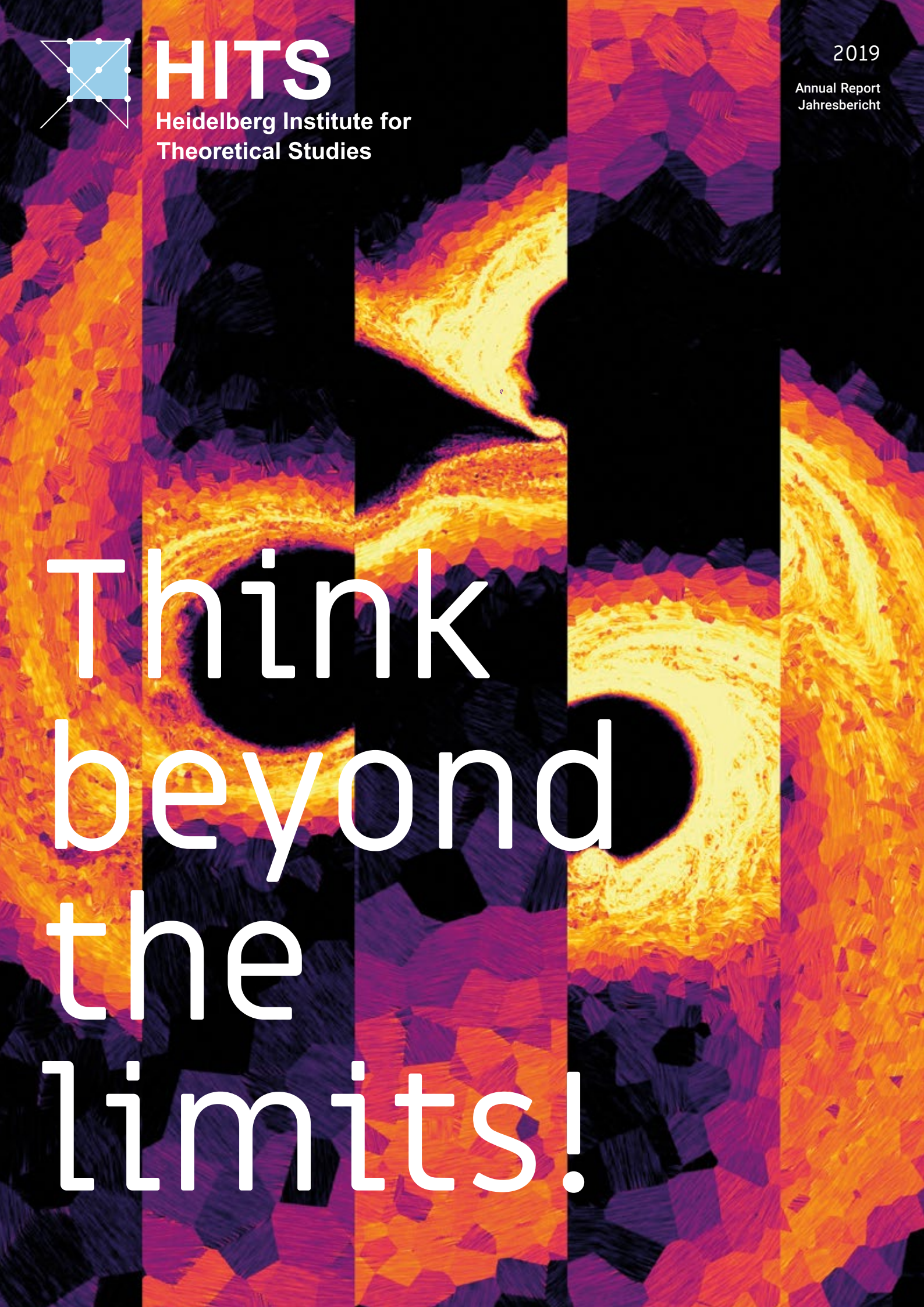
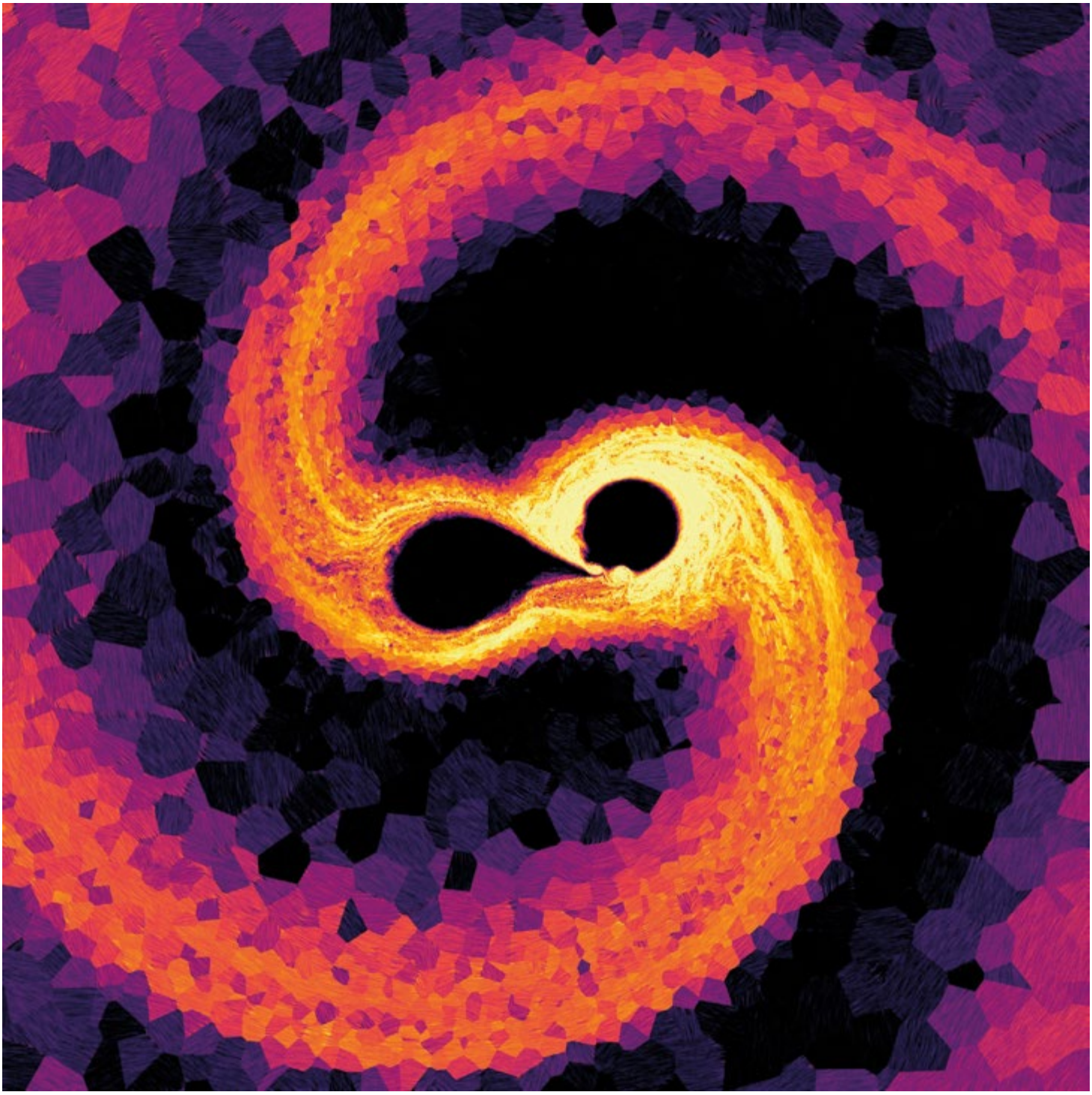

HITS
Heidelberg Institute for
Theoretical Studies
2019
Annual Report
Jahresbericht
Think
beyond
the
limits!

The irresistible pull: The simulation shown here marks the birth of a magnetic star such as Tau Scorpii. The image is a cut through the orbital plane where the coloring indicates the strength of the magnetic field and the hatching represents its field lines (cf Chapter 2.10, p 65).
(Picture: Ohlmann/Schneider/Röpke).

Einfach unwiderstehlich: Die Simulation zeigt die Entstehung eines Magnetsterns, wie zum Beispiel Tau Scorpii. Auf der Abbildung ist ein Schnitt durch die Bahnebene zu sehen. Die Färbung zeigt die Stärke des Magnetfelds, die Schraffierung stellt die Feldlinie dar (vgl. Kapitel 2.10, S 65).
(Bild: Ohlmann / Schneider / Röpke).



1 Think beyond the limits!

PD Dr. Wolfgang Müller
(Scientific Director / Institutssprecher)

Dr. Gesa Schönberger
(Managing Director / Geschäftsführerin)

Science emerges from conversation – this quote comes from physicist and Nobel Prize winner Werner Heisenberg. His book "Physics and Beyond" (German: "Der Teil und das Ganze") consists of reconstructions of conversations with other scientists in which new ways of thinking were developed.

Science emerges from conversation. At HITS, Heisenberg's concept proves itself time and again, especially last year. Many individuals who work here are visiting scientists who take advantage of the opportunity to conduct research together with HITS scientists for a certain period of time. We believe that this collaboration is an important element for the success of the Institute, and we want to further increase the attractiveness of our program to scientific guests.

In addition, speakers from Germany as well as from all over the world come to us regularly, be it to give smaller lectures within the Institute or to give a lecture in the public HITS Colloquium *(see Chapter 5.2)*. Colloquium lectures are streamed and posted on YouTube, and a discussion often develops after the lecture, which helps not only the questioners but also the lecturers to further their research.

The public lectures take place at the Studio Villa Bosch. The handicapped-accessible conference center is located directly between HITS and the Villa Bosch and shares common property with both. In addition to the Studio's unique surroundings, it offers the technical facilities to conveniently organize meetings with up to 90 participants and to provide full catering on site. A large foyer and three lecture rooms of various sizes are available for the meetings.

HITS makes intensive use of the Studio Villa Bosch. In addition to colloquia, monthly lab meetings also take place in the Studio. These meetings have a central function as they are used for important organizational matters and – even more importantly – are used by each research group to present its progress to the entire Institute once a year via two talks. These talks offer the opportunity to develop interdisciplinary ideas. The Lab Meeting is complemented by an internal Scientific Seminar Series, whose talks usually address a more specialist audience. Another opportunity for internal exchange is the "HITS Interdisciplinary Meeting," at which important topics concerning the scientific community are discussed.

The 2018 Meeting contributed substantially to the development of our 2019 Mission Statement.

HITS additionally organizes events with broader visibility, such as the Open House and the Alumni Meeting with former HITSters, both of which take place once every two years *(see chapter 5.3)*.

Our list would not be complete without the scientific events that HITS organizes with the professional support of the events team. From small project meetings to larger events, such as the international COMBINE meeting and the "Sino-German Symposium," HITS organized a good dozen events in 2019, with well over 300 participants in total *(see chapter 5.1)*. Many individuals fondly remember participating in a workshop at HITS. The setting, organization, and fast accommodation to the wishes of our guests are often praised in addition to the scientific quality of the events themselves.

Science emerges from conversation – that is, via synchronous communication. But does science require physical proximity? As we complete this annual report, measures to contain the coronavirus are in full swing. We all hope for a swift and well-planned end to these measures. Despite all the suffering of those with the virus and despite all the stress caused by *physical distancing*, opportunities are emerging for communication between scientists: For the first time, all collaboration partners are (almost) equally far apart. A video conference spanning two districts of Heidelberg feels exactly the same as a conference between Manchester and Heidelberg. New formats are being tested: How do 50 people meet meaningfully online for scientific exchange that does not take the form of a lecture? How can online communication create a sense of cohesion within working groups? How can we strike the right balance between small and large meetings? Moreover, how can we find the time to "get the real work done"? The current situation is forcing us to try out many new things. We are interested to see how communication within science will develop "after Corona."

Science emerges from conversation, especially when scientists from different disciplines talk with one another. HITS has been dedicated to this idea from the very beginning – in keeping with the motto "Think Beyond the Limits," which was also part of its predecessor institutions, EML European Media Laboratory (beginning in 1997) and EML Research (beginning in 2003). The change of name from EML Research gGmbH to HITS gGmbH in 2010 represented a significant expansion of the original concept: The goal was for HITS to develop into an interdisciplinary research institute with a focus on *data sciences*. From five research groups in 2010 HITS has 2019 expanded into eleven groups that consist of approximately 100 scientists from twenty countries, one-third of whom are women. A total of 122 people (as of the end of December 2019) work at HITS, including the service sectors of the administration, IT, and the communication and events team.

HITS is closely associated with Heidelberg University and the Karlsruhe Institute of Technology (KIT), which – in addition to the HITS-Stiftung – are also shareholders of the Institute *(see Chapter 10)*. This close association is reflected in joint appointments: At present, five group leaders at HITS also hold a professorship at one of the two universities. Additionally, two associated group leaders are professors at Heidelberg University, and one group leader is an honorary professor at the University. Two further joint appointments are planned for 2020.

As a well-established Institute that is closely linked to other research institutions, we will continue to organize scientific events in the future and – in line with the founding concept – expand our guest program to ensure that *science continues to emerge from conversation* at our Institute.

Wissenschaft entsteht im Gespräch – dieses Zitat stammt von dem Physiker und Nobelpreisträger Werner Heisenberg. Sein Buch „Der Teil und das Ganze" besteht aus Rekonstruktionen von Gesprächen mit anderen Wissenschaftlern, in denen neue Denkansätze entstanden.

Wissenschaft entsteht im Gespräch. Am HITS bestätigt sich diese Erkenntnis Heisenbergs immer wieder und ganz besonders im letzten Jahr. Viele, die hier arbeiten, sind Gastwissenschaftler und Gastwissenschaftlerinnen, die die Möglichkeit nutzen, für eine bestimmte Zeit mit den HITS-Wissenschaftler/-innen gemeinsam zu forschen. Wir sind uns einig, dass dies ein wichtiger Baustein für den Erfolg des Instituts ist, und wir wollen die Attraktivität unseres Programms für wissenschaftliche Gäste weiter steigern. Daneben kommen immer wieder Vortragende aus Deutschland, aber auch aus aller Welt zu uns: sei es, um institutsinterne kleinere Vorträge zu halten, sei es für das öffentliche HITS-Kolloquium ***(siehe Kapitel 5.2)***. Kolloquiums-Vorträge werden gestreamt und auf Youtube gestellt, und oft entspinnt sich nach dem Vortrag eine Diskussion, die nicht nur die Fragenden, sondern auch die Vortragenden wissenschaftlich weiterbringt.

Die öffentlichen Vorträge finden im Studio Villa Bosch statt. Das barrierefreie Tagungszentrum liegt zwischen dem HITS und der Villa Bosch und teilt sich mit beiden den Garten. Es bietet neben der einzigartigen Umgebung die technischen Möglichkeiten, Treffen mit bis zu 90 Teilnehmer/-innen bequem zu organisieren und diese dort auch zu bewirten. Hierzu stehen ein großes Foyer und drei Vortragsräume verschiedener Größe zur Verfügung. HITS nutzt das Studio Villa Bosch intensiv. Neben den Kolloquien finden hier auch die monatlichen Lab Meetings statt. Diese haben eine zentrale Funktion. Wichtige organisatorische Punkte werden hier verhandelt, aber noch wichtiger ist es, dass sich jede Arbeitsgruppe einmal im Jahr dem gesamten Institut mit zwei Vorträgen vorstellt. Diese Vorträge bieten die Möglichkeit, interdisziplinäre Ideen zu entwickeln. Das Lab Meeting wird durch das interne Scientific Seminar ergänzt, dessen Vorträge meist ein fachnäheres Publikum ansprechen. Eine weitere Gelegenheit zum internen Austausch ist das „HITS Interdisciplinary Meeting", bei dem gemeinsam wichtige Themen rund um den Wissenschaftsbetrieb diskutiert werden. Als ein Ergebnis des Meetings 2018 wurde die Entwicklung unseres Leitbildes 2019 entscheidend vorangebracht.

Darüber hinaus organisiert das HITS Events mit Außenwirkung, wie alle

zwei Jahre den Tag der offenen Tür und für ehemalige HITSter das Alumni-Treffen ***(siehe Kapitel 5.3)***. Unsere Auflistung wäre nicht vollständig ohne die wissenschaftlichen Veranstaltungen, die das HITS mit professioneller Unterstützung des Event-Teams durchführt.
Von kleineren Projekttreffen bis zu großen Events wie dem internationalen COMBINE-Meeting und dem „Sino-German Symposium" hat das HITS 2019 ein gutes Dutzend Veranstaltungen mit weit über 300 Teilnehmer/-innen veranstaltet ***(siehe Kapitel 5.1)***. Vielen bleibt ein Workshop am HITS im Gedächtnis. Gelobt wird neben der wissenschaftlichen Qualität die Umgebung, die Organisation und die gute und schnelle Reaktion auf Wünsche unserer Gäste.

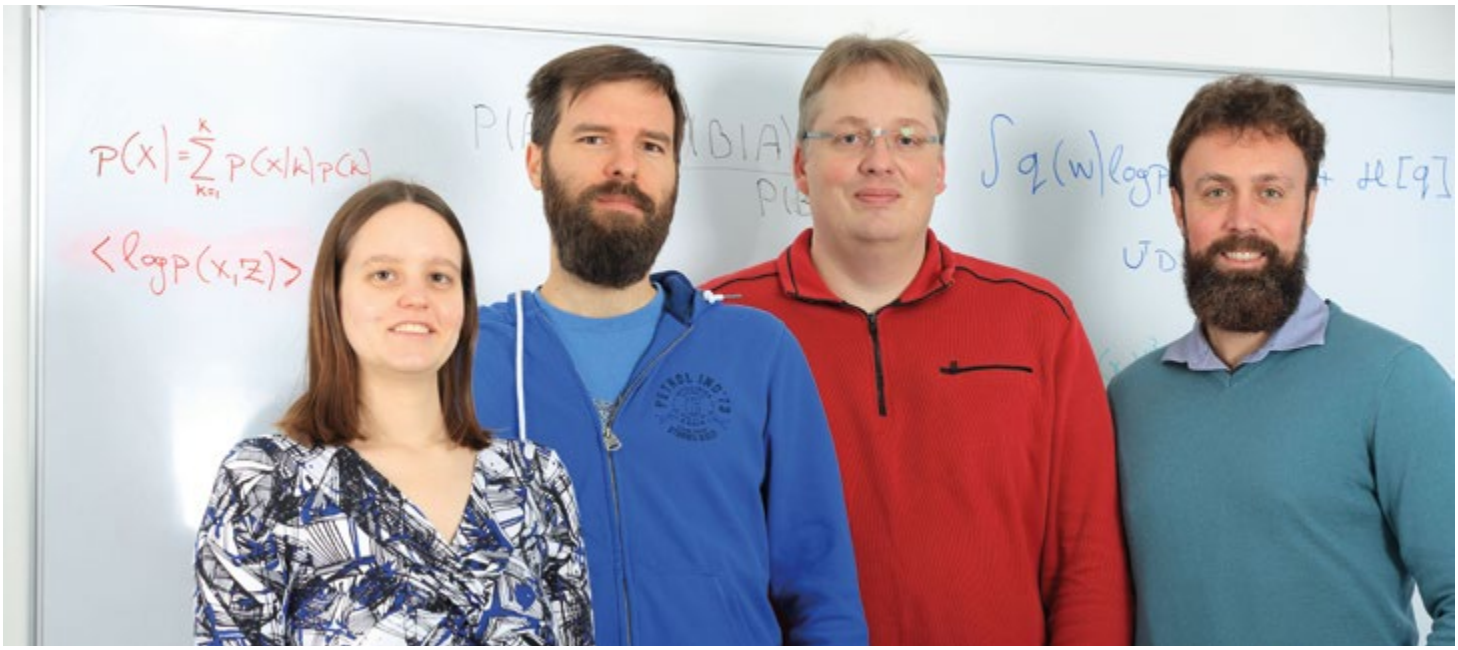
Wissenschaft entsteht im Gespräch, also durch synchrone Kommunikation. Aber braucht Wissenschaft physische Nähe? Während wir diesen Jahresbericht fertigstellen, sind Maßnahmen zur Eindämmung des Coronavirus in vollem Gange. Wir alle hoffen auf ein schnelles, gut durchdachtes Ende der Maßnahmen. Trotz allem Leid für die Patient/-innen, trotz aller Belastungen durch das *physical distancing* zeichnen sich für die Kommunikation zwischen Wissenschaftler/-innen auch Chancen ab: Kooperationspartner/-innen sind erstmals alle (fast) gleich weit entfernt. Die Videokonferenz zwischen zwei Heidelberger Stadtteilen fühlt sich genauso an wie die Konferenz zwischen Manchester und Heidelberg. Neue Formate werden ausprobiert: Wie treffen sich 50 Personen sinnvoll online zum wissenschaftlichen Austausch, jenseits von Vorträgen? Wie lässt sich durch Online-Kommunikation Nähe innerhalb von Arbeitsgruppen erzeugen? Wie finden wir das richtige Maß zwischen kleinen und großen Meetings? Und: Wie verschaffen wir uns Zeit, um „etwas zu schaffen"? Die jetzige Situation zwingt uns, vieles auszuprobieren. Wir sind gespannt, wie sich die Kommunikation innerhalb der Wissenschaft „nach Corona" entwickelt.

Wissenschaft entsteht im Gespräch. Vor allem dann, wenn Wissenschaftlerinnen und Wissenschaftler aus verschiedenen Disziplinen miteinander reden. Diesem Gedanken hat sich das HITS von Anbeginn verschrieben – ganz im Sinne des Mottos „Think Beyond the Limits!", das schon die Vorgängerinstitutionen EML European Media Laboratory (seit 1997) und EML Research (seit 2003) begleitet hat. Mit der Namensänderung von EML Research gGmbH zu HITS gGmbH im Jahr 2010 verband sich eine wesentliche Erweiterung des ursprünglichen Konzeptes: HITS sollte sich zu einem interdisziplinär arbeitenden Forschungsinstitut entwickeln, bei dem *Data Sciences* im Zentrum stehen. Aus den fünf Forschungsgruppen im Jahr 2010 sind inzwischen elf Gruppen geworden, mit rund 100 Wissenschaftlerinnen und Wissenschaftlern aus zwanzig Ländern, davon ein Drittel Frauen. Insgesamt arbeiten am HITS 122 Personen (Stand: Ende Dezember 2019), einschließlich der Service-Einheiten aus Verwaltung, IT und dem Kommunikations- und Event-Team.

Das HITS ist eng mit der Universität Heidelberg und dem KIT Karlsruhe verbunden, die neben der HITS-Stiftung ebenfalls Gesellschafter des Instituts sind ***(siehe Kapitel 10)***. Das schlägt sich in gemeinsamen Berufungen nieder: Zurzeit haben fünf Gruppenleiter/-innen des HITS zugleich eine Professur an einer der beiden Universitäten inne, hinzu kommen zwei assoziierte Gruppen und eine Honorarprofessur an der Universität Heidelberg. Zwei weitere gemeinsame Berufungen sind für das Jahr 2020 geplant.

Gut aufgestellt und eng mit anderen Forschungseinrichtungen verbunden werden wir auch in Zukunft wissenschaftliche Veranstaltungen organisieren und – ganz im Sinne des Gründungskonzepts – unser Gästeprogramm erweitern, damit bei uns weiterhin *Wissenschaft im Gespräch entsteht.*

2 Research
2.1 Astroinformatics (AIN)



Group Leader
Dr. Kai Polsterer

Staff members
Dr. Nikos Gianniotis
Dr. Antonio D'Isanto
Markus Nullmeier (from June until August 2019)

Scholarship holder
Erica Hopkins

Student assistant
Fenja Kollasch (since May 2019)

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 at high spatial, spectral, and temporal resolution. In addition, new, untapped wavelength regimes have yet to be investigated by new survey telescopes that are dedicated to 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 time-series analyses and redshift models based on photometric measurements. These tools will prove critical to the analysis of data in large upcoming 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. Other interests of the Astroinformatics group include morphologically classifying galaxies based on imaging data as well as similarity measurements in high-dimensional dataspaces.

Probabilistic inference of black-hole mass in reverberation mapping

Direct observations of super-massive black holes in the center of galaxies – as performed by the Event Horizon Telescope – go hand in hand with high costs and complicated tools of analysis. However, some physical quantities can be indirectly constrained by combining probabilistic modeling with low-resolution observations. In our work, we address the potential problem of underestimated black-hole masses in active galactic nuclei (AGNs). For example, we are currently studying the object Mrk509, a luminous Seyfert 1 galaxy located at a distance of 145 mega parsec (Mpc), by using photometric reverberation mapping of the accretion disk.

Our work is motivated by the desire to understand the physical structures believed to power AGNs, particularly the accretion disk (AD) that surrounds super-massive black holes. Spatially resolving the engine of the AGN is a task that requires very high resolution, which is achievable only via the combined use of a very large number of radio telescopes placed all around the world. A viable alternative is reverberation mapping (RM), which functions independently of the spatial resolution of the observing instrument and instead relies on estimating time delays between changes in the AD continuum and the emission lines from the photo-ionized clouds in the broad line region (BLR). Knowledge of time delays helps us estimate the distance of the BLR clouds to the accretion disk, thereby enabling greater insight into the geometry of an AGN. Time delays at different continuum bands shed further light on the temperature distribution across the disk and on its size.

However, in certain cases, time delays extracted from observations seem to suggest disk sizes that are larger (or, equivalently, black-hole masses that are larger) than anticipated by standard thin-disk theory, thereby leading to a disagreement. Interestingly, similar observations have also been made in microlensing studies of luminous quasars. Observations of our object of study, the Mrk509 galaxy, have also indicated a disagreement between standard disk theory and the disk size suggested by the time delays obtained through reverberation mapping.

One possible method of resolving this apparent disagreement is to revise the physical model that describes AGNs and especially their geometry: By assuming a thin-disk geometry with a low inclination (as opposed to no inclination), we can reconcile the discrepancy between observation and theory. The aim of our work in 2019 was therefore to investigate and quantify the plausibility of the hypothesis of an inclined geometry.

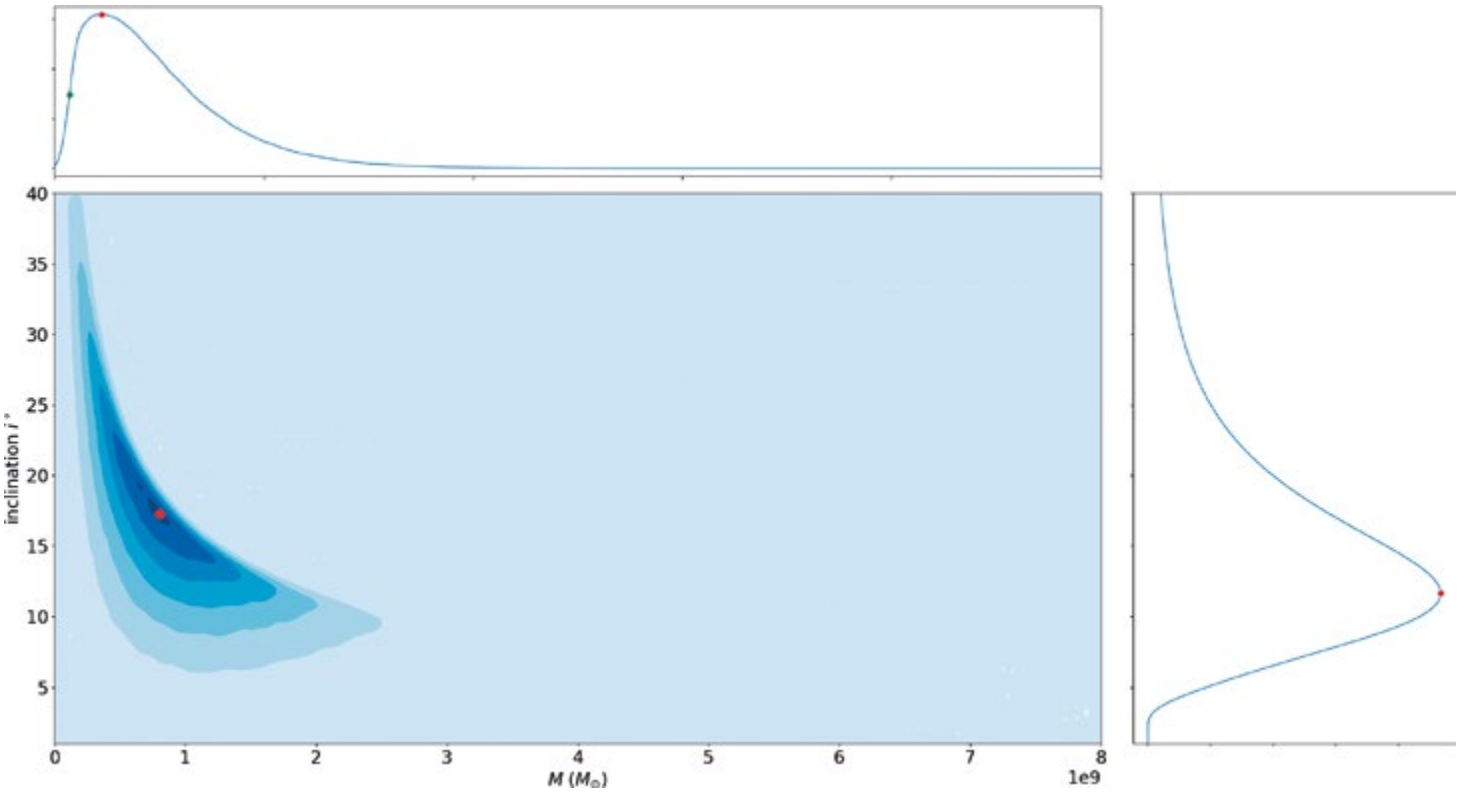


Figure 1: Joint probability density for mass and inclination along with marginals for the Mrk509 galaxy. The plot reveals how much the data support mass–inclination pairs as candidate solutions. The marginal distribution at the top shows the support for inclination angles. Similarly, at the right, support for black hole masses is shown. Modes of the densities are indicated with red dots. The green dot in the upper plot indicates the previously estimated black-hole mass as reported in: Peterson BM et al.: "Central Masses and Broad-Line Region Sizes of Active Galactic Nuclei. II. A Homogeneous Analysis of a Large Reverberation-Mapping Database." The Astrophysical Journal, Volume 613, Issue 2, pp. 682-699, 2004.)

To achieve this goal, we formulated a probabilistic model that describes the range of plausible values of the desired quantities of black-hole mass and inclination given the observed data. In this model, black-hole mass serves as a proxy for disk size, with higher black-hole-mass values indicating larger disk sizes. The formulated model relies on physical domain knowledge that describes how the observed time delays and luminosities mathematically interact with the physical parameters of black-hole mass and inclination. The formulation takes measurement errors in delays and photometry into account, and the model is capable of producing a joint probability density for the quantities of black-hole mass and inclinations, thereby fully quantifying the uncertainty of our estimates. The probabilistic analysis can also be conceived of as a method of checking how much support the observed data lend to each possible mass–inclination pair. Figure 1 displays the estimated joint probability density of mass inclination along with the marginals for mass and inclination. The data support relatively low inclinations, which also strengthens our hypothesis of a nearly face-on low inclination and an upward revised black-hole mass (and hence a larger disk). The most likely black-hole mass–inclination pair is 7.22×10^8 solar masses, 18.19° is shown as a red dot. The range of probable values for mass and inclination can be found in the marginal densities, where the most likely values are marked as a red dot. We note that in general, the mode of the joint density does not coincide with the mode of the marginals. Additionally, we plotted a previously reported estimate for mass as a green dot. It is evident that the previous black-hole mass estimate is less likely under the given model assumptions. As a consistency check, in Figure 2, we plotted the time-delay predictions that our model delivers when we plug in the inferred distribution of parameters presented in Figure 1, and we found that the observed time delays are within the predictive distribution of the model.

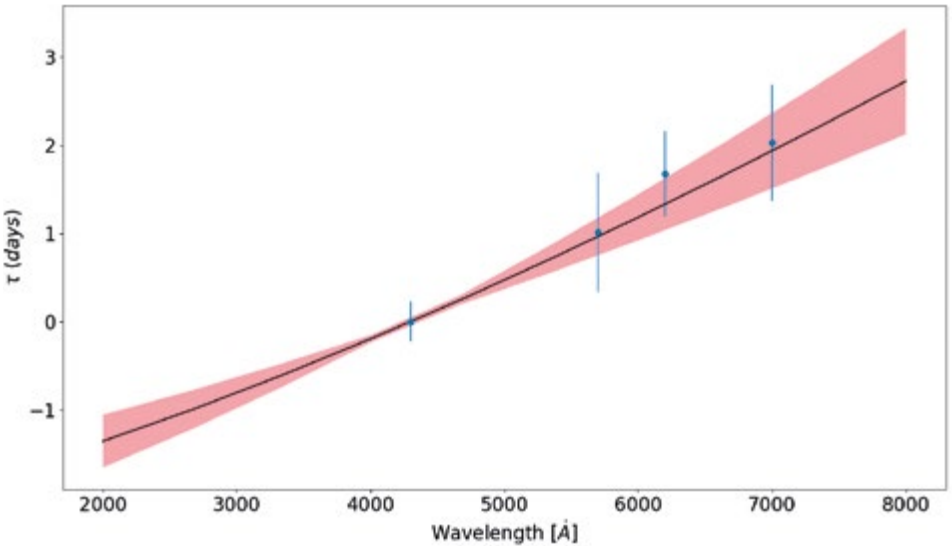


Figure 2: Predictive distribution of the model along with observed time delays in blue with error bars. The black line is the mean prediction, while the red shaded area corresponds to ±1 standard deviation from the mean.

Infrastructure for exploring astronomical spectra

Data archives have a long history in astronomy. In the past, mostly photographic plates – in addition to a few catalogs – were preserved. With the advent of digital detectors, more and more targeted digital observations and large area surveys have been taken and archived. This data explosion in astronomy demands the development of new techniques in terms of infrastructure, access, and analysis. The virtual observatory alliance aims at unifying access to all data archives through standardization and the provisioning of the necessary infrastructure. This work forms part of the ESCAPE project, whose goal is to establish a large European collaboration cluster that faces the new challenges created by data-driven research in astronomy and astroparticle physics – namely dealing with complex data workflows, infrastructural issues, data- and software interoperability, and FAIR data access.

The first phase of work at HITS focused on the development of novel explorative access methods for the astronomical data stored in the ESO archive. Explicit, criterion-based, positionally indexed searches currently form the standard in data-retrieval practice. Searching by structure as well as by similarity when providing examples is the first task in providing implicit methods of data retrieval. Thus far, we have worked on the development of a prototype (see Figure 3) based on dimensionality-reduction methods that allows for compressed visualization, inspection, and interaction with spectral observations.

Dimensionality-reduction methods project data items from a high-dimensional space to a low-dimensional one while aiming to preserve the similarity of the high-dimensional data, which means that similar objects should be placed near one another. This method enables scientists to browse massive datasets that are ordered by structural similarities to find classes, outliers/anomalies, and scientifically relevant objects in the dataset.

The dimensionality-reduction model employed for the prototype is based on an autoencoder – that is, an

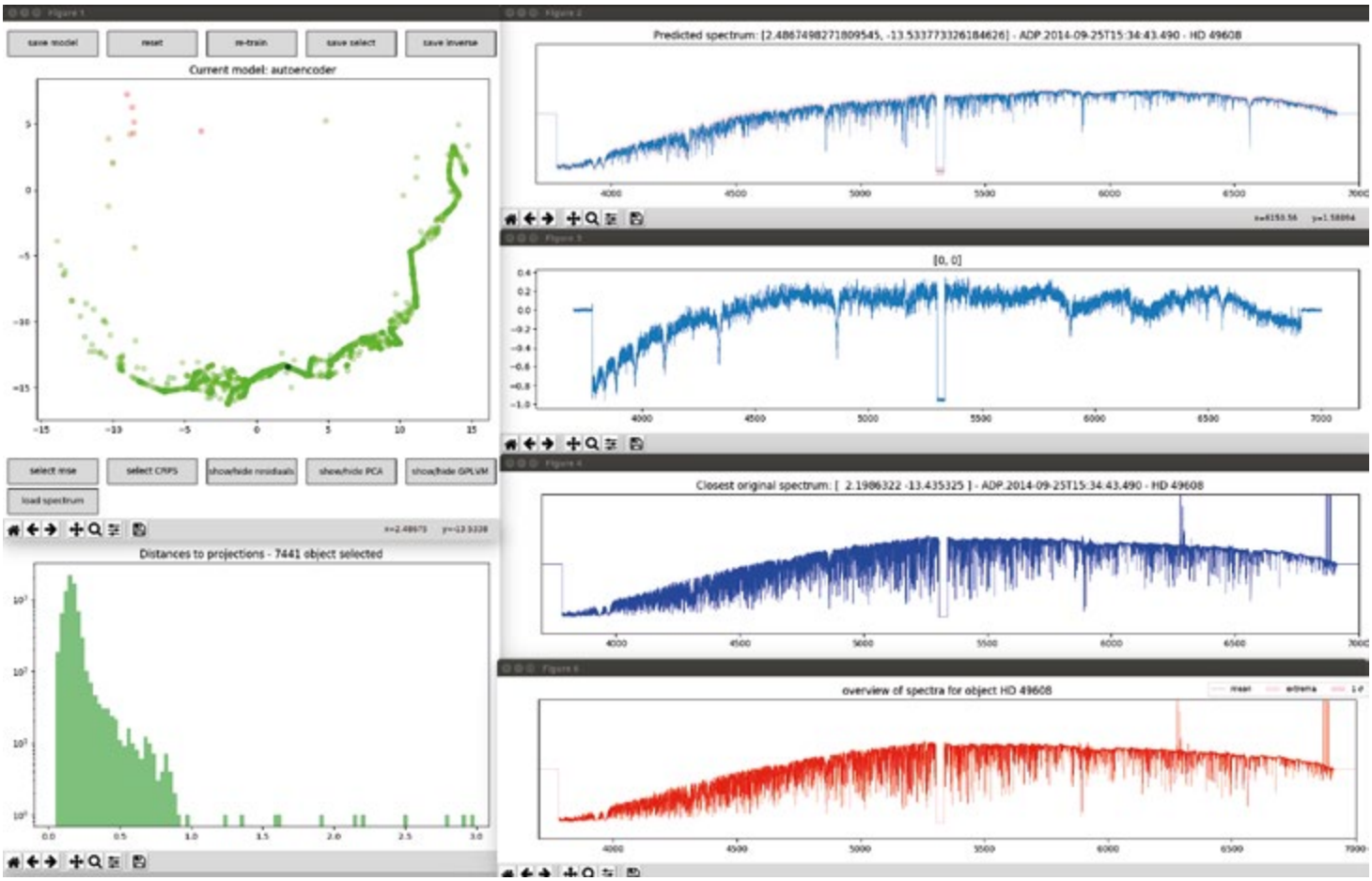


Figure 3: Main interface of the prototype for the exploration of spectra, which shows projections with the buttons that activate different functionalities (top left), reconstructed spectra, original spectra, spectral characterization of the selected sub-region and of the spectra belonging to a unique source (right), as well as a histogram (bottom left) that shows and selects sources based on their reconstruction error.

unsupervised neural network that is able to compress the original data into a low-dimensional representation. By generating a reconstruction from the low-dimensional space and comparing it with the original data, the network learns a low-dimensional representation that still allows for reconstructing the most relevant information from the input data.

When developing the prototype, we began by using high-resolution HARPS spectra. The data retrieval and preparation phase was particularly important as it allowed us to understand the requirements of the data-access process and also led to a further improvement of the user interfaces of the archive. As access to data and the development of standards form part of the tasks within ESCAPE, this phase was given a great deal of attention. For example, many spectral sources were observed multiple times, and HARPS was thus specifically dedicated to searching for exoplanets and to detecting velocity shifts in spectral lines. To compress the visualization, a unique source had to be identified. A hierarchical representation of the spectra – first by target and then by observation date – was required.

The autoencoder model was trained with the entire dataset of spectra by using the HITS GPU cluster, whereas the catalog containing the unique spectra was only used for visualizing the results (i.e., the projections) and for an interactive update when retraining the model. This method enabled the visualization and retraining to be performed on a standard laptop. The prototype allows for inspecting the projections obtained by the autoencoder and the corresponding reconstructions in real time. The model could clearly also be used to project data onto a low-dimensional space with more than two dimensions. Hence, in order to be able to present the projections on a screen, a two-dimensional space was selected. The prototype is connected to the most important VO tools (Aladin, Topcat, Splat) in order to be able to interact with the original spectral data and to obtain further information on the sources. Additional features are also available, such as the possibility of visualizing projections for all spectra of a particular source in order to retrain and update the model in real time, to select particular subsets for export or exchange, and to interactively further refine the selection. It is also possible to import spectra, to calculate their projections and their reconstruction, and to generate a catalog with the spectra's closest neighbors in the projected space. In addition to the autoencoder model, the prototype provides various other dimensionality-reduction methods, such as Principal Component Analysis and the Gaussian Process Latent Variable Model.

When inspecting the projections, a clear sequence in structural similarities is visible (see Figure 4). This sequence corresponds to the main spectral classes of stars, which was confirmed by checking the spectral classification from Simbad and over-plotting it as color.

The next steps in the project will involve refining the analysis on the current dataset and testing the prototype with other datasets. Testing additional models for dimensionality reduction and incorporating uncertainties will also be considered and are currently being discussed. We furthermore envision offering the tool as a web service in order to grant user access to astronomical data in a novel and efficient way.

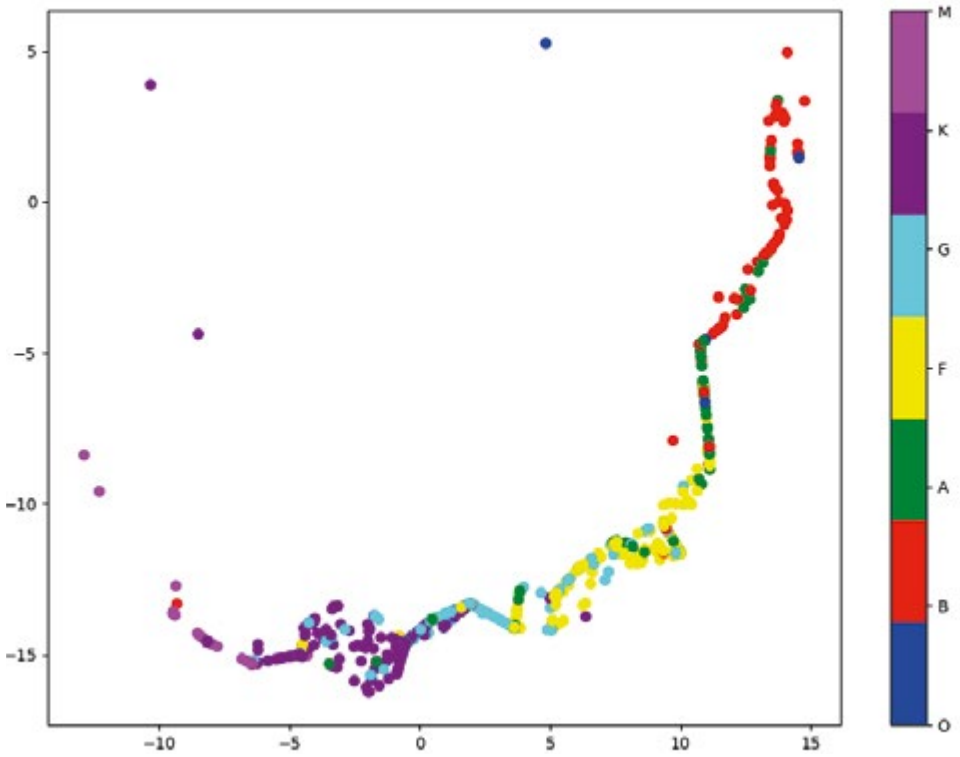


Figure 4: Projections for the HARPS spectra with the spectral class obtained by Simbad over-plotted in order to demonstrate the correlation between automatically derived structural sequences in the projected space and the spectral stellar classes.

Giant radio galaxies

Galaxies that have a strong signal in the radio spectrum emit powerful jets from the supermassive black hole (SMBH) in the center of the host galaxy. These jets are often large and easily exceed the size of the center of the galaxy, with some able to grow to extremely massive sizes. These giant radio galaxies (GRGs) are some of the largest objects in the Universe and are defined as a radio galaxy with an extent of at least 0.7Mpc. For reference, this is around 20–30 times larger than the diameter of the Milky Way. As a result, they are also extremely rare, with only a few dozen examples known. The mechanisms that allow for this extreme growth are currently being debated. One theory suggests that the interstellar medium (ISM) that surrounds these galaxies has a particularly low density and provides less resistance to their growth, while another theory suggests that these galaxies are very old and represent typical radio galaxies that have grown extremely large over an extended period of time. The currently small sample size renders it difficult to determine which theory, if either, could explain these galaxies.

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. Der technologische Fortschritt ermöglichte den Bau neuer Detektoren und innovativer Instrumente sowie das Design neuer Teleskope. Dadurch können Astronomen nun Objekte mit bisher unerreichtem Detailreichtum und in neuen Wellenlängenbereichen beobachten. Mit speziell dafür vorgesehenen Teleskopen wird der Himmel wiederholt durchmustert und die so gewonnen Daten werden frei zur Verfügung gestellt. Durch unsere Forschung ermöglichen wir es Wissenschaftlern, diese riesigen Datenmengen durch neue Analysemethoden explorativ und unvoreingenommener zu erschließen und somit effizienter zu nutzen.

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

In collaboration with Tim Galvin at CSIRO, we have found 17 GRGs, 16 of which are – to our knowledge – previously unidentified. We discovered the GRGs in the Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) survey, which is already 20 years old. The GRGs range from 0.7 Mpc to 1.1 Mpc in size and were discovered by using a rotationally invariant, self-organizing map (SOM), for which a GPU accelerated-software implementation was developed at HITS. A SOM is a dimensionality-reduction technique that reduces the high-dimensional data to a set of prototypes arranged in a two-dimensional space, thereby allowing us to learn which morphologies exist in the dataset. We used these prototypes to narrow the data to cases that could potentially be GRGs. In so doing, we considered extended radio objects that contain multiple radio sources. From there, we selected the objects for which the center source aligned with an infrared object in the Wide-field Infrared Survey Explorer (WISE) survey and for which the infrared object had an associated redshift, which could be converted into a distance. Next, we determined the physical size of the radio object and whether it qualified as a GRG.

In Figure 5, four of the new GRGs that we discovered are displayed. The figure reveals the extent of each GRG along with the radio components in FIRST and the predicted IR position and the associated probable IR source in WISE. In some cases, other nearby radio components exist that we believe to be related to the GRG. The GRGs vary in size but are generally a couple of arcseconds in diameter.
In the future, we hope to expand this work by using much newer and deeper radio data. The FIRST survey ended over 20 years ago, and since then, our technology has improved drastically. The new generation of radio telescopes – the Square Kilometer Array (SKA) – is currently being built (mainly in Australia and South Africa) and is set to completely revolutionize radio astronomy. About 2.5 million observed radio sources currently exist, less than one million of which have been observed in FIRST. The SKA is expected to observe around 70 million sources, which will open up a new domain of radio astronomy and can hopefully help shed light on what causes these radio galaxies to grow so tremendously large.

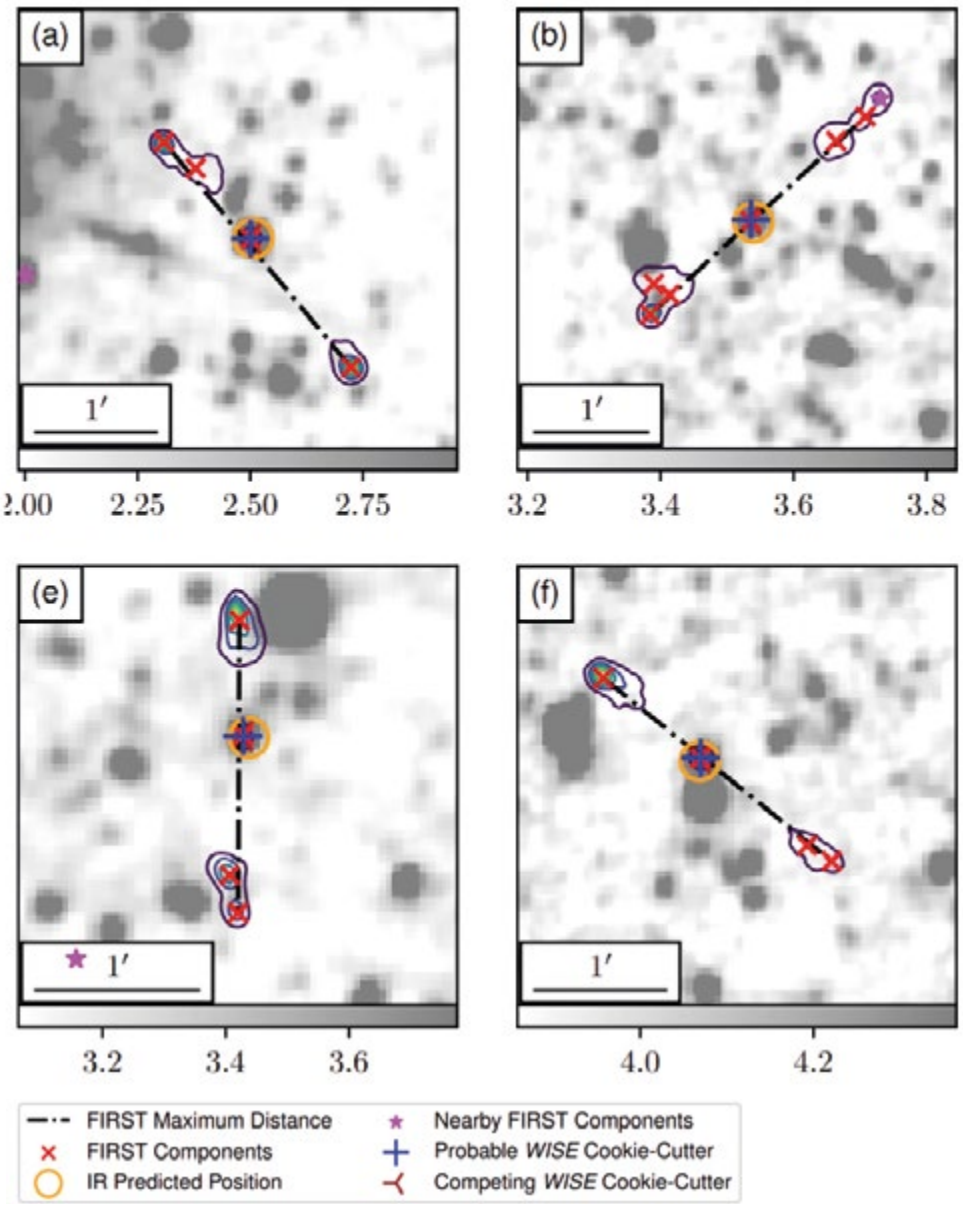


Figure 5: Four of the 17 GRGs we discovered in the FIRST survey. The gray scale represents the WISE data, with the contours referring to the radio fluxes of FIRST.

2 Research

2.2 Computational Carbon Chemistry (CCC)

Group Leader
Dr. Ganna Gryn'ova (since April 2019)

Staff members
Dr. Christopher Ehlert (since October 2019)
Anna Piras (since October 2019)

Scholarship holder
Oğuzhan Kucur (since September 2019)

computational carbon chemistry

The Computational Carbon Chemistry (CCC) group was established in April 2019 when Ganna (Anya) Gryn'ova came to HITS to lead the new junior group. The other three current members joined the group in September–October 2019. The CCC group uses theoretical and computational chemistry to explore and exploit diverse functional organic materials.

Organic materials can be broadly defined as materials that consist primarily of carbon atoms, with millions of examples already known and many more theoretically possible. This diversity of composition and structure translates to a plethora of shapes, colors, and other physical properties as well as to a corresponding breadth of practical applications in the chemical industry, medicine, electronics, etc.

In the CCC group, we are primarily interested in applying diverse state-of-the-art computational chemistry techniques to simulate and rationalize the chemical behavior of various organic molecules and compounds with the ultimate goal of building structure–property–function frameworks. We always pursue the fundamental physical principles that determine chemical behavior while keeping the targeted practical applications and the associated experimental testing in mind.

In this report, we document the early stages of the CCC group's research agenda in the fields of graphene chemistry and free radical chemistry. Following a brief introduction and a description of the motivation behind this research, we outline the preliminary work and future goals in two key areas: (i) employing graphene-based materials in the detection and transformation of nitroaromatic pollutants and (ii) designing unconventional free radical species for applications in molecular electronics.

Graphene chemistry: Accurate modeling and applications in environmental remediation

Ganna Gryn'ova, Christopher Ehlert, and Anna Piras

Graphene is an allotropic form of carbon consisting of a monolayer two-dimensional hexagonal lattice. Together with its many chemically modified derivatives (graphene oxide, reduced graphene oxide, doped and/or functionalized graphene oxide), graphene has recently emerged as an attractive alternative to silicon and transition-metal-based materials in the areas of efficient and selective carbocatalysis for a broad range of chemical reactions. Further advantages of graphene-based materials (GBMs) include their natural biocompatibility and nontoxicity, their infinite structural variability, their mechanical flexibility, their unique electronic and optical properties, their relative ease of fabrication and functionalization, and their inherent molecular nature. Biomass transformation and environmental protection represent some of the most attractive practice-oriented applications of GBMs and are the central focus of our research. We are currently developing this area in three parallel directions, as discussed below.

