HITS

Heidelberg Institute for Theoretical Studies

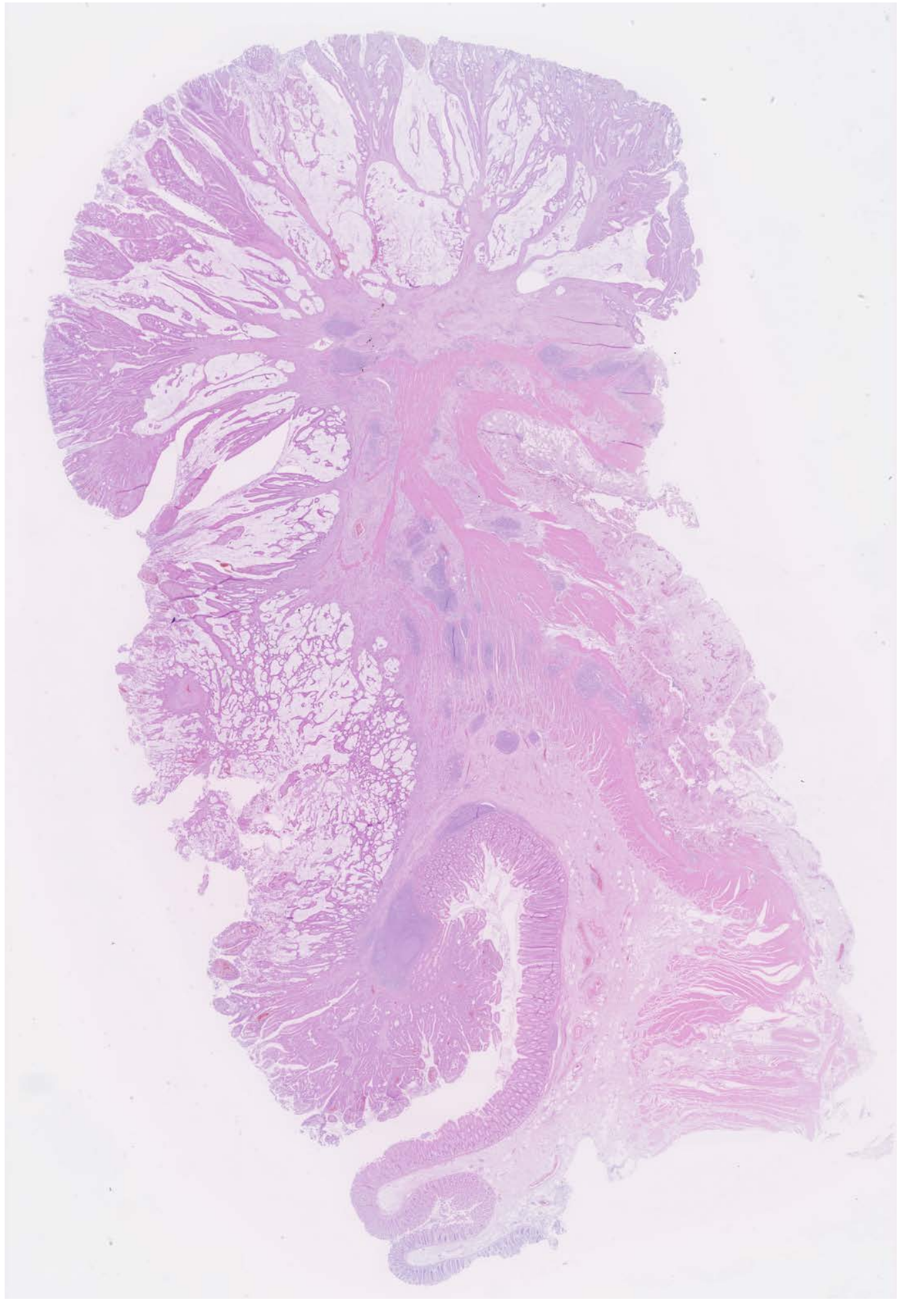
2021

Annual Report
Jahresbericht

Think beyond the limits!

