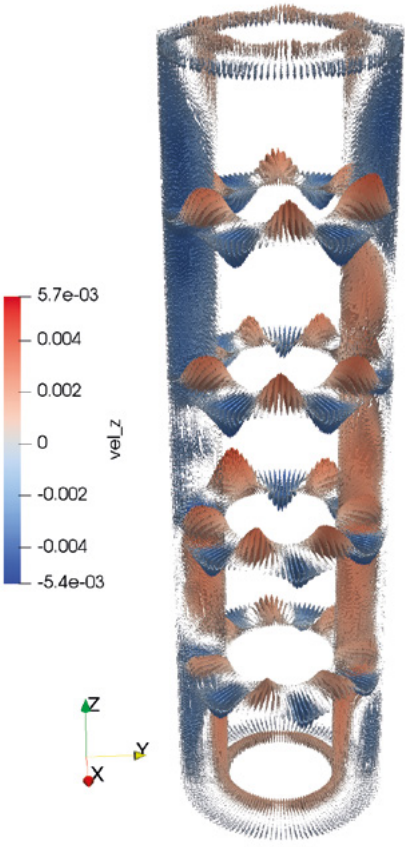



*Figure 14: Fluid velocity field in vertical cylinder annulus, under influence of temperature and pressure gradient between inner and outer wall.*

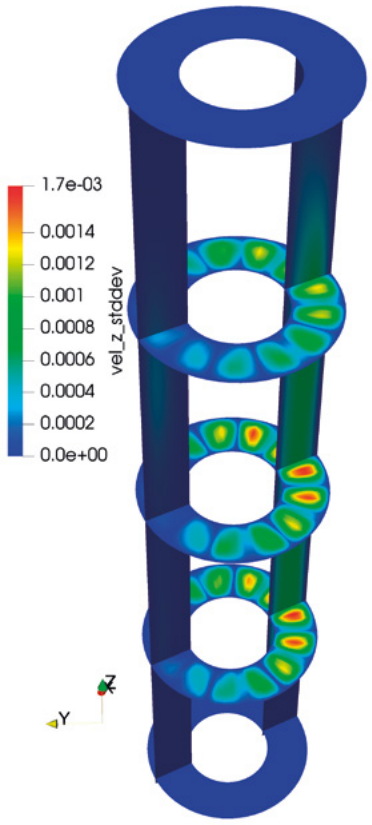



*Figure 15: Standard deviation of fluid velocity field in vertical cylinder annulus, under influence of uncertain temperature difference between inner and outer wall.*

Moreover, since the number of vortices – and therefore also the hydrodynamic behavior – varies for different temperature gradients and fluid parameters (which are not exactly known), we employ uncertainty quantification methods to obtain reliable results.

*Figure 15* illustrates the standard deviation of the fluid's axial velocity under an uncertain temperature difference between the inner and outer walls.

## Parasites discovered in fossil fly pupae via mathematical algorithms and software for digital reconstruction

Parasitic wasps have existed for many millions of years. Thanks to ultrafast X-ray imaging, it has been possible – for the first time – to prove the presence of parasitic wasps inside their hosts. Four extinct and previously unknown wasp species have been newly discovered and described, with our group developing dedicated mathematical algorithms and the software for the digital reconstruction.

The studies were based on over 1,500 fossils that belong to the collections of the Natural History Museum in Basel (Switzerland) and the Naturhistoriska Riksmuseet in Stockholm (Sweden). The fossils consist mostly of mineralized fly pupae, which were found in late 19th-century phosphorite mines in Quercy in southern France. Basel entomologist Eduard Handschin described them in detail as early as 1944 and pointed to the special importance of the externally inconspicuous pieces, which were only about three millimeters long. Handschin had suspected the contours of a parasitic wasp in a fly pupa that had been estimated to be 34–40 million years old, but this hypothesis could not be definitively proven with the methods used at the time.

Using synchrotron X-ray microtomography, it is now possible to gain access to the insides of objects that are millions of years old. The inner structures can be observed non-invasively and in three dimensions. The fly

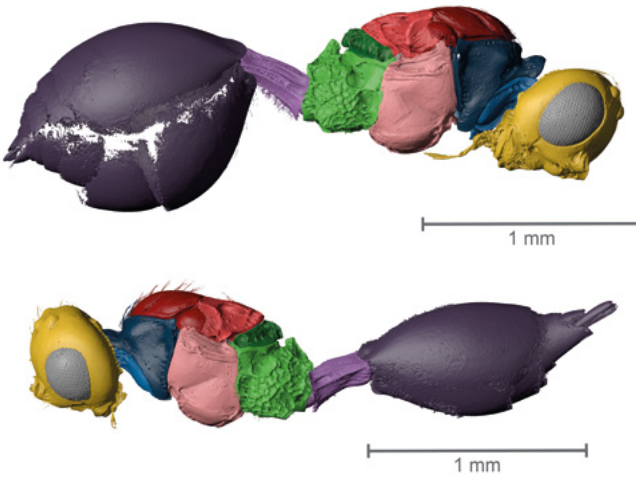



*Figure 16 + Figure 17: Digital reconstruction of parasitic wasp Xenomorphia resurrecta with the application Biomedisa: male (top) and female.*

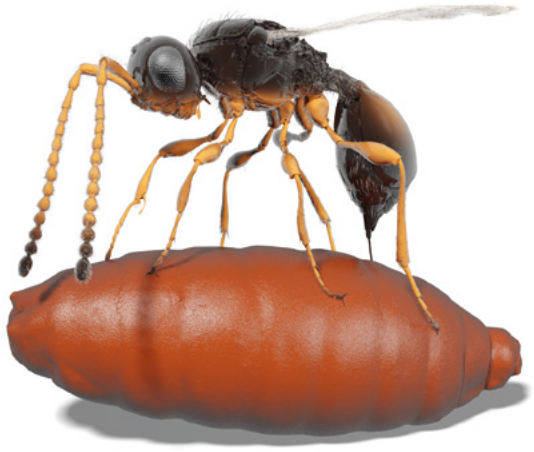


*Figure 18: Digitally resurrected: The parasitic wasp Xenomorphia resurrecta deposits an egg in a fly pupa. (Copyright Thomas van de Kamp, KIT; Nature Communications.)*

pupae were X-rayed at the UFO high-speed tomography station of the KIT synchrotron. The project was also coordinated at the Karlsruhe Institute for Technology (KIT). After imaging the fly pupae, the parasitic wasps were digitally reconstructed in high resolution (see *Figures 16*, *17* and *18*).

This reconstruction was carried out using the Biomedical Image Segmentation App (Biomedisa), which was developed by the associated researcher Philipp Lösel. Biomedisa is able to generate ultrafast analyses of very large sets of three-dimensional imaging data similar to that produced by synchrotron X-ray microtomography. In addition, individual objects are labeled in the data in order to present them three-dimensionally and to use them for further analyses. This segmentation is frequently still carried out manually due to the complexity of observing the object. It usually takes several weeks or even months in order to build the 3D model of a very large set of data. On the basis of high-performance computer technologies working in parallel, Biomedisa's segmentation procedure reduces the work of a few weeks to a few hours.

Thanks to the collaboration between different disciplines, we were able to discover 55 parasitation events and describe four extinct wasp species that were unknown until now. The four different species are parasites that develop inside their host – in this case, the fly pupa – which lived around 40–23 million years ago. The project results provide important evidence on the evolution of parasitism, which is widespread and essentially characterizes ecosystems. Today, about 50% of all animal species are deemed to be parasites. The connection between biodiversity and parasitism is shown particularly clearly in the insect order Hymenoptera, to which wasps belong.

The research findings were published in "Nature Communications" *[van de Kamp, 2018]*, see also the cover image of this Annual Report.

# 2 Research

# 2.4 Groups and Geometry (GRG)

## 2.4
## Groups and Geometry (GRG)

The research group "Groups and Geometry" works 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 the object unchanged. These transformations can be composed – i.e., applied one after the other – and form what is called a group.

In the 19th century, mathematician Felix Klein proposed a new definition of geometry: Geometry is 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 newly discovered hyperbolic geometry, and projective geometry, which has its origins in the study of perspective in art and is not based on the measurement of distances but rather on incidence relations.
Even more importantly, Felix Klein's concept fundamentally changed our view of geometry in mathematics and theoretical physics and continues to influence it today.

In our research group, we investigate various mathematical problems in the fields of geometry, topology, and dynamics that involve the interplay between spaces – such as manifolds or metric spaces – and groups, which act as groups of symmetries on these spaces.

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

*Symmetrien spielen eine zentrale Rolle 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. Jh. entwickelte der Mathematiker Felix Klein einen neuen Begriff der Geometrie: Geometrie ist das Studium der Eigenschaften eines Raumes, die invariant sind unter einer gegebenen Gruppe von Transformationen. Kurz gesagt: Geometrie ist Symmetrie.*

*Mit diesem Konzept vereinheitlichte Klein die klassische Euklidische Geometrie, die damals gerade neu entdeckte hyperbolische Geometrie als auch 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 auf dem Gebiet der Geometrie, Topologie, sowie der dynamischen Systeme, in denen das Zusammenspiel zwischen Räumen und Gruppen, die auf diesen als Symmetriegruppen wirken, zentral ist.*

## Group Leader

Prof. Dr. Anna Wienhard

## Staff Members

Mareike Pfeil
Florian Stecker
Nicolaus Treib *(until June 2018)*

## Visiting Scientists

Prof. Dr. Maria Beatrice Pozzetti
Dr. Daniele Alessandrini
Dr. Shinpei Baba *(until April 2018)*
Dr. Jonas Beyrer *(since October 2018)*
Dr. Valentina Disarlo
Dr. Federica Fanoni *(until September 2018)*
Dr. Sourav Ghosh *(until September 2018)*
Dr. Gye-Seon Lee
Dr. Andreas Ott
Dr. Andy Sanders
Johannes Horn
Christoph Karg
Evgenii Rogozinnikov
Anna Schilling

## Students

Sven Grützmacher *(until September 2018)*
Clemens Fruböse *(since December 2018)*
Lukas Sauer *(since December 2018)*
Menelaos Zikidis

## The Illumination Problem

Imagine you are standing in a room with mirrored walls. The room might be very oddly shaped – polygonal or even with curved walls. Is it always possible to illuminate the entire room with a single source of light? This question is known as the "Illumination Problem". It and related problems have captivated mathematicians since the 1950s and remain an active area of research today. Though easy to state, their solution often is very involved or not known at all. The first answer to the Illumination Problem (reached in the late 1950s) was negative: Roger Penrose found a counterexample. In his "unilluminable room" *(Figure 19)*, there is always a dark region, no matter where you put the candle. The upper and the lower wall of the room are arcs of ellipses with foci at the green points. For different locations of the candle (marked with red dots), there is always a region in the room that remains dark. But how about polygonal rooms? It took four decades until this question, as well, was answered negatively: George Tokarsky found a 26-sided polygonal room *(Figure 2)* for which two points exist that do not illuminate each other *[Tokarsky, G. Polygonal Rooms Not Illuminable from Every Point, American Mathematical Monthly (1995) 102(10):867-879]*. In other words: There is a point for the light source for which a dark spot exists.

To understand how the Illumination Problem can be approached, let us analyze the situation. A light source – for instance, a candle – is located at some point in the room. The light radiates in all directions. Once a light beam meets the mirrored walls, it is reflected by the law of reflection: The angle of incidence equals the angle of reflection. The beam is sent back into the room, where it eventually hits another wall and is reflected again. This procedure repeats infinitely many times. When we only consider one light beam that travels from the light source in a given direction and want to describe its path, the setup is analogous to another situation we know from everyday life: billiards.

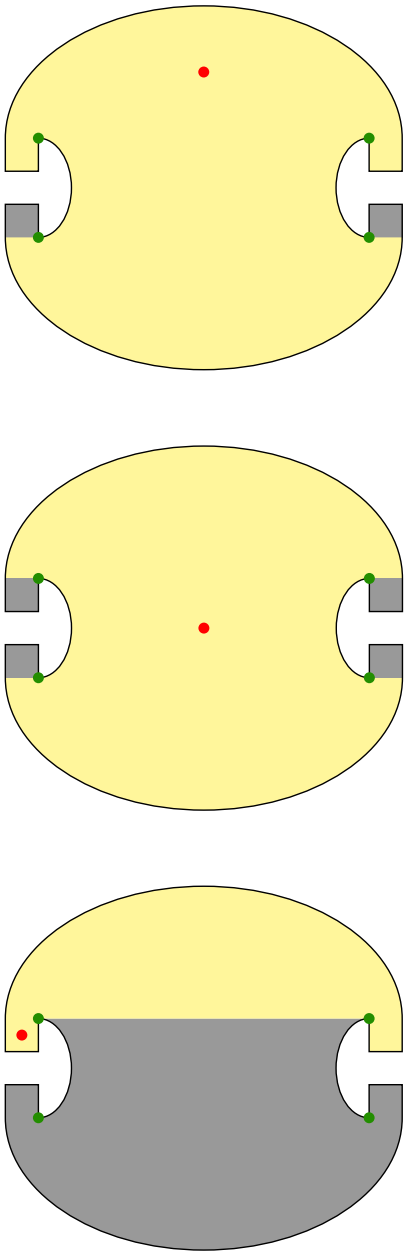


*Figure 19: Penrose's unilluminable room – for different locations of the candle (red), there is always a region that is not illuminated (grey).*

## How mathematicians play billiards

Consider a simplified version of a game of billiards: We assume that there is just one player, just one ball that is point-sized, no friction, and no holes. This game of billiards seems to be rather boring upon first analysis, but it helps to understand what is happening on the billiard table. The billiard table corresponds to the dark room in the Illumination Problem. The point at which the ball rests at the beginning is the light source, and the direction in which the ball is struck provides the direction of the light beam. When the ball hits a boundary, it bounces off in accordance with the law of reflection – just as the light beam does on the mirrored walls. The Illumination Problem translates into the question: "Is it possible to strike the ball from a given position to any other position on the table?" For a rectangular billiard table, the answer is fairly obvious. However, as soon as we consider non-convex tables, the answer becomes trickier. Let us have a closer look at where the billiard ball moves. It draws a path on the table that we call an orbit. The orbit is determined by the starting point, the direction in which the ball moves, and by the shape of the table. This is an example of a dynamical system: We have a given initial situation and that enables us to determine how the situation changes. In our setup, we know the starting point and the initial direction of the ball and can use this information to determine where the ball will be at a later point in time. Dynamical systems appear in physics, for example, when we describe a swinging pendulum, and they also come up in many other parts of our lives. For instance, the development of a population can be described as a dynamical system. Now, let us again examine the dynamical system in which we are interested: the billiard. For the beginning, we consider a rectangular billiard table. What can happen? Either the ball returns to where it began with the same direction in which it began, or it does not. In the first case, we say that the orbit closes. Otherwise, the orbit has infinite length and comes arbitrarily close to any point on the table. We then say that it is dense on the table. Given the starting point and the direction in which the ball begins, can we say which of the two cases will happen?

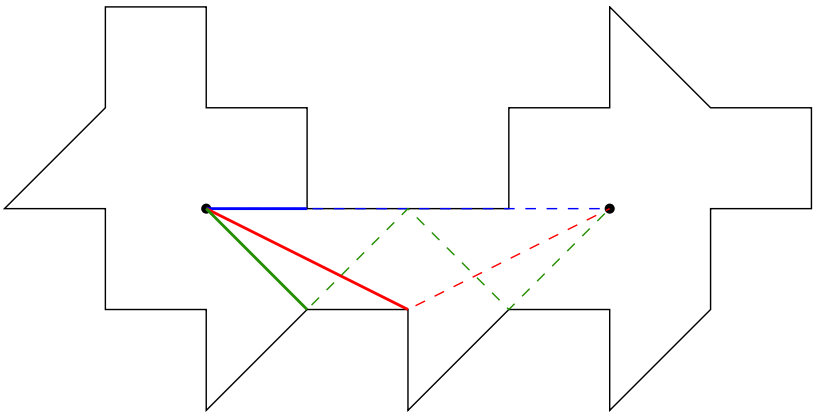


*Figure 20: Tokarsky's unilluminable room – the two points do not illuminate each other. Every light ray from one to the other meets some corner at which it cannot be reflected.*

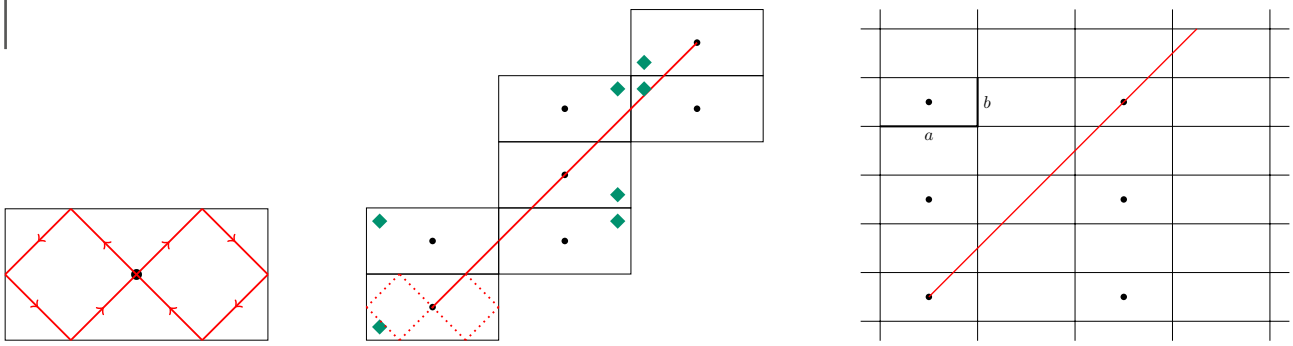



*Figure 21: Unfolding the orbit on a rectangular billiard table.*

## Reflect rather than rebound

To investigate this question, we use a trick: Instead of letting the ball bounce off at the boundary of the table, we reflect the entire table on the side where the ball hits. The orbit then goes straight into the reflected copy of the table. When the ball hits the next boundary, we reflect again. Repeating this procedure, step by step, we unfold the orbit onto a straight line in the plane. *Figure 21* illustrates this concept. It shows the original orbit on the billiard table *(Figure 21, left)* and the unfolded version *(Figure 21, middle)*. For every copy of the table, we have to memorize how it corresponds to the original table. We thus placed green markers in one corner of the rectangle. There are four different types of copies that correspond to the reflections at the four sides of our rectangular table. By reflecting, we can cover the entire plane with copies of the table. If the orbit is closed on the table, the unfolded orbit in the plane meets one of the copies of the starting points that lies in a copy of the table with the same orientation as the original one (drawn in red in *Figure 21, right*). Thus, we are only interested in copies of the starting point that lie in copies with the same orientation of the sides as the original table. In *Figure 21 (right)*, we deleted all copies of the starting points and thus no longer need the markings. Denoting the lengths of the sides of the billiard table with $a$ and $b$ ($a > b$), we see that the orbit is closed if and only if the slope of the orbit is a rational multiple of the ratio of the sides $\frac{b}{a}$. Otherwise, the orbit never closes up and instead remains dense on the billiard table. The procedure of reflecting the billiard table instead of letting the ball rebound might appear complicated at first sight, but it helped us to easily classify when the orbit closes.

Just as the room in the Illumination problem does not need to have a rectangular base, we can also consider billiard tables with different shapes. We begin by examining a triangular table with interior angles 30°, 60°, and 90°. Following our strategy of "reflect rather than rebound," we reflect the table. This time, there are twelve essentially different copies that form a hexagon. Again, we are only interested in copies of the starting point that lie in tables with the correct orientation. In every hexagon, there is exactly one such table, thus exactly one admissible copy of the starting point. *Figure 22* illustrates the situation. Similar considerations as before lead us to the conclusion that the orbit is closed if and only if the slope is a rational multiple of $\sqrt{3}$.

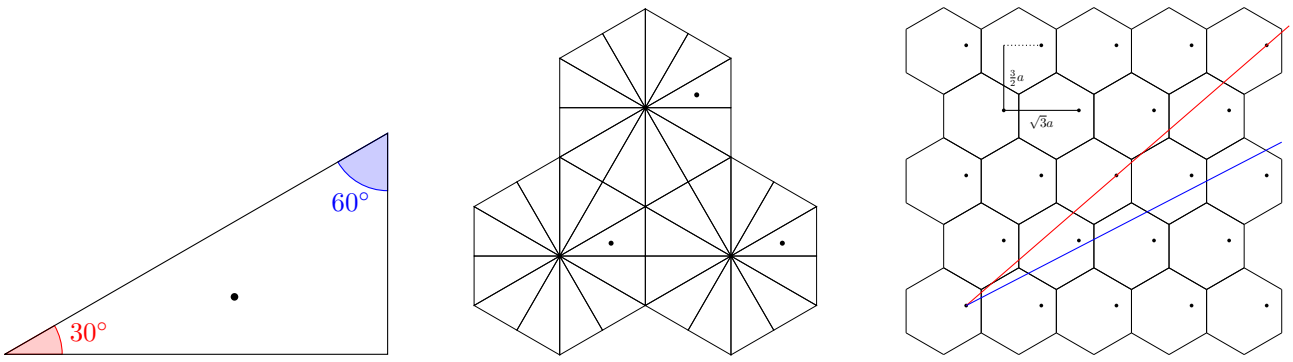



*Figure 22: Unfolding the orbit on a triangular billiard table. On the right, "a" denotes the length of the hypotenuse of the triangle.*

## From billiards to translation surfaces

Until now, our strategy of reflecting the table seems to have worked well. Let us now examine another triangular billiard table, this time with interior angles 22.5°, 67.5°, and 90°. For this table, we need 16 copies of the triangle to obtain all possible orientations. These copies form a regular octagon, but in contrast to the hexagon, it is not possible to cover the plane with regular octagons. We cannot apply the same considerations as before. What can we do instead? We can still add octagons along the orbit (see *Figure 23*). Now, all octagons that appear are translated copies of the first octagon and can be identified with it using translations. Under this identification, the ball leaves the original octagon at the right side and enters it again from the left with the same direction. Why? The unfolded orbit enters the second octagon from the left, which can be identified with the original octagon via a translation. When the orbit leaves the octagon again through another wall, it enters

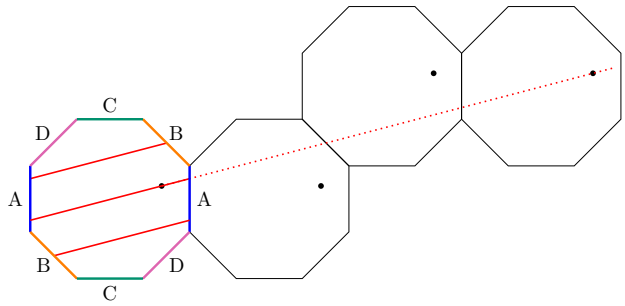



*Figure 23: The octagons are translated copies of the first one and can be identified with it, which results in identifying opposite sides of the octagon, marked with the same color and letter.*

again from the corresponding opposite side. In *Figure 23*, the sides that are identified are marked with the same color. This procedure provides us with all necessary information about the orbit inside the original octagon. The octagon in which we identify sides pairwise via translations is an example of a translation surface. The advantage of using translation surfaces to understand billiards is that on them, the orbit is always parallel and never intersects itself.

The rectangular billiard from above can be viewed in the setting of translation surfaces, as well. We first reflect the table until we have all possible orientations. In this case, four copies of the table again form a rectangle whose side

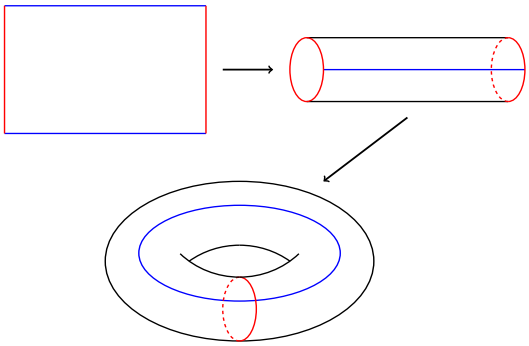


*Figure 24: Identifying opposite sides of a rectangle to obtain a torus.*

lengths are two times the side lengths of the original rectangle. Next, we identify opposite sides (*Figure 24*). The resulting translation surface is a torus and can be visualized as the surface of a donut. In this case, we have several ways of drawing the orbit. The first one is on the original billiard table with rebounding and self-intersections. The second option is on the translation surface, viewed as a polygon and for which the orbit is always parallel to itself. Finally, we can view the orbit on the torus and see how it winds around it. Strictly speaking, this option is the same as the second one but is visualized in three dimensions. *Figure 25* displays 3D prints of the three different ways of drawing the orbit.

Translation surfaces provide a powerful tool for investigating billiards on various tables. For instance, given a direction $\theta$ on the table, it is possible to divide the table into two kind of components: periodic and non-periodic. These components have the following property: Whenever an orbit in direction $\theta$ begins in a periodic component, it closes up; when it begins in a non-periodic component, the orbit has infinite length, never leaves this component, and is dense in it.

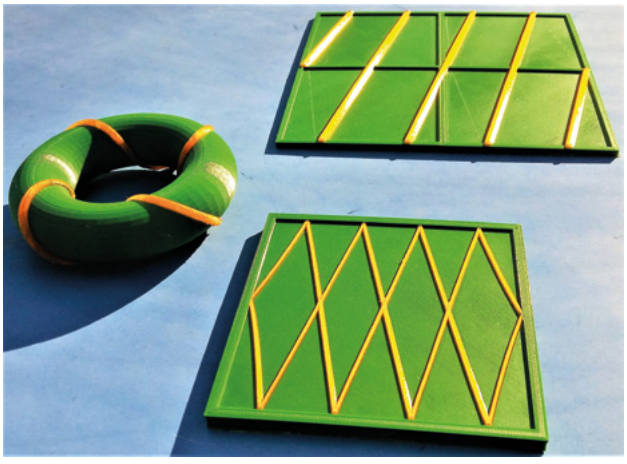


*Figure 25: Three different ways of viewing the orbit of a rectangular billiard.*

## The moduli space of translation surfaces

Translation surfaces are not only a tool for understanding billiards but are themselves an interesting object of study. The *moduli space* is the space of all possible translation surfaces of a fixed type. What do we mean by the type of a translation surfaces? We have already seen two different examples of translation surface: the torus, that is a rectangle with opposite sides identified, and the octagon, also with pairwise identified sides. In both cases, all vertices of the polygon are identified to the same point, which we call a *singularity*. The *cone angle* of the singularity is the sum of the interior angles. Both the torus and the octagon have only one singularity with cone angle $2\pi$ for the torus and $6\pi$ for the octagon *(Figure 26)*. Furthermore, these two examples are topologically different: If we visualize the identification of sides by gluing, the resulting surface has one hole for the torus and two holes for the glued octagon. The number of holes is called the *genus* of the surface. The genus, the number of singularities, and the cone angles together form the type of a translation surface. We thus only want to examine translation surfaces with a fixed genus and a fixed number of singularities with given cone angles. Moreover, we fix the area of the surface to be 1. Such a translation surface can be represented by a polygon in $\mathbb{R}^2$ with area 1 and with sides identified pairwise. We can apply a matrix in $SL(2, \mathbb{R})$ – a 2 by 2 matrix with real entries and determinant one – to a polygon in $\mathbb{R}^2$ and obtain another polygon in $\mathbb{R}^2$ – a sheared, stretched, or rotated version of the original one (see *Figure 27*). However, parallel sides remain parallel, and the fact that the matrix has determinant 1 guarantees that the new polygon has area 1 again. By identifying sides, this polygon represents a translation surface. In this way, we can apply a matrix in $SL(2, \mathbb{R})$ to a translation surface and obtain another translation surface. We say that $SL(2, \mathbb{R})$ *acts on the moduli space*. However, not every non-trivial matrix acts non-trivially: The matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ maps the square S with vertices (0,0), (1,0), (1,1), and (0,1) to the parallelogram P with vertices (0,0), (1,0), (2,1), and (1,1). As polygons, the two are different, but if we cut P along the red dashed line from (1,0) to (1,1) and glue the right triangle that we obtain to the left side, we obtain a square (see *Figure 27*). This process demonstrates that the corresponding translation surfaces of S and P are identical. Thus, the matrix maps the corresponding translation surface to itself.

Alex Eskin, Maryam Mirzakhani, and Amir Mohammadi investigated the action of $SL(2, \mathbb{R})$ on the moduli space and found that the closure of the union of all possible images of a fixed translation surface under the $SL(2, \mathbb{R})$ action is a "nice" subset of the moduli space *[Eskin A, Mirzakhani M, Mohammadi A. Isolation, equidistribution and orbit closures for the $SL(2, \mathbb{R})$ action on moduli space, Ann. Math. (2015) 182:673-721]*. Their result was a major breakthrough in the field and has wide-reaching implications. This result inter alia led to Mirzakhani's being awarded the Fields Medal, one of the most prestigious awards in mathematics.

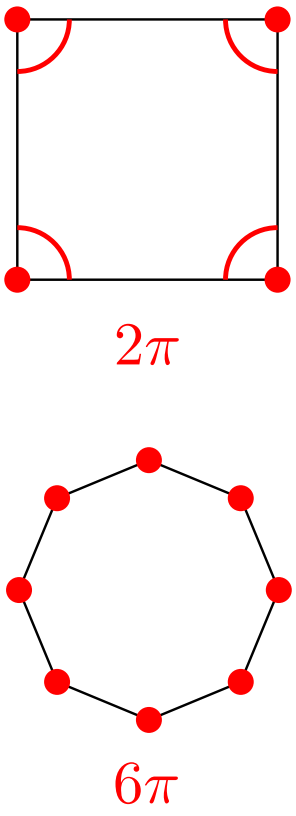



*Figure 26: Cone angles for translation surfaces obtained from gluing a rectangle and an octagon.*

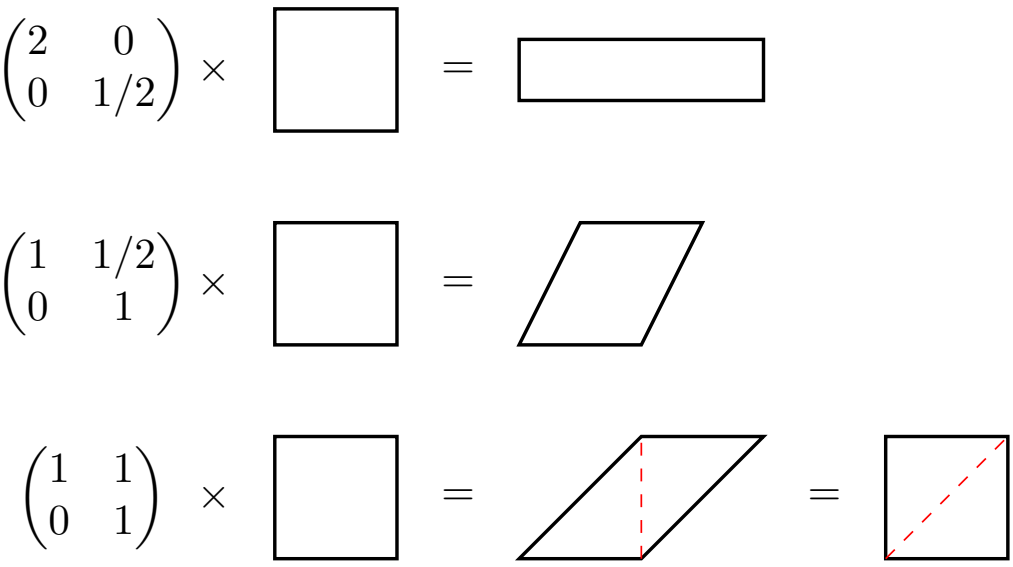


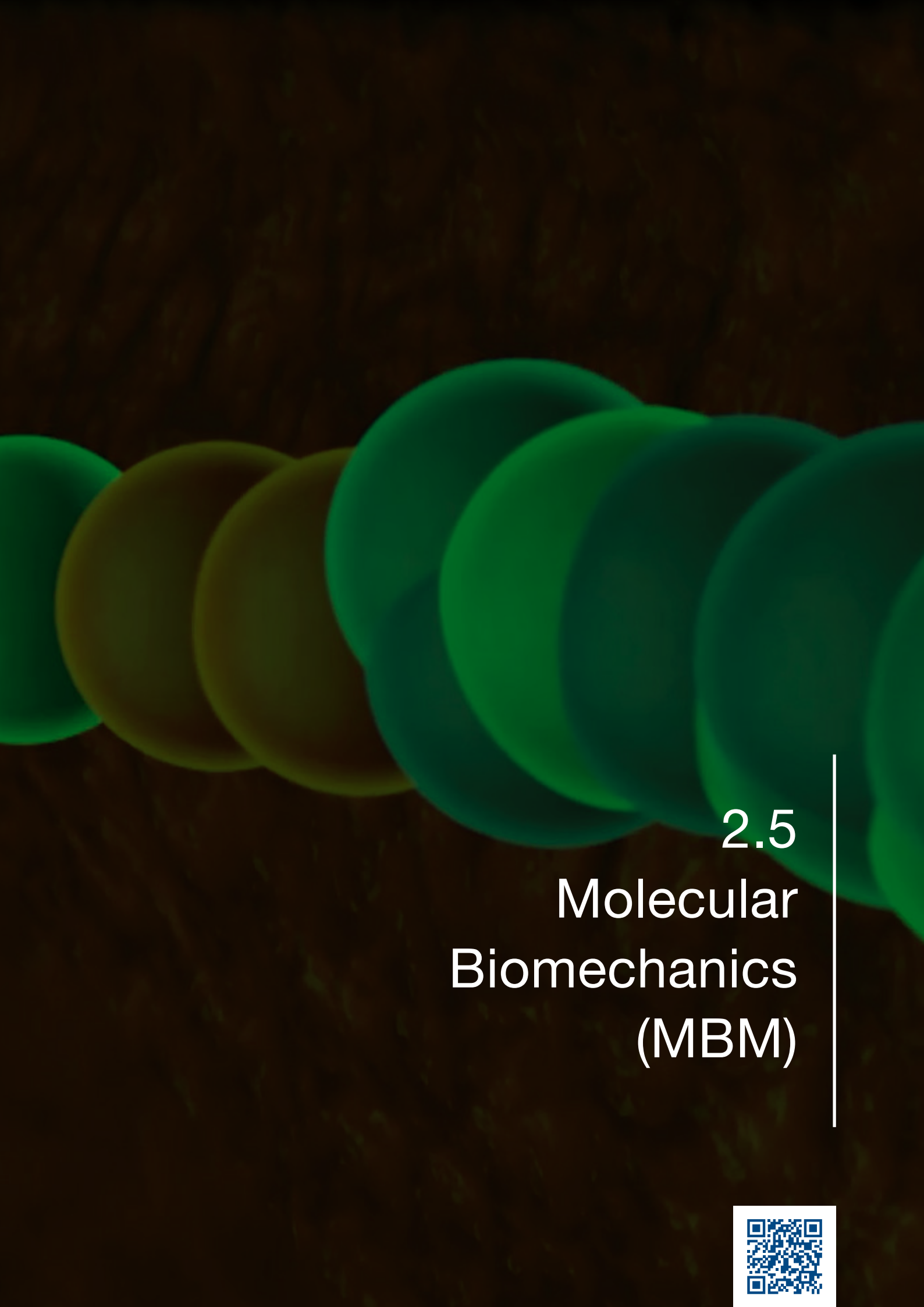
*Figure 27: Applying a matrix to a polygon in $\mathbb{R}^2$, we obtain a sheared, stretched, or rotated version of it.*

## The Illumination Problem revisited

Now, let us return to the illumination problem. Remember: The question was if it is always possible to illuminate a room with mirrored walls with a single source of light, and Tokarsky provided a counterexample by finding a polygonal room and two points in it that do not illuminate each other. Recently, Samuel Lelièvre, Thierry Monteil, and Barak Weiss proved that for a polygonal room with only rational angles, there are always only finitely many points with this property. In their paper "Everything is illuminated", the authors used the results from Eskin, Mirzakhani, and Mohammadi on the $SL(2, \mathbb{R})$ action on the moduli space to demonstrate that any light source illuminates the entire room – except, perhaps, for finitely many points *[Lelièvre S, Monteil T, Weiss B. Everything is illuminated, Geometry & Topology (2016) 20(3):1737-1762]*. Note that in order to investigate a property of a single translation surface – here, illuminability – the authors used results on the moduli space of all translation surfaces. This technique of considering a more general case instead of a concrete one not only applies to translation surfaces but can also be used in other contexts.

In our group, we work not only with moduli spaces of translation surfaces but mostly with moduli spaces of other geometrical objects: moduli spaces of surfaces with a given genus with or without punctures, cusps, or boundary, or moduli spaces of higher-dimensional objects. One element of our research is dynamical systems on moduli spaces. How can a geometric object be deformed? What path does this deformation draw in the moduli space? We also investigate how groups act on the moduli spaces – similar to $SL(2, \mathbb{R})$ above, which acts on the moduli space of translation surfaces. This procedure helps us to obtain a deeper understanding of the moduli spaces and the objects they contain.

# 2 Research



# 2.5 Molecular Biomechanics (MBM)

## 2.5
## Molecular Biomechanics (MBM)

Proteins are the working horses of living systems. In the past decades, we have learned much about how proteins look, move, and work. A rather new aspect to proteins is that many of them display very surprising properties once mechanical force acts on them, which has important consequences for their function in the biological cell. The aim of the Molecular Biomechanics group is to decipher the mechanical function of proteins. We use Molecular Dynamics simulations, coarse-grained simulations, and continuum mechanics approaches to address this issue on multiple time- and length scales.

In 2018, we advanced the field of mechanobiology by various aspects of protein mechanical function, a few of which are highlighted here. We made important progress in the field of mechano-sensing in flowing blood and tackled the question of how the steady flow of blood – as well as its varying physical properties, such as the mean velocity or the shear gradient – is used as a stimulus in biochemical processes occurring in blood. In collaboration with Phil Hogg (Sydney), we could pinpoint two distinct mechanisms that render blood proteins sensitive toward the forces present in blood flow.

First, integrin – a pivotal receptor protein that occurs inter alia at the platelet surface – catches its binding partners in a force-dependent manner. Secondly, von Willebrand factor – a protein we have been working on for nearly a decade – features disulfide bonds that regulate its mechanical response to blood flow. In collaboration with Edward Lemke (EMBL Heidelberg), we were also able to establish that seemingly similar intrinsically disordered proteins of nuclear pores follow very distinct binding mechanisms with their receptors. Finally, we developed a new and very versatile analysis method based on contacts within and among proteins that we now use in virtually all protein-related projects in the group and that we predict to be very valuable to the greater Molecular Dynamics community.

*Proteine sind die Arbeitstiere lebender Systeme. In den vergangenen Jahrzehnten haben wir viel darüber gelernt, wie sie aussehen, sich bewegen und arbeiten. Ein relativ neuer Aspekt dabei ist, dass viele Proteine äußerst überraschende Eigenschaften zeigen, wenn mechanische Kräfte auf sie einwirken, was wiederum wichtige Auswirkungen auf ihre Funktion in der biologischen Zelle hat. Das Ziel der MBM-Gruppe ist, die mechanische Funktion von Proteinen zu entschlüsseln. Um dies auf unterschiedlichen Zeit- und Längenskalen zu untersuchen, verwenden wir Molekulardynamik-Simulationen, grobauflösende Simulationen und Ansätze aus der Kontinuumsmechanik.*

*2018 haben wir in der Mechanobiologie erhebliche Fortschritte gemacht, was die mechanische Funktion von Proteinen betrifft. Einige von ihnen stellen wir hier näher vor. Auf dem Gebiet der Mechanosensorik in fließendem Blut konnten wir wichtige Beiträge liefern. Wie wirkt der konstante Blutfluss – und seine wechselnden physikalischen Eigenschaften, wie z. B. die mittlere Geschwindigkeit oder der Schergradient – als Impuls in biochemischen Prozessen, die im Blut vorkommen? In Zusammenarbeit mit Phil Hogg in Sydney konnten wir zwei unterschiedliche Mechanismen genauer bestimmen, die Blutproteine empfindlich gegenüber den Kräften im Blutfluss machen.*

*Zum einen untersuchten wir Integrin, ein zentraler Rezeptor, der unter anderem auf der Oberfläche von Thrombozyten vorkommt und dessen Bindung an seine Partner kraftabhängig ist. Zum anderen arbeiteten wir weiterhin intensiv am Von-Willebrand-Faktor, einem Protein, dessen Disulfidbindungen seine mechanische Reaktion auf den Blutfluss regulieren. Wir konnten außerdem in Zusammenarbeit mit Edward Lemke vom EMBL Heidelberg nachweisen, dass scheinbar ähnliche intrinsisch ungeordnete Proteine von Kernporen ganz bestimmten Bindungsmechanismen mit ihren Rezeptoren folgen. Schließlich gelang es uns, eine neue und äußerst vielseitig einsetzbare Analysemethode zu entwickeln, die auf Kontakten innerhalb und zwischen Proteinen basiert, die wir von nun an in praktisch allen protein-bezogenen Projekten der Gruppe verwenden und bei der wir davon ausgehen, dass sie für die breite Forschungsgemeinschaft auf dem Gebiet der Molekulardynamik von Nutzen sein wird.*

## Group Leader

Prof. Dr. Frauke Gräter

## Staff Members

Dr. Csaba Daday
Florian Franz
Ana María Herrera-Rodríguez
Fabian Kutzki
Isabel Martin *(since February 2018)*
Dr. Vedran Miletić *(until April 2018)*
Dr. Fan Jin
Dr. Agnieszka Obarska-Kosinska
Christopher Zapp

## Students

Svenja de Buhr *(since September 2018)*
Maximilian Seidel *(March – May 2018)*
Tobias Jäger *(until May 2018)*
Salvador Ramirez Ciarreta
Benedikt Rennekamp *(since January 2018)*
Anna Schröder *(since December 2018)*
Steven Sheridan
Frieda Sorgenfrei *(March – October 2018)*
Odysseas Vavourakis *(April – June 2018)*
Christopher Lance *(until April 2018)*

## Guest Scientists

Prof. Dr. Fei Xia *(August 2018)*
Dr. Kevin Hinkle *(June – July 2018)*
Dr. Krisztina Fehér *(since April 2018)*
Dr. Lipi Thukral *(October 2018)*

## Disulfide bonds: molecular switches in blood

**F. Gräter (with C. Aponte-Santamaría, University of los Andes, Bogotá, Colombia)**

Disulfide bonds play key roles in proteins in terms of not only stabilizing their structure but also influencing their function. In collaboration with Philip Hogg's (Sydney) group, we found that disulfide bonds regulate two key processes during the coagulation cascade: the activation of the von Willebrand factor adhesive protein and the control of integrin de-adhesion.

Von Willebrand factor (VWF) is an enormous extracellular protein that plays a key role in primary hemostasis, during which it adheres platelets at sites of vascular injury to stop bleeding. Its activation is mediated by mechanical stress imposed by the flowing blood. Platelets adhere to the VWF A1 domain, and we previously demonstrated that VWF's neighbor domain (VWF A2) is able to autoinhibit its binding via force-dependent specific inter-domain interactions. In this new study *[Butera et al., 2018]*, we used a combination of experimental biophysical techniques and molecular dynamics (MD) simulations and demonstrated that a vicinal disulfide bond – located at the C-terminal part of the A2 domain – mediates the activation of its A1. More specifically, only when the bond is cleaved does the A2 domain bind to the A1 domain and block platelet GPIb binding *(Figure 28A)*. Our simulations indicate that the cleavage of the disulfide bond modifies the structure *(Figure 28B)* and molecular stresses *(Figure 29)* of the A2 domain in a long-range allosteric manner, which provides a structural explanation for the redox control of the autoinhibition. Our findings thus demonstrate that a disulfide bond switch regulates the mechanopresentation of VWF.

Integrins are a group of cell-surface proteins that play different roles in the cells and become activated by interaction with other protein-ligand partners. The integrin αIIbβ3 – found at the surface of blood platelets – plays a key role in blood clotting by forming the clots upon interactions with a ligand protein called fibrinogen. In this study *[Passam et al. 2018]*, we employed a series of experimental and computational methods to demonstrate that

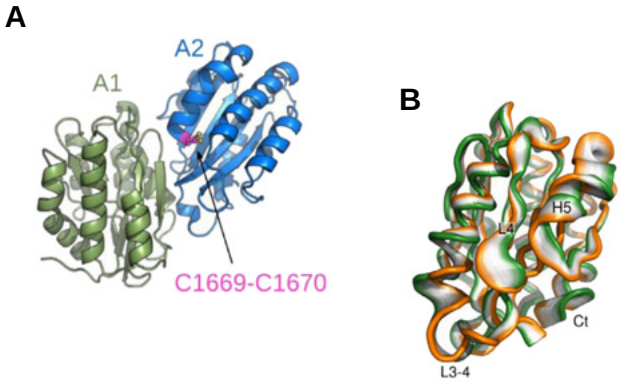



*Figure 28: Disulfide-bond switch in von Willebrand factor (VWF). A. Force-dependent interactions between the VWF A2 (blue) and A1 (green) domains are mediated by the vicinal disulfide bond C1669-C1670 (magenta) located at the C-terminus of A2. B. Interpolation of structures sampled by the A2, ranging from the oxidized state (green) to the reduced state (orange).*

yet another partner – named ERp5 – mediates the interaction of integrin αIIbβ3 with fibrinogen by cleaving a disulfide bond present at the integrin. Our data demonstrate that ERp5-mediated disulfide-bond cleavage occurs only once integrin has released fibrinogen and that this process is augmented by the application of shear forces, just as it occurs in nature in flowing blood. MD simulations reveal how the cleavage of the disulfide induces long-range allosteric effects within the βI-domain, which mainly affect the metal-binding sites, resulting in the release of fibrinogen *(Figure 30)*.

Overall, our interdisciplinary approach – which combined simulations and experiments – demonstrated that the mechano-redox control of protein-protein interactions thereby crucially modulates functional (de-)adhesion in blood.

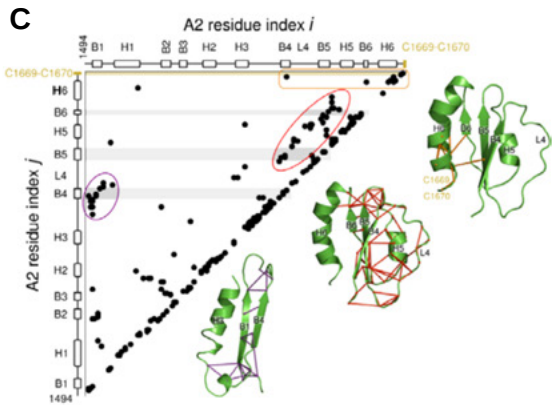



*Figure 29: Residue pairs (i,j) in the A2 domain, which displayed a pronounced increase in mechanical stress (pair-wise force > 90 pN) upon the reduction of the vicinal disulfide bond. Pairs are highlighted with the colored links in protein illustrations at the right.*

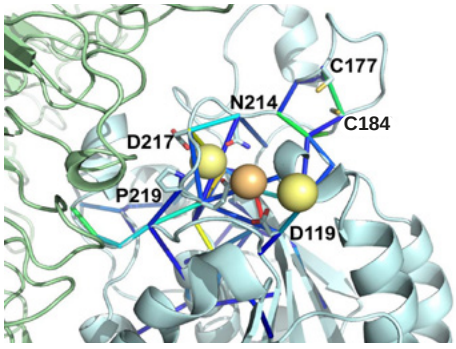



*Figure 30: Disulfide-bond redox state controls de-adhesion of integrin αIIbβ3. Reduction of the Cys177-Cys184 disulfide bond causes allosteric differences in the mechanical stress of integrin, as indicated by the sticks (colored by the force difference).*

## A figure worth 1,000 words

**Frauke Gräter and Csaba Daday (with Davide Mercadante, University of Zürich, Switzerland)**

Proteins constantly move and change their conformation. Molecular dynamics typically answers the question of what the possible conformations of proteins are. However, proteins have a highly complicated and crowded structure, and understanding changes in their behavior is a challenging task due to the high number of coordinates that need to be monitored. Digesting the large amount of molecular data often involves creative 3D visualization, but even with considerable effort, important details can be missed. This fact led to a dichotomous problem: Not only was data visualization a challenge, but scientists also ran the risk of overlooking aspects of their own results. A novel tool called CONAN (CONtact ANalysis) – developed at "Molecular Biomechanics" at HITS – can alleviate these issues through compressing this 3D data into simpler 2D images – called contact maps – that capture the key interactions.