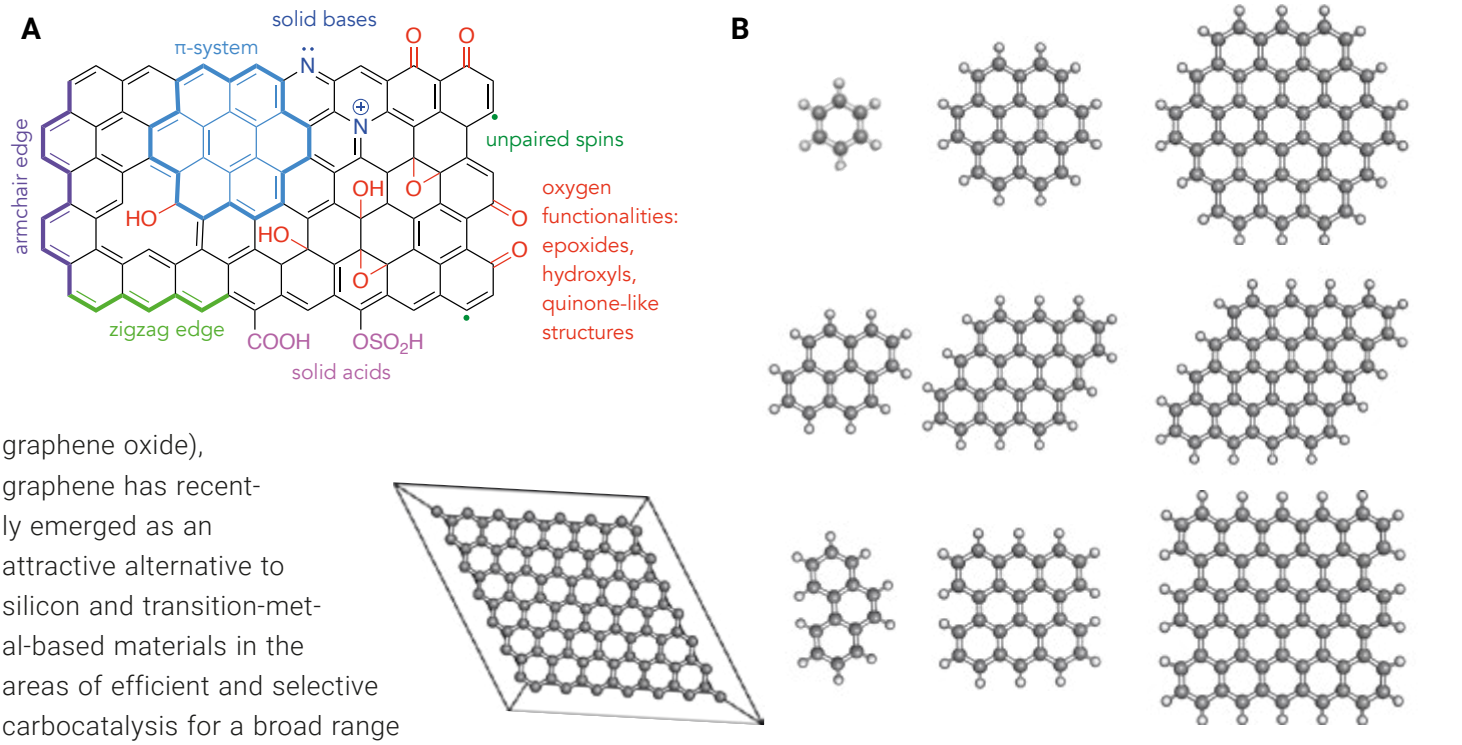


Figure 6: A Common active sites in graphene-based catalysts. B A range of computational models of pristine graphene.

First, the validity of our in silico results and predictions depends on the adequate choice of a model system and the accuracy of the computational technique. Extensive models involving the two-dimensional periodicity of a graphene-based catalyst in order to avoid edge effects are often feasible only with relatively cheap methods, such as the local-density approximation and the generalized gradient approximation density functional theory (DFT), which are not always sufficiently accurate for molecular systems. Much more sophisticated and reliable hybrid DFT functionals and ab initio methods can be applied to small-molecule models that reproduce the local chemistry of the active sites and capture the intricate details of the elementary reaction steps while unavoidably sacrificing the realistic description of the extended catalytic system. Furthermore, the computations for either approach should ideally be performed for every possible catalytic site – that is, for each potential active site and its immediate chemical surroundings, which can vary significantly in terms of the nature, quantity, and position of the defect units (Figure 6A). To address this ambiguity, we are currently conducting an extensive benchmarking study of representative graphene chemistries (the adsorption of small molecules, one-electron oxidation, and reduction) across a spectrum of model systems using a variety of theoretical approaches (Figure 6B). Computational results arising from multiple combinations of models and methods are being validated across levels of accuracy and – ultimately – against available and consistent experimental data. Our main goal is to establish a reliable and transferable in silico protocol for modeling fundamental chemical processes involving graphene with the best accuracy-to-efficiency ratio. Later in 2020, we will also contribute an invited focus article on the computational modeling of the chemistry of graphene-based materials.

Our second avenue of research in this area involves using GBMs to

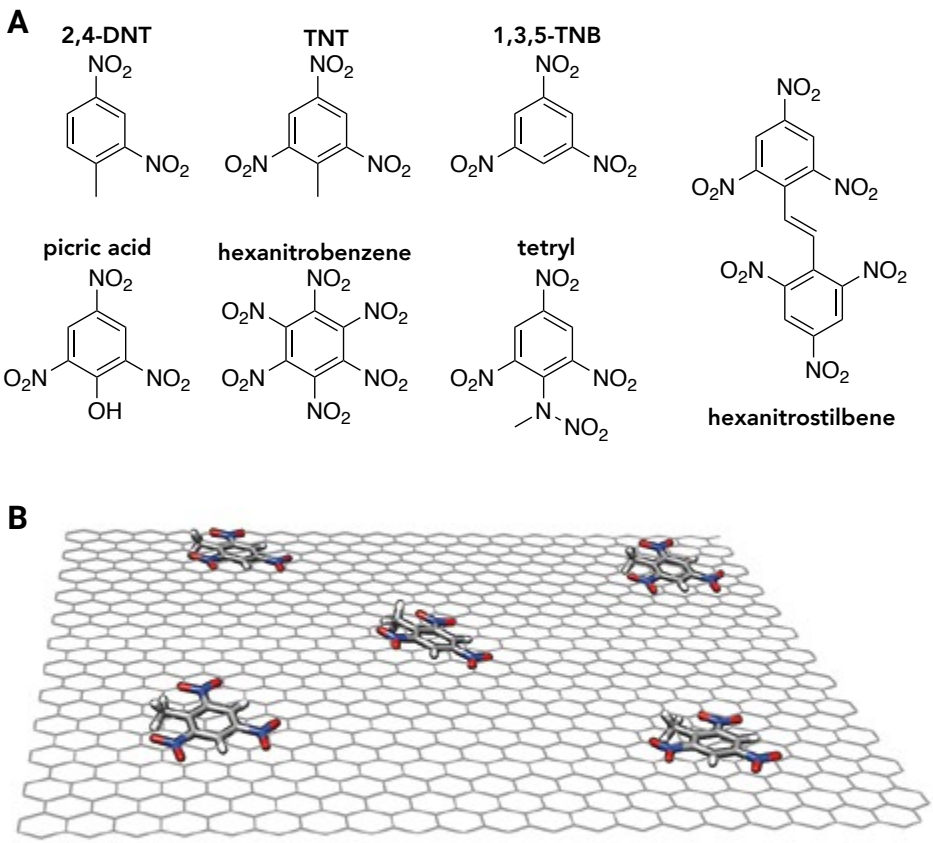


Figure 7: A Examples of common toxic and mutagenic nitroaromatic pollutants (with their colloquial names) from the EU- and USA priority pollutants lists. B Snapshop of a periodic simulation of TNT adsorption on pristine graphene.

detect nitrogen-containing aromatic compounds (NACs, Figure 7A), which are the main components of military explosives due to their ability to release large amounts of heat via self-oxidation. Contamination by NACs is a serious hazard to human health since nitroaromatic compounds are easily adsorbed through the skin and have a high bio-accumulation rate that leads to liver damage and other pathologies. The detection of small quantities of NACs – especially in water – is therefore of great importance for security reasons, particularly in forensics and anti-terrorism investigations, and is necessary to classify contaminated sites and proceed with removal and purification efforts. Electrochemical detection, which is based on the reduction of the nitro groups to amino functionalities, offers the possibility of an on-site real-time analysis with high sensitivity, a low limit of detection (LOD), a large linear range, and a relatively low cost of the apparatus. The development of the graphene-based electrochemical sensors, in particular, is attracting interest due to GBMs' exceptional electrochemical and mechanical properties, which allow for the production of cheap, robust, and highly sensitive devices. A number of GBMs have been exploited in the electrochemical detection of NACs and demonstrate highly diverse performance in terms of sensitivity and selectivity. Nevertheless, to the best of our knowledge, no theoretical studies have yet been performed that shed light on this behavior. Our approach to fill this research gap is to rationalize these observations from the perspective of non-covalent interactions (NCIs) between the nitroaromatic analyte and the graphene substrate. We are currently employing a range of quantum-chemical methods to simulate the adsorption of NACs on GBMs as well as to visualize, quantify, and analyze the corresponding NCIs (Figure 7B). In this manner, we aim to establish a relationship between the surface chemistry of graphene-based electrochemical sensors and the LOD of the detection method and ultimately to deduce a set of design guidelines for custom graphene-based sensing materials for the highly sensitive and selective detection of different NAC pollutants.

Finally, we aim to uncover the mechanisms by which graphene and its derivatives catalyze the reductive transformation of harmful nitroaromatic pollutants into valuable aminoaromatic synthons (Figure 8A). Typically, the decomposition of these persistent and non-biodegradable water- and soil contaminants requires aggressive and/or costly reagents, such as transition metal ions and H_2O_2 in Fenton process as well as noble metal catalysts for electrochemical reduction. Graphene-based catalysts represent an attractive green and sustainable alternative to these methods. However, several intertwined issues hinder progress in this field: (i) In some systems, the observed catalytic effects are due to trace metals rather than to the graphene-based material itself, and (ii) for the vast majority of carbocatalytic processes, the actual mechanisms and chemical nature of the active sites remain largely unknown or speculated. Our goal is to employ computational modeling to resolve these problems. To achieve this goal, we are currently quantifying and analyzing the adsorption energies of the key intermediates in nitrobenzene reduction on diverse defect structures in pristine graphene and on various doped graphenes (Figure 8B). In combination with available experimental evidence, this procedure allows for determining the relationship between the nitrobenzene•••GBM energetics and the GBMs'

respective catalytic activity. Ultimately, clarifying the key mechanistic aspects of the observed catalysis and definitively establishing the associated active sites would provide clear insight into the role of metal impurities. Furthermore, linking the chemical structure of the material to its catalytic activity can yield invaluable catalyst design principles and lay the groundwork for a systematic improvement of catalytic efficiencies and selectivities.

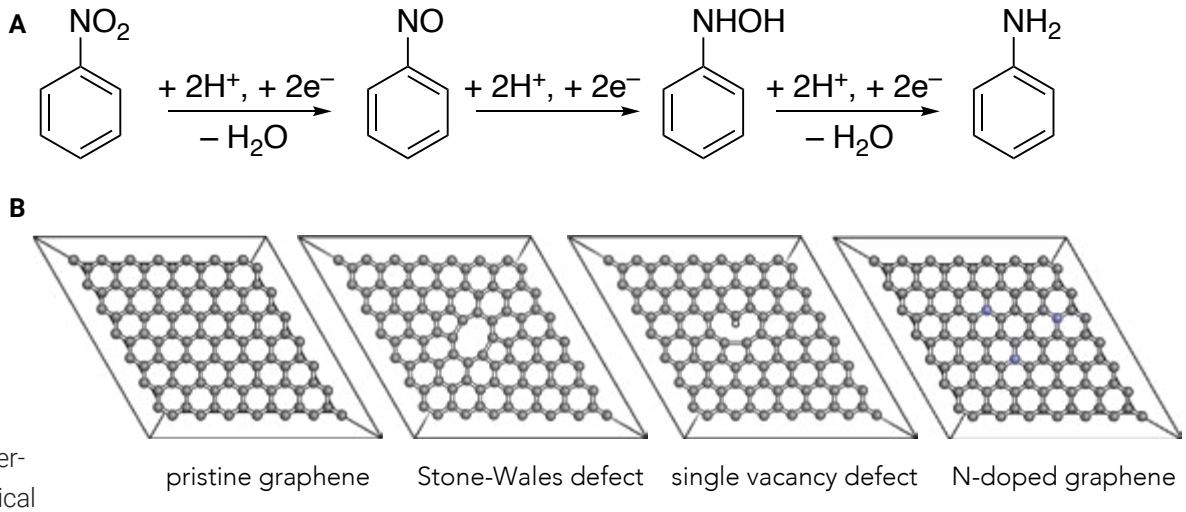


Figure 8: A General scheme of a stepwise nitrobenzene reduction via key intermediates. B Exemplary unit cells of various graphene-based catalysts.

Free radicals in molecular electronics

Ganna Gryn'ova and Oğuzhan Kucur

According to the International Union of Pure and Applied Chemistry, a free radical is a molecular entity that possesses an unpaired electron. These species are ubiquitous in nature and exist in environments ranging from a human body to interstellar space. Moreover, they are employed in a broad range of chemical fields, from natural product synthesis and control radical polymerization to mass spectrometry and electron paramagnetic resonance analyses. In the CCC group, we are especially interested in several aspects of radical chemistry, as outlined below.

One of our research foci involves modeling, analyzing, and designing non-conventional non-aufbau free radicals. The aufbau principle dictates that molecular orbitals are filled with electrons in an order of increas-

Die HITS-Juniorgruppe Computational Carbon Chemistry (CCC) wurde im April 2019 von Ganna (Anya) Gryn'ova gegründet, drei weitere Mitglieder kamen im September bzw. Oktober 2019 hinzu. Die Forschungsgruppe arbeitet mit den neuesten Methoden der theoretischen und computergestützten Chemie, um funktionale organische Materialien zu untersuchen und auszuwerten.

Grob gesprochen bestehen organische Materialien vornehmlich aus Kohlenstoffatomen. Dazu gibt es bereits Millionen von bekannten Beispielen, viele weitere sind theoretisch möglich. Die Vielfältigkeit bei Zusammensetzung und Struktur führt zu einer Fülle von Formen, Farben und anderen physikalischen Eigenschaften sowie zu einer entsprechenden Bandbreite praktischer Anwendungsmöglichkeiten u.a. in der chemischen Industrie, der Medizin und der Elektronik.
In unserer Forschungsgruppe arbeiten wir hauptsächlich mit den neuesten Methoden der Computerchemie, um das chemische Verhalten organischer Moleküle und Komponenten zu simulieren und zu rationalisieren. Unser vorrangiges Ziel dabei ist die Bildung von Struktur–Eigenschafts–Beziehungen. Dabei konzentrieren wir uns auf die fundamentalen physikalischen Gesetzmäßigkeiten, die das chemische Verhalten bestimmen, ohne die gezielten praktischen Anwendungen und die damit verbundenen experimentellen Prüfungen aus den Augen zu verlieren.

In diesem Bericht dokumentieren wir die Anfänge der CCC-Forschungsgruppe im Bereich Graphen-Chemie und der Chemie der freien Radikale. Nach einer kurzen Einführung und Beschreibung der unserer Forschung zugrundeliegenden Motivation fassen wir die bisherige Forschungsarbeit kurz zusammen und geben einen Ausblick auf zwei zukünftige Schwerpunktbereiche: (i) den Einsatz graphen-basierter Materialien beim Nachweis und der Transformation nitroaromatischer Gefahrstoffe und (ii) das Design unkonventioneller freier Radikalarten für die Anwendung in der molekularen Elektronik.

ing energy, which implies that the unpaired electron in free radicals occupies the orbital with the highest energy (Figure 9A). However, a number of species have been reported over the years to violate this principle, thereby allowing for an unconventional application in molecular electronics (Gryn'ova et al., WIREs Comput. Mol. Sci., 2015, 5, 440). We are currently pursuing this research in several directions. First, we aim to definitively establish the requirements for constructing such non-aufbau free radicals. We have formulated a working hypothesis regarding which chemical factors lead to this unconventional electronic structure and have begun testing it in silico. Second, we are exploring the relationship between the electronic configuration and the electroluminescence properties of various derivatives of the polychlorotriphenylmethyl (PTM) radical (Figure 9B). Specifically, recent studies suggest that the introduction of different donor groups to the PTM structure enhances its photostability in organic light-emitting diodes (OLEDs) that display near-infrared emission. Different combinations within the aforementioned radical-donor couples result in distinct redox properties that are symptomatic of the non-aufbau occupation pattern. We have conducted an extensive review of experimental data and derived a pattern of chemical behavior in such couples, which we are currently analyzing in silico in order to derive the design framework for non-aufbau radical OLEDs.

We are also interested in using free radicals as components of single-molecule junctions (SMJs). SMJs represent a powerful tool for exploring intimate details of electron transport on a truly molecular level as well as a unique functional architecture with an inherent capability to respond to external physical and chemical stimuli. The presence of unpaired electron(s) within conducting wires allows for adding a spin-coupled dimension and the associated magnetic properties to the transport (Figure 9C). Research in this direction is in its infancy, with limited experimental studies of such assemblies having appeared in recent years. To advance this field, we aim to formulate a theoretical framework for designing molecular magnetic junctions with optimized spin-filtering efficiencies and considerable magnetoresistances. To this end, we are currently preparing an invited perspective article on open-shell single-molecule junctions in which we discuss the challenges to fabricating and characterizing these fascinating systems and – where possible – extract the relevant structure-property patterns. Subsequent research in this area will involve in silico testing of these patterns, modeling the designed radical junctions, and customizing their charge- and spin transport properties.

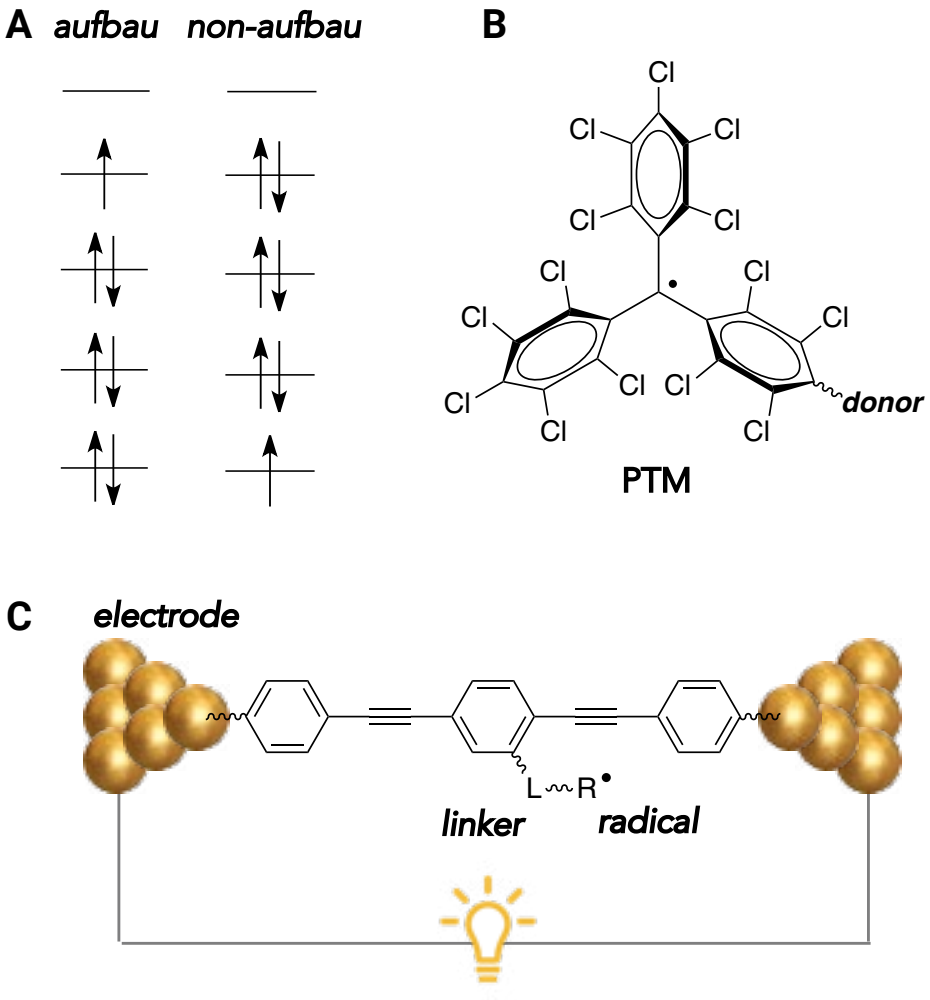
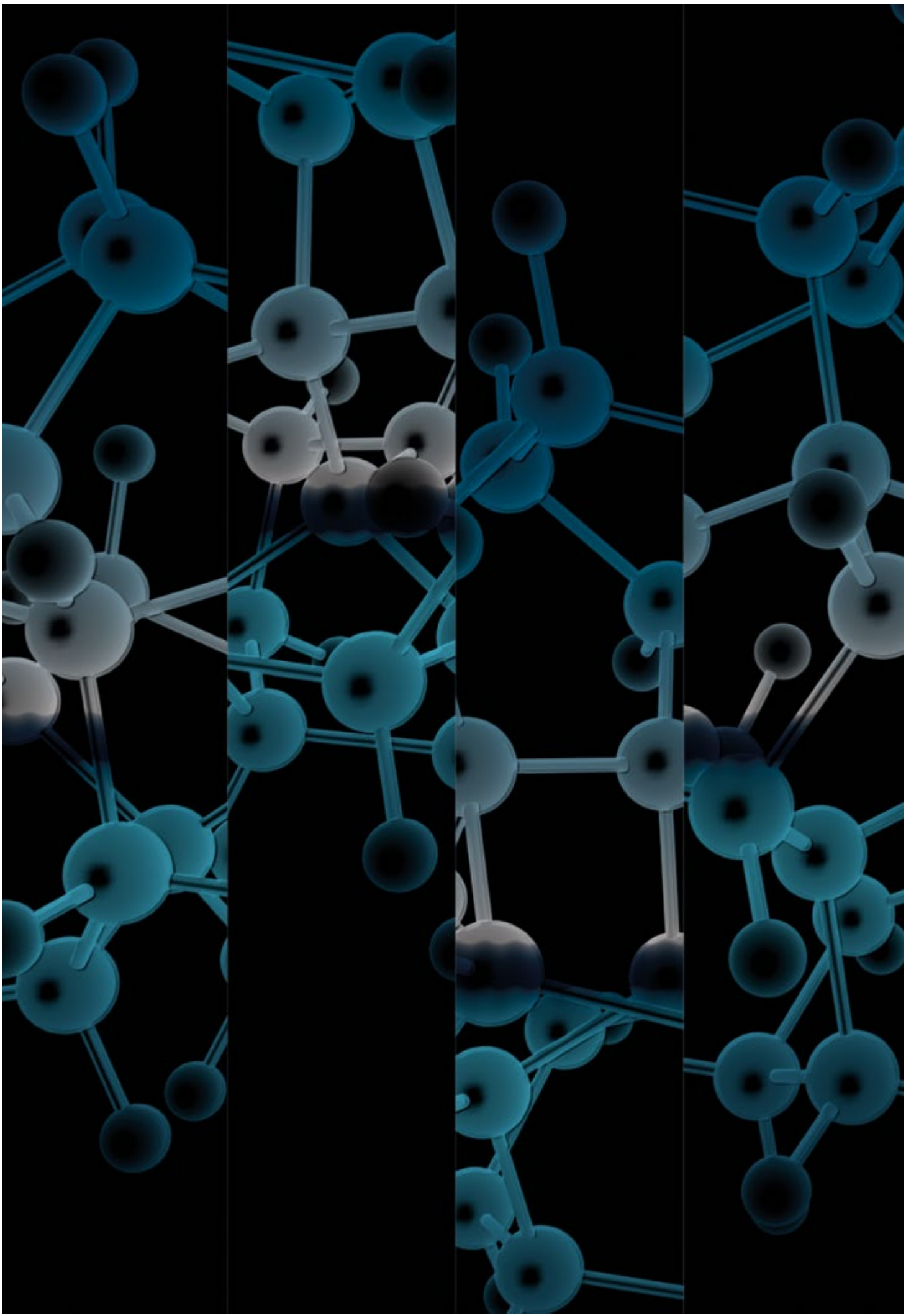
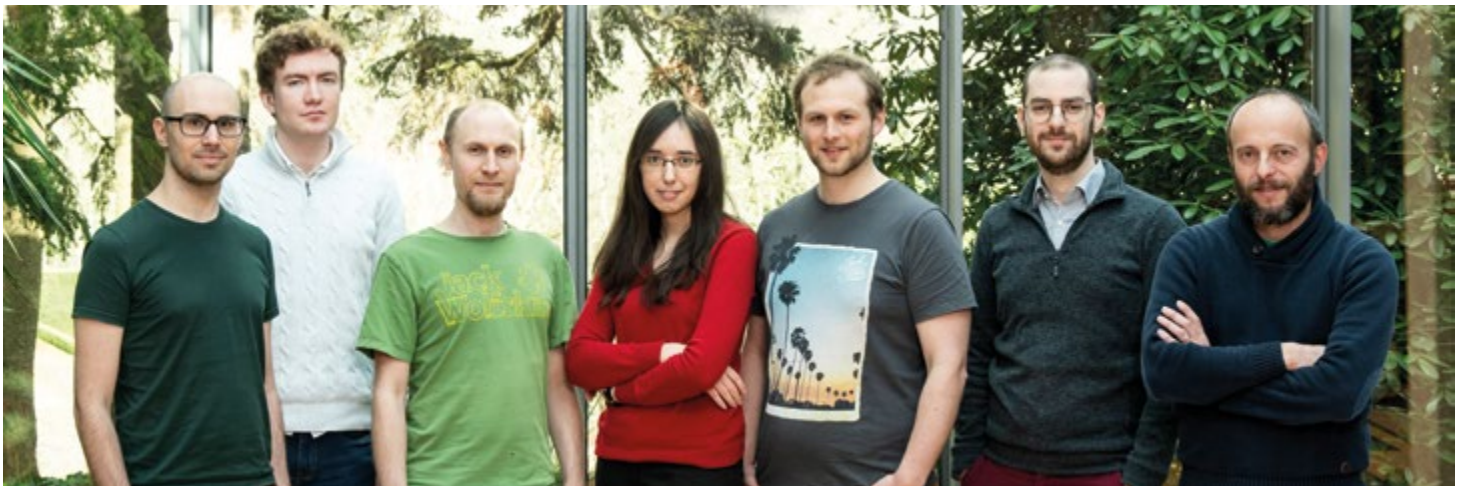


Figure 9: A Aufbau and non-aufbau molecular orbital configurations. B Polychlorotriphenylmethyl radical with a donor substituent. C Schematic open-shell single-molecule junction.



2 Research

2.3 Computational Molecular Evolution (CME)



Group Leader
Prof. Dr. Alexandros Stamatakis

Staff members
Pierre Barbera
Benjamin Bettisworth
Lucas Czech (until June 2019)
Dr. Alexey Kozlov (staff scientist)
Benoit Morel

Scholarship holder
Sarah Lutteropp (HITS Scholarship)

Visiting scientist
Aggelos Koropoulis (until January 2019)

Students
Ivo Baar
Rudolf Biczok (until January 2019)
Paula Breitling (until June 2019)
Johanna Wegmann (until November 2019)
Adrian Zapletal (until November 2019)
Dimitri Höhler (since December 2019)
Lukas Hübner (since December 2019)

The Computational Molecular Evolution 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
- discrete algorithms on trees
- population genetics.

In the following section, we outline our current research activities, which lie at the interface(s) between computer science, biology, and bioinformatics. The overall goal of the group 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."

What happend in the lab in 2019?

In the winter of 2018/2019, 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://cme.h-its.org/exelixis/web/teaching/courseEvaluations/Winter18_19.pdf). Enrollment in our course at KIT further increased over the winter semester of 2019/20 to over 40 students.

During the summer semester of 2018, we again taught our main seminar, "Hot Topics in Bioinformatics." Rudolf Biczok, Paula Breitling, and Johanna Wegmann all successfully defended their master's theses at the Department of Computer Science at KIT. For the first time ever since its establishment as an Emmy-Noether junior research group in 2008, more female than male students graduated from the lab.

In 2019, a total of three KIT master's students joined the lab either as student programmers or to work on their master's theses.
One highlight of 2019 was the summer school on Computational Molecular Evolution, for which Alexis again served as co-organizer. The course took place for the 11th time this year at the European Bioinformatics Institute in Hinxton, UK. CME lab members Lucas Czech and Benoit Morel contributed substantially to the success of the course as teaching assistants (see Chapter 5.1.6).
Alexis was listed on the Clarivate Analytics highly cited researchers list for the fourth year in a row as well as for the second consecutive year on the new cross-field category, which comprises researchers with a focus on interdisciplinary research (see Chapter 9.5). An independent bibliometric study [J.P.A. Ioannidis et al. A standardized citation metrics author database annotated for scientific field (2019), PLOS Biology, 17(8): e3000384] conducted in 2019 by researchers at Stanford using additional metrics revealed that Alexis is among the top 0.01% of scientists in his field based on his impact.

Finally, our substantial and long-term efforts to re-engineer and re-write some of our most widely used software tools were completed with the publications of Modeltest-NG [Darriba, 2019] and RAxML-NG in Molecular Biology and Evolution and Bioinformatics, respectively (corresponding preprints were mentioned in previous annual reports).

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

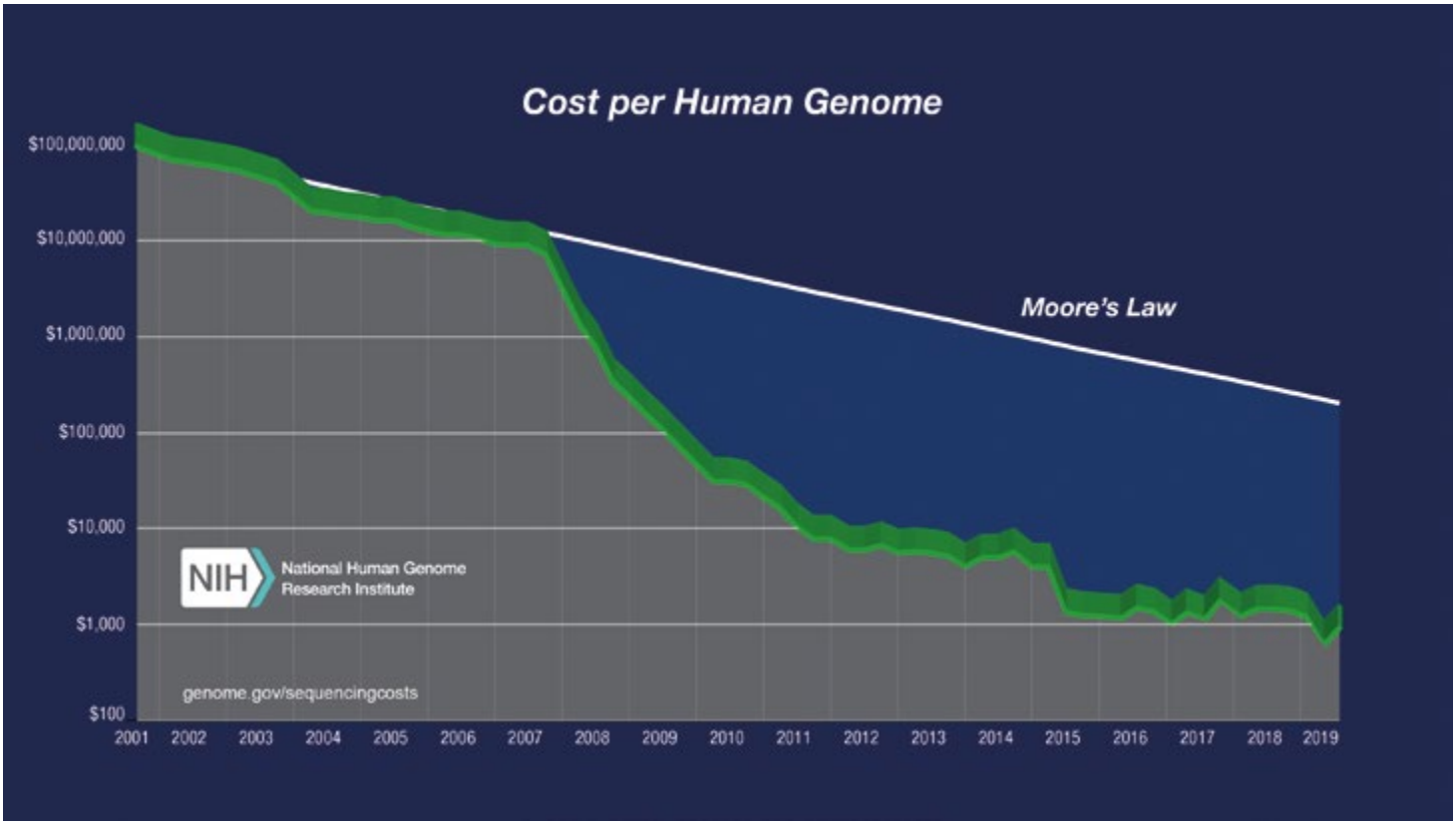


Figure 10: "Cost per Genome": The cost of sequencing a genome is decreasing at a faster rate than is the cost of computation (source: National Human Genome Research Institute (NHGRI).

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 Figure 10).

We are thus faced with 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.

To address the complexity of evolution, we have recently begun investigating the properties of so-called phylogenetic networks – that is, we are considering more complex and realistic evolutionary scenarios that cannot be captured via a simple binary tree.

Another relatively novel line of research in this area is that of gene tree/species tree reconciliation. In this field, we intend to take into account discordances between the phylogenies of individual genes or gene families and the underlying species tree that are driven by evolutionary phenomena such as gene duplication, gene loss, and lateral gene transfer.
Another challenge to 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. One question that arises here is how and whether existing methods can be used or need to be adapted to work on longer sequence reads. We investigated this issue in the context of a taxonomic and phylogenetic analysis of eukaryotic diversity [Jamy, 2019] in collaboration with Jamy Mahwash, a former participant in our 2018 summer school on computational molecular evolution (Mahwash and Alexis can be seen here: https://www.youtube.com/watch?v=yjzwqzaLUX-A&feature=youtu.be in a local news report about the 2018 summer school).

Overall, phylogenetic trees (evolutionary histories of species) and the application of evolutionary concepts in general 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 past climatic changes, and to analyze microbial diversity in the human gut.

GeneRax – A tool for gene tree inference under gene duplication, loss, and transfer that takes into account the species tree

We have recently become interested in developing so-called 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, which is particularly common, for instance, in bacterial evolution.

In general, individual gene trees are frequently not identical to the species tree due to the presence of the aforementioned evolutionary events. The main goal of gene tree/species tree reconciliation approaches is thus to explain and resolve these differences between the gene trees and the species trees using appropriate likelihood-based models.
One of the reasons for focusing on this problem is that most available reconciliation tools are neither efficient nor user-friendly, nor are they particularly scalable. In other words, great potential for making substantial contributions to this area exists.
In 2018, we initially addressed and solved the problem of efficiently inferring gene trees in large compute clusters, which constitutes a necessary pre-processing step. In 2019, we

focused on correcting the inferred gene trees given a fixed, externally provided species tree. Next year, we also plan to work on methods to improve or infer the species tree.
In 2019, we developed and released an open-source code called GeneRax [Morel, 2019]. This code represents the first comprehensive maximum likelihood-based- and species tree-aware gene tree inference software. It simultaneously accounts for substitutions at the sequence level via the so-called phylogenetic likelihood component and gene level events, such as duplication, transfer, and loss via the so-called reconciliation likelihood component. In other words, we optimized gene family tree topologies with respect to a composite reconciliation and phylogenetic likelihood. GeneRax can infer rooted gene trees on an almost arbitrarily large number of multiple sequence alignments for gene families and a given rooted species tree. A gene family is a set of sequences from similar genes with respect to function and sequence similarity and can be generated via the aforementioned evolutionary process of gene duplication.

We revealed that in comparison with competing tools on simulated data, GeneRax infers trees that are the closest to the true tree in 90% of simulations with respect to the standard topological distance measure in phylogenetics.
We also demonstrated that for empirical datasets, GeneRax is the fastest among all tested methods when beginning with the plain set of multiple sequence alignments for gene families. Since we cannot rely on any ground truth for empirical data because the true history of gene duplication-, loss-, and transfer events is unknown, we revealed that GeneRax is able to infer trees with the highest likelihood score under our reconciliation model. In other words, we demonstrated that the heuristics that we developed are "good" at optimizing the composite phylogenetic and reconciliation-likelihood score. GeneRax also implements an elaborate parallelization scheme. In a showcase analysis, we were able to demonstrate that GeneRax only required 8 minutes to complete gene tree inferences and reconciliations for 1,099 Cyanobacterial gene families comprising between 37 and 130 sequences on 512 CPU cores on the HITS cluster. Apart from the short time-to-completion, we also demonstrated that GeneRax exhibits "good" parallel scalability (see Figure 11). GeneRax is available as an open source code under GNU GPL at https://github.com/BenoitMorel/GeneRax. It is worth emphasizing that GeneRax constitutes the first one-stop shop for gene tree reconciliation as – unlike all competing tools – it does not require any external programs for pre-processing. We believe that this substantial usability improvement – coupled with its high parallel efficiency and good accuracy will render GeneRax a popular tool.

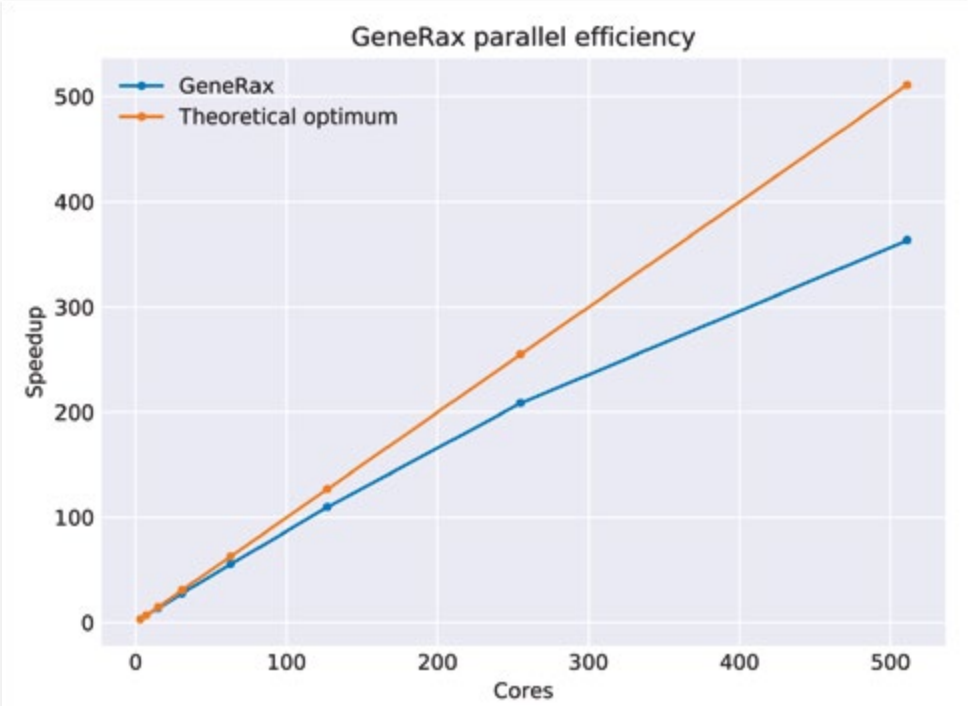


Figure 11: Parallel speedup of GeneRax on the empirical Cyanobacteria dataset comprising 1,099 gene families and using between 4 and up to 512 cores.

Efficient computation of the transfer bootstrap

The phylogenetic bootstrap procedure for obtaining branch support values on a phylogenetic tree has been known since the 1980s. In the classic setting, there is a reference tree and a set of bootstrapped (BS) trees. To calculate BS support, it is necessary to calculate the frequency of occurrence of splits (bipartitions of the set of species contained in the reference tree by cutting the tree at an inner branch) that are induced by the reference tree in the set of BS trees. In 2018, the so-called transfer BS expectation (TBE) method was introduced [F. Lemoine, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data, (2018) Nature 556:452-456].

Not only does this new procedure for calculating BS support take into account the presence/absence of a bipartition of the reference tree in a BS tree, but bipartitions that are similar to the reference bipartition also contribute to the TBE support value. We believe that the TBE method constitutes a reasonable adaptation of BS support calculations, particularly for trees that comprise hundreds to

thousands of species that typically exhibit a weak phylogenetic signal. However, the computation of the TBE is substantially more compute-intensive than is the classic BS. We thus set forth to devise an algorithmically and technically optimized implementation of the TBE in a classic algorithm engineering project [Lutteropp, 2019]. For empirical as well as random tree sets with varying species counts, our implementation is up to 480 times faster than the original implementation (called Booster) provided in the aforementioned Nature paper (see Figure 12). Furthermore, our implementation only requires memory that is linear in the number of species in the phylogeny, which leads to 10- and up to 40-fold memory savings. Finally, we have already integrated our efficient TBE implementation into our flagship tool, RAxML-NG.

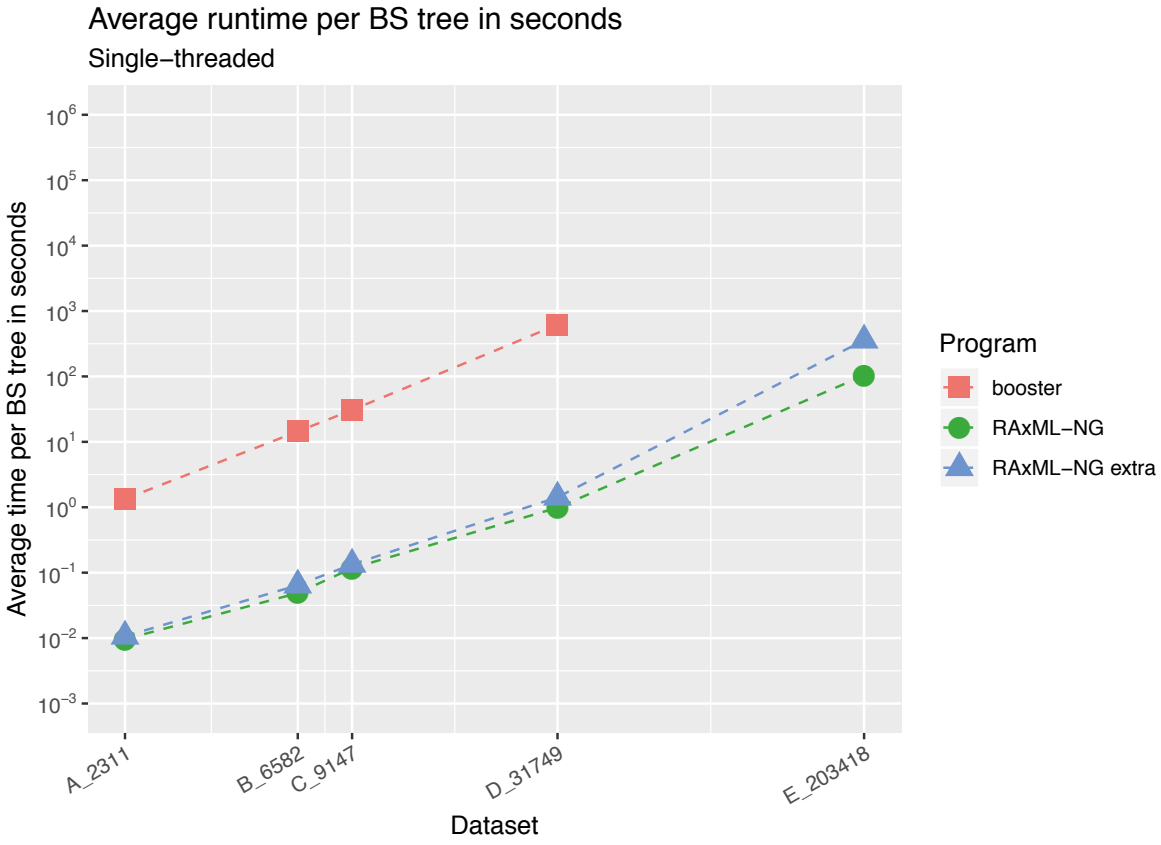


Figure 12: Average sequential runtime per BS tree in seconds with and without computing additional statistics (RAxML-NG extra/RAxML-NG) over increasing tree sizes. On the E_203418 dataset, Booster ran out of memory.

Softwipe: A scientific software quality-assessment tool and benchmark

In an interview with a journalist on our initial work on software quality in the area of evolutionary biology ("The State of Software for Evolutionary Biology", Molecular Biology and Evolution, 35(5):1037-1046, 2018), we were asked if there is a tool available for conducting our software quality analyses in a more systematic and automated manner.

As such a tool did not yet exist, we decided to begin developing such a framework and an appropriate benchmark that includes scientific software from several application areas via the help of our student programmer Adrian Zapletal (a former participant in all of our courses, practicals, and seminars at KIT) and Carsten Sinz (KIT CS faculty, expert on code verification).
Our software quality ranking system computes relative ranks for a plethora of code-quality indicators (e.g., systematic usage of assertions, memory behavior, compiler warnings, cyclomatic complexity, etc.) and subsequently simply computes the average over these ranks to derive a global ranking. The code and benchmark are available at https://github.com/adrianzap/softwipe/wiki/Code-Quality-Benchmark. A screenshot of the current ranking is provided in Figure 13.

program	overall	compiler_and_sanitizer	assertions	cppcheck	clang_tid
genesis 0.22.1	9.0	8.0	10.0	8.2	9.4
hyperphylo	8.8	8.5	10.0	8.8	8.6
kahypar	8.7	6.7	10.0	8.5	9.5
repeatscounter	7.9	9.2	0.0	10.0	10.0
raxml-ng 0.8.1	7.3	9.5	4.3	6.0	9.4
dawg 1.2	7.0	6.6	2.3	7.5	10.0
swarm 3.0.0	6.4	10.0	0.3	9.6	0.2
treerecs 1.0	6.3	7.5	1.8	6.5	6.8
samtools 1.9	5.7	7.7	1.3	7.0	6.2
seq-gen 1.3.4	5.7	7.7	0.0	6.8	8.3
sf	5.7	6.9	1.1	4.7	8.3
iqtree 1.6.10	5.2	0.0	4.3	4.0	3.9
vsearch 2.13.4	4.9	4.3	0.0	8.3	0.0
minimap 2.17-r943	4.7	3.1	2.6	5.3	2.2
clustal omega 1.2.4	4.6	5.6	3.1	6.7	4.5
phyml 3.3.20190321	4.6	8.9	4.9	4.6	0.7
prank 0.170427	4.6	4.3	5.0	0.0	7.3
tcoffee	4.4	4.8	0.0	5.7	8.2
gadget 2	4.2	7.4	0.0	6.9	4.9
ms 7.429	4.2	6.5	0.0	0.0	5.1
mrbayes 3.2.6	4.1	9.3	1.5	8.2	7.3
bpp 3.4	3.7	8.7	0.0	1.7	7.4
cellcoal 1.0.0	3.6	9.2	0.0	6.1	0.0
mafft	3.4	8.2	0.0	6.5	5.1
athena	3.3	4.0	0.0	0.0	3.9
indelible 1.03	1.8	0.0	0.0	0.3	2.7

Figure 13: Softwipe software quality ranking of various scientific software tools from the areas of Bioinformatics, Astrophysics, and Computer Science.

Die Gruppe rechnerbasierte Molekulare Evolution (CME) 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
- Diskrete Algorithmen auf Bäumen
- Methoden der Populationsgenetik.

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".

2 Research

2.4 Computational Statistics (CST)

Group Leader
Prof. Dr. Tilmann Gneiting

Staff members
Dr. Timo Dimitriadis (since June 2019)
Kira Feldmann (until February 2019)
Dr. Sebastian Lerch (until June 2019)
Johannes Resin

Scholarship holder
Patrick Schmidt (HITS scholarship)

Visiting scientists
Prof. Dr. Sándor Baran (July 2019)
Dr. Sebastian Lerch (since July 2019)
Peter Vogel (until February 2019)
Eva-Maria Walz (since March 2019)

The Computational Statistics group at HITS was established in November 2013, when Tilmann Gneiting was appointed as Group Leader in addition to Professor of Computational Statistics at the Karlsruhe Institute of Technology (KIT). The group's research focuses on the theory and practice of forecasting.

As the future is uncertain, forecasts should be probabilistic in nature, which means that they should take the form of probability distributions over future quantities or events. Accordingly, we have recently been 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 relationships with meteorologists at KIT and at the European Centre for Medium-Range Weather Forecasts.

Figure 14: Participants of the mini symposium on statistical postprocessing of ensemble weather forecasts held on 29 July 2019 at HITS.

General news

The focus of our research remains on the theory and practice of forecasting. While our work is deeply rooted in probability theory and mathematical statistics, we maintain close contacts to applied communities in areas including economics, epidemiology, meteorology, and seismology. Our interdisciplinary collaborations involve joint work with meteorologists at the European Centre for Medium-Range Weather Forecasts (ECMWF) in Reading in the United Kingdom and within the collaborative research center "Waves to Weather," which is shared by the Karlsruhe Institute of Technology (KIT), the University of Mainz, and the University of Munich. We are delighted to report that the German Research Foundation (Deutsche Forschungsgemeinschaft) has renewed funding for "Waves to Weather" (SFB/TRR 165) for a further four-year period through June 2023.

The ties to our home university of KIT remain strong. Various CST team members have taught lecture- and problem-classes on topics in statistics and machine learning within the KIT Faculty of Mathematics, where we are affiliated with the Institute for Stochastics. In September, the Faculty of Mathematics hosted the Annual Meeting of the German Mathematical Society (Deutsche Mathematikervereinigung, DMV) in its newly renovated, award-winning building on KIT's Campus South. At the conference, CST Group Leader Tilmann Gneiting presented a public lecture on the mathematical aspects of weather prediction. In October, we welcomed CST alumnus Fabian Krüger to his new position as a tenure-track professor in the Faculty of Economics at KIT.

In keeping with the time-honored tradition, an integral facet of our work involves intense disciplinary and interdisciplinary scientific exchange on various occasions. Throughout 2019, we were again happy to welcome scientific guests from all over the world. A mini symposium on the topic of the statistical postprocessing of ensemble weather forecasts held on 29 July on the HITS premises served as a highlight during the year. Figure 14 shows the participants, who discussed methodological advances that lie at the interface of meteorology and statistics, at the main entrance to the HITS building.

Tilmann Gneiting was invited by the International Institute of Forecasters (IIF), the preeminent international organization devoted to the science of forecasting, to present a keynote lecture at the International Symposium on Forecasting (ISF) 2019 in Thessaloniki, Greece, and to serve as instructor of the second iteration of the associated IIF Summer School. The IIF Summer School is a two-day course that precedes the symposium and provides an in-depth analysis of a cutting-edge topic in forecasting from one of the ISF invited speakers. The IIF Summer School was held from 15–16 June with the theme of "Probabilistic Forecasting" and included seven lectures on topics ranging from mathematical foundations and software solutions to case studies in economics and meteorology, all of which were supplemented by practical sessions under the direction of Johannes Resin. The Summer School participants constituted a highly engaged and motivated interdisciplinary audience of 36 PhD students, postdocs, and junior faculty from all

Figure 15: Participants and instructors from the IIF Summer School on "Probabilistic Forecasting," held from 15–16 June 2019 in Thessaloniki, Greece (picture: Monash University, Victoria, Australia)

over the world, several of whom had received travel grants from the IIF. Figure 15 shows the Summer School students and instructors enjoying a mid-afternoon break. On 19 June, Tilmann Gneiting presented a plenary lecture on the topic of receiver operating characteristic (ROC) curves and their uses in forecast evaluation, as illustrated in Figure 16. The lecture was based on joint work with Peter Vogel, who pursued PhD studies within the "Waves to Weather" collaborative research center and defended his PhD thesis at KIT in early 2019.

Forecast evaluation with proper scoring rules

As noted, forecasts should be probabilistic in nature, meaning that they should take the form of probability distributions over future quantities or events. Accordingly, we have recently been witnessing a shift of paradigms to probabilistic forecasts in various fields, including meteorology, hydrology, economics, and demography. In typical practice, a plethora of models and data sources can be used to generate probabilistic forecasts, and the evaluation and selection among competing forecasting methods is hence critically important. In this context, scoring rules serve as the major workhorses in the quantification of predictive performance. In a nutshell, a scoring rule assigns a numerical penalty based on the probabilistic forecast and the event or value that materializes. Scores are then averaged over forecast cases, and the forecast technique with the lowest mean score is deemed the best. A critically important requirement for a scoring rule is that it be "proper" in the sense that a forecaster should minimize the expected penalty by issuing a forecast that is in line with his or her best judgment [Gneiting T, Raftery AE. Strictly proper scoring rules, prediction and estimation. Journal of the American Statistical Association (2007) 102:359-378]. Proper scoring rules have also been claimed to provide a "truth serum," and this guiding principle is referred to as the "proper scoring paradigm." Figure 17 illustrates the proper scoring rules that are most commonly used in practice, namely the continuous ranked probability score (CRPS) and the logarithmic score (LogS).

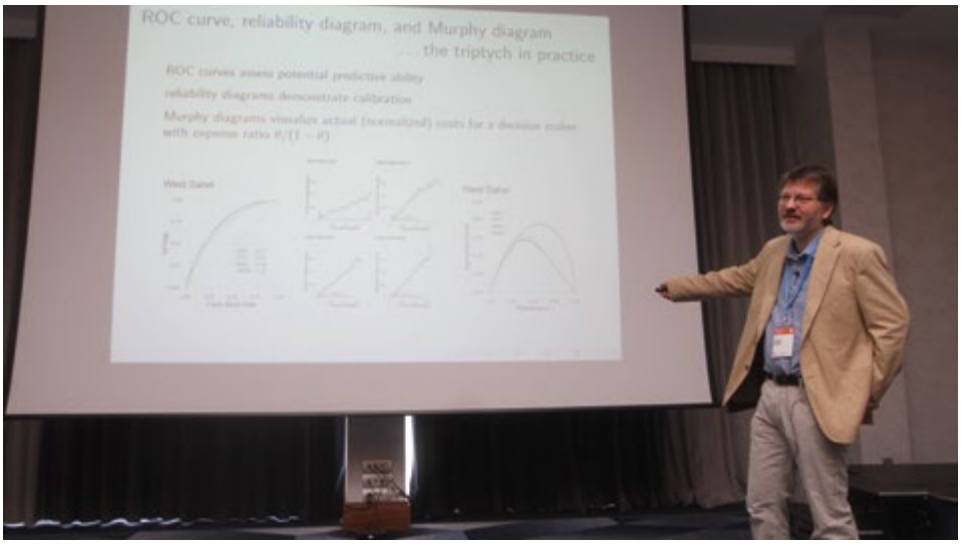


Figure 16: Group Leader Tilmann Gneiting presents a keynote lecture on receiver operating characteristic (ROC) curves at the International Symposium on Forecasting (ISF) on 19 June 2019 (picture: Monash University, Victoria, Australia)

The CST group has spent many years developing a theory, methodology, and software solutions for proper scoring rules. A paper by Sebastian Lerch together with HITS alumni Alexander Jordan and Fabian Krüger [Jordan, Krüger, and Lerch, 2019] describes the group's scoringRules package within the R environment for statistical computing and graphics, which was designed as a library for computing proper scoring rules under a comprehensive range of forecast distributions that cover many applied settings. In contrast to existing software, a key novelty of the scoringRules package is its extensive coverage of the continuous ranked probability score (CRPS), which is attractive for both practical and theoretical reasons. The CRPS is reported in the same unit as the predictand – for example, in the unit of degrees Celsius for temperature forecasts – and can be interpreted as a generalization of the classical absolute error (AE) measure. However, the more widespread use of the CRPS has been hampered by computational challenges. In deriving closed-form solutions and providing efficient numerical implementations, the scoringRules package enables its routine use in applied work.