Mathematical Oncology: Colon cancer is one of the most common cancers worldwide. Due to the microscopic size of the morphological units in the colon from which cancer originates, the processes during cancer initiation remain clinically obscure. HITS researchers from the Data Mining and Uncertainty Quantification group (DMQ) have used mathematical models and computer simulations to model biological processes behind the initial steps of colon cancer development to unravel the otherwise invisible cancer formation process. This picture shows a section through a colon cancer, where different areas represent different ramifications of cancer development and histopathology differentiation patterns. This visible diversity reflects the mutational diversity of these cancers, which can be described by mathematical models at different scales (cf chapter 2.5, pp. 32–37).

Mathematische Onkologie: Dickdarmkrebs ist weltweit eine der häufigsten Krebsarten. Aufgrund der mikroskopischen Größe der Strukturen im Dickdarm, aus denen Krebs entsteht, sind die genauen biologischen Prozesse während der Krebsentstehung klinisch unklar. HITS-Forschende der Data Mining and Uncertainty Quantification Gruppe (DMQ) modellierten mit Hilfe von mathematischen Modellen und Computersimulationen diese biologischen Prozesse der frühen Darmkrebsentstehung. Die Simulationen helfen dabei, den sonst unsichtbaren Vorgang der Krebsentstehung zu entschlüsseln. Das Bild zeigt einen Schnitt durch einen Dickdarmkrebs, bei dem verschiedene Bereiche unterschiedliche Verzweigungen der Krebsentwicklung und histopathologische Differenzierungsmuster darstellen. Diese sichtbare Vielfalt spiegelt die Mutationsvielfalt dieser Krebsarten wider, die durch mathematische Modelle auf verschiedenen Skalen beschrieben werden kann (Siehe Kapitel 2.5, pp. 32–37).



1 Think beyond the limits!

Dr. Gesa Schönberger
(Managing Director / Geschäftsführerin)

Prof. Dr. Frauke Gräter
(Scientific Director / Institutssprecherin)

Communicating, interacting, and working with scientists from distant disciplines is an often-cited and well-known hurdle of interdisciplinary research.

HITS removes structural hurdles by uniting an unusually broad range of research fields under one roof. On the HITS campus, astrophysicists and computational linguists run into one another in the same hallway, and the uncertainty of COVID predictions is outlined on the same board as are molecular orbitals. But what remains are the cultural hurdles – that is, the differences in language and teachings that exist between different scientific cultures. Overcoming these hurdles remains a daily goal for us at HITS.

Looking back on 2021, the Institute has made progress in the direction of interdisciplinarity: As part of the HITS Lab, we launched promising interdisciplinary projects that are now bearing fruit. That is how Federico López, Michael Strube (both Natural Language Processing; see Chapter 2.9), and Anna Wienhard (Groups and Geometry; see Chapter 2.6) – in collaboration with Beatrice Pozzetti (Heidelberg University) and Steve Trettel (Stanford University) – transferred geometric concepts from pure mathematics to machine learning. The result was that deep learning becomes more robust thanks to special symmetrical geometries. Given the rapid spread of deep learning in research and in our everyday lives, this is a conceptually groundbreaking result – and simultaneously extremely useful. This success was only possible because our mathematicians and computational linguists were able to overcome disciplinary hurdles.

The HITS Lab promotes innovative research projects between two or more groups with the aim of strengthening interdisciplinarity. As part of the HITS Lab, a total of 16 researchers worked successfully on four projects in 2021. For the coming years, we have set the goal of further promoting interdisciplinarity and filling the HITS Lab with more life – as well as with promising new projects!

The broad range of research at HITS is evident when reviewing our COVID-19 research from 2021: Among other things, the challenges of calculating family trees for current virus variants were made clear, molecular interactions of the COVID spike protein were simulated (see Chapter 2.8), and our COVID predictions were expanded to include a reliable estimate of the hospitalization incidence (see Chapter 2.4).

Another interdisciplinary project – a collaboration between Heidelberg University, the Karlsruhe Institute of Technology, and HITS – was launched in October 2021 as a strategic research initiative: Entitled SIMPLAIX, the project combines expertise from the field of multiscale simulation with that from the field of machine learning. The initiative thus enables new ways of exploring complex molecules and materials across scales and of improving the properties of these molecules and materials. The project is being funded by the Klaus Tschira Foundation for an initial period of three years. SIMPLAIX is coordinated by our group leaders Rebecca Wade (Molecular and Cellular Modeling) and Frauke Gräter (Molecular Biomechanics), who – together with group leader Ganna Gryn'ova (Computational Carbon Chemistry) – are part of this collaboration (see Chapters 2.8 and 6).