Contact maps measure inter-residue distances, thereby compressing 3D structures into 2D images, as shown in *Figure 31*. This process often facilitates data interpretation and makes important changes easier to spot. These contact maps have usually only been used to study single-protein structures as a single snapshot, but they can in fact easily be obtained for many structures, resulting in a contact map movie. This analysis somehow extends the saying, "A figure is worth 1,000 words," into the dynamic regime since it creates a multitude of possible contact-map snapshots from one simulation, thereby enabling the identification of conformational subpopulations and transitions.

Until now, analysis methods based on contact maps have been widely used only for understanding single structures, such as those in the Protein Data Bank (PDB). Even when the methods were generalized for dynamic simulations, the implementations were often various "ad hoc" analysis scripts since there wasn't a standardized tool. This meant that the measured quantities and definitions were inconsistent and that results were not directly comparable. The new tool "CONAN" *[Mercadante et al. 2018]*, however, is a standardized, easy-to-use package that allows for several different types of analyses, including principal-component analysis and cluster analysis. Developed by Csaba Daday and Frauke Gräter of the Molecular Biomechanics group as well as former group member Davide Mercadante, this tool therefore fills a gap and offers a comprehensive, user-friendly program that requires no programming experience and can help scientists who perform molecular-dynamics calculations to understand and present their data. This tool will hopefully lead to a more widespread use of these measures and a more uniform set of definitions. The tool is open access and free of use, and the team constantly optimizes the software and is open to feedback from the community.

CONAN is freely available at *https://github.com/HITS-MBM/conan/tree/master/docs*.

Examples and illustrations can be found on our blog at *https://contactmaps.blogspot.de/* and on our YouTube channel at *tinyurl.com/Con4nMD*.

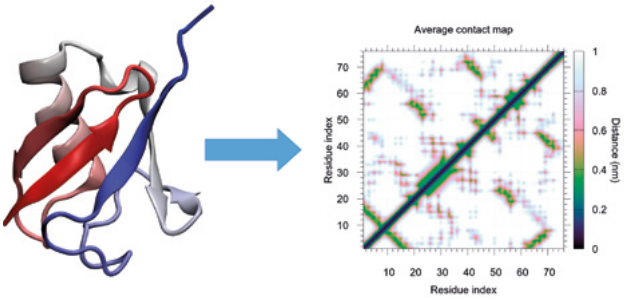


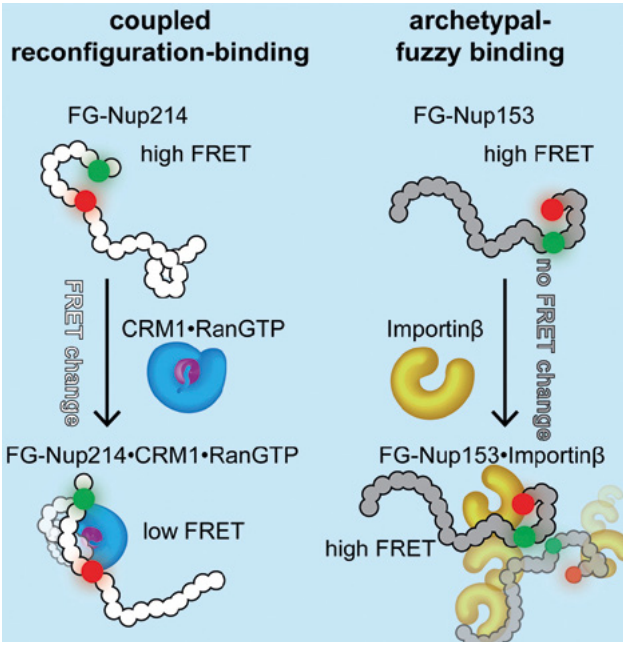

*Figure 31: The dynamics of a 3D structure (left) is transformed into an animated 2D contact map, which allows for an easier overview of inter-residue interactions (right).*

## Two Differential Binding Mechanisms of FG-Nucleoporins and Nuclear Transport Receptors

**Frauke Gräter (with Davide Mercadante, University of Zürich, Switzerland)**

The past decade has seen the breakdown of one of biology's long-standing paradigms, which states that macromolecular structure is necessarily related to function. A large part of organismal proteomes is now considered intrinsically disordered as intrinsically disordered proteins (IDPs) do not show a stable formation of tertiary or secondary structure within their functional environment. Nevertheless, they are still able to perfectly function inside cells and fulfill complicated roles ranging from cell signaling to chaperoning and molecular transport.

Over the past years, we have investigated a particular class of IDPs called FG-nucleoporins (FG-Nups). FG-Nups associate inside the nuclear pore complex (NPC) in a spaghetti-like hydrogel and interact specifically with nuclear transport proteins that facilitate the controlled shipping of cellular components into or out of the nucleus. This interaction is mediated by binding moieties composed of only phenylalanine-glycine (FG) repeats, and the nuclear transport occurs efficiently within 3–5 milliseconds.




In this contribution *[Tan et al. 2018]*, we tackled the following question: "Is the conformational-independent binding of nuclear transport receptors (NTRs) common to all FG-nucleoporins?" The question is extremely relevant as active transport through the NPC proceeds through different phases: docking, passage through the hydrogel, and release. This final step is mediated by the association of RanGTP with the NTR bound to the FG-Nups. It has been speculated that aside from the nucleoporins, which form the central core of the NPC that composes the permeability barrier, other nucleoporins at the edges (either cytoplasmic or nucleoplasmic) of the NPC may instead have other, specific functionalities. FG-Nup214 has been suggested to take part specifically in the discharge of cargo on the cytoplasmic side of the pore and is simultaneously important in the recruitment of the DEAD-box helicase Ddx19 for the formation of ribonuclease particles. How does FG-Nup214 interact with cargo proteins? Does it follow the same paradigm observed for hydrogel-forming FG-Nups in the central barrier? By testing the binding of FG-Nup214 to the nuclear transport receptor CRM1 via a combination of molecular dynamics (MD) simulations, smFRET, and stopped-flow spectroscopy, we discovered that FG-Nup214 – unlike other FG-Nups – undergoes a conformational transition when binding to nuclear transport receptors *(Figure 32)*.

*Figure 32: Schematic representation of the differences between the mechanisms of binding of a central pore nucleoporin vs. Nup214. The conformational-dependent association is opposed the conformational-independent binding of a central pore that forms nucleoporin FG-Nup153.*

We discovered that the half-life of the CRM1-Nup214 complex is – on average – ~60 times longer than that of FG-Nup153, a central pore nucleoporin. This higher half-life is favored by a mechanism of recognition encoded within the intrinsically disordered partner, which displays a differential binding affinity for the different subsites that dock FG-repeats along its sequence. This mechanism is mediated by the formation of transient secondary structure elements that dock stably into place and flank binding FG-repeats.

This new mechanism of FG-Nups that binds nuclear transport receptors reveals – for the first time – that disorder within the same class of proteins can mediate more than one binding mechanism and that the formation of a fuzzy, short-lived, and conformation-independent complex is not the only means of favoring nuclear transport *(Figure 33)*.

Future research will be devoted to understanding whether this mechanism of action is also peculiar in other FG-Nups located strategically along the NPC so as to facilitate particular processes (such as engagement and disengagement of binding partners from the NPC), thereby enhancing the overall efficiency of nuclear transport *(Figure 33)*.

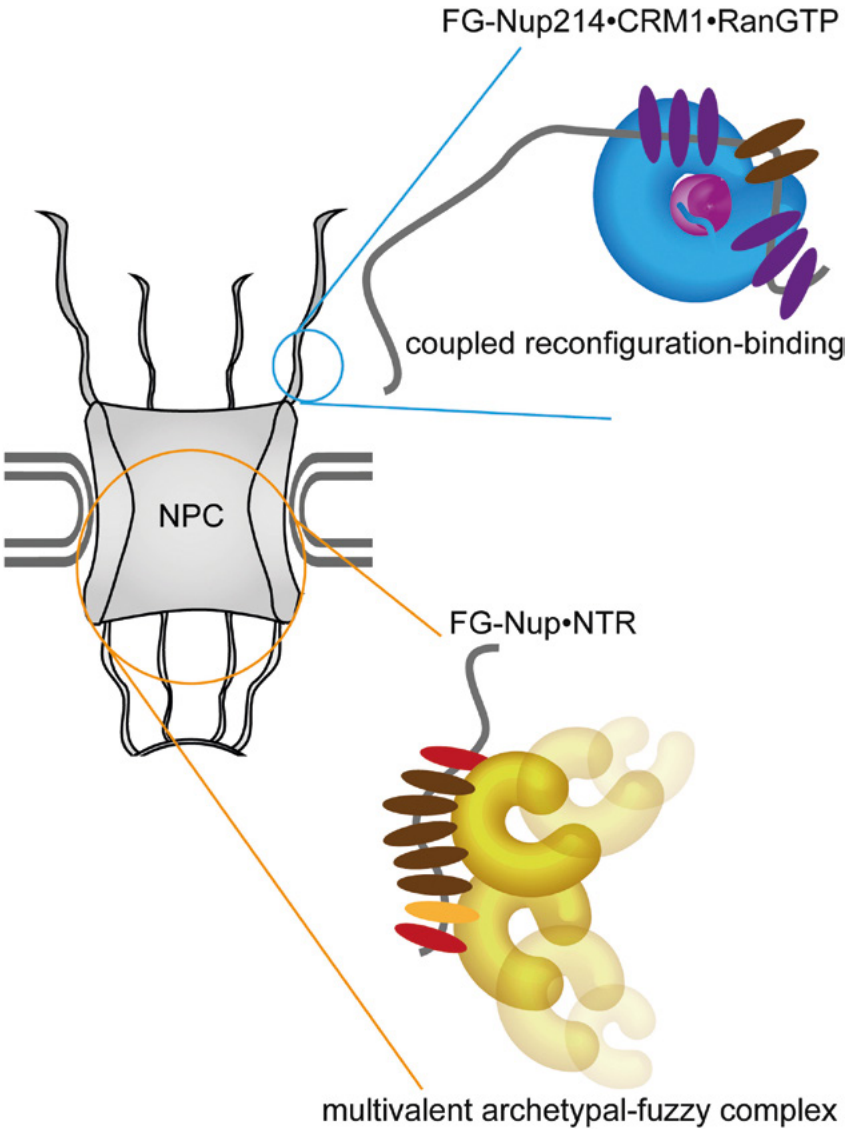
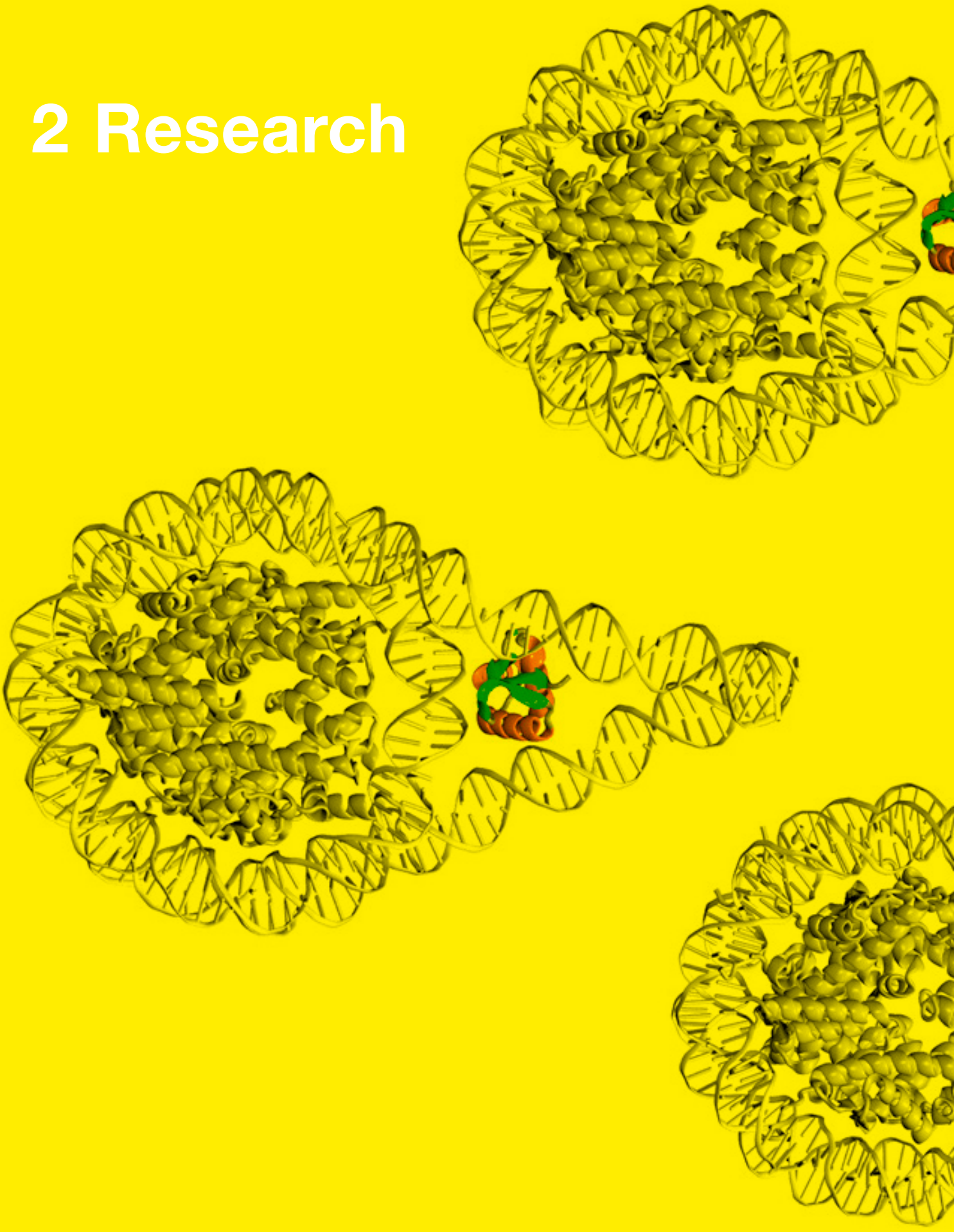
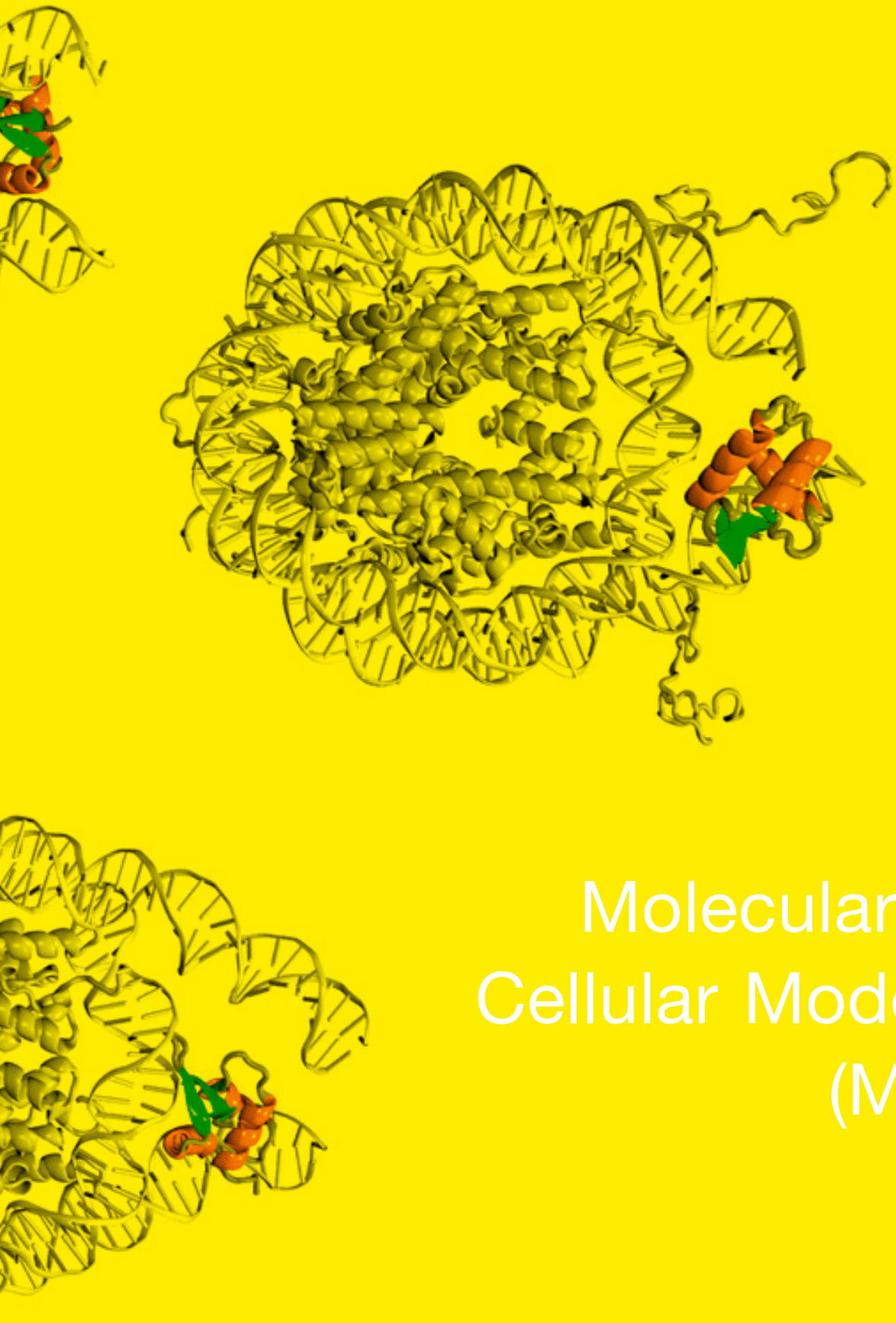



*Figure 33: Differential Binding Modes of FG-Nup•NTR Interactions. In contrast to the previously reported archetypal-fuzzy binding mechanism of FG-Nup153·Importinβ, disordered FG-Nup214 undergoes a coupled reconfiguration binding with the NTR CRM1·RanGTP complex, indicating that variations in the type and extent of individual FG·NTR interactions can drastically change the mechanism of binding between the FG-Nup and its NTR.*

# 2 Research





# 2.6 Molecular and Cellular Modeling (MCM)

# 2.6 Molecular and Cellular Modeling (MCM)

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

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

These questions are illustrative of the types of problems that we address in our projects by developing and applying computational approaches to study 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 detail some of the results achieved this year. These achievements demonstrate the types of methods we develop to study macromolecular interactions and their application to problems in biology, biotechnology, and drug design. Following a general overview, we focus on projects on (i) the computational prediction of drug-target binding kinetics, (ii) chromatosome formation through linker histone-nucleosome interactions, and (iii) multiresolution modeling of biomolecular complexes.

*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 Vorhersage mittels Computer von Wirkstoff-Protein Bindungskinetik, (ii) der Bildung des Chromatosoms durch Linker Histon-Nukleosom Wechselwirkungen, und (iii) der Multiresolution- Modellierung von biomolekularen Komplexen.*

## Group Leader

Prof. Dr. Rebecca Wade

## Staff Members

Christina Athanasiou *(since September 2018)*
Dr. Neil Bruce
Dr. Daria Kokh
Dr. Mehmet Öztürk *(until May 2018)*
Dr. Ghulam Mustafa *(until May 2018)*
Ina Pöhner
Dr. Stefan Richter
Dr. Kashif Sadiq
Alexandros Tsengenes *(since September 2018)*

## HITS Scholarship Holder

Gaurav Ganotra

## Visiting Scientists

Lucas Gasparello Viviani *(since September 2018, Fapesp Scholarship, University of São Paulo, Brazil)*
Dr. Goutam Mukherjee *(BIOMS fellow, Heidelberg University)*
Dr. Prajwal Nandekar *(Heidelberg University)*
Dr. Ariane Nunes-Alves *(CellNetworks Postdoctoral Fellow, Heidelberg University)*
Katarzyna Świerkula *(August – September 2018, Erasmus Student, University of Warsaw, Poland)*
Prof. Huan-Xiang Zhou *(January – May 2018, Romberg Professor, Heidelberg University, on sabbatical from University of Chicago, USA)*

## Students

Lukas Adam *(since December 2018)*
Jan-Niklas Dohrke *(March – July & October – November 2018)*
Patrick Friedrich *(until November 2018)*
Tom Kaufmann *(May – July 2018)*
Bastian Kister *(October – December 2018)*
Jui-Hung Yuan

# 2.6 Molecular and Cellular Modeling (MCM)

## General news

This year, we welcomed two new arrivals from the University of São Paulo, Brazil: Dr. Ariane Nunes-Alves joined the group as a CellNetworks postdoctoral fellow at the beginning of the year, and Lucas Gasparello Viviani came in the autumn as a visiting doctoral student with the support of a Fapesp fellowship to learn how to apply our TRAPP methodology to his studies on enzyme inhibitors. We also welcomed two new arrivals from the Academy of Athens, Greece: Christina Athanasiou and Alexandros Tsengenes joined us as doctoral students working on the Euroneurotrophin International Training Network (ITN), an interdisciplinary EU-supported project that began this year and is focused on the discovery of neurotrophin mimetics as therapeutic agents against brain diseases.

We hosted Prof. Huan-Xiang Zhou from the University of Chicago, USA, in the group for four months on a sabbatical as the 2018 Romberg Professor (IWR, Heidelberg University). Huan-Xiang is a renowned theoretical biophysicist, and – apart from stimulating lectures and discussions – his visit led to the start of some new joint projects. Katarzyna Świerkula – a master's student working with Dr. Joanna Panecka-Hofman in Warsaw – visited in the summer as an Erasmus student to work on a joint project with Joanna on parasitic pteridine reductase dynamics. Dr. Mehmet Ali Öztürk defended his doctorate on linker histone-nucleosome binding and went on to postdoctoral research at the University of Freiburg. Dr. Ghulam Mustafa left to commence a postdoctoral position at the German Cancer Research Center (DKFZ) in Heidelberg. Jan-Niklas Dohrke conducted his bachelor's thesis in Molecular Biotechnology in the summer, and Patrick Friedrich completed his master's thesis in Physics at the end of the year. Several master's students from Heidelberg University completed internships in the group during the year: Lukas Adam and Jan-Niklas Dohrke (Molecular Biotechnology), Bastian Kister (Systems Biology), Tom Kaufmann (Physics), and Jui-Hung Yuan (Scientific Computing). Dr. Kashif Sadiq was awarded an "Experiment!" grant from the Volkswagen Foundation on "RNA Epicatalysis" to apply simulation approaches in the investigation of how ribonucleoprotein granules affect enzyme kinetics, and he began work on the project in the spring. We continued to work in the Human Brain Project, a large EU-supported infrastructure project whose third phase began in April 2018, in which we are contributing to the development of multiscale molecular simulation techniques and their application to investigate the molecular basis of neurotransmitter signaling.

Moreover, we organized the fourth Biological Diffusion and Brownian Dynamics Brainstorm, BDBDB4, in the Studio Villa Bosch in April 2018. About 60 scientists from around the world attended this very successful meeting (see *Chapter 5.1.1*).

## Computational prediction of drug-target binding kinetics

Traditionally, target-based drug discovery has focused on the identification of compounds that bind to their target – usually a protein – with high affinity. Recently, drug-binding kinetics – which underlie the affinity – have become increasingly seen as important factors for drug efficacy and safety. Consequently, there is great interest in the development of *in silico* methods to predict drug-binding kinetics and to derive QSKRs (Quantitative Structure-Kinetics Relationships). Over the past few years, a range of computational methods has been developed and applied to study drug-binding kinetics (see *[Bruce, 2018]* for a review). However, many of these methods are very computationally intensive and have not been widely tested on drug-like molecules. We are currently using the systematically measured datasets of drug-target binding kinetics from partners in the K4DD (Kinetics for Drug Discovery; *www.k4dd.eu*) project as a basis for developing and validating several new computational approaches to estimating binding kinetic parameters that are designed to be computationally efficient and to reveal QSKRs. Our methods are described along with those of other researchers in KBbox (*kbbox.h-its.org*), a webserver that provides a toolbox of computational methods for studying the kinetics of molecular binding.

In 2018, we published the τRAMD method for estimating relative residence times (τ) from a set of Random Acceleration Molecular Dynamics simulations *[Kokh, 2018]*. We developed the method for ease of use and to enable the computation of long residence times from short molecular

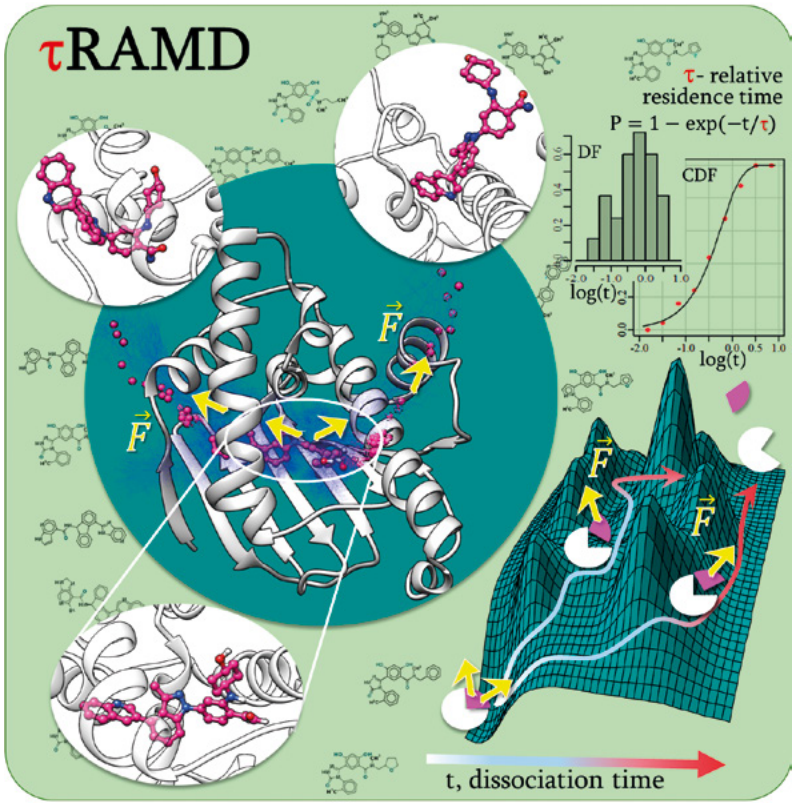



*Figure 34: Illustration of the τRAMD method, which is based on the Random Acceleration Molecular Dynamics technique and designed for the computation of the relative target residence times (τ) of drug-like compounds [Kokh, 2018]. (This image appeared on the July 2018 cover of J. Comput. Theor. Chem.)*

dynamics simulations. The short simulations are performed with the addition of a randomly oriented force on the ligand so that the egress of a ligand from a protein-binding site can be observed within about a nanosecond (i.e., in a much shorter period than the true unbinding time). The software for the τRAMD method is freely available. We obtained high predictive performance for a diverse set of 70 inhibitors of the cancer target heat shock protein 90 by using experimental data measured by collaborators at Merck KGaA (Darmstadt), Sanofi-Aventis Deutschland (Frankfurt am Main), and Sanofi R&D (Vitry-sur-Seine, France). In ongoing work, we have successfully applied the τRAMD method to diverse sets of compounds that bind a range of therapeutically important target proteins. Analysis of dissociation trajectories revealed ligand egress routes and features that affect ligand unbinding rates. These results suggest that τRAMD will be widely applicable as a computationally efficient aid to improving drug residence times during lead optimization.

The determinants of drug-target binding kinetics are currently poorly understood and will – in general – emerge from a complex combination of ligand- and target properties. However, some simple determinants of drug-binding kinetics can be identified. In work conducted along with Stefan Knapp and colleagues at Frankfurt University as well as scientists at Bayer AG (Berlin), we found that the interaction between a halogen moiety on an inhibitor and an aromatic residue in the target protein can significantly increase inhibitor residence time *[Heroven, 2018a,b]*. By characterizing halogen–aromatic π interactions in a set of haspin kinase–inhibitor complexes by means of kinetic, thermodynamic, and structural measurements along with quantum-mechanical- and classical binding energy calculations, we found that inhibitor residence times markedly increased with the size and polarizability of the inhibitor halogen atom.

To assess whether QSKRs can be derived simply from the modeled structures of protein-ligand complexes, we applied the COMparative BINding Energy (COMBINE) analysis method *[Ganotra, 2018]* (see *Figure 35 (i)*). We originally developed COMBINE analysis to estimate the relative binding affinities of a congeneric series of protein inhibitors. Here, we applied COMBINE analysis to derive QSKRs for the dissociation rate constants ($k_{off}$) of diverse sets of inhibitors of heat shock protein 90 (HSP90) and HIV-1 protease (*Figure 35 (ii)*). We derived protein-specific scoring functions by correlating $k_{off}$ rate constants with a subset of weighted interaction energy components determined from the energy-minimized structures of drug–protein complexes. As the QSKRs derived for these sets of chemically diverse compounds have good predictive ability and provide insights into important drug–protein interactions for optimizing $k_{off}$, COMBINE analysis offers a promising approach for binding kinetics-guided lead optimization.

(i)

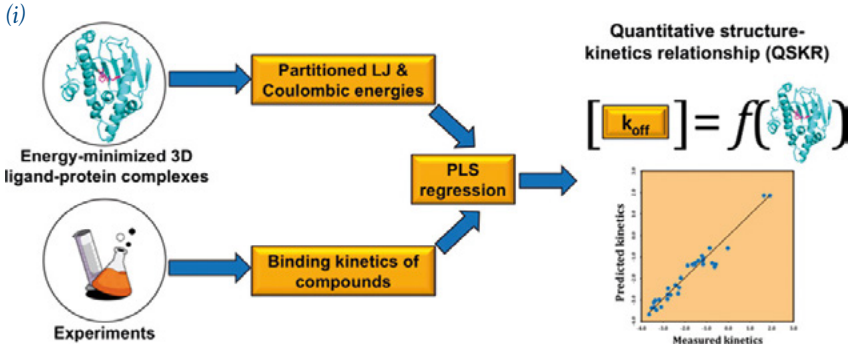
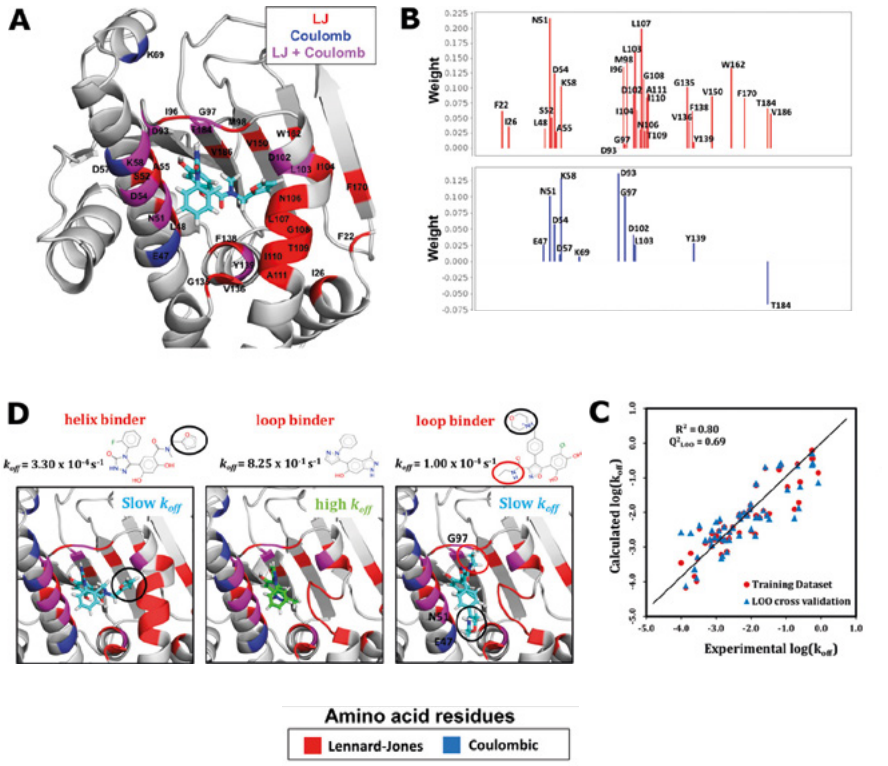



*Figure 35 (i): Schematic diagram illustrating the application of COMBINE analysis to the derivation of QSKRs. A model is derived for a training set of compounds with measured off-rates and modeled structures of protein-ligand complexes by partial least squares projection to latent structures (PLS). The model is then applied to predict the off-rates of a test set of compounds with modeled structures of protein-ligand complexes.*

(ii)




*(ii) COMBINE analysis model for the koff rate constants of HSP90 inhibitors. (A) Ribbon representation of the protein with a representative ligand (cyan) showing the 30 Lennard-Jones (LJ, red) and 12 Coulombic (blue) protein residue–inhibitor interaction energy terms selected for the model. (B) Weights for different LJ and Coulombic interaction energy contributions derived from the PLS analysis. (C) Plot of calculated vs. experimental log(koff) values for the training dataset (R2 = 0.80) and leave-one-out cross-validation (Q2 = 0.69). The straight line corresponds to y = x. (D) Comparison of the binding modes and the key interactions for three compounds. Hydrophobic moieties (shown with a black circle in the left panel) of helix-binders occupy a transient hydrophobic cavity formed by the helix conformation of HSP90 and mediate strong LJ interactions with hydrophobic residues. Most of the loop binders are smaller in size and dissociate faster (middle panel). Some of the more slowly dissociating loop-binders have additional polar moieties (marked with red and black circles in the right panel) that mediate additional electrostatic interactions with the binding-site residues [Ganotra, 2018]. (Reproduced with permission from ACS).*

## Chromatosome formation through linker histone-nucleosome interactions

In eukaryotes, DNA – which is about 2 m long in humans – must be dynamically packed such that specific genes for transcription can be accessed when required. The DNA is wrapped around core histone protein oligomers to form nucleosomes that are connected by linker DNA and that pack together to form chromatin. The linker histones (LH) are proteins that assist in compacting DNA by binding to nucleosomes. The complex of a LH with a nucleosome and its linker DNA arms is known as a chromatosome, the fundamental unit of chromatin structure. There is growing interest in understanding the functional roles of LH proteins and the effects of epigenetic variation on chromatin structuring. While we and others have previously modeled chromatosome structures, the first structures of LH-nucleosome complexes have only recently been determined. We systematically compared five three-dimensional structures of LH-nucleosome complexes determined by a variety of experimental methods *[Öztürk, 2018a]*. Our analysis suggests that the different structures of LH-nucleosome complexes that were revealed in these studies can be reconciled by a paradigm shift away from the concept of "the structure of the chromatosome" and toward "the structural ensemble distributions of individual chromatosomes," in which alternative configurations of LH-nucleosome complex structures can exist. These configurations differ in the position and orientation of the LH globular domain with respect to the nucleosome (see *Figure 36*). Our analysis also highlights the importance of the comprehensive documentation of protein- and DNA sequences and post-translational modifications in future studies of LH-nucleosome complexes.

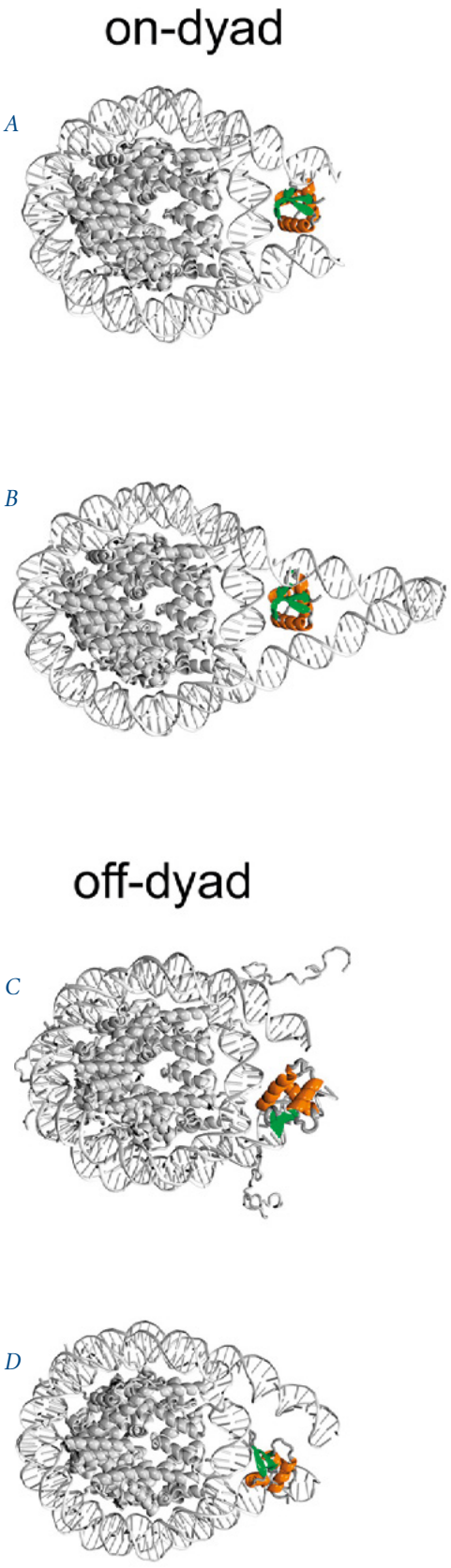



*Figure 36: Four recently determined three-dimensional structures of Linker Histone (LH)-Nucleosome complexes display off-dyad ((A) D. melanogaster and (B) H. sapiens H1.4) and on-dyad ((C) G. gallus H5 and (D) X. laevis H1) positioning of the globular domain (GD) of the LH with respect to the dyad axis of the nucleosome. LH proteins are shown in cartoon representation and colored and labeled according to secondary structure elements. DNA is shown in light gray, and core histones are shown in dark gray [Öztürk, 2018a]. (Reproduced and adapted with permission from Structure, Cell Press.)*

To investigate the possible effects of mutations and post-translational modifications of LHs on LH-nucleosome binding, we modeled variants of the D. melanogaster H1 globular domain (GD) and applied Brownian dynamics simulations to dock the D. melanogaster LH GD and a chicken LH GD to a nucleosome (see *Figure X*). We found that certain single-point mutations and post-translational modifications of the LH proteins can significantly affect chromatosome structure. These findings indicate that even subtle differences in LH sequence can significantly shift the chromatosome structural ensemble and thus have implications for chromatin structure and transcriptional regulation.

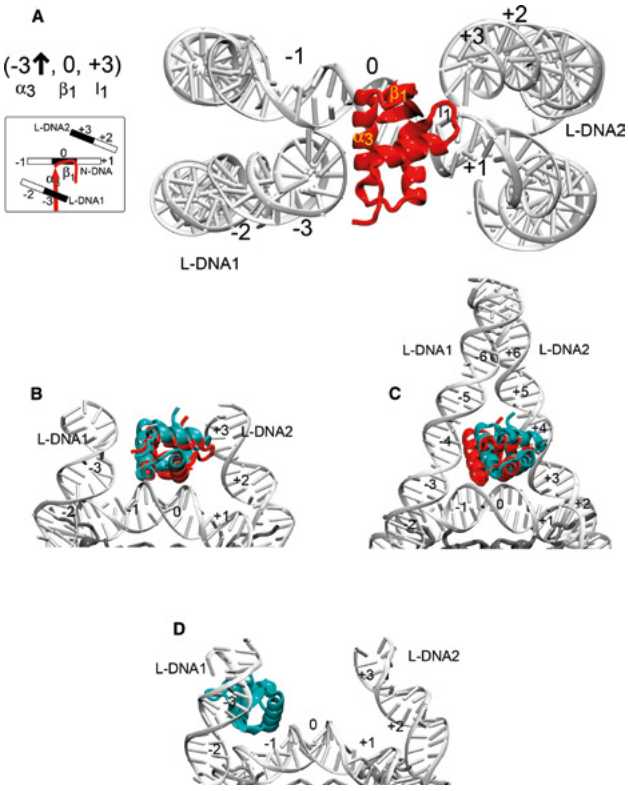



*Figure 37: Representative LH-nucleosome encounter complexes from Brownian Dynamics docking simulations. LH globular domains are shown in cartoon representation and are depicted in red for reference crystal structures and in cyan for docking results. (A) The crystal structure of the complex formed by chicken gH5 bound to a 147 bp Widom 601 DNA sequence nucleosome. The classification of the on-dyad configuration as (–3↑, 0, +3) is illustrated. (B) A representative structure from the largest diffusional encounter-complex cluster from the docking of chicken gH5 to a slightly opened structure of the nucleosome derived by normal mode analysis from the crystal structure shown in (A). The gH5 has a Cα RMSD of 3.6 Å from the crystal structure and the same docked on-dyad configuration. (C) A representative structure from the second largest encounter-complex cluster for Xenopus gH1 docked to the nucleosome structure has a Cα RMSD of 5.5 Å from the crystal structure and the same docked on-dyad configuration. (D) A representative structure from the encounter-complex cluster with the greatest population from docking chicken gH5 to a snapshot representing the average structure from an MD simulation of the nucleosome. The docked configuration is off-dyad [Öztürk, 2018b]. (Reproduced with permission from Biophysical Journal, Cell Press.)*

## Multiresolution modeling of biomolecular complexes

While it is often possible to apply established atomic-detail protein-modeling- and ligand-docking approaches (see, e.g., *[Begolo, 2018]*, *[Sultan, 2018]*), the size and complexity of some biological systems demand the application of multiresolution approaches. Here, we describe the application of atomic-detail- and coarse-grained modeling to study actin filament stability and cytochrome P450-membrane interactions.

In a collaborative project with Ross Douglas and Freddy Frischknecht at the Center for Infectious Diseases at Heidelberg University Medical School, we aimed to understand the determinants of the differences in actin filament stability between mammals and malaria parasites (see *Figure 39 [Douglas, 2018]*). Malaria parasites of the genus Plasmodium move ten times faster through the skin than immune cells, whose job it is to capture such pathogens. Experiments revealed one possible reason: differences between parasites and mammals in the stability and rates of formation and degradation of actin filaments, which are important to the structure and movement of cells. Monomers of the actin protein assemble into long filaments. By making chimeric actin proteins containing sequences from both parasite- and rabbit actins, it was possible to alter the movement and survival of the parasites. To investigate the underlying mechanism, we used atomic-detail models of actin filaments to derive coarse-grained models, with which we performed computer sim-

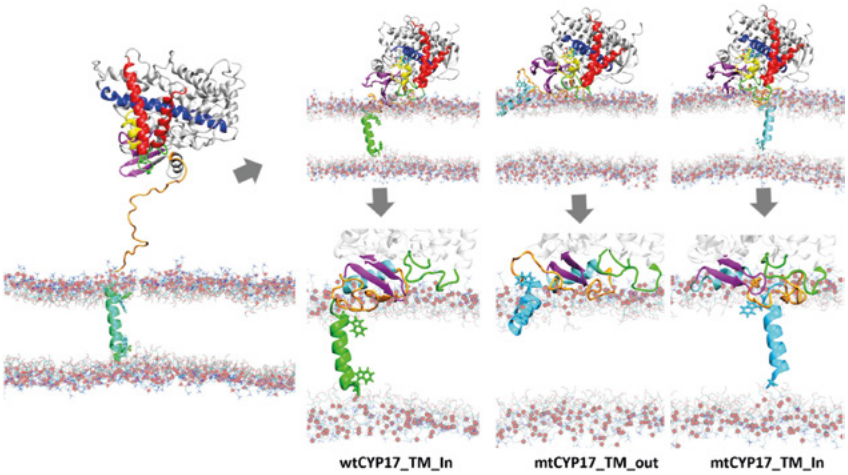



*Figure 38: Positioning of the TM-helix, linker, and the globular domain before and after coarse-grain simulations of CYP 17A1. Left: initial representative superimposed frames for the simulations of the wildtype (green TM-helix) and the mutant (cyan TM-helix) proteins. Right: structures from the final snapshots of three simulations are displayed for the full system (above) and with a close-up showing the position of the TM-helix (below) and differences in the arrangement of the linker (orange) with respect to the F-G loop (green) . The position of the linker on the right would hinder the access of substrates from the membrane to the active site. (Mustafa, G., Nandekar, P., Camp, T., Bruce, N., Gregory, M., Sligar, S., and Wade, RC: Influence of Transmembrane Helix Mutations on Cytochrome P450-Membrane Interactions and Function. Biophys J., 2019, 116, 419. Reproduced with permission from Biophysical Journal, Cell Press, CC-BY Licence.)*

ulations of actin filament dynamics *(see Figure 39, right)*. These simulations allowed us to identify residues at the interface between actin monomers that were important for differences in stability between parasite- and mammalian actin filaments. These findings may provide a basis for discovering anti-malarial agents that selectively target parasite actin and affect either the building or the breakdown of the filament and consequently also the mobility of the parasite.

Human cytochrome P450 (CYP) enzymes play an important role in the metabolism of drugs, steroids, fatty acids, and xenobiotics. Microsomal CYPs are anchored in the endoplasmic reticulum membrane by an N-terminal transmembrane (TM) helix that is connected to the globular catalytic domain by a flexible linker sequence (see *Figure 38*). However, the structural and functional importance of the TM-helix is unclear. We investigated the effect of mutations in the N-terminal TM-helix residues of two human steroidogenic enzymes, CYP 17A1 and CYP 19A1, which are major drug targets for cancer therapy *[Mustafa et al., Biophys J., 2019, 116, 419]*. These mutations were originally introduced to increase the expression of the proteins in Escherichia coli. To investigate the effect of the mutations on protein-membrane interactions and function, we carried out coarse-grained and all-atom molecular dynamics simulations of the CYPs in a phospholipid bilayer. The orientations of the globular domain in the membrane observed in the simulations were confirmed by linear dichroism measurements in a Nanodisc by our collaborators, Stephen Sligar and his colleagues at the University of Illinois, USA. Notably, mutations in the TM-helix of CYP 17A1 led to a gradual drifting of the TM-helix out of the hydrophobic core of the membrane. This instability of the TM-helix could affect interactions with the allosteric redox partner, cytochrome b5, required for CYP 17A1's lyase activity. Furthermore, the simulations revealed that the mutant TM-helix influenced the membrane interactions of the CYP 17A1 globular domain, sometimes obstructing the substrate access tunnel from the membrane to the CYP active site, and thereby indicating a possible effect on enzyme function.