In sum, the paper discusses implementation- and usage details for the package, presents case studies from meteorology and economics, and reviews pertinent background literature. In a joint paper with Jonas Brehmer at the nearby University of Mannheim [Brehmer and Gneiting, 2019], we demonstrate that subject to customary regularity conditions, any scoring rule can be made proper by applying an extant, very general construction that rests on optimal strategies, typically called Bayes acts. We refer to this principle as "properization." In a nutshell, properization proceeds as follows:

- Given any scoring rule and the forecast distribution at hand, compute the Bayes act. The Bayes act is the probabilistic forecast that minimizes the expected penalty under the assumption that the forecast distribution at hand is valid
- Compute the "properized" score as the original score applied to the Bayes act.

The construction is very general and

depends only on the existence of Bayes acts, for which we supply theoretical guarantees. In the paper, we formulate sufficient conditions under which Bayes acts exist and scoring rules can be made proper. The paper also discusses examples from the recent literature and applies the construction to create new types – and reinterpret existing forms – of proper scoring rules. As it turns out, the technique has found widespread, implicit use in the transdisciplinary literature on proper scoring rules, where our unified approach yields simplified, shorter, and considerably more instructive and transparent arguments than do extant methods. We anticipate additional, important uses of properization in a wide range of applied settings in which scoring rules need to be tailored to the specific needs of forecast users.

How accurate are state-of-the-art temperature forecasts?

In a joint paper with the European Centre for Medium-Range Weather Forecasts (ECMWF), we investigated the accuracy of probabilistic temperature forecasts from the Centre's world-leading ensemble system [Feldmann, Richardson, and Gneiting, 2019]. Based on our work, forecasts of surface temperature one day in advance are accurate to about one degree Celsius!

To provide a background, despite their undisputed successes, ensemble forecasts from numerical weather prediction (NWP) models continue to be subject to systematic deficiencies. Typically, they are biased relative to surface-weather observations, and the ensemble output underestimates the uncertainty inherent to the forecast. Statistical postprocessing aims to compensate for these shortcomings by calibrating the ensemble output. Any application and evaluation of postprocessing methods relies on the availability of training and verification data. The choice of which data to use is of critical importance, and a fundamental decision must be made regarding whether to use gridded data or station-based data. Weather observation stations are scattered across the globe and cluster in more densely populated and more developed parts of the world, whereas coverage is sparse over the oceans, in polar regions, and over large parts of Africa. In contrast, so-called analyses combine weather observations with past forecasts to provide a retrospective, gridded best estimate of the state of the atmosphere with full global coverage.

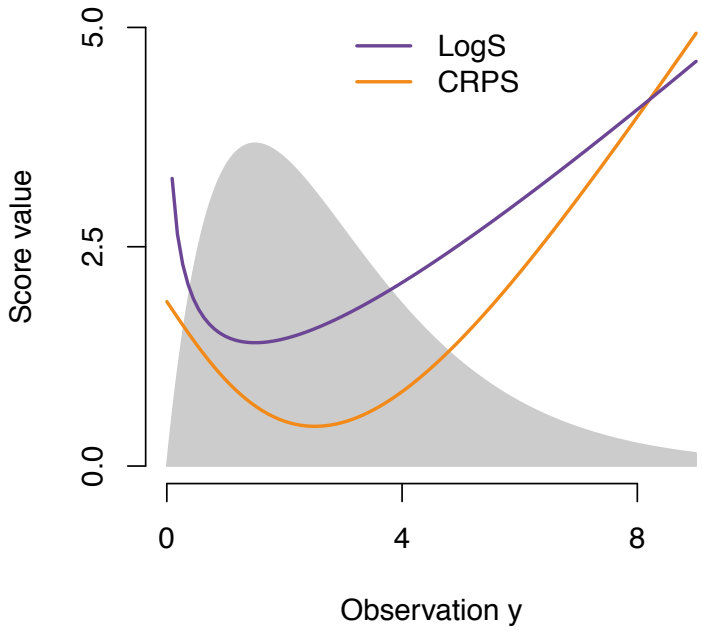


Figure 17: Continuous ranked probability score (CRPS) and logarithmic score (LogS) when the probabilistic forecast is a gamma distribution. A scaled version of the predictive density is shown in gray. Source: Jordan, Krüger, and Lerch (2019)

In our paper, we considered forecasts of surface temperatures issued across the world between 1 November 2016 and 7 December 2017 that were valid at forecast lead times from between 1 and 15 days.

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 trans-disziplinären 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/-innen am KIT und am Europäischen Zentrum für mittelfristige Wettervorhersagen.

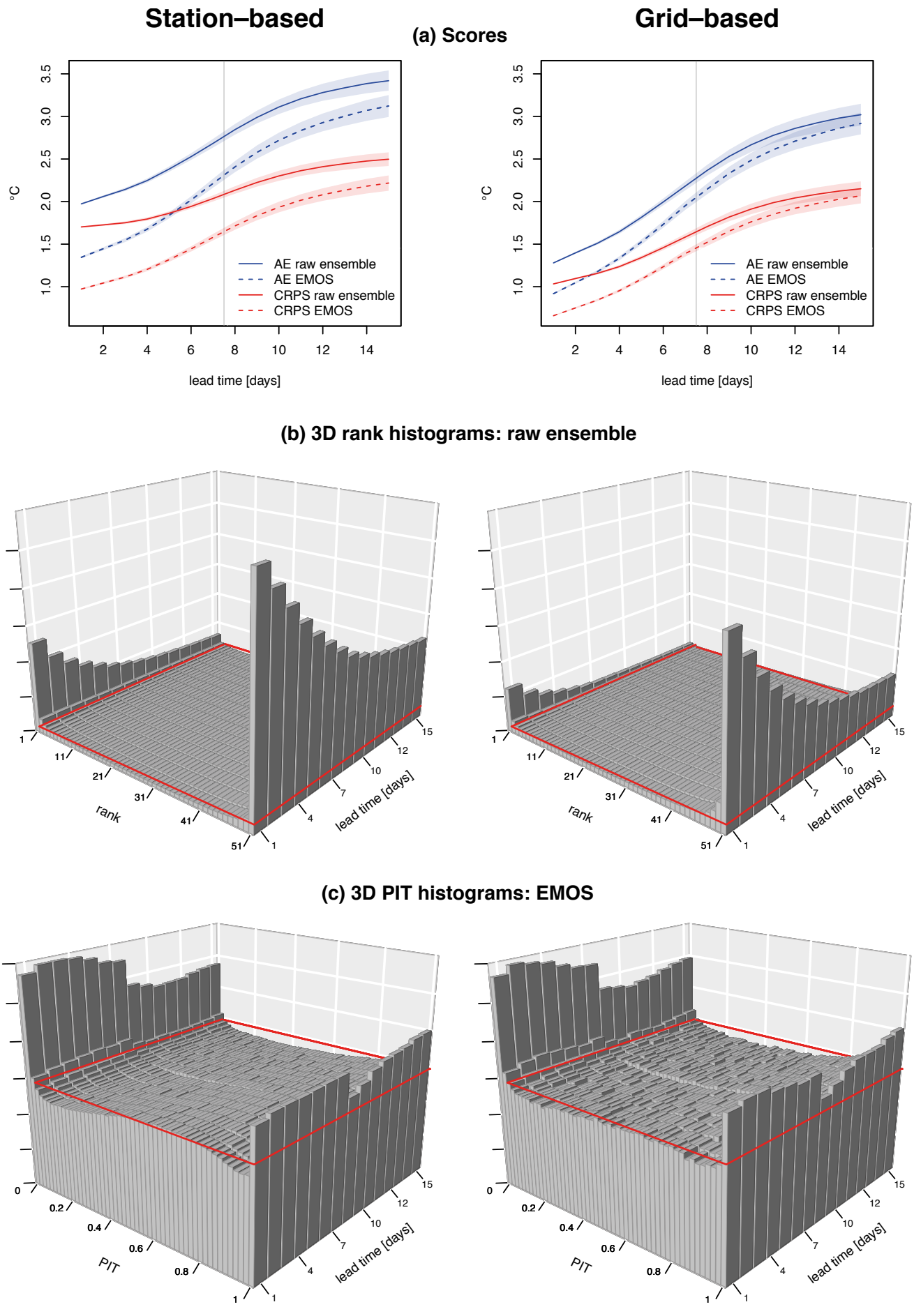


Figure 18: Comparison of station-based forecasts of surface temperature to matched grid-based forecasts in terms of (a) CRPS and absolute error (AE), (b) rank histograms for the ECMWF raw ensemble, and (c) PIT histograms for postprocessed forecasts at lead times between 1 and 15 days. The thick red lines in the three-dimensional histograms correspond to uniformity. Source: Feldmann, Richardson, and Gneiting (2019).

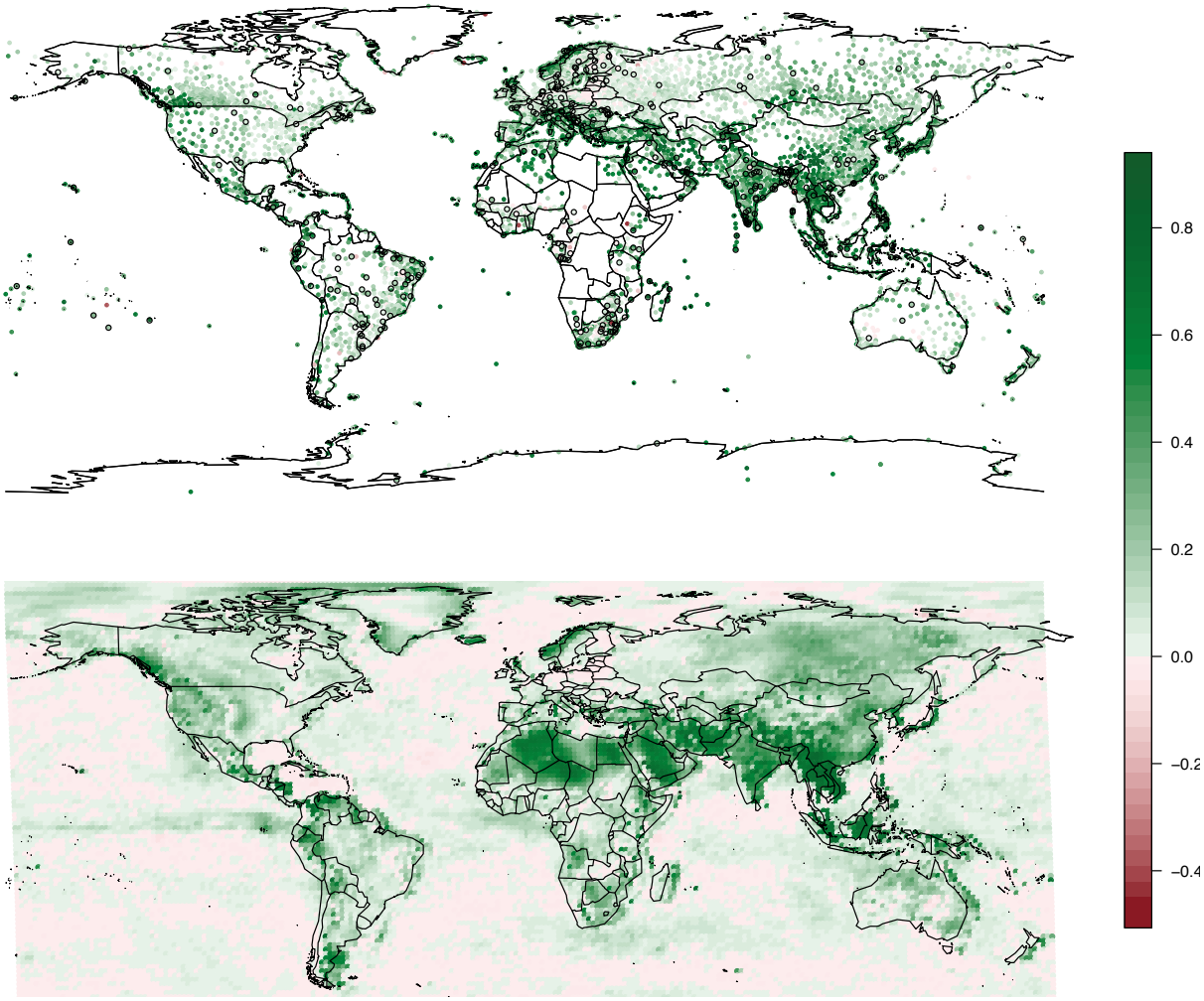


Figure 19: CRPS skill for station-based (top panel) and grid-based (bottom panel) postprocessed forecasts of surface temperature relative to the raw ECMWF ensemble at a lead time of three days. Source: Feldmann, Richardson, and Gneiting (2019).

For postprocessing, we employed the ensemble model output statistics (EMOS) approach developed by Gneiting et al. [Gneiting T, Raftery AE, Westveld AH, Goldman T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Review (2005) 133:1098--1118], which uses training data from a rolling temporal window. To enable meaningful comparisons of grid- versus station-based approaches, we matched every single available forecast case at a weather station to a forecast case at a grid point.

This comparison is illustrated and detailed in Figure 18. A first, crucial insight from panel (a) is that statistical postprocessing improves the raw ensemble forecast. At all lead times, the benefits of postprocessing are greater when it is performed on station data as opposed to when using the gridded analysis. At one day in advance, the raw ensemble forecast shows a mean CRPS and mean absolute error (AE) of 1.03 and 1.28 degrees Celsius on the grid and of 1.72 and 1.97 degrees Celsius at the stations, respectively. Postprocessing with EMOS reduces these numbers to 0.66 and 0.92 degrees Celsius on the grid and to 0.97 and 1.35 degrees Celsius at the stations, respectively. The fact that mean CRPS values below a single degree Celsius were achieved both on the grid and at the stations is remarkable and indicative of the vast improvements in NWP.

Panels (b) and (c) show probability integral transform (PIT) and verification rank histograms. For calibrated, statistically valid probabilistic forecasts, these histograms are uniform, whereas U-shapes indicate underestimation of the uncertainty that is inherent to the forecast. To simultaneously accommodate all lead times from 1 to 15 days, we employed three-dimensional versions of the histograms. The rank histograms in panel (b) reveal that the raw ECMWF ensemble underestimates uncertainty, particularly at smaller lead times, with the underdispersion being more pronounced in the station-based approach. While the PIT histograms for the EMOS postprocessed forecasts in panel (c) are much closer to uniformity, a small degree of underdispersion remains.

For an analysis of any spatial patterns in the predictive performance, Figure 19 shows the global distribution of the CRPS skill score for EMOS postprocessed forecasts relative to the raw ECMWF ensemble, at an exemplary lead time of three days. Positive and negative skill correspond to better performance for the postprocessed forecasts and for the raw ensemble, respectively. The top and bottom panels concern the station-based approach and the grid-based approach, respectively. Overall, postprocessing has a thoroughly positive effect, with skill scores that are overwhelmingly positive. The benefits are strongest along the west coast of the Americas and Scandinavia and in tropical and subtropical areas, such as Northern Africa, the Arabian Peninsula, India, Southeast Asia, and Japan.

2 Research

2.5 Data Mining and Uncertainty Quantification (DMQ)

Group Leader
Prof. Dr. Vincent Heuveline

Scholarship holders
Vijayasarathi Janardhanam (until April 2019)
Alejandra Jayme

Staff members
Dr. Chen Song
Philipp Lösel (since October 2019)
Philipp Gerstner

Visiting scientists
Saskia Haupt
Sotirios Nikas
Jonas Kratzke (until March 2019)
Nils Schween (until May 2019)
Prof. Dr. Jiawei Zhang

Students
Charlotte Boys
Jonas Roller

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 Dr. Vincent Heuveline.

DMQ's research focus lies in gaining knowledge from extremely large and complex datasets through data-mining techniques. 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 2019, DMQ focused on research activities in the areas of mathematical oncology, machine learning for healthcare and power system applications, as well as energy-aware programming.

Mathematical oncology – how mathematics can help in the fight against cancer

Cancer is one of the leading causes of disease-related deaths worldwide and is caused by alterations of the genome, a highly complex carrier of genetic information. Cancer can thus develop through many different molecular ways, thereby rendering its clinical management extremely complex.

Although large amounts of data on cancer genetics and molecular characteristics are currently available and growing exponentially, adequately interpreting these data remains a major challenge.

By building mathematical models, we are able to gain insights into complex processes. We can simultaneously analyze different biological concepts and test various medical hypotheses about cancer evolution and cancer immunology by changing the parameters and initial conditions in our mathematical models. Based on existing clinical and molecular data, we can model the effectiveness of current clinical approaches. Furthermore, we can use the new findings to guide the development of innovative clinical procedures for the diagnosis, prevention, and treatment of cancer. In so doing, it is possible to efficiently translate novel basic science insights into clinical management procedures while decreasing the amount of time- and resource-consuming empirical research and minimizing the burden imposed by animal experiments and clinical trials (see Figure 20).

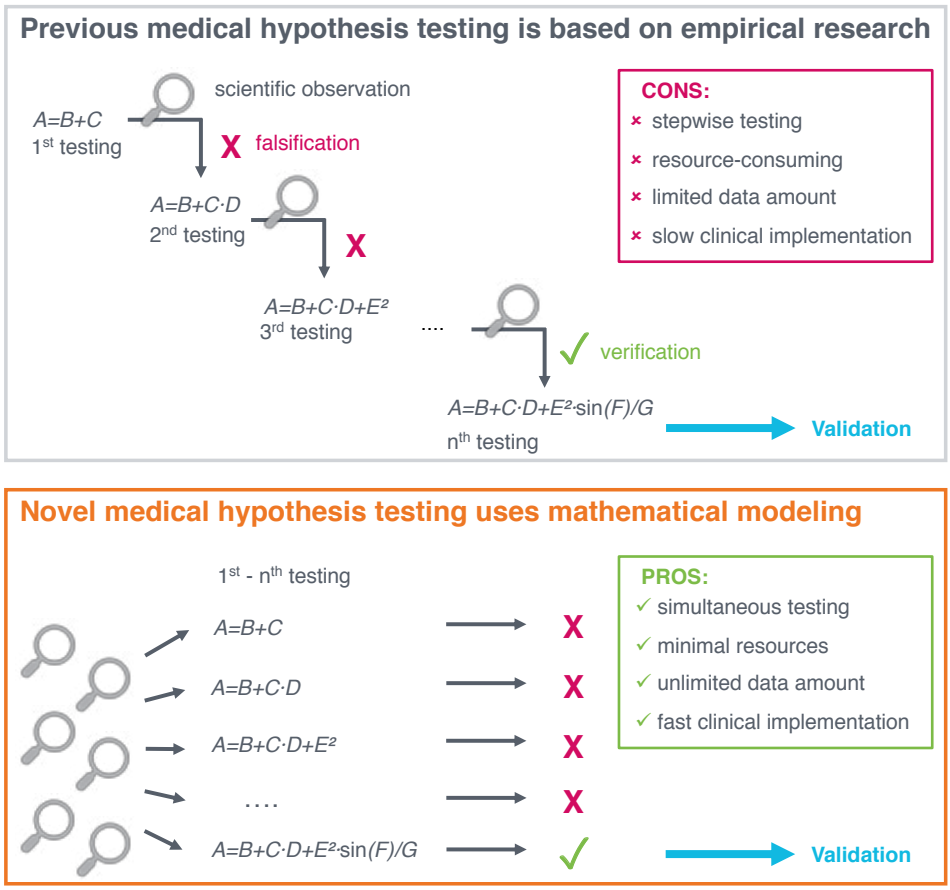


Figure 20: The use of mathematical modeling for medical hypothesis testing leads to a fast and cost-efficient clinical implementation of the novel research results.

DMQ / EMCL works in close collaboration with the Department of Applied Tumor Biology, University Hospital Heidelberg, and aims to establish mathematical models of cancer initiation, evolution, and immunology by focusing on hereditary forms of cancer, which are responsible for 5–10% of the worldwide tumor burden and tremendously increase the lifetime cancer risk of affected individuals. Hereditary cancers are an ideal case for mathematical modeling as they reflect the general principles of cancer evolution but have a known cause of initiation and enable multiple tumors to be studied in parallel. Lynch syndrome (LS) is the most common inherited cancer syndrome yet remains largely underdiagnosed. LS predisposes affected individuals to developing cancer in the large bowel and other organs and can only be partially prevented, even if the syndrome has been diagnosed. A strong medical need for better diagnosis, prevention, and treatment of LS thereby exists.

In [Bläker et al., 2019], we evaluated two molecular diagnostic approaches to LS. Cost-effectiveness and accuracy analyses revealed for the first time that diagnostic procedures should be tailored to patients' age. By adapting the current clinical procedure accordingly,

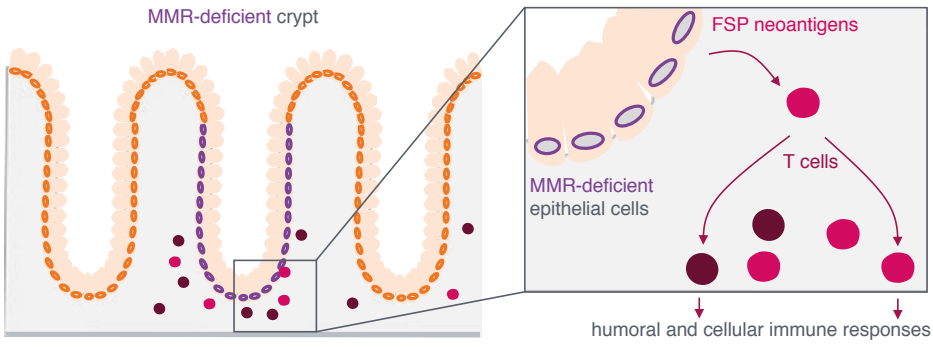


Figure 21: In LS patients, altered crypts in the human large bowel can generate neoantigens, even before a clinically manifest tumor develops. Strong immune responses (T-cell activity) have been detected in healthy LS patients.

both the risk of overlooking LS patients and the costs of the clinical procedure can be significantly decreased.
When designing novel prevention approaches, including cancer-preventive vaccination, it is crucial to understand the interaction between the immune system and emerging precancerous cells. In [Ballhausen et al., 2019], we used bioinformatics and data analysis techniques to predict a set of alterations in the genome of LS individuals with a likely driver function during cancer evolution. Furthermore, these mutations were found to trigger strong immune responses, which can even be detected in tumor-free LS individuals (see Figure 21). This information will be used to design innovative approaches to cancer prevention in the framework of LS in order to provide a proof of concept that cancer prevention by vaccination is feasible.

Machine learning for predicting employee absences

To remain competitive and effective in the market, companies seek to reduce costs and maximize profit. Employee absenteeism – whether justified or not – is a crucial factor in reaching this goal. Within the KIPROSPER project, we study this issue by using machine learning techniques to gain insight on employee absences. Specifically, we aim to predict absences given employee data such as demographics, medical and behavior history, self-reported job satisfaction, and work-structure information. Furthermore, we analyze which of these employee data are significant factors in this prediction.

The task is treated as both a regression problem and a classification problem. Regression and classification approximate a mapping function from input, independent variables (features) to output, dependent variables targets, while their difference lies in the nature of the output target variable.
Regression models provide a numerical and continuous value for an employee's predicted absences (e.g., in hours or days), whereas classification models categorize employees with labels (e.g., predicted absences of 2–5 days, or medium risk). Regression output is important for customized or individual employee care, whereas management may be more interested in identifying potentially critical groups.

In this work, four machine learning models are studied: Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN). With the exception of Linear Regression, all models can perform both regression and classification. Linear Regression assumes a linear relationship between features and targets. The linear function parameters are estimated from training data by minimizing the residual sum of square errors between the observed targets and predicted targets. Support Vector Machine aims to construct a hyperplane or a set of hyperplanes to be used to perform the required task. In regression, the hyperplane should be within acceptable distance of all training data points. In classification, the hyperplane separates the different classes with as much distance as possible. Random Forest constructs a multitude of decision trees that are trained on the input data. The output is the mean prediction for regression and the mode label for classification. Random Forest minimizes overfitting by using bootstrap aggregation of the training data and a random subset of features for each tree. Artificial Neural Networks are composed of layers of connected nodes through which information travels from the input layer (the first layer) to the output layer (the last layer) via one or more (hidden) layers. The connections between nodes have weights that

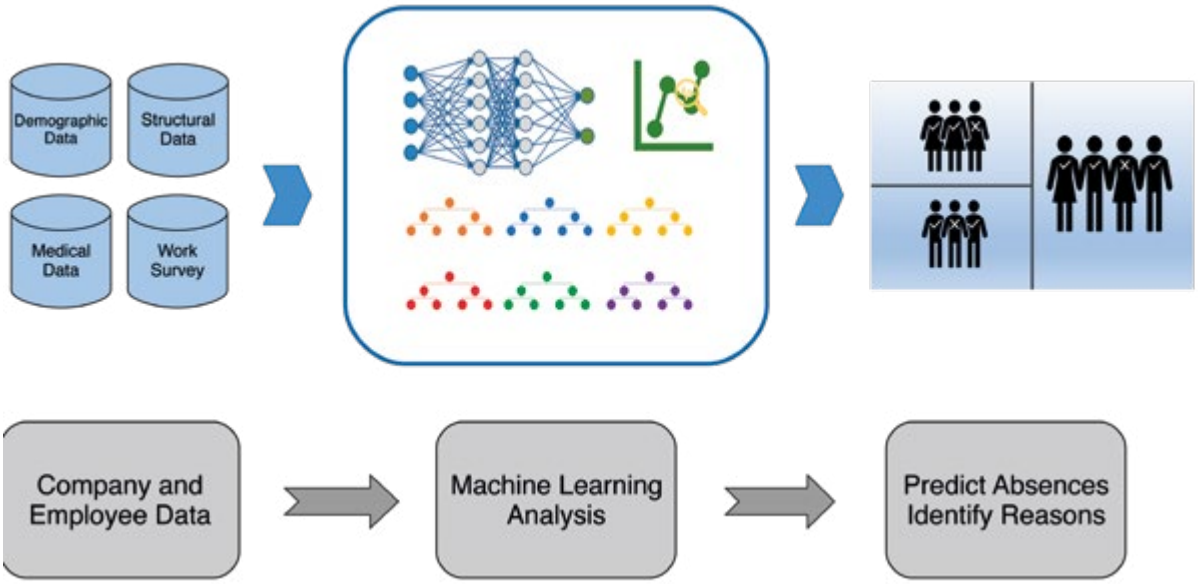


Figure 22 : Machine learning for predicting employee absences.

adjust as learning proceeds through several traversals of the layers. Artificial Neural Networks perform tasks without programmed rules and prior knowledge and instead automatically learn from the examples or training data that they are given to process.

In our experiments, the training data consist of features that are categorized into medical measurements, demographic data, work structural data, self-reported health behavior, self-reported work characteristics, and health- or behavior history and the target output (i.e., the number of days of absence or the range of the number of days corresponding to absence risk level). Only data without missing values are considered, and values are scaled to prevent large numbers from influencing the model. The metrics used for performance measurement are R^2 for regression, and accuracy (i.e., the ratio of correct predictions to total predictions) for classification. Furthermore, due to limited training data size, cross-validation is employed to generalize model evaluation.

Results show that Random Forest and Artificial Neural Networks perform particularly well, with an average accuracy rate of 74%. Current work is now focused on using these models on feature sensitivity analysis to find out how input features influence the output target. This task is important for its application in scenario management, risk assessment and decision making. The two approaches being studied are a derivative-based method and an input perturbation method. In the first approach, we look at derivatives as representations of the variations in the output parameters with respect to small changes in each input parameter. The second approach introduces small perturbation on each input of the network and the corresponding change in the output is measured.

Extraction of spatial-temporal features of bus loads in electric grids

Predicting the bus load is a fundamental issue for ensuring secure and reliable power systems. The goal is to estimate the power load in advance (e.g., in hours or days) such that the dispatch of power plants can be optimized. Thus far, most predictions in this area have focused on point-wise prediction based on previous data on each bus without considering their spatial relationship.

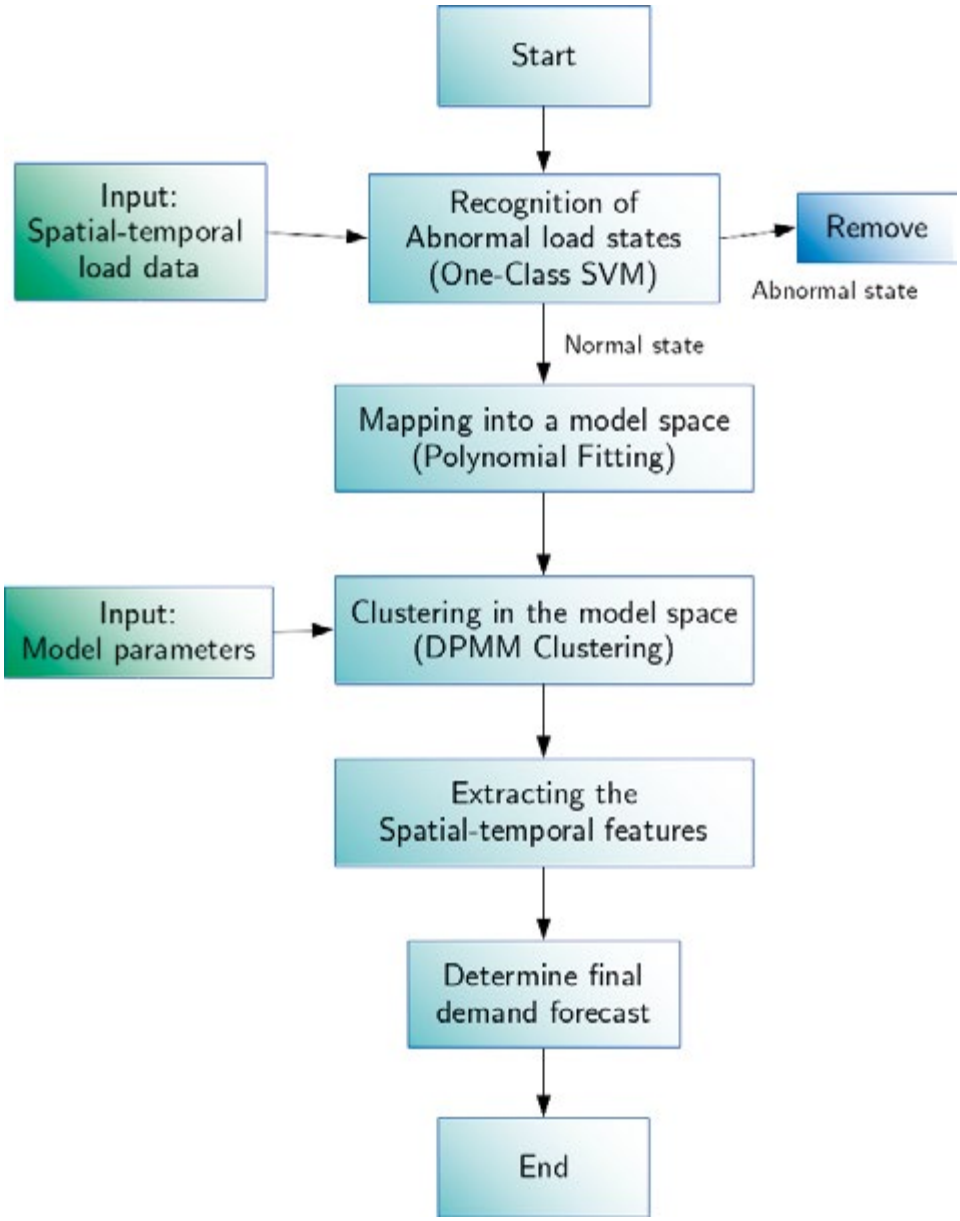


Figure 23: Spatial-temporal feature extraction workflow.

The dispatch therefore cannot be optimized with respect to the locality of power plants, especially considering that the consumption of renewable energy also has to be assured. Moreover, the abnormal states in the network are more difficult to detect if the spatial correlation is neglected because the load distribution of a single bus is quite regular and has a very small quantity of outliers. Hence, the extraction of spatial-temporal characteristics from previous data plays an important role in power system management because the stress level of the power flow is strongly influenced by the spatial-temporal distribution of bus loads, and the security constrained unit commitment (SCUC) also requires spatial-temporal features of bus loads to perform the limits of transmission flows and evaluate

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.

2019 wurde dazu in der Gruppe in folgenden Anwendungsbereichen gearbeitet: mathematische Onkologie, Maschinelles Lernen für Anwendungen in der Medizin und für Energiesysteme, sowie energieeffiziente Programmierung.

security margins with possible load mismatches between demands and supplies.

A great deal of research has been devoted to extracting key characteristics for the single bus load prediction. This research has mostly focused on statistical methods for short-term load forecasting. Recently, machine learning techniques – such as the Support Vector Regression Model, the Random Forest Method, the Kalman filter method, and Artificial Neural Networks (ANNs) – have also been employed. However, in many cases, the multiple bus loads are not mutually independent. For example, if two closed buses share similar conditions (e.g., population size, the intensity of manufacturing activities, weather conditions, etc.), their loads are fully coupled. Therefore, to achieve a good prediction, the spatial-temporal patterns of multiple loads are needed. Two major reasons currently exist for the lack of research on this issue. The first is that extracting union features of multiple temporal bus loads during the data flow process is a challenging task. The second is that the description and visualization of the extracted spatial-temporal features remain problematic due to the couplings and interactions between the spatial and temporal domain.

To overcome the difficulties mentioned above, we propose using a hybrid method composed of forecast-aided one-class learning and clustering in dynamic model space. We first apply a one-class support vector machine (SVM) to detect the variation of overall operation behaviors of the power system. We then use Dirichlet Process Mixture Model (DPMM) clustering to capture the spatial-temporal features of the normal load states. In general, the pattern recognition methods cannot be applied directly to the feature extraction of data flow (e.g., spatial-temporal loads), and we hence use the polynomial function as a surrogate model in order to represent the input data by model parameters as it can capture the variation trend of time series loads and suppress irrelevant components, such as fluctuation and noise. Next, the stable and parsimonious model parameters – instead of the original load data – are clustered by means of DPMM clustering in the model space. Learning in a model space helps to capture the dynamic characteristics (i.e., increasing/decreasing trend) of the bus load profiles. In addition, it has been proven that better clustering performance would be achieved in the model space than in the data space since the data are linearly inseparable in most cases but become sparse and separable when they are mapped into a model space.

GPU Mekong Project – simplified multi-GPU energy-aware programming

In the exascale computing era, reducing power consumption is considered highly critical. However, a reduction of power consumption should not go hand in hand with a reduction in performance. We thus address compiler optimizations that reduce power consumption without sacrificing performance.

Graphic processing units (GPUs) are powerful computing accelerators that are being pervasively used in areas such as scientific computing, machine learning, and data analytics. There is a steadily increasing interest in using multi-GPUs in parallel both to overcome the memory limitations resulting from a single device and to further improve performance.

Figure 24: The Mekong project aims to transform a single data stream into a large number of smaller streams that embrace smaller islands (computational units, memory) that mostly operate independently except for interactions, such as data distribution, communication, and synchronization. © Visible Earth, NASA

The DMQ/EMCL group works together with the Computing System Group (CSG) from the Institute of Computer Engineering at Heidelberg University (ZITI) on the Mekong project. The main objective of Mekong is to develop a compiler that allows for a simplified path to scale out the execution of GPU programs from one GPU to almost any number, independent of whether they are located within one host or distributed at the cloud- or cluster level.
In this context, we developed our instrumentation to measure both the compute performance (execution time) and the power consumption of Nvidia GPU programs. Power consumption measurements were conducted by accessing the internal GPU powermeter by using the Nvidia Management Library (NVML). To this end, we used the CUDA Flux profiler to collect instruction information from the compiler. We trained our prediction models by using CUDA Flux profiler information as model features and by using execution time or power consumption as model targets for each GPU program. Our machine learning models enable portable and fast predictions for the execution time and power consumption of GPU programs among different GPUs that use only hardware-independent features. In [Braun et al., 2020], we present our models, which yield good prediction results and can therefore be used in compiler analysis to optimize energy efficiency as needed. Finally, we developed GPU mini-apps (small showcases) for evaluating the subsequent energy efficiency optimizations of the compiler.

Currently, work is underway to quantify the power consumption of data transfers and NUMA- (non-unified memory access) related effects in multi-GPU programming during numerical simulations.

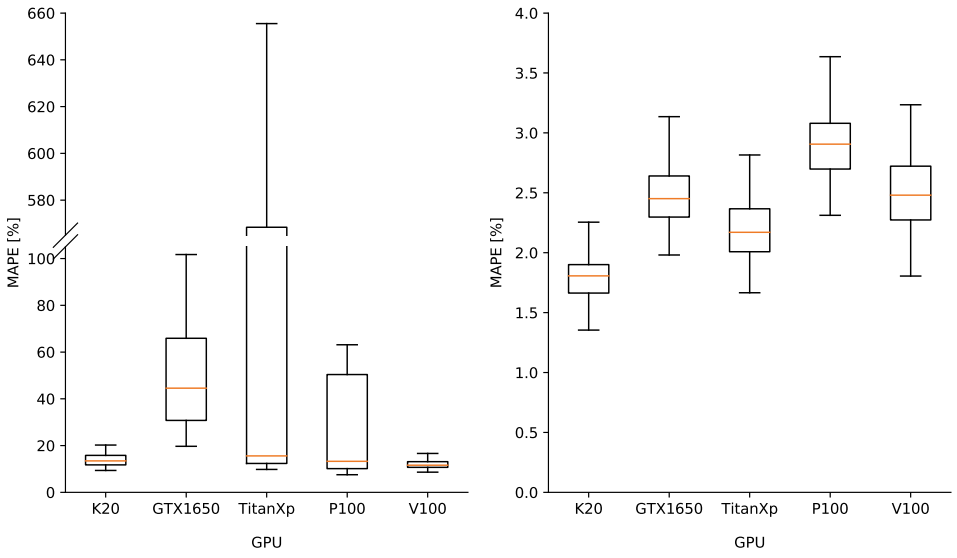
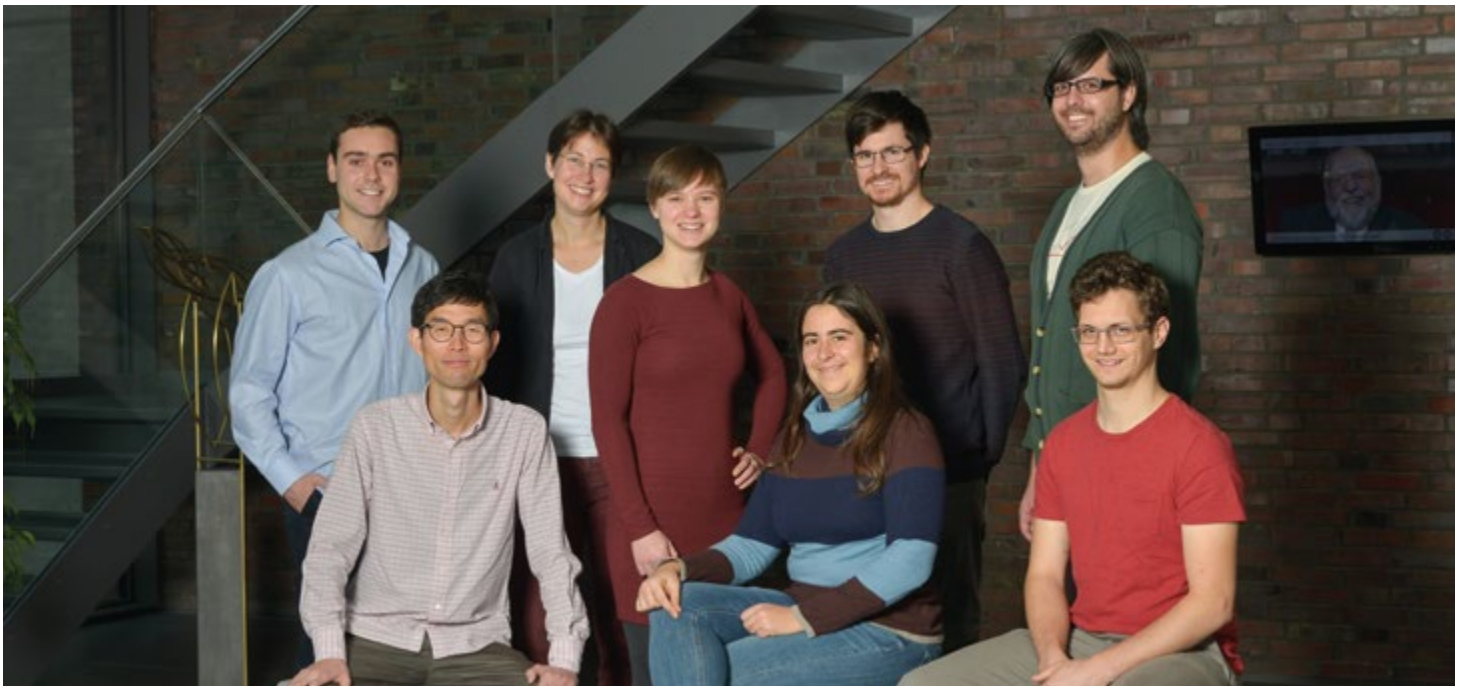


Figure 25: Portability of time- (left) and power (right) prediction across different GPUs: mean absolute percentage error (MAPE) scores for all iterations of nested cross-validation with the median-, first-, and third quartile.

2 Research

2.6 Groups and Geometry (GRG)



Group Leader
Prof. Dr. Anna Wienhard

Staff members
Johannnes Horn (since October 2019)
Mareike Pfeil
Florian Stecker (until August 2019)

Visiting scientists
Prof. Dr. Maria Beatrice Pozzetti
Dr. Daniele Alessandrini (until August 2019)
Dr. Jonas Beyrer
Dr. Nguyen-Thi Dang (since October 2019)
Dr. Valentina Disarlo
Dr. Gye-Seon Lee (until August 2019)
Dr. Andreas Ott
Dr. Andrew Sanders
Christoph Karg (until September 2019)
Marta Magnani (since August 2019)
Evgenii Rogozinnikov
Anna Schilling

Students
Clemens Fruböse (until August 2019)
Lukas Sauer
Menelaos Zikidis

The Groups and Geometry research group 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 – that is, 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 as the study of all properties of a space that are invariant under a given group of transformations. In short, geometry is symmetry. This concept unified the fields of classical Euclidean geometry, hyperbolic geometry (a new field at the

time), 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 have an influence 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.

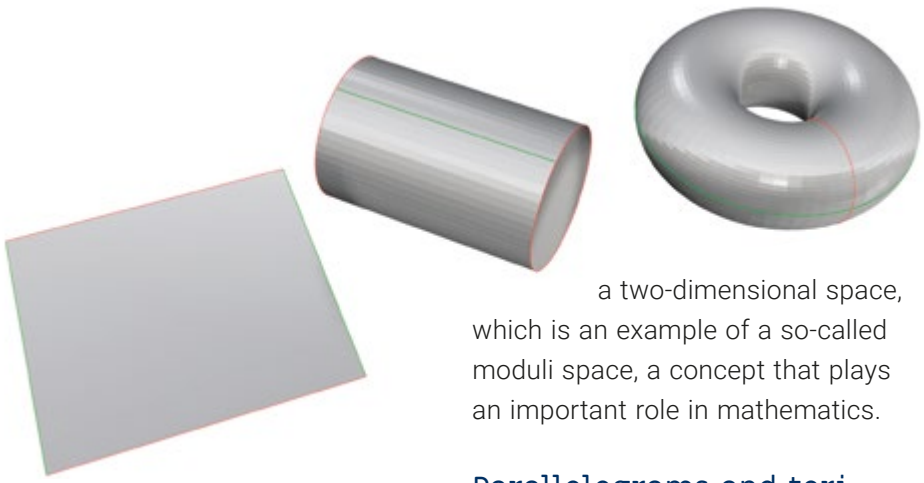


Figure 26: Gluing a Torus from a Parallelogram.

Torus - from donuts to moduli spaces

One of the most fundamental mathematical objects is the torus. In daily life, we encounter the torus as the surface of a donut, which has become a famous example used to illustrate the point of view of topologists, who make no distinction between the donut and the coffee mug. For a topologist, two-dimensional surfaces in three-dimensional space differ only in their number of holes. Both the donut and the coffee mug have one hole and are hence topologically identical. As geometers, we are interested in tori with richer structure. The operation of addition can be introduced on the torus, and similar to other surfaces, we can equip tori with a measurement of angle. Mathematicians refer to this extra data as a conformal structure, and not all two-dimensional tori can be identified with respect to this structure. In fact, the collection of all two-dimensional tori with a conformal structure can be parametrized by a two-dimensional space, which is an example of a so-called moduli space, a concept that plays an important role in mathematics.

Parallelograms and tori

In order to define extra structures on a torus, a good mathematical model is needed. A torus can be constructed by gluing a parallelogram in the plane, as illustrated in Figure 26. A cylinder is obtained by gluing the left (green) edge to the right (green) edge. Next, gluing the ends of the cylinder along the red curve results in a torus. A torus can thus be modeled by a parallelogram in the plane whose opposite sides are identified. Now, structures can be defined on the torus by using features of the Euclidean plane.

Addition on the torus

The structure of addition on a torus can be defined by the addition of vectors in the plane (see Figure 27). Let us fix a vertex O on the parallelogram. Taking two points – P and Q – in the parallelogram, we can define their sum in the plane as the concatenation of the position vectors from O to P and from O to Q and denote by R the resulting point on the plane. Let us think of the plane as being tessellated by copies of the parallelogram obtained by translating it along its sides. The point R might not lie on the original parallelogram, but it lies on one of the translated parallelograms. We have constructed the torus by gluing opposite sides and

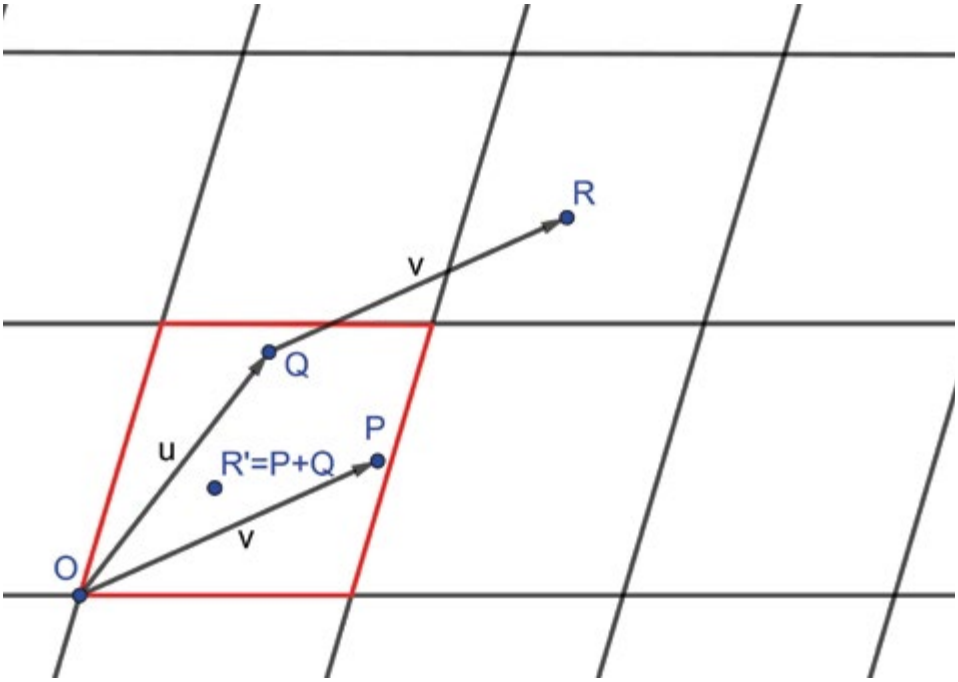


Figure 27: Addition on the Torus.

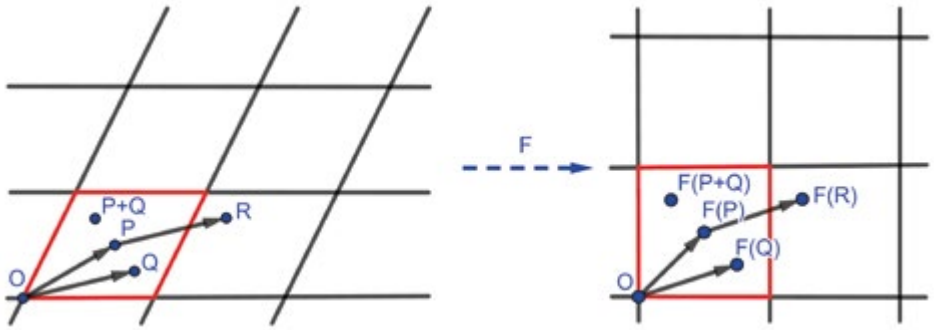


Figure 28: Identification of Abelian Tori by Shearing.

thereby identifying all these parallelograms. Any point R on the plane is identified to a point R' in the original parallelogram that represents the same point on the torus. Setting P+Q=R' defines the addition of points on the torus.

A structure of addition is what mathematicians refer to as the structure of an abelian group, and we refer to a torus with a fixed structure of addition as an abelian torus. However, this group structure does not distinguish between different tori obtained in this way. We can map a tessellation by a parallelogram S1 onto a second tessellation by a parallelogram S2 with a function F by shearing, as illustrated in Figure 28. This identification preserves the structure of addition – that is, it does not matter whether we first take P+Q on the left and then map it to the right or whether we first map P and Q to the right separately and then take the sum of the resulting points F(P)+F(Q). This property is expressed by the formula F(P+Q)=F(P)+F(Q). In the same way, we can define a map from the tessellation with respect to S2 to the tessellation with respect to S1 and thereby identify the corresponding abelian tori.

Measuring angles on the torus

The torus inherits even more structure from the Euclidean plane. In the Euclidean plane, we can measure angles, lengths, and areas, which allows us to define the same measurements on the torus. Let us focus on the measurement of angles. The sides of the parallelogram are identified via translations, and these translations clearly preserve angles. The torus hence inherits a measurement of angle, which mathematicians refer to as a conformal structure. We refer to a torus with a fixed measurement of angle as a conformal torus. The question thus arises: Is it possible to distinguish between tori that come from different parallelograms according to their conformal structure? The answer is yes. In the end, it boils down to the fact that shearing does not preserve angles, and the map F above thus does not identify the conformal structures.

Such a question is more generally referred to as a moduli problem. Whenever mathematicians define an object, they like to understand how many of these objects exist and how they are related to one another. The example above demonstrates that all abelian tori can be identified such that the structure of addition is preserved. The moduli space of abelian tori is hence a single point. If the mathematical objects are instead conformal tori, then there is a two-dimensional space that parametrizes these objects.

To illustrate this feature, it is necessary to find a standard model for parallelograms in the plane. Taking a parallelogram S, we can translate, rotate, and scale it such that

i. one edge is identified by line e between the origin O and the point (1,0) and
ii. the entire parallelogram lies above the horizontal axis (see Figure 29).

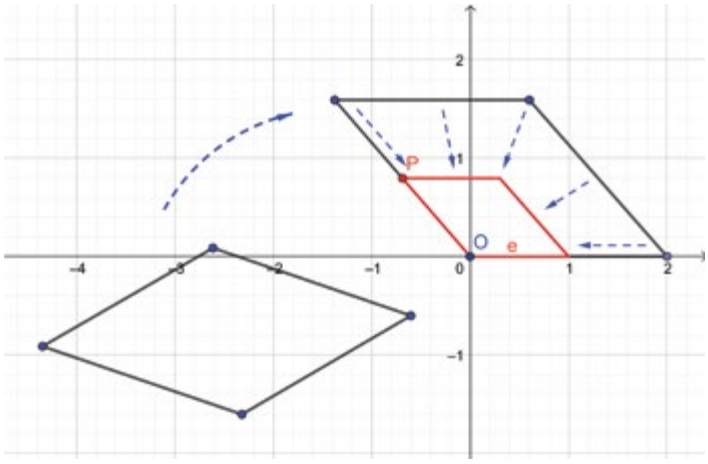


Figure 29: Identifying Parallelograms via Points in the Upper Half-Plane.

Translating, rotating, and scaling do not change angles, and these operations hence do not change the conformal structure of the associated torus. Under these conditions, the parallelogram is uniquely defined by vertex P, which is connected to the origin by the second, free edge. The set of points above the horizontal axis (not including the axis itself) is called the upper half-plane and is denoted by H2. A unique torus can be associated to every point P in H2 by gluing the parallelogram with edges e and [O,P] to it, and any conformal torus can be represented in this way. The parametrization problem can thus be reduced to a question of which parallelograms with property i, ii can be identified with one another

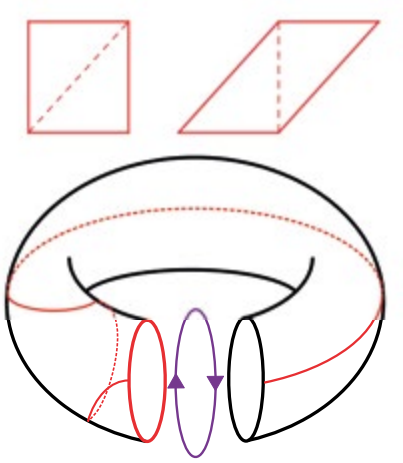


Figure 30: Dehn Twist on Parallelogram and Torus.

while preserving their angles. These identifications come from cutting and gluing. For example, we can also obtain the torus defined by the unit square via the parallelogram in Figure 30. This parallelogram is obtained from the unit square by cutting it along its diagonal and rearranging the two triangles. This operation corresponds to cutting the torus along the curve corresponding to edge e, twisting one side of the cylinder 360°, and gluing the edges back together, thereby forming a so-called Dehn twist (see Figure 30). This cutting and gluing does not change the way we measure angles, and hence, two conformal tori are identified whenever they differ in these operations. This identification is the last one we have to make before we can finally obtain a model for the moduli space of conformal tori by identifying points in H2 for which the corresponding parallelograms are identified by cutting and gluing. The result is illustrated in Figure 31. The gray area in H2 is chosen such that only the points on its boundary can be identified by cutting and gluing. Similar to the definition of the torus by identifying the opposite sides of a parallelogram, we can obtain a model for the moduli space by identifying the boundary of the gray region. The interesting geometry of the moduli space itself should also be noted (see the yellow structure in Figure 31).