Personal interaction enables hurdles to be overcome. Unfortunately, digital conference systems cannot replace such interaction. In order to attract excellent young researchers and internationally renowned scientists alike to our institute, we launched two new programs in 2021: The HITS Independent Postdoc Program offers young ambitious scientists a 2-year research stay to work on an independent project. We expect to welcome the first person to HITS within the scope of this program in 2022. The Klaus Tschira Guest Professorship Program was also initiated in 2021. For this program, we have already selected

two individuals as visiting professors for 2022. We look forward to interacting personally with our future guests and to growing from their scientific insights.

2021 was a year in which we were pleased to see one highlight after another. Here is just a selection: In January 2021, the new Stellar Evolution Theory group (SET) was launched under the direction of Fabian Schneider (see Chapter 2.12), who also received the Ludwig Biermann Prize from the Astronomical Society during the year. HITS alumnus Volker Springel – a former group leader and now a Director at the Max Planck Institute for Astrophysics in Garching as well as a HITS Fellow since 2018 – was awarded the Gottfried Wilhelm Leibniz Prize by the German Research Foundation for his pioneering work in the field of numerical astrophysics. Alexandros Stamatakis (Computational Molecular Evolution; see Chapter 2.3) became a Highly Cited Researcher for the sixth time in a row, making him one of the 1,000 most cited scientists in his field. Anna Wienhard was awarded an ERC Advanced Grant for her project "PosLieRep – Positivity in Lie Groups and Representation Varieties." In addition, Ganna Gryn'ova's work paid off, and we are proud to report that she will receive an ERC Starting Grant beginning in 2022 for her project PATTERNCHEM. With Ganna's grant, six of our thirteen HITS group leaders will be either ERC grant recipients or ERC beneficiaries by the beginning of 2022.

It is important to us to make research understandable. While 2021 was again marked by the COVID-19 pandemic, we were nevertheless able to present our research not only to interested specialist groups, but also to the general public. Both the Computational Carbon Chemistry group and the Molecular and Cellular Modeling group took part in Explore Science 2021 in Mannheim with presentations and hands-on stations on the subject of chemistry. To strengthen the transfer of information and knowledge, we expanded our website in 2021 to include HITS software highlights and HITS publications. There, you will also find news about the HITS Lab and information on new programs for fostering scientific exchange.

We are pleased to look back on another strong year at HITS – a year that our founder, Klaus Tschira, would have truly enjoyed, we are sure. Indeed, breaking through scientific and interdisciplinary hurdles remains our main objective.

Mit Wissenschaftlerinnen und Wissenschaftlern aus weit entfernten Disziplinen zu kommunizieren, zu interagieren und zu arbeiten, stellt eine oft zitierte und wohlbekannte Hürde interdisziplinärer Forschung dar.

HITS schafft strukturelle Hürden ab, indem es eine ungewöhnlich große Breite an Forschungsrichtungen unter einem Dach vereint. Hier treffen wir auf dem gleichen Flur die Astrophysikerin und den Computerlinguisten, an der gleichen Tafel wird die Unsicherheit der Covid-Vorhersage neben molekularen Orbitalen skizziert. Was aber bleibt, sind die kulturellen Hürden – die Unterschiede in Sprache und Lehre der Wissenschaftskulturen. Diese zu überwinden, fordert uns täglich heraus.

HITS kann im Rückblick auf 2021 feststellen, in Richtung Interdisziplinarität vorangeschritten zu sein: Im Rahmen des HITS Lab haben wir vielversprechende, interdisziplinäre Projekte auf den Weg gebracht, die nun erste Früchte tragen. So haben Federico López, Michael Strube (beide Natural Language Processing, Kapitel 2.9) und Anna Wienhard (Groups and Geometry, Kapitel 2.6) - in Zusammenarbeit mit Beatrice Pozzetti (Universität Heidelberg) und Steve Trettel (Stanford University) - geometrische Konzepte aus der reinen Mathematik in das maschinelle Lernen übertragen. Das Resultat: Deep Learning wird durch spezielle symmetrische Geometrien robuster. Angesichts der rasanten Ausbreitung von Deep Learning in der Forschung und in unserem Alltag ist das ein konzeptionell bahnbrechendes und gleichzeitig extrem nützliches Ergebnis. Der Erfolg war nur möglich, weil unsere Mathematiker/-innen und Computerlinguist/-innen disziplinäre Hürden überwinden konnten.