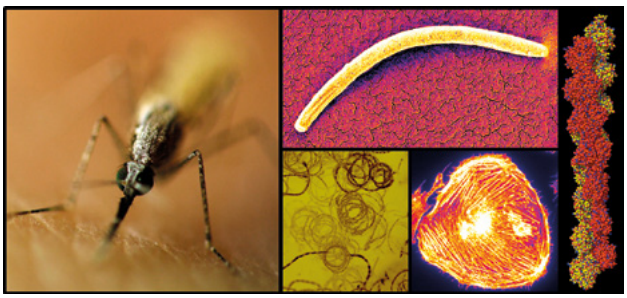
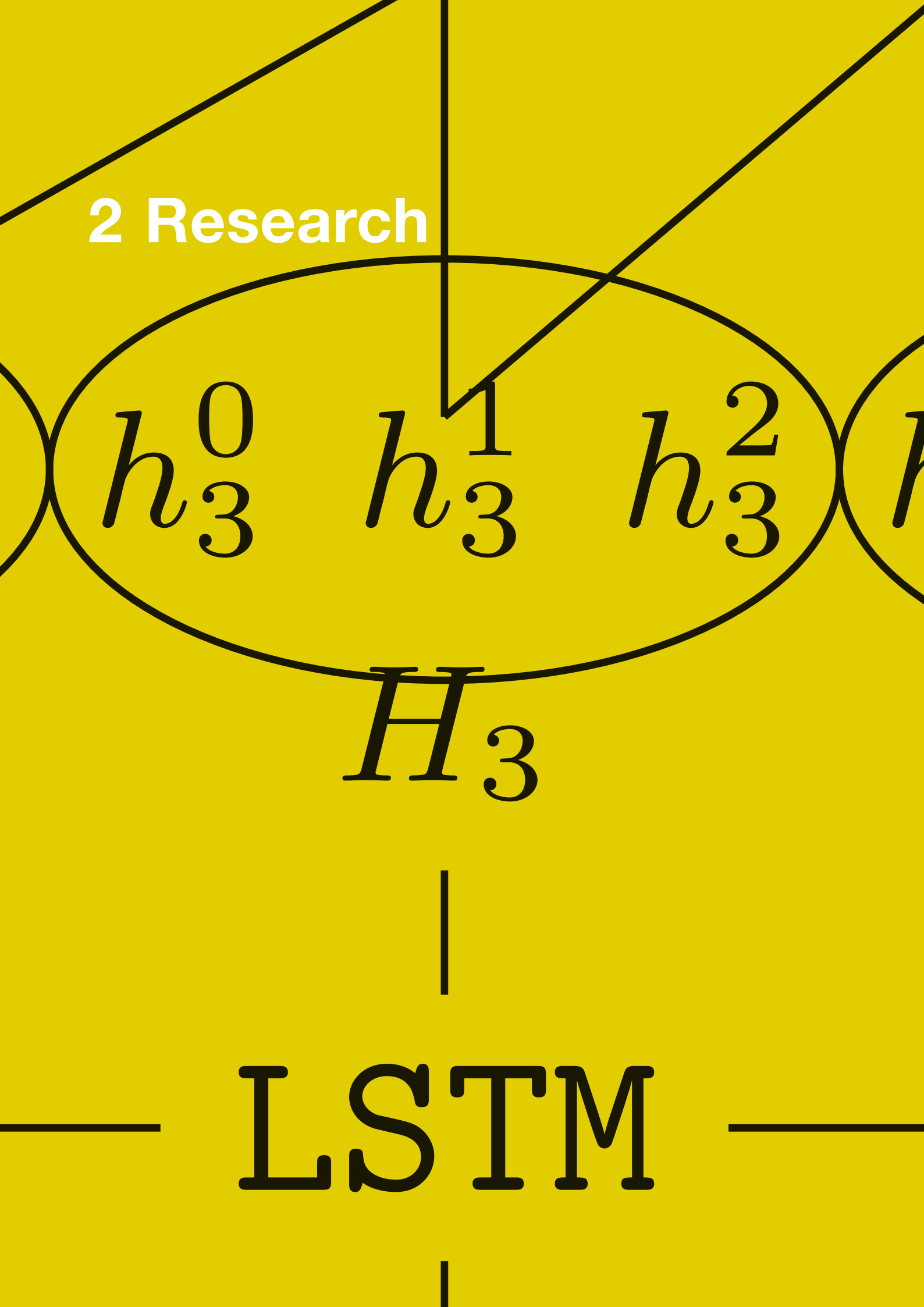
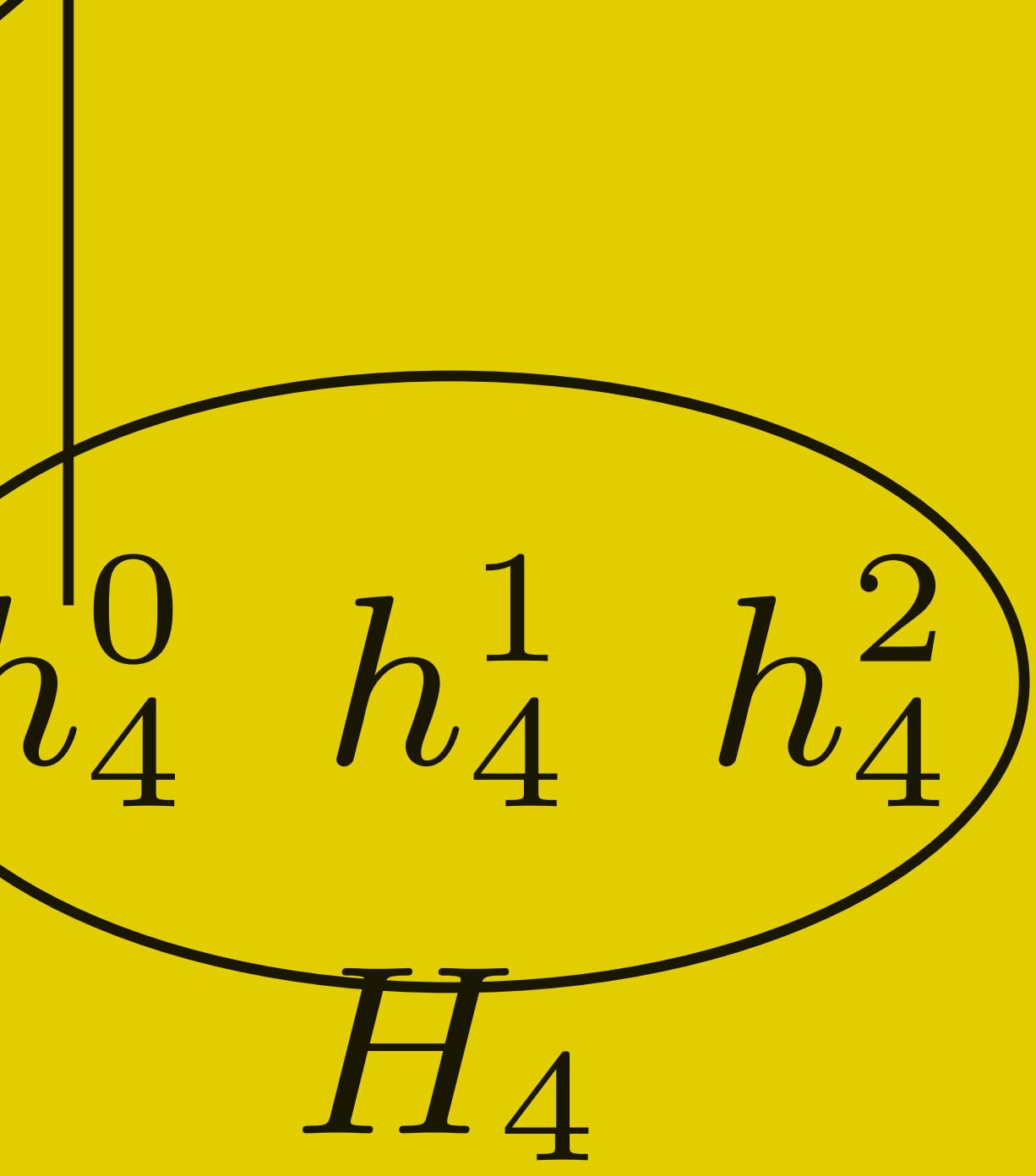


*Figure 39: Mosquitoes (left) inject malaria parasites (top middle) into skin. The parasites move very rapidly (bottom middle left) using a protein that is very similar to the one our cells (lower middle right) use to construct their form and contract: actin (right). Douglas et al. found that certain inter-subunit contacts in the parasite protein are responsible for the different behavior of the actin in parasites [Douglas, 2018]. Image: Heidelberg University Hospital / HITS / ZMBH.*

## 2 Research






2.7

# Natural Language Processing (NLP)

LSTM

## 2.7 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 2018, Daraksha Parveen successfully defended her thesis, entitled "A Graph-based Approach for the Summarization of Scientific Articles", at which point she had already begun a new job as a Research Scientist at Microsoft Research Hyderabad. In 2018, the PhD students Nafise Moosavi, Mohsen Mesgar, and Benjamin Heinzerling submitted their theses, which will be defended in 2019. Nafise and Mohsen have already joined another lab as postdoctoral researchers. New PhD students Kevin Mathews (metonymy resolution), Federico López (entity linking), Sungho Jeon (modeling local coherence), and Mehwish Fatima (automatic summarization of scientific articles) also joined the group. They were accompanied by guest scientist Mahdi Mohseni (abstractive headline summarization) and the intern Ivan Sekulić (NLP and mental health).

2018 proved to be a successful year in terms of publications. Together with her advisers Katja Markert (former Humboldt scholar at HITS, now University of Heidelberg) and Michael Strube, former NLP group member Yufang Hou published her thesis paper on "Unrestricted Bridging Resolution" in the NLP community's top journal, "Computational Linguistics." Both Nafise Moosavi and Mohsen Mesgar presented well-received papers at the top conference EMNLP in Brussels.

Michael Strube co-organized the "Second ACL Workshop on Ethics in NLP", which was held during the NAACL conference in New Orleans, Louisiana, USA (see *Chapter 5.1.3*).

*Natural Language Processing (NLP) ist ein interdisziplinäres Forschungsgebiet, das mit Methoden der Informatik linguistische Fragestellungen 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 2018 verteidigte Daraksha Parveen ihre Dissertation zum Thema „A Graph-based Approach for the Summarization of Scientific Articles". Zu diesem Zeitpunkt arbeitete sie schon als Research Scientist bei Microsoft Research Hyderabad. Im Jahr 2018 hatte die NLP-Gruppe drei weitere Doktoranden, Nafise Moosavi, Mohsen Mesgar und Benjamin Heinzerling, die ihre Dissertationen einreichten. Alle drei werden 2019 verteidigen. Nafise und Mohsen sind schon als Postdoktoranden in anderen Forschungsgruppen tätig. Um sie zu ersetzen kamen Kevin Mathews (Metonymieauflösung), Federico López (Entity Linking), Sungho Jeon (Modellierung lokaler Kohärenz) und Mehwish Fatima (Automatische Zusammenfassung wissenschaftlicher Artikel) als Doktoranden neu in die Gruppe. Zu ihnen gesellten sich der Gastwissenschaftler Mahdi Mohseni (Abstraktive Zusammenfassung) und der Praktikant Ivan Sekulić (NLP und geistige Gesundheit).*

*Wie in den vergangenen Jahren publizierten wir erfolgreich. Das ehemalige NLP-Gruppenmitglied Yufang Hou (Promotion 2016) veröffentlichte mit ihren Betreuern Katja Markert (ehemals Humboldt-Stipendiatin am HITS, jetzt Professorin in Heidelberg) und Michael Strube einen ihre Dissertation zusammenfassenden Artikel in „Computational Linguistics", der wichtigsten Zeitschrift des Fachgebiets. Sowohl Nafise Moosavi als auch Mohsen Mesgar präsentierten gut aufgenommene Arbeiten bei der Top-Konferenz EMNLP. Schließlich ko-organisierte Michael Strube die zweite Auflage des Workshops über Ethik in der NLP während der NAACL-Konferenz in New Orleans, Louisiana, USA.*

## Group Leader

Prof. Dr. Michael Strube

## Staff Members

Benjamin Heinzerling *(since July 2018)*
Alex Judea *(until April 2018)*
Dr. Mark-Christoph Müller
Ivan Sekulić *(since September 2018)*

## Scholarship Holders

Kevin Alex Mathews *(HITS Scholarship, since January 2018)*
Sungho Jeon *(HITS Scholarship, since November 2018)*
Mohsen Mesgar *(HITS Scholarship, until June 2018)*
Nafise Moosavi *(HITS Scholarship, until February 2018)*

## Visiting Scientists

Mehwish Fatima *(since November 2018)*
Benjamin Heinzerling *(PhD Student, Research Training Group AIPHES, until June 2018)*
Federico López *(PhD Student, Research Training Group AIPHES, since April 2018)*
Mahdi Mohseni

## Students

Nadia Arslan *(since October 2018)*
Leo Born *(until April 2018)*
Jason Brockmeyer *(since May 2018)*
Fabian Düker *(since May 2018)*
Raphael Schumann *(until March 2018)*
Julia Suter *(until September 2018)*
Isabell Wolter *(until January 2018)*

## Local Coherence Modeling

**(Mohsen Mesgar)**

Coherence is a crucial factor for a well-written text, and the flow of topics should be smooth and easy to follow. A coherence model has applications in NLP tasks, such as readability assessment and essay scoring. We address the problem of local coherence modeling, which captures text relatedness at the level of sentence-to-sentence transitions.

In *[Mesgar and Strube, 2018]*, we proposed a deep learning approach to solve this problem. Unlike other existing local coherence methods, our model does not depend on the output of any other text-processing tool, such as a coreference resolution system or a syntactic parser. The model considers words in both their sentence- and discourse context.

*Figure 40* depicts the structure of our model. We represent words in a text by word embeddings, which are trained via vast text corpora. Pre-trained word vectors encode semantic relations between words in a vector space. We employ a Recurrent Neural Network (RNN) layer with Long Short-Term Memory (LSTM) cells to combine word embeddings. An RNN transfers a sequence of word embeddings associated with words in a sentence into an array of vectors, which are known as RNN states. These RNN states capture the semantics of a sentence after processing each word in the sentence. We compare the RNN states of each sentence in a text with the RNN states of the immediately preceding sentence. Two adjacent sentences relate to each other via the pair of states that have the maximum similarity. We compute the average vector of the RNN states that connect two adjacent sentences to represent salient information that links the sentences. We then capture variations in salient information among sentences by computing the similarity between any two adjacent average vectors. Afterward, a Convolutional Neural Network (CNN) extracts patterns of changes in salient information through sentences in a text. The output of the CNN layer is a vector that captures patterns of information flow in a text. This vector represents the coherence of the input text.

A coherence vector can be mapped to a coherence score via a dense neural layer or be directly integrated as a feature vector into NLP applications. We applied these approaches to readability assessment and essay-scoring tasks to evaluate our coherence model. On the readability assessment task, our model set a new state of the art. In the essay-scoring task, our model improved the performance of an existing strong essay scorer.

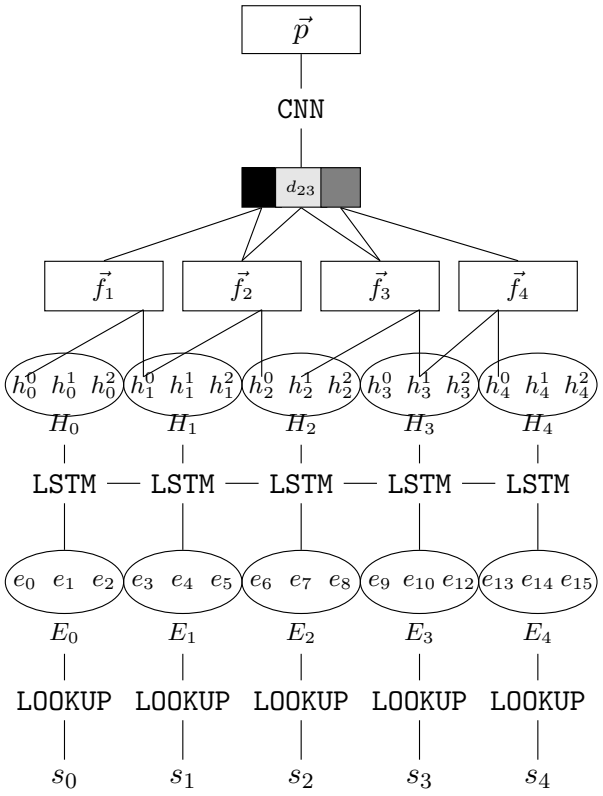



*Figure 40: An illustration of our coherence model. e stands for word embeddings associated with each word in the input text. h depicts the hidden state in the LSTM states of sentences s. The two states in the LTSM states of two subsequent sentences s that have the greatest similarity are selected to connect sentences. Vectors f represent the salient information that relates two subsequent sentences. d represents the similarity between two such vectors f. Different shades of gray reveal different degrees of similarity. The CNN layer encodes patterns of information flow through sentences as vector p.*

## Fine-Grained Entity Typing

**(Federico López)**

"Paul robbed John. He was arrested." What are Paul and John? We could easily say that an entity that robs and is arrested is a person; however, these types are not accurate enough for some Natural Language Processing applications. In order to provide a deeper understanding of texts, it is useful to more precisely determine the semantic classes of entities mentioned in unstructured text with more fine-grained information. We could say that Paul is a criminal and John is a victim since he was robbed.

Entity typing is an information-extraction task that seeks to assign one or more category labels to a mention given its context. Initial approaches aimed to recognize four broad classes (namely *person, location, organization, and other*), and more recent studies have extended the inventory of types to about one hundred. In our model, we propose increasing the set of types to several thousand. These fine-grained types provide much more useful information that can be used in context-sensitive and entity-focused downstream tasks, such as coreference resolution, relation extraction, and question answering.

By applying word embeddings with different neural-network methods, it is possible to project words (or in this case, sentences) together with the categorical labels (types) into a joint vector space. Finally, we can make use of algebraic techniques to operate in this space and look for the closest neighbors of these vectors in order to assign the labels.

The types, such as writer, artist, city, country, government, and organization, are arranged in a hierarchical manner. The hierarchy is given by hypernym/IS-A relations. That is, a writer is a type of artist, which is a type of person. The key idea is to find a representation – or a geometry – that leverages this property. Therefore, as the target vector space, we chose a hyperbolic space (a space of constant negative curvature) in which tree-like structures can be easily modeled. Hence, with this representation, we can simultaneously model the hierarchy in the inventory and the semantic similarity among the types.

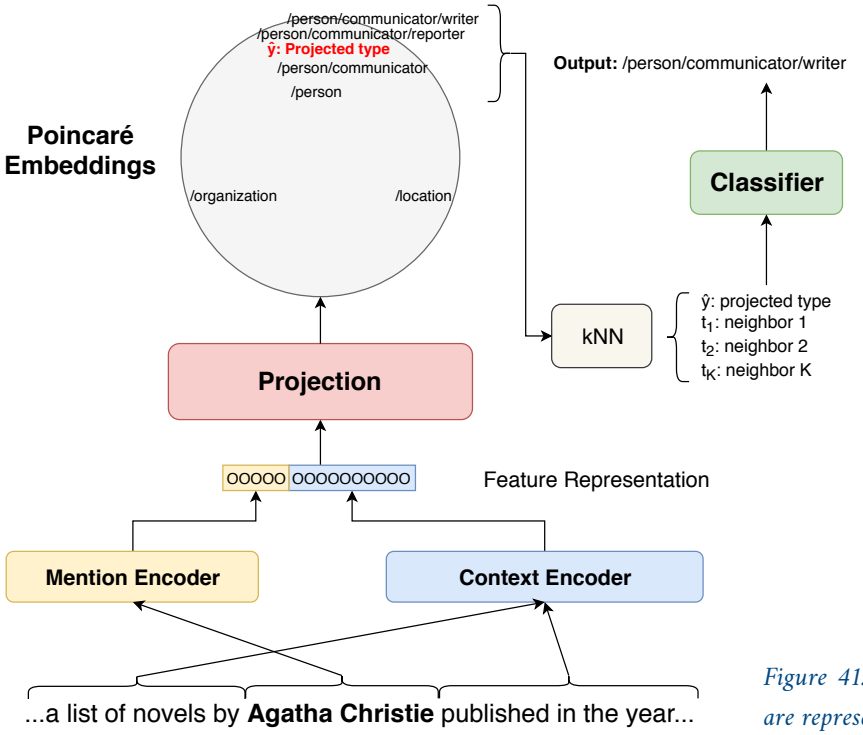



*Figure 41: Model Architecture. Poincaré embeddings are represented within a hyperbolic space.*

## Subword Embeddings in 275 Languages

**(Benjamin Heinzerling)**

This year, the NLP group released a new version of its multilingual resource BPEmb, a collection of subword embeddings in 275 languages. Subword-based approaches to NLP aim to alleviate the problem of unknown words by automatically analyzing a word's constituents based on the intuition that word meaning can be inferred from subwords in many cases *[Heinzerling and Strube, 2018]*. For example, the suffix -shire in the English word *Melfordshire* indicates a location, and the suffix *-osis* in Myxomatosis indicates an illness, as does *-ose* in the German word *Myxomatose*.

We employ an unsupervised subword-segmentation algorithm – Byte Pair Encoding (BPE) – to automatically find subwords for a given word and then train embeddings for each subword using Wikipedia as a background corpus.

The updated version of last year's initial publication includes a Python package *(https://pypi.org/project/bpemb/)* and a website *(https://nlp.h-its.org/bpemb/)* with interactive visualizations, which allow the user to explore the embedding spaces in each language (see *Figure 42*).

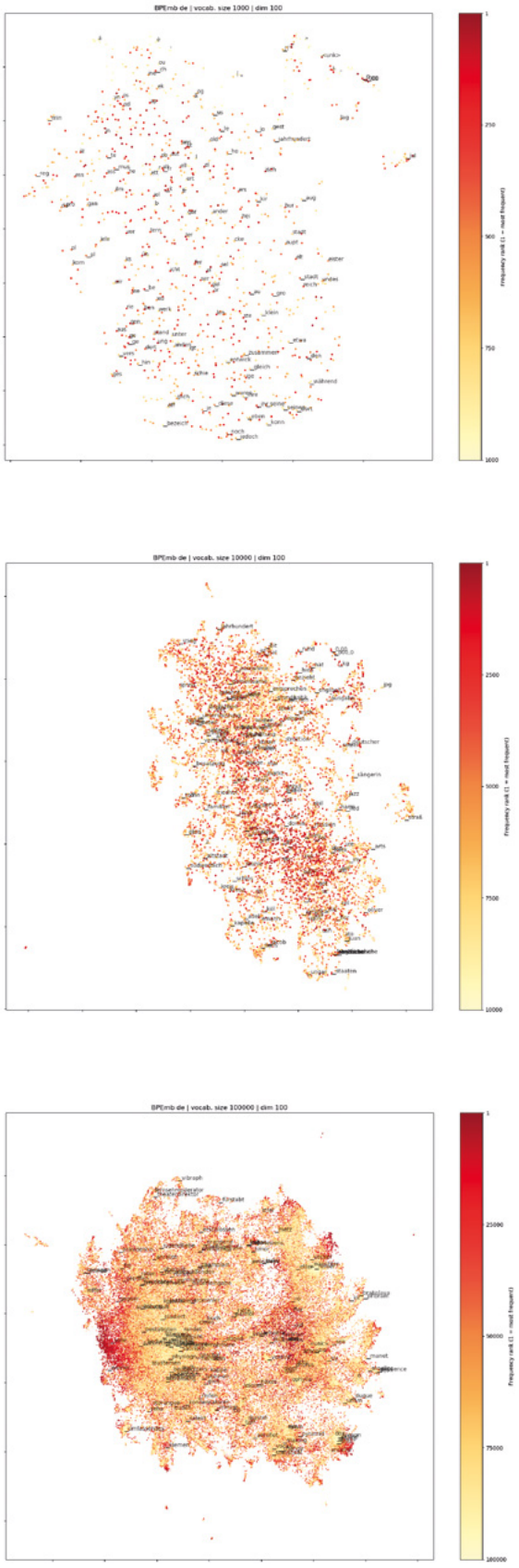


*Figure 42: Byte pair encodings of different vocabulary size. The larger the vocabulary, the more subwords are represented.*

For example, it is possible to search for subwords that are similar to the suffix -osis in English or -ose in German and find that this suffix is close – and thereby similar – to words and subwords related to disease (see *Figure 43*).

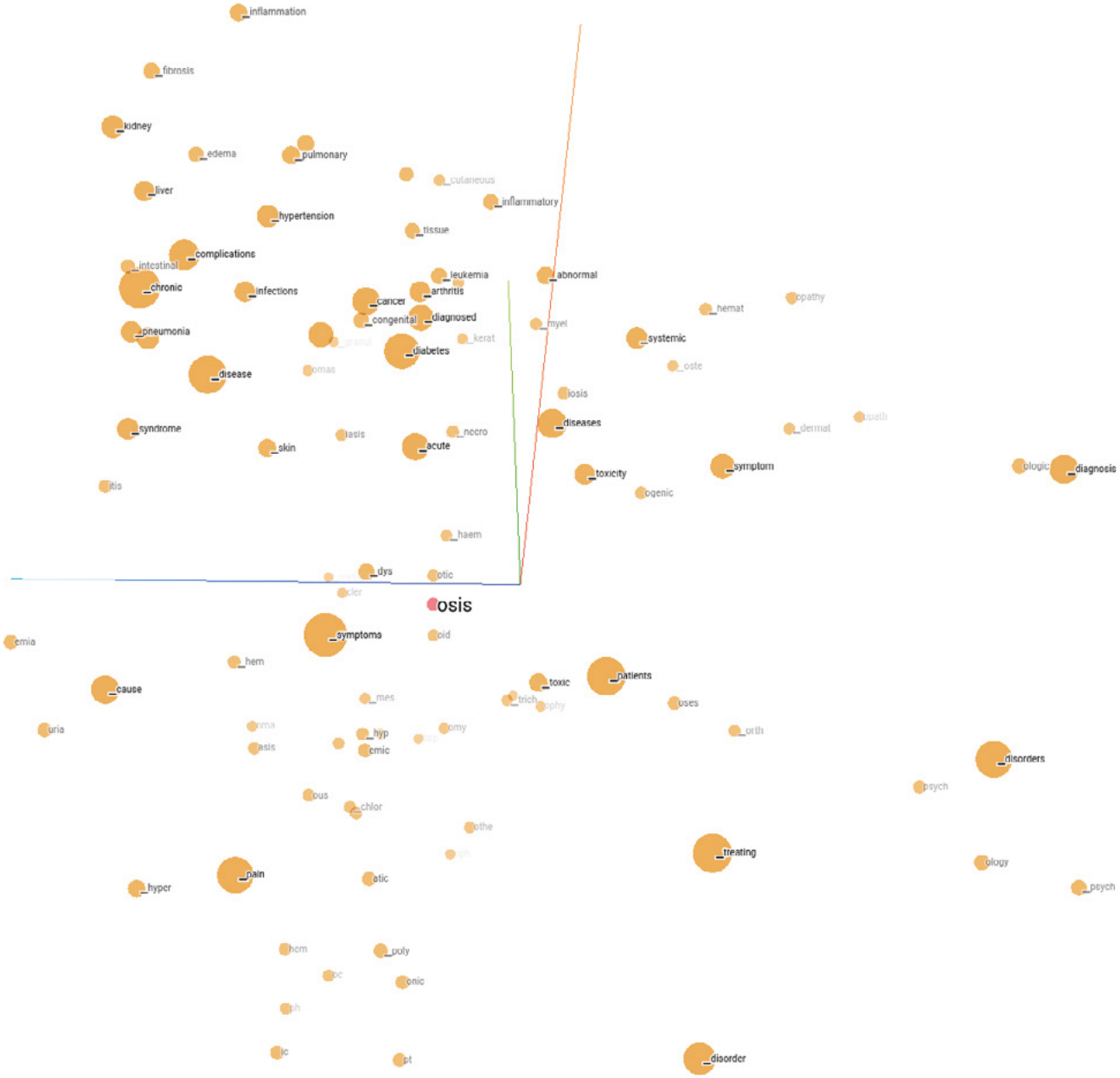



*Figure 43: Semantic space around the suffix -osis. Words and subwords related to diseases ending in -osis are located close to it.*

## Metonymy Resolution using Selectional Preferences

**(Kevin Mathews)**

Language is rich in various phenomena, such as ambiguity, symbolism, rhetoric, connotation, and euphemism. One example of such a phenomenon that is prevalent in written as well as spoken communication is metonymy. For instance, in (1), the words Nietzsche and Marx denote the philosophies developed by Friedrich Nietzsche and Karl Marx, respectively, instead of the individuals themselves. Here, the words Nietzsche and Marx are used metonymically.

**(1) I'm more inclined toward Nietzsche than Marx.**

Metonymy is loosely defined as the phenomenon in which a word in a sentence denotes an entity different from the one it commonly denotes. It is important for Natural Language Processing systems to resolve instances of metonymy.

A strong indicator of the metonymicity of a candidate word (i.e., whether the candidate word is an instance of metonymy or not) is the sequence of words adjacent to it. For instance, in (2), the words preceding the candidate word Tolstoy indicate that the subjects under discussion are related to books and the act of reading. As a result, the word Tolstoy is correctly interpreted as meaning books written by Leo Tolstoy instead of as the person himself.

**(2) I am in fact an omnivorous, voracious reader. I have since read Tolstoy, ...**

The restrictions imposed on the interpretation of a word by its nearby words are called its selectional preferences. In its current, early state, our metonymy resolution system leverages an existing resource that models selectional preferences to predict the metonymicity of a word based on its nearby words and the linguistic relationships between them. Although our system does not yet outperform the state of the art, it provides an insight into the usefulness of selectional preferences for metonymy resolution.

## NLP & Mental Health

**(Ivan Sekulić)**

According to the World Health Organization, up to one-quarter of the adult population in Europe either suffers or has suffered from an episode of some kind of mental disorder. These disorders represent a rising issue in modern society that requires researchers from various fields to gather information and evidence on mental conditions.

In Natural Language Processing, we gear our research toward a deeper understanding of mental health and the development of models for the early detection of various mental disorders. Everyday language reflects basic social and personality processes in addition to our mental health. We search for signals in language that distinguish people who might be affected by a mental disorder from those who are not.

We focus on bipolar disorder, an illness characterized by manic and depressive episodes that affects more than 60 million people worldwide. In *[Sekulić et al., 2018]*, we presented a preliminary study on bipolar disorder prediction from user-generated text on Reddit, which relies on users' self-reported labels. Benchmark classifiers outperformed the baselines and reached an accuracy of above 86%. Feature analysis demonstrated interesting differences in language use between users with bipolar disorder and the control group, such as the use of personal pronouns ('I,' 'you,' etc.) and emotion-expressive words. Future work will explore Deep Learning models for predicting the disorder and compare it to other mood disorders (e.g., clinical depression) with the goal of finding linguistic differences during manic and depressive episodes.

## Semantic Author Name Disambiguation

### (Mark-Christoph Müller)

Research on author name disambiguation at HITS was conducted as part of the "Scalable Author Disambiguation for Bibliographic Databases" (SCAD) project in cooperation with DBLP and zbMATH and was funded by the Leibniz-Gemeinschaft. 2018 was the third and final year of the project.

Our research focused on two areas: First, we conducted a follow-up study to our earlier work on semantics-based author name disambiguation [Müller, 2018]. The work addressed some of the previous work's open issues by performing a more fine-grained analysis of the effect of word-embedding-based word semantics on author name disambiguation, which included a reduction of the size of the dataset and a more stringent evaluation. We also employed a simpler machine-learning model to be able to better identify the different contributions of co-author overlap, string similarity, and semantic similarity. More importantly, we applied a novel, light-weight approach to detect high semantic similarity among publication titles. While still using word embeddings, the approach is based on similarity at the level of word pairs as opposed to that at the level of *averaged word sequences*.

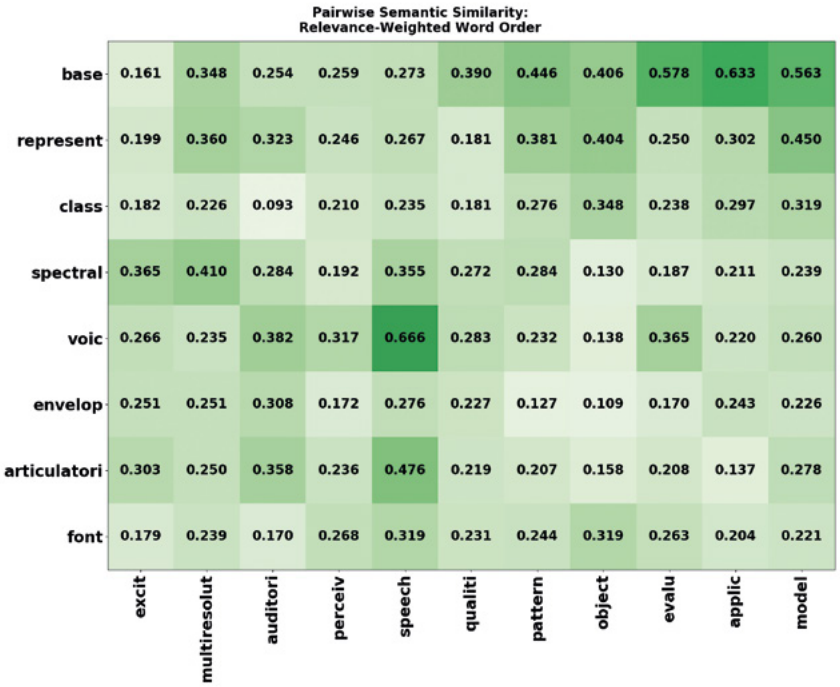



*Figure 44: Pairwise semantic similarity of words in two publication titles.*

The following is a pair of publication titles from our dataset:

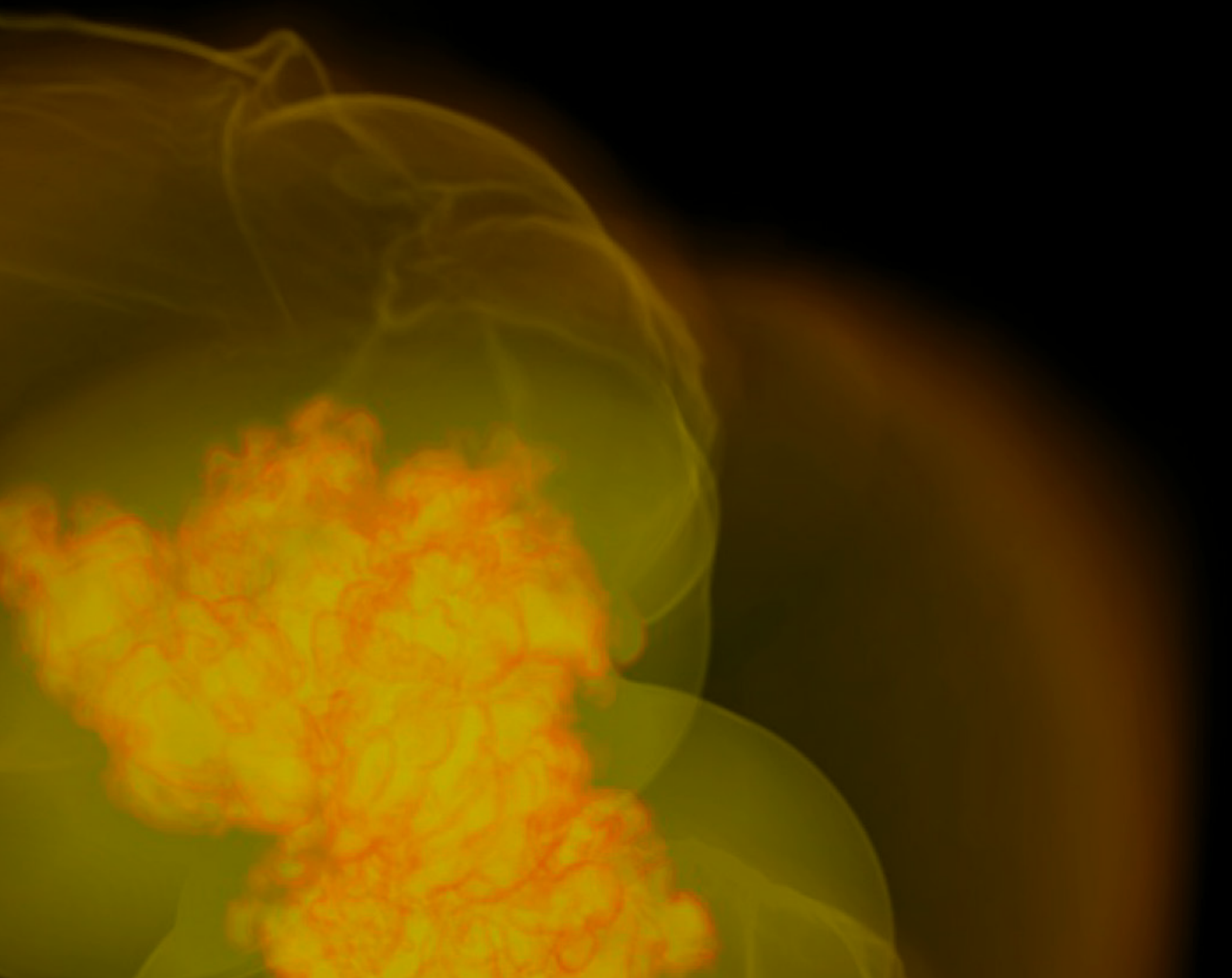
"Articulatory class based spectral envelope representation for voice fonts."

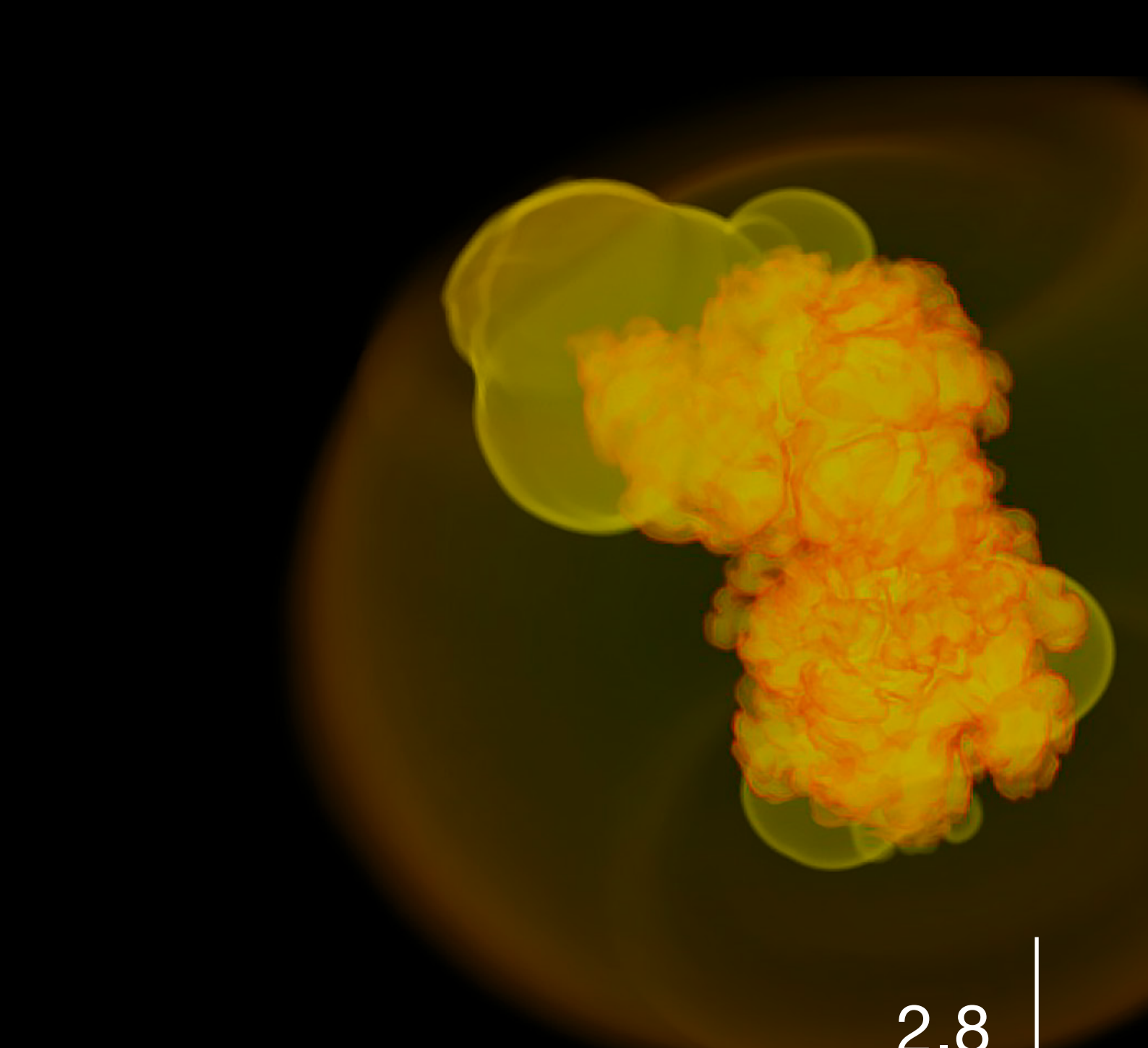
"A multiresolution model of auditory excitation pattern and its application to objective evaluation of perceived speech quality."

Pairwise computation of the semantic similarity of these two titles produces a matrix that can be visualized as a heat map. In order to reduce the effect of irrelevant words, title words along both axes were sorted according to their IDF weight. This process caused irrelevant yet potentially highly similar words (e.g., 'model' and 'base') to be clustered in the upper right area of the matrix, while the more important relations (e.g., 'voice' and 'speech') moved to the lower left area.

In our second research area, we developed, demonstrated *[Müller and Strube 2018]*, and made available a versatile tool for word embedding access. The tool was originally inspired by performance and flexibility considerations in the SCAD project and continues to be of instrumental importance there. However, it soon became clear that the tool addresses an issue that exists in many current NLP approaches, which are characterized by massive use of word embeddings as the de-facto standard method for the representation of word semantics.

# 2 Research





# 2.8 Physics of Stellar Objects (PSO)

## 2.8
## Physics of Stellar Objects (PSO)

"We are stardust" – the matter we are made of is largely the result of processing the primordial material formed during the Big Bang. All heavier elements originate from nucleosynthesis in stars and gigantic stellar explosions. How this material formed and how it is 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 and 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 and 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. 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 picture: Markus Kromer*

## Group Leader

Prof. Dr. Friedrich Röpke

## Staff Members

Dr. Andreas Bauswein *(until June 2018)*
Dr. Sebastian Ohlmann *(until April 2018)*

## Scholarship Holders

Leonhard Horst
Christian Sand *(since October 2018)*

## Visiting Scientists

Dr. Markus Kromer *(Gliese Fellowship, until September 2018)*
Dr. Fabian Schneider *(Gliese Fellowship, since October 2018)*
Florian Lach
Sabrina Gronow
Theodoros Soultanis *(since September 2018)*

## Student

Manuel Kramer *(since October 2018)*

## Theoretical modeling of supernova explosions in the PSO group

Supernovae are the brightest cosmic explosions and thus constitute prime targets of astronomy. Supernova observations – most prominently those performed by Tycho Brahe and Johannes Kepler – even helped to shape astronomy as a modern science. Contrary to what the name suggests, supernovae do not result from the birth of new stars but rather from the cataclysmic explosions of stellar objects that have reached the end of their evolution. The implications of supernovae are far-reaching for many fields of astrophysics: They are main sites of cosmic nucleosynthesis (a large fraction of the chemical elements forming our world were in fact produced in supernova explosions), they drive shock waves into the gas in galaxies and thereby trigger star formation, some of them form neutron stars or black holes, and a sub-class of them has been used to measure cosmic distances. The latter application led to the spectacular and unexpected discovery of the accelerated expansion of the Universe (Nobel Prize in Physics 2011), which would not have been expected if the Universe were filled with normal matter and energy. Therefore, an enigmatic "dark energy" pushes it apart and must dominate its contents today. Most importantly for supernova researchers, however, the physical mechanisms of supernovae are a fascinating field of research that lies at the interface of fluid dynamics, nuclear physics, turbulence theory, combustion physics, neutrino physics, radiation transport, general relativity, and gravitational-wave physics.

This astrophysical significance of supernovae has been a driver for many astronomical campaigns that observe various aspects of supernovae, which has led to the collection of a wealth of data. Ultimately, astrophysicists strive for an understanding of the physical nature of these events. Unveiling their mechanism and interpreting the observed data requires the construction of theoretical models of supernova explosions. Given the complexity and great variety of physical effects at play, this ultimate goal cannot be achieved via a "pen-and-paper" approach. The sets of equations underlying the modeling can usually only be solved with numerical techniques and by performing realistic simulations in three spatial dimensions. Accounting for all relevant physics is extremely challenging in terms of the required super-computational resources.

This challenge is addressed by one of the main branches of research in the Physics of Stellar Objects group at HITS. By developing new modeling approaches, implementing them in numerical codes, and performing simulations at high-performance computer facilities, we aim to improve the physical understanding of supernovae.

This improved understanding covers various elements. One focus is on providing a better description of the explosion itself. The standard model for one class of supernovae – those of Type Ia – is that of a thermonuclear explosion in a compact astrophysical object: a white dwarf star. The resulting event is extremely bright. Observationally, a surprisingly high degree of homogeneity was found, which led to the application of Type Ia supernovae as distance indicators. But what drives their extreme brightness and the powerful explosions? White dwarf stars concentrate the mass of the Sun into the volume of a planet. They usually consist of carbon and oxygen. Under the extreme conditions in white dwarfs, this material can explosively "burn" in thermonuclear reactions into heavier elements. A main product of this process is $^{56}$Ni, an unstable isotope that decays radioactively to $^{56}$Co and finally to $^{56}$Fe. This process renders Type Ia supernovae the largest source of iron in the Universe and explains the extreme brightness of the events. The released radioactivity heats the ejected material, which radiates primarily in the optical wavebands for days, weeks, and months. Beyond this general concept, however, many of the details of Type Ia supernova explosions remain unclear. An important reason for this uncertainty is that the progenitors of these events have not yet been observed. We therefore do not know how the white dwarf star reaches conditions for triggering explosive burning. During the explosion, which lasts for a few seconds only, the density of the stellar material is very high such that it remains opaque to radiation. The actual explosion can thus not be observed directly. The explosive burning propagates over the white dwarf star as a combustion wave, but the details of its na-

ture remain uncertain. The PSO group therefore performs large-scale three-dimensional hydrodynamic simulations of the explosion process with an assumed progenitor structure on high-performance computers to explore different possibilities. An exploration of parameters and models for physical processes naturally has to test out the extremes. An example for such simulations is given in Section 1.

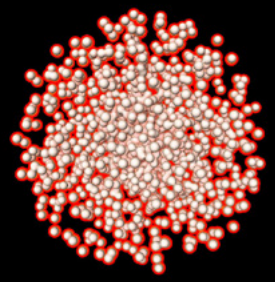
Without validation, however, such models are of little use for improving our understanding of supernovae. Only a comparison with astronomical data reveals whether the hypothesized physical explosion mechanism is actually realized in nature. For this comparison, we need to predict observables from the explosion models. One important method of achieving this prediction is by means of radiation transfer calculations. The propagation of electromagnetic radiation from its generation (by the radioactive decay of $^{56}$Ni in the case of Type Ia supernovae) is followed in simulations, where its interaction with matter from the explosion ejecta is modeled. Such calculations are extremely demanding due to the complex structure of the involved atoms and the general challenge of modeling a radiation field consistently. They were performed by the PSO group to predict observables from modeled Type Ia supernovae, and we provide an example in Section 1. Another example is discussed in Section 2 in the context of so-called pair-instability supernovae, which are conjectured to terminate the evolution of extremely massive stars and again test an extreme in the parameter space of stellar explosions. Here, the focus is on determining potential observational features to check whether such events really occur.

## Thermonuclear explosions of rapidly differentially rotating white dwarfs

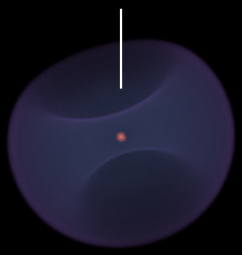
White dwarfs – the objects that can explode as Type Ia supernovae – are truly remarkable stars. In contrast to other, common stars, which are stabilized against their own gravitational pull by a pressure gradient created in nuclear burning, white dwarf stars are stabilized by a quantum mechanical effect: the Fermi pressure of degenerate electrons. This stabilization at very high densities is eternal and does not depend on a finite energy source, such as nuclear fusion. However, there is also a limit to this stabilization: In 1931, Chandraskekhar found that relativistic effects begin to dominate the stellar material beyond 1.4 solar masses, making it unstable collapse.

For a long time, the leading model of Type Ia supernovae assumed that the white dwarf accretes material from a companion star and sets off an explosion when it approaches this Chandrasekhar limit. In recent years, however, alternatives to this "standard model" have been proposed, a development to which the PSO group contributed. However, the 1.4 solar masses can still be considered the upper limit of material in the exploding star for very fundamental physical reasons.

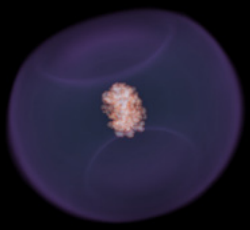
This picture was challenged by puzzling observations of extremely bright ("superluminous") Type Ia supernovae that were first published by Howell in 2006 and later found in about a dozen cases. Explaining their brightness required close to or even more than a Chandrasekhar mass of 56Ni to be synthesized in the thermonuclear explosion, which seems impossible given that other material is also seen in the spectra, meaning that the total mass must exceed this critical limit. There are, however, ways to increase the mass of a white dwarf star beyond the canonical 1.4 solar masses without losing stability. One way is via differential rotation. In this case, centrifugal forces add to the stabilization of the star against gravity, and its mass can reach above twice that of the sun. Such a star is not spherical; rather, it is highly deformed.