This is the first example of a moduli space of conformal structures on a surface. These moduli spaces for surfaces with a larger number of holes are the object of study in Teichmüller theory and have a very rich and well-studied structure. The rich structure of these moduli spaces motivated the study of more general analogues – so-called higher Teichmüller spaces – which parametrize more complicated structures on the surface and are a major field of research in the GRG group [Wienhard A. An Invitation to higher Teichmüller Theory. Proc. Int. Cong. of Math. (2018), Vol. 1: 1007–1034]. The generalization of identifications by cutting and gluing is described in terms of the action of another interesting mathematical object, the so-called Mapping Class Group. Valentina Disarlo is an expert on this topic within the GRG group.

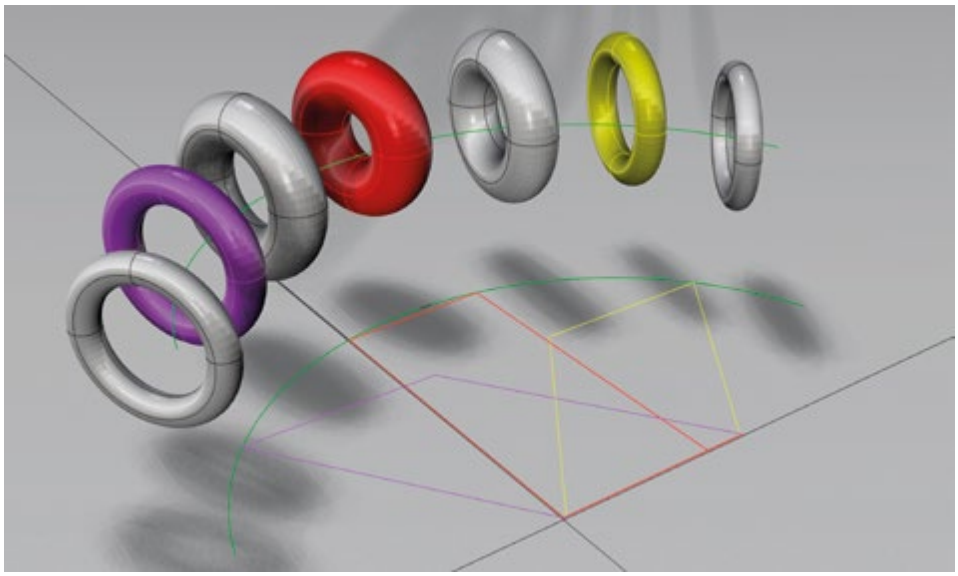


Figure 32: Torus Fibration over a Curve in the Upper Half-Plane.

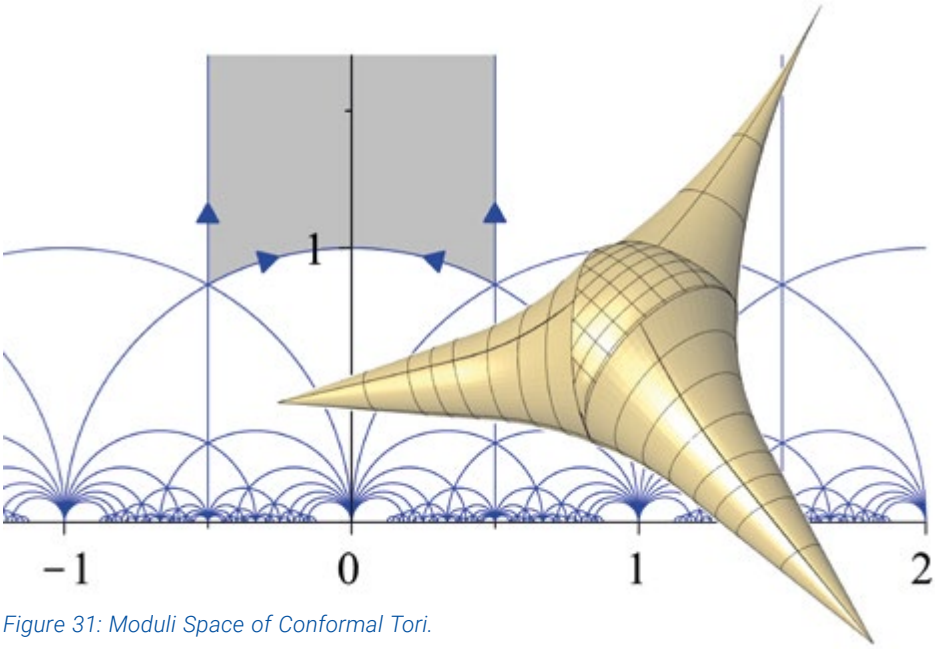


Figure 31: Moduli Space of Conformal Tori.

Torus fibrations and their mirror images

Torus fibrations

Tori play an important role in many areas of mathematics. We next describe another situation in which tori appear and that is also significant in theoretical physics, namely torus fibrations. A torus fibration is a family X of conformal tori that are parametrized by a parameter space B.

More specifically, for every b in B, the fiber Xb is a conformal torus that varies continuously with parameter b. X can be thought of as lying over parameter space B, which is hence referred to as the base of the torus fibration. The fibers can be imagined as lying vertically over the base like the fibers of a muscle.
An example of a torus fibration can

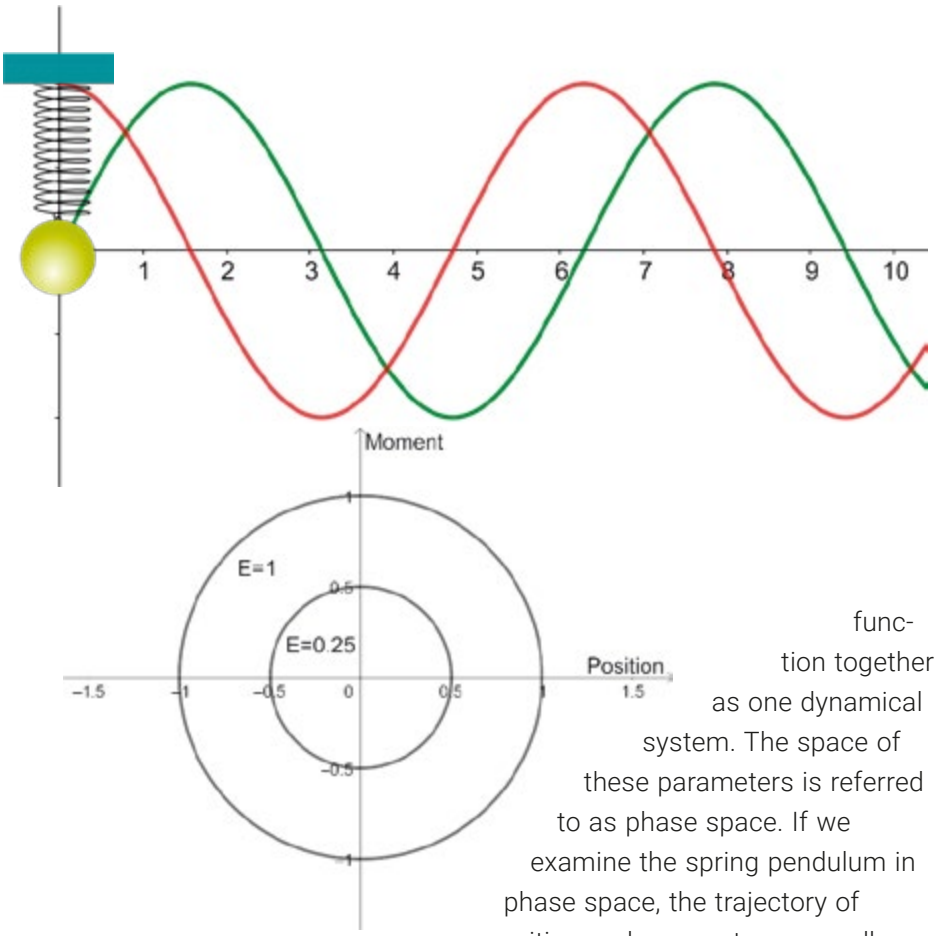


Figure 33: Spring Pendulum in Phase Space.

be given by using the description of the moduli space of conformal tori discussed above (Figure 32). As a base, we take a curve in H2. Every point on the curve defines a parallelogram and therefore also a conformal torus. The collection of all these tori defines a torus fibration. In most of the interesting cases, the parameter space B is of a higher dimension. The previous example could be extended by imagining a torus fibration over all of H2 that is defined in the same way.

Integrable systems

Torus fibrations appear naturally in the analysis of physical systems with predictable behavior. A simple example can be found in a spring pendulum (see Figure 33). Here, the height function is a sine function, and the momentum function is a cosine function. The central idea in Hamiltonian mechanics is the study of the position function and the momentum function together as one dynamical system. The space of these parameters is referred to as phase space. If we examine the spring pendulum in phase space, the trajectory of position and momentum over all times can be found to describe a circle. The radius of this circle corresponds to the amplitude of the oscillation or – in other words – to the energy of the system. If the energy E is viewed as a parameter in the positive real numbers, a circle fibration of phase space over this base is obtained. As demonstrated above, a torus consists of circles, and this family of circles can hence be viewed as a one-dimensional analogue of a torus fibration.

More generally, torus fibrations appear as so-called integrable systems, which are Hamiltonian systems with the maximal amount of conserved quantities. This property makes it possible to solve the systems explicitly. The solutions are given via integration, which is why the system is called integrable. In the previous example, the preserved quantity was energy. In general, when a system is completely integrable (i.e., the space has the maximal possible number of independent conserved quantities), it is described in phase space by a torus fibration according to a classical theorem by Liouville and Arnold. The base space is given by the parameters of the preserved quantities. By fixing values for all these quantities, the solutions sweep out a torus in phase space. An example in which a two-dimensional torus is obtained can be found in the motion of a rigid body about its center of mass. Here, the preserved quantities are energy and angular momentum, and the explicit solutions can be given by so-called elliptic integrals.

We recently demonstrated that several of the higher Teichmüller spaces mentioned above are in fact completely integrable systems [Sun Z, Wienhard A, Zhang T. Flows on the PGL(V)-Hitchin component. Arxiv 1709.03580].

Mirror symmetry and the SYZ conjecture

Another motivating factor behind studying torus fibrations is the

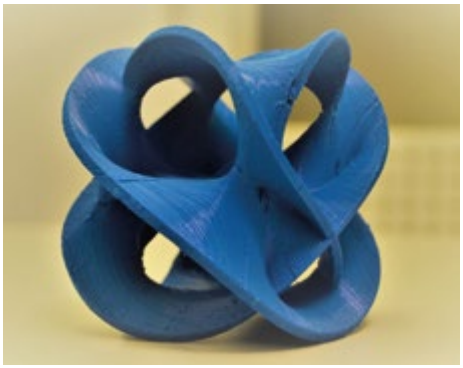


Figure 34: Two-dimensional Slice through a Quintic.

Die Arbeitsgruppe **Gruppen und Geometrie** arbeitet eng mit der Arbeitsgruppe „Differentialgeometrie“ an der Uni Heidelberg zusammen. Beide Gruppen werden von Prof. Dr. Anna Wienhard geleitet.
Symmetrien spielen eine zentrale Rolle in der Mathematik als auch in vielen Naturwissenschaften. In der Mathematik verstehen wir unter Symmetrien die Transformationen eines Objektes, die diese invariant lassen. Solche Transformationen lassen sich verknüpfen, d.h. hintereinander ausführen und bilden so eine Gruppe.
Im 19. Jahrhundert entwickelte der Mathematiker Felix Klein einen neuen Begriff der Geometrie: Geometrie ist das Studium der Eigenschaften eines Raumes, die invariant sind unter einer gegebenen Gruppe von Transformationen. Kurz gesagt: Geometrie ist Symmetrie.
Mit diesem Konzept vereinheitlichte Klein die klassische Euklidische Geometrie, die damals gerade neu entdeckte hyperbolische Geometrie als auch die projektive Geometrie, die aus dem Studium der perspektivischen Kunst erwuchs und die nicht auf dem Messen von Abständen, sondern auch Inzidenzrelationen beruht. Noch wichtiger ist, dass Felix Kleins Konzept unser Verständnis von Geometrie in der Mathematik und der theoretischen Physik grundlegend verändert hat und bis heute prägt.
Unsere Arbeitsgruppe beschäftigt sich mit verschiedenen mathematischen Forschungsfragen aus dem Gebiet der Geometrie, Topologie, sowie der dynamischen Systeme, in denen das Zusammenspiel zwischen Räumen und Gruppen, die auf diesen als Symmetriegruppen wirken, zentral ist.

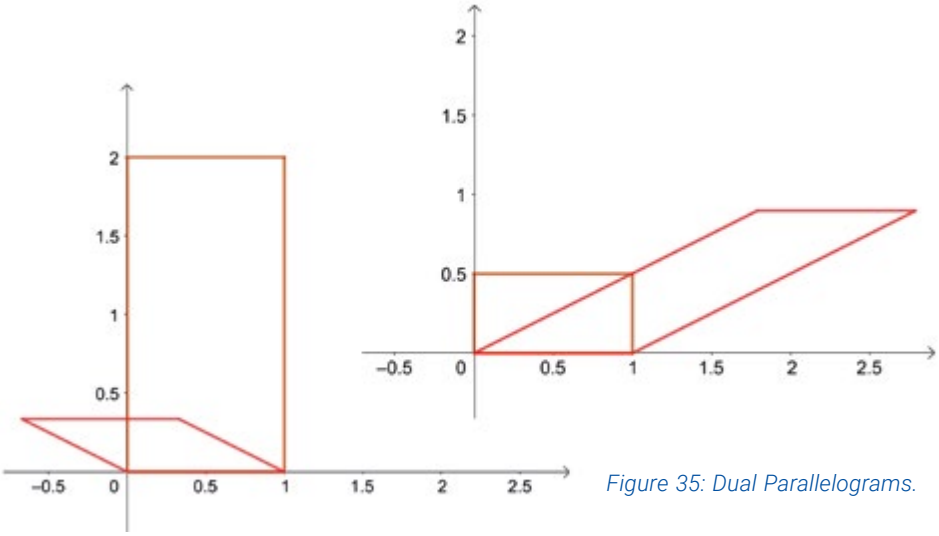


Figure 35: Dual Parallelograms.

Strominger–Yau–Zaslow (SYZ) conjecture, which provides an interpretation of mirror symmetry by using torus fibrations. Mirror symmetry is a duality between different physical models of string theory. This physical duality has surprising implications for the mathematical objects with which it is concerned: so-called Calabi–Yau manifolds. The simplest example of a Calabi–Yau manifold is the two-dimensional conformal torus discussed above, which is the only two-dimensional example. However, many more examples can be found in higher dimensions. Another famous example – in 6 dimensions – is the quintic (see Figure 34). Mirror symmetry implies that every Calabi–Yau manifold X should have a mirror Calabi–Yau manifold X’ such that certain geometrical data are exchanged. Evidence for the veracity of this conjecture can be found in statistics on big datasets of known Calabi–Yau manifolds; however, a general method of producing the mirror partner of a given Calabi–Yau X has not yet been found. One proposal is the SYZ conjecture, which provides an interpretation of mirror symmetry in terms of a duality of torus fibrations [Strominger A, Yau S, Zaslow E. Mirror symmetry is T-duality. Proceedings of the winter school on mirror symmetry, Cambridge (2001): 333-347.]. Not only does this conjecture enable the translation of mirror symmetry in a geometric statement, but it also proposes a constructive way of finding the mirror Calabi–Yau.

The duality of a single two-dimensional conformal torus can be described in terms of a duality of the corresponding parallelograms, as illustrated in Figure 35. In the simplest case, a parallelogram can be described by a point on the y-axis in H2 at distance l from the origin. The dual parallelogram can be described by a point on the y-axis at distance 1/l. If we imagine the torus as illustrated in Figure 26, this means that the circumference of the vertical circles on the torus is inverted. For a general point P in H2, we first have to scale it by 1/l2, where l is the distance to the origin, and then reflect it across the y-axis. The resulting point in H2 describes the dual conformal torus. In Figure 35, the red and brown parallelograms are both dual.

The SYZ conjecture proposes that given a Calabi–Yau manifold X, which is a torus fibration over a base B, it is possible to describe the mirror partner X’ via a torus fibration over the same base with fibers given by the dual tori. However, for many Calabi–Yau manifolds, it is not possible to find an honest torus fibration, and the SYZ conjecture therefore cannot be applied to construct the mirror dual. The best one could hope for is a fibration in which almost all fibers are tori.

Singular fibers

Our framework must be slightly extended to allow for fibrations for which there is some subset S of base B of a strictly smaller dimension above which the fibers are not tori. S is thus referred to as the singular locus, and the fibers Xs of s in S are referred to as singular fibers. However, these singular fibers are not arbitrary spaces because they are surrounded by torus fibers. It is possible to deform a torus into such a space. Figure 36 illustrates a case in which a horizontal curve on the torus is shrunk until it collapses to a point. The resulting object is called a nodal torus. The vertical circle in Figure 36 can be observed to persist through the deformation as a remainder of the geometry of the torus. One method for understanding the mirror symmetry of such fibrations is to apply the SYZ conjecture to the regular fibers and to describe a way in which to fill in the missing singular fibers of the mirror dual. Extending the SYZ conjecture to this setting remains an active field of research with the goal of providing a true mathematical understanding of mirror symmetry. The solution is known for the example of a torus fibration with a single nodal torus over the complex plane: It is the mirror dual of itself [Auroux D. Mirror symmetry and T-duality in the complement of an anticanonical divisor. J. Gökova Geom. Topol. (2007), GGT 1: 51-91.].

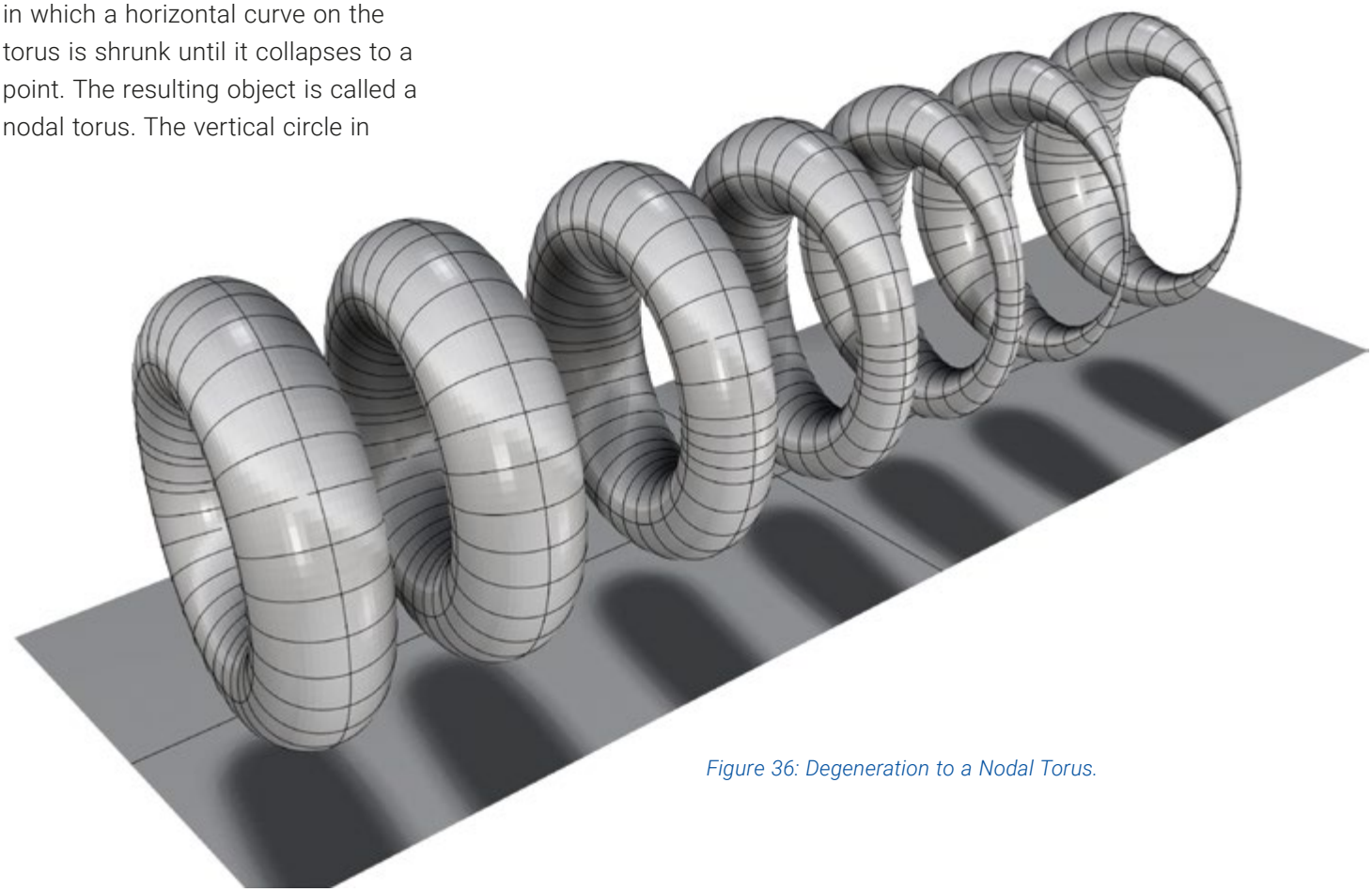
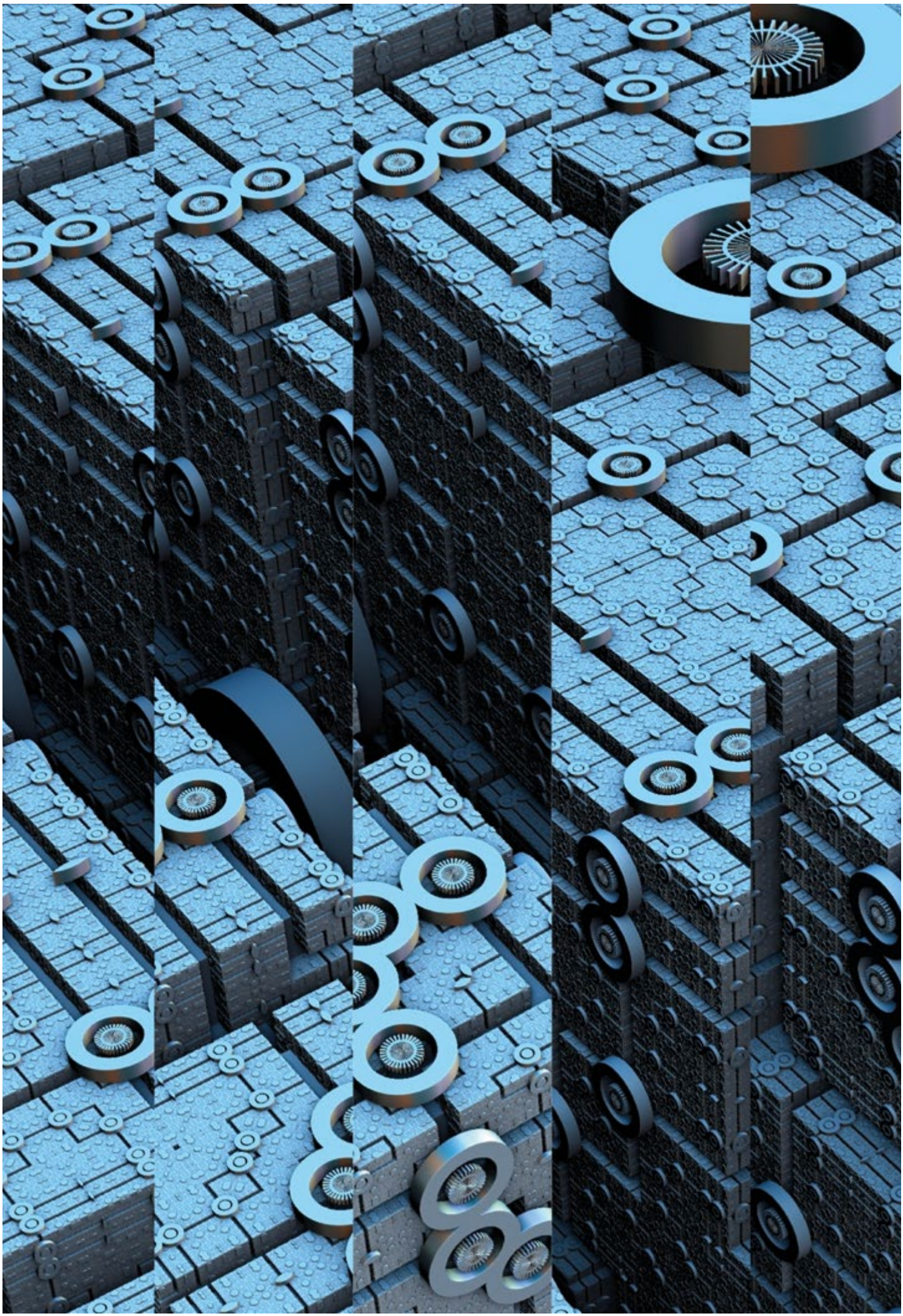


Figure 36: Degeneration to a Nodal Torus.



2 Research

2.7 Molecular Biomechanics (MBM)

Group Leader
Prof. Dr. Frauke Gräter

Staff members
Matthias Brosz (since August 2019)
Dr. Csaba Daday (until April 2019)
Florian Franz
Dr. Markus Kurth (since March 2019)
Isabel Martin
Dr. Nicholas Michelarakis (since May 2019)
Benedikt Rennekamp (since August 2019)
Ana María Herrera-Rodríguez
Fabian Kutzki
Dr. Fan Jin
Dr. Agnieszka Obarska-Kosinska (until July 2019)
Christopher Zapp

Students
Benedikt Rennekamp (until June 2019)
Martin Richter (since May 2019)
Anna Schroeder
Paula Weidemueller (May–August 2019)

Visiting scientists
Dr. Krisztina Fehér (until December 2019)
Prof. Kevin Hinkle (Humboldt Research Fellow, June–August 2019)

All proteins are regulated. They can be switched on and off by binding other – smaller or larger – molecules, by modifying their amino acid alphabet, or by being produced or degraded more readily. Mechanical force can also switch proteins on or off, just like biochemical stimuli. In the Molecular Biomechanics group, we aim to understand the role of mechanical signals in protein assemblies and protein materials, and we employ Molecular Dynamics simulations and other computational techniques to study the consequences of mechanical force on protein structure and dynamics. Below, we highlight some of our studies in 2019. Von Willebrand factor is a protein that is only active under the high shear forces that are present in flowing blood

close to injuries. Along with collaborators from the European Molecular Biology Laboratory (EMBL) in Heidelberg, we were able to shed light on one part of von Willebrand factor that had previously represented a blind spot in our understanding, namely the so-called C4 domain. Exactly how force effects the C4 domain of healthy and diseased mutants differently is one major outcome of this collaborative study. In order to better directly monitor the consequences of flow on proteins, we devised a simulation protocol that can very efficiently mimic elongational or more complex flows around proteins, such as von Willebrand factor. Whenever we subject proteins to force in the computer, be it by directly pulling on them or by subjecting them to flow, we typically use forces higher than those that occur in vivo or than those used in experiments in order to keep the computational expense at reasonable levels. We were thus able to dissect the artifacts from overly strong pulling in simulations of protein unfolding. In this Annual Report, we highlight our recent results for the force-dependent behavior of talin and vinculin, two major players at focal adhesion sites, which are force-transduction sites between cells and their surroundings.

Hinging motions within von Willebrand factor control blood coagulation

Camilo Aponte-Santamaría and Frauke Gräter

Von Willebrand factor (VWF) is an enormous extracellular protein that plays a key role in primary hemostasis by adhering platelets at sites of vascular injury in order to stop bleeding. Its activation is mediated by mechanical stress imposed by flowing blood. Platelets bind to two different VWF domains: A1 and C4. A mutation in the latter has been shown to pathologically increase the binding of platelets and intriguingly has been associated with repeated myocardial infarction, particularly in younger women.

A joint effort involving eight research laboratories and combining structural biology, functional assays, and expertise in computer simulations was able to unravel the molecular determinants of this clinically relevant process. The structure of the C4 domain was determined at atomistic resolution by using nuclear magnetic resonance. The C4 domain is largely stabilized by disulfide bonds and contains two hinged subdomains, the flexibility of which regions is remarkably high. Computer simulations helped to determine how the clinically related mutation altered the hinge motion between these regions.

In summary, this combined approach involving experiments and simulations provided the molecular basis for understanding the interaction of platelets with adhesive blood proteins and connected this information with a clinically relevant cardiological pathology.

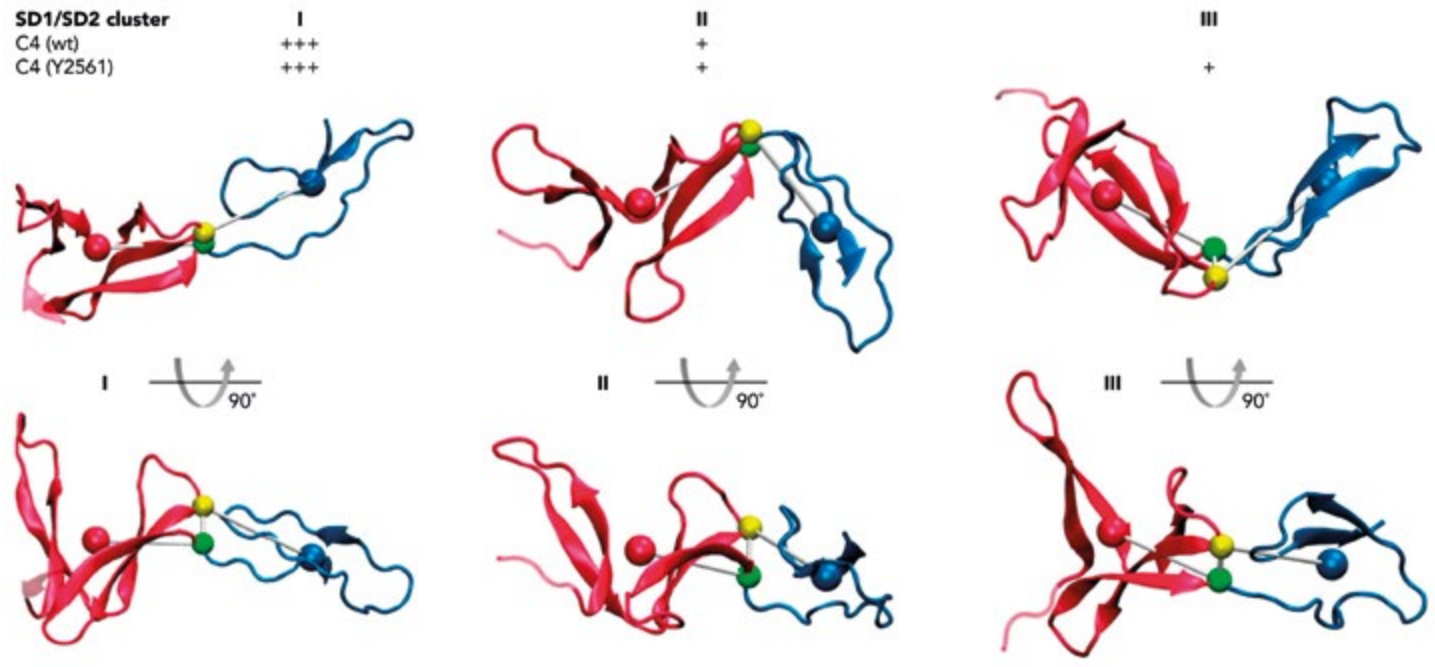


Figure 37: Structure and dynamics of the C4 von Willebrand factor (VWF) domain was determined by nuclear magnetic resonance experiments and molecular dynamics simulations. The structure revealed a hinged two-domain arrangement (SD1: red; SD2: blue) with a remarkable hinge flexibility. Three main conformations were sampled (I, II, and III). Simulations demonstrated the shift in the population of these states imparted by the clinically relevant gain-of-function mutation of residue Y2561. Figure adapted from Figure 3 in ref [Xu, 2019].

Molecular Dynamics simulations of molecules in uniform flow

Ana M. Herrera-Rodríguez, Vedran Miletić, Camilo Aponte-Santamaría, Frauke Gräter

How biomolecules behave under flow is an issue relevant to many biological systems, from blood vessels to glands. The dynamic flow response of biomolecules has been computationally studied primarily on a mesoscopic-to-macroscopic scale via a plethora of highly advanced hydrodynamic simulation techniques. However, exactly how explicit water molecules that flow around a biological macromolecule alter its dynamics on the atomistic scale has only been incidentally addressed.

In our paper, we presented a simple extension of Molecular Dynamics simulations that are used to study molecules in uniform flow and demonstrated its use for a small peptide extended by flowing water. We implemented our method in the widely used Gromacs programme to find widespread use within the community. We extended the pulling code in Gromacs, and the implementation was tested via a rectangular water box (Figure 38, top). Force was applied to water molecules within the slice boundaries after applying a weighting scheme. A smoothed distribution of external force along the flow direction was then applied (see in Figure 38, bottom). At each MD step, a dynamic reassignment of the molecules belonging to the slice was performed, and the weighting process was repeated according to the number of water molecules found in this region [Herrera-Rodriguez, 2019].

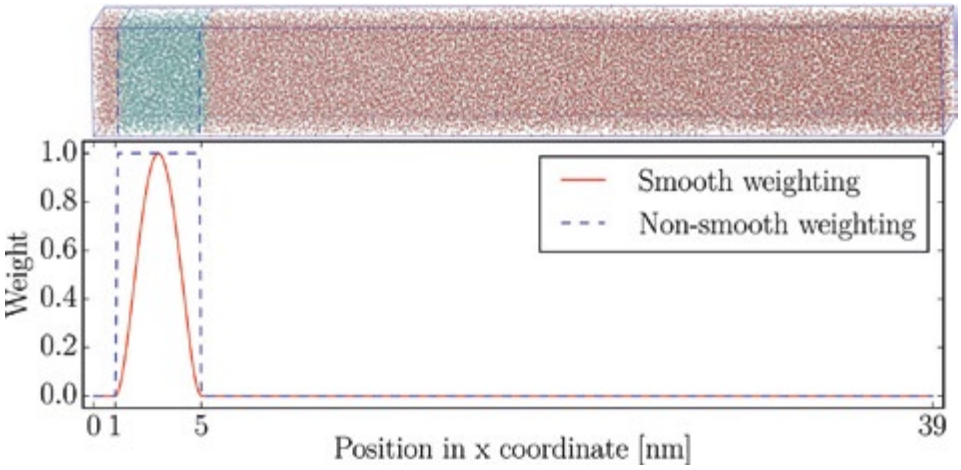


Figure 38 Example setup for flow MD simulations. a) Rectangular water box with approximately 32,000 water molecules. Water molecules within the slice are highlighted in cyan. b) Slice weighting function considering smooth (red) and non-smooth (blue) options plotted along the x coordinate of the simulation box. The smoothing allows for avoiding abrupt discontinuities in the applied force at the slice interfaces.

Nanoscale flow can play a role not only in the most commonly addressed cases (e.g., for blood) but also more generally within cell compartments and the extracellular space (e.g., synovial fluid, collagen matrices, etc.).

Three factors of force propagation: Location, location, location!

Csaba Daday and Frauke Gräter

Atomic force microscopy (AFM) is a rapidly improving experimental technique that allows us to understand – inter alia – how individual proteins react to mechanical force. However, the output of these experiments can be relatively coarse: The breaking of entire proteins or of protein domains can be detected, but the mechanism cannot be determined, and smaller changes cannot be detected. Molecular dynamics (MD) simulations can be used to alleviate these issues, but this process imposes certain technical issues. One of these technical issues is that if we want to simulate the unfolding process in a reasonable amount of time, we need to perform these simulations at forces and velocities that are orders of magnitude greater than the experimental ones.

In experiments, typical speeds for pulling apart proteins lie in the order of 1 micrometer/second: that is, 1 kilometer per 32 years. Simulations are usually about 1 million times faster (1 meter/second), which is similar to the walking speed of a human being. Understandably, as we perform these simulations, we have worried that proteins could behave differently in these two scenarios. We thus set out to investigate one source of the problem: If we pull a protein too fast, it is possible for the parts closer to the pulling point to break faster than the "shielded parts" in the middle.

Our system was composed of four identical copies of a protein domain that were connected, and we pulled at velocities from 1 cm/s (the speed of an average snail) up to 10 m/s (the speed of a fast athlete sprinting). We observed that the external domains always unfolded first in the fastest case, but this effect was not preserved in the slowest cases. We

Figure 39: Four copies of the same system subjected to an external force. The dark blue domains, which are closer to the point of force application, are subject to more apparent force than the green ones.

identified an apparent "signal" of force propagation of about 1–100 m/s, which seems fast in our day-to-day life but is much slower than the speed of sound (about 3,000 m/s in biological matter), which is the mechanical equivalent of the speed of light for electrical signals (e.g., radio waves). Our data can be useful for future studies that aim to compare similar experiments and simulations, yet they also pose an interesting theoretical puzzle as to the origin of this signal [Sheridan, 2019].

Alle Proteine sind reguliert. Sie können ein- und ausgeschaltet werden, indem andere - kleinere oder größere - Moleküle gebunden werden, indem ihr Aminosäurealphabet verändert wird oder indem sie schneller hergestellt oder abgebaut werden. Mechanische Kräfte können ebenso wie biochemische Reize Proteine ein- oder ausschalten. In der **Gruppe Molecular Biomechanics** wollen wir die Rolle mechanischer Signale in Proteinen und Proteinmaterialien verstehen. Wir verwenden Molekulardynamik-Simulationen und andere Computer-basierte Methoden, um die Konsequenzen mechanischer Kraft auf die Proteinfunktion zu untersuchen.
Wir stellen in diesem Jahresbericht einige unserer Studien aus dem Jahr 2019 vor. Der Von-Willebrand-Faktor ist ein Protein, das nur unter hohen Scherkräften aktiv ist, wie sie zum Beispiel in fließendem Blut in der Nähe von Verletzungen vorkommen. Gemeinsam mit Mitarbeiter/-innen des European Molecular Biology Laboratory (EMBL) in Heidelberg konnten wir einen Teil des Von-Willebrand-Faktors beleuchten, der bislang ein blinder Fleck war, die sogenannte C4-Domäne. Wie sich Kraft unterschiedlich auf die C4-Domäne gesunder und kranker Mutanten auswirkt, ist ein wichtiges Ergebnis dieser kollaborativen Studie. Um die Folgen des Flusses auf Proteine besser direkt beobachten zu können, haben wir ein Simulationsprotokoll entwickelt, das Kräfte so modelliert, wie sie auf Proteine wie den von Willebrand-Faktor durch den Blutfluss wirken. Wann immer wir Proteine im Computer einer Kraft aussetzen, sei es durch direktes Ziehen oder durch Fluss, verwenden wir normalerweise höhere Kräfte als in vivo oder in Experimenten, um den Rechenaufwand auf einem vernünftigen Niveau zu halten. In Simulationen der Proteinentfaltung konnten wir die Artefakte von zu starkem Ziehen trennen. Zuletzt präsentieren wir unsere jüngsten Ergebnisse für das kraftabhängige Verhalten von Talin und Vinculin, zwei Hauptakteuren an fokalen Adhäsionen, den Kraftübertragungsstellen zwischen Zellen und ihrer Umgebung.

Talin impacts force-induced vinculin activation by 'loosening' the vinculin inactive state

Florian Franz and Frauke Gräter

Focal Adhesions (FA) are large, multi-protein complexes that connect the cytoskeleton to the extracellular matrix. Their adhesive functionality is tightly regulated by mechanical stress. A key component of FA-associated mechanosensing is vinculin, which plays a critical role in embryonic development and different disease states. Vinculin consists of a globular head, a proline-rich neck region, and a rod-like tail domain that contains binding sites for many other cytoplasmic proteins. Vinculin can assume either a closed ("inactive") or open ("active") conformation.
In the cell cytoplasm, vinculin almost exclusively assumes the inactive conformation. To play its important role in the functionality of FAs, the protein needs to open into the active state.
The underlying activation mechanism, however, has yet to be fully understood.

We employ Molecular Dynamics (MD) simulations in order to demonstrate that vinculin activation is greatly facilitated by its binding on talin's vinculin binding site. Our steered MD simulations reveal that the force required for vinculin activation is drastically reduced by more than 50% upon the formation of the vinculin–talin complex, which is mainly accounted for by a weekend interface between vinculin's head- and tail domain. The lower

activation energy renders thermal activation in a physiological context much more likely. This would enhance the integration of vinculin into the focal adhesion machinery.

Interestingly, after talin dissociation, vinculin returns to its native conformation on a submicrosecond time scale, with 60% of its native contacts restored. Our results suggest a rapid dynamic equilibrium between 'tight' and 'loosened' inactive vinculin, which depends on talin and determines the level of mechanical stress required for activation. Our study has important implications for the understanding of mechano-sensing mechanisms at FAs [Franz, 2020].

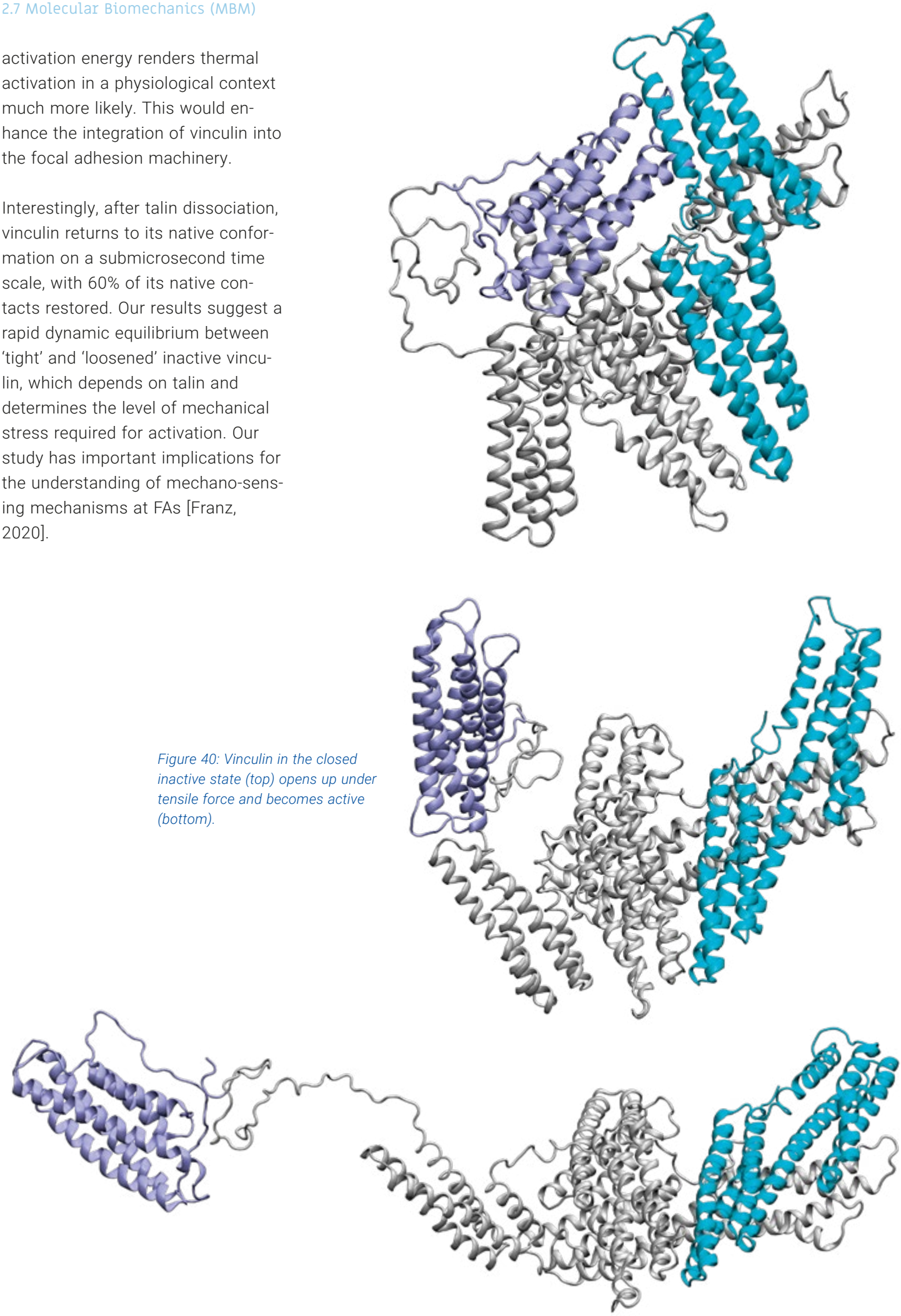
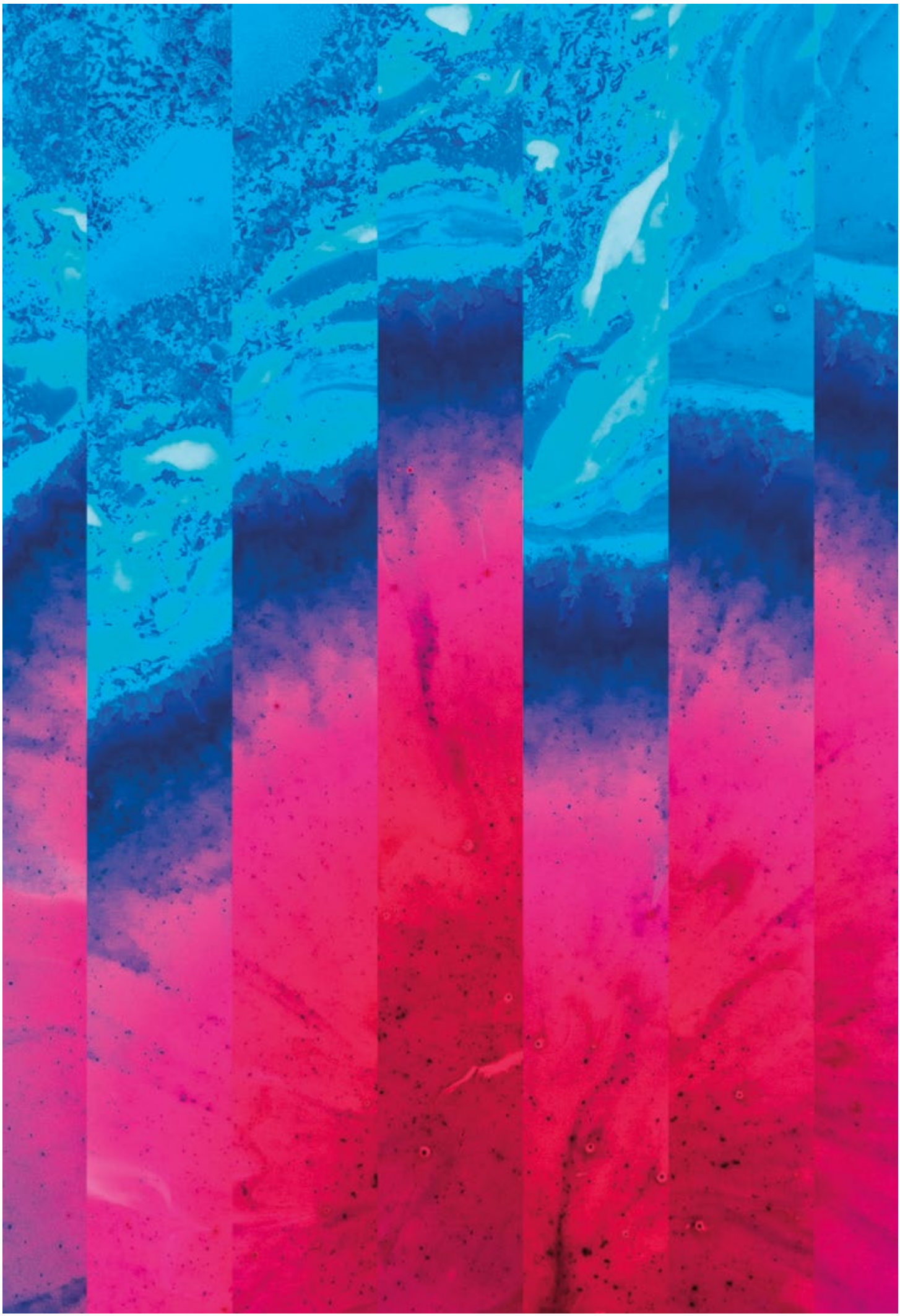


Figure 40: Vinculin in the closed inactive state (top) opens up under tensile force and becomes active (bottom).



2 Research

2.8 Molecular and Cellular Modeling (MCM)

Group Leader
Prof. Dr. Rebecca Wade

Staff members
Christina Athanasiou
Dr. Neil Bruce (until June 2019)
Manuel Glaser (since November 2019)
Dr. Daria Kokh
Abraham Muniz Chicharro (since October 2019)
Dr. Stefan Richter
Dr. Kashif Sadiq
Alexandros Tsengenes

HITS scholarship holders
Gaurav Ganotra (until May 2019)

Visiting scientists
Madhura De (Heidelberg University, since October 2019)
Lorenzo Fabbri (University of Trento, Italy, from March to May 2019)
Dr. Goutam Mukherjee (Heidelberg University)
Dr. Prajwal Nandekar (Heidelberg University, until February 2019)
Dr. Ariane Nunes-Alves (Capes-Humboldt fellowship)
Lucas Gasparello Viviani (University of Sao Paulo, Brazil, until January 2019)

Students
Lukas Adam (until May 2019)
Patrick Friedrich (until March 2019)
Sungho Han (from October to December 2019)
Anton Hanke (from April to July 2019)
Stefan Holderbach (from May to July 2019)
Konstantinos Mavridakis (since November 2019)
Daniel Saar (from July to September 2019)
Philipp Ullmann (from March to July 2019)
Jui-Hung Yuan (until June 2019)

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

In the MCM group, we are primarily interested in understanding how biomolecules interact. What determines the specificity and selectivity of a drug–receptor interaction? How can proteins assemble to form a complex, and what shape can this complex take? How is the assembly of a complex influenced by the crowded environment of a cell?

What makes some binding processes quick and others slow? How do the motions of proteins affect their binding properties? These questions are illustrative of the types of problems 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 that are based on physics and bio-/chemo-informatics. The broad spectrum of techniques employed range from interactive, web-based visualization tools to atomic-detail molecular simulations.
In this report, we outline some of the results achieved this year. These achievements demonstrate the types of methods we develop to study macromolecular interactions and their application to problems in biology, biotechnology, and drug design. Following a general overview of what was new in the group in 2019, we focus on projects concerning (i) the prediction of drug–target binding kinetics, (ii) the prediction of protein binding pocket druggability, and (iii) multiscale molecular modeling.

What was new in 2019?

In 2019, we began working on two new research projects. Informatics4Life (I4L: informatics4life.org) is a Heidelberg-based initiative funded by the Klaus Tschira Foundation that supports joint research by computational and clinical researchers and focuses on cardiovascular research. As part of I4L, we are working on a project with Patrick Most (Heidelberg University Hospital) on the structure-based modeling and development of peptide-based therapeutics against striatal muscle disorders. In 2019, we also initiated a bilateral German–Indian project involving Heidelberg University and the research group of Dr. Abhay Sangamwar at the National Institute for Pharmaceutical Education and Research (NIPER) in India on the simulation, design, and characterization of drug–polymer interactions for stable drug formulation.

During the year, we welcomed several new members to the group: Manuel Glaser, who works on the I4L project, and Madhura De and Abraham Muniz Chicharro, as doctoral students. Ariane Nunes-Alves began research supported by her Capes–Humboldt postdoctoral fellowship. She wrote about her experiences as a Brazilian scientist working in Germany in an article published in the Journal of Chemical Information and Modeling [Nunes-Alves, 2019]. Goutam Mukherjee won a poster prize in the summer at the 19th International Conference on Biological Inorganic Chemistry in Interlaken, Switzerland, for his poster on simulations of cytochrome P450-cytochrome P450 reductase interactions and electron transfer (see Chapter 9.5). Rebecca Wade and Outi Salo-Ahen (an MCM alumna, currently Professor of Pharmacy, Abo Akademi University, Turku, Finland) guest-edited a special issue of the journal 'Molecules' on the state-of-the-art in 'Molecular Modeling in Drug Design' [Wade, 2019].

In the autumn, Gaurav Kumar Ganotra defended his doctorate on methods of computing ligand–protein-binding kinetics. Jui-Hung Yuan completed his master's thesis in Scientific Computing on the development of a machine learning method to identify protein binding pockets with high potential for binding to drug molecules. In the summer, Anton Hanke and Philipp Ullmann completed their bachelor's theses in Molecular Biotechnology and Biosciences, respectively. Lorenzo Fabbri, a master's student from the University of Trento, Italy, visited during the summer as an Erasmus student to work on machine learning of drug binding kinetics. Several master's students from Heidelberg University completed internships in the group during the year: Lukas Adam and Stefan Holderbach (Molecular Biotechnology), Daniel Saar (Biochemistry), and Sungho 'Bosco' Han (Molecular and Cellular Biology).

Computational prediction of drug–target binding kinetics

Growing evidence suggests that the efficacy of a drug can be correlated with its target binding kinetics, which means that target-based drug design procedures should not aim simply to discover compounds that bind tightly (i.e., with high affinity or a favorable binding free energy) to their molecular target but also to discover compounds that bind and unbind at optimal rates. Consequently, there is a need for accurate methods to compute binding kinetic parameters.