Das HITS Lab fördert innovative Forschungsprojekte zwischen zwei oder mehr Gruppen, um Interdisziplinarität zu stärken und Hürden zu überwinden. Im Rahmen des HITS Lab haben 2021 insgesamt 16 Forschende an vier Projekten erfolgreich gearbeitet. Für die kommenden Jahre haben wir uns vorgenommen, Interdisziplinarität weiter voranzutreiben und das HITS Lab verstärkt mit Leben - und vielversprechenden Projekten - zu füllen.

Die ganze Breite der Forschung am HITS wird sehr deutlich, wenn wir unsere Covid-19-Forschung des vergangenen Jahres Revue passieren lassen: Hier wurden unter anderem die Herausforderungen bei der Berechnung von Stammbäumen aktueller Virusvarianten aufgezeigt, molekulare Interaktionen des Covid Spike-Proteins simuliert (siehe Kapitel 2.8) und die Covid-Vorhersage um eine belastbare Schätzung der Hospitalisierungsinzidenz erweitert (siehe Kapitel 2.4).

Ein weiteres interdisziplinäres Projekt, eine Zusammenarbeit zwischen der Universität Heidelberg, dem Karlsruher Institut für Technologie (KIT) und HITS, wurde im Oktober 2021 als strategische Forschungsinitiative ins Leben gerufen: SIMPLAIX vereint die Expertisen auf dem Gebiet der Multiskalensimulation mit solchen des maschinellen Lernens. Die Initiative ermöglicht so neue Wege, um komplexe Moleküle und Materialien skalenübergreifend zu erforschen und ihre Eigenschaften zu verbessern. Das Projekt wird von der Klaus Tschira Stiftung für vorerst drei Jahre gefördert. Unsere HITS-Gruppenleiterinnen Ganna Gryn'ova und Frauke Gräter sind Teil dieser Zusammenarbeit, Rebecca Wade ist zugleich Sprecherin (siehe auch Kapitel 2.8 und 6).

Persönlicher Austausch überwindet Hürden. Das können digitale Konferenzsysteme nicht ersetzen. Um exzellente junge, oder auch bereits international renommierte Wissenschaftlerinnen und Wissenschaftler für unser Institut zu gewinnen, haben wir 2021 zwei neue Programme ins Leben gerufen: Das „HITS-Independent-Postdoc-Programm" bietet jungen, ambitionierten Forschenden einen zweijährigen Aufenthalt für ein unabhängiges Forschungsvorhaben. Die erste Person in diesem Programm erwarten wir 2022 am HITS. Das „Klaus Tschira Gastprofessor/-innen-Programm" wurde ebenfalls 2021 gestartet. Hier haben wir bereits zwei Personen für 2022 ausgewählt. Wir freuen uns auf den persönlichen Austausch mit unseren zukünftigen Gästen und auf ihre wissenschaftlichen Impulse.