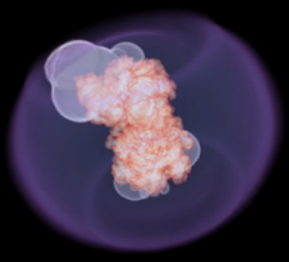
1600 ignition kernels
in sphere with $R = 180$ km



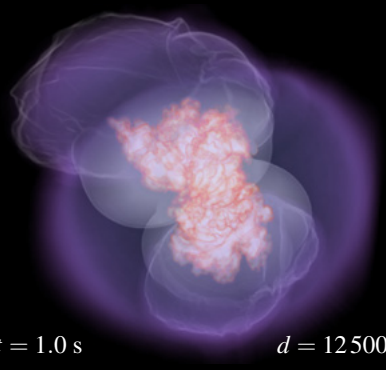
$t = 0$ s $d = 10\,500$ km



$t = 0.6$ s $d = 9\,700$ km



$t = 0.9$ s $d = 9\,200$ km



$t = 1.0$ s $d = 12\,500$ km

Together with an international team of scientists, the PSO group performed simulations of thermonuclear explosions of such white dwarf configurations *[Fink , 2018]*, an example of which is shown in *Figure 45*. Here, the thermonuclear combustion wave is assumed to ignite at the center of the deformed star in multiple sparks (1st panel of *Fig. 45*) and propagates outward as a flame with subsonic velocities. The propagation of such flames is normally accelerated by hydrodynamic instabilities and turbulence. In conventional models for Type Ia supernovae, this phase leads to a significant expansion of the star. In the case of strong rotation, however, the accelerating instabilities are suppressed in the equatorial direction by angular momentum conservation. Therefore, the flame spreads faster toward the poles of the star (2nd panel of *Fig. 45*). Due to this anisotropic combustion, the overall energy generation in the burning – and consequently, the expansion of the star – is far less efficient than in isotropic models. In a later phase, the combustion wave undergoes a transition to a supersonic detonation (white surface in the 3rd and 4th panels of *Fig. 45*). Because the density of the rather unexpanded star remains very high, this detonation burns the material in large parts into $^{56}$Ni and produces a very bright event.

This result is confirmed by radiative transfer calculations based on our explosion models. The predicted events reach the brightness of the extreme class of Type Ia supernovae. Due to the high energy release, however, the ejecta move very fast, which is reflected in the spectra determined via radiative transfer calculations. However, the velocities inferred from the observed superluminous supernovae are rather low.

We therefore find that differentially rotating white dwarfs as progenitors of thermonuclear supernovae cannot explain the observed superluminous events. These are not produced by explosions synthesizing more than 1.4 solar masses of heavy nuclei, and our results thereby contradict the hypothesis of "super-Chandrasekhar mass explosions". Other effects – such as the interaction of the ejecta with circumstellar material – may lead to the extreme brightness in such objects. The studied models, however, explore an extreme part of the parameter space and make specific predictions. Because these models find no match among currently known Type Ia supernovae, we conclude that progenitor stars of these astrophysical explosions are very unlikely to feature strong differential rotation. This conclusion, in turn, has implications for the mechanism of mass transfer to the supernova progenitor star and the mechanism of angular momentum transport in it.

Fig. 45: Simulation of the thermonuclear explosion of a differentially rotating white dwarf star. Figure from [Fink, 2018] . (Reproduced with permission from Astronomy & Astrophysics, © ESO.)

## OGLE14-073 – A promising pair-instability supernova candidate

Very different instability is expected to occur in stars with masses well above one hundred times that of the Sun and that contain few heavy elements (astronomically called "metals"). Under conditions reached in their centers, pairs of electrons and their antiparticles (positrons) can form, which leads to a sudden drop in pressure that supports the star against gravitational collapse. The compression of the material accelerates nuclear burning in a thermonuclear runaway process that leads to an explosion of the entire massive star. This scenario suffers from two missing observations: (i) that of suitable progenitor stars and (ii) that of the actual explosions.

Although current knowledge suggests that sufficiently massive stars can form, it is unclear whether they remain massive enough in their evolution to undergo such a Pair Instability Supernova (PISN). Stellar winds are likely to remove substantial parts of the mass unless the material has a low metallicity and winds are inefficient. Deciding on the existence of suitable progenitor stars (i) is difficult, and it therefore appears promising to search for their bright explosions (ii). The problem with this approach is that firm theoretical predictions of what to look for remain missing.

In collaboration with an international team, PSO researcher Dr. Markus Kromer worked on finding a clearer picture of what to expect from PISN events *[Kozyreva, 2018]*. The ARTIS radiative transfer code enabled the prediction of detailed spectra from explosion simulations (see *Fig. 46*), which reproduced the main features of an observed PISN candidate: the bright Type II supernova OGLE14-073. In combination with other observational signatures, satisfactory fits to OGLE14-073 for a pair-instability supernova in a massive progenitor star with an initial mass of 150 solar masses and low metallicity (Z=0.001) were found, thereby rendering OGLE14-073 one of the most promising candidates for this theoretically predicted supernova explosion mechanism.

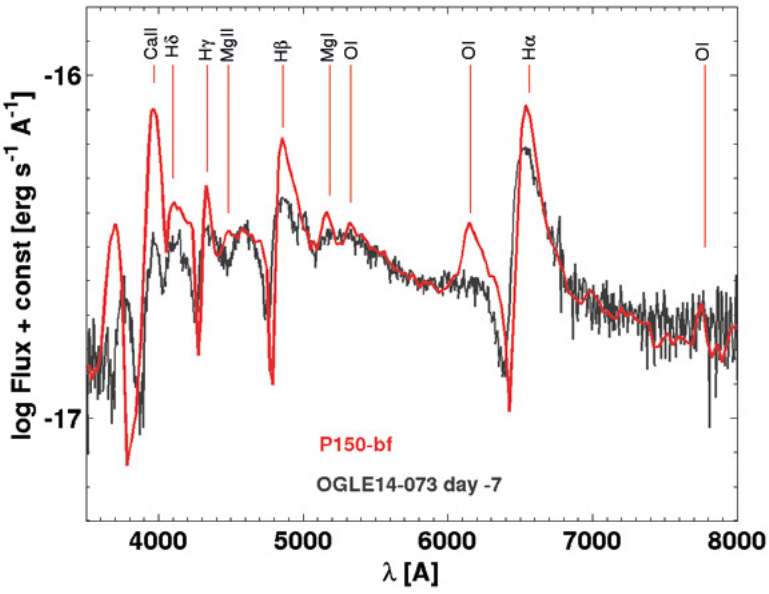



*Fig. 46: Predicted spectra for a pair instability supernova (red) compared with observational data from OGLE14-073 (black). Figure from [Kozyreva, 2018].*

# 2 Research

# 2.9 Scientific Computing (SCO)

## 2.9
## Scientific Computing Group (SCO)

The Scientific Computing Group 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
- 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: The group changed its name effective January 1st, 2019. To better reflect its main research area, it is now called the "Computational Molecular Evolution" (CME) group.

*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: Die Gruppe ändert zum 1. Januar 2019 ihren Namen: Sie heißt künftig – passend zu ihren Hauptforschungsfeldern – „Computational Molecular Evolution" (CME).*

## Group Leader

Prof. Dr. Alexandros Stamatakis

## Staff Members

Pierre Barbera
Benjamin Bettisworth *(since September 2018)*
Lucas Czech
Dr. Alexey Kozlov *(staff scientist)*
Benoit Morel

## Scholarship Holder

Sarah Lutteropp *(HITS Scholarship)*

## Visiting Scientists

Aggelos Koropoulis *(since October 2018)*
Josefin Stiller *(November 2018)*

## Students

Ivo Baar *(since May 2018)*
Rudolf Biczok *(until July 2018)*
Paula Breitling *(since October 2018)*
Sebastian Giesse *(until July 2018)*
Fernando Ramirez *(until February 2018)*
Tobias Ribizel *(until August 2018)*
Axel Trefzer *(until April 2018)*
Johanna Wegener *(since December 2018)*
Adrian Zapletal *(since May 2018)*

## What happened at the lab in 2018?

In the winter of 2017/2018, Alexis, Benoit, 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 highly 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/Winter17_18.pdf). Enrollment in our course at KIT further increased over the winter semester of 2018/19 to approximately 35 students.

During the summer semester of 2018, we were involved in several teaching activities. We again taught our main seminar, "Hot Topics in Bioinformatics." In collaboration with Prof. Sanders at the Institute for Theoretical Informatics, we also offered a bioinformatics programming practical trying to address a challenging data distribution problem with our students (paper in preparation).

Alexis received a teaching award from the dean of the computer science faculty for the programming practicum held in the summer term of 2017, marking the second time he has received an award for this specific practicum and the third time overall that he has received a teaching award at KIT.

2018 was also a very important year for Alexey Kozlov, who successfully defended his PhD thesis and fortunately decided to remain at HITS as an SCO staff scientist.

Sebastian Giesse successfully completed his master's thesis on "Inferring Species Trees by Minimizing Quartet-based Uncertainty" at the Department of Computer Science at KIT. Linda Rülicke also finished her master's thesis on "Design and Implementation of I/O-Efficient Taxa Quartets Counting in the Context of Phylogenetic Analysis" at the University of Frankfurt (co-supervised by Prof. Dr. U. Meyer).

In 2018, a total of four KIT master's students joined the lab either as research assistants or to work on their master's theses. Apart from keeping up the constant flow of students from KIT, we were also pleased that two of the four students were females.

Toward the end of the year, we began hosting Aggelos Koropoulis – a master's student in computer science from the University of Crete – as an intern working on hardware aspects of phylogenetic inference. His main supervisor is former SCO postdoc Pavlos Pavlidis, who now leads a junior research group in Crete. In November, we also hosted Josefin Stiller, a postdoc from Copenhagen, who is coordinating the data analyses for the project on 350 bird genomes, which is a follow-up project to our 2014 Science paper on the evolution of approximately 50 bird genomes.

One highlight that recurred in 2018 was the summer school on Computational Molecular Evolution, for which Alexis again served as the main organizer. The course took place for the 10th time this year at the Hellenic Center for Marine Research in Heraklion, Crete. Former lab members Paschalia Kapli and Pavlos Pavlidis as well as current PhD student Pierre Barbera contributed substantially to the success of the course as teaching assistants (see *Chapter 5.1.2*).

Finally, the year was marked by the creation and presentation of the full proposal entitled "Algorithm Engineering for the Scalability Challenge" for the German Excellence Initiative by colleagues stemming mostly from KIT's Department of Computer Science. Unfortunately, this proposal was ultimately rejected, but the core team of PIs is currently investigating alternative avenues for funding the project. A positive side-effect of the failed proposal is that Alexis established substantially closer research links with the computer science department at KIT.

## 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, programs that disentangle 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. Today, 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 and algorithms for phylogenetic and population-genetic analyses.

Following the introduction of so-called short-read sequencing machines (machines used by biologists in the wet lab to extract DNA data from organisms), which can generate over 10,000,000 short DNA fragments (each containing between 30 and 400 DNA characters), the community as a whole is now 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 (see https://www.genome.gov/sequencingcosts/).

We are thus facing a scalability challenge – that is, we are constantly trying to catch up with the data avalanche and make molecular data analysis tools more scalable with respect to dataset sizes. At the same time, we also want to implement more complex and hence more realistic and compute-intensive models of evolution.

Inversely, we also need to devise methods for reducing the number of data – that is, to assemble data subsets such that they can be analyzed in a reasonable amount time while maintaining a high level of accuracy. Our PhD student Sarah is currently investigating such methods. For instance, computing only one phylogenetic tree from the aforementioned dataset of 350 bird genomes would require 1.2 million core hours and approximately 30TB of main memory. While computing such a tree is still feasible in principle, we are currently attempting to devise novel methods that avoid such excessive resource requirements.

In this context, we are also investigating novel data compression methods. For example, together with our student programmer Axel Trefzer, we have developed algorithms for compressing phylogenetic trees as generated by Bayesian Markov Chain Monte Carlo tools for phylogenetic inference (see paper preprint at https://www.biorxiv.org/content/early/2018/10/11/440644).

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 constantly changing. This output requires the continuous development of new algorithms and tools to filter, puzzle together, and analyze these molecular data.

Phylogenetic trees (evolutionary histories of species) are important in numerous domains of biological and medical research. Programs for tree reconstruction that have been 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 – inter alia – 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.

## ParGenes – reconstructing gene trees in parallel

Given the genomic data (or a subset thereof) for a certain number of species under study, there are two main approaches to reconstructing a phylogenetic tree provided that the individual genes have already been identified.

First, it is possible to concatenate all these genes to build a so-called supermatrix and then reconstruct a phylogeny on this supermatrix by deploying – for example – our RAxML-NG tool on a supercomputer.

Second, it is possible to initially reconstruct individual per-gene phylogenies for each gene separately and independently and to subsequently use all gene trees to reconstruct a species tree via so-called gene tree/species tree reconciliation methods.

We have recently become interested in gene tree/species tree reconciliation methods as they can model additional biological phenomena that occur during the course of evolution, such as gene loss, gene duplication, and lateral gene transfer. Moreover, most available reconciliation tools are neither efficient nor particularly scalable such that there is great potential for making substantial contributions to this area.

It is important to note that the individual gene trees are frequently not identical to the species tree due to the presence of gene duplication, gene loss, and lateral gene transfer. The main goal of gene tree/species tree reconciliation approaches is thus to explain and resolve these differences.

While investigating the topic, we observed the lack of a fundamental data pre-processing tool required for executing reconciliation analyses. There was no convenient or efficient tool available for inferring per-gene phylogenies, and we therefore decided to implement such a tool prior to addressing the development of novel reconciliation algorithms. While the problem might appear trivial (embarrassingly parallel) at first sight, it is not, for per-gene tree inference times can vary substantially depending on the length of the gene as well as on the number of species that have this gene or for which we have data on the particular gene. In other words, per-gene dataset sizes vary both in terms of sequence length and the number of sequences contained therein. In turn, this variation impacts the per-gene inference times as well as the number of cores that can be efficiently utilized in parallel to obtain 'good' parallel efficiency for a tree search on a single gene. We are thus facing a classic load balancing problem when inferring per-gene phylogenies. In other words, we need to decide in which order and with how many cores to infer per-gene phylogenies given a resource allocation of – for example – 24 hours and 512 cores.

To address this issue, we developed and released an open-source tool called ParGenes that automatically determines the best-fit model for each gene (using our tool Modeltest-NG), infers a phylogeny under this model (using our tool RAxML-NG), and automatically orchestrates the load balancing. ParGenes allows users to infer accurate per-gene trees on hundreds or thousands of gene datasets via one single Message Passing Interface (MPI) program invocation.

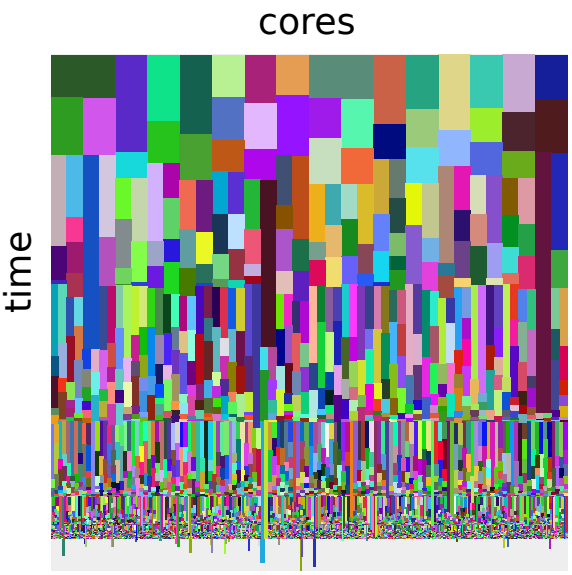



*Figure 47: Pargenes – load distribution graph of per-gene tree inference jobs.*

On an empirical dataset containing about 8,800 genes, we obtained a core utilization efficiency of 93% on 512 cores, as visualized in *Fig 47*. Each colored block in the figure represents a per-gene tree inference job and includes the number of cores allocated to it (x-axis, block width) as well as its run time relative to the overall ParGenes execution time (y-axis, block height). The gray blocks depict idle time.

While ParGenes mostly represents solid engineering work (apart from the scheduling algorithm), we believe that it might become quite popular as a convenient tool for this purpose had clearly been missing before its development.

## Scalable phylogenetic placement

In 2018, we also published our new, completely re-designed, and highly efficient tool EPA-NG (Evolutionary Placement Algorithm-Next Generation) for phylogenetic placement. In phylogenetic placement, instead of building a comprehensive phylogenetic tree from scratch, we place anonymous DNA sequences (e.g., as obtained by sequencing gut bacterial communities) onto a given phylogenetic reference tree with known sequences, which enables us to calculate a distribution of the anonymous sequences over the reference tree for a given bacterial sample. Moreover, we can compare two or more such samples (e.g., from a healthy or sick individual or before and after treatment) by placing the corresponding sequences onto the same underlying reference phylogeny. Methods for post-analyzing such phylogenetic placements are described in the following section.

Due to the decreasing sequencing cost, there was an urgent need to develop EPA-NG in order to keep up with the molecular data avalanche. To this end, Pierre re-engineered the original EPA algorithm that had been part of

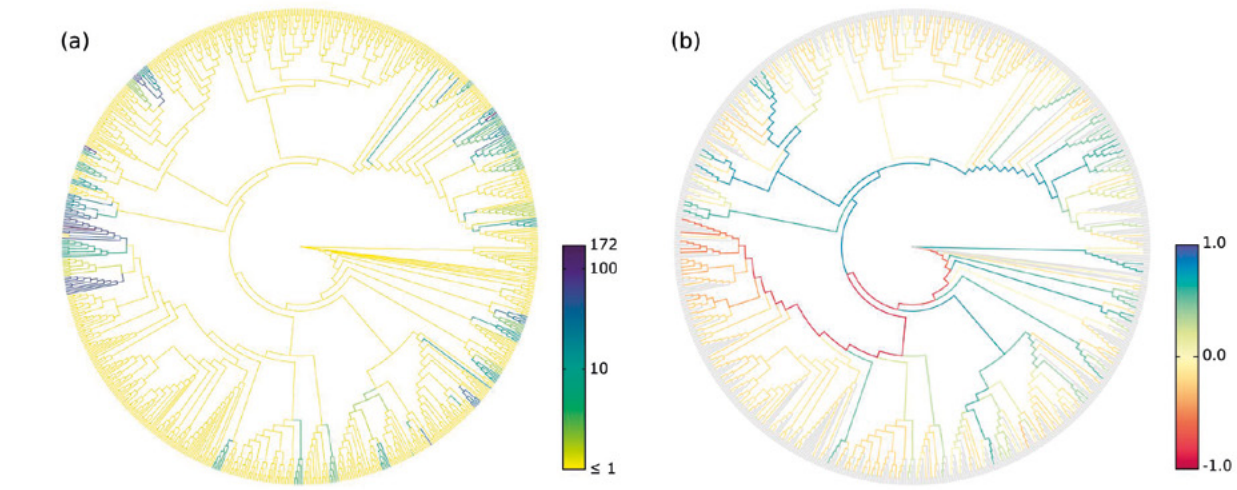



*Figure 49: Visualization of bacterial vaginosis data.*

RAxML. He initially developed novel heuristics, integrated our highly efficient low-level library for phylogenetic likelihood calculations, and optimized the EPA-specific likelihood calculations. These sequential optimizations yielded a speedup of up to a factor of 30 compared with the previous EPA implementation in RAxML as well as with the competing program, pplacer. More importantly, Pierre also developed a highly efficient parallelization approach that scales well up to 2,048 cores (see *Fig 48*). To showcase the scalability of EPA-NG, we placed 1 billion sequences from the Tara Oceans Project (a project that collects bacterial samples from oceans across the world) onto a reference tree with 3,748 species in just under 7 hours by using 2,048 cores. Toward the end of 2018, we were also asked to collaborate on a project that placed 3 billion eukaryotic sequences on a reference tree with approximately 800 species within the framework of the UniEuk initiative (see https://unieuk.org/ for details).

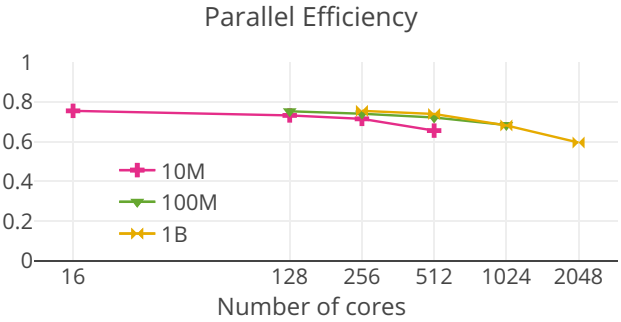



*Figure 48: Parallel efficiency of EPA-NG.*

Due to the substantial run time and scalability improvements, the pre-processing step of phylogenetic placement has now become the bottleneck – that is, the alignment of anonymous sequences to the reference multiple sequence alignment with our PaPaRa (Parsimony-based Phylogeny-aware Read aligner) tool. We therefore plan to investigate novel heuristics to speed up PaPaRa.

## Postprocessing, visualizing, and analyzing phylogenetic placements

Apart from the placement process per se, existing tools for post-analyzing the results of a phylogenetic placement run also face scalability challenges. Therefore, Lucas Czech re-implemented existing methods more efficiently in addition to developing novel methods for this purpose. More specifically, Lucas introduced methods for visualizing differences between samples and their correlation with associated meta-data on the reference phylogeny in addition to introducing methods for clustering similar samples by using a variant of the well-known k-means clustering method.

An example of such a visualization is provided in *Fig 49* for an empirical dataset containing bacterial vaginosis data. The depicted phylogeny is the reference phylogeny, and the color-coding scheme highlights the differences between the bacterial sample placements of healthy and sick individuals.

To demonstrate the scalability of his post-analysis methods, Lucas also applied them to the aforementioned Tara Oceans dataset, which was also used to assess the scalability of EPA-NG (for details, see https://www.biorxiv.org/content/early/2018/06/14/346353).

# 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


# 2.10 Scientific Databases and Visualization (SDBV)

# 2.10
# Scientific Databases and Visualization (SDBV)

Our mission is to make data accessible and comprehensible to people and software alike. How accessibility is defined and what is important for scientific data management are constantly changing. In the past few years, the buzzword "FAIR" has both shaped the discussion in the field and had a profound influence on what is asked of data management.

FAIR is an acronym that refers to making data "findable, accessible, interoperable, and reusable". There are 15 rules describing FAIR that cover data descriptions (metadata) and their quality, persistent identifiers of data, and licensing (knowing what can be reused).
A key part of making data FAIR is *biocuration* – that is, 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 rather altruistic yet that simultaneously has uncertain long-term benefits into an immediately gratifying experience?
There are numerous open-ended challenges in the field. We tackle these challenges both through concrete projects with real people and real data needs as well as through 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.

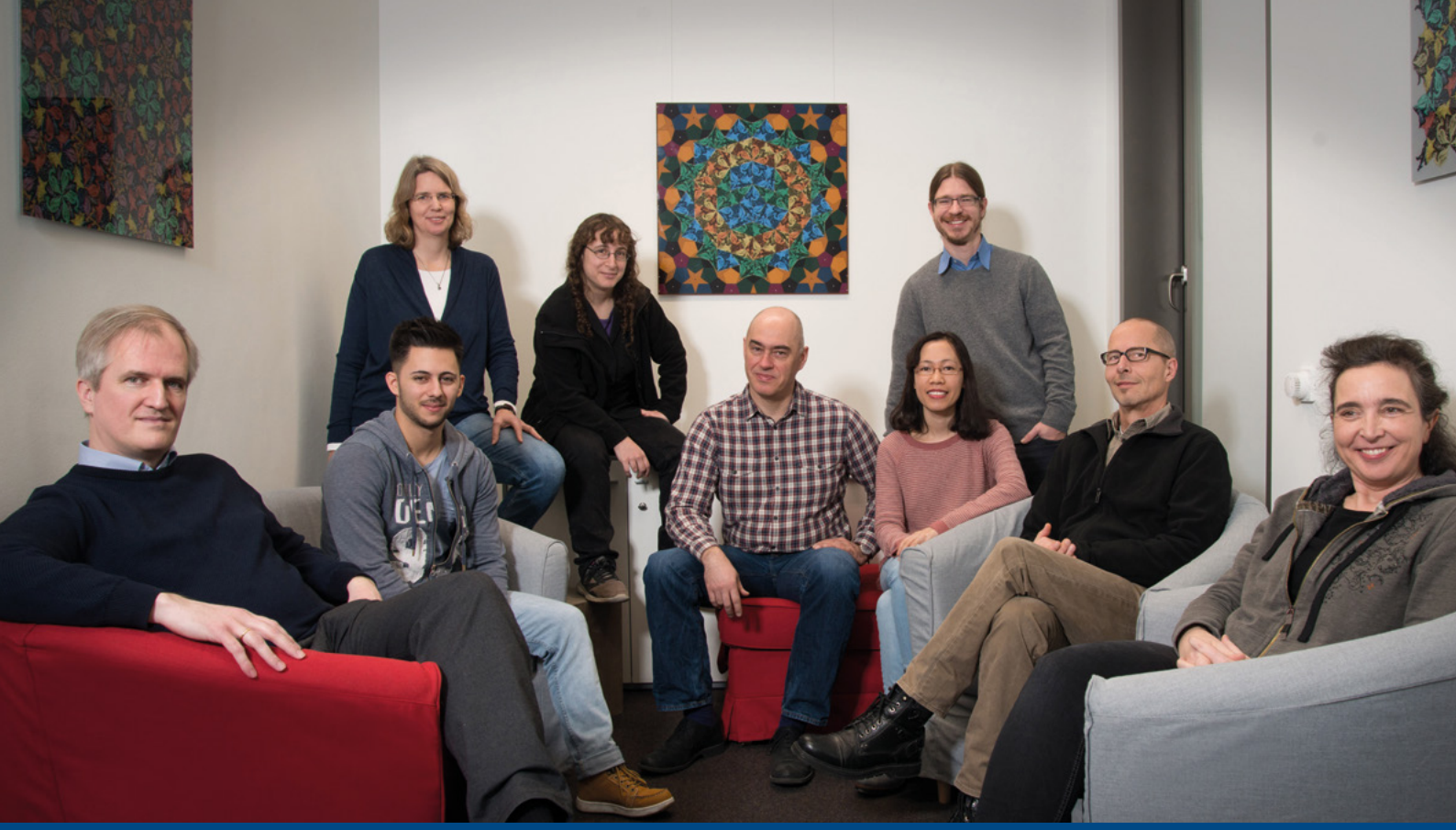
*Unsere Mission ist, Daten für Menschen und Maschinen zugänglich zu machen. Was „zugänglich" heißt und was davon für wissenschaftliches Datenmanagement wichtig ist, ist ständigen Veränderungen unterworfen. In den letzten Jahren hat die Abkürzung „FAIR" die Diskussion bestimmt und die Anforderungen für wissenschaftliches Datenmanagement stark beeinflusst. FAIR bedeutet ausgeschrieben „Findable, Accessible, Interoperable, Reusable". Es gibt 15 Regeln, die beschreiben, was FAIR bedeutet. Sie berühren Datenbeschreibungen (Metadaten) und ihre Qualität, permanente Identifikatoren für Daten und auch Lizenzierungen, die bestimmen, was weiter genutzt werden kann.*

*Ein wichtiger Teil von FAIR wird durch Biokuratierung geleistet, also Anreicherung, Strukturierung und Vernetzung von Daten. Was bedeutet Kuratierung, und was ist ausreichende Datenqualität? Auf diese Fragen pragmatische Antworten zu finden, ist immer wieder eine Herausforderung.*

*Biokuratierung ist inhärent beschränkt, zumindest wenn wenige Menschen Daten für viele kuratieren. Diese Beschränkung bringt uns in ein anderes Feld. Wie können wir Kuratierungslast auf andere verlagern? Wie können wir Forscherinnen und Forscher in die Lage versetzen, ihre eigenen Daten zu kuratieren? Wie können wir Selbstkuratierung so vereinfachen, dass sie von den Nutzern angenommen wird? Wie können wir eine Aufgabe, die hauptsächlich anderen hilft, für die, die sie durchführen, möglichst befriedigend gestalten?*

*Es gibt verschiedene schwierige Aufgaben in diesem Feld. Wir stellen uns dieser Herausforderung in konkreten Projekten sowie im Design von Werkzeugen und Entwicklung.*

*Seit ihrer Gründung profitiert die Gruppe davon , dass sie Informatiker und Wissenschaftler anderer Disziplinen zusammenbringt. Gemeinsam bauen wir Datensammlungen und Werkzeuge, die Teil deutscher, europäischer und internationaler Dateninfrastrukturen sind.*



## Group Leader

PD Dr. Wolfgang Müller

## Staff Members

Martin Golebiewski
Dr. Sucheta Ghosh
Xiaoming Hu *(since November 2018)*
Dr. Olga Krebs
Dr. Hadas Leonov *(until November 2018)*
Dr. Maja Rey
Diego Sanchez *(until November 2018)*
Dr. Andreas Weidemann
Dr. Ulrike Wittig

## Students

Marcel Petrov
Jumana Abu-Khader

## The Liver Systems Medicine network (LiSyM)

Last year, we continued our work on LiSyM, the Liver Systems Medicine network. LiSyM is a research network of more than 20 institutions that were brought together for 5 years by a 20-million-euro funding program from 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 the development and progression of many diseases. These models will facilitate decision-making at the patient's bedside and 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. Due to its vital functions, the liver is prone to many diseases, which, in turn, lead to serious problems worldwide.

More than 5 million people in Germany suffer from diseases of the liver, with non-alcoholic fatty liver disease (NAFLD) rapidly becoming the most prevalent form. 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 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. The molecular mechanisms that trigger the establishment of NAFLD in affected patients are elucidated via a systematic analysis.

### Pillar II: Chronic Liver Disease Progression

Chronic liver disease (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 (at the 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 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 MREs, 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 in a comprehensive evaluation of altered liver functions.

In addition, four junior research groups funded by LiSyM work in close collaboration with the pillars. There is also some degree of collaboration between the pillars. The scientific project management coordinates these actions and is supported and complemented by the central data management for the entire network, which is the task of SDBV in LiSyM, to which Hadas Leonov (software developer) and Martin Golebiewski (requirements engineer and user contact) contributed in 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: named patient data that can only be shared within an organization.
- Pseudonymized or anonymized data that can be shared across organizational boundaries.
- Information that would be legal to share but that may or may not be restricted in sharing (e.g., by IP considerations).

We run a 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.

### Technical work

In the course of preparing for the mid-term evaluation of LiSyM, numerous milestones were reached through a collaboration on both a service- and a technical level.

### Read/Write API

A cornerstone of our federated approach is the use of a Web API (i.e., an Application Programming Interface). The API of SEEK provides users access to SEEK via other programs. Fitting a web application with an API essentially means serving out pages in a format that is more adapted to machine processing. For this purpose, the state of the art is the use of JavaScript Object Notation (JSON).

JSON is essentially a list of attribute-value pairs in which a value can again be a list of attribute-value pairs, and so on. As a consequence, complex data structures can be stored. They can also be easily generated and consumed in a wide variety of platforms and programming languages.

A write API is particularly important for uploading and sharing. In collaboration with our partners from Manchester, we took the write API to a demonstrable state that was shown at the evaluation meeting. Matthias König built a demonstrator at Humboldt University Berlin, Germany, whose foundation lay in advice and the example code that we provided.

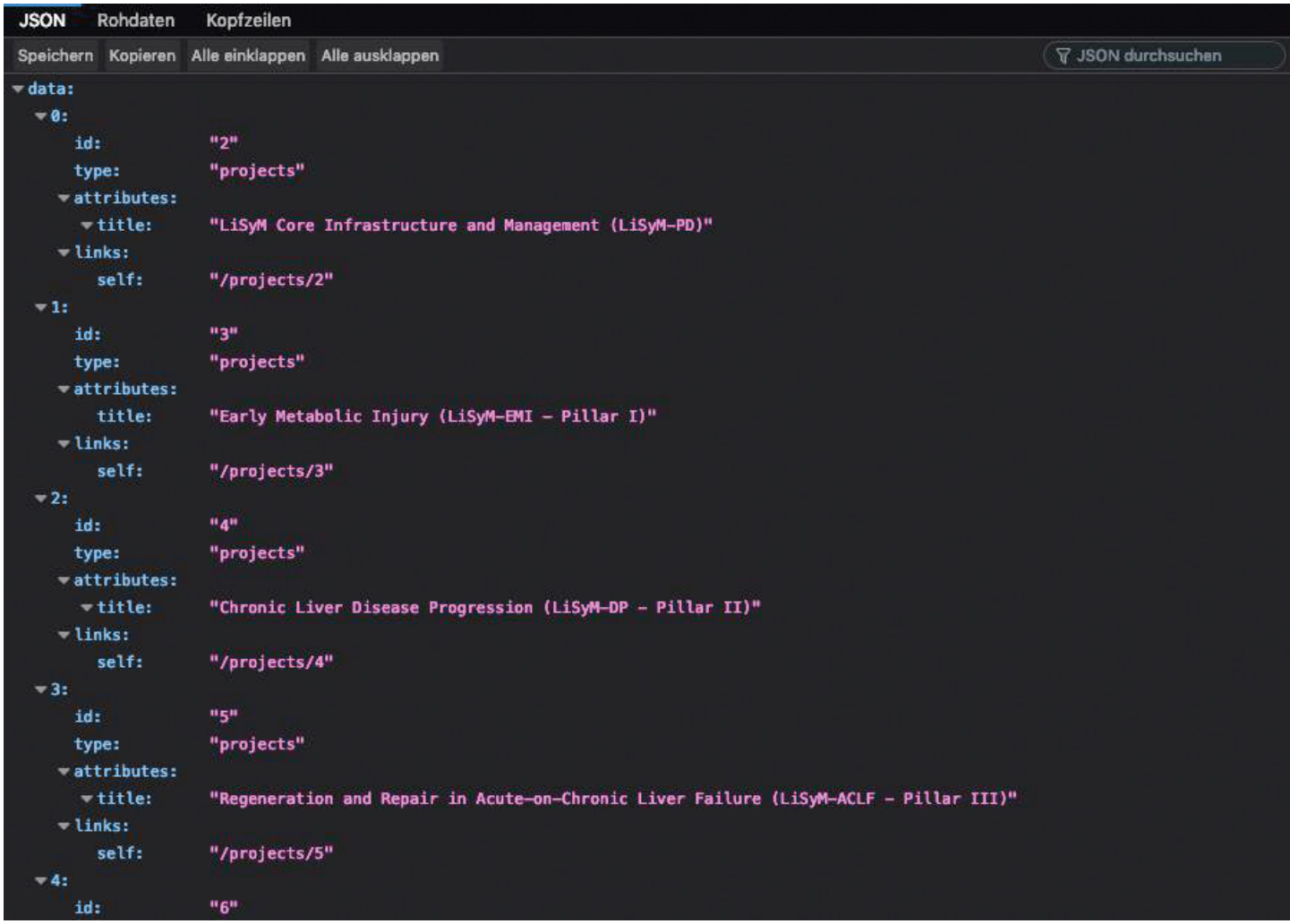



*Fig. 50: Hierarchical view of JSON data containing a list of LiSyM subprojects.*

We also built a small demonstrator for the script-based privacy-preserving distributed processing of sample-related data. This demonstrator was run by Christian Hudert at Charité, Berlin, Germany, and applied to data by Prof. Dr. Jochen Hampe at University Hospital Dresden, Germany.

Finally, on behalf of the Klingmüller Department, we were able to demonstrate the import of data that reside in the openBIS server hosted by the German Cancer Research Center (DKFZ) in Heidelberg, Germany.

### Service

To provide consulting support and service to LiSyM users, we answered many emails from users and addressed their support needs. We complemented our support communication via phone calls and participated in several LiSyM meetings.

An important part of the service was consulting users on how to apply domain-specific standards for data formats as well as providing information on metadata and semantic content descriptions. Consequently, we now continue our active work and/or leadership in several standardization initiatives and organizations (e.g., within ISO and COMBINE (Computational Modeling in Biology Network)), and we thereby provide a link between these standard-developing bodies and our SEEK users.

### Training

The requirements that were implemented as described above had been gathered together with "PALs" – a team of scientists who work in the LiSyM projects and represent different parts of the projects – as well as the main scientific directions, experimentalists, modelers, and clinicians. Last year, we had three opportunities to meet PALs from the project: first at the midterm evaluation meeting, second at a tutorial held in conjunction with SBMC, and third at a hands-on data-management training as part of the FAIRDOM meeting as well as at a modeling tutorial that served as a satellite to the 19th International Conference on Systems Biology (ICSB) in October 2018.

### Policy-building and implementation

In addition to maintaining a common data policy that was signed by all PIs from LiSyM, we created a LiSyM data management task force with colleagues from the clinics (Prof. Dr. Frank Lammert, Saarland University Hospital, Homburg, Germany), experimentation (Prof. Dr. Ursula Klingmüller, German Cancer Research Center, DKFZ, Heidelberg, Germany), and modeling (Prof. Dr. Jens Timmer, University of Freiburg, Germany). As first output it built Standard Operating Protocols for the exchange of clinical data that may be constrained due to legal and ethical regulations. Such protocols are important in answering the questions of who is allowed to share data and who is allowed to accept it.

## SABIO-RK extraction of enzyme-function data from STRENDA DB

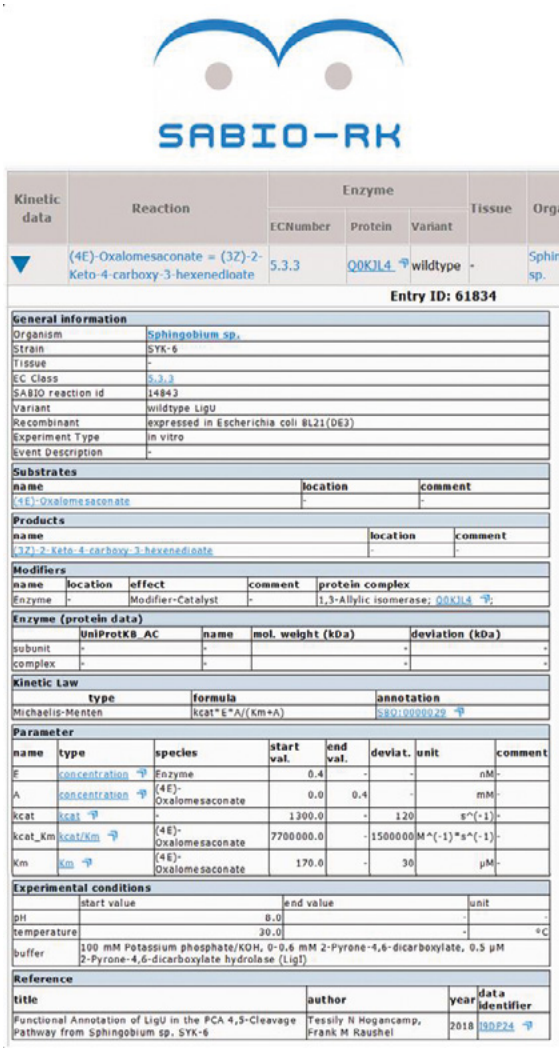
Enzyme-function data (e.g., enzyme kinetics) are of significant importance in understanding biological processes. SABIO-RK (http://sabiork.h-its.org) is a well-established, manually curated database for biochemical reactions and their kinetic properties with a focus on allowing experimentalists to gain further knowledge on enzymatic activities and reaction properties as well as on the supporting computational modeling in order to create models of biochemical reactions and complex biological networks.

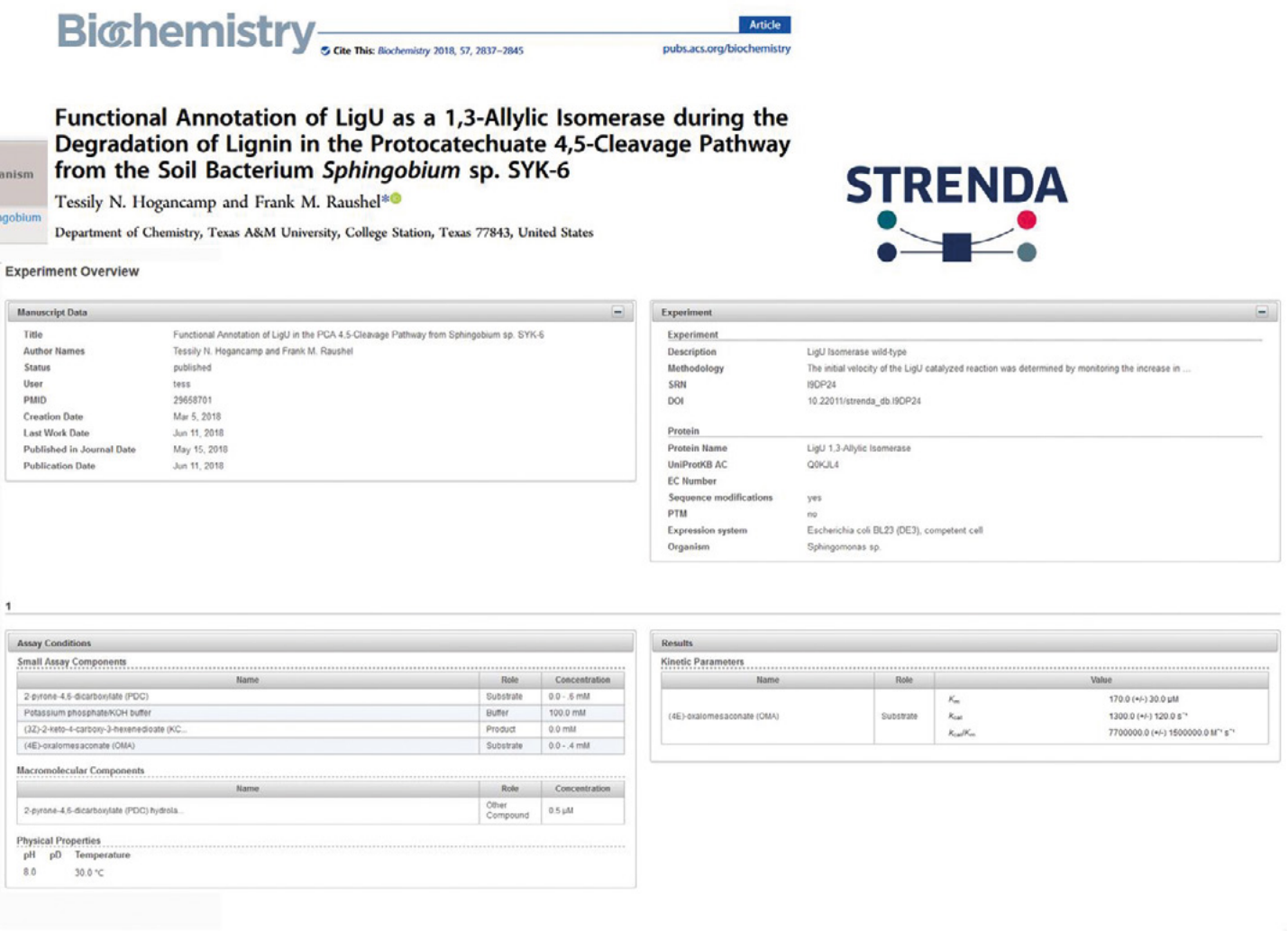
Data in SABIO-RK are mainly manually extracted from the literature but can also be uploaded directly from laboratories or other resources via the standard data-exchange format SBML (Systems Biology Mark-up Language). The quality of the data in SABIO-RK therefore highly depends on the quality of the corresponding publications used for data extraction. Biocurators are often confronted with missing and incomplete information in publications [Halling et al., 2018]. Since this manual data-extraction process requires biological expertise and is very time consuming, SABIO-RK supports initiatives within the biocuration community to improve the quality of publications and to involve authors of publications in the process of structuring and standardization their own published data. SABIO-RK therefore supports the STRENDA (STandards for Reporting ENzymology DAta) initiative (http://www.beilstein-strenda.org), which defines guidelines for reporting enzyme-function data in publications in order to increase the reusability of data for databases and modeling tools as well as to improve the reproducibility of results for experimentalists. The paper-submission process STRENDA recommends that authors submit enzyme-function data to STRENDA DB (http://www.strenda-db.org) [Swainston et al., 2018], which automatically checks the manuscript data for functional enzyme data – including kinetic parameters and experimental conditions on compliance – against the STRENDA guidelines and ensures that all required enzyme data and metadata are supplied. By simultaneously allowing the submission of manuscripts to a journal and the insertion of the corresponding enzyme-function data into a database, STRENDA DB supports the journal-reviewing process. STRENDA DB offers standardized data sheets gener-

ated for each submission and provides reviewers with a structured overview of the data to be published.
Only after the peer-reviewing process and the final publication of the paper are the data in STRENDA DB made publicly available, and they can be automatically extracted by other databases, such as SABIO-RK. After uploading the data from STRENDA DB via the SBML-upload tool to SABIO-RK, biocurators check and further annotate the data with references to controlled vocabularies, ontologies, and external database identifiers (if necessary) and finally publish the data in SABIO-RK online. Each database entry refers and links back to STRENDA DB as the original source of the data.

Thus SABIO-RK allows researchers to query these recently published data and to compare them with previous publications. Data in SABIO-RK contain annotations for controlled vocabularies and ontologies, are highly interlinked with many other databases, and are integrated into the data workflow of several modeling- and simulation tools. A flexible way of exporting database search results using web services or in a table-like format is provided. This method offers the possibility of using the recently published data to create computational models or to design new experimental setups.

Consequently, by extracting data from STRENDA DB, the SABIO-RK database populates its database content with very recent publications, and the manual curation process benefits from the import of reaction-kinetics data in a structured and standardized format.
This entire procedure allows for structured and standardized data storage during the publication process, improves the data-extraction- and curation process for other databases, and gives the authors more responsibility for final data quality.