However, their computation poses a challenge because the rates at which proteins associate to and dissociate from their binding partners – such as drugs – vary over a wide range of timescales. Recently, the increasing interest in drug–target binding kinetics has led to the development of a plethora of new methods aimed at computing rate constants for receptor–ligand binding processes. These methods have varying ranges of application, degrees of accuracy, and computational requirements.

In order to help researchers decide which method might be suitable for their projects, we developed KBbox (kbbox.h-its.org), a web server that guides users in choosing the methods they should consider on the basis of the information they wish to obtain,

the data they currently have available, and the computational resources to which they have access [Bruce, 2019a]. KBbox provides information on the toolbox of available methods, their associated software tools, an expanding list of curated examples of published applications, and tutorials that explain how to apply some of the methods (see Figure 41). KBbox was designed to allow for the easy addition of new methods, tools, and examples as they are developed and published.

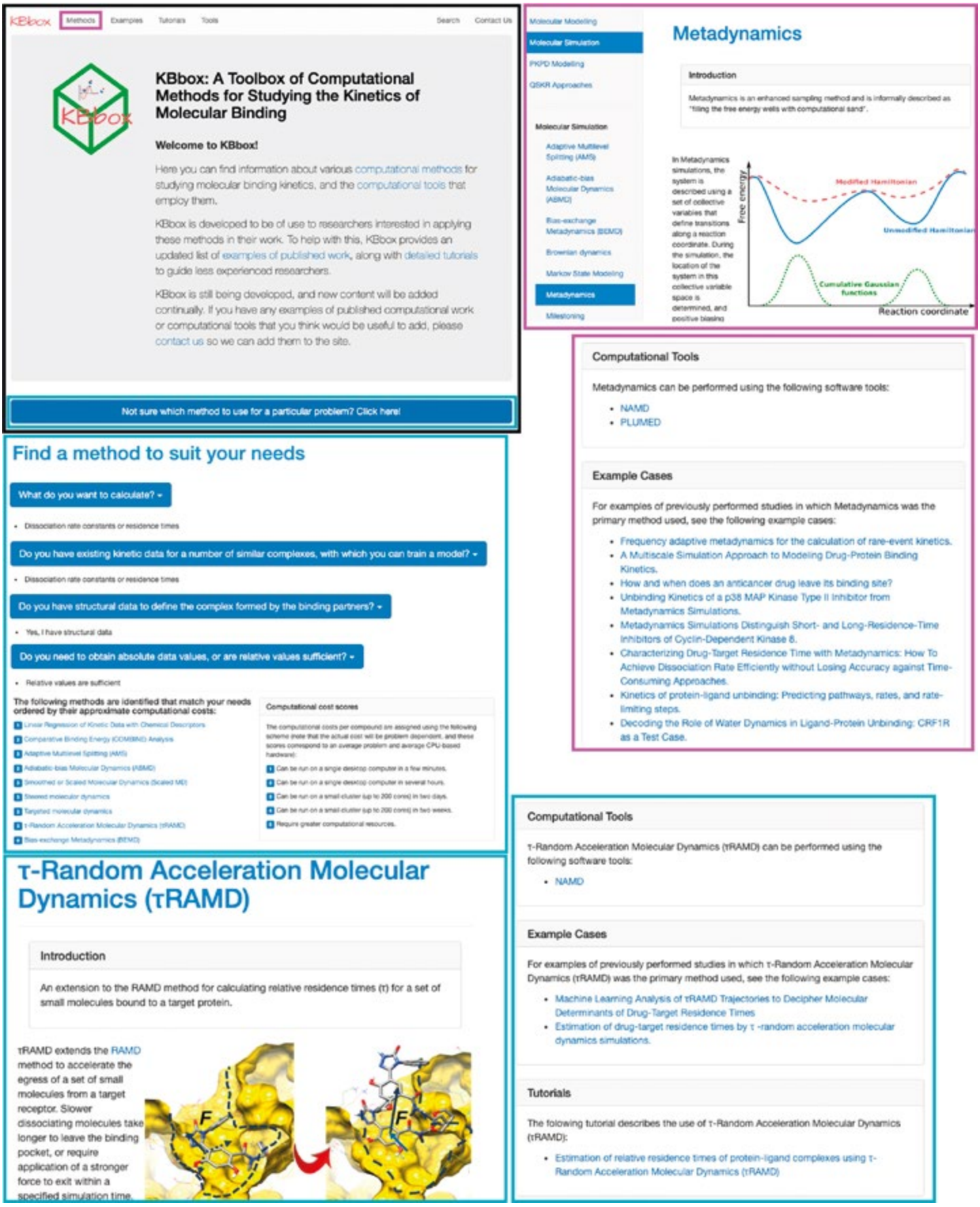


Figure 41: KBbox (kbbox.h-its.org) is a toolbox of computational methods for studying the kinetics of molecular binding. The figure displays screenshots that illustrate two example use cases. Reproduced with permission from the American Chemical Society [Bruce, 2019a].

Thus, KBbox is continually updated in order to add information on new publications, methods, and tutorials.

In 2019, we also continued to work on developing methods for computing drug–target binding kinetics by focusing on Brownian dynamics-based methods for association rates and molecular dynamics-based methods for dissociation rates. We previously developed the τRAMD method for estimating relative residence times (τ) from a set of Random Acceleration Molecular Dynamics simulations. We developed this method for ease of use and to make it possible to compute long ligand–protein residence times from short molecular dynamics simulations. In the τRAMD method, a large number of random acceleration MD (RAMD) simulations are performed in which ligand dissociation from the protein target occurs in times of about a nanosecond (much shorter than the true unbinding time) due to the application of an additional, randomly oriented force to the ligand. The length of the RAMD simulations is used to deduce τ. The RAMD simulations also provide information on ligand egress pathways and dissociation mechanisms. We initially implemented the τRAMD method for use with the NAMD molecular dynamics simulation package. To improve computational efficiency and the accessibility of the method to users, we have now implemented it in the GROMACS software.

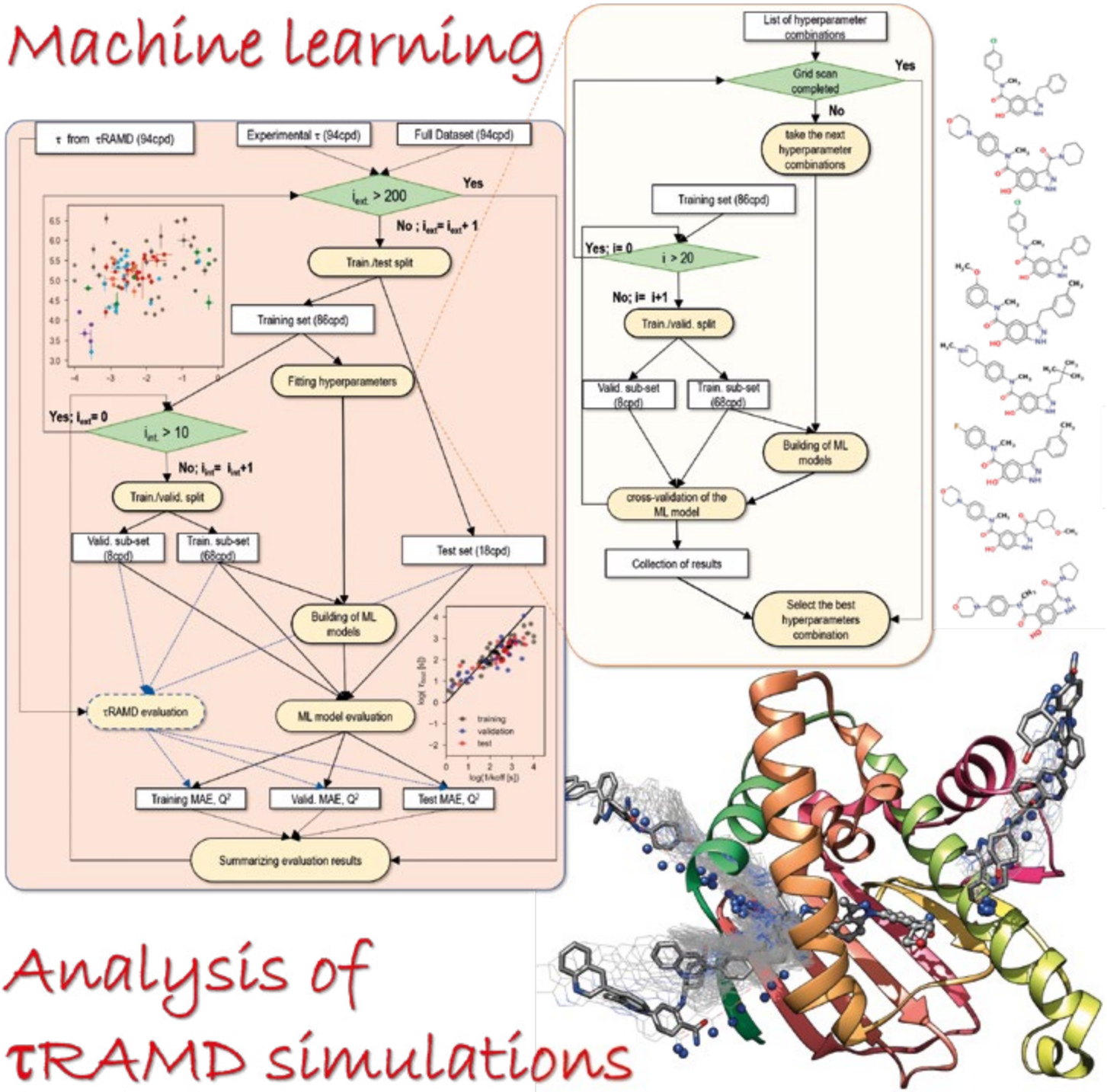


Figure 42: Illustration of the application of machine learning techniques (top left) to the analysis of τ-Random Acceleration Molecular Dynamics (τRAMD) simulated trajectories of ligand–protein dissociation (bottom right) for a set of compounds (top right). The use of machine learning techniques improves the τRAMD predictions of drug–target residence time (τ) and helps in deciphering the molecular determinants of τ [Kokh, 2019].

In addition, we developed a machine learning analysis of protein–ligand binding contacts in RAMD trajectories that can be used to improve predictions of residence time and to decipher the determinants of drug–target residence times (see Figure 42) [Kokh, 2019].

We demonstrated that regression models built on protein–ligand interaction fingerprints of ligand dissociation trajectories result in robust estimates of residence time for a set of 94 drug-like inhibitors of the anti-cancer target (heat shock protein 90 (HSP90)), even for compounds for which the length of the RAMD trajectories does not provide a good estimation of residence time [Kokh, 2019]. We thus found that machine learning helps in overcoming inaccuracies in the modeling of protein–ligand complexes due to incomplete sampling or force field deficiencies. Moreover, the approach facilitates the identification of features that are important for residence time. In the case of HSP90, for example, we observed that the interactions of a particular aromatic amino-acid residue in the binding site with bound inhibitors play a key role in slowing their dissociation. We expect the combination of the τRAMD simulation procedure with machine learning analysis to be generally applicable as an aid to target-based lead optimization.

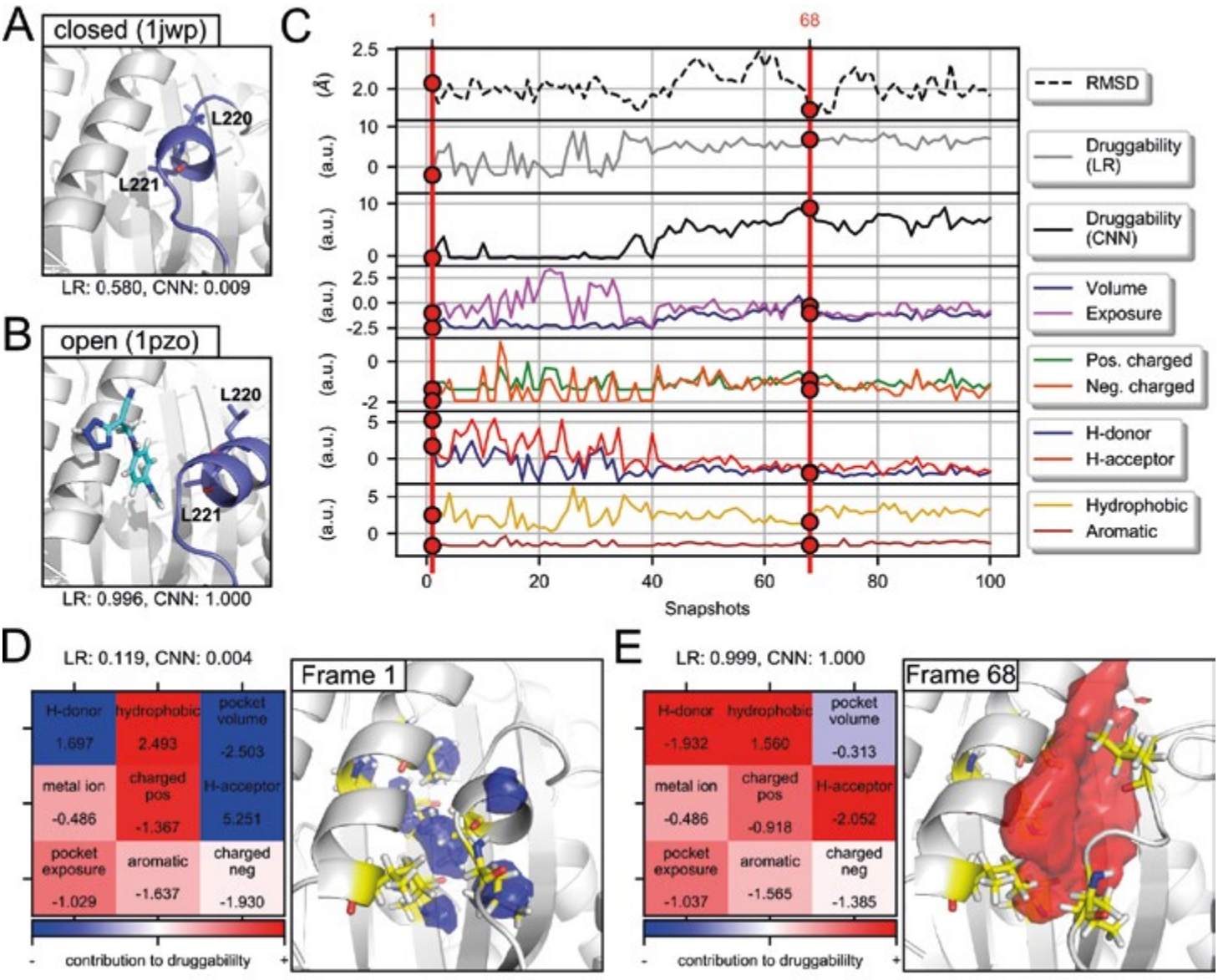


Figure 43: Example of the use of the new TRAPP-based druggability-scoring module to monitor variations in druggability along a molecular dynamics simulation of the enzyme β-lactamase. (A,B) Crystal structures of β-lactamase with (A) a closed and (B) an open transient pocket with a ligand (cyan) bound. (C) Traces of druggability and other physicochemical properties along a simulated trajectory (generated with the L-RIP enhanced sampling molecular dynamics simulation method), beginning from the closed conformation shown in panel A in frame 1. The upper plot depicts the deviation of the structure from the open conformation shown in panel B. Frame 68 is the closest to the structure in panel B and has a high predicted druggability. (D, E) The druggability scores (shown above the heat maps) and the corresponding computed global (TRAPP-LR heat maps) and grid-based LRP (TRAPP-CNN) properties for frames 1 (D) and 68 (E). Red denotes greater druggability, while blue denotes lower druggablity. Reproduced with permission from the American Chemical Society [Yuan, 2020].

Prediction of the druggability of flexible protein binding pockets

A key step in target-based approaches to drug discovery is identifying a pocket on a suitable target protein and assessing its druggability – that is, its potential to bind to a drug-like molecule. While several methods have been developed to compute druggability scores for static structures of proteins, few methods account for the dynamic nature of protein structures. This dynamic nature is important because the druggability of a protein binding pocket may vary due to changes in its shape that arise from the motions of the protein. For example, the crystal structure of a protein might reveal a binding pocket, but the protein might be able to adopt an arrangement in which this pocket has a higher druggability score. If such pocket shapes can be identified, they can be targeted in the design of drugs. We previously developed TRAPP (TRAnsient Pockets in Proteins, trapp.h-its.org) to enable the exploration of different protein conformations, the analysis of binding pocket flexibility and dynamics, and the extraction of spatial and physicochemical information on the binding pocket conformations. In 2019, we used machine learning approaches to develop a method to assess binding pocket druggability in TRAPP [Yuan, 2019].

We developed two statistical models – a logistic regression model (TRAPP-LR) and a convolutional neural network model (TRAPP-CNN) – for predicting druggability and how it varies with changes in the spatial and physicochemical properties of a binding pocket. Both models are integrated into TRAPP. Moreover, the models – which were trained on publicly available and self-augmented datasets – show equivalent or superior performance compared with existing methods on test sets of protein crystal structures and have sufficient sensitivity to identify potentially druggable protein conformations in trajectories from molecular dynamics simulations. Furthermore, visualization of the evidence that supports the decisions of the models in TRAPP facilitates the identification of the factors that affect the druggability of protein binding pockets.

Multiscale molecular modeling

As computing power and software improve, the size of the molecular systems that can be simulated in full atomic detail increases. Nevertheless, it is often necessary to combine atomic-detail modeling with coarse-grained or continuum models in order to study biological systems on relevant spatial and temporal scales. For example, in order to model and simulate cytochrome P450 enzymes in a membrane environment with their reductase protein, we combine Brownian dynamics docking with coarse-grained and atomic-detail molecular dynamics simulations [Mustafa, 2019 a,b], [Mukherjee, 2019].

One of the aims of the EU-supported Human Brain Project (www.humanbrainproject.eu) is to provide a research infrastructure for multilevel modeling and multiscale simulation of the brain. In this project, we focus on combining workflows that employ kinetic modeling to investigate biochemical processes in signaling

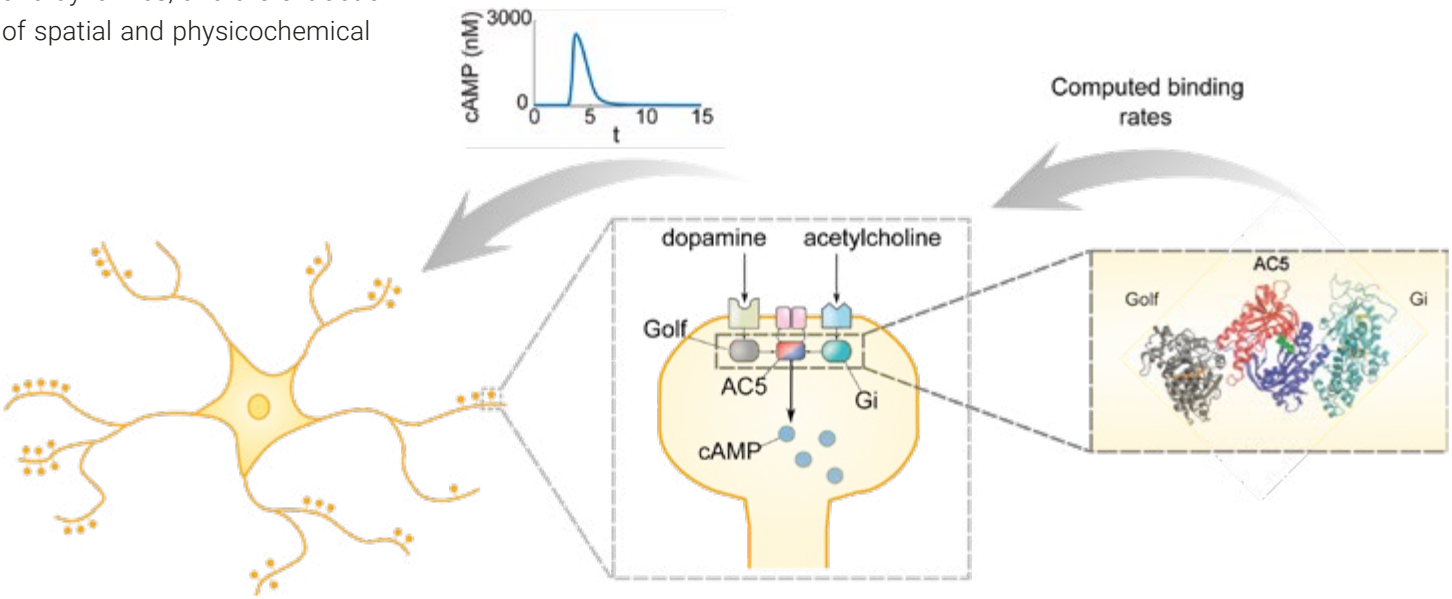


Figure 44: Bridging scales to simulate signaling through a synaptic signal transduction network in order to investigate the molecular mechanisms behind brain plasticity. Protein structure-based simulations (right) were performed to investigate the stability of protein complexes and to compute protein-protein association rate constants. This molecular-level information was used to build a kinetic model of the signal transduction network that includes the regulation of adenylate cyclase 5 (AC5) by various proteins (Golf, Gi) (center). This model revealed how the regulation of AC5 enables the coincidence detection of neuromodulators – such as dopamine and acetylcholine – in striatal neurons (left). [Bruce, 2019b] [Image: Daniel Trpevski, Karolinska Institute, Stockholm]

networks with molecular simulation workflows based on information on protein structures. The molecular simulations enable the prediction of the topology of the networks along with the parameters needed to model them. We employed this multiscale approach in a collaboration with the groups of Jeanette Hellgren Kotaleski (KTH, Stockholm, Sweden), Paolo Carloni (Forschungszentrum Jülich), and Ursula Röthlisberger (EPFL, Lausanne, Switzerland) to study a synaptic signal transduction network that is important for brain plasticity and influences processes such as learning and memory formation.

The strengthening of the corticostriatal synapses in response to stimuli – which is important for reinforcement learning – depends on the second messenger, cAMP, whose synthesis is catalyzed by the enzyme adenylyl cyclase 5 (AC5), which is itself regulated by stimulatory and inhibitory proteins (see Figure 44). To investigate how AC5 can act as a coincidence detector that only promotes cellular responses when convergent regulatory signals occur close in time, a kinetic network model constrained by results from protein-structure-based simulations was built.

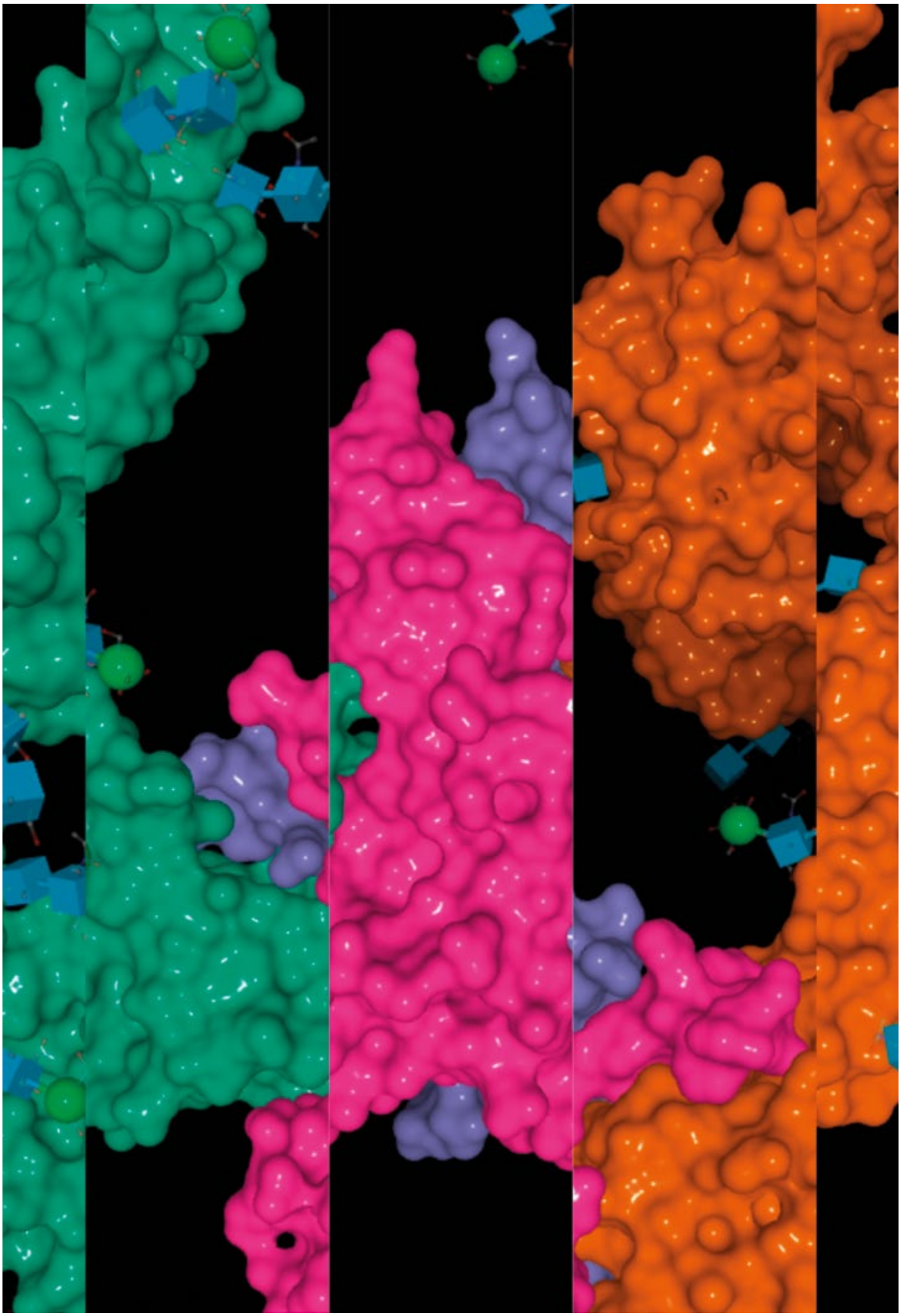
Molecular dynamics simulations were used to investigate the stability and activity of complexes of AC5 with regulatory proteins, and Brownian dynamics simulations were performed to estimate protein association rate constants. The results revealed how the formation of an inactive complex of AC5 with both stimulatory and inhibitory proteins is crucial to the ability of AC5 to detect coincident neuromodulatory signals in striatal neurons [Bruce, 2019b].

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 **Gruppe Molecular and Cellular Modeling (MCM)** 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?

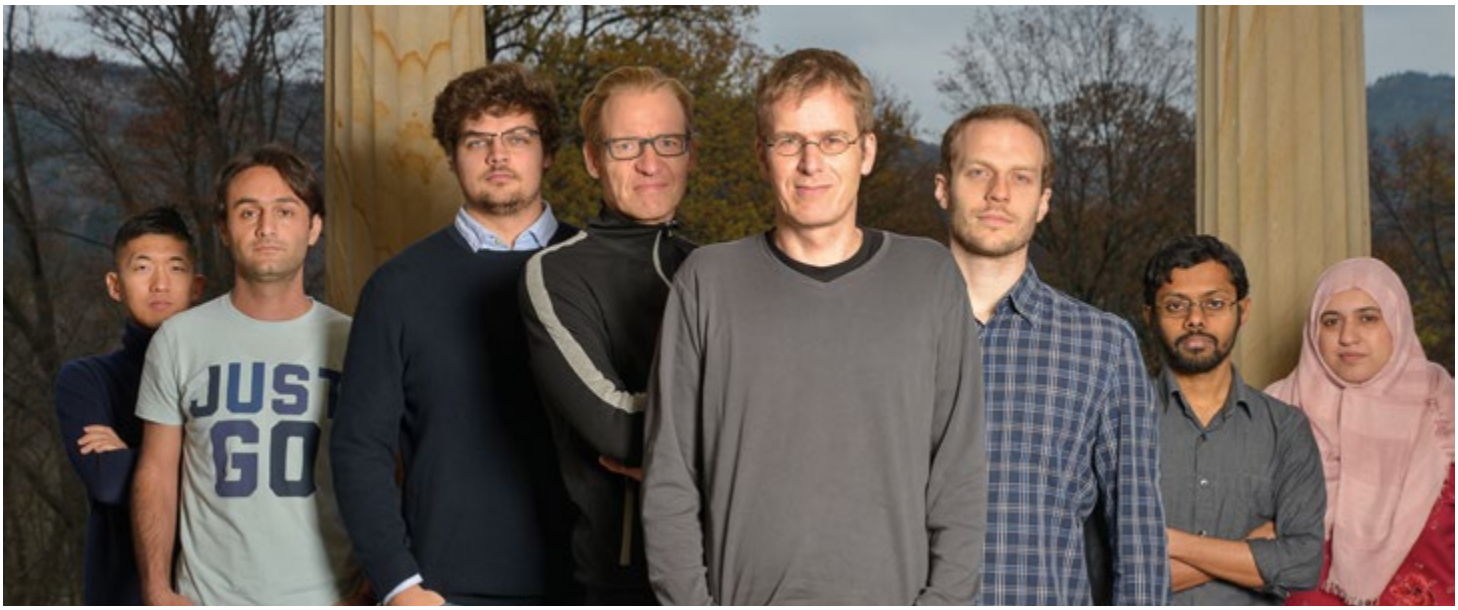
Diese Fragen sind beispielhaft für die Art von Problemen, die wir in unseren Projekten durch die Entwicklung und Anwendung rechnerischer Methoden zur Untersuchung biomolekularer Strukturen, Dynamik, Wechselwirkungen und Reaktionen behandeln. In enger Zusammenarbeit mit Experimentatoren verwenden wir 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 Vorhersage der ‚Druggability' von Protein Bindungstaschen, und (iii) der molekularen Multiskalenmodellierung.



2 Research

2.9 Natural Language Processing (NLP)



Group Leader
Prof. Dr. Michael Strube

Staff members
Benjamin Heinzerling (until March 2019)
Dr. Mark-Christoph Müller

Scholarship holders
Haixia Chai (HITS Scholarship, since April 2019)
Kevin Alex Mathews (HITS Scholarship)
Sungho Jeon (HITS Scholarship)

Visiting scientists
Mehwish Fatima (PhD Student, HEC-DAAD Scholarship)
Federico López (PhD Student, Research Training Group AIPHES)
Ivan Sekulić (Intern, until June 2019)

Students
Nadia Arslan
Jason Brockmeyer
Fabian Düker
Lucas Rettenmeier (since July 2019)

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 2019, Benjamin Heinzerling and Nafise Moosavi successfully defended their theses. Benjamin worked on neural network-based approaches to entity linking and created the resource BPEmb, a collection of tokenization-free pre-trained subword embeddings for 275 languages. BPEmb is particularly popular with researchers who work on low-resource languages. Benjamin is currently a postdoctoral researcher at the RIKEN AIP and Tohuko University in Japan. Nafise's thesis deals with robustness in coreference resolution and demonstrates that current deep neural-network-based approaches to coreference resolution cannot be applied across genres and domains. Based on this insight, Nafise developed methods to overcome these restrictions, thereby making state-of-the-art approaches more robust. She now works as a postdoctoral researcher at the UKP Lab at the Technical University of Darmstadt. The NLP group's intern, Ivan Sekulić,

joined the informatics department at the University of Lugano, Switzerland, as a PhD student. Haixia Chai, another PhD student who had previously worked in industry for several years, and master's student Lucas Rettenmeier, who – interestingly – is working on a master's degree in physics, also joined the NLP group in 2019. Haixia is currently conducting work on coreference resolution, and Lucas is writing his thesis on using word embeddings as a tool to gain insight into linguistic change over time.

The NLP group continued to publish at first-rate conferences, such as ACL, with first-authored papers by Nafise Moosavi and Benjamin Heinzerling in 2019. However, there was more to come: Federico López won a best-paper award at the ACL Workshop on Representation Learning for his work on hyperbolic neural networks for entity linking, and Mark-Christoph Müller won a best-poster- and demonstration award at the Conference on Theory and Practice of Digital Libraries. Finally, Michael Strube was named ACL Fellow by the Association for Computational Linguistics.

A lightweight, fully hyperbolic neural model for hierarchical multi-class classification (Federico López)

"Paul robbed John. He was arrested." What are Paul and John? As an entity that robs and is arrested, we could easily identify Paul as a person. However, this types are not accurate enough for some NLP 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 via more fine-grained information. Thus, more specifically, we could say that Paul is a criminal and John – since he was robbed – is a victim. 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). More recent studies have extended the inventory of types to several thousand. These fine-grained types provide much more useful information that can be implemented in context-sensitive and entity-focused downstream tasks, such as coreference resolution, relation extraction, and question answering.

The key is to find a representation – or a geometry – that leverages the hierarchical structure of the type of inventories for the task of fine-grained entity typing. We therefore use a hyperbolic space as the target vector space in which to represent the data. These hyperbolic spaces have drawn attention in the area of network science as they are appropriate for modeling hierarchical data. Hyperbolic geometry is a non-Euclidean geometry that studies spaces of constant negative curvature and in which tree-like structures can be easily modeled. Hyperbolic geometry can hence easily and simultaneously model the hierarchy in the inventory as well as the semantic similarity among the types.

In our work, we propose using a fully hyperbolic model for multi-class, multi-label classification that performs all operations in hyperbolic space. We evaluate the proposed model on datasets for fine-grained entity typing and compare it with different baselines that operate under Euclidean assumptions. The hyperbolic model shows performance that is on par with state-of-the-art methods

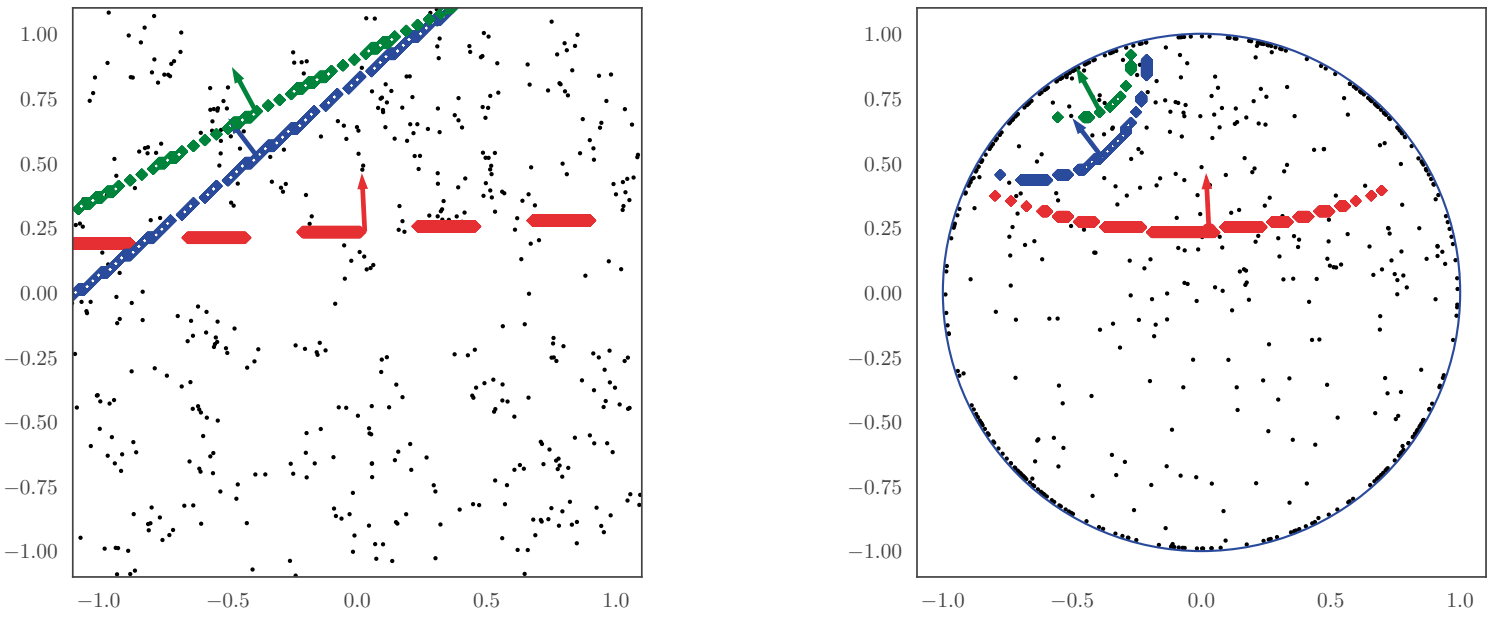


Figure 45: (a and b): Classification hyperplanes projected in 2D for Euclidean (a) and hyperbolic space (b).

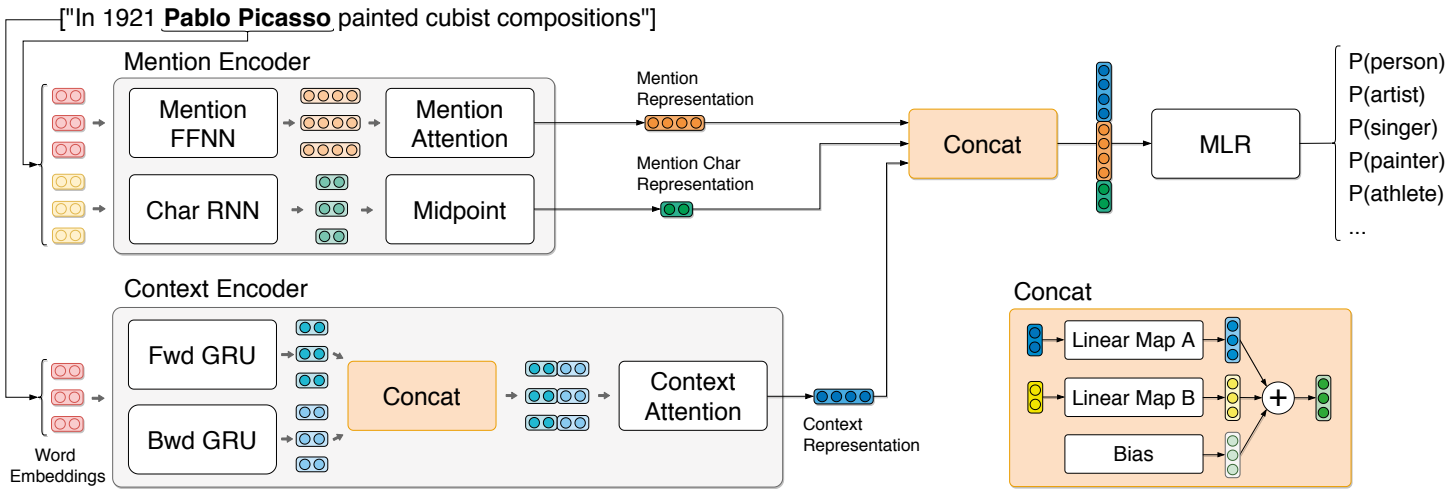


Figure 46: Overview of the proposed neural model for hierarchical multi-class classification.

of fine-grained classification and that displays a remarkable reduction of parameter size due to the efficiency of the hyperbolic representations.

WiMCor: A large corpus of metonymy (Kevin Mathews)

Metonymy is a figure of speech in which an entity is referred to by another, related entity. For instance, the government of a country can be represented metonymically by the name of the country (e.g., Washington for the US government), and an event can be represented metonymically by the location of the event (e.g., Vietnam for the Vietnam war). Metonymy is common in verbal as well as written communication and is also a universal feature across natural languages.

Datasets containing samples of real-world text are important to the study of languages. The existing datasets on metonymy, namely SemEval and ReLocar, are small and do not sufficiently cover the various ways in which metonymy can be observed in real-world text. In addition, the samples in these datasets lack sufficient label granularity. For instance, the location Delft can be labeled as a place or – more specifically – as a city. This is what we mean by label granularity. As a result, there is disagreement over the labels in these datasets.

We developed a new dataset of metonymy called WiMCor (Wikipedia Metonymy Corpus). Our dataset construction mechanism is semi-automatic in nature, with minimal human intervention. The main idea behind our approach is the use of Wikipedia disambiguation pages. These disambiguation pages list different senses of ambiguous entities, such as Washington, and provide links to Wikipedia articles corresponding to each sense of the entity. Through our approach, we identified entities that are ambiguous due to metonymy with the help of commonly observed patterns of metonymy. Subsequently, using these entities, we generated samples from the corresponding Wikipedia pages. The current version of the dataset consists of more than 200,000 samples from the English-language Wikipedia. Each sample contains one or more sentences and three labels of varying granularity: coarse-grained, medium-grained, and fine-grained.

WiMCor represents an improvement over existing datasets in various aspects, such as the number of samples, the length of the samples (see Figure 47), and label granularity and is also a testament to the

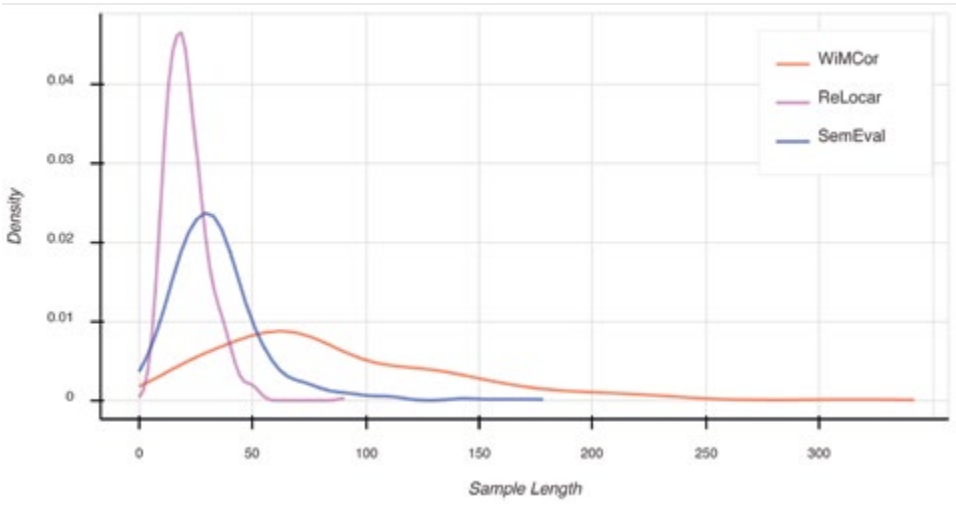


Figure 47.: Comparison of length distributions of 500 samples randomly selected from WiMCor, SemEval, and ReLocar. The length of a sample is the number of words it contains. The samples in WiMCor are substantially longer than those in SemEval and ReLocar.

richness and variety of metonymy. We conducted a manual evaluation of WiMCor to estimate the amount of noise in the dataset and ensure the reliability of the automatically assigned labels. We also used WiMCor to create benchmarks for the task of automatic metonymy resolution.

A major drawback of our dataset is that the samples are generated from Wikipedia entities, which are proper nouns, such as Germany and Heidelberg. However, metonymy is a complex phenomenon that can manifest itself in various ways, including as multi-word expressions and common nouns. Despite this limitation, the new dataset should aid in the study of metonymy and automatic metonymy-resolution systems. Finally, the multilingual nature of Wikipedia offers the possibility to extend WiMCor to many other languages, such as German, Chinese, and Hindi.

NoisyCoref: A benchmark dataset for coreference resolution in noisy text (Haixia Chai)

With the advent of contextualized embeddings, such as ELMo and BERT, neural coreference resolution systems now achieve the highest scores on the datasets introduced by the CoNLL 2012 English coreference resolution shared task. However, these notable improvements to the standard CoNLL evaluation test set cannot persist on noisy user-generated content from social-media platforms, where the text may contain typographical errors, code-switching, and deliberate obfuscation and the vocabulary may mismatch with the training data of the systems (e.g., using antonyms or hypernyms).

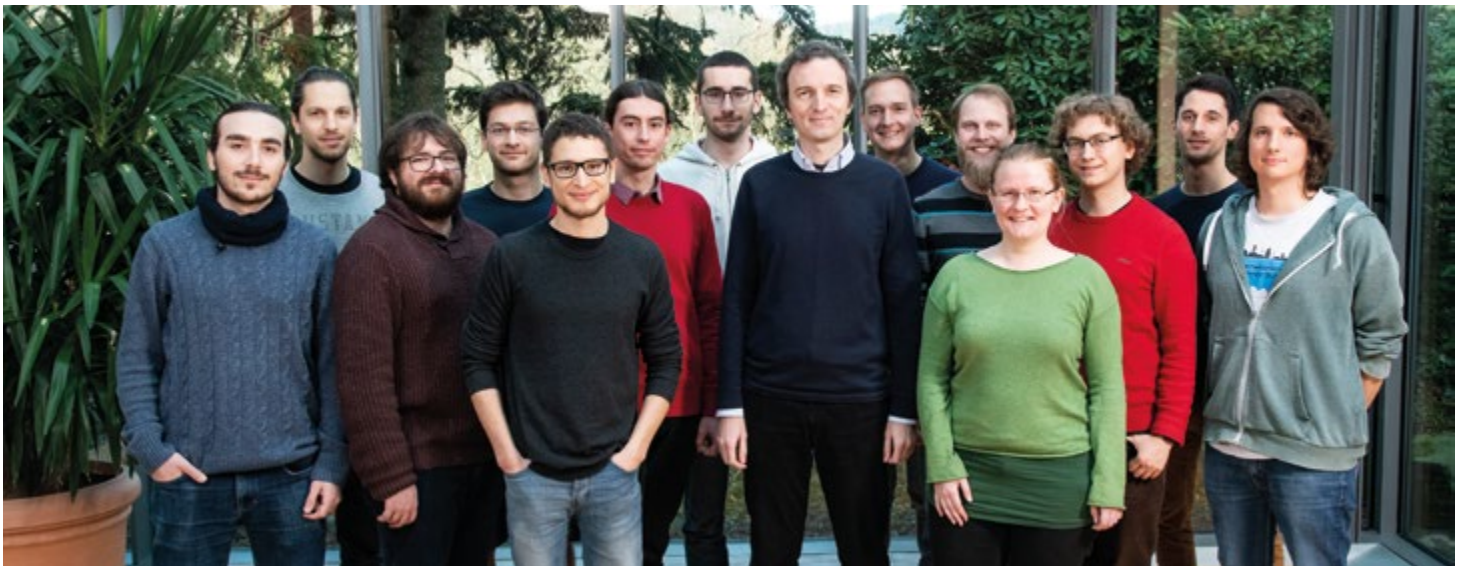
In our work, we propose using an artificially created, large-scale adversarial benchmark dataset that covers a wide range of noise phenomena (e.g., typographical errors, word replacement, and code-switching). Although this dataset contains only perturbations that are only barely perceptible by humans, our findings reveal that existing off-the-shelf coreference resolution systems perform dramatically worse, which indicates that this new benchmark can provide a testbed for systems concerning the handling of noisy user-generated text.

We compare existing coreference resolution systems via our benchmark dataset, including non-neural, statistical, and neural systems. Experiments have demonstrated that current state-of-the-art system performance drastically decreases when confronted with noisy data. We have also shown that different attack scenarios may have different impacts on various systems, which indicates that they are vulnerable in various ways, with code-switching and visual attacks causing the greatest damage. Our benchmark dataset provides a testbed on which to foster future research with the goal of making coreference resolution systems more robust.

Natural Language Processing (NLP) ist ein interdisziplinäres Forschungsgebiet, das mit Methoden der Informatik linguistische Fragegestellungen bearbeitet. Die NLP Gruppe entwickelt Methoden, Algorithmen und Tools zur automatischen Analyse von Sprache. Sie konzentriert sich auf die Diskursverarbeitung und verwandte Anwendungen, wie zum Beispiel automatische Zusammenfassung und Lesbarkeitsbewertung.
Im Jahr 2019 verteidigten Benjamin Heinzerling und Nafise Moosavi ihre Dissertationen. Benjamin wendete Neuronale Netzwerke auf die computerlinguistische Aufgabe des Entity Linking an. Dabei erstellte er die Ressource BPEmb, die Word Embeddings für 275 Sprachen zur Verfügung stellt und sich insbesondere als nützlich für Sprachen mit wenig existierenden Ressourcen herausgestellt hat. Benjamin ist jetzt Postdoc bei RIKEN AIP und der Tohuko University in Japan. Nafises Doktorarbeit erforscht die Robustheit von Koreferenzresolutionssystemen. Sie zeigt, dass existierende Koreferenzresolutionssysteme, die auf tiefen neuronalen Netzwerken basieren, nicht auf neue Genres und Domänen angewendet werden können. Deshalb entwickelte sie für ihre Dissertation Methoden, um Koreferenzresolutionssysteme robuster zu machen. Nafise ist jetzt Postdoc im UKP Lab an der Technischen Universität Darmstadt. Ivan Sekulić, Praktikant in der NLP Gruppe bis Juni 2019, ist jetzt Doktorand an der Universität Lugano. Doktorandin Haixia Chai, die zuvor in der Industrie arbeitete, und Master Student Lucas Rettenmeier – interessanterweise wird er seinen Abschluß in Physik machen – sind neue Gruppenmitglieder. Haixia arbeitet über Koreferenzresolution während Lucas Word Embeddings verwendet, um historischen Sprachwandel zu untersuchen.
Die NLP Gruppe publizierte im Jahr 2019 wieder bei erstklassigen Konferenzen wie der ACL, so mit zwei Papieren deren Erstautoren Nafise Moosavi und Benjamin Heinzerling waren. Das war aber nicht alles: Federico López gewann den Best Paper Award beim ACL Workshop über Representation Learning, während Christoph Müller den Best Poster and Demonstration Award bei der Konferenz über „Theory and Practice of Digital Libraries" gewann. Und schließlich wurde Michael Strube zum ACL Fellow von der Association for Computational Linguistics ernannt.

2 Research

2.10 Physics of Stellar Objects (PSO)



Group Leader
Prof. Dr. Friedrich Röpke

Staff members
Dr. Róbert Andrássy
Dr. Johann Higl (since October 2019)
Javier Morán Fraile (since October 2019)

Scholarship holders
Leonhard Horst (HITS Scholarship)
Christian Sand (HITS Scholarship)

Visiting scientists
Dr. Fabian Schneider (Gliese Fellowship)
Sabrina Gronow
Florian Lach
Theodoros Soultanis (IMPRS-PhD-Student at MPIA Heidelberg)
Maryam Modjaz (until August 2019)

Students
David Bubeck (since May 2019)
Manuel Kramer (until October 2019)
Melvin Moreno
Patrick Ondratschek (since October 2019)

"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 in 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 electro-magnetic spectrum and are fundamental building blocks of galaxies and larger cosmological structures. With the help of extensive numerical simulations, the 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 thermo-nuclear 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 that take place 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. Using 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.

The irresistible pull: When massive stars collide

All neutron stars are magnetic, but some are more magnetic than others. These more magnetic neutron stars are the so-called magnetars, the strongest magnets in the Universe. The reason for their exceptionally large magnetic field is most likely that they formed in supernovae of already highly magnetized stars. But how do these massive stars acquire their large magnetic field? With the help of large computer simulations, researchers from Heidelberg, Oxford, and Garching developed a model that demonstrates that these strong fields can be formed in stellar mergers. The results were published in the scientific journal ***Nature*** in its issue from 9 October 2019 [Schneider et al., 2019].

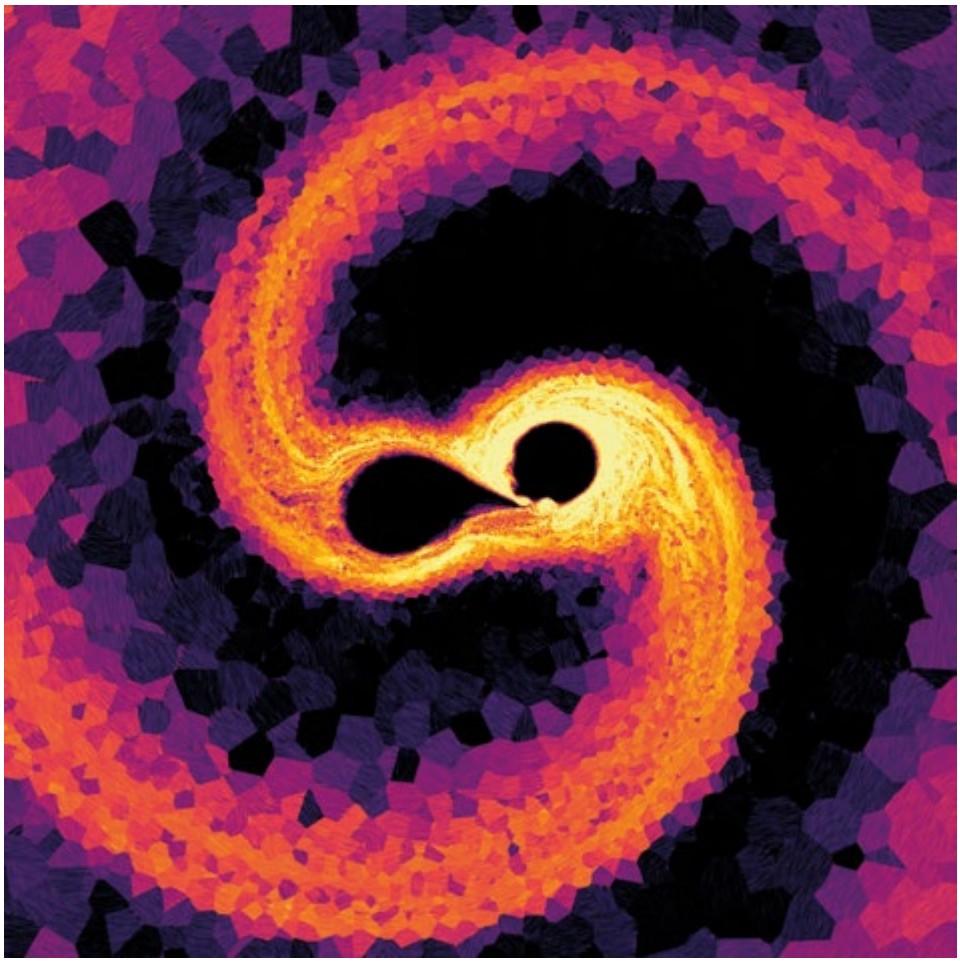


Figure 48: Birth of a magnetic star – The simulation shown marks the birth of a magnetic star such as Tau Scorpii. The image is a cut through the orbital plane, with the coloring indicating the strength of the magnetic field and the hatching representing its field lines. (Credit: Ohlmann/Schneider/Röpke)

Our Universe is threaded by magnetic fields. For example, the Sun has a turbulent envelope in which its magnetic field is continuously generated. However, more massive stars do not have such an envelope and are therefore not expected to show magnetic surfaces. Nevertheless, about 10 percent of these more massive stars have a strong, large-scale surface magnetic field whose origin has eluded scientists since its discovery in 1947 by Babcock. It is these stars that are believed to form highly magnetic neutron stars when they explode in supernovae.