2021 war ein Jahr, in dem wir uns über ein Highlight nach dem anderen freuen konnten - hier nur eine Auswahl: Im Januar 2021 nahm die neue Stellar Evolution Theory-Gruppe unter der Leitung von Fabian Schneider (siehe Kapitel 2.12) ihre Arbeit auf, Fabian Schneider erhielt im Laufe des Jahres zudem den Ludwig-Biermann Preis der Astronomischen Gesellschaft. HITS-Alumnus Volker Springel, ehemaliger Gruppenleiter, jetzt Direktor am Max-Planck-Institut für Astrophysik in Garching und seit 2018 HITS Fellow, wurde für seine wegweisende Arbeit auf dem Gebiet der Numerischen Astrophysik mit dem Gottfried Wilhelm Leibniz-Preis der Deutschen Forschungsgemeinschaft ausgezeichnet. Alexandros Stamatakis (Computational Molecular Evolution, Kapitel 2.3) wurde zum sechsten Mal in Folge „Highly Cited Researcher" und zählt damit zu den 1.000 am meisten zitierten Forschenden seines Feldes. Anna Wienhard hat für ihr Projekt „PosLieRep – Positivity in Lie Groups and Representation Varieties" einen ERC Advanced Grant gewonnen. Darüber hinaus hat sich die Arbeit für Ganna Gryn'ova (Computational Carbon Chemistry, Kapitel 2.2) ausgezahlt, und wir freuen uns mit ihr über einen 2022 beginnenden ERC Starting Grant für ihr Projekt „PATTERNCHEM". Damit zählen zu Beginn des Jahres 2022 sechs unserer dreizehn HITS Gruppenleiter/-innen zu ERC-Grant-Gewinnern oder sind als Begünstigte an einem solchen beteiligt.

Es ist uns ein wichtiges Anliegen, Forschung verständlich zu machen. Obwohl 2021 erneut ein von der Covid-19-Pandemie bestimmtes Jahr war, haben wir unsere Forschung nicht nur in interessierten Fachkreisen, sondern auch in der Öffentlichkeit vorgestellt. So beteiligten sich die Gruppen Computational Carbon Chemistry (CCC) und Molecular and Cellular Modeling (MCM) mit Präsentationen und Hands-on-Stationen zum Thema „Chemie" bei der Explore Science 2021 in Mannheim. Um den Informations- und Wissenstransfer zu stärken, erweiterten wir 2021 unsere Webseite um die HITS Software Highlights und HITS Publikationen. Dort finden sich auch Aktuelles zum HITS Lab und Informationen zu den neuen Programmen für wissenschaftlichen Austausch.

Wir blicken auf ein starkes HITS-Jahr zurück, an dem unser Gründer und Stifter Klaus Tschira seine wahre Freude gehabt hätte, dessen sind wir uns sicher. Wissenschaftliche und interdisziplinäre Hürden zu durchbrechen, bleibt unser Hauptanliegen.

2 Research

2.1 Astroinformatics (AIN)

Group leader
Dr. Kai Polsterer

Staff members
Dr. Nikos Gianniotis
Dr. Antonio D'Isanto (until April 2021)
Dr. Jan Plier
Dr. Franciso Pozo Nuñez (since April 2021)

Scholarship holder
Erica Hopkins (until May 2021)

Student assistant
Fenja Kollasch

In recent 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 higher spatial, spectral, and temporal resolutions. In addition, new, untapped wavelength regimes along with other messengers – such as gravitational waves and astro-particles – are now granting more complete observational access to the Universe than ever before.

The Astroinformatics group deals with the challenges of analyzing and processing complex, heterogeneous, and large datasets. Our scientific focus in astronomy is on evolutionary processes and extreme physics in galaxies, such as those found around active super-massive black holes in the centers of galaxies. Driven by these scientific interests, we develop new methods and tools that we share with the community. From a computer-science perspective, we focus on time-series analyses, sparse-data problems, morphological classification, the proper evaluation and training of models, and the development of exploratory research environments. These methods and tools will prove critical to the analysis of data in large upcoming survey projects, such as SKA, Gaia, LSST, and Euclid.

Our ultimate goal is to enable scientists to analyze the ever-growing amount of information in a bias-free manner.

Inferring Physical Parameters of Supermassive Black Holes via Light Curves

Active Galactic Nuclei (AGN) are supermassive black holes at the center of galaxies. By accreting material, these AGN exhibit extreme luminosities and can thus be observed at great distances, thereby enabling the study of evolutionary processes in the Universe.

Photometric reverberation mapping is a technique used to indirectly determine the radial extent of the accretion disk (AD) by measuring time delays between light curves observed in different continuum bands. Quantifying constraints to the efficiency and accuracy of time-delay measurements is critical to recovering the AD size:luminosity relation, to estimating the corresponding black hole mass, and hence, to potentially using quasars as standard cosmological candles to determine distances.