Biochemistry

Article

Cite This: Biochemistry 2018, 57, 2837−2845

pubs.acs.org/biochemistry

Functional Annotation of LigU as a 1,3-Allylic Isomerase during the Degradation of Lignin in the Protocatechuate 4,5-Cleavage Pathway from the Soil Bacterium *Sphingobium* sp. SYK-6

Tessily N. Hogancamp and Frank M. Raushel*

Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

STRENDA

Experiment Overview

Manuscript Data

| | |
|---|---|
| Title | Functional Annotation of LigU in the PCA 4,5-Cleavage Pathway from Sphingobium sp. SYK-6 |
| Author Names | Tessily N. Hogancamp and Frank M. Raushel |
| Status | published |
| User | tess |
| PMID | 29658701 |
| Creation Date | Mar 5, 2018 |
| Last Work Date | Jun 11, 2018 |
| Published in Journal Date | May 15, 2018 |
| Publication Date | Jun 11, 2018 |

Experiment

Experiment

| | |
|---|---|
| Description | LigU Isomerase wild-type |
| Methodology | The initial velocity of the LigU catalyzed reaction was determined by monitoring the increase in ... |
| SRN | I9OP24 |
| DOI | 10.22011/strenda_db.I9OP24 |

Protein

| | |
|---|---|
| Protein Name | LigU 1,3-Allylic Isomerase |
| UniProtKB AC | Q0KJL4 |
| EC Number | |
| Sequence modifications | yes |
| PTM | no |
| Expression system | Escherichia coli BL23 (DE3), competent cell |
| Organism | Sphingomonas sp. |

1

Assay Conditions

Small Assay Components

| Name | Role | Concentration |
|---|---|---|
| 2-pyrone-4,6-dicarboxylate (PDC) | Substrate | 0.0 - .6 mM |
| Potassium phosphate/KOH buffer | Buffer | 100.0 mM |
| (3Z)-2-keto-4-carboxy-3-hexenedioate (KC... | Product | 0.0 mM |
| (4E)-oxalomesaconate (OMA) | Substrate | 0.0 - .4 mM |

Macromolecular Components

| Name | Role | Concentration |
|---|---|---|
| 2-pyrone-4,6-dicarboxylate (PDC) hydrola... | Other Compound | 0.5 µM |

Physical Properties

| pH | pD | Temperature |
|---|---|---|
| 8.0 | | 30.0 °C |

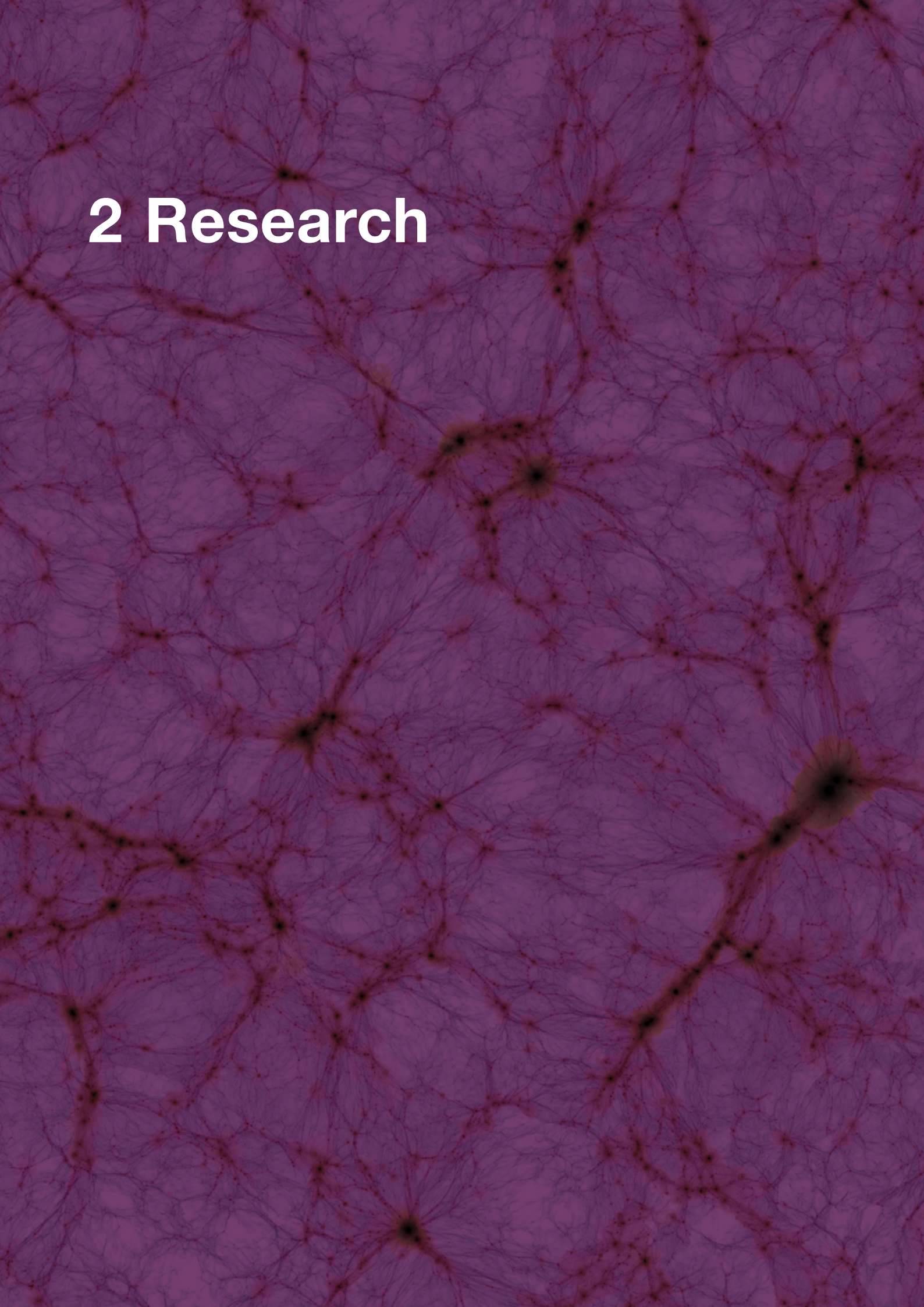
Results

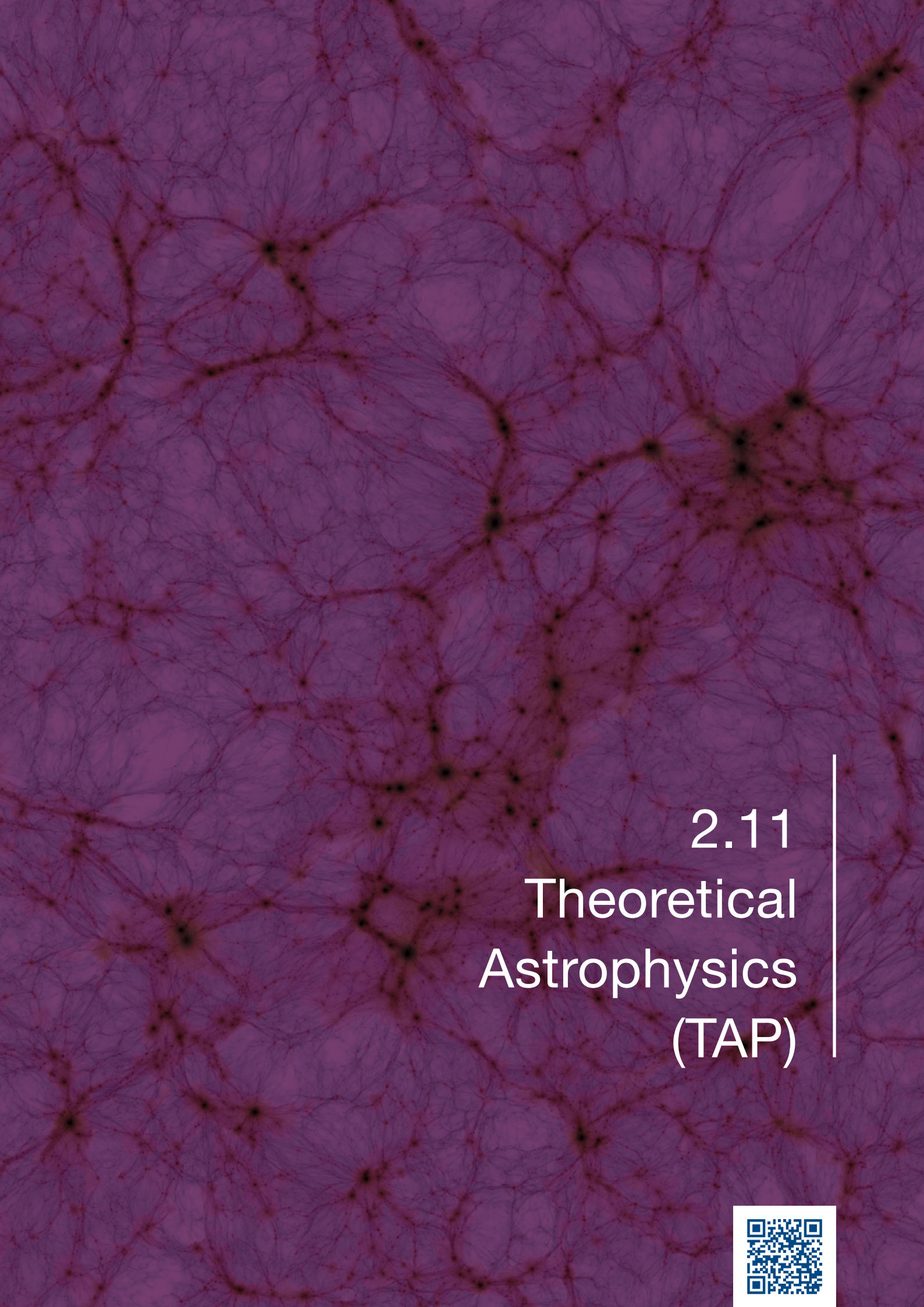
Kinetic Parameters

| Name | Role | Value |
|---|---|---|
| (4E)-oxalomesaconate (OMA) | Substrate | $K_m$ 170.0 (+/-) 30.0 µM<br>$k_{cat}$ 1300.0 (+/-) 120.0 $s^{-1}$<br>$k_{cat}/K_m$ 7700000.0 (+/-) 1500000.0 $M^{-1}$ $s^{-1}$ |

*Figure 51: Screenshots of a selected publication in the Biochemistry journal as well as from the corresponding data inserted in STRENDA DB and SABIO-RK.*

# 2 Research





# 2.11 Theoretical Astrophysics (TAP)

## 2.11
## 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 such as our own Galaxy, a busy metropolis of more than one hundred billion stars. We also aim to constrain the properties of dark matter and dark energy, the two enigmatic matter- and energy components that dominate today's Universe and constitute some of the most fundamental problems in modern physics.

A prominent role in our work is played by numerical simulations of both the collisionless and hydrodynamical type on a variety of scales. To this end, we develop novel numerical schemes that can be used efficiently on very large supercomputers with the goal of exploiting them at their full capacity for linking the initial conditions of the Universe with its complex, evolved state today. Using simulations, we are able to study how diverse physical processes relevant in structure formation interact in a complex and highly non-linear fashion. A current priority in our group is to incorporate physics into our models which is thought to be important but has thus far often been neglected, such as supermassive black hole formation, cosmic rays, and radiative transfer. In this report, we highlight a few results from our work in the first half of 2018 in an exemplary fashion.

At the end of July 2018, the TAP group at HITS dissolved as group leader Volker Springel assumed his new position as Max-Planck director in Garching.

*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. 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 supermassereiche Schwarze Löcher, kosmische Strahlen oder Strahlungstransport. In diesem Bericht stellen wir beispielhaft einige Ergebnisse unserer Arbeit aus der ersten Hälfte des vergangenen Jahres vor.*

*Im Juli 2018 hat sich die TAP-Gruppe am HITS aufgelöst, da Gruppenleiter Volker Springel eine neue Position als Max-Planck Direktor in Garching angetreten hat.*


## Group Leader

Prof. Dr. Volker Springel *(until July 2018)*

## Postdocs

Dr. Robert Grand *(until October 2018)*
Dr. Rüdiger Pakmor *(until July 2018)*
Dr. Christine Simpson *(until July 2018)*
Dr. Dandan Xu *(until July 2018)*
Dr. Freeke van de Voort *(until November 2018)*
Dr. Thomas Guillet *(until August 2018)*
Dr. Felipe Goicovic *(until August 2018)*

## Students

Rainer Weinberger *(until July 2018)*
Jolanta Zjupa *(until October 2018)*
Svenja Jacob *(until October 2018)*
Sebastian Bustamante

## Mock Gaia stellar catalogues from the AURIGA cosmological simulations

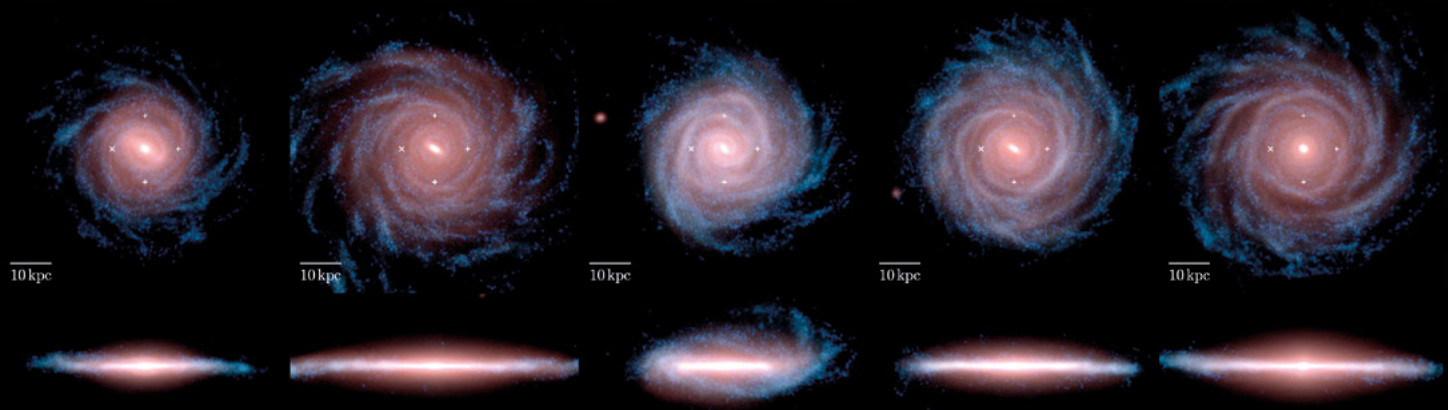
Over the next five years, our view of the Milky Way galaxy will be revolutionized by the European Space Agency's cornerstone Gaia satellite, which aims to provide positions and velocities for billions of stars in the Galaxy – a 10,000-fold increase in sample size and 100-fold increase in precision over its predecessor, Hipparcos. The second Gaia data release (DR2, from April 25, 2018) has already provided astrometric and photometric data in three bands for ∼1.4 billion sources across the entire sky. A fraction of this dataset also contains measurements for radial velocities as well as extinction and effective temperatures. In combination with several major current and future spectroscopic surveys, subsequent Gaia data releases will produce additional data for tens of millions of stars that include chemical abundances, radial velocities, and stellar ages.

In principle, this vast amount of high-dimensional empirical information about the stellar component of our Galaxy holds the key to unveiling its current state through the precise identification of disc-, bulge-, and halo substructures as well as its formation history. However, this interpretation requires some form of modeling. A crucial aspect in the quest to draw reliable conclusions lies in understanding the limitations, biases, and quality of the observational data. Specifically, the effects of survey selection functions, sample size, survey volume, the accuracy of phase space and spectroscopic measurements, dust obscuration, and image crowding influence inferences as to the true phase-space distribution of stars.

A pragmatic solution to these problems is to generate and analyze synthetic Milky Way catalogues cast in the observational frame of the survey. "Mock catalogues" of this general type were first used in cosmology in the 2000s and have now become an essential tool for the design and analysis of large galaxy- and quasar surveys. Realistic mock catalogues provide assessments of an instrument's capabilities and biases, tests of statistical modeling techniques applied to realistic representations of observational data, and detailed comparisons between theoretical predictions and observations.

Mock stellar catalogues based on full hydrodynamical cosmological simulations are particularly appealing for fulfilling these goals as they can provide us with a window into how different types of stars that originate from cosmological initial conditions are distributed in phase space. Given that the details of these distributions depend on the formation history of the Milky Way, multiple mock catalogues derived from simulations that span a range of formation histories are desirable when studying many aspects of disc and halo formation in detail.

*Figure 52: Face-on- and edge-on projected stellar densities at z = 0 for the six high-resolution AURIGA simulations from which we construct mock catalogues. The images are a projection of the K-, B-, and U-band luminosity of stars, shown in the red-, green-, and blue color channels, respectively. Younger (older) star particles are therefore represented by bluer (redder) colors. The cross in each panel (leftmost white symbol) indicates the default solar position.*



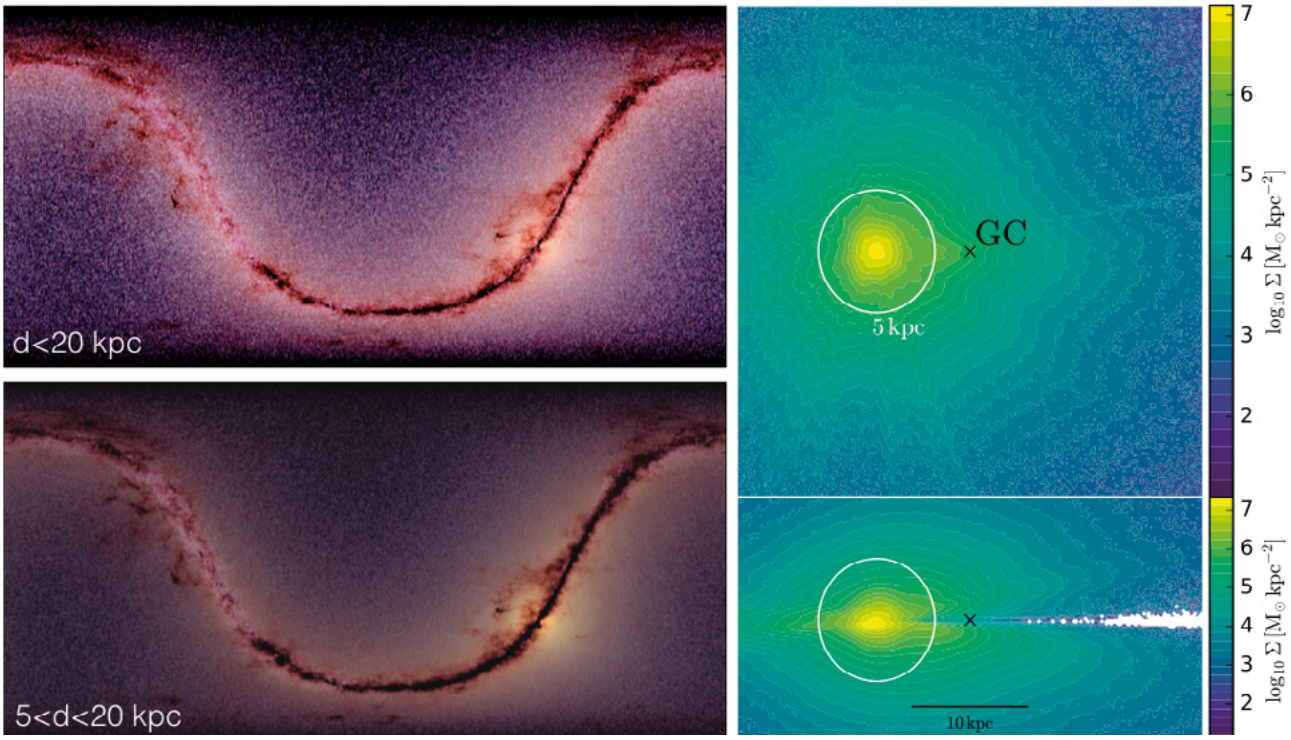



*Figure 53: Three-color all-sky maps in heliocentric equatorial coordinates of the default HITS-MOCK for Au-24. The x and y axes represent right ascension (RA) and declination (DEC), respectively. The upper image shows the stellar light distribution for all stars up to 20 kpc of heliocentric distance, whereas the lower image displays the map for stars with between 5 and 20 kpc of heliocentric distance. Right: contour maps showing the projected face-on- (top panel) and edge-on (bottom panel) stellar-mass surface density, respectively, with annotations for the galactic center and 5 kpc of heliocentric distance.*

Our AURIGA simulation suite provides an ideal set of simulations for this purpose. It consists of 40 Milky Way mass haloes simulated at high resolution with a comprehensive galaxy-formation model, including physical processes from active galactic nuclei and stars. These simulations (illustrated in *Figure 52*) have been shown to produce disc-dominated, star-forming, late-type spiral galaxies that are broadly consistent with a plethora of observational data, such as star-formation histories, abundance-matching predictions, gas fractions, sizes, and rotation curves of L∗ galaxies. Furthermore, they are sufficiently detailed to address questions related to chemodynamical properties of the Milky Way, such as the origin of the chemical thin-thick disc dichotomy, the properties of the stellar halo, and the formation of bars, spiral arms, and warps. The confluence of these advanced simulation techniques with the new Gaia- and ground-based data will transform the understanding of our Galaxy in its cosmological context at a fundamental level.

In a dedicated study *[Grand et al., 2018]*, we thus constructed two sets of mock-Gaia DR2 stellar catalogues generated from the AURIGA cosmological simulations that differ in their phase-space sampling techniques: One is denoted HITS-MOCKS, the other ICC-MOCKS. These catalogues (see *Figure 53*) contain the true and observed phase-space coordinates of stars, their Gaia DR2 errors, magnitudes in several passbands, metallicities, ages, masses, and stellar parameters. We demonstrated that a powerful use of the mock catalogues is to compare them with the intrinsic simulation data from which they were generated in order to acquire predictions of how accurately physical properties are reproduced and to determine what kind of data should be studied from the Gaia survey to target specific questions. In a practical application, we revealed that in contrast to typical disc setups in many idealized N-body simulations, the AURIGA simulations predict that young stars (∼ a few hundred Myr old) make up flared distributions (increasing disc height with increasing radius), which are well-traced by B- and A-dwarf stars.

## The dependence of cosmic-ray-driven galactic winds on halo mass

Galactic winds play an important role in the formation and evolution of galaxies. Observations demonstrate that they are common at higher redshifts as well as in star-bursting galaxies in the local Universe. Galactic winds might be able to transport chemically enriched material from the star-forming disk to the circum-galactic medium and help to explain the observed metal abundances there. Moreover, the wind material is at least temporarily unavailable for star formation in the disk. The last point, in particular, makes galactic winds crucial in simulations of galaxy formation that typically suffer from an overproduction of stars unless very strong feedback models are invoked.

Thus far, most simulations (especially on cosmological scales) have employed empirical models to drive winds. In order to improve these heuristic subgrid prescriptions, better knowledge of the physical driving mechanisms behind the winds is essential. Most models are based on the notion that some aspect of stellar feedback drives the winds, but exactly which part of the feedback physics launches the outflow remains unclear. One possibility is the direct thermal and mechanical energy input from (several) supernovae. Furthermore, the radiation pressure from young stars might be able to accelerate the gas, although the required opacity is still subject to debate. The alternative possibility of driving galactic winds with cos-

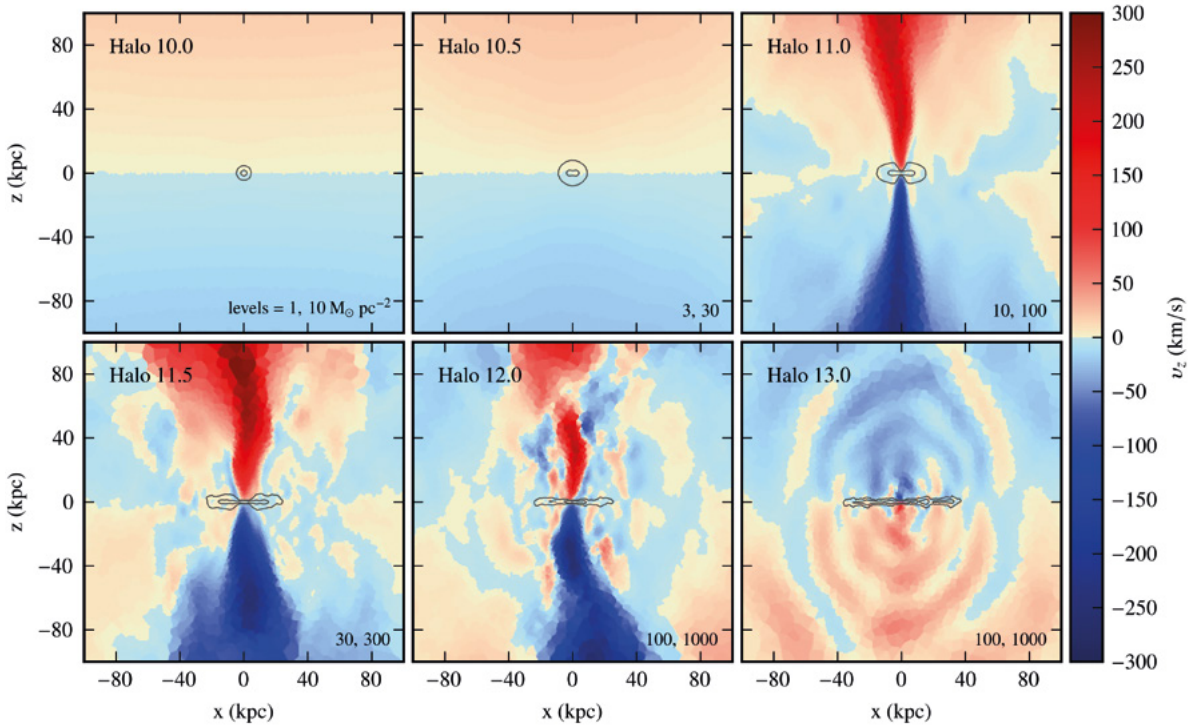
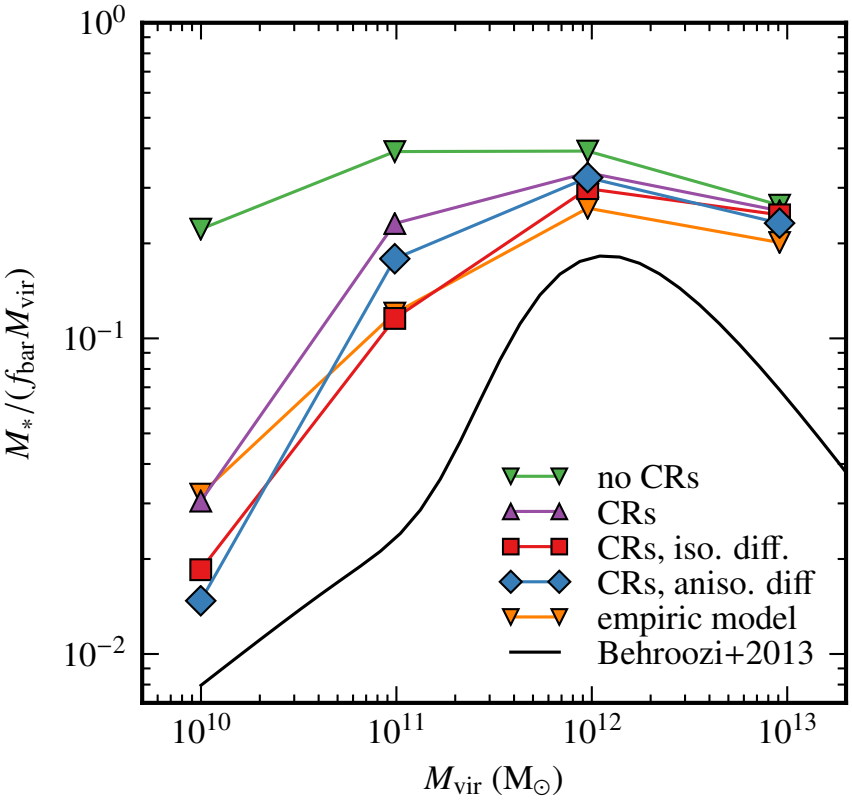



*Figure 54: Projections of the z-component of the velocity after 6 Gyrs in our fiducial runs with isotropic CR diffusion. The mass of the displayed galaxies increases from left to right and top to bottom. The contours show the gas-surface density in the central 4 kpc of the galaxy. The two galaxies with the lowest masses display slow, unstructured outflows, whereas the galaxies with intermediate masses develop biconical outflows with higher velocities. If the mass of the galaxy is further increased, the outflow becomes weaker and is completely absent in the highest-mass galaxy.*

mic rays (CRs) has come increasingly into focus in recent years. CRs are relativistic particles that permeate the interstellar medium with an energy density that is comparable to the thermal energy density and the energy density in magnetic fields. CRs interact with the thermal gas via magnetic fields, which leads to additional, effective pressure. Therefore, gradients in the CR pressure exert a force on the gas.

However, CRs can only efficiently drive winds if they can move relative to the thermal gas. In this case, extended CR pressure gradients form above and below the disk can accelerate the gas. The required transport mechanism depends on the detailed physics of CR propagation. It can be modeled as either diffusion or streaming. Earlier magnetohydrodynamic simulations demonstrated that at least one of these transport mechanisms is required to produce galactic outflows.

*Figure 55: Star-formation efficiency after 3 Gyrs as a function of halo mass. This star-formation efficiency is reduced in the presence of CRs, especially if CR-driven outflows develop in the lower-mass haloes in the simulations with CR diffusion. This reduction gives the dependence of star-formation efficiency on halo mass a similar shape as the one required by direct observational modeling (black line).*




Nevertheless, a driving mechanism that generates an outflow in one galaxy might create winds with vastly different properties in a galaxy with higher or lower mass or not drive an outflow at all. We therefore studied in detail exactly which galaxies can produce CR-driven winds and how the wind properties depend on halo mass *[Jacob et al., 2018]*. To this end, we simulated a set of idealized, isolated galaxies that included CR diffusion (*Figure 54*). We varied the virial mass of the galaxy between $10^{10}$ and $10^{13}$ $M_\odot$ and tested different aspects of CR physics, such as isotropic and anisotropic diffusion. Moreover, we compared our results with observations and empirical wind models. Interestingly, we found that the star-formation efficiency declines strongly for low-mass galaxies due to the presence of CRs (*Figure 55*), roughly in the manner required by observational inferences based on abundance-matching models. This finding strongly supports the notion that CRs play an important role in the regulation of star formation, especially in low-mass galaxies.

## Faraday rotation maps of disc galaxies

The magnetic field in the Milky Way and in nearby disk galaxies of similar mass is in equipartition with thermal-, turbulent-, and cosmic-ray energy densities. Therefore, understanding the magnetic field of disk galaxies is necessary to understand their dynamical evolution as well as to model anisotropic transport processes of CRs along magnetic field lines.

However, the magnetic field of galaxies is very difficult to observe directly; instead, a number of indirect tracers of the magnetic field are employed to infer its properties in galaxies. One widely used observational tracer is Faraday rotation of polarized radio-continuum sources at frequencies of a few GHz, at which Faraday depolarization is weak. The corresponding signal carries information about magnetic field strength as well as the structure of the magnetic field. Together with other tracers, Faraday rotation is widely used to build and test models of the global magnetic field of the Milky Way.

The magnetic field is often deemed to be dominated by either a dipole- or a quadrupole field on large scales, as expected from idealized galactic dynamo models. Nevertheless, many observations point to a more complex structure. One major advantage of Faraday rotation over other tracers – such as polarized and unpolarized synchrotron emission – is that it does not depend on the properties of the population of cosmic-ray electrons in the observed galaxy, which is typically

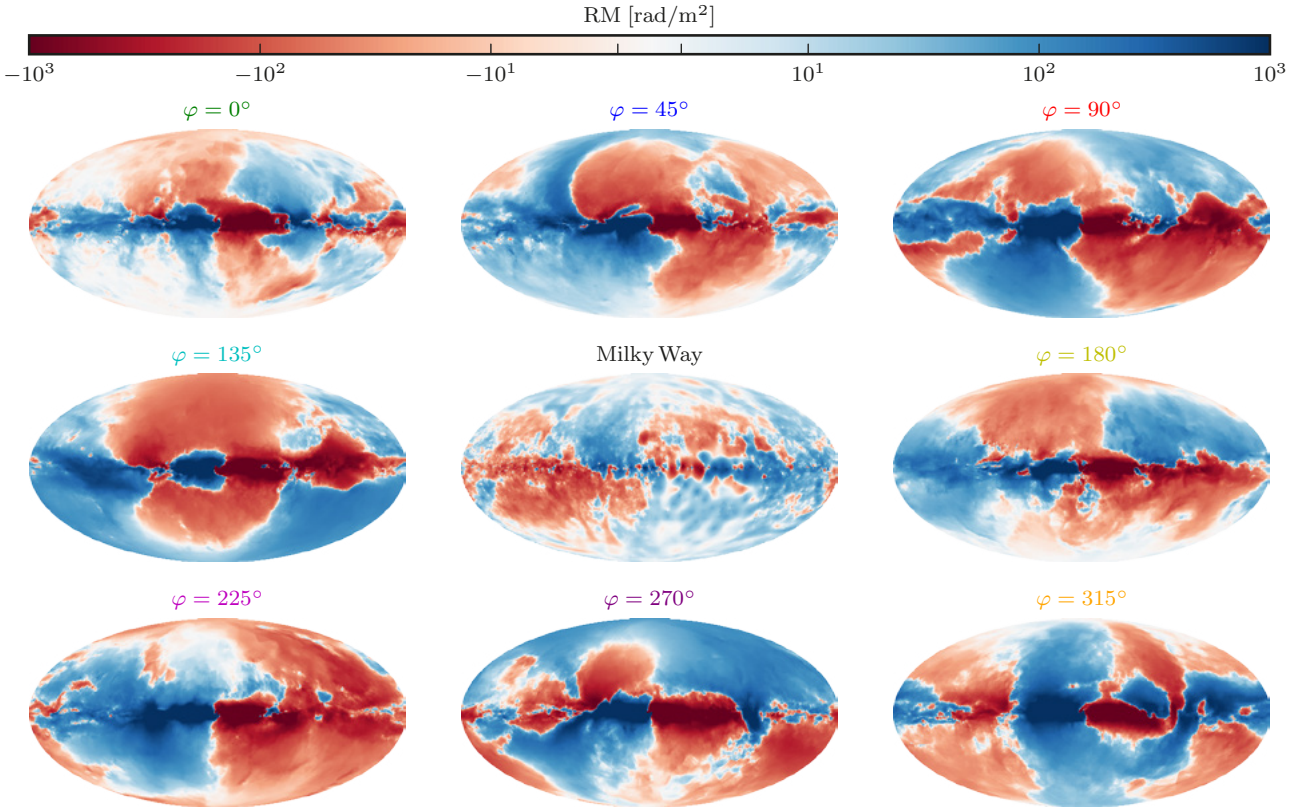



*Figure 56: The reconstructed Faraday rotation map of the Milky Way (central map) and all-sky Faraday rotation maps of Au-6 for an observer at different positions in the midplane of the disk at a radius of 8 kpc. The azimuthal positions φ on this circle are separated by 45 degrees. The maps are integrated to a depth of 20 kpc. Contributions from larger distances are negligible.*

also poorly known. Instead, Faraday rotation depends on thermal electron density, which can be measured independently of tracers of the ionized medium. However, the interpretation of observed Faraday rotation can also be non-trivial, and numerical simulations of galaxies that include magnetic fields are thus needed to better understand the observed data.

Until recently, the evolution of magnetic fields could only be simulated in idealized simulations of isolated disk galaxies that ignored both their complicated evolutionary history and their cosmological environment. Recently, however, high-resolution cosmological simulations of disk galaxies that follow the evolution of magnetic fields from the formation of the galaxies to z = 0 have become feasible, which we realized for the first time in the AURIGA simulations.

In a new study by the TAP group *[Pakmor et al., 2018]*, we also – and for the first time – computed synthetic Faraday rotation maps for polarized background sources of a fully cosmological high-resolution simulation of a Milky Way-like disk galaxy for different observer positions both within and outside the Galaxy (*Figure 56*). Interestingly, we found that the strength of the Faraday rotation of our simulated galaxies for a hypothetical observer at the solar circle was broadly consistent with the Faraday rotation seen for the Milky Way. The same held for an observer outside the Galaxy and the observed signal of nearby spiral galaxy M51 (*Figure 57*). However, we also found that the structure and angular power spectra of the synthetic all-sky Faraday rotation maps varied strongly with azimuthal position along the solar circle. We argue that this variation is a result of the structure of the magnetic field of the Galaxy, which is dominated by an azimuthal magnetic field ordered on scales of several kpc yet has radial and vertical magnetic field components that are only ordered on scales of 1–2 kpc.

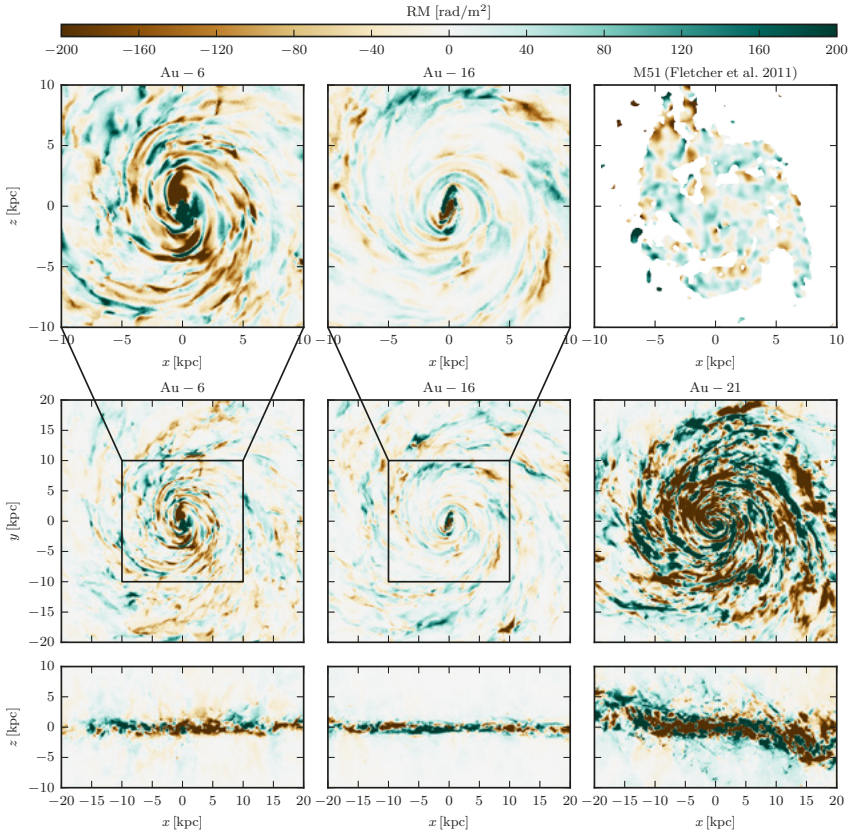
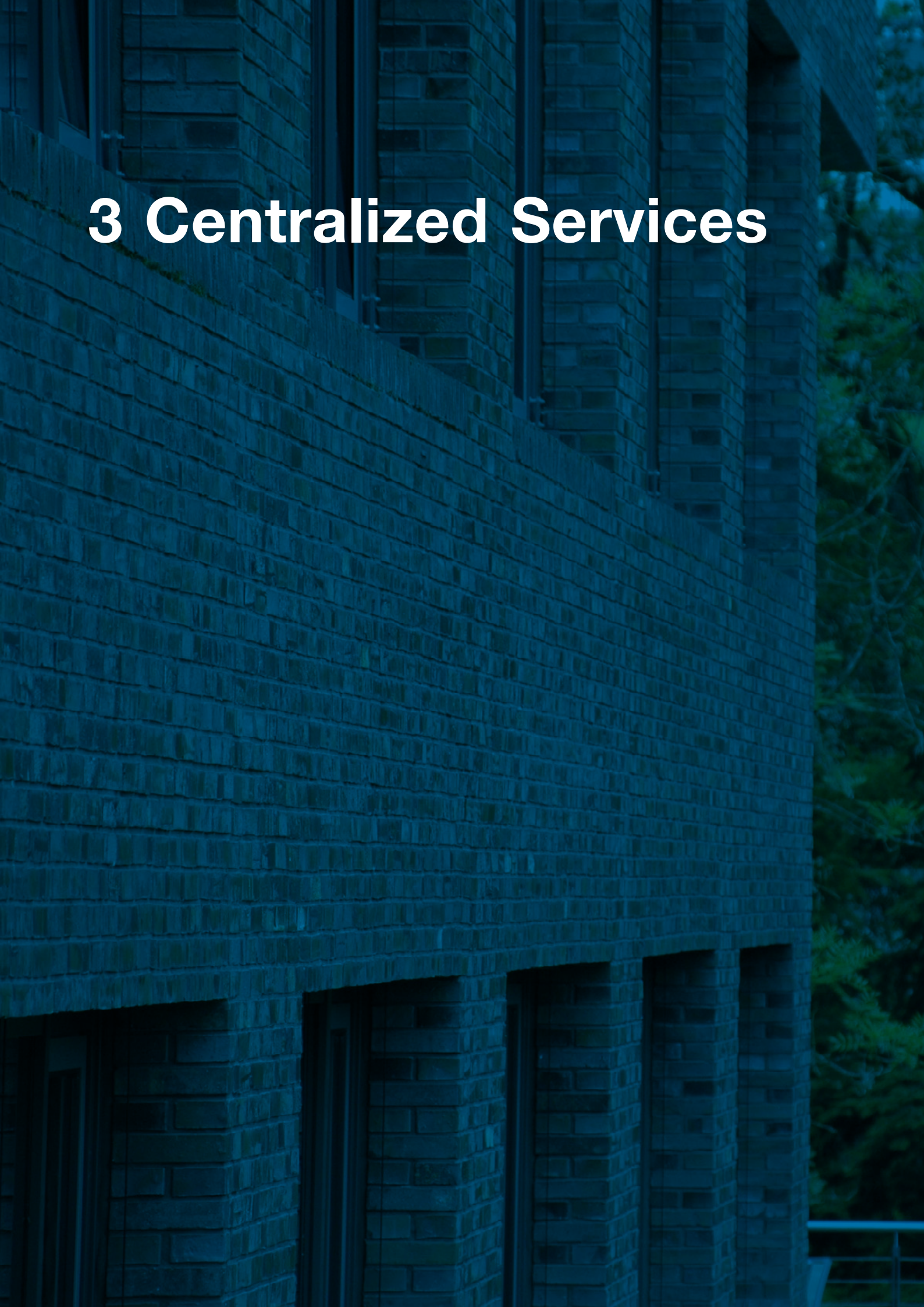
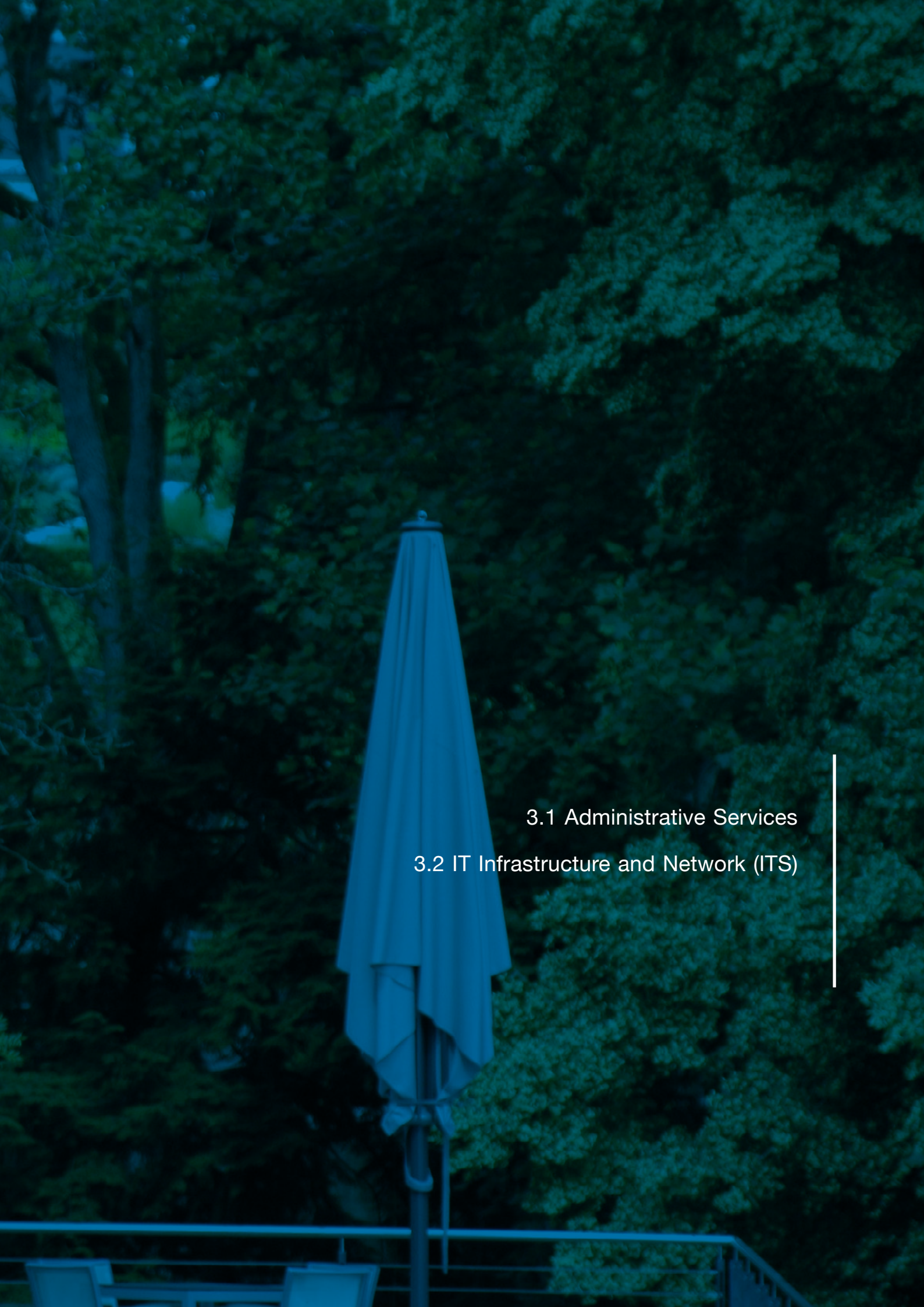



*Figure 57: Faraday rotation maps as seen by an external observer. The top panel displays zoom-ins of face-on Faraday rotation maps of halos Au-6 and Au-16 and the observed Faraday rotation map of M51 (all three are displayed on the bottom panels) for wavelengths of between 3cm and 6cm and assuming a distance to M51 of 7.97 Mpc.*

# 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)*
Frauke Bley *(human resources, since October 2018)*
Christina Bölk-Krosta *(controlling)*
Benedicta Frech *(office)*
Ingrid Kräling *(controlling)*
Thomas Rasem *(controlling)*
Rebekka Riehl *(assistant to the managing director and human resources)*
Stefanie Szymorek *(human resources)*
Irina Zaichenko *(accounting)*

### Student

Lilly Börstler

The HITS administration supports the scientific groups in almost all administrative tasks. It takes care of the day-to-day operations at the two HITS sites, supervises human resources and accounting, clarifies legal issues, and supports the organization of events.

At the end of 2018, HITS comprised three buildings at two locations in which ten research groups consisting of around 120 people – including 35 doctoral students – sat and worked. At the beginning of the year, twelve scientific research groups remained at HITS, two of which – the Computational Biology group and the Theoretical Astrophysics group – left the institute during the course of the year. A group reaches its end when the group leader leaves; however, the administration usually continues to be involved with the group via final tasks, such as settling accounts, completing reports, issuing certificates, and performing tax write-offs.

In order to fill the vacant positions, two joint appointments were prepared with Heidelberg University and the Karlsruhe Institute of Technology, and an internal opening appointment was made for a new junior group. At the end of 2018, the process for the latter group had already reached the stage at which a contract could be drawn on the topic of "Computational Carbon Chemistry," and a promising young scientist was found in computational chemist Ganna Gryn'ova. The appointment and selection procedure also kept the HITS administration busy, with preparations for the start of the new group already underway. The state is one of the greatest driving forces behind the administration, regularly calling on it to act on new legal requirements. In 2018, this included the EU General Data Protection Regulation and new directives regarding the posting of workers' activities in other European countries.

After rooms had become available, several HITS groups changed locations within the two buildings on the HITS premises along the Schloss-Wolfsbrunnenweg. Preparations were also made to renovate a wall in the HITS garden that had become dilapidated due to water pressure from the slope.

In the area of human resources, we worked intensively on the revision of forms and other documents with the aim of reducing the administration's workload and bureaucracy where possible. In the autumn of 2018, Frauke Bley was recruited as a new member of staff, and her presence is sure to significantly fortify our human-resources work. At the same time, we worked on a new concept for welcoming and better integrating new people at HITS.

In the area of controlling, we were able to re-allocate tasks. By separating preparatory accounting from the main department, we were able to significantly reduce the controllers' workload and free them up for other urgent tasks. For example, we participated in the preparations for the conversion of the software for financial and personnel accounting in close cooperation with our service provider, KT Abrechnungsdienste. The actual changeover is planned for 2019/2020. At the same time, we continued to work on the HITS reporting system, which should now involve the regular preparation of reports for the management, the shareholders' meeting, and the HITS research groups according to a fixed schedule. The day-to-day business of controlling in 2018 also saw the conclusion of two ERC Grants, including auditing. After the opening of the ESO Supernova Planetarium and Visitor Centre in April 2018 in Garching, the project could begin to come to an end (see *Chapter 6*).

For the administration, the end of a project is often far from the end of their work, but these final activities usually take place behind closed doors.

## Group Leader

Dr. Ion Bogdan Costescu

## Staff Members

Dr. Bernd Doser *(Software Developer)*
Cristian Huza *(System Administrator)*
Norbert Rabes *(System Administrator)*
Andreas Ulrich *(System Administrator)*
Taufan Zimmer *(System Administrator, since April 2018)*

At the very beginning of the year, the IT world was shaken by the announcement of the processor flaws called Meltdown and Spectre, which allow for otherwise non-accessible data to be read in certain circumstances. As fixes began to appear, it became clear that most of them can introduce significant performance losses. As a result, we needed to perform a thorough analysis of the available information and of the compromise between security and performance (particularly regarding our HPC environment) and to gradually roll out the necessary software updates.

Begun at the end of 2017, the replacement of the firewall raised unexpected compatibility issues. Later, it also became apparent that the new hardware was not running stable, and it thus had to be replaced several times. As the firewall is a central component of the network setup, careful planning on our side was required to avoid disruptions to the traffic, particularly long interruptions in access to and from the Internet. In parallel, we also increased the redundancy of our Internet connection by linking to a second point of access in the regional academic BelWü network. Finally, we completed the Wi-Fi coverage inside the HITS buildings and will be looking to add coverage for some outside areas in the near future.

Our HPC environment grew with the addition of more than 400 cores that use the newest Intel Skylake CPU architecture. By attaching them to the existing FDR Infiniband network, they gained access to our large data storage systems and became able to communicate with low latency. Independent of this extension, the scheduling of the cluster jobs raised many discussions. On one hand, the default scheduling features of SLURM turned out to be insufficient for our needs, leading us to develop another priority scheme. On the other hand, the previous addition of a large number of NVIDIA GPUs (see *Annual Report 2017*) promoted them to an important computational resource at the institute. It then became apparent that the rather interactive nature of working with neural networks for machine-learning purposes does not mix well with the more predictable and long-running molecular dynamics simulations when sharing the same GPU-equipped nodes. Furthermore, even the various machine-learning applications run with widely varying efficiency, thereby rendering job scheduling difficult. Thus, it will be necessary to further refine the scheduling policies in the near future.