As early as over a decade ago, strong magnetic fields were suggested to be capable of being produced from the collision of two stars. Thus far, however, researchers have not been able to test this hypothesis because they have not had the necessary computational tools. However, thanks to the novel AREPO code, developed by Volker Springel and his group during their time at HITS, it is now possible to simulate the merger of two massive stars. The simulations were set up and run by the PSO group on the computing clusters of HITS and reveal that a strong magnetic field is indeed produced due to the strong shear and the large turbulence present in the merger of the two stars (Figure 48). Stellar mergers occur frequently, and about 10 percent of all massive stars in the Milky Way are thought to be the product of such mergers – a good match with the occurrence rate of magnetic stars.

When stars merge, they appear younger than they really are. This phenomenon is well known, and such stars are called blue stragglers.

In 2016, the magnetic star Tau Scorpii (τ Sco) was discovered to be a blue straggler that appears anomalously young as compared with other stars in its birth stellar association. If τ Sco is indeed the product of a merger, this formation process would nicely explain the anomaly. Moreover, it has been suggested that τ Sco may have obtained its strong magnetic field as a result of the merger, and the new simulations demonstrate exactly this finding.

At the end of its life, τ Sco will explode in a supernova when its core collapses, and it will most probably leave behind a highly magnetized neutron star. Such magnetars are thought to have the strongest magnetic fields in the Universe – up to one hundred million times stronger than the strongest magnetic field ever produced by humans. Indeed, the new simulations by the PSO group reveal that the generated magnetic field could be sufficient to explain the exceptionally strong magnetic fields inferred to exist in magnetars. This hypothesis renders the merger model a promising option for explaining the origin of such extremely strong magnetic fields.

An explosive result

Nuclear physicists from Denmark, Finland, Germany, and the United States succeeded in experimentally determining nuclear processes that operate at conditions ten million times denser and 25 times hotter than those at the center of our Sun. A result of the measurement is that intermediate-mass stars are very likely to explode and not – as assumed until now – to collapse, which was confirmed with three-dimensional computer simulations by members of the PSO group together with Dr. Samuel Jones, a PSO alumnus now working at the Los Alamos National Laboratory, USA. The findings were published in the scientific magazine Physical Review Letters [O. Kirsebom et al., 2019].

Depending on their mass, stars take different evolutionary paths. Low-mass stars, such as our Sun, eventually become white dwarfs. Massive stars, by contrast, terminate their lives in spectacular explosions known as supernovae, leaving behind either a neutron star or a black hole. But where is the dividing line between these two options? The fate of intermediate-mass stars, which weigh 7–11 times as much as the Sun, has remained unclear, which is surprising since intermediate-mass stars are prevalent in the Galaxy. The question has been addressed in previous simulations of the PSO group (see, e.g., the HITS Annual Report 2016).

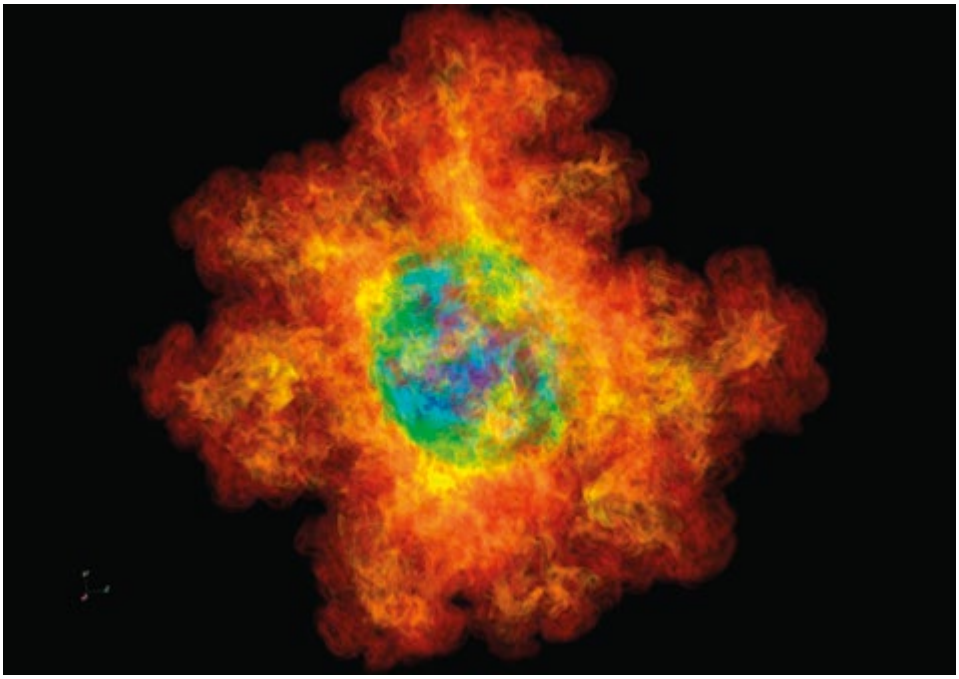


Figure 49: Snapshot from a three-dimensional simulation of a thermonuclear explosion in a high-density Chandrasekhar-mass oxygen-neon white dwarf star. The electron fraction of the material is color-coded, with bluish colors indicating highly neutronized material (image credit: Samuel Jones).

The fate of intermediate-mass stars depends on a small detail, namely how readily electrons are captured on the isotope neon-20 in the stellar core. Depending on this electron capture rate, the star will either be disrupted in a thermonuclear explosion or collapse to form a neutron star. In 2014, theorists pointed out that the strongly suppressed and hence previously ignored and experimentally unknown transition between the ground states of neon-20 and fluorine-20 was the final piece of information needed to determine the electron capture rate at the astrophysical conditions reached in intermediate-mass stars. Now, an international collaboration of physicists has succeeded in obtaining the first accurate determination of this important rate through a combination of careful measurements of the beta-decay of fluorine-20 on the one hand and theoretical calculations on the other. The experiment took place under conditions far more peaceful than those found in stars, namely at the Accelerator Laboratory of the University of Jyväskylä, Finland. The measurements found a surprisingly strong transition between the ground states of neon-20 and fluorine-20 that leads to electron captures on neon-20 that occur at lower densities than previously believed. What does this imply for the fate of the star? A detailed three-dimensional simulation similar to that shown in Figure 49 was performed by the Physics of Stellar Objects Group at HITS and by a group at Los Alamos.

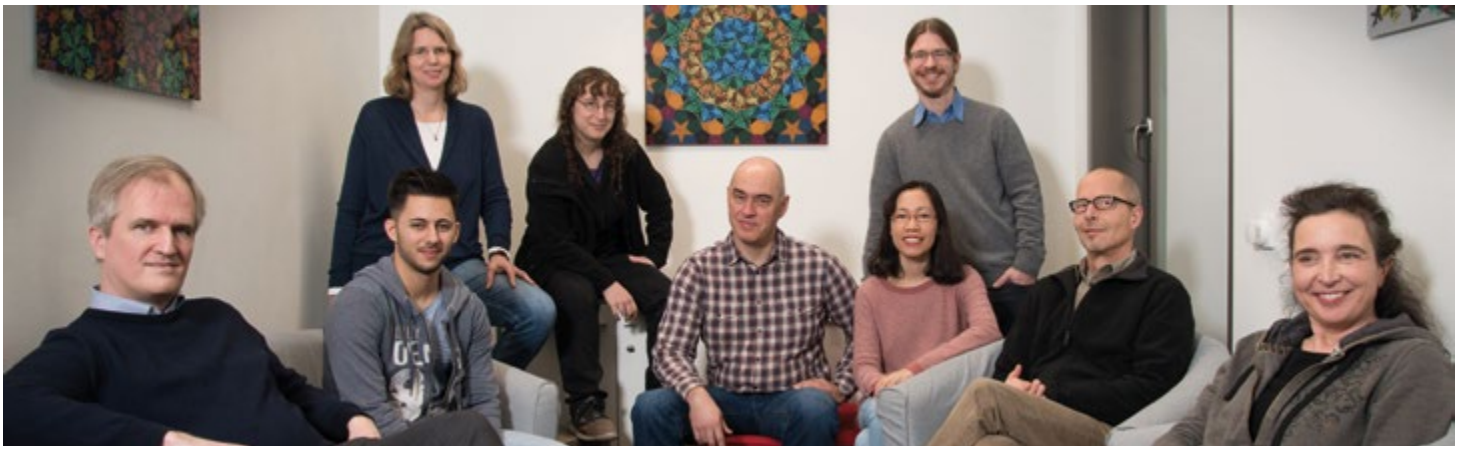
„Wir sind Sternenstaub" – die Materie, aus der wir geformt sind, ist zum großen Teil das Ergebnis von Prozessierung des primordialen Materials aus dem Urknall. Alle schwereren Elemente stammen aus der Nukleosynthese in Sternen und gigantischen stellaren Explosionen. Wie dieses Material gebildet wurde und wie es sich im Universum verteilt, stellen für Astrophysiker fundamentale Fragen dar.
Sterne sind fundamentale Bausteine von Galaxien und aller größeren kosmologischen Strukturen. Gleichzeitig machen stellare Objekte das Universum für uns in astronomischen Beobachtungen überhaupt erst sichtbar. Sterne scheinen im optischen und anderen Teilen des elektromagnetischen Spektrums. Am Ende ihrer Entwicklung kollabieren massereiche Sterne zu Neutronensternen oder Schwarzen Löchern. Eine Verschmelzung solcher kompakten Objekte wurde kürzlich mit Hilfe von Gravitationswellen beobachtet, die ein neues Fenster für astronomische Beobachtungen des Universums öffnen.
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.
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 unsere 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.
Eine weitere Komplikation, die in klassischen Sternentwicklungsmodellen nur sehr grob angenähert werden kann, ist die Binarität. Wohl wegen des Beispiels unserer Sonne tendieren wir oft dazu, Sterne als isolierte Objekte zu sehen; tatsächlich findet man die meisten von ihnen jedoch in Systemen mit zwei oder sogar mehr Sternen. Einige von diesen wechselwirken miteinander, und das hat weitreichende Auswirkungen auf ihre weitere Entwicklung. Solche Interaktionen sind inhärent mehrdimensional und können in klassischen Modellen nicht konsistent behandelt werden. Die PSO-Gruppe führt dreidimensionale Simulationen zu stellaren Wechselwirkungen durch, um neue Einsichten in diese entscheidenden Phasen der Entwicklung von Sternsystemen zu gewinnen.
Das dritte Forschungsfeld der PSO 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 strömungsdynamische 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.

The simulation demonstrates that in contrast to previous beliefs, the star is more likely to be disrupted by a thermonuclear explosion than to collapse into a neutron star.

The new result is the final piece of the puzzle required to understand the nuclear processes that take place in the interior of the star during the last moments of its life. Since thermonuclear explosions eject much more material than do those triggered by gravitational collapse, the results have implications for galactic chemical evolution [Jones et al, 2019a]. The ejected material is rich in titanium-50, chromium-54, and iron-60 [Jones et al, 2019b]. Therefore, the unusual isotopic ratios of titanium and chromium found in some meteorites – in addition to the iron-60 discovered in deep-sea sediments – could have been produced by the explosion of an intermediate-mass star from our galactic neighborhood in either the distant (billions of years) or the not-so-distant (millions of years) past.
In the light of these new findings, the most probable fate of intermediate-mass stars seems to be a thermonuclear explosion, which produces a subluminous Type Ia supernova and a special oxygen-neon-iron white dwarf. Detection or non-detection of such white dwarfs in the future would provide important insights into the explosion mechanism.

2 Research

2.11 Scientific Databases and Visualization (SDBV)



Group Leader
PD Dr. Wolfgang Müller

Staff members
Dr. Dorotea Dudaš
Dr. Sucheta Ghosh
Martin Golebiewski
Xiaoming Hu
Dr. Olga Krebs
Ina Pöhner
Dr. Maja Rey
Natalia Simous (since April 2019)
Dr. Andreas Weidemann
Benjamin Winter (since July 2019)
Dr. Ulrike Wittig

Student
Marcel Petrov (until September 2019)

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). More about FAIR see below.
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.

What is FAIR?

Regular readers of the HITS Annual Report will be familiar with the term "FAIR data." FAIR stands for our target properties when maintaining scientific data, namely that data should be Findable, Accessible, Interoperable, and Reusable. FAIR data is actually quite similar to money. We want money to be findable, for example, by knowing that it is in the bank. We also want it to be accessible, meaning that we need to have a means of getting at it, typically via standardized methods, such as with an ATM and a credit card. Once we have our money, we want it to be interoperable, which means that if we want to purchase a train ticket at the railway station, we need the ticket machine to use the properties of money, such as the size and shape of coins. Finally, we have license to reuse our money, meaning that once we have money, we can reuse it however we wish. One interesting anecdote about money involves Dr. Brian May, astrophysicist and guitarist of the rock band Queen, who uses sixpence coins to pick the strings of his guitar. This quirk also highlights the challenges of reusing data and procedures as the coins that Brian May uses are no longer minted.

Leaving the realm of money and music behind, we now turn to translation: The findability of data is assured by storing it in repositories that provide search- and long-term, unique identifiers. Accessibility mainly means that data can be accessed via a standard protocol (e.g., FTP or HTTP) as opposed to via regimes such as "send an email to the corresponding author and then wait for an unspecified amount time." Interoperability mainly concerns data description: Are the methods, the conditions, and the data format itself described such that the data can be combined with other data? Reusability concerns the license that is attached to the data description. Without a license, it is not possible to know whether the data can be reused. All FAIRness criteria are useless without a license.
Measuring data FAIRness is a current topic of research since F and A as well as the presence of useful licenses are easily testable. However, I and some aspects of R largely depend on the quality of the data description (metadata), which depends on expectations. Exactly what metadata can be considered good enough, and what metadata are too strict? There are ways that are using standards as a basis.
We now turn to the topic of data management. Someone has to make the data FAIR, which is ideally done by the person who generated the data through measurement or by the person who created the model. Below, we present two tools that we developed in 2019 to help make data FAIR.

Aiming at FAIR data collection with RightField

Given that we have a rough understanding of what we want to achieve, the question is: How do we make data FAIR? The way to do so is (i) to facilitate the maintenance of a structure and (ii) to enable the addition of metadata to the original data (i.e., by annotating data to additional information).
The Scientific Databases and Visualization (SDBV) group has been working on two tools designed to facilitate data FAIRness.

RightField has been part of the FAIRDOM project's platform for some time. The FAIRDOM project (FAIR Data, Operations, and Models) currently involves partners from the UK, Germany, Switzerland, Belgium, Norway, and the Netherlands, with the UK, Germany, and Switzerland serving as funded partners. Within the project, RightField enables us to "hide" ontology terms within Excel workbooks, thereby solving a recurrent problem: Despite the fact that most non-computer scientists may have heard of ontologies, they generally do not know how to use them. RightField enables scientists to use ontologies without needing to know about them.

As an example, let us consider an experimentalist, Alice, who wants to share data in a FAIR manner. Alice normally uses Excel files to sum up experiments. A data curator, Bob, sets up a RightField template, which is an "empty" data file that has a table structure adapted to Alice's experiments. The RightField template also contains selection lists, with each selection list containing terms of an ontology.

An ontology is a formal representation of knowledge. In the present context, we are mostly interested in taxonomies (i.e., hierarchies of concepts). The selection lists in the RightField template correspond to lists of ontology terms, which have a standardized meaning; therefore, instead of using descriptions that are only readable by humans, the RightField template produces descriptions that are readable by both machines and humans alike.
This brings us to our contribution: Within the FAIRDOM project, RightField has thus far enabled only the selection of subtrees of taxonomies, which could translate to extremely long selection lists. We thus added functionality to select single items, thereby increasing the number of ontologies that can be used productively within RightField templates.

Our addition was solicited and is part of a larger endeavor: a common FAIR data use case with the BioInfra.Prot team in Bochum. The BioInfra.Prot de.NBI service center provides a statistics consulting service through which scientists can share statistical evaluations with BioInfra.Prot and

inquire as to whether the evaluations are correct or request a statistical evaluation of a given experiment. In order to document the scientists' work, improve their processes, and make their results more reproducible, statistical ontologies and a documentation format designed with the help of RightField are used.

INCOME and template checking

RightField is the method of choice when a template is already known at the time of data creation. Filling in a pre-existing form is the easiest way to adhere to the agreed form, and RightField enables the easy addition of ontology terms. However, what should Alice do when she has a great deal of data and wants to adhere to a new template?

In this case, Alice should transform her datasheet. To do so, she can use the OpenRefine tool. We did not contribute to the development of this tool, and a more detailed description would thus lie outside the scope of the present report.

However, once Alice has transformed her data into her chosen template, a new problem arises that is again recurrent in the life sciences: Many wet-lab biologists and programmers have a different understanding of the meaning behind "adherence to a template," and these numerous differences that can often only be understood with training. In our work with actual users, we are sometimes confronted with a situation that requires a bit of communication until a data file actually matches a given template.

This situation is unfortunate on two levels: (i) It requires a lot of time to deal with it, which (ii) leads to a negative atmosphere between the receiver of the data on the one hand and the generator of the data on the other hand.

A tool that checks for compliance and generates advice on how to fix this would remove any negativity from the interaction and be much more precise than a human could be. We have thus begun the development of such a tool as part of the INCOME project. INCOME aims to bring the community together by making models more commensurable.
We launched the tool at the INCOME Hackathon, which was held in October 2018, and continued its development in March 2019 (in Frankfurt) as well as in November 2019 (in Berlin). The tool has several requirements: It should be

- platform-independent,
- simple to install in user space,
- simple to develop, maintain, and extend, and
- simple to use.

Platform independence and simple installation were solved by basing our work on Anaconda, a Python distribution that exists for Apple, Linux, and Windows. Anaconda can be installed by "normal" users without administrative rights, thereby rendering Python is a convenient development language.

SABIO-VIS: Transforming SABIO-RK tabular results into a visual experience

SABIO-RK is our database for reaction kinetics data. It 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 underlying computational modeling in order to create models of biochemical reactions and complex biological networks.

"Manual curation" refers to the idea that humans make data FAIR. Most of the time, this process involves extracting data from the literature, but data can also be uploaded directly from the lab or collected from other resources via the standard data-exchange format SBML (Systems Biology Mark-up Language).

The quality of the data in SABIO-RK therefore significantly depends on the quality of its sources. Advances in using STRENDA-DB as a basis for some SABIO-RK data were discussed in the Annual Report 2018. This year's installment focuses on the visualization of SABIO-RK data.

SABIO is an acronym for System for the Analysis of BIOlogical pathways. The suffix "-RK" refers to Reaction Kinetics, which is the main purpose of SABIO-RK. Pathway data are highly interlinked networks of chemical reactions, each of which can be described by substrates, products, enzymes, rates, and conditions. SABIO-RK thus inherently stores graphs with multi-dimensional labels. For several years, SABIO-RK has offered searches in which some results are visualized and plots are clickable in order to enable users to refine their searches. SABIO-VIS, a BMBF-funded project at HITS that was launched in 2019, goes far beyond the limits of SABIO-RK by combining tools devised for the exploration of multi-dimensional data in order to enable the visualization of multi-dimensional data such that they can be interactively explored. Multi-dimensional exploration has two sets of advantages: On one hand, it enables improved data discovery. This feature is particularly interesting if exact matches do not exist in the database and only similar matches can be found. On the other hand, multi-dimensional exploration also enables the detection of patterns as well as outliers.

One challenge to multi-dimensional exploration is its efficiency as it is necessary to break down what is needed for visualization such that corresponding queries can be run quickly. A second challenge is the variable number of parameters that describe items in SABIO-RK. A non-technical challenge involves the highly complex architecture of the code and integrating it into another, long-living code.

The example below depicts SABIO-RK data visualized across a variety of dimensions. All graphs show the same data, and hovering over something in one graph highlights the same SABIO-RK item in the corresponding graph. This technique – called brush-and-link – is key to browsing multi-dimensional data.

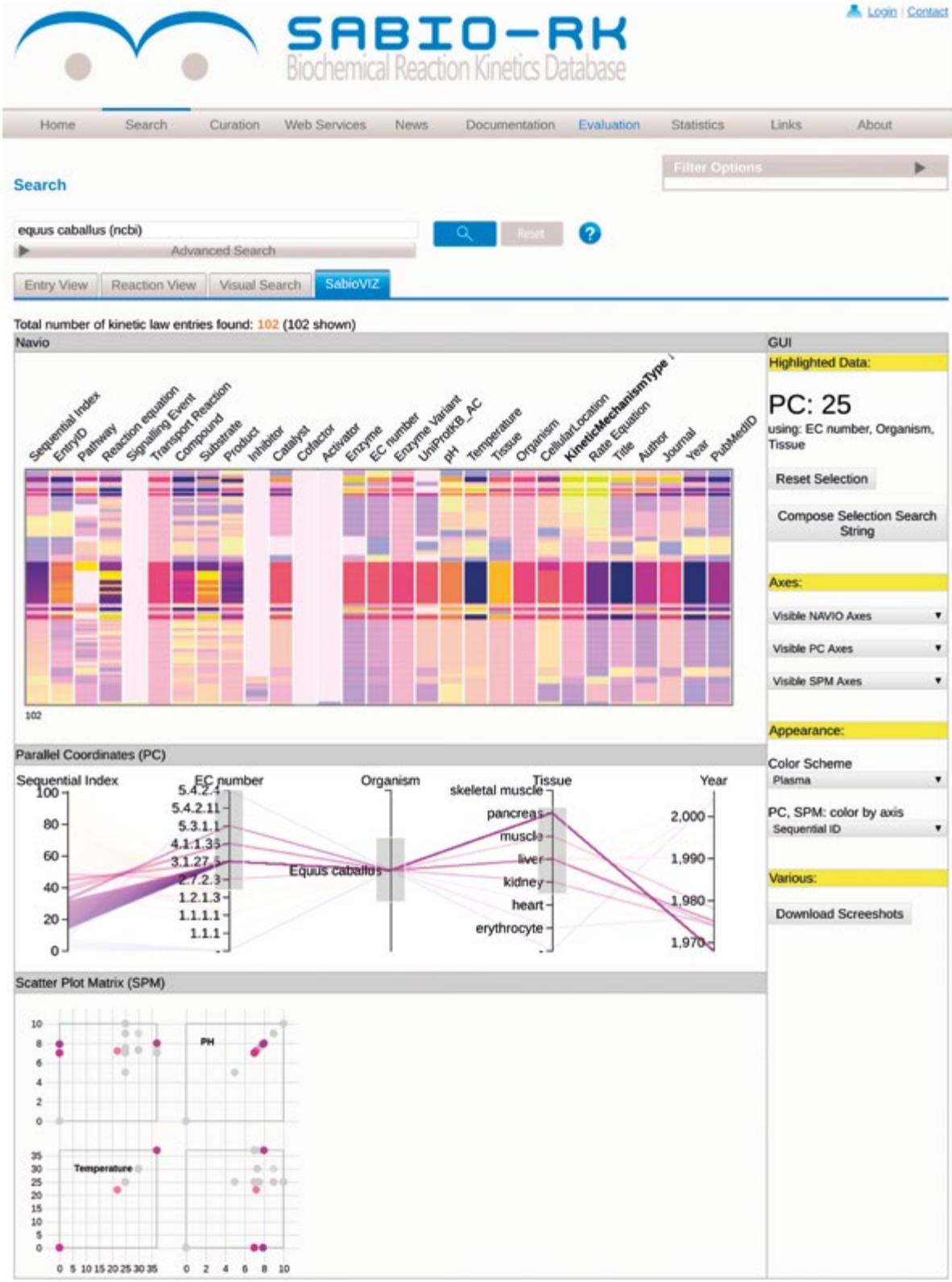


Figure 50: The SABIO-RK interface with data visualized across a variety of dimensions.

In our prototypes, we combine heat maps, parallel coordinate plots, and linked scatter plots. Heat maps are basically tables in which values are replaced by colors that correspond to categorical values or number ranges. Parallel coordinate plots visualize each point in multi-dimensional space as a polyline that spans parallel axes. Parallel coordinate plots are highly useful for exploring and restricting result sets as they enable regions on each axis to be marked simply, which limits the result to the lines that pass through the region. Scatter plots are two-dimensional graphs that display two-dimensional points directly. Linked scatter plots are groups of such plots that belong together. Each scatterplot makes use of two dimensions through which each data point can be visualized.

In combination, these plots offer intuitive insights into the results of SABIO-RK queries.

Aiming at a FAIR future for data services at HITS

Above, we described some of our goals regarding the collection and use of FAIR data. Our work in this field forms part of national and transnational initiatives that aim to create an integrated, interoperable portfolio of services for scientists. The value of a service thus lies not only in the service itself but also in the ways it can be combined with other services. We are therefore currently working on two groups of activities:

SABIO-RK enables reaction kinetics data to be found and interoperated with SBML-enabled tools as it exports data in SBML. SABIO-RK respects the MIRIAM guidelines (Minimal Information Required In the Annotation of Models) and thereby ensures that the data are in line with the COMBINE set of standards (COMBINE = COmputational Modeling in Biology NEtwork). The annotation to relevant data sources and codes (e.g., the EC codes or UniProt) links the data to relevant ontologies and data sources. The actual linking is performed and verified by professional curators.

The FAIRDOM project seeks to build a pathway from experimental data collection to data submission and to large-scale, specialized forms of data collections, such as UniProt. The main challenge in this process involves lowering the entry bar to FAIR data creation and verifying FAIRness. The difficulty in meeting this challenge lies in the fact that data and data descriptions (metadata) are highly diverse. Significant differences thereby exist in the employed methods and thus also in describing them. We continue to investigate how to build tooling that is flexible and powerful enough yet does not require many months of configuration by users.

Our services are typically embedded in a pipeline that spans data collection, data storage, metadata collection and creation, analytic processing, and data publication.

We keep this specific goal in mind in our participation in ELIXIR CONVERGE.

Within Germany, the key topics of the coming years will be aiming to sustain (i) de.NBI services (de.NBI is the German Network for Bioinformatics Infrastructure) and (ii) the German National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur, NFDI), both of which are in flux. However, due to our work in the past 10 years, we are well-positioned to play a strong role in both endeavors.

Das Ziel der **Scientific Databases and Visualization Gruppe (SDBV)** 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 und in der Entwicklung von Werkzeugen.

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.

3 Centralized Services

3.1 Administrative Services

The HITS administration supports the scientific groups in nearly all administrative tasks. It takes care of day-to-day operations at both HITS locations, manages human resources and accounting, clarifies legal issues, and assists the communications team in organizing events.

HITS encompasses two locations (its headquarters at Schloss-Wolfsbrunnenweg and a location at the Mathematikon in Neuenheimer Feld) and three buildings, in which eleven research groups were active at the end of 2019. Together with its non-academic staff, HITS employs a total of 122 people, 37% of whom are women.
In April 2019, the new junior group "Computational Carbon Chemistry" began its work under the leadership of computational chemist Ganna Gryn'ova. The HITS administration made diligent preparations for the group so that Ganna could concentrate fully on building it up.

At the same time, preparations were underway to establish a new group in the field of "Computational Astrophysics" that is linked to a professorship at Heidelberg University. A joint appointment committee consisting of scientists from the Department of Physics and Astronomy at Heidelberg University, the Center for Astronomy of Heidelberg University (ZAH), and HITS met with the top candidates for a two-day symposium at the Studio Villa Bosch and on the HITS campus. The new group leader will join HITS in the course of 2020. The administration will also make all necessary preparations for the start of this group.

Group Leader
Dr. Gesa Schönberger

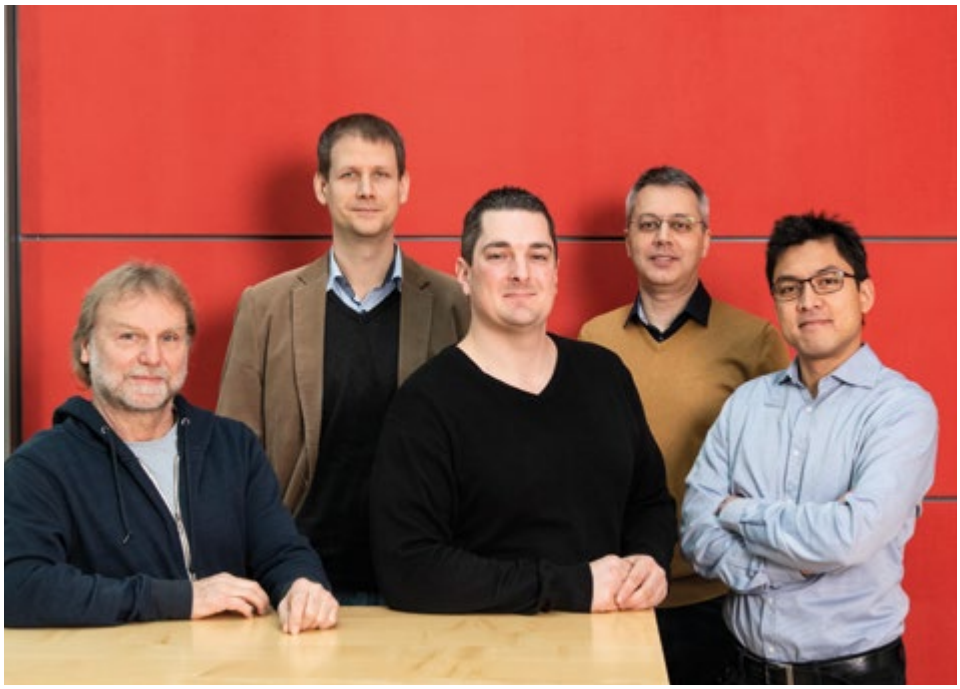
Staff members
Christina Blach (office)
Frauke Bley (human resources)
Christina Bölk-Krosta (controlling)
Benedicta Frech (office)
Ingrid Kräling (controlling)
Thomas Rasem (controlling)
Rebekka Riehl (human resources and Assistant to the Managing Director)
Stefanie Szymorek (human resources)
Irina Zaichenko (accounting)

Student
Lilly Börstler

One major innovation in 2019 was the introduction of the "Welcome Day." Three times a year, the new employees who have joined the Institute are comprehensively informed about the structures, offers, and key contact people at HITS in addition to the Institute's involvement with the Klaus Tschira Group.

Some additional innovations also became necessary last year due to governmental regulations, including the provisions of the Posted Workers Directive for activities in other European countries, the General Data Protection Regulation, and the Trade Secrets Act ("Geschäftsgeheimnisgesetz"). The HITS administration has always worked on all of these innovations with the aim of providing HITS scientists with the best possible support and as little red tape as possible in order to enable them to fully concentrate on their research.

3.2 IT Infrastructure and Network



Group Leader

Dr. Ion Bogdan Costescu

Staff members

Dr. Bernd Doser | Senior Software Developer
Cristian Huza | System Administrator (until July 2019)
Norbert Rabes | System Administrator
Andreas Ulrich | System Administrator
Taufan Zimmer | System Administrator
Christian Zimmermann | System Administrator (February – April 2019)

Following up on work that began in 2018, we further increased the reliability of our connection to the regional academic BelWü network by replacing the central routers with a better-performing, redundant setup, that also provides reserves for the future. We will round out this project by bringing into service a new underground fiber connection, which is planned for the very beginning of 2020. HITSters' access to the Internet as well as other scientists' access to the information and services hosted by HITS should thus be protected against most kinds of network failures.

One such source of information provided by HITS is our website, which has grown quite complex over the years. In order to better serve the needs of the scientific groups while simultaneously improving performance and security, we handed over the technical maintenance to a specialized company. The website was migrated to a new WordPress setup and relaunched with a new design. As part of this project, between the beginning and early summer of 2019, we were heavily involved in the preparation, migration, and testing phases.

During the first part of last year, we upgraded the computers and software used by the Administration and Communications Department, thereby providing our colleagues with a homogeneous and up-to-date work environment as well as increased mobility. This project included rolling out Office 365 non-profit licenses, after HITS had been recognized by Microsoft as a non-profit institution. Later, during the summer, HITS – together with its sister companies – was subject to a Microsoft license check, which we successfully passed.

As we aim for a tighter integration of our clusters at HITS and URZ, we upgraded the connection between the two sites to 40Gbit/s and added a second link for redundancy. At the same time, we made 396 additional Intel Skylake cores available to our scientists. The increase might appear small compared with that of some previous years due to our planning of a complete upgrade of the HPC environment at HITS at the beginning of 2020. This upgrade will be one the main subjects of the next annual report.

4 Communication and Outreach

HITS Communications team in 2019 (f.l.t.r.): Olexandr Golovin, Peter Saueressig, Isabel Lacurie, Angela Michel.

Head of Communications
Dr. Peter Saueressig

Staff members
Isabel Lacurie
Angela Michel (since January 2019)

Student
Olexandr Golovin (since April 2019)

The HITS communications team works to make "HITS" synonymous with excellence in interdisciplinary research. The small Institute is perfectly suited for young, aspiring researchers, and the communications team aims to draw the attention of the media to the excellence of these HITS scientists in their respective fields, which range from covering the nano- to the galactic scales. Moreover, we strive to spark enthusiasm for science among school students and the general public alike through our outreach activities. In 2019, we assumed the additional responsibility of organizing scientific events, including workshops and conferences at HITS. Some of last year's highlights are presented below.

2019 was a very busy year for the communications department. We developed a new HITS corporate design as part of our branding strategy that applies to our digital products, such as our website and our social media channels, as well as the our print products, such as our newsletter "The Charts" and, of course, the Annual Report that you are reading right now. The new corporate design became fully implemented by the end of the year. In January, we were happy to welcome a new permanent member to the communications team: Angela Michel, an experienced science communicator with a long track record not only in science writing but also in event management. Angela is responsible for the new task of organizing scientific workshops and conferences at HITS and the Studio Villa Bosch together with Isabel Lacurie, who is also in charge of HITS' online activities and internal communications. Our student worker, Alex Golovin, who is currently writing his master's thesis in astrophysics, rounded out the team in April.

Events, events, events

In 2019, we organized a dozen scientific workshops and conferences on the HITS campus and at Studio Villa Bosch with the help of the administration's front office and in large part thanks to Christina Blach (see the list of events in Chapter 5.1). Moreover, our annual events took place beginning with the national "Girls' Day" on 28 March. Girls' Day aims to broaden the minds of young girls and pique their interest in – for example – a STEM subject, such as research at HITS. In 2019, twelve girls between the ages of ten and fifteen visited us. Scientists from the GRG and PSO groups offered small-scale, hands-on workshops to show the girls what the daily work and life of a researcher are like in fields such as mathematics and astrophysics.
From 22–26 May, 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, exhibits, and presentations designed to get youngsters interested in the natural sciences. In 2019, Explore Science's motto was "Time," and the event attracted more than 47,000 visitors from all over the region. HITS offered two hands-on stations – a brachistochrone and a "snail run" – at which the school kids could determine if they were faster than a snail in terms of their relative size. Researchers from five groups (MCM, MBM, CST, DMQ, GRG) as well as members of the administration and the communications teams participated in the event.

A bit exhausted, but happy: The HITS team after the run.

Stefan Richter (MCM) explaining the different time scales.

On 5 July, a team of 24 HITSters participated in the "NCT-Lauf" ("NCT Run"), an annual charity run organized by the National Center for Tumor Diseases Heidelberg (NCT) that boasted more than 5,000 runners in 2019. The participants could choose between a round course, a 10-kilometer track, and a half marathon. HITS management reimbursed the starting fee for its runners and provided a free running shirt. Angela Michel organized the participation of the HITS runners, who did remarkably well: Lucas Rettenmeier (NLP) won the 10-km run with a time of 35.29 minutes. Moreover, Florian Stecker (GRG) came in 2nd, Michael Strube (NLP) came in 8th, and Daria Kokh (MCM) and Alex Golovin (Communications) both came in 17th place, all in their respective age categories.

Every two years, HITS invites its alumni to a two-day meeting in Heidelberg. Last year, almost 80 alumni and HITSters met in the Studio Villa Bosch and on the HITS campus (see Chapter 5.3).
Later in the summer, the Institute again participated in the International Summer Science School Heidelberg (ISH), at which Erica Hopkins and her colleagues from the AIN group tutored two school students from France and Great Britain. The students worked with PINK, a software tool developed by the AIN group that helps astronomers analyze the morphology of galaxies.

In late September, the 7th installment of the Heidelberg Laureate Forum (HLF) took place, and we again had

Lucas Czech gave a talk on "The Revolution of Evolution" at the "Talk im Hirsch" event.

the opportunity to welcome a group of young researchers from all over the world (see Chapter 6).
In addition, HITSters participated in special outreach events in 2019: Lucas Czech (CME) appeared in the "Talk im Hirsch" ("Talk at the Hirsch") event in Schwetzingen, a well-known live talk show with guests from politics, show business, and science. Last but not least, two HITSters participated in an outreach event format that has been around for some time: "Astronomy on tap," at which topics of astronomy are "served" in a pub via short talks, along with a glass of beer or two. Antonio D'Isanto and Erica Hopkins mesmerized the audience in the Heidelberg pub "O'Reilly's" with lively and funny talks about stars, galaxies, and artificial intelligence.

Erica Hopkins and Antonio D´Isanto during their "Astronomy on Tap" talks.

Spreading the news of scientific excellence

Nowadays, everyone is talking about AI, but machine learning methods have been applied at HITS for several years now. Moreover, the interdisciplinary exchange of scientific ideas between groups has proven highly fruitful. Together with meteorologists at the Karlsruhe Institute of Technology KIT, Sebastian Lerch (CST) has developed a neural network that provides better temperature predictions than do classical standard methods (see Chapter 2.4). Lerch's research is part of the SFB 165 Transregional Collaborative Research Center "Waves to Weather" (W2W), and his research was inspired by an internal seminar on machine learning at HITS held by Antonio D'Isanto (AIN). Following the KIT press release, Sebastian Lerch talked about his method in the "Campus Report" radio broadcast.

In another press release, we spread the news about findings on the neuronal circuits of a brain structure involved in memory, behavior, and reward learning. Rebecca Wade was part of an international team that published its results in the PLOS Computational Biology Journal (see Chapter 2.8).
HITS research hit the front page of

the scientific journal *Nature* in October: PSO group leader Friedrich Röpke and Fabian Schneider (Heidelberg University) developed a model that demonstrates that the strong magnets in the Universe can be formed in stellar mergers (see Chapter 2.10 and the title image of this report).

The excellent quality of HITS research was again reflected in the "Highly Cited Researchers" report by Clarivate Analytics, with Alexandros Stamatakis (CME) ranking among the scientists most cited in their field and year of publication.

Media relations: Journalist-in-Residence program

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

In June 2019, HITS began a new call for applications for the years 2020 and 2021 that was announced internationally for the fourth time. Isabel Lacurie and Peter Saueressig advertised the program at the World Conference of Science Journalists in Lausanne, Switzerland, where over 1,300 journalists from 82 countries met on the shoreline of Lake Geneva from 1–5 July.

Immediately following the conference, HITS welcomed its eighth Journalist in Residence, award-winning journalist Andreas von Bubnoff from New York City, USA, for a three-month stay instead of the originally planned five months due to his new affiliation with Rhine-Waal University in Kleve, Germany. During his time with us, von Bubnoff held an internal seminar with the HITSters entitled "Getting the story, and getting it right: How science journalists find and report stories."

At the end of his stay, von Bubnoff gave a public talk on "the pursuit of truth" about science journalism in the "post-truth" era.

Candidates from six continents applied to become the HITS "Journalist in Residence" for the following two years. In November 2019, the selection committee chose Siobhan Roberts (Toronto, Canada) for 2020 and Carl Smith (Sydney, Australia) for 2021.

Andreas von Bubnoff at the World Conference of Science Journalists in Lausanne. After the conference, he came to HITS as the eigth Journalist in Residence.

5 Events

5.1 Conferences, Workshops & Courses

5.1.1 "Towards Simulating Cell Membranes" workshop

6-7 February 2019
Heidelberg, Studio Villa Bosch

The workshop "Towards Simulating Cell Membranes: Closer to Reality" focussed on new developments in the area of membrane biology, using molecular dynamics simulations and integrative computational techniques. More than 30 researchers from Europe, Asia and America participated in the workshop that covered broadly themes specific to membrane signaling, membrane mechanics, and recent trends in integrating experimental data into computational structure and function that will aid in the prediction of proteins in/at membranes.

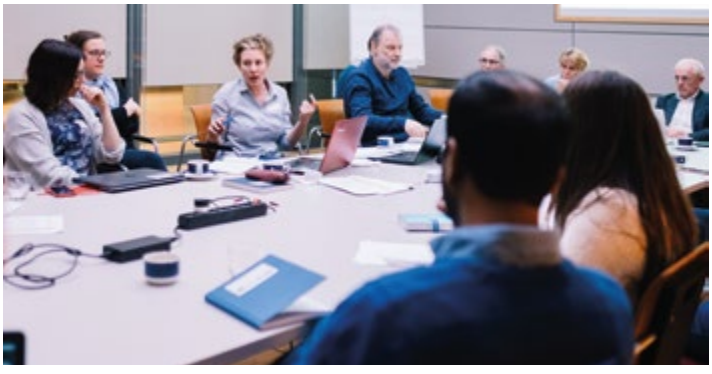
The event was organized by HITS researcher Frauke Gräter (MBM group) and Dr. Lipi Thukral, CSIR-Institute of Genomics and Integrative Biology, Delhi, India.

5.1.2 MESI-STRAT Annual Meeting

14–15 March 2019
Studio Villa Bosch, Heidelberg, Germany

MESI-STRAT
Systems Medicine of Metabolic-Signaling networks - A New Concept for Breast Cancer Patient Stratification

The MESI-STRAT consortium explores the interplay of breast cancer metabolism and oncogenic signaling (MEtabolic SIgnaling) by using systems medicine approaches. MESI-STRAT develops new models for the knowledge-based STRATification of patients into subgroups with different endocrine therapy resistance mechanisms.
The first annual meeting took place from 14–15 March 2019 at the Studio Villa Bosch in Heidelberg. For two days, all MESI-STRAT partners and the international advisory board met to present and discuss the results of all work packages. Together, they also developed a concrete plan of action for the following year that will ensure that all milestones are fulfilled. Fruitful workshops with the members from the international advisory board enabled extensive integration of their expertise and resources into ongoing MESI-STRAT activities. In addition, sessions on the importance of open access publication and FAIR data management were held. The Scientific Databases and Visualization group (SDBV) hosted and organized the event.



5.1.3 "Data to Knowledge": E-Science Days

27–29 March 2019
Heidelberg University, Heidelberg, Germany

The E-Science-Days 2019 "Data to Knowledge" took place from 27–29 March 2019 at Heidelberg University. The event concentrated on the quest for knowledge contained in data: How is new knowledge obtained from data? How is this knowledge shared and preserved? Major topics included research data management as well as new concepts, techniques, infrastructures, and services for different research areas – from bioscience to the humanities.

Event formats included workshops, tandem talks, and a panel discussion. Moreover, there was also room for demo tables, "birds of a feather" discussions, and so-called lightning talks on topics related to E-Science. HITS Scientific Director Wolfgang Müller gave a lightning talk on "de.NBI-SysBio FAIR Data Management for the Life Sciences Using the FAIRDOMHub." The opening session in the "Neue Aula" (the new auditorium) was introduced and presented by Prof. Dr. Vincent Heuveline, CIO of Heidelberg University and leader of the Data Mining and Uncertainty Quantification (DMQ) group at HITS. Prof. Dr. Bernhard Eitel, Rector of Heidelberg University, welcomed the participants. Then, Theresia Bauer, Minister of Science for Baden-Württemberg, delivered her welcome address. "Data science is a topic of high impact," she proclaimed as she encouraged the participants to use the power of data to create new knowledge.

Vincent Heuveline at the opening session.

Theresia Bauer during her welcome address.

The E-Science-Days are organized by the bwFDM-Info II project comprising Karlsruhe Institute of Technology, Konstanz University, and Heidelberg University. The event receives funding from the Ministry of Science, Research and Arts Baden-Württemberg.

5.1.4 de.NBI SIG and CCU meeting

8–9 April 2019
Studio Villa Bosch, Heidelberg, Germany

The "German Network for Bioinformatics Infrastructure – de.NBI" is a national, academic, non-profit infrastructure supported by the Federal Ministry of Education and Research that provides bioinformatics services to users in life sciences research and biomedicine both in Germany and throughout Europe. The partners organize training events, courses, and summer schools on tools, standards, and compute services provided by de.NBI to assist researchers in more effectively exploiting their data.
In April, the de.NBI Special Interest groups (SIG) and the Central Coordination Unit (CCU) met at the Studio Villa Bosch. In a number of separate and joint sessions, the SIG groups discussed topics such as data management and service monitoring. The CCU meeting closed the event, which was organized by SDBV group leader and HITS Scientific Director Wolfgang Müller, who is a member of CCU and deputy chair of SIG 4 "interoperability and data management."

5.1.5 Symposium "Computational Astrophysics"

2–3 May 2019
Studio Villa Bosch, Heidelberg, Germany
Heidelberg University and HITS jointly organized a two-day international symposium on computational astrophysics at the Studio Villa Bosch. Distinguished researchers from Europe, Asia, America, and Australia gave lectures on and discussed current topics from this research area ranging from star formation and asteroseismology to black holes and galaxy formation. The Q&A discussions with the audience supported the common notion of Heidelberg as being the Germany city with the highest density of astronomers.

5.1.6 Wellcome Trust Advanced Course on Computational Molecular Evolution

13–24 May 2019, Wellcome Genome Campus, Hinxton, UK

The EMBL-EBI Advanced Course on Computational Molecular Evolution took place for the 11th time at the European Bioinformatics Institute in Hinxton, UK, 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, Brian Moore, Bruce Rannala, and Ziheng Yang.

Alexandros Stamatakis (CME group leader) has been a co-organizer of the event at Hinxton since 2011. CME PhD students Lucas Czech and Benoit Morel also contributed substantially to the success of the course as teaching assistants.
The course will be offered next in Heraklion, Greece, in May 2020, and the dates for the 2021 course in Hinxton have already been determined.

5.1.7 "Electron-Capture-Initiated Stellar Collapse" workshop

20–24 May 2019, Lorentz Center, Leiden University, Leiden, Netherlands

The PSO group co-sponsored the international workshop on "Electron-Capture-Initiated Stellar Collapse" held at the Lorentz Center of Leiden University, the Netherlands. In this fruitful meeting, scientists discussed the progenitor systems, strategies for modeling the expected supernova events, and implications for observations and cosmic nucleosynthesis of the evolutionary end stages of stars in the range between 7 and 10 solar masses. Some of the discussions from the workshop even resulted in a publication (see [Jones, 2019]).

5.1.8 PoLiMeR kick-off meeting

26–31 May 2019, Studio Villa Bosch and Bioquant (Heidelberg University), Heidelberg, Germany

Metabolic diseases are a burden on the European population as well as to its healthcare system. Individual differences with respect to history, lifestyle, and genetic makeup have been increasingly recognized to affect disease progression and treatment response. A Systems Medicine approach based on computational models and fed with individual patient data has the potential to provide the basis for a personalized diagnosis and treatment strategy. The PoLiMeR consortium (Polymers in the Liver: Metabolism and Regulation) has found that the inherited, liver-related diseases of glycogen and lipid metabolism serve as the ideal starting point for innovative research training in personalized 'Systems Medicine.'

More than 30 participants met in Heidelberg for the kick-off meeting, which consisted of training sessions, teaching seminars, and talks. The invited speakers were Adriano Henney (University of Luxembourg) and Ursula Kummer (Heidelberg University), and the meeting was organized by SDBV members Olga Krebs and Wolfgang Müller.

PoLiMeR is funded through the EU Marie Skłodowska-Curie Innovative Training Network (ITN), which drives scientific excellence and innovation. ITNs bring together universities, research institutes, and industry and clinical partners from across the world to train researchers at the doctorate level.

5.1.9 10 years of COMBINE: Better standards in systems biology

15–19 July 2019, Studio Villa Bosch, Heidelberg, Germany

Systems biology uses simulated computer models to investigate biological processes and entire organ systems. Researchers obtain data from different sources and in various formats. In order to combine these data and make them truly "predictable," numerous data standards have been developed in recent years that cover subfields of research.

Organizer Martin Golebiewski showing the COMBINE anniversary shirt.

These subfields include "Systems Biology Markup Language" (SBML) as a representation format for biological computer models and "Systems Biology Graphical Notation" (SBGN) for visualization. The problem, however, is that the standards were developed independently of one another. This is where the international initiative "Computational Modeling in Biology Network" (COMBINE) comes in. COMBIME aims at harmonizing these standards. For 10 years, experts from this international network have been meeting worldwide at annual conferences. This year, two HITS scientists – Martin Golebiewski and Wolfgang Müller (both from Scientific Databases and Visualization, SDBV) – hosted "COMBINE 2019," which took place for the tenth time in total and the second time at HITS. "COMBINE 2019" brought together more than 100 experts with the goal of advancing standards in systems biology and synthetic biology. Topics such as computer-aided physiology, modeling tools, reproducibility, and the visualization of biological networks were tackled by more than 25 invited speakers from all over the world as well as through numerous lectures, poster presentations, and discussion rounds. A special focus was placed on the applications of systems biology for personalized medicine. COMBINE's history dates back to its co-founder Michael Hucka (California Institute of Technology, Pasadena, CA, USA), who is also considered the "father of SBML." Hucka gave a positive summary of the initiative´s previous work: "Today, there are eight recognized COMBINE standards as well as additional standardization initiatives, including initiatives from the pharmaceutical and neurosciences sectors."

5.1.10 "FAIR Data Infrastructures for Biomedical Communities" workshop

10 September 2019, Dortmund University of Applied Sciences and Arts, Dortmund, Germany

This workshop introduced the activities of different projects, initiatives, and research sites in Germany with regard to FAIR (Findable, Accessible, Interoperable, and Re-usable) data management within the biomedical research communities. Sharing experiences enables common concepts and a joint strategy for FAIR data and corresponding standards to be developed.
HITSter Martin Golebiewski (SDBV) organized the workshop during the 64th annual conference of the German Association for Medical Informatics, Biometry, and Epidemiology (GMDS). Martin also leads the hosting GMDS working group "FAIR data infrastructures for biomedical informatics," which was formed in early 2019. The group bundles activities for health-related FAIR data infrastructures in the fields of biomedical research, clinical data management, and medical informatics in Germany. Furthermore, it aims to provide a platform for information exchange concerning the best practices for and implementation of FAIR data management in these fields. The group additionally aims to construct a network of FAIR data management providers in these fields. By bridging between providers and potential users of such data management platforms, user requirements are expected to become better integrated in the development of data management tools.

5.1.11 COMBINE & de.NBI Tutorial: "Modeling and Simulation Tools in Systems Biology"

31 October 2019, OIST, Okinawa, Japan

At the 20th International Conference on Systems Biology (ICSB) at the Okinawa Institute of Science and Technology Graduate University (https://www2.aeplan.co.jp/icsb2019/), Martin Golebiewski (SDBV group) organized and coordinated the one-day COMBINE & de.NBI tutorial workshop entitled "Modeling and Simulation Tools in Systems Biology" together with David Nickerson (Auckland Bioengineering Institute, New Zealand). (http://co.mbine.org/events/tutorial2019). As a conference satellite, the tutorial offered young scientists from 12 different countries the opportunity to participate in lectures, software demonstrations, and hands-on sessions that showed them how to set up computer models of biological networks and simulate these models in different software platforms. Moreover, domain-specific community standard formats were demonstrated to be crucial to the data- and model exchange. These types of modeling standards for the life sciences are defined by the COMBINE network (http://co.mbine.org). COMBINE board member Martin Golebiewski provided an overview of the network's standards and their use. The tutors for the conference included Andreas Weidemann (SDBV) as well as HITS alumni Jürgen Pahle and Sven Sahle (Heidelberg University).

5.1.12 "Molecular Stresses" workshop

4–5 November 2019, Studio Villa Bosch, Heidelberg, Germany

On 4–5 November, the Molecular Biomechanics (MBM) group organized the international workshop "Molecular Stresses," which dealt with new and unifying concepts for calculating and understanding molecular stresses in complex materials. Such molecular stresses can stem from quantum chemical calculations as well as from atomistic or coarse-grained Molecular Dynamics simulations. These stresses can be inherent to the system (i.e., they can reflect intrinsic tension) or build upon external mechanical perturbations.
At HITS, experts from the relevant scientific areas – such as computational physics, materials science, and biological matter – came together for the workshop, which comprised invited and contributed talks as well as many discussions among speakers and participants. Additionally, in a practical session, the participants learned and applied the force distribution analysis tool, which was developed in the MBM Group. The workshop was organized by Isabel Martin and Frauke Gräter (both MBM).

5.1.13 Sino-German Symposium on "Uncertainty Quantification for Engineering and Industrial Application"

2–6 December 2019, Studio Villa Bosch, Heidelberg, Germany

In December 2019, the first Sino-German Symposium took place at HITS in Heidelberg. The event was organized by Vincent Heuveline (leader of the DMQ group at HITS and a professor at Heidelberg University), Chen Song (a researcher at both HITS and Heidelberg University), and Jiawei Zhang (a professor at Xi'an University of Technology). During the week, more than 30 scientists gave highly exciting talks about their current research. Various topics were discussed during the symposium, including solvers and preconditioners, multilevel Monte-Carlo techniques, machine learning, and Bayesian inference. Several topics of industrial applications were also discussed, such as smart grid, power flow, medical applications, and material sciences. The main emphasis lay on addressing the issues of reliable numerical simulation and optimization techniques in the context of innovative engineering and industrial applications.
In order to render the first Sino-German symposium the nucleus of a long-term and regular workshop series and Sino-German cooperation platform, the last day of the symposium was dedicated to evaluating the first event and to establishing a framework for future exchange and collaborations.

5.1.14 "Supernovae and Stellar Hydrodynamics" workshop

16–17 December 2019, Studio Villa Bosch, Heidelberg, Germany

On 16–17 December 2019, the PSO group organized the 14th Würzburg Workshop on Supernovae and Stellar Hydrodynamics in Heidelberg at the Studio Villa Bosch. More than 30 researchers from universities and institutions in Germany, the UK, Japan, Sweden, Hungary, the US, and Germany gathered to discuss new developments in the field and to collaborate on new and existing projects. These researchers included several HITS alumni from the PSO and TAP groups. The participants had detailed discussions on 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.