In order to estimate black hole mass, we need to postulate a physical mechanism that connects the physical properties of the AGN to the observed light curves. Roughly speaking, we assume that the light curves that we observe are the result of AGN emissions that we have not yet observed and that are processed by the accretion disk of the AGN. Both the physics governing the behavior of the accretion disk and the associated transfer function are dictated by the Shakura–Sunyaev model. From a statistical point of view, we treat the unobserved AGN signal as a latent signal and model it using a Gaussian process (GP). Thus, our physically motivated statistical model states that the multiple AGN light curves that we observe are the result of the convolution of the transfer function and a signal that stems from the latent signal, which is modeled as a GP. The statistical nature of our model allows us to calculate a probability distribution – instead of point estimate – for the black hole mass (i.e., a posterior distribution given the observed data) that fully characterizes the uncertainty of our estimate.

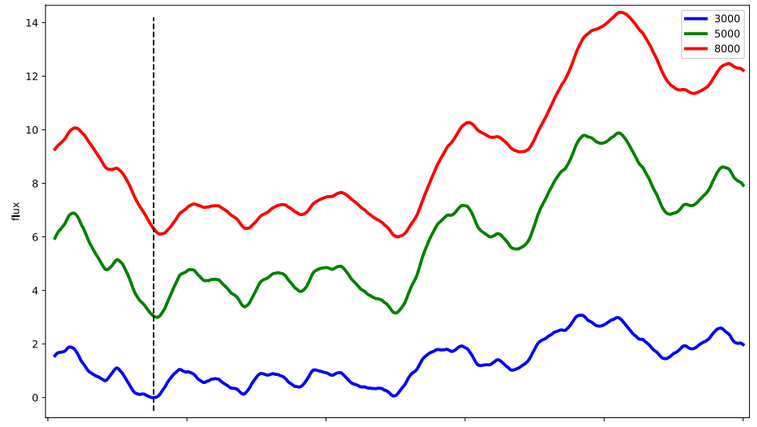


Figure 1: Light curves at 3,000 Ångström (Å; blue), 5,000 Å (green), and 8,000 Å (red) simulated with our physically inspired Gaussian-process model.). The plot reveals how strongly correlated the light curves are because they display very similar peaks and dips. However, the red light curve appears smoother than the blue light curve. We note, for instance, how the two peaks of the blue light curve at ~80 days are not as pronounced when we look for them in the red light curve at the same time point. We also note a delay effect: Notice how the dip in the blue light curve – marked by the horizontal dashed line – occurs a bit earlier than the corresponding dip in the red light curve. This indicates that the red light curve is a bit delayed with respect to the blue light curve.

Following the above-mentioned considerations, the resulting model has two interesting properties: (1) a multi-output property that allows the model to capture correlations between the multiple observed light curves and (2) a physically motivated kernel function that endows the outputs of the GP with characteristics such as delays and features that smooth out with increasing wavelengths exhibited by real light curves. Figure 1 displays light curves generated by our GP model.

The known elements in our estimation problem are the functional form of the physical transfer function and – of course – the observed AGN light curves at different bandwidths. The unknowns in the problem are the physical parameters of the transfer function – that is, mass and accretion rate – and the latent signal itself. Of these quantities, mass and accretion rate are of primary interest to us, while the inference of the latent signal is not important in this problem setting. Modeling the latent signal as a GP allows us to easily marginalize it from the model (i.e., to calculate a model average that results after considering all possible realizations of the latent signal), and it is thus of no further concern to us. Hence, we are left with only the two unknown parameters to estimate – that is, mass and accretion rate. We approximate the joint posterior distribution of mass and accretion on a two-dimensional grid defined by the joint support. As a preliminary example, we applied our model to AGN object Mrk279. Figure 2 (see next page) displays the joint posterior distribution of mass and accretion rate along with their marginals at the sides.