# 4 Communication and Outreach


KLAUS TSCHIRA

# 4 Communication and Outreach

*With paper, CD slivers, and scissors: Building your own spectrograph at Explore Science*

The HITS communications team works on drawing the attention of the media to the excellence of HITS scientists in their respective fields, which range from the molecular to the galactic levels. At the same time, we aim to establish "HITS" as a brand name for a small albeit excellent interdisciplinary research institute that is perfectly suited for young, aspiring researchers. Moreover, we strive to spark enthusiasm for science among both school students and the general public alike through our outreach activities. Some of this year's highlights are presented below.

In 2018, we were happy to initiate our media activities with press releases on long-term research projects in the Computational Biology group, namely the decoding of the genomes from the Axolotl salamander and the flatworm Schmidtea mediterranea, both of which are relevant organisms for regenerative medicine. This collaboration with partners in Germany and Austria resulted in two "Nature" publications in late January that caused a worldwide media echo.

Only one week later, the Theoretical Astrophysics group (TAP) published initial results from the "IllustrisTNG" project. The group shared new insights into the formation and evolution of galaxies via three joint papers with researchers from Germany and the USA. Again, the international media response was enormous. In the last five years, the "Illustris" collaboration led by Volker Springel has produced the most detailed computer simulations of the Universe. It was thus not surprising that this scientific endeavor was recognized by Germany's postal service: On December 18, the "Deutsche Post" issued an official stamp to honor the "Illustris Simulation."


*Stamp of approval: The German postal service has issued a special stamp to honor the "Illustris" simulation. (Image courtesy of the GerDesign, Andrea Voß-Acker, Wuppertal/Germany, and the Illustris collaboration)*

Two other notable HITS press releases in 2018 dealt with research on parasites, but from two different angles: In their joint publication with other Heidelberg institutes, the Molecular and Cellular Modeling group (MCM) suggested a new starting point for therapy for the Malaria parasite. Concurrently, the Data Mining and Uncertainty Quantification group (DMQ) developed algorithms and software for the digital reconstruction of parasitic wasps that existed as early as several million years ago (see *title picture of this Annual Report*). In their joint work with KIT researchers, the group emphasized the fact that parasitism is widespread and exacts a tremendous impact on ecosystems.

This year again bore witness to the excellent quality of HITS research, as was reflected in the "Highly Cited Researchers" report by Clarivate Analytics. Tilmann Gneiting (Computational Statistics group, CST) and Alexandros Stamatakis (Scientific Computing group, SCO) ranked among the scientists most cited in their subject fields and year of publication.


*The "Girls' Day" 2018 at HITS.*

## Regular outreach and special events: From spectrographs to eye trackers

In 2018, the HITS communication team covered and organized a remarkable number of scientific and outreach events. As in previous years, we offered regular outreach events, beginning with the national "Girls' Day" on April 26. The goal of this yearly event is to broaden the minds of young girls and interest them in – for example – a STEM subject, such as research at HITS. In 2018, thirty girls between the ages of ten and fifteen visited us. Scientists from five research groups (CST, MCM, PSO, SCO, and TAP) offered small-scale, hands-on workshops to show the girls what the daily work and life of a researcher is like in fields ranging from bioinformatics to astrophysics.

A special event in 2018 was the one-week "helpING" academy for female high-school graduates, which focused on computer science and technology and was initiated by the Federal Ministry of Education and Science (BMBF). HITS contributed a workshop on May 23, with short talks by young researchers Sarah Lutteropp (SCO), Ina Pöhner (MCM), and Freeke van de Voort (TAP), who immerged into deep discussions with the 22 participants from all over Germany and served as role models for the young women, who were investigating possible future career paths.

From June 13–17, the "Explore Science" event again took place in Mannheim's Luisenpark. The event is geared toward 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 2018, Explore Science's motto was "Astronomy," which attracted more than 50,000 visitors from all over the region. Inside the HITS tent, visitors could build their own spectrographs and throw "meteorites" on sand to find out how the different patterns of craters emerge.

*Discussing the pros and cons of an academic career: Participants of the helpING academy*

# 4 Communication and Outreach

*One session at the "HITS Fest".*

This year, the International Building Exhibition (IBA) in Heidelberg celebrated its 5th anniversary. The IBA's motto, "Wissen schafft Stadt" ("city of science"), alludes to the university and the science institutes in Heidelberg. A "knowledge pearls" exhibition with exhibits from institutes all over Heidelberg was organized in an interim presentation. HITS provided a 3D print of a simulated protein from the Molecular and Cellular Modeling group with an explanation in German that conveyed HITS' hands-on view on data science to the broad public.

Speaking of the broad public: HITS organizes an open-house event for the public at large every two years. In 2018, the diverse program of the event included hands-on stations for kids, a slot-car racing set, presentations by HITS research groups, and four talks on topics ranging from evolution to forecasting (see *Chapter 5.3*).

The most special event this year was an internal one initiated by Scientific Director Michael Strube: the "HITS Fest" on July 23, an interdisciplinary scientific workshop consisting of a colloquium talk by Matthias Scheffler (see *Chapter 5.2*) followed by a bottom-up interactive session with different topics covering scientific areas as well as careers and social areas. Both scientific and non-scientific staff participated in this session. The HITS communication team organized the entire event, including the "pub quiz" that followed the barbecue at the end of the event.

Later that summer, the institute again participated in the International Summer Science School Heidelberg (ISH), in which the SDBV group tutored four school students from Israel, the U.K., Germany, and Ukraine who were working on a study for visual search. Using eye trackers, they acquired visual data, converted it into numeric data, and analyzed it.

From September 3–7, the AIN group at HITS hosted the International "Astroinformatics" Conference, which took place in Germany for the first time (see *Chapter 5.1.5*). We covered the entire conference, organized the video streaming of 35 talks, and assisted the organizers with manifold tasks.

In late September, the 6th installment of the Heidelberg Laureate Forum (HLF) took place, and we again had the opportunity to welcome a group of young researchers from all over the world (see *Chapter 6*).

Finally, we helped organize the workshop on Computational Topology and Topological Data Analysis on November 19, held by Anna Wienhard and the GRG group (see *Chapter 5.1.7*).


*Outreach: Yay or nay? – One of the topics discussed at the "HITS Fest".*

*HITS Communications team in 2018 (f.l.t.r.): Julia Klawitter, Peter Saueressig, Isabel Lacurie.*

## Media relations: Journalist in Residence program

*Kerstin Hoppenhaus, HITS Journalist in Residence 2018.*

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 geared toward science journalists and offers them a paid sojourn at HITS. During their stay, they can learn more about data-driven science and get to know researchers and new research topics without the pressure of the "daily grind".

In July 2018, HITS welcomed its seventh Journalist in Residence, award-winning German TV journalist Kerstin Hoppenhaus from Berlin (Germany), for six months.

During her stay, Kerstin Hoppenhaus held an internal seminar with the HITSters entitled "Death by simplification – Communicating complexity and uncertainty on the Internet." Moreover, she gave a public talk on the future of audio-visual media in the post-television era. Kerstin Hoppenhaus also dealt with machine learning and neural networks via numerous long conversations with HITS researchers from the fields of mathematics to natural language processing and astroinformatics. She was thus able to delve into "Artificial Intelligence", a topic with a history at the institute and that will be the theme of the German science year 2019.

### Head of Communications

Dr. Peter Saueressig

### Staff Member

Isabel Lacurie

### Student

Julia Klawitter *(until October 2018)*

# 5 Events

5.1 Conferences, Workshops & Courses

5.2 HITS Colloquia

5.3 HITS Open House Event


*BDBDB4 poster, the poster session and the participants of BDBDB4.*

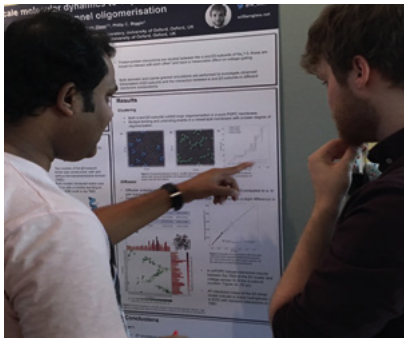
## 5.1.1 BDBDB4: Biological Diffusion and Brownian Dynamics Brainstorm 4

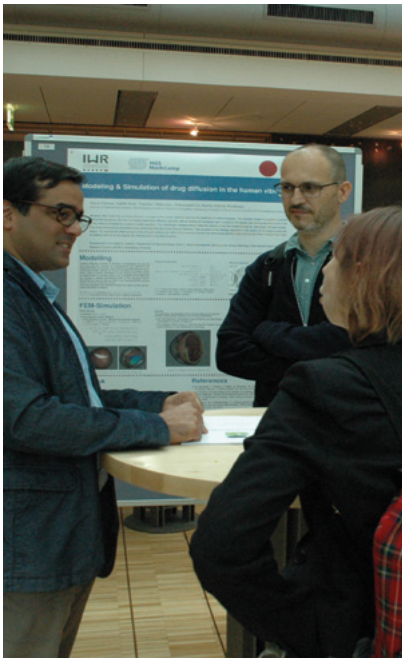
April 16-18, 2018, Studio Villa Bosch, Heidelberg, Germany

As for previous BDBDB meetings, the goal of BDBDB4 was to provide a forum for intensive, informal discussions about the state-of-the-art in studying biomolecular diffusion. This is currently a particularly exciting and active research area because of recent advances in experimental techniques, such as super-resolution microscopy, and in modeling and simulation approaches that exploit advances in computing hardware, such as virtual reality glasses and GPUs. Presentations in the form of talks, posters and software demonstrations covered Brownian Dynamics and related simulation methodologies, along with recent theoretical and experimental advances in studying diffusion from molecular to cellular levels. The workshop brought together about 60 theoreticians and experimentalists from around the world. For more details, see the website: *http://bdbdb.h-its.org*

Organizers: Rebecca Wade (HITS/ Heidelberg University), Rommie Amaro (University of California San Diego, USA), Ulrich Schwarz (Heidelberg University, Germany), Huan-Xiang Zhou (University of Chicago, USA), Neil Bruce, Gaurav Ganotra and Stefan Richter (Molecular and Cellular Modeling group, HITS)

Sponsors: BIOMS Center for Modelling and Simulation in the Biosciences, Heidelberg; Heidelberg Institute for Theoretical Studies; the National Biomedical Computation Resource (NCBR), San Diego.




*Teaching the talents: During a tutorial session.*


*Talking on television: Alexandros Stamatakis giving an interview on Greek TV.*

## 5.1.2 EMBO Practical Course on Computational Molecular Evolution

May 6–17, 2018,
Heraklion, Greece

The EMBO Practical Course celebrated its 10th anniversary at the Hellenic Institute of Marine Research near Heraklion, Crete, and provided graduate and postgraduate researchers with the theoretical knowledge and practical skills to carry out molecular evolutionary analyses on sequence data.

The course offered students the opportunity for direct interaction with leading scientists and authors of famous analysis tools in evolutionary bioinformatics from across the globe, including Maria Anisimova, Nick Goldman, Bruce Rannala, and Ziheng Yang.

Alexandros Stamatakis (SCO group leader) was the main organizer of the event. We received approximately 170 applications for the 35 available places. Former SCO postdocs Pavlos Pavlidis and Paschalia Kapli as well as current SCO PhD student Pierre Barbera contributed substantially as teaching assistants. HITS was again the main co-sponsor of the course.

This year, the school also received substantial coverage by the local press in Crete:

A report about the school was shown on local TV news: *https://www.youtube.com/watch?v=yjzwqzaLUXA&feature=youtu.be*

And Alexis gave a 10-minute interview on a live TV news show (beginning around minute 15): *http://www.cretetv.gr/9188-krhth-shmera-16052018/*

Local newspapers – including Patris, the largest local newspaper on Crete – also reported on the school.

*https://www.cretalive.gr/science/therino-scholeio-sthn-krhth-gia-thn-ypologistikh-moriakh-ekseliksh*

*https://www.patris.gr/2018/05/15/diethnis-kyklos-mathimaton/*

*https://www.patris.gr/2018/05/21/alexandros-stamatakis-to-kleidi-gia-tin-ereyna-kai-toys-ereynites/*

## 5.1.3 Second ACL Workshop on Ethics in Natural Language Processing

June 5, 2018, New Orleans, USA

After a successful first Workshop on Ethics in NLP at the EACL 2017 in Valencia, Spain, NLP group leader Michael Strube got together with philosopher Mark Alfano (TU Delft) and NLP researchers Dirk Hovy (Bocconi University, Milan) and Margaret Mitchell (Google) to organize the second workshop on this topic. While the first workshop introduced ethics to NLP researchers from a variety of perspectives, the second one had a decidedly philosophical and legal slant since it featured more invited talks by prominent researchers from these fields and fewer contributions from an NLP perspective. After the invited talks and regular paper presentations, the workshop concluded with a reversed science café to discuss ethical questions in smaller groups. The results were summarized and presented by students, whose travel to the workshop we could sponsor due to generous funding by Google, Bloomberg, and HITS.

Thankfully, our message has already been heard and partially accepted by the NLP community. One of the best-paper awards at the EMNLP 2017 conference went to a work on reducing gender bias, a topic which had been introduced to the NLP community by our first workshop. The NAACL 2018 conference featured a well-received panel on Ethics in NLP. The NAACL 2019 conference is slotted to have a track dedicated to ethics, bias, and fairness. Ethics is also a topic that is beginning to show up in selected NLP curricula around the globe.

*Ganna (Anya) Gryn'ova during the symposium.*

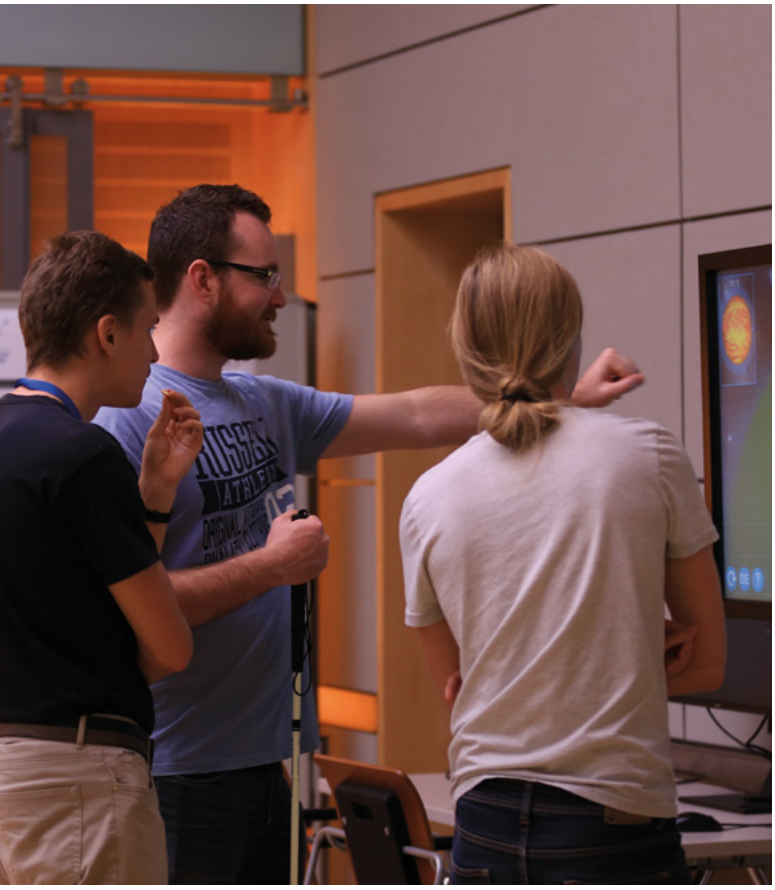
## 5.1.4 Symposium "Computational Materials Science"

July 24-25, 2018, Studio Villa Bosch, Heidelberg, Germany

HITS organized a symposium on the topic of computational materials science. The goal of the event was to select the group leader for a new junior research group at the institute. Six candidates from five countries contributed to the symposium, which was composed of research talks by the candidates, individual meetings with the committee and with HITS group leaders, and Q & A meetings on administrative issues with the management.

By chance, the symposium began one day after the "HITS Fest", an interdisciplinary scientific workshop for all HITSters (see *Chapter 4*) that most of the candidates attended. The candidates were thus able to learn more about the institute and to meet HITS' scientific and non-scientific staff in a more informal atmosphere.

After the two-day event, the committee chose Dr. Ganna (Anya) Gryn'ova from the Ecole polytechnique fédérale de Lausanne, Switzerland, as the new group leader. Dr. Gryn'ova will initiate the junior group "Computational Carbon Chemistry" in April 2019.



*At the demo and poster session.*

## 5.1.5 Astroinformatics Conference

September 3-7, 2018, Studio Villa Bosch, Heidelberg, Germany

The international "Astroinformatics 2018" conference was held in the Studio Villa Bosch, Heidelberg, from September 3–7, 2018. Scientists from all over the world met to exchange views on the newest and most successful methods of machine learning in an effort to advance the exploration of the Universe. The conference was organized by HITS researchers Dr. Kai Polsterer, Antonio D Isanto, Erica Hopkins, and Dr. Nikos Gianniotis (all from the Astroinformatics group) in cooperation with Prof. Joachim Wambsganß (Heidelberg University) and Dr. Coryn Bailer-Jones (Max Planck Institute for Astronomy).

The "Astroinformatics" conference – which is devoted to the scientific exploitation of the fast-growing volumes of data in astronomy – is one of the most important events in this field. At the conference, scientists discussed topics including novel database systems, visualization and augmented reality, artificial intelligence, and the reproducibility of research results. Moreover, the organizers had prepared tutorials and "birds-of-a-feather" sessions about machine learning.

The conference is hosted at a different international location once every year. This year, it took place in Germany for the first time.

*The participants at the Astroinformatics conference 2018.*

Among the speakers were Prof. Ray Norris (CSIRO + Western Sydney University, Australia), Prof. Pavlos Protopapas (Harvard University, USA), Prof. Katharina Morik (TU Dortmund, Germany), and Prof. George Djorgovski (California Institute of Technology).

See also the conference talks on the HITS YouTube channel.

*Kai Polsterer (left) during a coffee break.*

## 5.1.6 Workshop NFDI4Life (National German Research Data Infrastructure for the Life Sciences Consortium)

October 2nd, 2018, Studio Villa Bosch, Heidelberg, Germany

The SDBV group of HITS organized a workshop for the NFDI4Life initiative – a consortium of more than 20 partnering institutions and organizations – with the aim of preparing a national German research data infrastructure for the life sciences. In this meeting, which was locally coordinated by HITSter Martin Golebiewski, about 60 consortium partners and invited high-ranking experts from different subdomains of the life- and data sciences gathered to discuss the planning of such a data infrastructure network.

NFDI4Life brings together data scientists from research communities across the life sciences with a particular focus on the subdomains of biology, epidemiology, nutrition, public health, agricultural, and environmental science as well as biodiversity research. This initiative formed in the context of the planned National Research Data Infrastructure (NFDI) in Germany. Following the recommendations given by the German Council for Scientific Information Infrastructures (RfII), the Joint Science Conference of federal and state governments decided to establish such an infrastructure to support the research landscape in Germany.

The integration of data along the life science subdomains requires as a prerequisite that all data be made FAIR (findable, accessible, interoperable, and re-usable). NFDI4Life is dedicated to lowering the barrier of FAIR data generation and re-use for researchers. To reach these goals, NFDI4Life addresses method development in addition to domain-specific- and cross-domain standardization processes and deploys tools to support automated processes and workflows.

Following a greeting by HITS from Dr. Wolfgang Müller, the meeting was opened by a welcome address from

*Taking a break: The NFDI4Life consortium.*

Dr. Simone Schwanitz from the State Ministry of Baden-Württemberg for Sciences, Research and Arts. Subsequently, Prof. Dietrich Rebholz-Schuhmann – Scientific Director of ZB MED (Cologne) – summarized the NFDI4Life plans to be discussed at the meeting. After a keynote presented by Prof. Wolfgang Marquardt from the RfII and Chairman of the Board of Directors of the Research Center in Jülich, high-ranking invited experts gave short kick-off speeches: Prof. Ulrich Sax (University Medical Center Göttingen), Dr. Niklas Blomberg (Director of the European Bioinformatics Infrastructure ELIXIR, UK), Dr. Johannes Keizer (Independent Consultant for Agriculture, Food and Wine Information, Berlin), Prof. Martin Wagner (University of Veterinary Medicine, Vienna, Austria), Prof. Stefan Hornbostel (German Centre for Higher Education Research and Science Studies (DZHW), Hannover), and Prof. Alette Bonn (German Centre for Integrative Biodiversity Research (iDiv), Leipzig).

The afternoon was organized as World-Cafés – short discussion rounds of smaller expert groups that focused on one topic each to discuss the plans of the NFDI4Life initiative, which provided the consortium with valuable feedback for their ambitious endeavors. ■

## 5.1.7 Workshop on Computational Topology and Topological Data Analysis

November 19, 2018, Studio Villa Bosch, Heidelberg, Germany

The HITS Colloquium Talk by Herbert Edelsbrunner (ISTA) on "Stochastic geometry with topological flavor"

*The topology workshop group picture.*

(see *Chapter 5.2*) was the kick-off for a one-day workshop on Computational Topology and Topological Data Analysis organized jointly with the newly funded Cluster of Excellence "STRUCTURES: a unifying approach to emergent phenomena in the physical world, mathematics, and data" at Heidelberg University. Edelsbrunner gave a wonderful introduction to the world of Delaunay Triangulations, alpha-shapes, and their statistical properties. The second talk was by Heather Harrington (Oxford) on "Comparing models and data using computational algebraic geometry and topology" and highlighted several applications to questions in medicine. Ulrich Bauer's (Munich) talk on "Persistent homology: from theory to computation" provided a quick introduction to the theory of persistent homology, which was proceeded by a talk by Egor Shelukhin "On persistence modules in symplectic topology".
The workshop hosted over 50 participants from local research institutions as well as a few foreign guests. The inspiring talks led to fruitful discussions during the extended breaks and over lunch, thereby rendering the one-day experiment a great success. ■

## 5.1.8 Workshop on Supernovae and Stellar Hydrodynamics

December 17-18, 2018, Studio Villa Bosch, Heidelberg, Germany

The PSO group organized the 13th Würzburg Workshop on Supernovae and Stellar Hydrodynamics in Heidelberg at the Studio Villa Bosch. About 30 researchers from international universities and institutions gathered to discuss new developments in the field and to collaborate on new and existing projects, among them several HITS alumni from the PSO and the TAP groups. The participants discussed in detail the modeling of Type Ia supernova explosions, dynamical processes in stellar evolution, mergers and common envelope phases in binary stellar systems and neutron star merger events. ■

## Lars Fischer

**Spektrum der Wissenschaft, Heidelberg, Germany**

January 22, 2018: To Boldly Go – the Art of Being Human in the Age of Science

## Prof. Dr. Bin Yu

**Department of Statistics and EECS, University of California, Berkeley, USA**

February 28, 2018: Three Principles of Data Science: Predictability, Stability and Computability

## Prof. Dr. Wolfgang Hillebrandt

**Max Planck Institute for Astrophysics, Garching, Germany**

April 23, 2018: Measuring the expansion rate of the Universe: Is "Hubble's constant" constant?

## Prof. Dr. Sergei L. Kosakovsky Pond

**Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, USA**

May 28, 2018: Beyond software tuning: scaling up comparative coding sequence analysis using approximations and models that adapt their complexity to the data

## Prof. Dr. Peter Sanders

**Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany**

June 18, 2018: Engineering Scalable Algorithms

## Prof. Dr. Matthias Scheffler

**Fritz Haber Institute of the Max Planck Society**

July 23, 2018: Making the Data Revolution happen – Turning Billions of Data from Computational Materials Science into Knowledge by Artificial Intelligence Branches of the Tree of Life

## Prof. Dr. Kai Johnsson

**Department of Chemical Biology, Max Planck Institute for Medical Research, Heidelberg, Germany**

September 17, 2018: Artificial sensor proteins for applications in basic research and medicine

## Prof. Dr. Thomas Janka

**Max Planck Institute for Astrophysics, Garching, Germany**

October 15, 2018: 3D Core-Collapse Supernova Modeling and Applications to Cas A and other Supernova Remnants

## Prof. Dr. Herbert Edelsbrunner

**Institute of Science and Technology (IST), Klosterneuburg, Austria**

November 19, 2018: Stochastic Geometry with Topological Flavor

*Lars Fischer*

*Prof. Dr. Bin Yu*

*Prof. Dr. Wolfgang Hillebrandt*

*Prof. Dr. Sergei L. Kosakovsky Pond*

*Prof. Dr. Peter Sanders*

*Prof. Dr. Matthias Scheffler*

*Prof. Dr. Kai Johnsson*

*Prof. Dr. Thomas Janka*

*Prof. Dr. Herbert Edelsbrunner*

## 5.3
## HITS Open House Event

On July 7, 2018, HITS opened its doors to the public and showed interested people from Heidelberg and the surrounding area the life and work of an international research institute.

*Benjamin Heinzerling (Natural Language Processing) in conversation with a very young visitor.*

*A special treat for kids: the face paint station.*

The program included science talks, research presentations and hands-on stations at several locations on the HITS campus, and guided tours through the HITS main building. On this hot summer's day, the HITS campus was abuzz with people eager to get to know and interact with HITS researchers and to learn about their scientific work. The Open House event offered a plethora of activities for families: They could try building a spectrograph, produce geometrical soap bubbles, play a statistical forecast game, or create 3D molecules with a 3D printer.

As a special treat for kids and adults as well, an artist applied face paint and painted tattoos with scientific motifs including stars, galaxies, and molecules.

The outstanding hands-on station was directly linked to artificial intelligence through its slot-car racing set, which had been constructed by Kai Polsterer and his team from the Astroinformatics group. A plethora of people gathered around this spot and tried to beat the machine-learning system that controlled one of the cars.

In addition to the hands-on stations, Lucas Czech (Scientific Computing group), Kai Polsterer (Astroinformatics group), Sebastian Lerch (Computational Statistics group), and Volker Springel (Theoretical Astrophysics group) gave talks tailored to laypeople. The scientists spoke about evolution and computer science, machine learning in astronomy, and the science of weather forecasting and galaxy formation. The talks took place in the Klaus Tschira

*Geometrical soap bubbles.*

*In full speed mode: the slot-car racing set.*

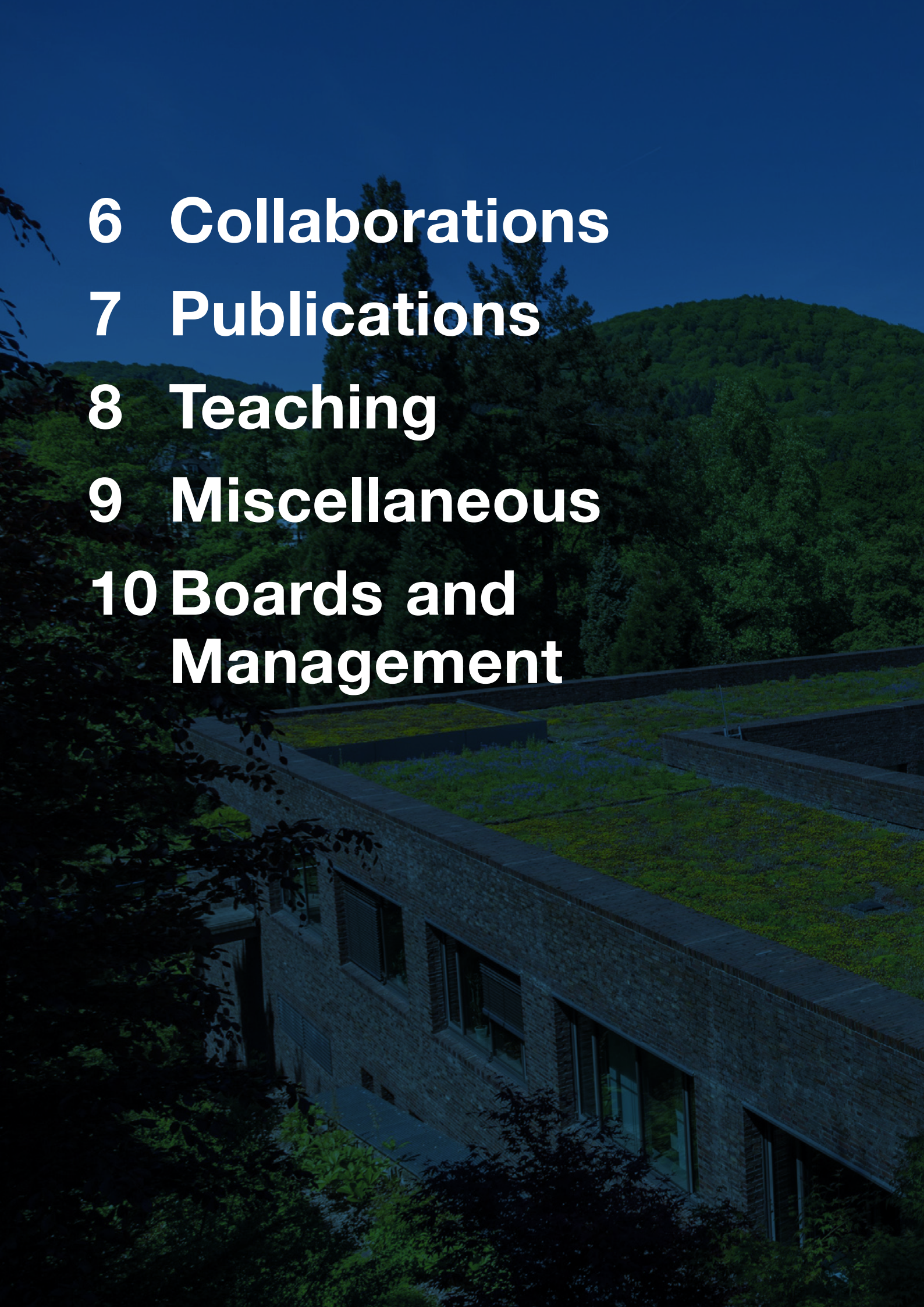

*The HITS campus during the Open House event.*

Library in Villa Reiner, which was constantly filled with visitors, some of whom listened to all the talks in a row. In one of his last presentations as a group leader at HITS, Volker Springel took visitors on a journey through the history of our Universe, shedding light on the mysteries of dark matter, dark energy, and black holes. All talks proved to be fun for visitors and speakers alike.

*Volker Springel (Theoretical Astrophysics).*


*Lucas Czech (Scientific Computing).*

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



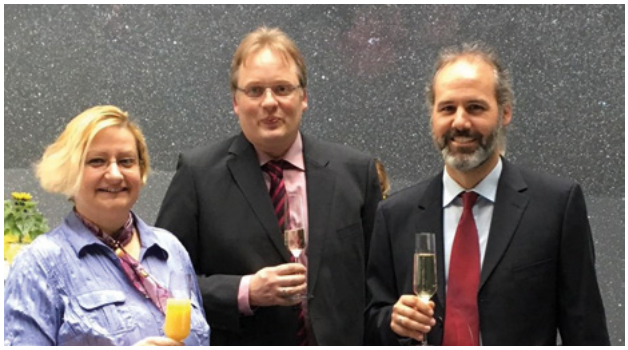
*The 360-degree tilted planetarium dome during a planetarium screening. (Picture: ESO/P. Horálek)*

## ESO Supernova

Stars, space, and the Universe hold an almost magical attraction. While astronomy is a vivid and rapidly evolving science, informing the public of researchers' latest discoveries is often difficult as the requirements of these scientists' daily work generally limit such opportunities. However, planetariums and exhibitions can bridge the gap between astronomical research and an interested public.

### A fruitful collaboration

After three years of construction, the *ESO Supernova Planetarium & Visitor Centre* was inaugurated on April 26, 2018. The new center is located at the European Southern Observatory headquarters in Garching near Munich, Germany. The state-of-the-art astronomy center was tailored to match both the public's fascination with and the significance of astronomy and astrophysics. It is the product of a collaboration between HITS and ESO, the foremost intergovernmental astronomy organization in Europe and by far the world's most productive ground-based astronomical observatory. The Klaus Tschira Foundation funded the construction of the premises, and ESO runs the facility. From the opening in April until the end of 2018, more than 55,000 people visited the ESO Supernova. The ESO team put together a supplemental program including special tours as well as planetarium shows and workshops for kindergartens, elementary-, and secondary-school students. Training sessions for teachers are also offered at the Supernova.



*The HITS team (f.l.t.r.): Dorotea Dudaš, Kai Polsterer and Volker Gaibler.*

### Virtual reality and interactive simulations

The ESO Supernova project group at HITS guided the exhibitions' conception and contributed scientific and technical expertise. Moreover, the group explored and developed interactive exhibits specifically for the ESO Supernova. Aided by computer simulations, virtual reality, and innovative computer graphics, visitors can dive into a world that is both diverse and spellbinding. They discover astronomy at both an individual and personal level, deepen their understanding of it, and share their experience and fascination with others. The project's two developers – Dr. Dorotea Dudaš and Dr. Volker Gaibler – possess a broad range of professional expertise, from computer graphics and numerical mathematics to astronomy and theoretical astrophysics. Dr. Kai Polsterer, leader of the Astroinformatics group, served as Project Manager.

*The ESO Supernova Planetarium & Visitor Centre by night. (Picture: ESO/P. Horálek)*

*The ESO Supernova exhibition. (Picture: ESO/P. Horálek)*

*HITS group leader Vincent Heuveline talking to the young researchers. (Picture: Kreutzer/HLFF)*

## Heidelberg Laureate Forum

The Heidelberg Laureate Forum (HLF) has taken place annually in Heidelberg since 2013. 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.

### 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, as was intended by its founder, Klaus Tschira.

### HITSters meet the young researchers

As in the previous five years, HITS hosted a group of young researchers from HLF 2018 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.

*Figure 85: The HLF young researchers at HITS.*


*Impressions from the visit of the young researchers at HITS. (Pictures: Kreutzer/HLFF, HITS)*

*HITSter Kira Feldmann (CST, left) during the poster session. (Picture: Kreutzer/HLFF)*

# 7
# Publications


**Alessandrini D.** Higgs bundles and geometric structures on manifolds. Prepr. (2018) arXiv:1809.07290.

**Alessandrini D, Lee GS, Schaffhauser F.** Hitchin components for orbifolds. Prepr. (2018) arXiv:1811.05366.

**Alfano M, Hovy D, Mitchell M, Strube M (Eds.).** "Proceedings of the 2nd ACL Workshop on Ethics in Natural Language Processing, New Orleans, Louis., 5 June 2018" (2018).

**Arnault J, Rummler T, Baur F, Lerch S, Wagner S, Fersch B, Zhang Z, Kerandi N, Keil C, Kunstmann H.** Precipitation sensitivity to the uncertainty of terrestrial water flow in WRF-Hydro: An ensemble analysis for Central Europe. J. Hydrometeorol. (2018) 19:1007–1025.

**Baba J, Kawata D, Matsunaga N, Grand RJJ, Hunt JAS.** Gaia DR1 Evidence of Disrupting the Perseus Arm. Astrophys. Journal (2018) 853:L23.

**Ballas S, Danciger J, Lee GS.** Convex projective structures on nonhyperbolic three-manifolds. Geom. Topol. (2018) 22:1593–1646.

**Bannister A, Müller FG, Müller MC, Teschke O.** On the road to MSC 2020 (editorial). Newsl. Eur. Math. Soc. (2018) pp. 3–4.

**Baran S, Lerch S.** Combining predictive distributions for the statistical postprocessing of ensemble forecasts. Int. J. Forecast. (2018) 34:477–496.

**Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A.** Epa-ng: Massively parallel evolutionary placement of genetic sequences. Syst. Biol. (2018) p. advance online access.

**Barbot T, Lee GS, Valério VP.** Pappus theorem, Schwartz representations and Anosov representations. Ann. Inst. Fourier (Grenoble) (2018) 68:2697–2741.

**Barna B, Szalai T, Kerzendorf WE, Kromer M, Sim SA, Magee MR, Leibundgut B.** Type Iax supernovae as a few-parameter family. Mon. Notices Royal Astron. Soc. (2018) 480:3609–3627.

**Barnes DJ, Vogelsberger M, Kannan R, Marinacci F, Weinberger R, Springel V, Torrey P, Pillepich A, Nelson D, Pakmor R, Naiman J, Hernquist L, McDonald M.** A census of cool-core galaxy clusters in IllustrisTNG. Mon. Notices Royal Astron. Soc. (2018) 481:1809–1831.

**Bass D, Czech L, Williams BA, Berney C, Dunthorn M, Mahé F, Torruella G, Stentiford GD, Williams TA.** Clarifying the relationships between microsporidia and cryptomycota. J. Eukaryot. Microbiol. (2018) pp. 773–782.

**Begolo D, Vincent IM, Giordani F, Pöhner I, Witty MJ, Rowan TG, Bengaly Z, Gillingwater K, Freund Y, Wade RC, Barrett MP, Clayton C.** The trypanocidal benzoxaborole AN7973 inhibits trypanosome mRNA processing. PLoS Pathog. (2018) 14:e1007315.

**Beyrer J.** A complete description of the antipodal set of most symmetric spaces of compact type. Osaka J. Math (2018) 55:567–586.

**Beyrer J, Fioravanti E.** Cubulations and cross ratios on contracting boundaries. Prepr. (2018) arXiv:1810.08087.

**Bozzola G, Stergioulas N, Bauswein A.** Universal relations for differentially rotating relativistic stars at the threshold to collapse. Mon. Notices Royal Astron. Soc. (2018) 474:3557–3564.

**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, Treib N.** Schottky presentations of positive representations. Prepr. (2018) arXiv:1807.05286.

**Bustamante S, Sparre M, Springel V, Grand RJJ.** Merger-induced metallicity dilution in cosmological galaxy formation simulations. Mon. Notices Royal Astron. Soc. (2018) 479:3381–3392.

**Butera D, Passam F, Ju L, Cook KM, Woon H, Aponte-Santamaría C, Gardiner E, Davis AK, Murphy DA, Bronowska A, Luken BM, Baldauf C, Jackson S, Andrews R, Gräter F, Hogg PJ.** Autoregulation of von Willebrand factor function by a disulfide bond switch. Sci. advances (2018) 4:eaaq1477.

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

**Choi S, Lee GS, Marquis L.** Deformations of convex real projective manifolds and orbifolds. In "Handbook of group actions. Vol. III". Int. Press, Somerville, MA (2018) vol. 40 of Adv. Lect. Math. (ALM), pp. 263–310.

**Correa CA, Schaye J, van de Voort F, Duffy AR, Wyithe JSB.** The impact of feedback and the hot halo on the rates of gas accretion on to galaxies. Mon. Notices Royal Astron. Soc. (2018) 478:255–269.

**Czech L, Stamatakis A.** Methods for automatic reference trees and multilevel phylogenetic placement. Bioinforma. (2018) p. advance online access.

**Czech L, Stamatakis A.** Scalable methods for post-processing, visualizing, and analyzing phylogenetic placements. bioRxiv (2018) p. 346353.

**Darriba D, Flouri T, Stamatakis A.** The state of software for evolutionary biology. Mol. Biol. Evol. (2018) 35:1037–1046.

**Davis TA, Bureau M, Onishi K, van de Voort F, Cappellari M, Iguchi S, Liu L, North EV, Sarzi M, Smith MD.** WISDOM Project - III. Molecular gas measurement of the supermassive black hole mass in the barred lenticular galaxy NGC4429. Mon. Notices Royal Astron. Soc. (2018) 473:3818–3834.

**D´Isanto, A, Cavuoti, S, Gieseke, F, Polsterer, KL.** Return of the features - efficient feature selection and interpretation for photometric redshifts. A&A (2018) 616:A97.

**D'Isanto A, Polsterer KL.** Photometric redshift estimation via deep learning. Generalized and pre-classification-less, image based, fully probabilistic redshifts. A&A (2018) 609:A111.

**Disarlo V.** Generalized stretch lines for surfaces with boundary. Oberwolfach Rep. No. 40 (2018) pp. 32–35.

**Disarlo V, Parlier H.** Simultaneous ips on triangulated surfaces. Mich. Math. J. (2018) 67:451–464.

**Disarlo V, Randecker A, Tang R.** Rigidity of the saddle connection complex. Prepr. (2018) arXiv:1810.00961.

**Douglas RG, Nandekar P, Aktories JE, Kumar H, Weber R, Sattler JM, Singer M, Lepper S, Sadiq SK, Wade RC, Frischknecht F.** Inter-subunit interactions drive divergent dynamics in mammalian and Plasmodium actin filaments. PLoS biology (2018) 16:e2005345.

**Duchesne B, Lécureux J, Pozzetti MB.** Boundary maps and maximal representations on infinite dimensional Hermitian symmetric spaces. Prepr. (2018) arXiv:1810.10208.

**Ehlert K, Weinberger R, Pfrommer C, Pakmor R, Springel V.** Simulations of the dynamics of magnetized jets and cosmic rays in galaxy clusters. Mon. Notices Royal Astron. Soc. (2018) 481:2878–2900.

**Ehm W, Krüger F.** Forecast dominance testing via sign randomization. Electron. J. Stat. (2018) 12:3758–3793.

**Eiserhardt WL, Antonelli A, Bennett DJ, Botigué LR, Burleigh JG, Dodsworth S, Enquist BJ, Forest F, Kim JT, Kozlov AM, et al.** A roadmap for global synthesis of the plant tree of life. Am. J. Bot. (2018) 105:614–622.

**Ergon M, Fransson C, Jerkstrand A, Kozma C, Kromer M, Spricer K.** Monte-Carlo methods for NLTE spectral synthesis of supernovae. A&A (2018) 620:A156.

**Fink M, Kromer M, Hillebrandt W, Röpke FK, Pakmor R, Seitenzahl IR, Sim SA.** Thermonuclear explosions of rapidly differentially rotating white dwarfs: Candidates for superluminous Type Ia supernovae? A&A (2018) 618:A124.

**Foerster JM, Pöhner I, Ullmann GM.** MCMap - A computational tool for mapping energy landscapes of transient protein-protein interactions. ACS Omega (2018) 3:6465–6475.

**Franz F, Aponte-Santamaría C, Daday C, Miletic V, Gräter F.** Stability of biological membranes upon mechanical indentation. The journal physical chemistry. B (2018) 122:7073–7079.

**Friedman R, Khalid S, Aponte-Santamaría C, Arutyunova E, Becker M, Boyd KJ, Christensen M, Coimbra JTS, Concilio S, Daday C, van Eerden FJ, Fernandes PA, Gräter F, Hakobyan D, Heuer A, Karathanou K, Keller F, Lemieux MJ, Marrink SJ, May ER, Mazumdar A, Naftalin R, Pickholz M, Piotto S, Pohl P, Quinn P, Ramos MJ, Schiøtt B, Sengupta D, Sessa L, Vanni S, Zeppelin T, Zoni V, Bondar AN, Domene C.** Understanding conformational dynamics of complex lipid mixtures relevant to biology. The J. membrane biology (2018) 251:609–631.

**Gall EEE, Kotak R, Leibundgut B, Taubenberger S, Hillebrandt W, Kromer M, Burgett WS, Chambers K, Flewelling H, Huber ME, Kaiser N, Kudritzki RP, Magnier EA, Metcalfe N, Smith K, Tonry JL, Wainscoat RJ, Waters C.** An updated Type II supernova Hubble diagram. A&A (2018) 611:A25.

**Ganotra GK, Wade RC.** Prediction of drug–target binding kinetics by comparative binding energy analysis. ACS Medicinal Chem. Lett. (2018) 9:1134–1139.

**Garrison-Kimmel S, Hopkins PF, Wetzel A, El-Badry K, Sanderson RE, Bullock JS, Ma X, van de Voort F, Hafen Z, Faucher-Giguère CA, Hayward CC, Quataert E, Kereš D, Boylan-Kolchin M.** The origin of the diverse morphologies and kinematics of Milky Way-mass galaxies in the FIRE-2 simulations. Mon. Notices Royal Astron. Soc. (2018) 481:4133–4157.

**Gawlok S, Heuveline V.** Nested schur-complement solver for a low-mach number model: Application to a cyclone-cyclone interaction. Prepr. Ser. Eng. Math. Comput. Lab (2018) Vol 02.

**Genel S, Nelson D, Pillepich A, Springel V, Pakmor R, Weinberger R, Hernquist L, Naiman J, Vogelsberger M, Marinacci F, Torrey P.** The size evolution of star-forming and quenched galaxies in the IllustrisTNG simulation. Mon. Notices Royal Astron. Soc. (2018) 474:3976–3996.

**Giesse, Sebastian:** "Inferring Species Trees by Minimizing Quartet-based Uncertainty", Master's thesis, Computer Science, Karlsruhe Institute of Technology, and HITS: Alexandros Stamatakis (2018).

**Gneiting T, Asher J, Carriquiry A, Davis R, Dawid AP, Efron B, Haberman S, Kou S, Newton M, Paddock S, Prewitt K, Raftery A, Stein M, Straf M.** Special section in memory of Stephen E. Fienberg (1942–2016). AOAS Editor-in-Chief 2013–2015. Annals Appl. Stat. (2018) 12:iii–x.

**Goicovic FG, Maureira-Fredes C, Sesana A, Amaro-Seoane P, Cuadra J.** Accretion of clumpy cold gas onto massive black hole binaries: a possible fast route to binary coalescence. Mon. Notices Royal Astron. Soc. (2018) 479:3438–3455.

**Golebiewski M.** Data formats for systems biology and quantitative modeling. In "Encyclopedia of Bioinformatics and Computational Biology", Eds. Ranganathan S, Gribskov M, Nakai K, Schönbach C. Elsevier, Oxford vol. 2, p. 884–893.

# 7
# Publications


**Goriely S, Bauswein A, Janka HT, Just O, Pllumbi E.** The r-process nucleosynthesis and related challenges. In "European Physical Journal Web of Conferences" (2018) vol. 165 of European Physical Journal Web of Conferences, p. 01025.

**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.

**Grand RJJ, Helly J, Fattahi A, Cautun M, Cole S, Cooper AP, Deason AJ, Frenk C, Gómez FA, Hunt JAS, Marinacci F, Pakmor R, Simpson CM, Springel V, Xu D.** Aurigaia: mock Gaia DR2 stellar catalogues from the AURIGA cosmological simulations. Mon. Notices Royal Astron. Soc. (2018) 481:1726–1743.

**Gupta A, Yuan T, Torrey P, Vogelsberger M, Martizzi D, Tran KVH, Kewley LJ, Marinacci F, Nelson D, Pillepich A, Hernquist L, Genel S, Springel V.** Chemical pre-processing of cluster galaxies over the past 10 billion years in the IllustrisTNG simulations. Mon. Notices Royal Astron. Soc. (2018) 477:L35–L39.

**Halling P, Fitzpatrick PF, Raushel FM, Rohwer J, Schnell S, Wittig U,Wohlgemuth R, Kettner C.** An empirical analysis of enzyme function reporting for experimental reproducibility: Missing/incomplete information in published papers. Biophys. Chem. (2018) 242:22–27.

**Heinzerling B, Strube M.** BPEmb: Tokenization-free pre-trained subword embeddings in 275 languages. In "Proceedings of the 11th International Conference on Language Resources and Evaluation, Miyazaki, Japan, 7–12 May 2018" (2018).

**Heroven C, Georgi V, Ganotra GK, Brennan P, Wolfreys F, Wade RC, Fernández-Montalván AE, Chaikuad A, Knapp S.** Halogen–aromatic _ interactions modulate inhibitor residence times. Angewandte Chemie Int. Ed. (2018) 57:7220–7224.

**Heroven C, Georgi V, Ganotra GK, Brennan P, Wolfreys F, Wade RC, Fernández-Montalván AE, Chaikuad A, Knapp S.** Halogenaromatische π-Wechselwirkungen modulieren die Verweilzeit von Inhibitoren. Angewandte Chemie (2018) 130:7338–7343.

**Hillebrandt W, Bulla M, Kromer M, Pakmor R, Röpke FK, Schrauth M, Sim SA.** The Progenitor Systems of Thermonuclear Supernovae. In "NIC Symposium 2018". NIC Symposium 2018, Jülich (Germany), 22-23 Feb 2018, Forschungszentrum Jülich GmbH, Zentralbibliothek, Jülich (2018) vol. 49 of NIC Series, pp. 99 – 106.

**Hoang DT, Vinh LS, Flouri T, Stamatakis A, von Haeseler A, Minh BQ.** Mpboot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. BMC Evol. Biol. (2018) 18:11.