5.2 HITS Colloquia

Prof. Dr. Alexander Mäder

Hochschule der Medien, Stuttgart, Germany
21 January 2019: How AI could help journalists

Prof. Dr. Carsten Rother

Visual Learning Lab Heidelberg, Heidelberg University, Germany
18 February 2019: Introducing the Visual Learning Lab Heidelberg – from Machine Learning for Inverse Problems to Learning Data Generation

Prof. Dr. Hadley Wickham

University of Auckland, New Zealand RStudio, Stanford University and Rice University, USA
12 March 2019: Data Science Challenges

Prof. Dr. Jørgen Ellegaard Andersen

Centre for Quantum Geometry of Moduli Spaces, Aarhus University, Denmark
13 May 2019: Quantum Geometry of Moduli Spaces in pure mathematics and with applications to Quantum Field Theory and macro-molecular biology

Prof. Dr. Rainer Malaka

Center for Computing Technologies (TZI), University of Bremen, Germany
12 July 2019: Empowering People with Digital Media: Playful and Natural Human Computer Interaction

Prof. Dr. Peter Hunter

Auckland Bioengineering Institute, University of Auckland, New Zealand
15 July 2019: Computational Physiology and the Physiome Project

Prof. Dr. Richard A. Bonneau

NYU Center for Data Science, New York, USA
9 September 2019: Large Scale Machine Learning Methods for Predicting Protein Functions from integration of Sequence, Structure and Networks

Prof. Dr. Andreas von Bubnoff

HITS Journalist in Residence 2019
7 October 2019: Auf der Suche nach der Wahrheit – Wissenschaftsjournalismus in Zeiten von posttruth und reproducibility crisis

Prof. Dr. Nicholas G. Reich

School of Public Health and Health Sciences, University of Massachusetts, Amherst, USA
21 October 2019: The Computational Science of Real-time Influenza Forecasting

Prof. Dr. Keith A. Crandall

Milken Institute of Public Health, The George Washington University, Washington DC, USA
18 November 2019: Computational Approaches for Characterizing Microbiome Diversity

Prof. Dr. Erik Lindahl

Department of Biochemistry and Biophysics, Stockholm University, Sweden
16 December 2019: Integrative Modeling of Allosteric Modulation in Pentameric Ligand-Gated Ion Channels

5.3 HITS Alumni Meeting

Twelve years after the first alumni meeting took place at Villa Bosch, HITS management invited alumni from HITS, its predecessor (EML Research), and EML for a two-day event on 12 & 13 July 2019.

The meeting format emphasized interaction between alumni and all current HITSters. The response proved impressive as more than 70 individuals registered for the meeting.

The event commenced on Friday with a public scientific colloquium in the Studio Villa Bosch given by Rainer Malaka, a professor at the University of Bremen, Germany, and an EML alumnus (1997–2006). In his talk, Dr. Malaka presented insights into the main topic: "Empowering people with digital media: Playful and natural human computer interaction."

After a break, HITS Managing Director Gesa Schönberger and Scientific Director Wolfgang Müller welcomed the participants to the event. Their welcome address was followed by a panel discussion on careers in which three alumni (who had remained in academia) talked about their career paths and life after leaving EML/HITS. These discussions were held by Agnieszka Bronowska (now at Newcastle University, UK), who had worked as a postdoc in the MBM group; Matthias Stein (now at the Max-Planck-Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany), a former postdoc in the MCM group; and Simone Ponzetto (University of Mannheim), who had been a doctoral student in the NLP group.

The vivid discussion was moderated by Michael Strube (NLP).
In the evening, alumni and HITSters met for an informal dinner at the "Wirtshaus Nepomuk" restaurant in Heidelberg's historic center.

On Saturday afternoon, an outdoor "grill & chill" party was held on the HITS campus that lasted until sundown. HITS chef Ralf Westermann and his team treated the participants to delicious food and beverages. Alumni and HITSters came with their families, so there was plenty of room for friendly discussions and laughs.

The "grill & chill" party on Saturday afternoon.

During the discussion: Matthias Stein, Simone Ponzetto, Agnieszka Bronowska, and Michael Strube (f.l.t.r.).

6 Collaborations

At the opening ceremony (f.l.t.r.): Manfred Salmhofer, Ralf Klessen, Anna Wienhard, and Bernhard Eitel, Rector of Heidelberg University.

STRUCTURES – Cluster of Excellence

Heidelberg University, one of the HITS shareholders, received support for two proposals from the Clusters of Excellence funding line within the framework of the Excellence Strategy from both the federal and state governments. "STRUCTURES" is one of these two proposals. It addresses specific, highly topical questions about the formation, role, and detection of structure in a broad range of natural phenomena, from subatomic particles to cosmology, from fundamental quantum physics to neuroscience. The new concepts and methods explored by the cluster are key to finding structures in large datasets and to realizing new forms of analogue computing.
The cluster's spokespersons are Manfred Salmhofer (Institute for Theoretical Physics), Ralf S. Klessen (Center for Astronomy), and Anna Wienhard (Mathematical Institute), who also leads the Groups and Geometry (GRG) group at HITS. Friedrich Röpke, head of the Physics of Stellar Objects (PSO) group at HITS, is also involved in this cluster.

Unifying fundamental concepts
Why are there planets rather than merely dust in the vicinity of stars? How did the 'dark ages' of the Universe end with the illumination of the first stars? How can we understand neuronal activity patterns in the brain? How does self-organization work in biophysical systems and in quantum matter? In all these phenomena, physical processes merge on a wide range of scales in terms of both time and length. While essential for the richness of observed phenomena, this merging poses enormous challenges to producing a quantitative analysis. The central idea of "STRUCTURES" is that these seemingly very different phenomena share a common feature of unifying fundamental concepts, which we test in a controlled analysis of model systems. The combination of mathematical theory, large-scale numerical simulations, and physical computation via novel analogue computers – which is unique to Heidelberg – promises to lead to transformative progress. The close collaboration of physicists, mathematicians, and computer scientists plays a crucial role in "STRUCTURES."
At the opening ceremony on 15 July 2019, Anna Wienhard acted as presenter, and her PhD student Mareike Pfeil delivered a short talk on how mathematicians play billiards. The talk outlined the path from the billiards table to translational surfaces.

Mareike Pfeil during her talk.

Young researchers from the HLF at HITS.

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 one another 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 is the product of a joint initiative between HITS and the Klaus Tschira Foundation, which supports both the HLF and HITS. Andreas Reuter – founding Managing Director of HITS – has been involved since the very beginning. The Heidelberg Laureate Forum Foundation (HLFF) was founded by the Klaus Tschira Foundation in 2013. The Foundation organizes the HLF, and Andreas Reuter serves as Scientific Chairperson.

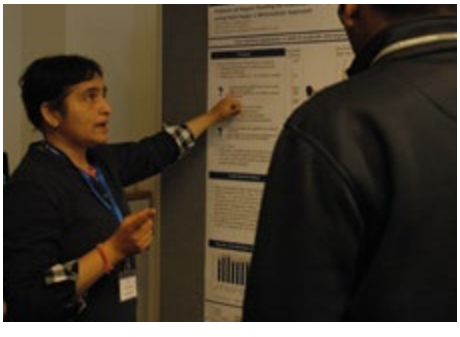
According to the agreement, HITS is in charge of continuing the scientific support of the networking event and is to continue to contribute its world-renowned expertise in mathematical and computational topics, as was intended by its founder, Klaus Tschira.

HITS Scientific Director Wolfgang Müller talks to the young researchers.

HITSters meet the young researchers

As in the previous six years, HITS hosted a group of young researchers from the HLF 2019 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.

Philipp Gerstner (Data Mining and Uncertainty Quantification group) during the poster session.



Some impressions from young researchers´ visit at HITS.

7 Publications

Alessandrini D (2019). *Higgs Bundles and Geometric Structures on Manifolds.* SIGMA 15 039: 32.

Alessandrini D, Guichard O, Rogozinnikov E, Wienhard A (2019). *Noncommutative coordinates for symplectic representations*. arXiv:math.DG,1911.08014.

Alessandrini D, DiSarlo V (2019). *Generalized stretch lines for surfaces with boundary*. arXiv:math.GT,1911.10431.

Ardevol-Pulpillo R, Janka H, Just O, Bauswein A (2019). *Improved leakage-equilibration-absorption scheme (ILEAS) for neutrino physics in compact object mergers.* Monthly Notices of the Royal Astronomical Society 485(4):4754-4789.

Baar I, Hübner L, Oettig P, Zapletal A, Schlag S, Stamatakis A, Morel B (2019). *Data Distribution for Phylogenetic Inference with Site Repeats via Judicious Hypergraph Partitioning.* IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp.175-184, IEE.E.

Ballhausen A, Przybilla MJ, Jendrusch M, Haupt S, Pfaffendorf E, Draxlbauer M, Seidler F, Krausert S, Ahadova A, Kalteis MS, Heid D, Gebert J, Bonsack M, Schott S, Bläker H, Seppälä T, Mecklin J, Broeke ST, Nielsen M, Heuveline V, Krzykalla J, Benner A, Riemer AB, von Knebel Doeberitz M, Kloor M (2019). *The shared neoantigen landscape of MSI cancers reflects immunoediting during tumor evolution.* biorxiv;691469v1,[Preprint].

Bauer MS, Baumann F, Daday C, Redondo P, Durner E, Jobst MA, Milles LF, Mercadante D, Pippig DA, Gaub HE, Gräter F, Lietha D (2019). *Structural and mechanistic insights into mechanoactivation of focal adhesion kinase.* Proc Natl Acad Sci USA 116(14):6766-6774.

Bauswein A (2019). *Equation of state constraints from multi-messenger observations of neutron star mergers.* Annals of Physics 411:167958.

Bauswein A, Stergioulas N (2019). *Spectral classification of gravitational-wave emission and equation of state constraints in binary neutron star mergers.* J. Phys. G: Nucl. Part. Phys. 46(11):113002.

Bauswein A, Bastian NF, Blaschke DB, Chatziioannou K, Clark JA, Fischer T, Oertel M (2019). *Identifying a First-Order Phase Transition in Neutron-Star Mergers through Gravitational Waves.* Phys. Rev. Lett. 122(6).

Beyrer J, Fioravanti E (2019). *Cross ratios on CAT(0) cube complexes and marked length-spectrum rigidity.* arXiv:math.GT,1903.02447.

Beyrer J (2019). *Marked length spectrum rigidity for actions on CAT(0) cube complexes.* (to appear in Oberwolfach reports /30).

Bläker H, Haupt S, Morak M, Holinski-Feder E, Arnold A, Horst D, Sieber-Frank J, Seidler F, von Winterfeld M, Alwers E, Chang-Claude J, Brenner H, Roth W, Engel C, Löffler M, Möslein G, Schackert H, Weitz J, Perne C, Aretz S, Hüneburg R, Schmiegel W, Vangala D, Rahner N, Steinke-Lange V, Heuveline V, von Knebel Doeberitz M, Ahadova A, Hoffmeister M, Kloor M (2019). *BRAF mutation testing of MSI CRCs in Lynch syndrome diagnostics: performance and efficiency according to patient's age.* medrxiv;19009274v1,[Preprint].

Brehm MA, Klemm U, Rehbach C, Erdmann N, Kolšek K, Lin H, Aponte-Santamaría C, Gräter F, Rauch BH, Riley AM, Mayr GW, Potter BVL, Windhorst S (2019). *Inositol hexakisphosphate increases the size of platelet aggregates.* Biochemical Pharmacology 161:14-25.

Brehmer J, Gneiting T (2019). *Properization: Constructing proper scoring rules via Bayes acts.* Ann Inst Stat Math.

Breitling P, Stamatakis A, Chernomor O, Bettisworth B, Reszczynski L (2019). *Empirical Analysis of Phylogenetic Quasi-Terraces*. biorxiv;810309v1,[Preprint].

Britavskiy N, Lennon DJ, Patrick LR, Evans CJ, Herrero A, Langer N, van Loon JT, Clark JS, Schneider FRN, Almeida LA, Sana H, de Koter A, Taylor WD (2019). *The VLT-FLAMES Tarantula Survey.* XXX. Red stragglers in the clusters Hodge 301 and SL 639. A&A 624:A128.

Bruce NJ, Narzi D, Trpevski D, van Keulen SC, Nair AG, Röthlisberger U, Wade RC, Carloni P, Hellgren Kotaleski J (2019). *Regulation of adenylyl cyclase 5 in striatal neurons confers the ability to detect coincident neuromodulatory signals.* PLoS Comput Biol 15(10):e1007382.

Bruce NJ, Ganotra GK, Richter S, Wade RC (2019). *KBbox: A Toolbox of Computational Methods for Studying the Kinetics of Molecular Binding.* J. Chem. Inf. Model. 59(9):3630-3634.

Burger M, Iozzi A, Parreau A, Pozzetti MB (2019). *Currents, Systoles, and Compactifications of Character Varieties.* arXiv:math.GT,1902.07680.

Caetano-Anollés G, Aziz MF, Mughal F, Gräter F, Koç I, Caetano-Anollés K, Caetano-Anollés D (2019). *Emergence of Hierarchical Modularity in Evolving Networks Uncovered by Phylogenomic Analysis.* Evol Bioinform Online 15:1176934319872980.

Calogeropoulou T, Magoulas GE, Pöhner I, Panecka-Hofman J, Linciano P, Ferrari S, Santarem N, Jiménez-Antón MD, Olías-Molero AI, Alunda JM, Cordeiro da Silva A, Wade RC, Costi MP (2019). *Hits and Lead Discovery in the Identification of New Drugs against the Trypanosomatidic Infections.* Medicinal Chemistry of Neglected and Tropical Diseases: Advances in the Design and Synthesis of Antimicrobial Agents, Eds. Jayaprakash V, Castagnolo D, Özkay Y, CRC Press, chapter 10, pp 185-231.

Collier B, Sanders A (2019). *(G,P)-opers and global Slodowy slices*. arXiv:math.DG,1911.12844

Comte N, Morel B, Hasic D, Guéguen L, Boussau B, Daubin V, Penel S, Scornavacca C, Gouy M, Stamatakis A, Tannier E, Parsons DP (2019). *Treerecs: an integrated phylogenetic tool, from sequences to reconciliations.* biorxiv;782946v1,[Preprint]

Czech L, Barbera P, Stamatakis A (2019). *Genesis and Gappa: Processing, Analyzing and Visualizing Phylogenetic (Placement) Data.* biorxiv;647958v3,[Preprint].

Daday C, Mateyka LM, Gräter F (2019). *How ARVC-Related Mutations Destabilize Desmoplakin: An MD Study.* Biophysical Journal 116(5):831-835.

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T (2019). *ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models.* Molecular Biology and Evolution 37(1):291-294.

Daskalopoulos G, Mese C, Sanders A, Vdovina A (2019). *Surface groups acting on CAT (-1) spaces.* Ergod. Th. Dynam. Sys. 39(7):1843-1856.

Dufton PL, Evans CJ, Hunter I, Lennon DJ, Schneider FRN (2019). *A census of massive stars in NGC 346.* Stellar parameters and rotational velocities. A&A 626:A50.

Fatima M, Müller M (2019). *HITS-SBD at the FinSBD Task: Machine Learning vs. Rule-based Sentence Boundary Detection.* In Proceedings of the 1st Workshop on Financial Technology and Natural Language Processing (FinNLP) @ IJCAI , Macau, China, 10-12 August, pp. 115-121.

Feldmann K, Richardson DS, Gneiting T (2019). *Grid- versus station-based postprocessing of ensemble temperature forecasts.* Geophys. Res. Lett. 46(13):7744-7751.

Ferrand G, Warren DC, Ono M, Nagataki S, Röpke FK, Seitenzahl IR (2019). *From Supernova to Supernova Remnant: The Three-dimensional Imprint of a Thermonuclear Explosion.* ApJ 877(2):136.

Franceschini F, Frigerio R, Pozzetti MB, Sisto A (2019). *The zero norm subspace of bounded cohomology of acylindrically hyperbolic groups.* Comment. Math. Helv. 94(1):89-139.

Galbany L, Ashall C, Höflich P, González-Gaitán S, Taubenberger S, Stritzinger M, Hsiao EY, Mazzali P, Baron E, Blondin S, Bose S, Bulla M, Burke JF, Burns CR, Cartier R, Chen P, Della Valle M, Diamond TR, Gutiérrez CP, Harmanen J, Hiramatsu D, Holoien TWS, Hosseinzadeh G, Howell DA, Huang Y, Inserra C, de Jaeger T, Jha SW, Kangas T, Kromer M, Lyman JD, Maguire K, Marion GH, Milisavljevic D, Prentice SJ, Razza A, Reynolds TM, Sand DJ, Shappee BJ, Shekhar R, Smartt SJ, Stassun KG, Sullivan M, Valenti S, Villanueva S, Wang X, Wheeler JC, Zhai Q, Zhang J (2019). *Evidence for a Chandrasekhar-mass explosion in the Ca-strong 1991bg-like type Ia supernova 2016hnk.* A&A 630:A76.

Galvin TJ, Huynh M, Norris RP, Wang XR, Hopkins E, Wong OI, Shabala S, Rudnick L, Alger MJ, Polsterer KL (2019). *Radio Galaxy Zoo: Knowledge Transfer Using Rotationally Invariant Self-organizing Maps*. PASP 131(1004):108009.

Gerstner P, Baumann M, Heuveline V (2019). *Analysis of the Stationary Thermal-ElectroHydrodynamic Boussinesq Equations.* EMCL PrePrint Series, No 01 [Preprint].

Gianniotis N (2019). *Mixed Variational Inference.* International Joint Conference on Neural Networks (IJCNN), pp.1-8, IEEE.

Goicovic FG, Springel V, Ohlmann ST, Pakmor R (2019). *Hydrodynamical moving-mesh simulations of the tidal disruption of stars by supermassive black holes.* Monthly Notices of the Royal Astronomical Society 487(1):981-992.

Golebiewski M (2019). *Data Formats for Systems Biology and Quantitative Modeling.* Encyclopedia of Bioinformatics and Computational Biology, pp.884-893, Elsevier.

Gryn'ova G, Corminboeuf C (2019). *Conceptual Framework of Organic Electronics.* chimia (aarau) 73(4):245-251.

Gryn'ova G, Corminboeuf C (2019). *Topology-Driven Single-Molecule Conductance of Carbon Nanothreads.* J. Phys. Chem. Lett. 10(4):825-830.

Gräter F, Li W (2019). *Studying Functional Disulphide Bonds by Computer Simulations.* Functional Disulphide Bonds 1967:87-113, Springer New York.

Heinzerling B, Strube M (2019). *Sequence Tagging with Contextual and Non-Contextual Subword Representations: A Multilingual Evaluation.* In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July – 2 August, pp. 273-291.

Hemri S (2019). *Multi-model Combination and Seamless Prediction.* In Handbook of Hydrometeorological Ensemble Forecasting, Eds: Duan Q, Pappenberger F, Thielen J, Wood A, Cloke HL, Schaake JC. Springer-Verlag, pp. 285-307.

Herrera-Rodríguez AM, Miletić V, Aponte-Santamaría C, Gräter F (2019). *Molecular Dynamics Simulations of Molecules in Uniform Flow.* Biophysical Journal 116(9):1579-1585.

Horowitz CJ, Arcones A, Côté B, Dillmann I, Nazarewicz W, Roederer I, Schatz H, Aprahamian A, Atanasov D, Bauswein A, Beers TC, Bliss J, Brodeur M, Clark JA, Frebel A, Foucart F, Hansen CJ, Just O, Kankainen A, McLaughlin GC, Kelly JM, Liddick SN, Lee DM, Lippuner J, Martin D, Mendoza-Temis J, Metzger BD, Mumpower MR, Perdikakis G, Pereira J, O'Shea BW, Reifarth R, Rogers AM, Siegel DM, Spyrou A, Surman R, Tang X, Uesaka T, Wang M (2019). *r-process nucleosynthesis: connecting rare-isotope beam facilities with the cosmos.* J. Phys. G: Nucl. Part. Phys. 46(8):083001.

Huang J, Brenna C, Khan AuM, Daniele C, Rudolf R, Heuveline V, Gretz N (2019). *A cationic near infrared fluorescent agent and ethyl-cinnamate tissue clearing protocol for vascular staining and imaging.* Nature Scientific Reports 9(1):521.

Huber S, Suyu SH, Noebauer UM, Bonvin V, Rothchild D, Chan JHH, Awan H, Courbin F, Kromer M, Marshall P, Oguri M, Ribeiro T, LSST Dark Energy Science Collaboration (2019). *Strongly lensed SNe Ia in the era of LSST: observing cadence for lens discoveries and time-delay measurements.* A&A 631:A161.

Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D, Burki F (2019/2020). *Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity.* Mol Ecol Resour.00:1-15.

John DN, Heuveline V, Schober M (2019). *GOODE: A Gaussian Off-The-Shelf Ordinary Differential Equation Solver.* In Proceedings of the 36th International Conference on Machine Learning, vol. 97, pp. 3152–3162, PMLR.

Jones S, Côté B, Röpke FK, Wanajo S (2019). *A New Model for Electron-capture Supernovae in Galactic Chemical Evolution.* ApJ 882(2):170.

Jones S, Röpke FK, Fryer C, Ruiter AJ, Seitenzahl IR, Nittler LR, Ohlmann ST, Reifarth R, Pignatari M, Belczynski K (2019). *Remnants and ejecta of thermonuclear electron-capture supernovae.* Constraining oxygen-neon deflagrations in high-density white dwarfs. A&A 622:A74.

Jordan A, Krüger F, Lerch S (2019). *Evaluating probabilistic forecasts with scoringRules.* J. Stat. Soft. 90(12).

Kirsebom OS, Jones S, Strömberg DF, Martínez-Pinedo G, Langanke K, Röpke FK, Brown BA, Eronen T, Fynbo HOU, Hukkanen M, Idini A, Jokinen A, Kankainen A, Kostensalo J, Moore I, Möller H, Ohlmann ST, Penttilä H, Riisager K, Rinta-Antila S, Srivastava PC, Suhonen J, Trzaska WH, Äystö J (2019). *Discovery of an Exceptionally Strong β -Decay Transition of F 20 and Implications for the Fate of Intermediate-Mass Stars.* Phys. Rev. Lett. 123(26).

Kokh DB, Kaufmann T, Kister B, Wade RC (2019). *Machine Learning Analysis of τRAMD Trajectories to Decipher Molecular Determinants of Drug-Target Residence Times.* Front. Mol. Biosci. 6, 36.

Kozlov AM, Stamatakis A (2019). *Using RAxML-NG in Practice.* [Preprint].

Laehnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Beerenwinkel N, Campbell KR, Mahfouz A, Pinello L, Skums P, Stamatakis A, Attolini CS, Aparicio S, Baaijens J, Balvert M, de Barbanson B, Cappuccio A, Corleone G, Dutilh BE, Florescu M, Guryev V, Holmer R, Jahn K, Lobo TJ, Keizer EM, Khatri I, Kiełbasa SM, Korbel JO, Kozlov AM, Kuo T, Lelieveldt BPF, Mandoiu II, Marioni JC, Marschall T, Mölder F, Niknejad A, Rączkowski Ł, Reinders M, de Ridder J, Saliba A, Somarakis A, Stegle O, Theis FJ, Yang H, Zelikovsky A, McHardy AC, Raphael BJ, Shah SP, Schönhuth A (2019). *12 Grand Challenges in Single-Cell Data Science.* e27885v3,[Preprint].

Lee G, Marquis L (2019). *Anti–de Sitter strictly GHC-regular groups which are not lattices.* Trans. Amer. Math. Soc. 372(1):153-186.

Linciano P, Pozzi C, dello Iacono L, di Pisa F, Landi G, Bonucci A, Gul S, Kuzikov M, Ellinger B, Witt G, Santarem N, Baptista C, Franco C, Moraes CB, Müller W, Wittig U, Luciani R, Sesenna A, Quotadamo A, Ferrari S, Pöhner I, Cordeiro-da-Silva A, Mangani S, Costantino L, Costi MP (2019). *Enhancement of Benzothiazoles as Pteridine Reductase-1 Inhibitors for the Treatment of Trypanosomatidic Infections.* J. Med. Chem. 62(8):3989-4012.

López F, Heinzerling B, Strube M (2019). *Fine-Grained Entity Typing in Hyperbolic Space.* In Proceedings of The Fourth Workshop on Representation Learning for NLP (Rep4NLP) @ ACL , Florence, Italy, 2 August , pp. 169-180.

Lutteropp S, Kozlov AM, Stamatakis A (2019). *A Fast and Memory-Efficient Implementation of the Transfer Bootstrap.* Bioinformatics, btz874.

Lyonnais S, Sadiq SK, Lorca-Oró C, Dufau L, Nieto-Marquez S, Escriba T, Gabrielli N, Tan X, Ouizougun-Oubari M, Okoronkwo J, Reboud-Ravaux M, Gatell JM, Marquet R, Paillart J, Meyerhans A, Tisné C, Gorelick RJ, Mirambeau G (2019). *The HIV-1 ribonucleoprotein dynamically regulates its condensate behavior and drives acceleration of protease activity through membrane-less granular phase-separation.* biorxiv;528638v2,[Preprint].

Marshall DL, Gryn'ova G, Poad BLJ, Bottle SE, Trevitt AJ, Coote ML, Blanksby SJ (2019). *Experimental Evidence for Long-Range Stabilizing and Destabilizing Interactions between Charge and Radical Sites in Distonic Ions.* International Journal of Mass Spectrometry 435:195-203.

Martynov D, Miao H, Yang H, Vivanco FH, Thrane E, Smith R, Lasky P, East WE, Adhikari R, Bauswein A, Brooks A, Chen Y, Corbitt T, Freise A, Grote H, Levin Y, Zhao C, Vecchio A (2019). *Exploring the sensitivity of gravitational wave detectors to neutron star physics.* Phys. Rev. D 99(10).

Moosavi NS, Born L, Poesio M, Strube M (2019). *Using Automatically Extracted Minimum Spans to Disentangle Coreference Evaluation from Boundary Detection.* In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July – 2 August, pp. 4168-4178.

Moraes CB, Witt G, Kuzikov M, Ellinger B, Calogeropoulou T, Prousis KC, Mangani S, Di Pisa F, Landi G, Dello Iacono L, Pozzi C, Freitas-Junior LH, Pascoalino BdS, Bertolacini CP, Behrens B, Keminer O, Leu J, Wolf M, Reinshagen J, Cordeiro-da-Silva A, Santarem N, Venturelli A, Wrigley S, Karunakaran D, Kebede B, Pöhner I, Müller W, Panecka-Hofman J, Wade RC, Fenske M, Clos J, Alunda JM, Corral MJ, Uliassi E, Bolognesi ML, Linciano P, Quotadamo A, Ferrari S, Santucci M, Borsari C, Costi MP, Gul S (2019). *Accelerating Drug Discovery Efforts for Trypanosomatidic Infections Using an Integrated Transnational Academic Drug Discovery Platform.* SLAS DISCOVERY: Advancing the Science of Drug Discovery 24(3):346-361.

Morel B, Kozlov AM, Stamatakis A, Szöllősi GJ (2019). *GeneRax: A tool for species tree-aware maximum likelihood based gene tree inference under gene duplication, transfer, and loss.* biorxiv;779066v1, [Preprint].

Mukherjee G, Nandekar P, Mustafa G, Richter S, Wade RC (2019). *A Multi-resolution Approach to the Simulation of Protein Complexes in a Membrane Bilayer.* High Performance Computing in Science and Engineering ' 18, pp. 505-514, Springer International Publishing.

Müller M, Bannister A, Reitz F (2019). *Off-the-shelf Semantic Author Name Disambiguation for Bibliographic Data Bases.* In Proceedings of the 23rd International Conference on Theory and Practice of Digital Libraries (TPDL), Oslo, Norway, 9-12 September, pp. 397-400.

Müller M (2019). *Semantic Matching of Documents from Heterogeneous Collections: A Simple and Transparent Method for Practical Applications.* In Proceedings of the RELATIONS Workshop on Meaning Relations between Phrases and Sentences @ IWCS , Gothenburg, Sweden, 23 May , pp. 34-41.

Mustafa G, Nandekar PP, Bruce NJ, Wade RC (2019). *Differing Membrane Interactions of Two Highly Similar Drug-Metabolizing Cytochrome P450 Isoforms: CYP 2C9 and CYP 2C19.* IJMS 20(18):4328.

Mustafa G, Nandekar PP, Camp TJ, Bruce NJ, Gregory MC, Sligar SG, Wade RC (2019). *Influence of Transmembrane Helix Mutations on Cytochrome P450-Membrane Interactions and Function.* Biophysical Journal 116(3):419-432.

Neal ML, König M, Nickerson D, Mısırlı 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 (2019). *Harmonizing semantic annotations for computational models in biology.* Briefings in Bioinformatics 20(2):540-550.

Nunes-Alves AN (2019). *From Brazil to Germany: Challenges and Advantages.* J. Chem. Info. Model. 60(2):449-451 74.

Nuñez FP, Gianniotis N, Blex J, Lisow T, Chini R, Polsterer KL, Pott J, Esser J, Pietrzyński G (2019). *Optical continuum photometric reverberation mapping of the Seyfert-1 galaxy Mrk509*. Monthly Notices of the Royal Astronomical Society 490(3):3936-3951.

Ott A (2019). *Transgression in bounded cohomology and a conjecture of Monod.* J. Topol. Anal.:1-40.

Owocki SP, Hirai R, Podsiadlowski P, Schneider FRN (2019). *Hydrodynamical simulations and similarity relations for eruptive mass-loss from massive stars.* Monthly Notices of the Royal Astronomical Society 485(1):988-1000.

Patrick LR, Lennon DJ, Britavskiy N, Evans CJ, Sana H, Taylor WD, Herrero A, Almeida LA, Clark JS, Gieles M, Langer N, Schneider FRN, van Loon JT (2019). *The VLT-FLAMES Tarantula Survey.* XXXI. Radial velocities and multiplicity constraints of red supergiant stars in 30 Doradus. A&A 624:A129.

Perron U, Kozlov AM, Stamatakis A, Goldman N, Moal IH (2019). *Modeling Structural Constraints on Protein Evolution via Side-Chain Conformational States.* Molecular Biology and Evolution 36(9):2086-2103.

Petit V, Wade GA, Schneider FRN, Fossati L, Kamp K, Neiner C, David-Uraz A, Alecian E, MiMeS Collaboration (2019). *The MiMeS survey of magnetism in massive stars: magnetic properties of the O-type star population.* Monthly Notices of the Royal Astronomical Society 489(4):5669-5687.

Pozzetti B, Sambarino A, Wienhard A (2019). *Anosov representations with Lipschitz limit set*. arXiv:math.DG,1910.06627.

Pozzetti B, Sambarino A, Wienhard A (2019). *Conformality for a robust class of non-conformal attractors.* arXiv:math.DG,1902.01303.

Renzo M, de Mink SE, Lennon DJ, Platais I, van der Marel RP, Laplace E, Bestenlehner JM, Evans CJ, Hénault-Brunet V, Justham S, de Koter A, Langer N, Najarro F, Schneider FRN, Vink JS (2019). *Space astrometry of the very massive ~150 M⊙ candidate runaway star VFTS682.* Monthly Notices of the Royal Astronomical Society: Letters 482(1):L102-L106.

Schmidt TT, Sharma S, Reyes GX, Gries K, Gross M, Zhao B, Yuan J, Wade RC, Chabes A, Hombauer H (2019). *A genetic screen pinpoints ribonucleotide reductase residues that sustain dNTP homeostasis and specifies a highly mutagenic type of dNTP imbalance.* Nucleic Acids Research 47(1):237-252.

Schneider FRN, Ohlmann ST, Podsiadlowski P, Röpke FK, Balbus SA, Pakmor R, Springel V (2019). *Stellar mergers as the origin of magnetic massive stars.* Nature 574(7777):211-214.

Schreiber F, Sommer B, Bader GD, Gleeson P, Golebiewski M, Hucka M, Keating SM, König M, Myers C, Nickerson D, Waltemath D (2019). *Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2019.* Journal of Integrative Bioinformatics 16(2).

Schween N, Meyer-Hübner N, Gerstner P, Heuveline V (2019). *A time step reduction method for Multi-Period Optimal Power Flow problems.* EMCL PrePrint Series, No 02 [Preprint].

Seelig T, Meyer A, Gerstner P, Meier M, Jongmanns M, Baumann M, Heuveline V, Egbers C (2019). *Dielectrophoretic force-driven convection in annular geometry under Earth's gravity.* International Journal of Heat and Mass Transfer 139:386-398.

Sekulić I, Strube M (2019). *Adapting deep learning methods for mental health prediction on social media.* In Proceedings of the 5th Workshop on Noisy User-generated Text, Hong Kong, 4 November, pp. 322-327.

Serdari D, Kostaki E, Paraskevis D, Stamatakis A, Kapli P (2019). *Automated, phylogeny-based genotype delimitation of the Hepatitis Viruses HBV and HCV.* PeerJ 7:e7754.

Sheridan S, Gräter F, Daday C (2019). *How Fast Is Too Fast in Force-Probe Molecular Dynamics Simulations?* J. Phys. Chem. B 123(17):3658-3664.

Song C, Heuveline V (2019). *Uncertainty Assessment of the Blood Damage in a FDA Blood Pump.* Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP), pp.640-652.

Stamatakis A (2019). *A review of approaches for optimizing phylogenetic likelihood calculations.* In Bioinformatics and Phylogenetics, pp. 1–19, Springer.

Stanford NJ, Scharm M, Dobson PD, Golebiewski M, Hucka M, Kothamachu VB, Nickerson D, Owen S, Pahle J, Wittig U, Waltemath D, Goble C, Mendes P, Snoep J (2019). *Data Management in Computational Systems Biology: Exploring Standards, Tools, Databases, and Packaging Best Practices.* Yeast Systems Biology 2049:285-314.

Suman SK, Daday C, Ferraro T, Vuong-Brender T, Tak S, Quintin S, Robin F, Gräter F, Labouesse M (2019). *The plakin domain of C. elegans VAB-10/plectin acts as a hub in a mechanotransduction pathway to promote morphogenesis.* Development 146(24):dev183780.

Taubenberger S, Floers A, Vogl C, Kromer M, Spyromilio J, Aldering G, Antilogus P, Bailey S, Baltay C, Bongard S, Boone K, Buton C, Chotard N, Copin Y, Dixon S, Fouchez D, Fransson C, Gangler E, Gupta RR, Hachinger S, Hayden B, Hillebrandt W, Kim AG, Kowalski M, Leget P, Leibundgut B, Mazzali PA, Noebauer UM, Nordin J, Pain R, Pakmor R, Pecontal E, Pereira R, Perlmutter S, Ponder KA, Rabinowitz D, Rigault M, Rubin D, Runge K, Saunders C, Smadja G, Tao C, Thomas RC (2019). *SN 2012dn from early to late times: 09dc-like supernovae reassessed.* Monthly Notices of the Royal Astronomical Society 488(4):5473-5488.

Timmes F, Fryer C, Timmes F, Hungerford AL, Couture A, Adams F, Aoki W, Arcones A, Arnett D, Auchettl K, Avila M, Badenes C, Baron E, Bauswein A, Beacom J, Blackmon J, Blondin S, Bloser P, Boggs S, Boss A, Brandt T, Bravo E, Brown E, Brown P, Bruenn S, Budtz-Jørgensen C, Burns E, Calder A, Caputo R, Champagne A, Chevalier R, Chieffi A, Chipps K, Cinabro D, Clarkson O, Clayton D, Coc A, Connolly D, Conroy C, Côté B, Couch S, Dauphas N, deBoer RJ, Deibel C, Denisenkov P, Desch S, Dessart L, Diehl R, Doherty C, Domínguez I, Dong S, Dwarkadas V, Fan D, Fields B, Fields C, Filippenko A, Fisher R, Foucart F, Fransson C, Fröhlich C, Fuller G, Gibson B, Giryanskaya V, Görres J, Goriely S, Grebenev S, Grefenstette B, Grohs E, Guillochon J, Harpole A, Harris C, Harris JA, Harrison F, Hartmann D, Hashimoto M, Heger A, Hernanz M, Herwig F, Hirschi R, Hix RW, Höflich P, Hoffman R, Holcomb C, Hsiao E, Iliadis C, Janiuk A, Janka T, Jerkstrand A, Johns L, Jones S, José J, Kajino T, Karakas A, Karpov P, Kasen D, Kierans C, Kippen M, Korobkin O, Kobayashi C, Kozma C, Krot S, Kumar P, Kuvvetli I, Laird A, Laming (J) M, Larsson J, Lattanzio J, Lattimer J, Leising M, Lennarz A, Lentz E, Limongi M, Lippuner J, Livne E, Lloyd-Ronning N, Longland R, Lopez LA, Lugaro M, Lutovinov A, Madsen K, Malone C, Matteucci F, McEnery J, Meisel Z, Messer B, Metzger B, Meyer B, Meynet G, Mezzacappa A, Miller J, Miller R, Milne P, Misch W, Mitchell L, Mösta P, Motizuki Y, Müller B, Mumpower M, Murphy J, Nagataki S, Nakar E, Nomoto K, Nugent P, Nunes F, O'Shea B, Oberlack U, Pain S, Parker L, Perego A, Pignatari M, Pinedo GM, Plewa T, Poznanski D, Priedhorsky W, Pritychenko B, Radice D, Ramirez-Ruiz E, Rauscher T, Reddy S, Rehm E, Reifarth R, Richman D, Ricker P, Rijal N, Roberts L, Röpke F, Rosswog S, Ruiter AJ, Ruiz C, Savin DW, Schatz H, Schneider D, Schwab J, Seitenzahl I, Shen K, Siegert T, Sim S, Smith D, Smith K, Smith M, Sollerman J, Sprouse T, Spyrou A, Starrfield S, Steiner A, Strong AW, Sukhbold T, Suntzeff N, Surman R, Tanimori T, The L, Thielemann F, Tolstov A, Tominaga N, Tomsick J, Townsley D, Tsintari P, Tsygankov S, Vartanyan D, Venters T, Vestrand T, Vink J, Waldman R, Wang L, Wang X, Warren M, West C, Wheeler JC, Wiescher M, Winkler C, Winter L, Wolf B, Woolf R, Woosley S, Wu J, Wrede C, Yamada S, Young P, Zegers R, Zingale M, Portegies Zwart S (2019). *Catching Element Formation In The Act ; The Case for a New MeV Gamma-Ray Mission: Radionuclide Astronomy in the 2020s.* Bulletin of the American Astronomical Society 51(3):2.

Torres-Rivas A, Chatziioannou K, Bauswein A, Clark JA (2019). *Observing the post-merger signal of GW170817-like events with improved gravitational-wave detectors.* Phys. Rev. D 99(4).

Wade RC, Salo-Ahen O (2019). *Molecular Modeling in Drug Design.* Molecules 24(2):321.

Watson D, Hansen CJ, Selsing J, Koch A, Malesani DB, Andersen AC, Fynbo JPU, Arcones A, Bauswein A, Covino S, Grado A, Heintz KE, Hunt L, Kouveliotou C, Leloudas G, Levan AJ, Mazzali P, Pian E (2019). *Identification of strontium in the merger of two neutron stars.* Nature 574(7779):497-500.

Wibberg D, Batut B, Belmann P, Blom J, Glöckner FO, Grüning B, Hoffmann N, Kleinbölting N, Rahn R, Rey M, Scholz U, Sharan M, Tauch A, Trojahn U, Usadel B, Kohlbacher O (2019). *The de.NBI / ELIXIR-DE training platform - Bioinformatics training in Germany and across Europe within ELIXIR.* F1000Res 8:1877.

Wodak SJ, Paci E, Dokholyan NV, Berezovsky IN, Horovitz A, Li J, Hilser VJ, Bahar I, Karanicolas J, Stock G, Hamm P, Stote RH, Eberhardt J, Chebaro Y, Dejaegere A, Cecchini M, Changeux J, Bolhuis PG, Vreede J, Faccioli P, Orioli S, Ravasio R, Yan L, Brito C, Wyart M, Gkeka P, Rivalta I, Palermo G, McCammon JA, Panecka-Hofman J, Wade RC, Pizio AD, Niv MY, Nussinov R, Tsai C, Jang H, Padhorny D, Kozakov D, McLeish T (2019). *Allostery in Its Many Disguises: From Theory to Applications.* Structure 27(4):566-578.

Xu E, Bülow Sv, Chen P, Lenting PJ, Kolšek K, Aponte-Santamaría C, Simon B, Foot J, Obser T, Schneppenheim R, Gräter F, Denis CV, Wilmanns M, Hennig J (2019). *Structure and dynamics of the platelet integrin-binding C4 domain of von Willebrand factor.* Blood 133(4):366-376.

Yuan J, Han SB, Richter S, Wade RC, Kokh D (2020). *Druggability Assessment in TRAPP using Machine Learning Approaches,* J. Chem. Inf. Model. 2020, 60(3):1685-1699.

Zhang W, Mu G, Song C, Yan G, Heuveline V (2019). *Extraction of Spatial-Temporal Features of Bus Loads in Electric Grids Through Clustering in a Dynamic Model Space.* IEEE Access 8:5852-5861.

Zhang Y, Xia K, Cao Z, Gräter F, Xia F (2019). *A new method for the construction of coarse-grained models of large biomolecules from low-resolution cryo-electron microscopy data.* Phys. Chem. Chem. Phys. 21(19):9720-9727.

Zhou X, Lutteropp S, Czech L, Stamatakis A, von Looz M, Rokas A (2019). *Quartet-Based Computations of Internode Certainty Provide Robust Measures of Phylogenetic Incongruence.* Systematic Biology 69(2):308–324.

Zhu Y, Heinzerling B, Vulić I, Strube M, Reichart R, Korhonen A (2019). *On the importance of subword information for morphological tasks in truly low-resource languages.* In Proceedings of the 23rd Conference on Computational Natural Language Learning, Hong Kong, 3-4 November, pp. 216-226.

8 Teaching

Degrees

Rudolf Biczok:
"Integration of internal and external gene expression and drug-perturbation data to empower novel immune therapies against Parkinson's Disease", Master thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2019).

Paula Breitling:
"Quantitave Analysis of Phylogenetic Quasi-Terraces", Master thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2019).

Svenja de Buhr:
"Investigating the First Steps in the Activation of Focal Adhesion Kinase Using Molecular Dynamics Simulations", Master thesis, Faculty of Bioscience, Heidelberg University (Ada Cavalcanti-Adam), and HITS: Frauke Gräter (2019).

Antonio D´Isanto:
"Probabilistic photometric redshift estimation in massive digital sky surveys via machine learning", Ph.D. thesis, faculty of physics and astronomy, Heidelberg University (Joachim Wambsganss) and HITS: Kai Polsterer (2019).

Gaurav Kumar Ganotra:
"Computational studies of drug-binding kinetics", Ph.D. thesis, Combined Faculty for the Natural Sciences and Mathematics, Heidelberg University, and HITS: Rebecca C. Wade (2019).

Paul Gontschar:
"The geometry of Higman's groups", Master thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Anna Wienhard, together with Thomas Haettel, University of Montpellier, France (2019).

Anton Hanke:
"Computational characterisation of protein homo- and heterodimer formation", Bachelor thesis, Molecular Biotechnology, Faculty of Biosciences, Heidelberg University, and HITS: Ariane Nunes-Alves and Rebecca C. Wade (2019).

Benjamin Heinzerling:
"Aspects of Coherence for Entity Analysis", Ph.D. thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube (2019).

Erica Hopkins:
"Semi-Automated Morphological Classification of All Radio Galaxies in the FIRST Survey Catalog Including Near-IR Counterparts in UKIDSS", Master thesis, faculty of physics and astronomy, Heidelberg University, Referees: Prof. Dr. Joachim Wambsganss and Dr. Coryn Bailer-Jones. Advisor at HITS: Kai Polsterer (2019).

Sascha Hund:
"Penetration Testing moderner Web-Anwendungen", Bachelor thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Vincent Heuveline (2019).

Manuel Kramer:
"Hydrodynamics of the Common Envelope Phase of sdB Stars", Master thesis, faculty of physics and astronomy, Heidelberg University, advisor: Friedrich Röpke (2019).

Nafise Sadat Moosavi:
"Robustness in Coreference Resolution", Ph.D. thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube (2019).

Marcel Petrov:
"Verbesserung des Nutzer-Engagements für SABIO-RK", Bachelor thesis, Bachelor of Arts DHBW Mosbach: Violette Lotz and HITS: Wolfgang Müller (2019).

Olaf Pichler:
"Smart Contracts for Distributed Databases", Master thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Vincent Heuveline (2019).

Benedikt Rennekamp:
"Simulating Bond Scission in Tensed Collagen: A Hybrid Monte Carlo / Molecular Dynamics Approach" Master thesis, Faculty of Physics, Heidelberg University (Ulrich S. Schwarz), and HITS: Frauke Gräter (2019).

Oskar Riedler:
"Lorentzian manifold with non-compact isometry group", Master thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Maria Beatrice Pozzetti (2019).

Robin Ruland:
"Künstliche neuronale Netze zur Klassifizierung von Hirntumoren auf Basis medizinischer Bilddaten", Bachelor thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Vincent Heuveline (2019).

Sebastian Scholz:
"Non-Intrusive Polynomial Chaos Expansion for Incompressible Viscous Flows", Master thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Vincent Heuveline (2019).

Tobias Schröder:
"Scaling Limits of Random Trees", Bachelor thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Anna Wienhard (2019).

Benedikt Schulz:
"Combining probability forecasts: A comprehensive comparison", Master thesis, Faculty of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting (2019).

Daniel Spitz:
"Universal structures far from equilibrium in alpha-complexes and persistent homology", Master thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Anna Wienhard, together with Jürgen Berges, Institute for Theoretical Physics, Heidelberg University (2019).

Florian Stecker:
"Domains of discontinuity of Anosov representations in flag manifolds and oriented flag manifolds", Ph.D. thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Anna Wienhard (2019).

Philipp Ullmann:
"Molecular dynamics simulation studies of gp41 cytoplasmic tail lateral embedding in phospholipid bilayer", Bachelor thesis, Biosciences, Faculty of Biosciences, Heidelberg University, and HITS: S. Kashif Sadiq and Rebecca C. Wade (2019).

Ferdinand Vanmaele:
"Coxeter groups, the Davis complex, and isolated flats", Bachelor thesis, Faculty of Mathematics and Computer Science, Heidelberg University: Maria Beatrice Pozzetti (2019).

Peter Vogel:
"Assessing predictive performance: From precipitation forecasts over the tropics to receiver operating characteristic curves and back", Ph.D. thesis, Faculty of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting (2019).

Eva-Maria Walz:
"A generalization of ROC curves", Master thesis, Faculty of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting (2019).

Johanna Wegmann:
"Technical and Algorithmic Optimization of PaPaRa", Master thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2019).

Jui-Hung Yuan:
"Druggability Assessment in TRAPP Procedure using Machine Learning Approaches", Master thesis, Master of Scientific Computing, Faculty of Mathematics and Computer Science, Heidelberg University, and HITS: Rebecca C. Wade and Daria Kokh (2019).

Lectures, Courses and Seminars

Peter Albers, Maria Beatrice Pozzetti:
Hauptseminar *„Geometrie"* (Geometry Seminar), Heidelberg University, winter semester 2019/20.

Peter Albers, Maria Beatrice Pozzetti, Anna Wienhard:
Hauptseminar *"Geometrie"* (Geometry Seminar), Heidelberg University, winter semester 2018/19. Hauptseminar *"Geometrie"* (Geometry Seminar), Heidelberg University, summer semester 2019.

Timo Dimitriadis:
Lecture and exercise course on *"Applied time series analysis and forecasting techniques"*, Universität Hohenheim, winter semester 2019.

Timo Dimitriadis (with Robert Jung):
Seminar in *"Time series and financial econometrics"*, Universität Hohenheim, winter semester 2019.

Valentina Disarlo:
Lecture *"Hyperbolic Surfaces"*, Heidelberg University, summer semester 2019.

Tilmann Gneiting:
Lecture course on *"Forecasting: Theory and practice II"*, Karlsruhe Institute of Technology, summer semester 2019.

Tilmann Gneiting (lectures) and Johannes Resin (exercises):
"Forecasting Summer School: Probabilistic forecasts", International Symposium on Forecasting (ISF), Thessaloniki, Greece, 15-16 June 2019.

Tilmann Gneiting and Johannes Resin:
Seminar on *"Statistical forecasting and classification"*, Karlsruhe Institute of Technology, winter semester 2019/2020.

Martin Golebiewski, Andreas Weidemann:
COMBINE & de.NBI Tutorial *"Modeling and Simulation Tools in Systems Biology"*, International Conference on Systems Biology (ICSB 2019), Okinawa, Japan, 31 October – 5 November 2019.

Frauke Gräter:
Contribution to lecture *"Computational biochemistry"* for biochemistry master students, winter semester 2019/20, Heidelberg University.
Contribution to lecture *"Biophysical Chemistry"* for Molecular Biotechnology bachelor students, winter semester 2019/20, Heidelberg University.
Lecture with practicals *"Physical Chemistry of Life II"* (with Michael Boutros), Heidelberg University, winter semester WS2019/20.

Frauke Gräter, Rebecca C. Wade:
M.Sc. Seminar course *"Machine Learning for the Biomolecular World"*, Heidelberg University, summer semester, 2019.

Frauke Gräter, Rebecca C. Wade, Kashif Sadiq, Fabian Kutzki, Benedikt Rennekamp and Florian Franz:
M.Sc. practical course *"Computational Molecular Biophysics"*, Heidelberg University, winter semester, 2018/2019.

Vincent Heuveline, Maximilian Hoecker:
Seminar *"IT Security"*, Heidelberg University, summer semester 2019.

Vincent Heuveline, Philipp Gerstner, Philipp Lösel:
Lecture *"Einführung in die Numerik"*, Heidelberg University, winter semester 2019.

Alexey Kozlov:
"RAxML-NG Introduction and Laboratory", 2019 Workshop on Phylogenomics, Český Krumlov, Czech Republic, 20 January – 2 February 2019.

Olga Krebs:
"FAIRDOM training", University of Newcastle, UK, 4-6 March 2019; *"FAIR data training"* at the MESI-STRAT annual meeting, Studio Villa Bosch, Heidelberg, 14-15 March 2019; *"Scientific Data Management using FAIRDOMHub"*, University of Bergen, Norway, 9-11 May 2019; FAIRDOM training at the University College London (UCL), UK, 11-12 November 2019; FAIRDOM training, at the Université Clermont Auvergne (UCA), France, 13 November 2019; *"FAIR data management"* training at CSIC Barcelona, Spain, 22-23 November 2019; *"Making your data FAIR"*, project training for the PoLiMeR project, Groningen, The Netherlands, 29 November 2019.

Florian Lach:
"Computational Astrophysics", Exercises and tutorial accompanying the lecture course, Heidelberg University, Heidelberg, summer semester 2019.

Sebastian Lerch:
Lecture course on *"Probability and statistics for computer scientists"*, Karlsruhe Institute of Technology, winter semester 2018/2019 and winter semester 2019/2020.
Lecture course on *"Methods of data analysis"*, Karlsruhe Institute of Technology, summer semester 2019.

Wolfgang Müller, Olga Krebs:
"Tools for collaboration and FAIR data management", PoLiMeR annual meeting, Heidelberg, Germany, 29 May 2019.

Andreas Ott:
Lecture *"Geometric Methods in Data Analysis"*, Heidelberg University, summer semester 2019.

Kai Polsterer:
Ringvorlesung: *"Deep Learning in Engineering and Natural Sciences. Machine learning in astronomy."* University of Bochum, Germany, 29 April 2019. Lectures and exercises on *"supervised and unsupervised learning with applications to astronomy."* Ecole Doctorale 182 on Data Mining, Machine Learning and Deep Learning, Strasbourg, France, May 2019. Tutorial *"Introduction to Artificial Neural Networks: how to code your own artificial brain from scratch."* ADASS XXIX, Groningen, The Netherlands, 6 October 2019.

Maria Beatrice Pozzetti:
Lecture *„Geometrische Gruppentheorie"*, Heidelberg University, winter semester 2019/2020.

Maria Beatrice Pozzetti, Andy Sanders, Johannes Horn:
Seminar *"Higgs Bundles"*, Heidelberg University, winter semester 2019/2020.

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

Maja Rey, Andreas Weidemann, Ulrike Wittig:
de.NBI Course *"Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOMHub/SEEK"*, Heidelberg, Germany, 18-20 March 2019.

Friedrich Röpke:
"Computational Astrophysics", lecture course, Heidelberg University, summer semester 2019. *"Fundamentals of Simulation Methods"*, lecture course, given jointly with Prof. Cornelis Dullemond, Heidelberg University, winter semester 2018/2019.

Friedrich Röpke, Fabian Schneider:
"The stellar cookbook: A practical guide to the theory of stars", lecture course, Heidelberg University, winter semester 2019/2020.

S. Kashif Sadiq, Rebecca C. Wade:
Ringvorlesung *"Structure and Dynamics of Biological Macromolecules"*, B.Sc. Biosciences, Heidelberg University, Summer Semester, 2019.

Christian Sand:
"Physik B", Exercises and tutorial accompanying the lecture course, Heidelberg University, summer semester 2019. "Physik A", Exercises and tutorial accompanying the lecture course, Heidelberg University, winter semester 2019/2020.

Andy Sanders:
Lecture *"Geometry of Quiver Varieties"*, summer semester 2019.

Fabian Schneider:
"Python programming for scientists", lecture course, Heidelberg University, summer semester 2019.

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

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, 2018/2019.

Alexandros Stamatakis, Ben Bettisworth, Alexey Kozlov, Pierre Barbera:
Lecture *"Introduction to Bioinformatics for Computer Scientists"*, computer science Master's program at Karlsruhe Institute of Technology, winter semester, 2019/2020.

Alexandros Stamatakis, Benoit Morel, Sarah Lutteropp, Pierre Barbera, Alexey Kozlov, Ben Bettisworth:
Seminar *"Hot Topics in Bioinformatics"*, computer science Master's program at Karlsruhe Institute of Technology, summer semester, 2019.

Alexandros Stamatakis, Lucas Czech, Benoit Morel:
Summer school *"Computational Molecular Evolution"*, Hinxton, UK, May 2019.

Michael Strube:
PhD Colloquium, Department of Computational Linguistics, Heidelberg University (Winter Semester 2018/2019). Seminar: *"Automatic Text Processing"*, Department of Computational Linguistics, Heidelberg University (Winter Semester 2018/2019). PhD Colloquium, Department of Computational Linguistics, Heidelberg University (Summer Semester 2019).