In addition, we are currently exploring the possibility of determining the accretion-disk size and the black hole mass using next-generation big-data surveys with a focus on the future public optical/near-infrared Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory, which

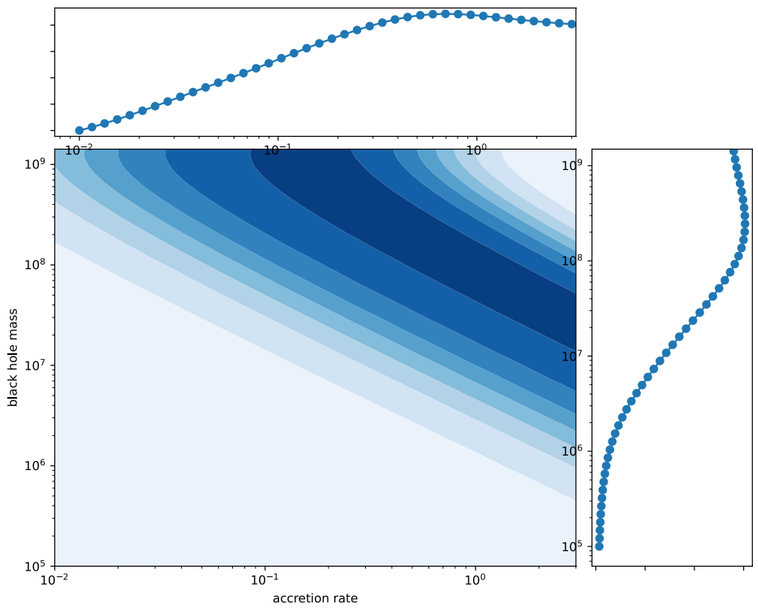


Figure 2 : Joint posterior distribution of mass and accretion rate along with marginals. Axes are in logarithmic scale.

will have observed about 10 million quasars in six broadband filters by the end of its 10-year operational lifetime. Therefore, we developed extensive simulations that take into account the characteristics of the LSST survey and the intrinsic properties of the quasars. The simulations will help us to characterize the light curves from which AD sizes and the black hole masses can be determined using various algorithms. Our model combines Shakura & Sunyaev's standard theory of an optically thick and geometrically thin AD with the dynamics of the Broad Line Region (BLR; Czerny et al. 2011), which is influenced by the radiation pressure of the AD (see Figure 3).

The UV/optical variations are driven by an X-ray-emitting corona located about 10 gravitational radii above the black hole. The variable X-ray emission is best described as a stochastic or chaotic process that is connected to the presence of turbulent magnetic fields, which produce fluctuations in the AD radiation. This stochastic input mechanism leads to optical light curves in quasars that are often well described by a random-walk process with a characteristic power spectral density. Figure 4 provides examples of our simulated random-walk light curves using the LSST bands.