**Hou Y, Markert K, Strube M.** Unrestricted bridging resolution. Comput. Linguist. (2018) 44:237–284.

**Hsueh JW, Despali G, Vegetti S, Xu D, Fassnacht CD, Metcalf RB.** Flux-ratio anomalies from discs and other baryonic structures in the Illustris simulation. Mon. Notices Royal Astron. Soc. (2018) 475:2438–2451.

**Hunt JAS, Hong J, Bovy J, Kawata D, Grand RJJ.** Transient spiral structure and the disc velocity substructure in Gaia DR2. Mon. Notices Royal Astron. Soc. (2018) 481:3794–3803.

**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.

**Ji L, Schilling AS**. Toric Varieties vs. Horofunction Compactifications of Polyhedral Norms. L'Enseignement Mathèmatique (2018) 63:375–401.

**John D, Schick M, Heuveline V.** Bayesian inference for estimating model discrepancy of an electric motor. PAMM (2018) 18:1–2.

**John D, Schick M, Heuveline V.** Learning model discrepancy of an electric motor with bayesian inference. Prepr. Ser. Eng. Math. Comput. Lab (2018) Vol. 01.

**Johnson KP, Dietrich CH, Friedrich F, Beutel RG, Wipfler B, Peters RS, Allen JM, Petersen M, Donath A, Walden KK, et al.** Phylogenomics and the evolution of hemipteroid insects. Proc. Natl. Acad. Sci. (2018) p. advance online access.

**van de Kamp T, Schwermann AH, dos Santos Rolo T, Lösel PD, Engler T, Etter W, Faragó T, Göttlicher J, Heuveline V, Kopmann A, Mähler B, Mörs T, Odar J, Rust J, Tan Jerome N, Vogelgesang M, Baumbach T, Krogmann L.** Parasitoid biology preserved in mineralized fossils. Nat. Commun. (2018) 9:3325.

**Kawata D, Allende Prieto C, Brook CB, Casagrande L, Ciucă I, Gibson BK, Grand RJJ, Hayden MR, Hunt JAS.** Metallicity gradient of the thick disc progenitor at high redshift. Mon. Notices Royal Astron. Soc. (2018) 473:867–878.

**Kawata D, Baba J, Ciucă I, Cropper M, Grand RJJ, Hunt JAS, Seabroke G.** Radial distribution of stellar motions in Gaia DR2. Mon. Notices Royal Astron. Soc. (2018) 479:L108–L112.

**Kerzendorf WE, Strampelli G, Shen KJ, Schwab J, Pakmor R, Do T, Buchner J, Rest A.** A search for a surviving companion in SN 1006. Mon. Notices Royal Astron. Soc. (2018) 479:192–199.

**Kirilin A, Strube M.** Exploiting a speaker's credibility to detect fake news. In Proceedings of the Workshop on Data Science, Journalism and Media, London, UK, 20 August 2018 (2018) 7 pages.

**Kokh DB, Amaral M, Bomke J, Grädler U, Musil D, Buchstaller HP, Dreyer MK, Frech M, Lowinski M, Vallee F, Bianciotto M, Rak A, Wade RC.** Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. J. Chem. Theory Comput. (2018) 14:3859–3869. PMID: 29768913.

**Kozlov A, Darriba D, Flouri T, Morel B, Stamatakis A.** Raxml-ng: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. bioRxiv (2018) p. 447110.

**Kozyreva A, Kromer M, Noebauer UM, Hirschi R.** OGLE14-073 - a promising pair-instability supernova candidate. Mon. Notices Royal Astron. Soc. (2018) 479:3106–3114.

**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.

**Lovell MR, Pillepich A, Genel S, Nelson D, Springel V, Pakmor R, Marinacci F, Weinberger R, Torrey P, Vogelsberger M, Alabi A, Hernquist L.** The fraction of dark matter within galaxies from the IllustrisTNG simulations. Mon. Notices Royal Astron. Soc. (2018) 481:1950–1975.

**Maat JR, Gianniotis N, Protopapas P.** Efficient optimization of echo state networks for time series datasets. In "2018 International Joint Conference on Neural Networks (IJCNN)". IEEE (2018) pp. 1–7.

**Maffione NP, Gómez FA, Cincotta PM, Giordano CM, Grand RJJ, Marinacci F, Pakmor R, Simpson CM, Springel V, Frenk CS.** On the relevance of chaos for halo stars in the solar neighbourhood II. Mon. Notices Royal Astron. Soc. (2018) 478:4052–4067.

**Marinacci F, Vogelsberger M, Kannan R, Mocz P, Pakmor R, Springel V.** Nonideal magnetohydrodynamics on a moving mesh. Mon. Notices Royal Astron. Soc. (2018) 476:2476–2492.

**Marinacci F, Vogelsberger M, Pakmor R, Torrey P, Springel V, Hernquist L, Nelson D, Weinberger R, Pillepich A, Naiman J, Genel S.** First results from the IllustrisTNG simulations: radio haloes and magnetic fields. Mon. Notices Royal Astron. Soc. (2018) 480:5113–5139.

**Maureira-Fredes C, Goicovic FG, Amaro-Seoane P, Sesana A.** Accretion of clumpy cold gas on to massive black hole binaries: the challenging formation of extended circumbinary structures. Mon. Notices Royal Astron. Soc. (2018) 478:1726–1748.

**Mercadante D, Gräter F, Daday C.** Conan: A tool to decode dynamical information from molecular interaction maps. Biophys. journal (2018) 114:1267–1273.

**Mesgar M, Strube M.** A neural local coherence model for text quality assessment. In "Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October – 4 November 2018" (2018) pp. 4328–4339.

**Mexner W, Bonn M, Kopmann A, Mauch V, Ressmann D, Chilingaryan SA, Jerome NT, van de Kamp T, Heuveline V, Lösel P.** Opengl Rapi-based analysis of large datasets in a cloud environment. In "Design and Use of Virtualization Technology in Cloud Computing". IGI Global (2018) pp. 161–181.

**Mohseni M, Faili H.** Title generation and keyphrase extraction from Persian scientific texts. In "Proceedings of the 9th International Symposium on Telecommunications, Tehran, Iran, 17-19 December, 2018" (2018).

**Moosavi NS, Strube M.** Using linguistic features to improve generalization in neural coreference resolvers. In "Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October – 4 November 2018" (2018) pp. 193–203.

**Morel B, Kozlov AM, Stamatakis A.** Pargenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. Bioinformatics 2018, 1-3, bty839, advance access publication date: 15 October 2018.

**Müller MC.** On the contribution of word-level semantics to practical author name disambiguation. In "Proceedings of the 18th ACM/IEEE Joint Conference on Digital Libraries, Fort Worth, Texas, 3–7 June 2018" (2018) pp. 367–368.

**Müller MC, Strube M.** Transparent, efficient, and robust word embedding access with WOMBAT. In "Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations, Santa Fé, New Mexico, 20–26 August 2018" (2018) pp. 53–57.

**Naiman JP, Pillepich A, Springel V, Ramirez-Ruiz E, Torrey P, Vogelsberger M, Pakmor R, Nelson D, Marinacci F, Hernquist L, Weinberger R, Genel S.** First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. Mon. Notices Royal Astron. Soc. (2018) 477:1206–1224.

**Neal ML, König M, Nickerson D, Misirli G, Kalbasi R, Dräger A, Atalag K, Chelliah V, Cooling MT, Cook DL, Crook S, de Alba M, Friedman SH, Garny A, Gennari JH, Gleeson P, Golebiewski M, Hucka M, Juty N, Myers C, Olivier BG, Sauro HM, Scharm M, Snoep JL, Touré V, Wipat A, Wolkenhauer O, Waltemath D.** Harmonizing semantic annotations for computational models in biology. Briefings Bioinforma. (2018).

**Nelson D, Kauffmann G, Pillepich A, Genel S, Springel V, Pakmor R, Hernquist L, Weinberger R, Torrey P, Vogelsberger M, Marinacci F.** The abundance, distribution, and physical nature of highly ionized oxygen O VI, O VII, and O VIII in IllustrisTNG. Mon. Notices Royal Astron. Soc. (2018) 477:450–479.

**Nelson D, Pillepich A, Springel V, Weinberger R, Hernquist L, Pakmor R, Genel S, Torrey P, Vogelsberger M, Kauffmann G, Marinacci F, Naiman J.** First results from the IllustrisTNG simulations: the galaxy colour bimodality. Mon. Notices Royal Astron. Soc. (2018) 475:624–647.

**Ott A.** Transgression in bounded cohomology and a conjecture of Monod. Prepr. (2018) arXiv:1811.07558.

**Öztürk MA, Cojocaru V, Wade RC.** Dependence of chromatosome structure on linker histone sequence and posttranslational modification. Biophys. J. (2018) 114:2363–2375.

**Öztürk MA, Cojocaru V, Wade RC.** Toward an ensemble view of chromatosome structure: A paradigm shift from one to many. Struct. (2018) 26:1050–1057.

**Pais M, Pfrommer C, Ehlert K, Pakmor R.** The effect of cosmic ray acceleration on supernova blast wave dynamics. Mon. Notices Royal Astron. Soc. (2018) 478:5278–5295.

**Pakmor R, Guillet T, Pfrommer C, Gómez FA, Grand RJJ, Marinacci F, Simpson CM, Springel V.** Faraday rotation maps of disc galaxies. Mon. Notices Royal Astron. Soc. (2018) 481:4410–4418.

**Panov IV, Glazyrin SI, Röpke FK, Blinnikov SI.** Nucleosynthesis during a Thermonuclear Supernova Explosion. Astron. Lett. (2018) 44:309–314.

# 7
# Publications


**Pantillon F, Lerch S, Knippertz P, Corsmeier U.** Forecasting wind gusts in winter storms using a calibrated convection-permitting ensemble. Q. J. Royal Meteorol. Soc. (2018) 144:1864–1881.

**Passam F, Chiu J, Ju L, Pijning A, Jahan Z, Mor-Cohen R, Yeheskel A, Kolšek K, Thärichen L, Aponte-Santamaría C, Gräter F, Hogg PJ.** Mechano-redox control of integrin de-adhesion. eLife (2018) 7.

**Pauli T, Burt TO, Meusemann K, Bayless K, Donath A, Podsiadlowski L, Mayer C, Kozlov A, Vasilikopoulos A, Liu S, et al.** New data, same story: phylogenomics does not support syrphoidea (diptera: Syrphidae, pipunculidae). Syst. Entomol. (2018) 43:447–459.

**Peters RS, Niehuis O, Gunkel S, Bläser M, Mayer C, Podsiadlowski L, Kozlov A, Donath A, van Noort S, Liu S, et al.** Transcriptome sequence-based phylogeny of chalcidoid wasps (hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. Mol. Phylogenetics Evol. (2018) 120:286–296.

**Pillepich A, Nelson D, Hernquist L, Springel V, Pakmor R, Torrey P, Weinberger R, Genel S, Naiman JP, Marinacci F, Vogelsberger M.** First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. Mon. Notices Royal Astron. Soc. (2018) 475:648–675.

**Pillepich A, Springel V, Nelson D, Genel S, Naiman J, Pakmor R, Hernquist L, Torrey P, Vogelsberger M, Weinberger R, Marinacci F.** Simulating galaxy formation with the IllustrisTNG model. Mon. Notices Royal Astron. Soc. (2018) 473:4077–4106.

**Psonis N, Antoniou A, Karameta E, Leaché AD, Kotsakiozi P, Darriba D, Kozlov A, Stamatakis A, Poursanidis D, Kukushkin O, et al.** Resolving complex phylogeographic patterns in the balkan peninsula using closely related walllizard species as a model system. Mol. Phylogenetics Evol. (2018) 125:100–115.

**Putzu M, Gräter F, Elstner M, Kubar T.** On the mechanism of spontaneous thiol-disulfide exchange in proteins. Phys. chemistry chemical physics: PCCP (2018) 20:16222–16230.

**Rasp S, Lerch S.** Neural networks for postprocessing ensemble weather forecasts. Mon. Weather Rev. (2018) 146:3885–3900.

**Ray A, Gräter F, Thukral L.** Probing molecular forces in multi-component physiological membranes. Phys. chemistry chemical physics: PCCP (2018) 20:2155–2161.

**Röpke FK, Berberich J, Edelmann PFV, Horst L, Jones SW, Michel A, Botto Poala A.** Towards Multidimensional Hydrodynamic Simulations of Stars. In "NIC Symposium 2018". NIC Symposium 2018, Jülich (Germany), 22 Feb 2018 - 23 Feb 2018, Forschungszentrum Jülich GmbH, Zentralbibliothek, Jülich (2018) vol. 49 of NIC Series, pp. 115 – 122.

**Röpke FK, Sim SA.** Models for Type Ia Supernovae and Related Astrophysical Transients. Space Sci. Rev. (2018) 214:72.

**Sadiq SK, Mirambeau G, Meyerhans A.** Equilibrium model of drug-modulated GagPol-embedded HIV-1 reverse transcriptase dimerization to enhance premature protease activation. AIDS Res. Hum. Retroviruses (2018) 34:804–807.

**Sanders A.** The pre-symplectic geometry of opers and the holonomy map. Prepr. (2018) arXiv:1804.04716.

**Sann M, Niehuis O, Peters RS, Mayer C, Kozlov A, Podsiadlowski L, Bank S, Meusemann K, Misof B, Bleidorn C, et al.** Phylogenomic analysis of apoidea sheds new light on the sister group of bees. BMC Evol. Biol. (2018) 18:71.

**Schmidt TT, Sharma S, Reyes GX, Gries K, Gross M, Zhao B, Yuan JH, Wade RC, Chabes A, Hombauer H.** A genetic screen pinpoints ribonucleotide reductase residues that sustain dNTP homeostasis and specifies a highly mutagenic type of dNTP imbalance. Nucleic Acids Res. (2019) 47:237–252.

**Schreiber F, Bader GD, Gleeson P, Golebiewski M, Hucka M, Keating S, Le Novère N, Myers CJ, Nickerson D, Sommer B, Waltemath D.** Specifications of standards in systems and synthetic biology: Status and developments in 2017. J. Integr. Bioinform. (2018) 15.

**Seelig T, Meyer A, Gerstner P, Meier M, Jongmanns M, Baumann M, Heuveline V, Egbers C.** Dielectrophoretic force-driven convection in annular geometry under earth's gravity. arXiv preprint arXiv:1812.05460 (2018).

**Sekulic I, Gjurkovic M, Šnajder J.** Not just depressed: Bipolar disorder prediction on Reddit. In "Proceedings of the 9th Workshop on Computational Approaches to Subjectivity and Social Media Analysis, Brussels, Belgium, 31 October 2018" (2018) pp. 72–78.

**Shao S, Cautun M, Frenk CS, Grand RJJ, Gómez FA, Marinacci F, Simpson CM.** The multiplicity and anisotropy of galactic satellite accretion. Mon. Notices Royal Astron. Soc. (2018) 476:1796–1810.

**Shen KJ, Boubert D, Gänsicke BT, Jha SW, Andrews JE, Chomiuk L, Foley RJ, Fraser M, Gromadzki M, Guillochon J, Kotze MM, Maguire K, Siebert MR, Smith N, Strader J, Badenes C, Kerzendorf WE, Koester D, Kromer M, Miles B, Pakmor R, Schwab J, Toloza O, Toonen S, Townsley DM, Williams BJ.** Three Hypervelocity White Dwarfs in Gaia DR2: Evidence for Dynamically Driven Double-degenerate Double-detonation Type Ia Supernovae. Astrophys. J. (2018) 865:15.

**Simpson CM, Grand RJJ, Gómez FA, Marinacci F, Pakmor R, Springel V, Campbell DJR, Frenk CS.** Quenching and ram pressure stripping of simulated Milky Way satellite galaxies. Mon. Notices Royal Astron. Soc. (2018) 478:548–567.

**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.

**Stamatakis A.** In "Population Genomics". Springer (2018) chapter Population and Evolutionary Genetic Inferences in the Whole-Genome Era: Software Challenges.

**Stecker F.** Balanced ideals and domains of discontinuity of Anosov representations. Prepr. (2018) arXiv:1810.11496.

**Stecker F, Treib N.** Domains of discontinuity in oriented flag manifolds. Prepr. (2018) arXiv:1806.04459.

**Sultan CS, Saackel A, Stank A, Fleming T, Fedorova M, Hoffmann R, Wade RC, Hecker M, Wagner AH.** Impact of carbonylation on glutathione peroxidase-1 activity in human hyperglycemic endothelial cells. Redox Biol. (2018) 16:113 – 122.

**Suter J, Strube M.** Extending and exploiting the entity graph for analysis, classification and visualization of German texts. In "Proceedings of the 14th Conference on Natural Language Processing (KONVENS), Vienna, Austria, 17–19 September 2018" (2018) pp. 136–140.

**Swainston N, Baici A, Bakker BM, Cornish-Bowden A, Fitzpatrick PF, Halling P, Leyh TS, O' Donovan C, Raushel FM, Reschel U, Rohwer JM, Schnell S, Schomburg D, Tipton KF, Tsai MD, Westerhoff HV, Wittig U, Wohlgemuth R, Kettner C.** STRENDA DB: enabling the validation and sharing of enzyme kinetics data. FEBS J. (2018) 285:2193–2204.

**Tan PS, Aramburu IV, Mercadante D, Tyagi S, Chowdhury A, Spitz D, Shammas SL, Gräter F, Lemke EA.** Two differential binding mechanisms of FG-nucleoporins and nuclear transport receptors. Cell reports (2018) 22:3660–3671.

**Torrey P, Vogelsberger M, Hernquist L, McKinnon R, Marinacci F, Simcoe RA, Springel V, Pillepich A, Naiman J, Pakmor R, Weinberger R, Nelson D, Genel S.** Similar star formation rate and metallicity variability time-scales drive the fundamental metallicity relation. Mon. Notices Royal Astron. Soc. (2018) 477:L16–L20.

**Trefzer A, Stamatakis A.** Compressing streams of phylogenetic trees. bioRxiv (2018) p. 440644.

**van de Voort F, Davis TA, Matsushita S, Rowlands K, Shabala SS, Allison JR, Ting YS, Sansom AE, van der Werf PP.** An ALMA view of star formation efficiency suppression in early-type galaxies after gas-rich minor mergers. Mon. Notices Royal Astron. Soc. (2018) 476:122–132.

**van de Voort F, Quataert E, Faucher-Giguère CA, Kereš D, Hopkins PF, Chan TK, Feldmann R, Hafen Z.** On the deuterium abundance and the importance of stellar mass loss in the interstellar and intergalactic medium. Mon. Notices Royal Astron. Soc. (2018) 477:80–92.

**Villaescusa-Navarro F, Genel S, Castorina E, Obuljen A, Spergel DN, Hernquist L, Nelson D, Carucci IP, Pillepich A, Marinacci F, Diemer B, Vogelsberger M, Weinberger R, Pakmor R.** Ingredients for 21 cm Intensity Mapping. Astrophys. J. (2018) 866:135.

**Vogel P, Knippertz P, Fink AH, Schlueter A, Gneiting T.** Skill of global raw and postprocessed ensemble predictions of rainfall over northern tropical Africa. Weather Forecast. (2018) 33:369–388.

**Vogelsberger M, Marinacci F, Torrey P, Genel S, Springel V, Weinberger R, Pakmor R, Hernquist L, Naiman J, Pillepich A, Nelson D.** The uniformity and time-invariance of the intra-cluster metal distribution in galaxy clusters from the IllustrisTNG simulations. Mon. Notices Royal Astron. Soc. (2018) 474:2073–2093.

**Wagner J.** Generalised model-independent characterisation of strong gravitational lenses. III. Perturbed axisymmetric lenses. A&A (2018) 615:A102.

**Wagner J, Liesenborgs J, Tessore N.** Model-independent and model-based local lensing properties of CL0024+1654 from multiply imaged galaxies. A&A (2018) 612:A17.

**Wagner J, Tessore N.** Generalised model-independent characterisation of strong gravitational lenses. II. Transformation matrix between multiple images. A&A (2018) 613:A6.

**Weinberger R, Springel V, Pakmor R, Nelson D, Genel S, Pillepich A, Vogelsberger M, Marinacci F, Naiman J, Torrey P, Hernquist L.** Supermassive black holes and their feedback effects in the IllustrisTNG simulation. Mon. Notices Royal Astron. Soc. (2018) 479:4056–4072.

**Wienhard A.** An invitation to higher Teichmüller theory. Proc. Int. Cong. Math. 2018, Rio de Janeiro (2018) 1:1007–1034.

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

**Zhang W, Mu G, Song C, Yan G, Heuveline V.** Extraction of spatial-temporal features of bus loads in electric grids through clustering in a dynamic model space. IEEE Transactions on Sustain. Energy (2019) Submitted.

**Zhu L, Ven Gvd, Bosch Rvd, Rix HW, Lyubenova M, Falcón-Barroso J, Martig M, Mao S, Xu D, Jin Y, Obreja A, Grand RJJ, Dutton AA, Macciò AV, Gómez FA, Walcher JC, García-Benito R, Zibetti S, Sánchez SF.** The stellar orbit distribution in present-day galaxies inferred from the CALIFA survey. Nat. Astron. (2018) 2:233–238.

**Zhu Q, Xu D, Gaspari M, Rodriguez-Gomez V, Nelson D, Vogelsberger M, Torrey P, Pillepich A, Zjupa J, Weinberger R, Marinacci F, Pakmor R, Genel S, Li Y, Springel V, Hernquist L.** Formation of a Malin 1 analogue in IllustrisTNG by stimulated accretion. Mon. Notices Royal Astron. Soc. (2018) 480:L18–L22.

**Zopf M, Botschen T, Falke T, Heinzerling B, Marasovic A, Mihaylov T, PVS A, Loza Mencía E, Fürnkranz J, Frank A.** What's important in a text? An extensive evaluation of linguistic annotations for summarization. In "Proceedings of the 5th International Conference on Social Networks Analysis, Management and Security, Valencia, Spain, 15–18 October 2018" (2018) pp. 272–277.

**Zuluaga JI, Bustamante S.** Magnetic properties of Proxima Centauri b analogues. Planet. Space Sci. (2018) 152:55–67.

# 8 Teaching

## Degrees

**Bauswein, Andreas:**
*"Neutron star mergers"*, Habilitation thesis, Department of Physics and Astronomy, Heidelberg University, 2018.

**Dohrke, Jan-Niklas:**
*"Modelling of Neurotrophin Receptors"*, Bachelor thesis, Molecular Biotechnology, Department of Physics and Astronomy, Heidelberg University and HITS: Rebecca Wade and Goutam Mukherjee, 2018.

**Friedrich, Patrick:**
*"A novel method to treat gated protein-ligand binding via on-the-fly conformational updates within rigid body BD simulations"*, Master thesis, Faculty of Physics, Heidelberg University and HITS: Rebecca Wade, Kashif Sadiq and Neil Bruce, 2018.

**Giesse, Sebastian:**
*"Inferring Species Trees by Minimizing Quartet-based Uncertainty"*, Master thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis, 2018.

**Jacob, Svenja:**
*"Cosmic ray feedback in galaxy formation and a numerical model for turbulence"*, Ph.D. Thesis, Physics, Department of Physics and Astronomy, Heidelberg University and HITS: Volker Springel, 2018.

**Jäger, Tobias:**
*"Mechanical properties of polymer nano-composites from multi-scale simulations"*, Master thesis, Department of Physics and Astronomy, Heidelberg University (Ulrich Schwarz) and HITS: Frauke Gräter, 2018.

**Kozlov, Alexey:**
*"Models, Optimizations, and Tools for Large-Scale Phylogenetic Inference, Handling Sequence Uncertainty, and Taxonomic Validation"*, Ph.D. thesis, Computer Science, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis, 2018.

**Kratzke, Jonas:**
*"Uncertainty quantification for fluid-structure interaction"*, Ph.D. thesis, Mathematics, Medicine, Heidelberg University: Vincent Heuveline, 2018.

**Öztürk, Mehmet Ali:**
*"A computational approach to decipher chromatosome structure determinants"*, Ph.D. thesis, Combined Faculty for the Natural Sciences and Mathematics, Heidelberg University and HITS: Rebecca Wade, 2018.

**Pakmor, Rüdiger:**
*"New methods in computational astrophysics and their application to galaxies and supernovae."* Habilitation thesis, Department of Physics and Astronomy, Heidelberg University, 2018.

**Parveen, Daraksha:**
*"A Graph-based Approach for the Summarization of Scientific Articles"*, Ph.D. thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2018.

**Rülicke, Linda:**
*"Design and Implementation of I/O-Efficient Taxa Quartets Counting in the Context of Phylogenetic Analysis"*, Master thesis, University of Frankfurt, Ulrich Meyer and HITS: Alexandros Stamatakis, 2018.

**Sand, Christian:**
*"Hydrodynamics of an AGB star during the Common Envelope Phase"*, Master thesis, Department of Physics and Astronomy, Heidelberg University, advisor: Friedrich Röpke, 2018.

**Song, Chen:**
*"Uncertainty Quantification for a Blood Pump Device with Generalized Polynomial Chaos Expansion"*, Ph.D. thesis, Mathematics, Medicine, Heidelberg University: Vincent Heuveline, 2018.

**Sorgenfrei, Frieda:**
*"SAS-6 structure and dynamics from Molecular Dynamics simulations"*, Master thesis, Faculty of Biosciences, Heidelberg University (Ulrich Schwarz) and HITS: Frauke Gräter, 2018.

**Vavourakis, Odysseas:**
*"A large-scale study of spectrin repeat mechanical unfolding with atomistic force-probe MD"*, Bachelor thesis, Faculty of Biosciences, Heidelberg University and HITS: Frauke Gräter, 2018.

**Zjupa, Jolanta:**
*"The impact of feedback on galactic and extra-galactic scales"*, Ph.D. thesis, Physics, Department of Physics and Astronomy, Heidelberg University and HITS: Volker Springel, 2018.

## Lectures, Courses, Seminars

**Daniele Alessandrini:**

*"Thurston's theory of surfaces"*, Heidelberg University and Karlsruhe Institute of Technology, winter semester 2017 / 2018.

**Shinpei Baba:**

Seminar *"Riemann Surfaces and Teichmüller Theory"*, Heidelberg University, winter semester 2017 / 2018.

**Andreas Bauswein:**

*"Kollidierende Neutronensterne im Computer"* Lehrerfortbildung (update for high school teachers), Haus der Astronomie, Heidelberg, January 18, 2018. *"Astronomical Techniques"*, Guest lecturer, lecture course, Heidelberg University, March 15, 2018. *"Computational Astrophysics"*, substitute lecturer, lecture course, Heidelberg University, June 19, 2018.

**Federica Fanoni:**

*"Translation surfaces"*, Heidelberg University, summer semester 2018.

**Tilmann Gneiting:**

Seminar on *"Statistical forecasting and classification"*, Karlsruhe Institute of Technology, winter semester 2017 / 2018. Lecture course on *"Time series analysis"*, Karlsruhe Institute of Technology, summer semester 2018. Lecture course on *"Forecasting: Theory and practice I"*, Karlsruhe Institute of Technology, winter semester 2018 / 2019.

**Martin Golebiewski, Maja Rey, Wolfgang Müller:**

LiSyM / de.NBI / ERASysAPP Tutorial: *"How to Share FAIR – The FAIRDOM Data and Model Management Practice"*, 7th Conference on Systems Biology of Mammalian Cells (SBMC 2018), Bremen, Germany, July 4 – 6, 2018.

**Martin Golebiewski, Maja Rey, Andreas Weidemann, Ulrike Wittig:**

COMBINE & de.NBI Tutorial *"Modelling and Simulation Tools in Systems Biology"*, International Conference on Systems Biology (ICSB 2018), Lyon, France, October 27 – November 1st, 2018.

**Martin Golebiewski, Xiaoming Hu:**

*"LiSyM Data Management"*, LiSyM network retreat, Hünfeld, Germany, November 26 – 28, 2018.

**Frauke Gräter:**

Contribution to lecture *"Computational biochemistry"* for biochemistry master students, Heidelberg University, winter semester 2018 / 2019. Contribution to lecture *"Biophysical Chemistry"* for Molecular Biotechnology bachelor students, Heidelberg University, winter semester 2018 / 2019.

**Frauke Gräter, Csaba Daday, Rebecca Wade:**

M.Sc. Seminar course *"Machine Learning for the Molecular World"*, Heidelberg University, summer semester, 2018.

**Frauke Gräter, Fan Jin, Fabian Kutzki, Krisztina Feher, Rebecca Wade:**

Practical Course on *"Computational Molecular Biophysics"*, Heidelberg University, winter semester, 2018 / 2019.

**Sabrina Gronow:**

*"Theoretical Astrophysics"*, Exercises and tutorial accompanying the lecture course, Heidelberg University, Heidelberg, winter semester 2017 / 2018. *"Physik A"*, Exercises and tutorial accompanying the lecture course, Heidelberg University, Heidelberg, winter semester 2017 / 2018. *"Physik B"*, Exercises and tutorial accompanying the lecture course, Heidelberg University, Heidelberg, summer semester 2018.

**Vincent Heuveline:**

Compact course *"Uncertainty Quantification and High Performance Computing"*, Heidelberg Graduate School MathComp, February 28, 2018.

**Vincent Heuveline, Simon Gawlok:**

Lecture *"Uncertainty Quantification 2"*, Heidelberg University, summer semester 2018.

**Vincent Heuveline, Artur Andrzejak:**

Seminar *"Software Security Engineering"*, Heidelberg University, summer semester 2018.

**Vincent Heuveline, Chen Song, Philipp Gerstner, Jonas Kratzke, Philipp Lösel:**

Seminar *"Advanced Topics in Uncertainty Quantification"*, Heidelberg University, winter semester 2018 / 2019.

# 8 Teaching

**Vincent Heuveline, Saskia Haupt, Chen Song, Philipp Gerstner, Jonas Kratzke, Philipp Lösel:**
Seminar *"Mathematical Methods for Medicine"*, Heidelberg University, winter semester 2018 / 2019.

**Vincent Heuveline, Maximilian Hoecker:**
Lecture *"IT Security"*, Heidelberg University, winter semester 2018 / 2019.

**Leonhard Horst:**
*"Theoretical Astrophysics"*, Exercises and tutorial accompanying the lecture course, Heidelberg University, winter semester 2017 / 2018.

**Olga Krebs:**
*"Systems biology and computer modeling in personalized medicine"* and *"Topological analysis of the stoichiometric models of metabolic networks"*, The 10th Young Scientists School *"Systems biology and Bioinformatics"* SBB-2018, Novosibirsk, Russia, August 27 – 31, 2018.

**Markus Kromer:**
*"Python programming for scientists"*, lecture course, Heidelberg University, winter semester 2017 / 2018. *"Python programming for scientists"*, lecture course, Heidelberg University, summer semester 2018.

**Florian Lach:**
*"Theoretical Astrophysics"*, Exercises and tutorial accompanying the lecture course, Heidelberg University, winter semester 2017 / 2018. *"Computational Astrophysics"*, Exercises and tutorial accompanying the lecture course, Heidelberg University, summer semester 2018.

**Gye-Seon Lee:**
RTG Lecture *"Coxeter groups and geometry"*, Heidelberg University and Karlsruhe Institute of Technology, summer semester 2018.

**Sebastian Lerch:**
Lecture course on *"Probability and statistics for computer scientists"*, Karlsruhe Institute of Technology, winter semester 2018 / 2019.

**Wolfgang Müller:**
*"Augmented Reality"*, Bachelor Seminar Media Informatics, University of Bamberg, Germany.

**Wolfgang Müller, Marcel Petrov, Sucheta Ghosh:**
XXIII International Summer Science School (ISH 2018), Heidelberg, Germany, July 30 – Aug 16, 2018.

**Wolfgang Müller, Ulrike Wittig:**
*"de.NBI Summer School 2018: Riding the Data Life Cycle"*, Braunschweig, Germany, September 3 – 7, 2018.

**Maria Beatrice Pozzetti:**
Seminar *"Groups acting on trees"*, Heidelberg University, summer semester 2018.

**Maria Beatrice Pozzetti, Daniele Alessandrini:**
*"Differentialgeometrie II – Symmetric spaces"*, Heidelberg University, winter semester 2018 / 2019.

**Johannes Resin:**
Exercise course on *"Probability and statistics for computer scientists"*, Karlsruhe Institute of Technology, winter semester 2018 / 2019.

**Maja Rey, Andreas Weidemann, Ulrike Wittig:**
de.NBI Course *"Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, SEEK"*, Magdeburg, Germany, April 24 – 26, 2018.

**Friedrich Röpke:**
*"Computational Astrophysics"*, lecture course, Heidelberg University, summer semester 2018. *"Fundamentals of Simulation Methods"*, lecture course, given jointly with Prof. Cornelis Dullemond, Heidelberg University, summer semester 2018.

**Friedrich Röpke, Markus Kromer:**
*"Theoretical Astrophysics"*, lecture course, Heidelberg University, Department of Physics and Astronomy, winter semester 2017 / 2018.

**Kashif Sadiq, Rebecca Wade:**
Ringvorlesung *"Structure and Dynamics of Biological Macromolecules"*, B.Sc. Biosciences, Heidelberg University, summer semester, 2018.

**Andrew Sanders:**
RTG Lecture *"An introduction to geometric representation theory"*, Heidelberg University and Karlsruhe Institute of Technology, winter semester 2018 / 2019.

**Theodoros Soultanis:**

*"Fundamentals of Simulation Methods"*, tutorial accompanying the lecture course, Heidelberg University, winter semester 2018 / 2019.

**Volker Springel:**

Study group *"Experimental Physics II: Electrodynamics and Optics"*, Department of Physics and Astronomy, Heidelberg University, April 2018 - July 2018.

**Alexandros Stamatakis, Benoit Morel, Alexey Kozlov, Pierre Barbera:**

Lecture *"Introduction to Bioinformatics for Computer Scientists"*, computer science Master's program at Karlsruhe Institute of Technology, winter semester 2017 / 2018.

**Alexandros Stamatakis, Benoit Morel, Sarah Lutteropp, Pierre Barbera, Alexey Kozlov:**

Seminar *"Hot Topics in Bioinformatics"*, computer science Master's program at Karlsruhe Institute of Technology, summer semester 2018.

**Alexandros Stamatakis:**

Practical *"Hands on Bioinformatics Practical"*, computer science Master's program at Karlsruhe Institute of Technology, summer semester 2018.

**Alexandros Stamatakis, Pierre Barbera:**

Summer school *"Computational Molecular Evolution"*, Heraklion, Crete, Greece, May 2018.

**Michael Strube:**

PhD Colloquium, Department of Computational Linguistics, Heidelberg University, winter semester 2017 / 2018. Seminar: *"NLP und Journalismus / Journalismus und NLP"*, Department of Computational Linguistics, Heidelberg University, winter semester 2017 / 2018. PhD Colloquium, Department of Computational Linguistics, Heidelberg University, summer semester 2018.

**Rebecca Wade:**

Module 4, *"Biomolecular Recognition: Modeling and Simulation"*, M.Sc. Molecular Cell Biology, Heidelberg University, March 19, 2018. Module 3, *"Protein Modeling"*, M.Sc. Molecular Cell Biology, Heidelberg University, April 27, May 7, 2018. Ringvorlesung *"Computational Biochemistry"*, *"Electrostatics and Solvation for Biomolecules"*, M. Sc. Biochemistry, Heidelberg University, November 7, 2018. Ringvorlesung *„Biophysik"*, *"Receptor-Ligand Interactions: Structure and Dynamics"*, B.Sc. Molecular Biotechnology, Heidelberg University, November 29, 2018.

**Rebecca Wade, Neil Bruce, Anna Feldman-Salit, Patrick Friedrich, Gaurav Ganotra, Daria Kokh, Goutam Mukherjee, Prajwal Nandekar, Ariane Nunes-Alves, Ina Pöhner, Stefan Richter, Jui-Hung Yuan, Frauke Gräter, Isabel Martin:**

B.Sc. Biosciences practical course *„Grundkurs Bioinformatik"*, Heidelberg University, January 29 – February 2, 2018.

**Rebecca Wade, Christina Athanasiou, Alexandros Tsengenes:**

*"Computer Assisted Drug Design"* Hands-on Course, EuroNeurotrophin ITN Project, 1st Doctoral Training Week, National Hellenic Research Foundation (NHRF), Athens, Greece, December 3 – 6, 2018.

**Anna Wienhard, Peter Albers, Mareike Pfeil:**

Seminar *"The symplectic structure of representation varieties"*, Heidelberg University, winter semester 2018 / 2019.

**Anna Wienhard, Andreas Ott:**

Hauptseminar *"Geometrie"*, Heidelberg University, summer semester 2018, winter semester 2018 / 2019.

**Anna Wienhard, Maria Beatrice Pozzetti, Jonas Beyrer:**

*"Junior Geometry Seminar"*, Heidelberg University, winter semester 2018 / 2019.

## 9.1 Guest Speaker Activities

**Daniele Alessandrini:**

*"Geometric Structures with Quasi-Hitchin Holonomy"*. Bers Seminar, CUNY Graduate Center, New York, USA, February 9, 2018. *"The relative PSL(2,R)-character varieties"*. Workshop on relative character varieties and parabolic Higgs bundles, Indio, California, USA, February 24 – March 3, 2018. *"Classification of real and complex projective structures with special holonomy"*. Conference: Representations of surface groups and beyond, Université de Lille, France, September 24 – 26, 2018; Seminar at IRMA, Université de Strasbourg, France, October 26, 2018. *"Domains of Discontinuity for (Quasi-)Hitchin representations"*. Seminar at University of Maryland, USA, April 9, 2018; Seminar at IHÉS, Paris, France, April 23, 2018.

**Andreas Bauswein:**

*"Neutron star mergers and the high-density equation of state."* Talk at EMMI NQM Seminar, GSI, Darmstadt, Germany, January 31, 2018. Talk at 3rd CBM-China workshop, Yichang, China, April 17, 2018. Colloquium talk at Koenigstuhl Colloquium, Heidelberg Germany, April 24, 2018. *"Neutron-star radius constraints from GW170817 and properties of GW signals from the postmerger phase"* talk at Neutron stars in Lisbon – The multi-messenger Universe after GW170817, Lisbon, Portugal, April 13, 2018. *"Postmerger physics: nature of the central object – kilonova ejecta"* talk at Third scientific workshop of the SVOM mission: disentangling the merging universe with SVOM, Les Houches, France, May 15, 2018. *"Collapse behavior and postmerger gravitational wave emission of neutron star mergers"*, seminar talk at Columbia University, New York, USA, May 29, 2018. *"EOS constraints from the collapse behavior and postmerger phase of neutron star mergers"* talk at The messenger roadmap: Nuclear astrophysics in the era of multi-messenger astronomy, New York, USA, May 30, 2018; talk at 12th Bonn workshop on the formation and evolution of neutron stars, Bonn, Germany, May 12, 2018; talk at EMMI Rapid Reaction Task Force: the physics of neutron star mergers at GSI/FAIR, Darmstadt, Germany, June 6, 2018. *"Postmerger gravitational wave emission"* talk at Path to Kilohertz Gravitational-Wave Astronomy, Perimeter Institute, Waterloo, Canada, June 11, 2018.

**Csaba Daday:**

*"Proteins under pressure: crash-testing cellular junctions in supercomputers"*, 5th Baden-Württemberg High Performance Computing Symposium, Freiburg, Germany, September 24, 2018. *"How fast is too fast in pulling simulations?"* Seminar talk, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, December 12, 2018.

**Antonio D'Isanto:**

*"The two worlds of photometric redshift estimation: feature-based and fully automatic models."* ITA "blackboard" Colloquium Heidelberg University, January 22, 2018. *"Convolutional neural networks: Theory and application in astronomy."* Lecture for the Astroinformatics course at the University of Naples Federico II, Physics Department, December 19, 2018.

**Valentina Disarlo:**

*"Generalized stretch lines for Thurston's distance"*. Conference: New trends in Teichmüller theory, Oberwolfach, Germany, September 2 – 8, 2018; Geometry of Teichmüller space and mapping class groups, University of Warwick, UK, April 9 – 13, 2018; Dynamics, Geometry and Groups Seminar, Queens University, Kingston ON, Canada, November 2018. *"Cubical Geometry in the polygonalization complex"*. Geometry and Topology Seminar, CUNY, New York City NY, USA, November 2018; Geometric Group Theory Seminar, McGill University, Montreal QC, Canada, November 2018. *"On the geometry of the flip graph"*. Postdoc Seminar, University of Toronto, Toronto ON, Canada, October 2018.

**Nikos Gianniotis:**

*"Linear Dimensionality Reduction for Time Series Visualisation"*. Seminar in Artificial Intelligence and Natural Computation, School of Computer Science, the University of Birmingham, UK, June 25, 2018.

**Tilmann Gneiting:**

*"Interpretation of point forecasts with unknown directive"*. Nelder Lecture, Imperial College, London, UK, June 11, 2018.

**Martin Golebiewski:**

*"How to Share Your Data FAIR –Integrated Data Management for Systems Medicine"*, Workshop *"Multi-Omics Data Analysis"*, 63. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Osnabrück, Germany, September 2 – 6, 2018.

**Frauke Gräter:**

*"Mechano-sensing through and at membranes"*, CECAM workshop, Lugano, Switzerland, January 10 – 12, 2018; International symposium on "Functionality controlled by organization in and between membranes", University of Göttingen, Germany, September 2018. *"Molecular high-five: fuzzy complex formation of IDPs with their receptors"*, Biophysical Society meeting, San Francisco, California, USA, February 17 – 21, 2018. *"Mechano-sensing at cell-cell junctions"*, colloquium talk at King's College, London, UK, March, 2018. *"Multi-scale simulations of silk mechanics"*, ESMC, Bologna, Italy, June, 2018. *"Multi-scale simulations for biomaterials"*, International Conference for Supercomputing, Frankfurt, Germany, June, 2018. *"Why playing soccer hurts: mechanoradicals in collagen"*, Symposium of the Volkswagen Foundation, Herrenhäuser Gärten, Hannover, Germany, November, 2018.

**Vincent Heuveline:**

*"Digitalisierung in der Wissenschaft: Status und Herausforderungen in der Metropolregion Rhein-Neckar"*, DHBW Mannheim, March 15, 2018. *"Large Scale Data Facility for Life Sciences"*, HeKKSaGOn Multidisciplinary Joint Workshop toward Fusions between Data and Mathematical Sciences, Osaka, Japan, April 11 – 14, 2018. *"Blockchain unchained"*, Opening event Innohub, Heidelberg, May 7, 2018. *"High performance computing for heart simulation: from cognition to uncertainty quantification"*, DZHK Retreat, Bad Schönborn, Germany, May 25 – 26, 2018. *"Facing the digital revolution: Digitalization as a driving force for innovation in research and industry"*, Freudenberg IT Meeting, Weinheim, Germany, September 17, 2018. *"Digitalisierung als Chance und Herausforderung: Was Wissenschaft und Industrie voneinander lernen können"*, MVV Enamic GmbH, Mannheim, Germany, September 27, 2018. *"The intersection of hybrid cloud, large scale data facility and AI: the benefits for universities"*, EUNIS IT Leaders Retreat, Barcelona/Tarragon, Spain, October 3, 2018. *"Digitalisierung für die Bildung: Chancen und Herausforderungen aus der Perspektive einer Forschungsuniversität"*, Advisory council meeting of Dieter von Holtzbrinck Stiftung, Stuttgart, Germany, October 11, 2018. Panel discussion on *"IT for Life Sciences"*, Zhongguancun Software Park Beijing, China, November 1, 2018.

**Johannes Horn:**

*"Limiting configurations over the singular locus of the Hitchin map"*. Oberseminar Geometrie at Karl-Albrechts-Universität zu Kiel, Germany, November 29, 2018.

**Daria Kokh:**

*"In silico Prediction of Relative Drug-Protein Residence Times using the RAMD approach"*, Boehringer-Ingelheim, Biberach, Germany, February 5, 2018. *"Prediction of Relative Drug-Protein Residence Times"*, 22th European Symposium on Quantitative Structure-Activity Relationships (EuroQSAR), Thessaloniki, Greece, September 16 – 20, 2018.

**Alexey Kozlov:**

*"Maximum-likelihood inference of cell phylogenies"*, Workshop on Single Cell Data Science: Making Sense of Data from Billions of Single Cells, Lorentz Center, Leiden, Netherlands, June 2018.

**Jonas Kratzke:**

*"Uncertainty quantification for the simulation of the human heart"*, Workshop at the Philips Research Laboratories, Hamburg, Germany, June 25, 2018.

**Markus Kromer:**

*"Progenitor models: explosions"*, talk at Observational Signatures of Type Ia Supernova Progenitors III, International Workshop, Lorentz Center, Leiden, Netherlands, February 5 – 9, 2018. *"Modelling SN Ia observables with ARTIS"*, talk at Radiation Transfer and Explosive Thermonuclear Burning in Supernovae, International Workshop, Weizmann Institute of Science, Rehovot, Israel, June 17 – 28, 2018.

**Gye-Seon Lee:**

*"Convex real projective Dehn filling"*. Topology Seminar, University of Texas at Austin, USA, February 26, 2018; Topology Seminar, KAIST, South Korea, May 18, 2018; Séminarie Géométries et Topologie, Institut de Mathématiques de Jussieu-Paris Rive Gauche, France, May 30, 2018; Workshop Hyperbolic 3-manifolds and related topics, POSTECH, South Korea, June 18, 2018; Séminaire HORUS, Université de Strasbourg, France, October 26, 2018. *"Exotic quasi-Fuchsian groups"*. Seminar in Mathematics, KIAS, South Korea, May 18, 2018; Conference: Representations of surface groups and beyond, Université de Lille, France, September 24 – 26, 2018; Geometry, Topology & Dynamics Seminar, Seoul National University, South Korea, October 17, 2018.

# 9 Miscellaneous

**Sebastian Lerch:**

*"Probabilistic weather prediction: From ensembles to neural networks"*, Meetup, Karlsruhe.ai, Karlsruhe, Germany, January 31, 2018.

**Philipp Lösel:**

*"The NOVA project: maximizing beam time efficiency through synergistic analyses of SRμCT data"*, Helmholtz-Zentrum Geesthacht GEMS Outstation: Materials Research and High Resolution Imaging, DESY Photon Science Users' Meeting, Hamburg, January 25, 2018, on invitation by Dr. Christina Krywka. *"Fast synchrotron X-ray imaging and semi-automated analysis of extant and fossil insects"*. BASF, Ludwigshafen, April 19, 2018, on invitation by Dr. Paul Birnbrich.

**Wolfgang Müller:**

*"FAIR Data Management"*, Charme Meeting on Teaching, Ljubljana, February 6, 2018. *"The FAIRDOM Project"*, RDA Workshop on Virtual Research Environment, Pisa, February 27 – 28, 2018.

**Andreas Ott:**

*"Higgs bundles and pleated surfaces"*. 9th Northern German Differential Geometry Day, June 8, 2018; Séminaire GT3 at University of Strasbourg, May 14, 2018.

**Kai Polsterer:**

*"Künstliche Intelligenz: Wie KI Astronomen bei der Arbeit hilft."*, *"Science Notes"* Event, Heidelberg, Germany, February 1, 2018; Planetarium Bochum, Germany, March 21, 2018. *"Machine Learning in Astronomy: Lessons learned from learning machines."* Colloquium, Kapteyn Astronomical Institute, Groningen, The Netherlands, February 19, 2018; Colloquium CADC, Victoria, Canada, November 5, 2018. *"Beginners Guide to machine learning in Astronomy"*. ADASS 2018, College Park, Maryland, USA, November 12, 2018.

**Maria Beatrice Pozzetti:**

*"Bounded cohomology of acylindrically hyperbolic groups"*. Conference Topological and Homological Methods in Group Theory, Bielefeld, Germany, April 5, 2018. *"Maximal representations"*. Conference New Methods for Zimmer's Conjecture, IPAM, Los Angeles, USA, January 25, 2018; RTG Colloquium Salt Lake City, USA, February 22, 2018; Berkeley, March 14, 2018. *"Critical exponent and Hausdorff dimension for Anosov representations"*. Heidelberg-Karlsruhe-Strasbourg Geometry Day XI, Strasbourg, France, June 22, 2018; Conference: Representation varieties and geometric structures in low dimensions, Warwick, UK, July 4, 2018; Bonn, Germany, November 8, 2018. *"Ultralimits of maximal representations"*. Santa Barbara, USA, February 13, 2018.