Jan Swoboda:
Seminar *"High dimensional geometry and probability with a view towards applications in data science"*, Heidelberg University, summer semester 2019. Lecture *"Geometrische Analysis"*, winter semester 2019/20.

Rebecca C. Wade:
Module 4, *"Biomolecular Recognition: Modeling and Simulation"*, M.Sc. Molecular Cell Biology, Heidelberg University, 13 March 2019. Module 3, *"Protein Modeling"*, M.Sc. Molecular Cell Biology, Heidelberg University, 10 & 22 May 2019. Ringvorlesung *"Computational Biochemistry"*, *"Electrostatics and Solvation for Biomolecules"*, M.Sc. Biochemistry, Heidelberg University, 28 October 2019. Ringvorlesung *„Biophysik"*, *"Receptor-Ligand Interactions: Structure and Dynamics"*, B.Sc. Molecular Biotechnology, Heidelberg University, 21 November 2019.

Rebecca C. Wade, Christina Athanasiou, Daria Kokh, Goutam Mukherjee, Ariane Nunes-Alves, Stefan Richter, S. Kashif Sadiq, Alexandros Tsengenes, Frauke Gräter, Isabel Martin:
B.Sc. Biosciences practical course *"Grundkurs Bioinformatik"*, Heidelberg University, 21-25 January 2019.

Rebecca C. Wade, Ariane Nunes-Alves, Goutam Mukherjee, Stefan Richter
Ph. D. practical course *"Computational analysis of protein binding properties"*, Heidelberg Biosciences International Graduate School (HBIGS), Heidelberg University, 8-9 October 2019.

Andreas Weidemann:
"SABIO-RK hands-on tutorial", International Conference on Systems Biology (ICSB 2019), Okinawa, Japan, 31 October 2019.

Anna Wienhard:
Bachelor-/Masterseminar *"Geometrie"* (Junior Geometry Seminar), winter semester 2018/19. Seminar *"Differentialgeometrie"*, summer semester 2019. Bachelor-/Masterseminar *"Geometrie"* (Junior Geometry Seminar), summer semester 2019.

Anna Wienhard, Peter Albers, Mareike Pfeil, Arnaud Maret:
Seminar *"Dynamics on Teichmüller spaces"*, Heidelberg University, summer semester 2019.

Anna Wienhard, Michel Bleher, Daniel Spitz:
Journal Club on *"Topics in Topological and Geometric Methods in Data Analysis"*, Heidelberg University, winter semester 2019/20.

Anna Wienhard, Roman Sauer:
RTG Lecture (Asymptotic Invariants and Limits of Groups and Spaces), Heidelberg University & Karlsruhe Institute of Technology, summer semester 2019; winter semester 2019/20.

9 Miscellaneous

9.1 Guest Speaker Activities

Daniele Alessandrini:
"Generalities on harmonic maps", Workshop on Harmonic maps and rigidity, Sisteron, France, 8 April 2019. Minicourse *„Higher-Teichmüller theory and geometric structures"*, Workshop Higher-Teichmüller theory and related topics, Pavia, Italy, 3-7 June 2019.

Jonas Beyrer:
"Marked length spectrum rigidity for actions on CAT(0) cube complexes", Oberseminar Gruppen und Geometrie, Bielefeld, Germany, 5 June 2019. Differentialgeometrie im Großen, MFO, Oberwolfach, Germany, 3 July 2019. Oberseminar Geometrie, Fribourg, Switzerland, 30 October 2019. LMU München, 5 November 2019.

Nguyen-Thi Dang:
"Mélange topologique du flot directionnel des chambres de Weyl", Ergodic theory and dynamical systems seminar, LAGA, University of Villetaneuse, France, 6 November 2019. Parole aux jeune chercheuses et chercheurs en géométrie et dynamique, University of Nancy, France, 26 November 2019. *"Topological mixing of the Weyl chamber flows"*, Spectral theory seminar, University of Paderborn, Germany, 8 October 2019. *"The top Lyapunov exponent"*, Arbeitsgemeinschaft, MFO, Oberwolfach, Germany, October 14-19, 2019. *"Topological dynamics of the Weyl chamber flows"*, Ergodic theory and dynamical systems seminar, University of Zürich, Switzerland, 11 November 2019.

Madhura De:
"Single-molecule FRET studies on mono- and tri-chromatosomes", Leiden Institute of Physics, Leiden University, Leiden, The Netherlands, 28 October 2019.

Nikos Gianniotis:
"Autoencoding eclipsing binaries", Astronomical Time Series Workshop, Max Planck Institute for Astronomy, Heidelberg, Germany, 24 January 2019. *"Mixed Variational Inference"*, Computer Science Distinguished Seminars, The University of Birmingham, UK., 26 February 2019. *"Laplaced Based Variational Inference"*, Machine Learning Tools for Research in Astronomy Workshop, Max Planck Institute for Astronomy, Ringberg, Germany, 13 December 2019.

Tilmann Gneiting:
"Receiver operating characteristic (ROC) curves". Workshop on Predictability, Dynamics and Applications Research Using the TIGGE and S2S Ensembles, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK, 4 April 2019. *"Receiver-operating characteristic (ROC)*

curves: What are they and what are they good for?" Keynote Lecture, International Symposium on Forecasting (ISF), Thessaloniki, Greece, 19 June 2019. *"Evaluating point forecasts"*. Amazon Web Services, Berlin, Germany, 22 July 2019. *"Evaluating classifier performance"*. GridKa School 2019 - The Art of Data, Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Germany, 26 August 2019. „Grenzen der Wettervorhersage: Eine mathematische Perspektive". Public Lecture, Annual Meeting of the German Mathematical Society, Jahrestagung der Deutschen Mathematiker-Vereinigung, Karlsruhe, Germany, 24 September 2019. *"Isotonic distributional regression"*. Universität Bielefeld, Bielefeld, Germany, 10 October 2019. Amazon Research Days, Berlin, Germany, 10 October 2019. *„Erfolge und Grenzen der Wettervorhersage: Eine mathematische Perspektive"*. Special Holiday Lecture, Universität Gießen, Germany, 6 December 2019.

Martin Golebiewski:
"Standardization Activities of ISO/TC 276 Biotechnology - WG 5 Data Processing and Integration", the 125th MPEG meeting, Marrakesh, Morocco, 14-18 January 2019; Meeting of ISO/TC 215 Health Informatics, Gothenburg, Sweden, 13-17 April 2019. *"Standardising Data and Models in the Life Sciences"*, EMBO Science Policy Workshop *"Genomic Heterogeneity in Cell Lines: Technical and Policy Responses"*, Heidelberg, Germany, 26-28 May 2019. *"Community standards in systems biology and systems medicine and their coordinated development"*, Workshop *"Multicellular model description and cross-simulator exchange"*, Dresden, Germany, 1-4 September 2019.

Frauke Gräter:
"HPC for biomaterials: why playing soccer hurts", R-CCS International Symposium, Kobe, Japan, 15-17 February 2019. *"Mechanoradicals in collagen from experiments and simulations"* Spring Meeting of German Physical Society, 31 March – 5 April 2019. *"Why playing soccer hurts: Discovering collagen radicals in simulations and experiments across scales",* EMBO workshop on *"Synergy of experiment and computation in quantitative systems biology"*, Nove Hrady, Czech Republic, 23 -28 June 2019. *"Bio/Soft Matter Simulations across Multiple Scales"*, Cecam School, Heidelberg, 18-20 September 2019. Workshop *"Mechanical forces in Biology: Theory and Simulation"*, University of Los Andes, Bogota, Colombia, 30 September – 2 October 2019.

Ganna Gryn'ova:
"Crossing Electronic Bridges: Computational Chemistry of Molecular Junctions", seminar talk at Monash University, Melbourne, Australia, 24 September 2019; seminar talk at Australian National University, Canberra, Australia, 27 September 2019. *"Topology-Driven Molecular Electronics Lacking π-Conjugation"*, 15th European Conference on Molecular Electronics, Linköping, Sweden, 27-31 August 2019; 9th Conference of the Asia-Pacific Association of Theoretical and Computational Chemists, Sydney, Australia, 30 September – 3 October 2019. *"Quantification of Fuzzy Chemical Concepts"*, Interdisciplinary Symposium: Chemical Concepts and Quantum Chemistry, Heidelberg, Germany, 18-20 November 2019.

Vincent Heuveline:
"Uncertainty Quantification and HPC Challenges and Perspectives for Medical Applications", 4EU Alliance Flagship Meeting, Heidelberg Germany, 16 January 2019. *"Data Science and High Performance Computing: Evolution or Revolution for Research?"*, Franco-German Science Circle Lecture, New Delhi, India, 20 February 2019. *"Uncertainty quantification with application to medical engineering: towards a more reliable medicine?"*, Keynote Lecture, GAMM Workshop, Pisa, Italy, 12 September 2019. *"Electric Network Optimization in Germany"*, invited talk at Xi'an University of Technology (XAUT), Xi'an, China, 28 October 2019. *"Hiflow3: An open source multi-purpose finite element software"*, invited talk at Peking University (PKU), Beijing, China, 1 November 2019. Fireside Chat *"Academia oder Firma",* Heidelberg, Germany, 11 November 2019. *„Was ist Fortschritt?",* Montagskonferenz, Institut für Übersetzen und Dolmetschen, Heidelberg University, Germany, 18 November 2019. *„Algorithmen & künstliche Intelligenz: Wahrer Fortschritt oder doch nur digitale Alchemie?"*, Lecture Series *„Dialog im Museum"*, Mercedes-Benz-Museum, Stuttgart, Germany, 3 December 2019. (Video: https://youtu.be/3fctDjZJxl0). *„Uncertainty Quantification for engineering and industry: perspectives and challenges for medical applications"*, Sino-German Symposium on UQ for Industrial Applications, Studio Villa Bosch, Heidelberg, Germany, 4 December 2019.

Daria Kokh:
"Exploring dynamics and kinetics of ligand unbinding in GPCRs", German workshop on "Structural prediction of membrane proteins", Forschungszentrum Jülich, 26-27 November, 2019.

Gye-Seon Lee:
"Hitchin components for orbifolds", SPP conference *"Geometry at infinity"*, University of Münster, Germany, 1 - 5 April 2019. *"Coxeter groups in real projective geometry"*, Perspectives on convex projective geometry, Sète, France, 24 - 28 June, 2019.

Sebastian Lerch:
"Statistics vs. machine learning for complex problems: White, black or grey boxes?" Panel discussion, American

Meteorological Society, 99th Annual Meeting, 18th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences, Phoenix, Arizona, United States, 9 January 2019. *"Forecaster's dilemma: Forecast evaluation and extreme events"*. VALPRED Workshop on the Assessment of Ensemble Forecasts, Aussois, France, 20 March 2019. International Conference on Extreme Value Analysis, Zagreb, Croatia, 2 July 2019. *"Neural networks for postprocessing ensemble weather forecasts"*. Helmholtz Digital Earth Workshop, Munich, Germany, 19 July 2019. Workshop on Machine Learning for Weather and Climate Modeling, Oxford, UK, 4 September 2019. *"Neural networks for postprocessing ensemble weather forecasts"*. Workshop on Data-driven Modeling in Fluid Mechanics, Karlsruhe, Germany, 17 September 2019. *"Machine learning for postprocessing: Paths and pitfalls"*. EUMETNET Workshop on Post-Processing, Bruxelles, Belgium, 10 December 2019.

Goutam Mukherjee:
"How Does CYP Sequence Affect CYP:CPR Complexation in a Phospholipid Bilayer and the Transfer of Electrons?" High Performance Computing in Science & Engineering - 22nd Results and Review Workshop of the HLRS, Stuttgart, 7 October 2019. *"Deciphering Biomolecular Recognition by Molecular Modeling Approaches."* CSIR-IMTECH Institute of Microbial Technology, Chandigarh, India, 10 December 2019. Department of Chemistry, IIT Ropar, Rupnagar, India, 11 December 2019.

Andreas Ott:
"Bounded cohomology in large degree", Geometry Seminar, University of Cologne, January 2019.
Seminar on Groups and Geometry, University of Geneva, May 2019.

Kai Polsterer:
"Machine Learning in Astronomy: accessing complex structures with unsupervised and deep-learning techniques", BASP, Villas-sur-Ollon, Switzerland, 4 February 2019. *"From Photometric Redshift to Improved Weather Forecasts"*, Astroinformatics 2019, Pasadena, USA, 16 July 2019. *"From Photometric Redshift to Improved Weather Forecasts: an interdisciplinary view on machine learning in astronomy"*, Herbsttagung der Astronomischen Gesellschaft, Stuttgart, Germany, 17 September 2019; Sino-German workshop in the area of Uncertainty Quantification in industrial applications, Heidelberg, Germany, 2 December 2019; Machine Learning Tools for Research in Astronomy Workshop, Max Planck Institute for Astronomy, Ringberg, Germany, 9 December 2019. *"Machine Learning in Astronomy: lessons learned from learning machines"*, Colloquium, University of Western Australia, Perth, Australia, 11 November 2019.

Maria Beatrice Pozzetti:
"Critical exponent and Hausdorff dimension for Anosov representations", Conference on Geometric Structures in Nice, France, January 2019. *"Orbit growth rate for maximal representations"*, Workshop Riemannian and Simplicial Volume, Karlsruhe, Germany, 8 April 2019. *"The critical exponent on higher rank Teichmüller spaces"*, Conference Aspects of Non-Positive and Negative Curvature in Group Theory, Luminy, France, 19 June 2019. *"Compactifications of Hitchin and maximal charachter variety"*, Arbeitstagung 2019, Bonn, Germany, 9 July 2019. Connections for Women: Holomorphic Differentials in Mathematics and Physics, MSRI, USA, 15 August 2019. *"Discrete subgroups of semisimple Lie groups"*, Forum des jeunes mathematicien-ne-sy, Paris, France, October 2019. *"Buildings and compactifications of character varieties"*, Conference Buildings, Varieties, and Applications, Leipzig, Germany, November 2019.

Friedrich Röpke:
"Testing Type Ia supernova progenitor channels in a consistent modeling pipeline", talk at the Progenitors of Type Ia Supernovae conference, Lijiang, China, 8 August 2019. Podium discussion at the Progenitors of Type Ia Supernovae conference, Lijiang, China, 8 August 2019.

Evgenii Rogozinnikov:
"Noncommutative coordinates for maximal symplectic representations", Séminaire GT3 (Geometry seminar), IRMA, Strasbourg, France, December 2019. Geometry, Symmetry and Physics Seminar, Rutgers University, USA, December 2019.

S. Kashif Sadiq:
"Modeling enzyme reaction-modulated assembly of phase-separated biomolecular condensates", CECAM Workshop - Challenges in Large Scale Biomolecular Simulation 2019: Bridging Theory and Experiments; Cargèse Institute of Scientific Studies, Cargèse, France, 13-17 May 2019.

Andrew Sanders:
"An introduction to (G; P)-opers", Geometry Seminar, Simons Center of Geometry and Physics, Stonybrook, New York, USA, February 2019. *„Harmonic maps and p-adic super-rigidity following Gromov and Schoen"*, Log cabin retreat: Harmonic maps and rigidity. Sisteron, France, April 2019. *"(G; P)-opers"*, Conference in honor of the 60th birthday of Yves Benoist, Cetraro, Italy. May 2019. *"Regular immersions in symmetric spaces and a priori inequalities"*, Conference on Higher Teichmüller theory and related topics, Pavia, Italy, June 2019.

Anna Sofie Schilling:
"Horofunction and Generalized Satake Compactifications of Symmetric Spaces", Conference Perspectives on convex projective geometry, Sète, France, June 24-28, 2019.

Chen Song:
"Uncertainty quantification for the simulation of the human heart", BIOTRONIK SE & Co. KG, Berlin, Germany, 11 June 2019. Physikalisch-Technische Bundesanstalt (PTB), 12 June 2019, Berlin, Germany, 2019. *"Application of uncertainty quantification for rotating device"*, invited talk at Xi'an University of Technology (XAUT), Xi'an, China, 28 October 2019. *"Hiflow3: An open source multi-purpose finite element software, Invited talk at Peking University (PKU)"*, Beijing, China, 1 November 2019. *"Uncertainty assessment of the blood damage in a FDA blood pump"*, invited talk at Sino-German Symposium on Uncertainty Quantification for Engineering and Industrial Applications, Heidelberg, 5 December 2019.

Alexandros Stamatakis:
"Hochleistungsrechnen könnte so schön sein ... wenn nur die Software nicht wäre", annual meeting of the DFN, Wissenschaftszentrum Bonn, Germany, December 2019.

Rebecca C. Wade:
"Towards computationally efficient approaches to study drug binding kinetics", CECAM Workshop on "Multiscale Modeling from Macromolecules to Cell: Opportunities and Challenges of Biomolecular Simulations", EPFL, Lausanne, Switzerland, 4-6 Feb 2019. *"Insights into the Membrane and Protein Interactions of Cytochrome P450 Enzymes from Molecular Simulations"*, Workshop on "Towards Simulating Cell Membranes: Closer to Reality", Studio Villa Bosch, Heidelberg, Germany, 6-7 Feb 2019. Colloquium, DynaMem Consortium, FIAS, Frankfurt, Germany, 27 March 2019. ICCP450, University of Queensland, Brisbane, Australia, 24 June 2019. *"Exploring the determinants of protein crowding effects by molecular simulation"*, Protein Society Meeting, Seattle, USA, 29 June-3 July 2019. *"Machine Learning in Computer-aided Drug Design"*, BioRN Annual Conference on "Artificial Intelligence meets Health", DKFZ, Heidelberg, 11 Nov 2019. *"Structure-based Approaches to the Discovery of Anti-Parasitic Compounds"*, CSIR-Institute of Genomics and Integrative Biology, New Delhi, India, 2 Dec 2019. *"Computational Approaches to Protein Dynamics and Binding Kinetics for Drug Discovery"*, Department of Chemistry, Indian Institute of Technology Delhi, New Delhi, India, 2 Dec 2019. AAPS Student Chapter, NIPER, S.A.S. Nagar, India, 4 Dec 2019. MM2019, Association of Molecular Modellers of Australasia (AMMA) Conference, Bintan, Indonesia, 4-8 Dec 2019.

Anna Wienhard:
"Hausdorff dimension and critical exponent", Dubrovnik IX, Topology and Dynamical Systems, Dubrovnik, Croatia, 28 June 2019. *"Hyperbolische Geometrische Strukturen - von der mathematischen Theorie bis zum maschinellen Lerrnen"*, Heidelberger Akademie der Wissenschaften, Heidelberg, Germany, July 2019. *"Non-commutative hyperbolic geometry"*, Colloquium, UC Berkeley, USA, October 2019. *"Higher Teichmüller spaces and positivity"*, QMAP Colloquium, UC Davis, 8 November 2019, Davis, USA. *"Higher Teichmüller spaces and Anosov representations"*, Workshop Holomorphic Differentials in Mathematics and Physics, MSRI, Berkeley, USA, 21 November 2019. *"Geometry, Topology, and Applications"*, Developments in the Mathematical Sciences, MPIMIS Leipzig, Germany, 27 November 2019.

Ulrike Wittig:
"Datenmanagement für Forschungsprojekte", Workshop "Persistente Identifikatoren (PID) - Services für Forschungsdaten im Bereich Lebenswissenschaften", Cologne, Germany, 23 January 2019. "FAIR BioData Management", Workshop "Ready for BioData Management?", Lisbon, Portugal, 2 July 2019.

9.2 Presentations

Talks

Robert Andrassy:
"Convection at the beginning and at the end of massive star evolution", talk at the Stellar Hydro Days V, Exeter, 26 June 2019. Seminar talk, Max Planck Institute for Astronomy, Heidelberg, Germany, 10 December 2019. *"Convective-reactive processes in evolved massive stars"*, talk at the Nuclear Physics in Astrophysics IX conference, Schloss Waldthausen, Budenheim, Germany, 18 September 2019.

Jonas Beyrer:
"Marked length spectrum rigidity for actions on CAT(0) cube complexes", Workshop on bounded cohomology, Heidelberg University, Heidelberg, 31 July 2019.

Csaba Daday:
"Mechanical stability of membranes upon indentation", Towards Simulating Cellular Membranes Workshop, Heidelberg, Germany, 9 February 2019.

Madhura De:
"Positioning the Linker Histone on a Chromatosome: What Role Does the DNA Play?", Multiscale Modeling of Chromatin: Bridging Experiment with Theory, Les Houches, France, 31 March –5 April 2019.

Timo Dimitriadis:
"Encompassing tests for higher-order elicitable functionals". Quantitative Finance and Financial Econometrics Conference, Marseille, France, 5 June 2019. Annual Conference of the International Association of Applied Econometrics, Nicosia, Cyprus, 25 June 2019. 34th Annual Congress of the European Economic Association, 72nd European Meeting of the Econometric Society, Manchester, UK, 29 August 2019. *"Estimating realized variance: An intrinsic time approach"*, Quantitative Finance and Financial Econometrics Conference, Marseille, France, 1 June 2019. 6th Hohenheim Finance Workshop, Bad Liebenzell, Germany, 6 June 2019. *"Evaluating the rationality of mode forecasts"*, HeiKaMEtrics Workshop, Mannheim, Germany. 6 September 2019. Statistische Woche, Universität Trier, Trier, Germany, 11 September 2019. *"Testing forecast rationality for measures of central tendency"*, 30th (EC)^2 conference on identification in macroeconomics, St. Anne's College, Oxford, UK, 12 December 2019.

Antonio D'Isanto:
"Is VO ready for machine learning?", IVOA InterOp May 2019, Paris, France, 13 May 2019. *"ESCAPE to victory - Building the infrastructure for the next generation astronomy"*, Astroinformatics 2019, Pasadena, USA, 26 June 2019; Meeting of the German Astronomical Society, Stuttgart, Germany, 17 September 2019.

Philipp Gerstner, Jonas Kratzke, Chen Song:
"Uncertainty quantification for the simulation of the human heart", Workshop on Cardiac Modeling, Bad Herrenalb, Germany, 15-17 April 2019.

Nikos Gianniotis:
"Mixed Variational Inference", International Joint Conference on Neural Networks, Budapest, Hungary, 18 July 2019.

Martin Golebiewski:
"Introduction into standards: Formal and de facto standards", EU-STANDS4PM Kick-Off Meeting, Brussels, Belgium, 7 March 2019. *"MulticellML - Standardising the exchange of multicellular models in systems medicine"*, HARMONY 2019 - 9th Hackathon on Resources for Modeling in Biology, California Institute of Technology (Caltech), Pasadena, California, USA, 25-29 March 2019. *"EU-STANDS4PM - A European standardization framework for data integration and data-driven in silico models for personalized medicine"*, HARMONY 2019 - 9th Hackathon on Resources for Modeling in Biology, California Institute of Technology (Caltech), Pasadena, California, USA, 25-29 March 2019; Meeting of ISO/TC 276 Biotechnology, Toronto, Canada, 2-6 December 2019. *"Two universes – one world: Community standards vs. formal standards in systems biology and systems medicine"*, COMBINE 2019: 10th Computational Modeling in Biology Network Meeting, Heidelberg, Germany, 15-19 July 2019. *"How to Share Your Data FAIR – Integrated Data Management for Systems Biology & Systems Medicine"*, International Conference on Systems Biology (ICSB 2019), Okinawa, Japan, 31 October – 5 November 2019. *"COMBINE - A Community of Communities"*, International Conference on Systems Biology (ICSB 2019), Okinawa, Japan, 31 October – 5 November 2019.

Martin Golebiewski and Wolfgang Müller:
"The LiSyM Data Management", Liver Systems Medicine (LiSyM) Jamboree, Leipzig, Germany, 9-10 May 2019.

Sabrina Gronow:
"A sub-Chandrasekhar mass white dwarf as possible progenitor for a thermonuclear explosion", talk at Supernova Remnants II - An Odyssey in Space after Stellar Death conference, Chania (Crete), Greece, 4 June 2019. *"Double detonations of sub-Chandrasekhar mass white dwarfs"*, talk at the Progenitors of Type Ia Supernovae conference, Lijiang, China, 8 August 2019.

Saskia Haupt:
"How mathematics can help in the fight against cancer", European Hereditary Tumor Group (EHTG) Meeting 2019, Barcelona, Spain, 17-20 October 2019.

Erica Hopkins:
"Can Crowdsourcing be Replaced by GPUs?", AG 2019 - Meeting of the German Astronomical Society, Stuttgart, Germany, 18 September 2019.

Leonhard Horst, Philipp Edelmann, Raphael Hirschi, Friedrich Röpke:
"Convective Boundary Mixing for He Shell-Burning" poster at the Stellar Hydro Days V, Exeter, UK, 24-28 June 2019.

Fan Jin:
"Molecular Mechanism for Phosphorylation of Intrinsically Disordered Proteins: Insights from Computer Simulations and Implications for Biological Function". Workshop on computer simulation and theory of macromolecules. Hünfeld, Germany, 22-23 March 2019.

Markus Kurth, Rainer Beck, Camilo A. Aponte-Santamaria, Britta Bruegger:
"Molecular dynamics simulation and photo-crosslinking reveal a specific cholesterol binding site for the Metabotropic glutamate receptor 2", Talk in the Biophysical Society Meeting, Baltimore, USA, 2-6 March 2019.

Florian Lach:
"Chandrasekhar Mass Explosions: Pure Deflagrations as a Model for Type Iax Supernovae", talk at the Progenitors of Type Ia Supernovae conference, Lijiang, China, 8 August 2019.

Florian Lach, Friedrich Röpke:
"Nucleosynthesis of Different Type Ia Supernova Explosions", poster at the Nuclear Physics in Astrophysics IX conference, Schloss Waldthausen, Budenheim, 15-20 September 2019.

Florian Lach, Friedrich Röpke, Markus Kromer:
"Chandrasekhar Mass Deflagrations as a Model for Type Iax Supernovae", poster contribution at the Supernova Remnants II -- An Odyssey in Space after Stellar Death conference, Chania (Crete), Greece, 3-8 June 2019.

Sebastian Lerch:
"Neural networks for postprocessing ensemble weather forecasts", American Meteorological Society, 99th Annual Meeting, 18th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences, Phoenix, Arizona, United States, 9 January 2019. European Geosciences Union General Assembly, Vienna, Austria, 11 April 2019. *"Predictive inference based on Markov chain Monte Carlo output"*, Annual Meeting of the German Mathematical Society, Karlsruhe, Germany, 25 September 2019.

Philipp Lösel:
"Biomedisa: 30 million-year-old fossils come back to life", HeKKSaGOn University Consortium, The 7th German – Japanese University Presidents' Conference, Heidelberg, Germany, 12-13 September 2019.

Isabel Martin:
"Force distribution in protein allosteric regulation: The case of Integrin-linked kinase", Molecular Stresses Workshop, Heidelberg, Germany, 4-5 November, 2019.

Nicholas Michelarakis:
"MARTINIZING Nafion", Toyota, Multiscale workshop, Heraklion, Greece, 10-11 October 2019.

Wolfgang Müller:
"Introduction to FAIR", Workshop "FAIR data infrastructures for biomedical communities", 64th Annual Conference of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS), Dortmund, Germany, 8-11 September 2019.

Sotirios Nikas:
"Towards a machine learning model for predicting time and power consumption of compute kernels", 9. HPC-Status-Konferenz der Gauß-Allianz, Paderborn, Germany, 17 October 2019.

Ariane Nunes-Alves:
"Effects of macromolecular crowding on the diffusion rates of enzyme substrates and drug-like molecules", Workshop on "Computer simulation and theory of macromolecules", Huenfeld, Germany, 22-23 March 2019. *"How can the cell environment affect the encounter of molecular binding partners? Effects of macromolecular crowding on the diffusion rates of enzyme substrates and drug-like molecules"*, 4th Kloster Schöntal Retreat of the DKFZ-ZMBH Alliance, Schöntal, Germany, 22-24 July 2019.

Mareike Pfeil:
"Wie Mathematiker Billard spielen", Opening Ceremony, STRUCTURES - Cluster of Excellence, Heidelberg University, Heidelberg, Germany, 15 July 2019.

Benedikt Rennekamp:
"Bond Scission in Tensed Collagen: A Hybrid Monte Carlo / Molecular Dynamics Approach", Workshop: Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, 22-23 March 2019.

Maja Rey:
"Data Management for Systems Biology Projects", de.NBI Course "Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOM-Hub/SEEK", Heidelberg, Germany, 18-20 March 2019. *"de.NBI-SysBio (ELIXIR Germany) serving Systems Biology community"*, ELIXIR Data Platform Kickoff Meeting, EBI Hinxton, UK, 15-16 October 2019.

Friedrich Röpke:
"Modeling explosions and nucleosynthesis in electron-capture induced thermonuclear supernovae", talk at the XIXth Workshop on Nuclear Astrophysics, Ringberg Castle, Tegernsee, Germany, 25 March, 2019. *"Simulating commmon envelope phases with AREPO"*, talk at the Common Envelope Evolution 2019 Workshop, Flatiron Institute, New York, USA, 7 May 2019. *"Electron-capture initiated stellar collapse"*, talk at the Electron-Capture-Initiated Stellar Collapse Workshop, Lorentz Center, Leiden, The Netherlands, 20 May 2019. *"Hydrodynamical models of binary star interactions",* talk at the Stellar Hydro Days V, Exeter, UK, 26 June 2019. *"Electron-capture induced thermonuclear supernovae: explosion and nucleosynthesis"*, talk at the Nuclear Physics in Astrophysics IX conference, Schloss Waldthausen, Budenheim, Germany, 19 September 2019.

Evgenii Rogozinnikov:
"Noncommutative coordinates for maximal symplectic representations", Omega Alpha Seminar, Heidelberg University, Germany, 5 December 2019.

Anna Sofie Schilling:
"Horofunction and Satake Compactification of Symmetric Spaces", Conference "Geometric Analysis meets Geometric Topology", Heidelberg University, Heidelberg, Germany, 25-28 February 2019.

Chen Song:
"Uncertainty assessment of the blood damage in a FDA blood pump", International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP), Crete, Greece, 25 June 2019.

Eva-Maria Walz:
"Generalization of receiver operating characteristic (ROC) curves – ROC movies, uROC and CPA", Amazon Research Days, Berlin, Germany, 15 October 2019.

Andreas Weidemann:
"SABIO-RK Introduction", International Conference on Systems Biology (ICSB 2019), Okinawa, Japan, 31 October 2019.

Ulrike Wittig:
"SABIO-RK Introduction", de.NBI Course "Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOMHub/SEEK", Heidelberg, Germany, March 18- 20, 2019. *"Collecting, Curating, Interlinking, and Sharing Data"*, Beilstein Enzymology Symposium 2019 "Molecular Functions, Catalysis and Regulation", Rüdesheim, Germany, 10-12 September 2019. *"Data Management Activities in de.NBI / Elixir-Germany"*, ELIXIR CONVERGE Data Management Network Workshop, Noordwijkerhout, The Netherlands, 14-15 October 2019. *"FAIR BioData Management"*, MIRACUM Data Management Workshop, Mannheim, Germany, 19-20 November 2019.

Adrian Zapletal:
"Data Distribution for Phylogenetic Inference with Site Repeats via Judicious Hypergraph Partitioning", HICOMB workshop, in conjunction with IEEE IPDPS conference, Rio de Janeiro, Brazil, May 2019.

Christopher Zapp:
"Mechanoradicals in tensed tendon collagen as a new source of oxidative stress", Multiscale Mechanochemistry & Mechanobiology, McGill University, Montreal, Canada, 29-31 July, 2019.

Posters

Svenja de Buhr and Frauke Gräter:
"Mechanical Stimuli during Retina Stem Cell Differentiation". 3D Matter Made to Order Young Scientists Retreat. Bad Herrenalb, Germany, 5-6 September 2019.

Csaba Daday and Frauke Gräter:
Unfolding Focal Adhesion Kinase: Getting Cellular Insight through AFM and MD Biophysical Society Meeting, Baltimore, 6 March 2019; Workshop on computer simulation and theory of macromolecules. Hünfeld, Germany, 22-23 March 2019.

Juliane Fluck, Iris Pigeot, Birte Lindstädt, Thomas Gübitz, Hajo Zeeb, Wolfgang Ahrens, Markus Löffler, Oana Brosteanu, Ulrich Lang, Constanze Curdt, Sebastian Claudius Semler, Oya Beyan, Hubertus Neuhausen, Jens Dierkes, Ulrich Sax, Harald Kusch, Henriette Senst, Thilo Muth, Sylvia Thun, Dietrich Kaiser, Jochen Dress, Wolfgang Müller, Martin Golebiewski:
"NFDI4Health: Ein Konzept für eine föderierte Forschungsdateninfrastruktur personenbezogener Gesundheitsdaten", 64th Annual Conference of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS), Dortmund, Germany, 8-11 September 2019.

Sucheta Ghosh, Maja Rey, Ulrike Wittig, Wolfgang Mueller:
"Bio-Entity Recognition for SABIO-RK Database", 12th International Biocuration Conference, Cambridge, UK, 7-10 April 2019.

Martin Golebiewski, Xiaoming Hu, 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", Liver Systems Medicine Network (LiSyM) Jamboree, Leipzig, Germany, 9-10 May 2019.

Martin Golebiewski, Xiaoming Hu, Olga Krebs, Maja Rey, Andreas Weidemann, Ulrike Wittig, Wolfgang Müller & the EU-STANDS4PM project and COMBINE community:
"FAIR data exchange in the life sciences by standardization of heterogenous data and models", International Conference on Systems Biology (ICSB 2019), Okinawa, Japan, 31 October - 5 November 2019.

Ganna Gryn'ova:
"Topology-Driven Molecular Electronics Lacking π-Conjugation", 9th Molecular Quantum Mechanics Conference, Heidelberg, Germany, 30 June – 5 July 2019; 10th Triennial Congress of the International Society for Theoretical Chemical Physics, Tromsø, Norway, 11-17 July 2019.

Saskia Haupt:
"Mathematical Modeling of the Pathogenesis of Mismatch Repair-Deficient Cancers", CSBC-PSON Mathematical Oncology Meeting 2019, Portland, USA, 13-14 May 2019.

Ana Maria Herrera Rodriguez and Frauke Gräter:
"The role of flow in the directed self-assembly of dragline spider silk proteins", Silk conference, Trento, Italy, 12-15 June 2019.

Fan Jin, Sebastian Richter, Frauke Gräter and Frauke Melchior:
"SUMO's intrinsically disordered N-termini are intramolecular regulators of SUMO-SIM interactions", EMBO workshop: Synergy of experiment and computation in quantitative systems biology, Nové Hrady, Czech Republic, 23-28 June 2019.

Fan Jin and Frauke Gräter:
"Molecular mechanisms of E-Cadherin mediated mechanosensing", CECAM Summer School: Applied mathematics and machine learning perspectives on Big Data Problems in Computational Sciences, Mainz, Germany. 30 Sept – 4 Oct 2019.

Olga Krebs, Martin Golebiewski, Stuart Owen, Maja Rey, Natalie Stanford, Ulrike Wittig, Xiaoming Hu, Katy Wolstencroft, Jacky L. Snoep, Bernd Rinn, Wolfgang Müller, Carole Goble:
"FAIRDOM approach for Systems Biology Data and Model Management", COMBINE 2019: 10th Computational Modeling in Biology Network Meeting, Heidelberg, Germany, 15-19 July 2019.

Olga Krebs, Martin Golebiewski, Stuart Owen, Maja Rey, Natalie Stanford, Ulrike Wittig, Xiaoming Hu, Katy Wolstencroft, Jacky L. Snoep, Bernd Rinn, Wolfgang Müller, Carole Goble:
"FAIRDOM approach for Systems Biology Data and Model Management", ISMB/ECCB 2019 - International Conference on Intelligent Systems for Molecular Biology & European Conference on Computational Biology Conference, Basel, Switzerland, 21-25 July 2019.

Markus Kurth, Rainer Beck, Camilo A. Aponte-Santamaria, Britta Bruegger:
"Probing the lipid environment of the Metabotropic glutamate receptor 2", Workshop: Towards Simulating Cell Membranes: Closer to Reality, Studio Villa Bosch, Heidelberg, 6-7 February 2019.

Isabel Martin and Frauke Gräter:
"Mechanosensing in the Focal Adhesions: Pseudokinase ILK studied with MD Simulations", Workshop on computer simulation and theory of macromolecules, Hünfeld, Germany, 22-23 March 2019.

Benedikt Rennekamp and Frauke Gräter:
"Mechnoradicals in tensed collagen in hybrid simulations", Summer School of the Max Planck School Matter to Life, Heidelberg, Germany, 2-6 September 2019; IWR School Machine Learning with Applications in Natural- and Life Sciences (ML4Nature), 23-27 September 2019, Heidelberg, Germany. *"Hybrid Simulations of Mechanoradicals in Proteins"*, Ringberg Symposium Matter to Life, Tagungsstätte Schloss Ringberg, Kreuth, Germany, 15-18 December 2019.

Ana Maria Herrera Rodriguez and Frauke Gräter:
"The role of flow in the directed self-assembly of dragline spider silk proteins", Silk conference, Trento, Italy, 12-15 June 2019.

Anna Sofie Schilling:
"Horofunction Compactification of Rn with Polyhedral Norms", European Women in Math, German Chapter Conference, Leipzig, Germany, 28-30 October 2019.

Andreas Weidemann, Maja Rey, Ulrike Wittig, Wolfgang Müller:
"SABIO-RK - a curated database for reaction kinetics", ELIXIR All Hands Meeting, Lisbon, Portugal, 17-20 June 2019.

Andreas Weidemann, Ulrike Wittig, Maja Rey, Wolfgang Müller:
"SABIO-RK - kinetic data for systems biology", COMBINE 2019: 10th Computational Modeling in Biology Network Meeting, Heidelberg, Germany, 15-19 July 2019.

Ulrike Wittig, Maja Rey, Andreas Weidemann, Wolfgang Müller:
"SABIO-RK: extraction of enzyme function data from STRENDA DB", Beilstein Enzymology Symposium 2019 "Molecular Functions, Catalysis and Regulation", Rüdesheim, Germany, 10-12 September 2019.

Andreas Weidemann, Ulrike Wittig, Maja Rey, Wolfgang Müller:
"SABIO-RK: a comprehensive reaction kinetics information system", International Conference on Systems Biology (ICSB 2019), Okinawa, Japan, 31 October – 5 November 2019.

Ulrike Wittig, Natalie Stanford, Stuart Owen, Olga Krebs, Martin Golebiewski, Maja Rey, Alan Williams, Finn Bacall, Jacky Snoep, Carole Goble, Wolfgang Müller:
"FAIRDOM: supporting FAIR data and model management", 12th International Biocuration Conference, Cambridge, UK, 7-10 April 2019.

Ulrike Wittig, Carsten Kettner, STRENDA Commission:
"STRENDA DB - The 'PDB' for Enzyme Function Data?", 12th International Biocuration Conference, Cambridge, UK, 7-10 April 2019.

Ulrike Wittig, Maja Rey, Andreas Weidemann, Wolfgang Müller:
"SABIO-RK: extraction of enzyme function data from STRENDA DB", 12th International Biocuration Conference, Cambridge (UK), 7-10 April 2019.

Demos

Alexandros Tsengenes, Christina Athanasiou:
Protein Modeling and Drug Design Demonstration, Researcher's Night, NHRF & Hellenic Cosmos, Athens, Greece, 20 & 27 September 2019.

9.3 Memberships

Tilmann Gneiting:
Editor-In-Chief, Annals of Applied Statistics (through 7 January 2019). 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 (through 7 January 2019).

Martin Golebiewski:
Convenor (chair) of ISO/TC 276 Biotechnology working group 5 *"Data Processing and Integration"*, International Organization for Standardization (ISO). Chair of the project group *"FAIR Data Infrastructures for Biomedical Informatics"* of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS). Member of the board of coordinators of COMBINE (Computational Modeling in Biology network). Co-chair of the ISO/IEC Joint Ad-hoc Group on Standardization of Genomic Information Compression and Storage (MPEG-G). 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 Richtlinienausschuss (German committee for engineering standards) VDI 6320 *"Datenmanagement im Bereich Life Sciences"*, Association of German Engineers (VDI). Member of the management committee of the European COST action CHARME (Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research).

Frauke Gräter:
Max Planck Fellow of the Max Planck School Matter to Life (since 2019). Fellow of the Marsilius Kolleg, Heidelberg University (2019-2020). 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). Member of the Biophysical Society. 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. Referee of the Biophysical Journal, Journal of the American Chemical Society, PLoS Journals, Nature Journals, Proceedings of the National Academy of Sciences, German Research Society (DFG).

Ganna Gryn'ova:
Affiliated junior research group leader, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University.

Sebastian Lerch:
Associate Editor, Monthly Weather Review. Guest Editor, Nonlinear Processes in Geophysics.

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.

Ariane Nunes-Alves:
Member of the Early Career Board of the Journal of Chemical Information and Modeling.

Kai Polsterer:
Vice president International AstroInformatics Association. Chair of the Knowledge Discovery in Databases Interest Group of the International Virtual Observatory Alliance. Member of the Astronomische Gesellschaft (A.G.). 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. Member of the DPG: Arbeitskreis Physik, moderne Informationstechnologie und Künstliche Intelligenz.

Friedrich Röpke:
Member of the Steering Board of the Collaborative Research Center SFB 881 *"The Milky Way System"*. Member of the Board of the International Max Planck Research School Heidelberg. Member of the Research Council of the Field of Focus 2, Heidelberg University. Member of the Directorial Board, Center for Astronomy at Heidelberg University (ZAH). Advisory board, *"Sterne und Weltraum"*. 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. Member of scientific committee of the SMPGD (Statistical Methods for Post Genomic Data analysis) workshop series.

Michael Strube:
Research Training Group 1994, Adaptive Preparation of Information from Heterogeneous Sources (AIPHES), TU Darmstadt/Heidelberg University/HITS. Associate Editor: Journal of Artificial Intelligence Research. Fellow of the Marsilius Kolleg, Heidelberg University (2019-2020).

Rebecca C. Wade:
Associate Editor: Journal of Molecular Recognition, PLOS Computational Biology. Editorial Board: BBA General Subjects; Journal of Chemical Information and Modeling; Journal of Computer-aided Molecular Design; International Journal of Molecular Sciences; Biopolymers; Protein Engineering, Design and Selection; Advances and Applications in Bioinformatics and Chemistry. Member of Scientific Advisory Council of the Leibniz-Institut für Molekulare Pharmakologie (FMP), Berlin-Buch. Member of Scientific Advisory Council of the Computational Biology Unit (CBU), University of Bergen, Norway. Member of Scientific Advisory Board of the Max Planck Institute of Biophysics, Frankfurt. 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, Informatics4Life Consortium, Interdisciplinary Center for Scientific Computing (IWR), DKFZ-ZMBH Alliance of the German Cancer Research Center and the Center for Molecular Biology at Heidelberg University (ZMBH).

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. Member of the Scientific Advisory Board Mathematisches Forschungsinstitut Oberwolfach. Member of the Scientific Committee Heidelberg Laureate Forum. Editor Geometry & Topology. Editor Annales scientifiques de l'école normale supérieure. Editor Geometriae Dedicata. Editor Forum Mathematicum. Editor Geometric and Functional Analysis. Editor Annales Henri Lebesgue. Editor Proceedings of the London Mathematical Society.

Ulrike Wittig:
Member of the STRENDA Commission (Standards for Reporting Enzymology Data).

9.4 Contributions to the Scientific Community

Program Committee Memberships

Sucheta Ghosh:
57th Annual Meeting of the Association for Computational Linguistics (ACL), Florence, Italy, 28 July – 2 August 2019. Interspeech, Graz, Austria, 15-19 September 2019. EMN-LP-IJCNLP 2019, Hong Kong, China, 3-7 Nov 2019.

Michael Strube:
Tutorial Chair at the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, Minnesota, 2-7 June 2019. Area Chair at the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July – 2 August 2019. Area Chair at the 2019 Conference on Empirical Methods in NLP and 9th International Joint Conference on Natural Language Processing, Hong Kong, China 3-7 November 2019.

Ulrike Wittig:
12th International Biocuration Conference, Cambridge, UK, 7-10 April 2019.

Workshop and Conference Organization

Madura De:
Organizer: 6th Heidelberg Forum for Young Life Scientists: 'Zooming Out: From Molecules to Patients', 6-7 June 2019, EMBL & DKFZ Heidelberg, Germany.

Martin Golebiewski:
Committee Meetings of ISO/TC 276 Biotechnology working group WG5 *"Data Processing and Integration"*, Tokyo, Japan, 10-15 June 2019, and Toronto, Canada, 2-6 December 2019. COMBINE 2019: 10th Computational Modeling in Biology Network Meeting, Heidelberg, Germany, 15-19 July 2019. EU-STANDS4PM workshop *"A European standardization framework for data integration and data-driven in silico models for personalized medicine"*, Heidelberg, Germany, 18 July 2019. Workshop *"FAIR data infrastructures for biomedical communities"*, 64th Annual Conference of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS), Dortmund, Germany, 10 September 2019. COMBINE & de.NBI Tutorial *"Modeling and Simulation Tools in Systems Biology"*, International Conference on Systems Biology (ICSB 2019), Okinawa, Japan, 31 October 2019.

Frauke Gräter:
Workshop on *'Towards simulating cell membranes: Closer to Reality?'* (with Lipi Thukral), Studio Villa Bosch, Heidelberg, Germany, 6-7 February 2019. Workshop on *'Molecular Stresses'*, Studio Villa Bosch, Heidelberg, Germany, 4-5 November 2019.

Vincent Heuveline:
50 Jahre Rechenzentrum, Festkolloquium, Heidelberg University, Germany, 11 October 2019. Vortragsreihe IT Forum 2019, Focus Blockchain, Heidelberg, Germany, 23 & 30 January 2019.

Vincent Heuveline, Chen Song:
Sino-German Symposium on UQ for Industrial Applications, Studio Villa Bosch, Heidelberg, Germany, 2-6 December 2019.

Olga Krebs:
FAIRDOM PALs and User Workshop, Heidelberg, Germany, 19 July 2019.

Andreas Ott:
Workshop *"Bounded Cohomology"*, Heidelberg University, Heidelberg, Germany, 29 July – 2 August 2019.

Kai Polsterer:
Organizing a session of the KDDIG at the International Virtual Observatory Alliance Interoperability Meeting in Paris, France, 12-17 May 2019. Member of the scientific organizing committee of Astroinformatics 2019, Pasadena, USA, 24-27 June 2019. Organizing the E-Science and E-Infrastructure splinter meeting at the AG-Tagung in Stuttgart, Germany, 16-20 September 2019. Member of the scientific organizing committee of ML@Ringberg, Germany, 9-13 December 2019.

Maria Beatrice Pozzetti:
"Geometric Analysis meets Geometric Topology" (with E. Mäder-Baumdicker, V. Disarlo, F. Fanoni and F. Dittberner), Heidelberg University, Heidelberg, Germany, 25-28 February 2019.
"Groups, spaces and geometries", conference in honor of Alessandra Iozzi (with C. Burrin, T. Hartnick, E. Kowalsky and A. Wienhard), FIM Zürich, Switzerland, 22-25 January 2019.

Friedrich Röpke:
Member of the Scientific Organizing Committee for the Electron-Capture-Initiated Stellar Collapse Workshop, Lorentz Center, Leiden, The Netherlands, 20-24 May 2019. Member of the Scientific Organizing Committee for the Progenitors of Type Ia Supernovae conference, Lijiang, China, 5-9 August 2019.

Friedrich Röpke, Robert Andrassy:
Member of the Scientific Organizing Committee for the Stellar Hydro Days V, Exeter, UK, 24-28 June 2019.

Friedrich Röpke, Sabrina Gronow, Robert Andrassy, Johann Higl, Christian Sand, Fabian Schneider, Leonhard Horst:
Members of the Organizing Committee for the *"14th Würzburg Workshop in Heidelberg"*, Studio Villa Bosch, Heidelberg, Germany, 16-17 December 2019.

Alexandros Stamatakis:
Co-organizer of 2019 Computational Molecular Evolution Summer School, Hinxton, UK.

Anna Wienhard:
"Groups, spaces and geometries", conference in honor of Alessandra Iozzi (with C. Burrin, T. Hartnick and E. Kowalsky), FIM Zürich, Switzerland, 22-25 January 2019. Workshop *"Geometry, Topology, and Computation"*, Heidelberg University, Heidelberg, Germany, 12-14 June 2019. *"Connection for Women: Holomorphic Differentials in Mathematics and Physics"* (with. L. Fredrickson, L. Holland, Q. Li), MSRI, Berkeley, USA, 15–16 August 2019. Program on *"Holomorphic Differentials in Mathematics and Physics"* (together with J.Athreya, S. Bradlow, S. Gukov, A. Neitzke, A. Zorich), MSRI, Berkeley, USA, August to December 2019. Thematic semester on *"Computational Methods in Topology, Geometry and Analysis"*, Excellence Cluster STRUCTURES (together with P. Albers, A. Marciniak-Czochra), Heidelberg University, Germany, 2019. Introductory Workshop: *"Holomorphic Differentials in Mathematics and Physics"* (with. J. Athreya, A. Neitzke, A. Zorich), MSRI, Berkeley, USA, September 2019.

Other contributions

Antonio D'Isanto:
5 outreach articles published on the Italian web journal *"Tom's Hardware"*.
"A new science is born - Astronomy in the era of AI", outreach talk in the *"Astronomy on Tap"* series, *"O'Reilly's"*, Heidelberg, Germany, 12 November 2019.

Erica Hopkins:
"Man vs Machine: Can Crowdsourcing be Replaced by Machine Learning?", outreach talk in the *"Astronomy on Tap"* series, *"O'Reilly's"*, Heidelberg, Germany, 3 December 2019.

Ina Pöhner:
'Campus Report': "Mathe gegen Parasiten', Campus Radio Interview, broadcast on 14 August 2019..

Rebecca C. Wade:
Guest Editor (Outi Salo-Ahen (Abo Akademi University, Turku, Finland)) of Special Issue in the journal 'Molecules' on 'Molecular Modeling in Drug Design'.

9.5 Awards

Federico López:
Best Paper Award at the 4th ACL Workshop on Representation Learning for NLP, Florence, Italy, 2 August 2019 (with Benjamin Heinzerling, Michael Strube).

Mark-Christoph Müller:
Best Poster and Demonstration Award at the 23rd International Conference on Theory and Practice of Digital Libraries, Oslo, Norway, 9-12 October 2019 (together with Adam Bannister and Florian Reitz)

Goutam Mukherjee:
Poster Prize, 19th International Conference on Biological Inorganic Chemistry. Interlaken, Switzerland, 11-16 August 2019.

Mareike Pfeil:
HITS *"Isabel Rojas Travel Award"* 2019 (award for research stays of up to three months at another institution).

Alexandros Stamatakis:
Highly Cited Researcher in Cross Field research, Clarivate Analytics, 2019.

Michael Strube:
ACL Fellow, Association for Computational Linguistics.

Anna Wienhard:
2019 Clay Senior Scholar and Co-Organizer of "Holomorphic Differentials in Mathematics and Physics", MSRI, Berkeley, USA.
2019 Elected Member to Berlin-Brandenburg Academy of Science
Third Party Funds: 2019 –2025 Cluster of Excellence *"STRUCTURES - a unifying approach to emergent phenomena in the physical world, mathematics, and complex data"* (together with R. Klessen, M. Salmhofer as Co-Spokesperson), Heidelberg University (see Chapter 6).

10 Boards and Management

The HITS Scientific Advisory Board at HITS 2019. From left to right: Wolfgang Müller (HITS Scientific Director), Tony Hey, Frauke Gräter (HITS Deputy Scientific Director), Alex Szalay, Victoria Stodden, Thomas Lengauer, Adele Goldberg, Barbara Wohlmuth, Dieter Kranzlmüller, Gert-Martin Greuel, Gesa Schönberger (HITS Managing Director), Jeffrey Brock.

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:

- **Prof. Dr. Jeffrey Brock**, Zhao and Ji Professor of Mathematics at Yale University, USA
- **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. Tony Hey**, Chief Data Scientist, Science and Technology Facilities Council, UK
- **Prof. Dr. Dieter Kranzlmüller**, Ludwig Maximilians University, Munich, Director of the Leibniz Super Computing Center, Germany (Chair, SAB)
- **Prof. Dr. Thomas Lengauer**, Max-Planck-Institute for Computer Science, Saarbrücken, Germany
- **Prof. Dr. Alex Szalay**, Johns Hopkins University, USA
- **Prof. Dr. Victoria Stodden**, School of Information Sciences, University of Illinois at Urbana-Champaign, USA
- **Prof. Dr. Barbara Wohlmuth**, Chair of Numerical Mathematics at the Technical University of Munich (TUM), Germany.

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
(until September 2019)

Prof. Dr. Jörg Pross
(since October 2019)
Vice-President of Research and Structure

Photo: © Philip Benjamin

Karlsruhe Institute of Technology (KIT)
Dr. Hanns-Günther Mayer
Director of Shareholdings
("Leitung Beteiligungen")

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:
PD Dr. Wolfgang Müller
(2019 – 2020)

Deputy Scientific Director:
Prof. Dr. Frauke Gräter
(2019 – 2020)

HITS

The Heidelberg Institute for Theoretical Studies (HITS) was established in 2010 by the physicist and SAP co-founder Klaus Tschira (1940–2015) and the Klaus Tschira Foundation as a private, non-profit research institute. HITS conducts basic research in the natural sciences, mathematics, and computer science, with a focus on the processing, structuring, and analyzing large amounts of complex data and the development of computational methods and software. The research fields range from molecular biology to astrophysics. The shareholders of HITS are the HITS-Stiftung, Heidelberg University, and the Karlsruhe Institute of Technology (KIT). HITS also cooperates with other universities and research institutes and with industrial partners. The base funding of HITS is provided by the HITS Stiftung with funds received from the Klaus Tschira Foundation. The primary external funding agencies are the Federal Ministry of Education and Research (BMBF), the German Research Foundation (DFG), and the European Union.

HITS gGmbH
Schloss-Wolfsbrunnenweg 35
D-69118 Heidelberg

Editor
Dr. Peter Saueressig
Head of Communications

Contact
info@h-its.org
Phone: +49 6221-533 533
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 | © 2020 HITS gGmbH

Twitter: @HITStudies
Facebook: /HITStudies
Youtube: /TheHITSters
LinkedIn: company/the-heidelberg-institute-for-theoretical-studies
Instagram: the_hitsters