**Friedrich Röpke:**

*"Simulations of thermonuclear explosions in white dwarf stars"*, talk at the WE Heraeus Seminar 660, Bad Honnef, Germany, January 22, 2018. *"HWB07: Towards multidimensional hydrodynamic simulations of stars"*, talk at NIC Symposium Jülich, Germany, February 23, 2018. *"Typ Ia Supernovae – wie explodierende Sterne unser kosmologisches Weltbild erschütterten"*, public talk at Planetarium Mannheim, Mannheim, Germany, March 7, 2018. *"Simulations of Type Ia supernova explosions"*, talk at Theoretical Astrophysics Seminar, University of Zurich, Zurich, Switzerland, March 16, 2018. *"Typ Ia Supernovae – Was uns Sternexplosionen über das Universum verraten"*, public talk for the Olbers-Gesellschaft Bremen, Germany, March 20, 2018; public talk for Sternfreunde Nordenham, Nordenham, Germany March 21, 2018. *"Models and simulations for Type Ia supernovae"*, Colloquium, Leibniz-Institut für Astrophysik Potsdam, Potsdam, Germany, May 24, 2018. *"Modeling thermonuclear combustion in supernovae"*, talk at Radiation Transfer and Explosive Thermonuclear Burning in Supernovae, International Workshop, Weizmann Institute of Science, Rehovot, Israel, June 21, 2018. *"Adventures in stellar hydrodynamics"*, talk at the 2nd ISSI Meeting *"Towards a New Generation of Massive Star Models"*, ISSI Bern, Switzerland, July 18, 2018. SOC review talk on the *"Supernovae and galactic chemical evolution"* session at Chemical evolution and nucleosynthesis across the Galaxy, Heidelberg, Germany, November 29, 2018.

**Kashif Sadiq:**

*"Towards multiscale spatiotemporal modeling of sub-cellular scale assembly"*, Department of Virology, Heidelberg University, January 30, 2018. *"Elucidating the role of the HIV-1 gp41 cytoplasmic tail from multiscale modeling"*, gp41 Cytoplasmic Tail Structure and Function Workshop, NIH – National Cancer Institute at Frederick, USA, April 26 – 27, 2018. *"Towards a computational molecular dynamics toolbox for structural biology"*, Blue Seminar, EMBL, Heidelberg, October 11, 2018. *"Modelling HIV-1 maturation: A journey through enzyme catalysis, protein assembly and RNA condensation"*, MRC Laboratory for Molecular Cell Biology, University College London, London, UK, October 29, 2018.

**Fabian Schneider:**

*"Weighing Stars"*, Public Lecture at Rijksmuseum Boerhaave as part of the Lorentz Center Workshop "Weighing Stars from Birth to Death", Leiden, The Netherlands, November 22, 2018.

**Chen Song:**

*"Unterstützung für schwache Herzen: Das zuverlässige Kunstherz aus dem Hochleistungsrechner"*, Doctoral Conferment Ceremony, Heidelberg University, Germany, December 2018.

**Alexandros Stamatakis:**

*"Introduction to Phylogenetics ... in 12 Minutes"*, Workshop on Single Cell Data Science: Making Sense of Data from Billions of Single Cells, Lorentz Center, Leiden, Netherlands, June 2018. *"Research Activities in the Exelixis Lab"*, Roche, Basel, Switzerland, November 2018.

**Volker Springel:**

*"Supercomputer simulations of the emergence of cosmic structures"*, University of Amsterdam, Netherlands, February 2018; ITC Colloquium, Harvard Center for Astrophysics, Cambridge, USA, March 2018; Physics Colloquium, Instituto de Fisica Teorica, Madrid, Spain, June 2018. *"Simulierte Universen: Ursprung und Schicksal unserer Milchstraße"*, Planetarium Mannheim, February 2018; Wissenschaftlicher Verein Mönchengladbach, April 2018; *"Multi-scale, multi-physics structure formation simulations"*, Stars, Planets, and Galaxies Conference, Harnack House Berlin, April 2018. *"Towards more scalable simulation codes"*, Ringberg Workshop on Computational Galaxy Formation, Castle Ringberg, Tegernsee, March 2018. *"Leuchtende und dunkle Strukturen des Kosmos"*, DLR Astroseminar 2018, Konferenzzentrum der Luftwaffe, Köln-Wahn, April 2018; Karl-Rahner Akademie Köln, April 2018. *"Simulating the galaxy population of Dark Energy Universes"*, The Dark Universe Workshop of Transregio TR33, Heidelberg, May 2018. *"Multi-scale, multi-physics structure formation simulations"*, Multi-scale physics of star formation and feedback during galaxy formation, Heidelberg, June 2018. *"Supercomputer insights into the messy physics of galaxy formation"*, Physics Colloquium, Heidelberg University, July 2018.

**Michael Strube:**

*"Data Mining und die Mensch-Maschine-Schnittstelle: Ein Blick in die Zukunft des Journalismus mit Daten"*, 8. Medientage der Deutschen Journalistinnen und Journalisten Union, Berlin, Germany, June 22, 2018. *"The Dark Side of NLP: Chances and Risks of Natural Language Processing"*, Symposium: 50 Jahre German Chapter of the ACM: Mensch-Sein mit Algorithmen, Heidelberg, Germany, September 21, 2018.

**Rebecca Wade:**

*"New Approaches for Computing Ligand-Receptor Binding Kinetics"*, Boehringer-Ingelheim, Biberach, Germany, February 5, 2018. *"Computational Mapping of Protein Binding Sites for Drug Discovery"*, DFG graduate college 1657 colloquium, Technical University Darmstadt, Germany, March 15, 2018. *"Computationally efficient approaches to estimate drug-target binding kinetic parameters"*, 2018 Workshop on Free Energy Methods, Kinetics and MSMs in Drug Design, Novartis Institute for Biomedical Research, Boston, USA, May 14 – 18, 2018. *"Towards simulation tools for studying macromolecular binding in crowded and confined environments"* Workshop on "Challenges in Large-Scale Biomolecular Simulations", Telluride, USA, June 16 – 21, 2018. *"Computational approaches to Protein Target Dynamics for Drug Discovery"*, Modeling and Design of Molecular Materials 2018 Conference, Polanica Zdrój, Poland, June 24 – 28, 2018. *"Insights into the Membrane and Protein Interactions of Cytochrome P450 Enzymes from Molecular Simulations"*, CECAM Workshop on "Frontiers and challenges of computing metals for biochemical, medical and technological applications 2018", ENSCP Chimie Paris Tech, Paris, France, July 11 – 13, 2018. *"Towards computationally efficient approaches to estimate drug-target binding kinetic parameters"*, 6th Annual CCPBioSim Meeting: Molecular Simulations in Drug Discovery and Development, Lady Margaret Hall, Oxford University, UK, September 5 – 8, 2018. *"A multi-resolution approach to the simulation of protein complexes in a membrane bilayer"*, 21st HLRS Results and Review Workshop on "High Performance Computing in Science and Engineering 2018", HLRS High Performance Computing Center, Stuttgart, Germany, October 4 – 5, 2018. *"Computational Approaches to Protein Dynamics and Binding Kinetics for Drug Discovery"*, Human Brain Project: Annual Meeting of the Co-Design Project 6 on "Modeling for Drug Discovery", Forschungszentrum Jülich, Germany, October 30 – 31, 2018. *"Computational Approaches to Protein Dynamics and Binding Kinetics for Drug Discovery"*, Department of Chemistry Colloquium, ETH, Zurich, Switzerland, December 11, 2018.

**Anna Wienhard:**

*"Symmetrien, Pflasterungen und andere Puzzles"*, Berlin-Brandenburgische Akademie der Wissenschaften, February 23, 2018. *"An invitation to higher Teichmüller theory"*, International Congress of Mathematicians 2018, Geometry Section, Rio de Janeiro, Brazil, August 1 – 9, 2018. *"Around higher Teichmüller theory"*, Workshop on Flat Surfaces and Algebraic Curves, Oberwolfach, Germany, September 18, 2018. *"Dynamics on the Hitchin component"*, Fields Medal Symposium in honor of Maryam Mirzakhani, Fields Institute, Toronto, Canada, November 5 – 9, 2018.

## 9.2 Presentations

### Demos

**Neil Bruce:**
*"webSDA: Prediction of the Structure of a Protein-Protein Complex and the Rate Constant of its Formation"*, Fourth Biological Diffusion and Brownian Dynamics Brainstorm Meeting (BDBDB4), Studio Villa Bosch, Heidelberg, Germany, April 16 – 18, 2018.

**Martin Golebiewski, Andreas Weidemann, Ulrike Wittig:**
*"SABIO-RK – Reaction Kinetics Database"*, COMBINE & de.NBI Tutorial *"Modelling and Simulation Tools in Systems Biology"*, International Conference on Systems Biology (ICSB 2018), Lyon, France, October 27, 2018.

**Maja Rey and Ulrike Wittig:**
FAIRDOM Workshop *"FAIR Data Management in Life-Sciences"*, International Conference on Systems Biology (ICSB 2018), Lyon, France, October 27, 2018.

### Talks

**Pierre Barbera:**
*"Analysis of microbial environments using phylogenetic placement"*, Station Biologique de Roscoff, France, April 2018.

**Andreas Bauswein:**
*"Neutron star mergers and the high-density equation of state"*, talk at EMMI NQM Seminar, GSI, Darmstadt, Germany, January 31, 2018; talk at 3rd CBM-China workshop, Yichang, China, April 17, 2018.

**Jonas Beyrer:**
*"Marked length spectrum rigidity of cubulations"*, Geometrie Seminar at Heidelberg University, November 6, 2018.

**Neil Bruce:**
*"Efficient simulation of protein diffusion in crowded and confined solutions"*, Fourth Biological Diffusion and Brownian Dynamics Brainstorm Meeting (BDBDB4), Studio Villa Bosch, Heidelberg, Germany, April 16 – 18, 2018. *"Potential Applications of Brownian Dynamics Simulations in Drug Design"*, Human Brain Project: Annual Meeting of the Co-Design Project 6 on *"Modeling for Drug Discovery"*, Forschungszentrum Jülich, Germany, October 30 – 31, 2018.

**Lucas Czech:**
*"Revolution der Evolution – Treffen sich ein Biologe und ein Informatiker"*, HITS open house event, Heidelberg, Germany, July 7, 2018.

**Csaba Daday:**
*"Reconciling AFM and MD: unfolding focal adhesion kinase"*, Biennial Meeting of the German Biophysical Society, Düsseldorf, Germany, September 17, 2018.

**Antonio D'Isanto:**
*"Return of the features"*, talk at the Astroinformatics Conference, Heidelberg, Germany, September 3 – 7, 2018.

**Valentina Disarlo:**
*"Cubical geometry in the polygonalisation complex"*, Conference: Nonpositively Curved Groups on the Mediterranean, Nahsholim, Israel, May 23 – 29, 2018. *"Thurston's distance for Teichmüller space"*, European Women in Mathematics: German Chapter Conference, Heidelberg, Germany, May 3 – 4, 2018.

**Krisztina Fehér:**
*"Atomistic simulations of immune stimulatory single stranded bacterial DNA"*, The 6th Annual CCPBioSim Meeting: Molecular Simulations in Drug Discovery and Development, Oxford, UK, September 5 – 7. 2018; Chemistry towards Biology (CTB9), Budapest, Hungary, September 25 – 27, 2018.

**Florian Franz:**
*"Punching membranes: How lipid bilayers withstand and propagate mechanical load"*, 62th Annual Biophysical Society Meeting. San Francisco, California, USA, February 17 – 21, 2018. *"The mechanical stability of lipid bilayers under force"*, Summer School: Computer simulations of biological membranes free energy calculations of biomolecular systems, Universidad de los Andes, Bogotá, Colombia, August 2018.

**Gaurav Ganotra:**
*"Application of COMparative BINding Energy (COMBINE) Analysis to prediction of drug-target binding kinetics"*, GCC 2018, 14th German Conference on Chemoinformatics, Mainz, Germany, November 11 – 13, 2018.

**Philipp Gerstner:**

*"Finite Element Simulation of Thermal Electro-Hydrodynamic Driven Flow in Annular Geometry"*, 20th International Couette Taylor Workshop, Marseille, France, July 11 – 13, 2018. *"Finite Element Approximation for Thermal Electro Hydrodynamical Boussinesq Equations"*, 13th World Congress on Computational Mechanics, New York City, USA, July 22 – 27, 2018.

**Nikos Gianniotis:**

*"Probabilistic Principal Component Analysis for Time Series"*, talk at the Astroinformatics Conference, Heidelberg, Germany, September 3 – 7, 2018.

**Martin Golebiewski:**

*"FAIRDOM on the last mile: data & model management and enrichment for the long tail"*, Elixir All Hands Meeting 2018, Berlin, Germany, June 4 – 7, 2018. *"Share FAIR – Data management and standards for Systems Medicine"*, 3rd Disease Maps Community Meeting (DMCM 2018), Institut Curie, Paris, France, June 21 – 22, 2018. *"Data Security and Privacy, Interoperability and Standardization for the Health EU FET Flagship"*, Health EU Flagship Meeting, Lausanne, Switzerland, June 28, 2018. *"FAIR data exchange in the life sciences by standardization of heterogenous data and multicellular models"*, COMBINE 2018, Boston, MA, USA, October 8 – 12, 2018. *"COMBINE and Its Standards"*, COMBINE 2018, Boston, MA, USA, October 8 – 12, 2018. *"Standardising Data and Models in Biotechnology and the Life Sciences"*, 1st EnzymeML workshop, Stuttgart, Germany, November 19 – 20, 2018. *"How to Share Your Data FAIR"*, LiSyM network retreat, Hünfeld, Germany, 26 – 28 November 2018.

**Martin Golebiewski, Wolfgang Müller:**

*"LiSyM Data Management"*, LiSyM Mid-Term Evaluation Meeting, Mannheim, Germany, May 2 – 3, 2018.

**Sabrina Gronow:**

*"Double detonations in sub-Chandrasekhar mass models"*, talk at the Supernova Workshop, Garching, June 18, 2018. *"Thermonuclear explosions of sub-Chandrasekhar mass white dwarfs"*, talk at the 21st European White Dwarf Workshop, Austin TX, USA, July 24, 2018. *"A possible progenitor of a supernova Ia: an explosion of a sub-Chandrasekhar mass white dwarf"*, talk at the International Conference of Physics Students, Helsinki (Finland), August 12, 2018. *"Double detonations of sub-Chandrasekhar mass white dwarfs"*, talk at Chemical Evolution and Nucleosynthesis Across the Galaxy, Heidelberg, November 28, 2018.

**Vincent Heuveline, Saskia Haupt:**

*"High Performance Computing for Uncertainty Quantification: Challenges and Perspectives for Flow Problems"*. SIAM Conference on Uncertainty Quantification (UQ18), Los Angeles, USA, 1 April 16 – 19, 2018.

**Erica Hopkins:**

*"Semi-automated morphological classification of radio galaxies in the FIRST survey"*, talk at the seminar at the Gravitational Lensing Group, Astronomisches Rechen-Institut (ARI), Heidelberg University, Germany, December 18, 2018.

**Ana Herrera-Rodriguez:**

*"Towards spider silk proteins self-assembly under uniform flow"*, Workshop on computer simulation and theory of macromolecules. Hünfeld, Germany, April 20 – 21. *"Multiscale simulation of directed spider dragline silk self-assembly by flow"*, ESMC 2018, 10th European Solid Mechanics Conference. Bologna, Italy, July 2 – 6.

**Leonhard Horst:**

*"3D Simulation of Dynamical Shear Instabilities in a Rotating Massive Star"*, talk at the 2nd ISSI Meeting *"Towards a New Generation of Massive Star Models"*, ISSI Bern, Switzerland, July 20, 2018.

**Fan Jin:**

*"How does phosphorylation affect intrinsically disordered proteins?"* Fourth Biological Diffusion and Brownian Dynamics Brainstorm Meeting (BDBDB4) Heidelberg, Germany, April 16 – 18, 2018.

**Daria Kokh:**

*"Prediction of Relative Drug-Protein Residence Times"*, Human Brain Project: Annual Meeting of the Co-Design Project 6 on *"Modeling for Drug Discovery"*, Forschungszentrum Jülich, Germany, October 30 – 31, 2018.

# 9 Miscellaneous

**Jonas Kratzke:**

*"Fluid-Structure Interaction with Uncertainty in Medical Engineering"*, SIAM Conference on Uncertainty Quantification (UQ18), Los Angeles, USA, April 16 – 19, 2018. *"A probability measure for aortic vessel wall overload based on fluid-structure interaction simulations"*, 6th European Conference on Computational Mechanics (ECCM 6) & 7th European Conference on Computational Fluid Dynamics (ECFD 7), Glasgow, United Kingdom, June 13, 2018.

**Olga Krebs:**

*"FAIRDOMHub: a repository and collaboration environment for sharing systems biology research"*, International Conference BGRS\SB-2018 – Bioinformatics of Genome Regulation and Structure \ Systems Biology, Novosibirsk, Russia, August 20 – 24, 2018. *"FAIR (Findable, Accessible, Interoperable, and Reusable) research data principles explained"*, The 10th Young Scientists School *"Systems biology and Bioinformatics SBB-2018"*, Novosibirsk, Russia, August 27 – 31, 2018. *"Scientific Data Management using FAIRDOMHub"*, European COST Action CHARME MC Meeting & Workshop , Valletta, Malta, October 1 – 3, 2018. *"FAIRDOMHub: a repository and collaboration environment for sharing systems biology research"*, COMBINE 2018, Boston, MA, USA, October 8 – 12, 2018.

**Florian Lach:**

*"Chandrasekhar Mass Deflagrations as a Model for Type Iax Supernovae"*, poster talk at the WE Heraeus Seminar 660 Supernovae – From Simulations to Observations and Nucleosynthetic Fingerprints, Bad Honnef, Germany, January 22, 2018. *"Chandrasekhar Mass Explosions: Pure Deflagrations as a Model for Type Iax Supernovae"*, talk at the Supernova Workshop Garching, Germany, June 18, 2018.

**Sebastian Lerch:**

*"Similarity-based semilocal estimation of post-processing models"*, German Probability and Statistics Days, Freiburg, Germany, February 27 – March 2, 2018. *"Forecaster's dilemma: Extreme events and forecast evaluation"*, International Conference on Computational and Methodological Statistics, Pisa, Italy, December 14 – 16, 2018.

**Philipp Lösel:**

*"Biomedisa: Fast and Accurate Segmentation of Fossil Insects from Synchrotron X-ray Microtomography Images"*, 13th World Congress in Computational Mechanics, New York City, USA, July 22 – 27, 2018.

**Benoit Morel:**

*"ParGenes: An integrated tool for model selection and maximum likelihood (ML) based phylogenetic inference on thousands of independent MSAs on clusters and supercomputers"*, Phylogenomics Symposium, Montpellier, France, October 2018.

**Ghulam Mustafa:**

*"Influence of mutations of the transmembrane-helix of CYP17A1 on catalytic domain-membrane interactions and function"*, 32nd Molecular Modelling Workshop (MMWS), Erlangen, Germany, March 12 – 14, 2018.

**Sotirios Nikas:**

Development and implementation of a temperature and energy monitoring system for HPC systems, 5. Seminar Energieeffizienz, Universitätsrechenzentrum Heidelberg, October 31, 2018.

**Ariane Nunes-Alves:**

*"Effects of macromolecular crowding on the diffusion rates of enzyme substrates and drug-like molecules"*, 22nd European Symposium on Quantitative Structure-Activity Relationships (EuroQSAR), Thessaloniki, Greece, September 16 – 20, 2018.

**Agnieszka Obarska-Kosinska:**

*"How collagen fibrils dynamically distribute and measure stress"*, 62nd Biophysical Society Annual Meeting, San Francisco, California, USA, February 17 – 21, 2018.

**Kai Polsterer:**

*"Machine Learning in Astronomy: Lessons learned from learning machines"*, introductory talk, Astroinformatics Conference, Heidelberg, Germany, September 3 – 7, 2018.

**Benedikt Rennekamp:**

*"Reactive Kinetic Monte Carlo / Molecular Dynamics (rKMC/MD) simulations – a hybrid scheme applied to tensed collagen"*, Seminar, Max Planck Institute for Biophysical Chemistry, Department of Theoretical and Computational Biophysics, Göttingen, Germany, December 5, 2018.

**Stefan Richter:**

*"Dicht gepackt – Proteine in der Zelle"*, Akademische Mittagspause, Fakultät für Biowissenschaften, Peterskirche, Heidelberg, Germany, June 29, 2018.

**Kashif Sadiq:**

*"Modelling protease-triggered HIV-1 infectivity using interaction particle-based reaction-diffusion (iPRD) simulations"*, BDBDB4 Biological Diffusion and Brownian Dynamics Brainstorm 4, Studio Villa Bosch, Heidelberg, Germany, April 16 – 18, 2018. *"Modeling Reaction-Triggered Oligomerization of Envelope Membrane Glycoproteins During HIV-1 Maturation"*, Hünfeld 2018 Workshop on Computer Simulation and Theory Of Macromolecules, Hünfeld, Germany, April 20 – 21, 2018.

**Fabian Schneider:**

*"Turbulent lives of massive stars"*, Seminar talk, ZAH/ARI, Heidelberg University, Germany, October 18, 2018.

**Chen Song:**

Uncertainty quantification for the reliable simulation of a blood pump device, SIAM Conference on Uncertainty Quantification (UQ18), Los Angeles, USA, April 16 – 19, 2018.

**Theodoros Soultanis:**

*"Open source codes – Einstein Toolkit"*, talk at 3rd HEL.A.S. summer school and DAAD school, Neutron stars and gravitational waves, Astronomical Observatory of AUTh, Thessaloniki, Greece, October 9, 2018. *"Threshold mass for prompt merger collapse"*, talk at 13th Würzburg Workshop in Heidelberg, Studio Villa Bosch, Heidelberg, Germany, December 18, 2018.

**Andreas Weidemann:**

*"SABIO-RK: a resource for manually curated biochemical reaction kinetics"*, 11th International Biocuration Conference, Shanghai, China, April 8 – 11, 2018. *"SABIO-RK"*, de.NBI Tutorial: Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, SEEK, Magdeburg, Germany, April 24 – 26, 2018.

**Ulrike Wittig:**

*"SABIO-RK – Finding, Curating, and Publishing Data"*, GFBio – de.NBI Summer School 2018 *"Riding the Data Life Cycle"*, Braunschweig, Germany, September 3 – 7, 2018. *"SABIO-RK – Data model and relation to EnzymeML"*, 1st EnzymeML workshop, Stuttgart, Germany, November 19 – 20, 2018.

**Christopher Zapp:**

*"Mechanosensation through Radicals in Tensed Collagen"*. 4th Bio-inspired Materials. Potsdam, Germany, March 19 – 22, 2018.

## Posters

**Neil Bruce, Daria Kokh, Martin Reinhardt, Rebecca Wade, Nuria Cirauqui Diaz, Elisa Frezza, Guillaume Launay, Richard Lavery, Juliette Martin, Sanja Zivanovic, Adam Hospital, Genís Bayarri, Francesco Colizzi, Robert Soliva, J. Lluís Gelpí, Modesto Orozco, Riccardo Capelli, Paolo Carloni and Michele Parrinello:**

*"Modelling and simulation in SP6: Recent advances in methods development at the molecular level."*, HBP Annual Summit, Maastricht, Netherlands, October 16 – 18, 2018.

**Neil Bruce, Martin Reinhardt, Rebecca Wade:**

*"Modelling Protein Diffusion Through Brownian Dynamics Simulations"*, Hünfeld 2018 Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, April 20 – 21, 2018. CECAM Workshop: Proteins in Realistic Environments: Simulation Meets Experiment, Stuttgart, Germany, May 23 – 25, 2018.

**Csaba Daday:**

*"When an Enzyme Self-Assembles on a Membrane: Focal Adhesion Kinase"*, Annual Meeting of the Biophysical Society, San Francisco, USA, February 19, 2018. *"CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps"*, Annual Meeting of the Biophysical Society, San Francisco, USA, February 19, 2018. *"Explaining cardiomiopathy-linked disease mutations in desmoplakin through simulations"*, International Meeting of SPP 1782, Halle an der Saale, Germany, July 4 – 6 2018.

**Olivia Eriksson, Alexandra Jauhiainnen, Daniel Trpevski, Andre Kramer, Parul Tewatia, Joao Santos, Neil Bruce, Rebecca Wade, Ursula Roethlisberger, Siri Van Keulen, Daniele Narzi, Paolo Carloni and Jeanette Hellgren Kotaleski:**

*"The challenge of building models of receptor induced cascades"*, HBP Annual Summit, Maastricht, Netherlands, October 16 – 18, 2018.

**Krisztina Fehér:**

*"Atomistic simulations of immune stimulatory single stranded bacterial DNA"*, ISQBP President's meeting 2018, Barcelona, Spain, June 19 – 21, 2018.

# 9 Miscellaneous

**Gaurav K. Ganotra, Kashif Sadiq, Imme Roggenbach, Kai Horny and Rebecca Wade:**
*"Computation of diffusional association rates (kon) for drug-like compounds using Brownian dynamics simulations"*, BDBDB4: 4th Biological Diffusion and Brownian Dynamics Brainstorm (BDBDB4), Studio Villa Bosch, Heidelberg, Germany, April 16 – 18, 2018.

**Gaurav K. Ganotra, Daria B. Kokh, Kashif Sadiq and Rebecca Wade:**
*"COMparative BINding Energy (COMBINE) analysis to predict drug-binding kinetics"*, 22nd European symposium on Quantitative Structure-Activity Relationships (EuroQSAR 2018), Thessaloniki, Greece, September 16 – 20, 2018.

**Martin Golebiewski, Hadas Leonov, Steffen Brinkmann, Wolfgang Müller:**
*"LiSyM Data Management"*, LiSyM Mid-Term Evaluation Meeting, Mannheim, Germany, May 2 – 3, 2018.

**Martin Golebiewski, Olga Krebs, Hadas Leonov, Stuart Owen, Maja Rey, Natalie Stanford, Andreas Weidemann, Ulrike Wittig, Katy Wolstencroft, Jacky L. Snoep, Carole Goble, Wolfgang Müller:**
*"Data Needs Structure: Data and Model Management for Systems Biology and Systems Medicine"*, 7th Conference on Systems Biology of Mammalian Cells, Bremen, Germany, July 4 – 6, 2018.

**Martin Golebiewski, Haralampos Hatzikirou, Wolfgang Müller, Lutz Brusch:**
*"FAIR data exchange in the life sciences by interoperability standards for heterogenous data and multicellular models"*, 5th e:Med Meeting on systems medicine, Berlin, Germany, September 24 – 26, 2018.

**Martin Golebiewski, Wolfgang Müller:**
*"The NormSys registry for modeling standards in systems and synthetic biology"*, International Conference on Systems Biology (ICSB 2018), Lyon, France, October 28 – November 1, 2018.

**Ana Herrera-Rodriguez:**
*"Multiscale molecular dynamics simulations of spider silk proteins self assembly"*, BDBDB4: 4th Biological Diffusion and Brownian Dynamics Brainstorm. Heidelberg, Germany, April 16 – 18, 2018.

**Erica Hopkins:**
*"PINK: more than a single color tool. Experimental data from UKIDSS, SDSS, and FIRST utilizing PINK's multiwavelenght extension"*, Astroinformatics Conference, Heidelberg, Germany, September 3 – 7, 2018.

**Leonhard Horst:**
*"Hydrodynamic Simulations of Stellar Interiors"*, E-Poster at International HPC Summer School on HPC Challenges in Computational Sciences, Ostrava, Czech Republic, July 8, 2018.

**Fan Jin and Frauke Gräter:**
*"How does phosphorylation affect intrinsically disordered proteins?"*, Gordon Research Seminar and Conference on Intrinsically Disordered Proteins, Les Diablerets, Switzerland, July 1 – 6, 2018.

**Daria B. Kokh, Julia Romanowska, Neil Bruce, Martin Reinhardt and Rebecca Wade:**
*"Modeling of diffusion-driven adsorption of biomolecules on solid surface"*, BDBDB4: 4th Biological Diffusion and Brownian Dynamics Brainstorm, Studio Villa Bosch, Heidelberg, April 16 – 18, 2018.

**Olga Krebs, Katy Wolstencroft, Stuart Owen, Martin Golebiewski, Natalie Stanford, Hadas Leonov, Finn Bacall, Alan R. Williams, Jacky L. Snoep, Wolfgang Müller, Bernd Rinn, Carole Goble:**
*"FAIRDOM: data and model management for systems biology projects"*, Advanced Lecture Course on Systems Biology, Innsbruck, Austria, February 28 – March 6, 2018.

**Olga Krebs, Martin Golebiewski, Natalie Stanford, Stuart Owen, Maja Rey, Andreas Weidemann, Ulrike Wittig, Katy Wolstencroft, Finn Bacall, Jacky L. Snoep, Carole Goble, Wolfgang Müller:**
*"FAIRDOMHub: a repository and collaboration environment for sharing systems biology research"*, International Study Group for Systems Biology (ISGSB) Conference 2018, Tromsö, Norway, September 24 – 28, 2018.

**Olga Krebs, Martin Golebiewski, Natalie Stanford, Stuart Owen, Maja Rey, Andreas Weidemann, Ulrike Wittig, Katy Wolstencroft, Jacky L. Snoep, Wolfgang Müller, Carole Goble:**
*"FAIRDOMHub for Findable, Accessible, Interoperable, and Reusable Research Data."* 17th European Conference on Computational Biology, Athens, Greece, September 8 – 12, 2018.

**Sébastien Lyonnais, Kashif Sadiq, Cristina Lorca-Oró, Andreas Meyerhans and Gilles Mirambeau:**
*"The HIV-1 ribonucleoprotein dynamically regulates its condensate behaviour and drives acceleration of protease activity through liquid phase separation"*, EMBO/EMBL Symposium: Cellular Mechanisms Driven by Liquid Phase Separation, EMBL, Heidelberg, May 14 – 17, 2018.

**Isabel Martin, Sara Wickstroem and Frauke Gräter:**
*"Mechanosensing in Focal Adhesions studied with Molecular Dynamics Simulations"* , BIOMS Symposium 2018, Heidelberg, Germany, October 1 – 2, 2018.

**Goutam Mukherjee, Prajwal P. Nandekar, Ghulam Mustafa, Rebecca Wade:**
*"How are Electrons Transferred from Cytochrome P450 Reductase to Cytochrome P450 Enzymes? Towards a Structural and Dynamic Understanding"*, Hünfeld 2018 Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, March 24 – 25, 2018. BDBDB4: Fourth Biological Diffusion and Brownian Dynamics Brainstorm Meeting, Heidelberg, Germany, April 16 – 18, 2018. 14th International Symposium on Cytochrome P450 Biodiversity and Biotechnology. York, United Kingdom, July 15 – 19, 2018.

**Ghulam Mustafa, Prajwal P. Nandekar, Rebecca Wade:**
*"Simulation of Human P450-membrane Interactions"*, 32nd Molecular Modelling Workshop (MMWS), Erlangen, March 12 – 14, 2018.

**Ariane Nunes-Alves, Neil Bruce, Gaurav K. Ganotra, Daria B. Kokh, Huan-Xiang Zhou, Rebecca Wade:**
*"Effects of macromolecular crowding on substrate diffusion rates"*, Biological Diffusion and Brownian Dynamics Brainstorm 4, Heidelberg, Germany, April 16 – 18, 2018. *"Effects of macromolecular crowding on the diffusion rates of enzyme substrates and drug-like molecules"*, Gordon Research Seminar on Computational Chemistry, West Dover, USA, July 21 – 22, 2018. Gordon Research Conference on Computational Chemistry, West Dover, USA, July 22 – 27, 2018. Center for Modelling and Simulation in the Biosciences Symposium 2018, Heidelberg, Germany, October 1 – 2, 2018.

**Agnieszka Obarska-Kosinska and Frauke Gräter:**
*"How collagen fibrils dynamically distribute and measure stress"*, Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, April 20 – 21, 2018.

**Mehmet Ali Öztürk, Rebecca Wade:**
*"Computation of FRAP recovery times for linker histone – chromatin binding on the basis of Brownian dynamics simulations"*, BDBDB4: 4th Biological Diffusion and Brownian Dynamics Brainstorm, Studio Villa Bosch, Heidelberg, Germany, April 16 – 18, 2018. Hünfeld 2018 Workshop on *"Computer Simulation and Theory of Macromolecules"*, Hünfeld, Germany, April 20 – 21, 2018.

**Mehmet Ali Öztürk, Vlad Cojocaru, Rebecca Wade:**
*"A computational approach to decipher chromatosome structure determinants"*, EMBO – EMBL Symposium: Principles of Chromosome Structure and Function, Heidelberg, Germany, September 5 – 8, 2018.

**Joanna Panecka-Hofman, Katarzyna Świerkula, Rebecca Wade:**
*"Investigating the dynamics of homotetramers and monomers of pteridine reductase 1, a folate pathway enzyme from trypanosomatid parasites"*, The HPC reporting session of the Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Poland, March 15 – 16, 2018.

**Joanna Panecka-Hofman, Katarzyna Świerkula, Rebecca Wade:**
*"Investigating the global and local dynamics of homotetrameric enzyme pteridine reductase by molecular dynamics and enhanced sampling simulations"*, Modelling and Design of Molecular Materials 2018 Conference, Polanica Zdroj, Poland, June 24 – 28, 2018.

**Benedikt Rennekamp, Frauke Gräter:**
*"Implementing Bond Ruptures in Biomolecules in MD Simulations"*, Workshop on Computer Simulations and Theory of Macromolecules, Hünfeld, Germany, April 20 – 21, 2018.

**Maja Rey, Ulrike Wittig, Andreas Weidemann, Wolfgang Müller:**
*"SABIO-RK: kinetic data for systems biology"*, 7th Conference on Systems Biology of Mammalian Cells, Bremen, Germany, July 4 – 6, 2018.

**Kashif Sadiq:**
*"Modeling Reaction-Triggered Infectivity during Retroviral Assembly and Maturation"*, 62nd Biophysical Society Meeting, San Francisco, USA, February 17 – 21, 2018. BIOMS Symposium, Heidelberg University, October 1 – 2, 2018.

**Kashif Sadiq, Rebecca Wade:**
*"Computing Protein-Ligand Binding Association Rate Constants by Combining Brownian Dynamics and Molecular Dynamics Simulations"*, 62nd Biophysical Society Meeting, San Francisco, USA, February 17 – 21, 2018.

**Kashif Sadiq, Daria B. Kokh, Gaurav K. Ganotra, Rebecca Wade:**
*"Computing protein-ligand association kinetics by combining Brownian dynamics and molecular dynamics simulations"*, BDBDB4: 4th Biological Diffusion and Brownian Dynamics Brainstorm, Studio Villa Bosch, Heidelberg, April 16 – 18, 2018.

**Katarzyna Świerkula, Rebecca Wade, Joanna Panecka-Hofman:**
*"Investigation of the trypanosomatid folate pathway through exploring pteridine reductase dynamics"*, BIOMS Symposium 2018, Heidelberg, Germany, October 1 – 2, 2018.

**Andreas Weidemann, Maja Rey, Ulrike Wittig, Wolfgang Müller:**
*"SABIO-RK – a manual curated database for reaction kinetics"*, 11th International Biocuration Conference, Shanghai, China, April 8 – 11, 2018.

**Andreas Weidemann, Frank Bergmann, Martin Golebiewski, Axel von Kamp, Maja Rey, Sven Sahle, Sven Thiele, Ulrike Wittig, Steffen Klamt, Ursula Kummer, Wolfgang Müller:**
*"de.NBI-SysBio: German de.NBI-node serving the Systems Biology Cycle"*, ELIXIR All Hands 2018, Berlin, Germany, June 4, 2018.

**Andreas Weidemann, Maja Rey, Ulrike Wittig, Wolfgang Müller:**
*"SABIO-RK – a curated database for reaction kinetics"*, 17th ECCB, Athens, Greece, September 8 – 12, 2018.

**Andreas Weidemann, Martin Golebiewski, Ulrike Wittig, Maja Rey, Wolfgang Müller:**
*"SABIO-RK – kinetic data for systems biology"*, International Conference on Systems Biology (ICSB 2018), Lyon, France, October 28 – November 1, 2018.

**Andreas Weidemann, Ulrike Wittig, Maja Rey, Wolfgang Müller:**
*"de.NBI-SysBio: Bioinformatics Services for Kinetic Data and Model Management"*, 5th de.NBI Plenary Meeting, Berlin, Germany, November 29 – 30, 2018.

**Christopher Zapp:**
*"Mechanosensation through Radicals in Tensed Collagen"*, 62th Annual Biophysical Society Meeting. San Francisco, California, US, February 17 – 21, 2018. Hands-on Workshop Density-Functional Theory and Beyond, Beijing, China, July 30 – August 10, 2018.

## 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). 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 COMBINE (Computational Modeling in Biology network). Member of the Richtlinienausschuss (German committee for engineering standards) VDI 6320 *"Datenmanagement im Bereich Life Sciences"*, Association of German Engineers (VDI). Co-chair Joint Ad-hoc Group on Standardization of Genomic Information Compression and Storage between ISO/IEC JTC 1/SC 29/ WG 11 (MPEG) and ISO/TC 276/WG 5. Member of the management committee and co-leader Working Group 1 (Community/platform-building) of the European COST action CHARME (Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research).

**Frauke Gräter:**
Member of BIOMS (Heidelberg Center for Modeling and Simulation in the Biosciences) Steering Committee. Faculty member, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University. Associated faculty member, HGS MathComp Graduate School, Heidelberg University. Faculty member, Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology (HBIGS), Heidelberg 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 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"*, DPG.

**Friedrich Röpke:**

Advisory board, *"Sterne und Weltraum"*. SpringerNature, Heidelberg, Germany. Member of the Organizing Committee for the Heidelberg Joint Astronomical Colloquium.

**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.

**Volker Springel:**

Member of the National Academy of Sciences, Leopoldina. Member of the Interdisciplinary Center for Scientific Computing (IWR), Heidelberg Heidelberg. Scientific Member of the Max Planck Institute for Astrophysics, Garching. 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.

**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.

**Anna Wienhard:**

Member of the Network Executive Committee of the *"Geometric Structures and Representation Varieties"* GEAR Network. Member of the Structure Committee of the International Mathematical Union. Fellow of the American Mathematical Society. Member of the Scientific Advisory Board Springer Lecture Notes in Mathematics. Member of the Scientific Advisory Board wissenschaftskommunikation.de. Member of the Scientific Advisory Board Ahlfors Bers Colloquium.

**Ulrike Wittig:**

Member of the STRENDA Commission (Standards for Reporting Enzymology Data).

## 9.4 Contributions to the Scientific Community

### Program Committee Memberships

**Daniele Alessandrini, Anna Wienhard:**

*"40. Süddeutsches Kolloquium über Differentialgeometrie"*, Heidelberg, June 29 – 30, 2018.

**Federica Fanoni, Maria Infusino, Anna Wienhard, Michael Winckler:**

*"German Chapter Conference of the European Women in Mathematics"*, Heidelberg, May 3 – 4, 2018.

**Martin Golebiewski:**

COMBINE 2018: 9th Computational Modeling in Biology Network Meeting, Boston, MA, USA, October 8 – 12, 2018.

**Frauke Gräter:**

Member of Scientific Steering Committee, PRACE. Member of the Board of Directors, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg. Member of the coordinating committee of the excellence cluster *"3D Matter Made to Order"* (KIT and Heidelberg University).

**Olga Krebs:**

The 10th Young Scientists School *"Systems Biology and Bioinformatics"* SBB-2018, Novosibirsk. 11th International Conference BGRS\SB-2018 – Bioinformatics of Genome Regulation and Structure \ Systems Biology, Novosibirsk, Russia, August 20 – 24, 2018.

**Wolfgang Müller:**

7th Conference on Systems Biology of Mammalian Cells, Bremen, Germany, July 4 – 6, 2018.

**Friedrich Röpke, Andreas Bauswein:**
Conference *"Chemical evolution and nucleosynthesis across the Galaxy"*, Heidelberg, November 26 – 29, 2018. *"Supernovae – From Simulations to Observations and Nucleosynthetic Fingerprints"*, 660. Wilhelm und Else Heraeus-Seminar, Physikzentrum Bad Honnef, Germany, January 21 – 24, 2018.

**Michael Strube:**
Area Chair at the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, Louisiana, USA, June 1 – 6, 2018.

**Rebecca Wade:**
Co-chair, Proteins/Structural Biology Theme, 17th European Conference on Computational Biology (ECCB 2018), Athens, Greece, September 9 – 12, 2018. Scientific Advisory Board: 22nd EuroQSAR Symposium, Greece, September 16 – 21, 2018.

## Workshop and Conference Organization

**Martin Golebiewski:**
Committee Meetings of ISO/TC 276 Biotechnology working group WG5 *"Data Processing and Integration"*, Beijing, China, June 13 – 15, 2018, Potsdam, Germany, December 10 – 12, 2018. COMBINE & de.NBI Tutorial *"Modelling and Simulation Tools in Systems Biology"*, International Conference on Systems Biology (ICSB 2018), Barcelona (Spain), September 16, 2016. NFDI4Life Workshop of the National German Research Data Infrastructure in the Life Sciences (NFDI4Life), Studio Villa Bosch, Heidelberg, Germany, October 2, 2018.

**Frauke Gräter:**
Scientific Committee BIOMS conference, Heidelberg, Germany, October 2018. Organizing committee *"Max Planck Croucher Symposium"*, Ringberg, Germany, June 2018. Organizer (w. N Pugno, Trento) *"Silk symposium"* at European Solid Mechanics Conference, Bologna, Italy, July 2 – 6, 2018.

**Vincent Heuveline:**
7th International Conference on High Performing Scientific Computing, Hanoi, Vietnam, March 19 – 23, 2018. Talk series *"IT Forum"*, Heidelberg University, Germany, started November 28, 2018.

**Vincent Heuveline, Nils Schween:**
2nd International Symposium on Energy System Optimization (ISESO), KIT Karlsruhe, Germany, October 10 – 11, 2018.

**Johannes Horn, Mareike Pfeil, Anna-Sofie Schilling, Florian Stecker, Anna-Maria Vocke:**
Girl's Day Workshop *"Mathematik und Muster: Was Omas Tapete mit Mathe zu tun hat"* at Mathematikon, Heidelberg University, April 26, 2018.

**Sebastian Lerch:**
Co-organizer, Workshop on Machine Learning and Neural Networks, September 24 – 26, 2018, Zugspitze, Germany.

**Wolfgang Müller:**
de.NBI metrics hackathon, Heidelberg, Germany, July 11, 2018. INCOME2018 Conference & Hackathon, Bernried, Germany, October 15 – 19, 2018. de.NBI FAIR Data Management Planathon, Kassel, Germany, November 19, 2018. Data Management Developers Foundry Workshop, Frankfurt, Germany, December 6 – 7, 2018.

**Ken'ichi Ohshika, Athanase Papadopoulos, Robert C. Penner, Anna Wienhard:**
Oberwolfach Workshop *"New Trends in Teichmüller Theory and Mapping Class Groups"*, Oberwolfach, Germany, September 2 – 8, 2018.

**Kai Polsterer:**
Session of the KDDIG at the International Virtual Observatory Alliance Interoperability Meeting in Victoria, Canada, May 27 – June 1, 2018.

**Friedrich Röpke, Andreas Bauswein, Sabrina Gronow, Christian Sand, Leonhard Horst, Fabian Schneider:**
*"12th Würzburg Workshop in Heidelberg"*, Studio Villa Bosch, Heidelberg, Germany, December 17 – 18, 2018.

**Alexandros Stamatakis:**
Main organizer of 2018 Computational Molecular Evolution Summer School, Heraklion, Greece, May 6 – 17, 2018.

**Volker Springel:**
AREPO Developer Workshop 2018, Heidelberg, Germany, June 4–5, 2018.

**Michael Strube:**
Program Co-Chair of the Second ACL Workshop on Ethics in Natural Language Processing, New Orleans, Louisiana, USA, June 5, 2018.

**Rebecca Wade, Neil Bruce, Gaurav Ganotra, Stefan Richter, Rommie Amaro (University of California San Diego), Ulrich Schwarz (Heidelberg University), Huan-Xiang Zhou (University of Chicago):**

BDBDB4: 4th Biological Diffusion and Brownian Dynamics Brainstorm, Studio Villa Bosch, Heidelberg, 16 – 18 April, 2018.

**Anna Wienhard:**

*"Heidelberg-Karlsruhe-Strasbourg Geometry Day"*, Heidelberg, Germany, November 30, 2018.

### Editorial Work

**Tilmann Gneiting:**

Editor-In-Chief: Annals of Applied Statistics.

**Michael Strube:**

Editorial Board: Dialogue & Discourse Journal.

**Rebecca Wade:**

Associate Editor: Journal of Molecular Recognition, PLOS Computational Biology. Editorial Advisor: 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: Geometry & Topology; Annales scientifiques de l'école normale supérieure; Geometriae Dedicata; Forum Mathematicum; Geometric and Functional Analysis; Annales Henri Lebesgue.

### Other contributions

**Antonio D'Isanto:**

15 outreach articles on astronomical subjects in the online newspaper Tom's Hardware Italia.

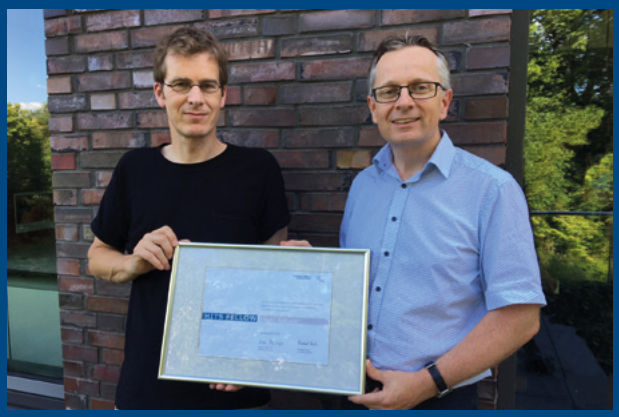
**Valentina Disarlo:**

Reporter for Oberwolfach Conference *"New Trends in Teichmüller Theory and Mapping Class Groups"*, September 2 – 8, 2018. Blogger and Editorial Board Member for the European Girls Math Olympiads (EGMO) 2018 Official Blog, April 9 – 15, 2018; Creator of the Series *"Women in Mathematics Beyond Stereotypes"* for EGMO 2018 Blog.

**Tilmann Gneiting:**

Highly Cited Researcher, category Mathematics, for 2006 – 2016, Clarivate Analytics, 2018.

## 9.5 Awards



**HITS Fellow Volker Springel**

In recognition of his *"exceptional scientific achievements and his service to the institute during his time at HITS,"* Prof. Volker Springel, head of the Theoretical Astrophysics group, was made an honorary *"HITS Fellow"* – the institute's highest honor – in a ceremony on July 23, 2018.

One week later, Volker Springel assumed his new role as Director of the Max Planck Institute for Astrophysics (MPA) in Garching, after more than eight years at HITS.

*"The new appointment represents a great achievement for Volker Springel as well as for us as a research institute,"* said HITS Scientific Director Prof. Michael Strube at the farewell ceremony. *"We are delighted that the concept of our founder, Klaus Tschira, has been realized: to bring brilliant young researchers to the institute, to let them conduct research with the greatest possible freedom, and to give them the opportunity to advance their careers."*

**Alexandros Stamatakis:**

Highly Cited Researcher, category Cross Field, for 2006 – 2016, Clarivate Analytics, 2018. Teaching award by the dean of computer science for master's level programming practical in summer 2017, Karlsruhe Institute of Technology, Germany.

**Volker Springel:**

Software Development Award 2018 of the German Astronomical Society. HITS Fellow, Heidelberg, Germany.

**Anna Wienhard:**

Clay Senior Scholar.

*The HITS Scientific Advisory Board and the HITS management 2018. From left to right: Michael Strube (HITS Scientific Director), Wolfgang Müller (HITS Deputy Scientific Director), Adele Goldberg, Thomas Lengauer, Gert-Martin Greuel, Dieter Kranzlmüller, Alex Szalay, Gesa Schönberger (HITS Managing Director).*

## 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 2018, 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, former Head of Department at Ministry of Science, Research and the Arts Baden-Württemberg, 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**

### Scientific Director:

**Prof. Dr. Michael Strube**
(2017 – 2018)

### Deputy Scientific Director:

**PD 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 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 | © 2019 HITS gGmbH*

**Twitter: @HITStudies**
**Facebook: /HITStudies**
**Youtube: /TheHITSters**
**LinkedIn: company/the-heidelberg-institute-for-theoretical-studies**
**Instagram: the_hitsters**