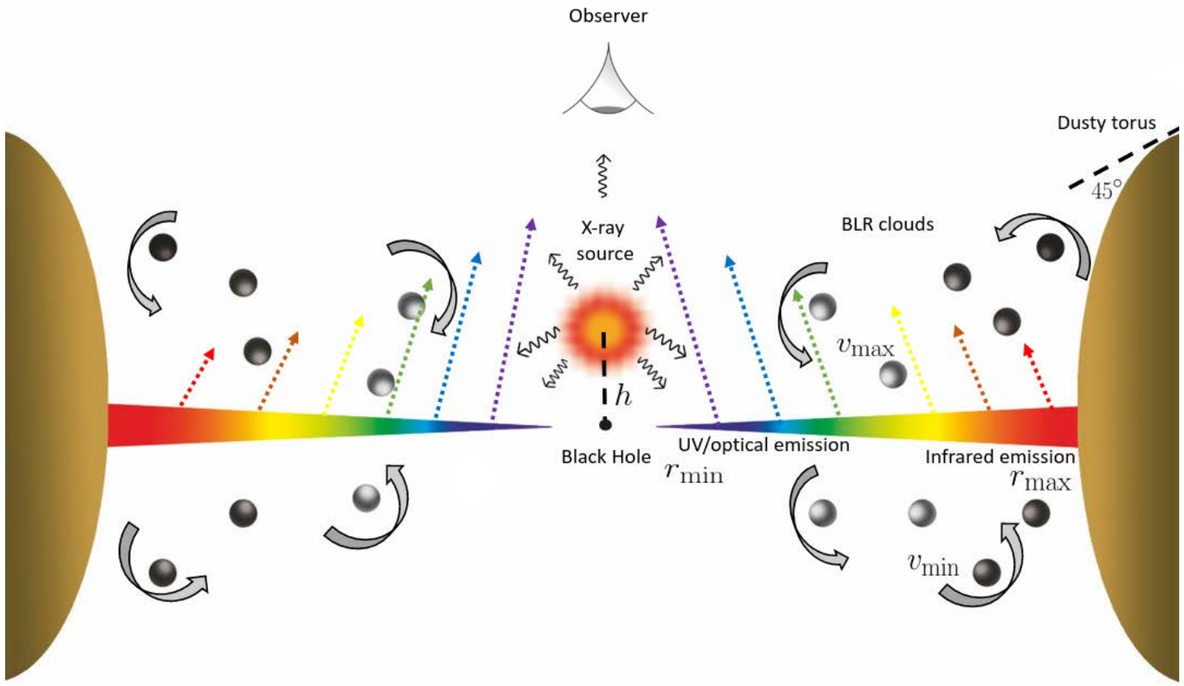


Figure 3: The physical AD+BLR model. X-ray photons from the corona are converted to UV/optical/infrared photons in the AD that reach the observer on Earth with a time delay. The further we move away from the black hole, the more the AD temperature drops below the sublimation temperature of the dust and the BLR material is lifted by local radiation pressure (Czerny et al. 2011). The material is exposed to irradiation from the AD, and the temperature rises above the sublimation temperature, resulting in vaporization of the dust and suppression of the radiation pressure. This effect causes the BLR gas to move up and down toward the surface of the AD.

The variability amplitude of a quasar depends on the quasar luminosity, the redshift, the wavelength, and the timespan of the observing program. We account for this effect in our simulations using the quasar variability–luminosity relation obtained with data from Pan-STARRS and SDSS. The emitted AD spectrum is calculated by integrating the Planck function of a black body from the radius of the innermost stable circular orbit to an outer radius and by assuming the AD's radial-temperature profile, which depends on the black hole mass and the mass accretion rate. The BLR contribution (i.e., the density and location of BLR clouds) determines whether we include the BLR-continuum emission in the total spectrum. Figure 5 provides an example of our simulated light curves and transfer functions for a particular quasar located at redshift z=0.5. As shown above, the determined AD time delays can be used to estimate black hole masses using the time-delay spectrum. This delay spectrum must be corrected for host-galaxy contamination and nuclear extinction. Our preliminary results indicate that we are able to estimate the masses of black holes with uncertainties of the order of 30% for LSST light curves, with sampling of ~2 days. Contamination from the BLR's diffuse continuum emission leads to time delays that are larger by a factor of ~1.5 in the u-band, as expected for the case of low-z quasars. We are currently investigating the impact of the BLR/AD interaction on the time delays.

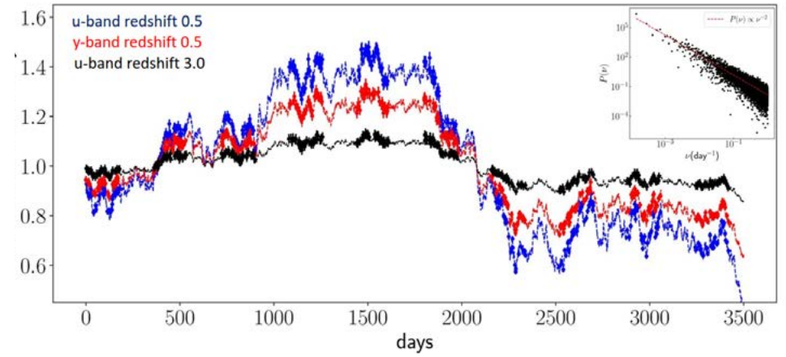


Figure 4: Examples of simulated light curves using a random walk process with a power spectrum of the form $P(\nu) \propto \nu^{-2}$ (top-right inset). The blue and red dotted lines correspond to the light curve of a quasar located at redshift z = 0.5 and observed through the u- and y-filters, respectively. For illustration, we show how the variability amplitude drastically changes if the quasar is located at z = 3.0 (black dotted line) in the u-band. The light curves were randomly sampled (dots), with an average sampling time of 5 days and a seasonal gap of 180 days. This particular example shows the light curves with a photometric signal-to-noise ratio of S/N = 60, which corresponds to measurement uncertainties at the ~2% level. The S/N calculations were performed using the total source counts, sky background contamination, instrumental noise, and gain.

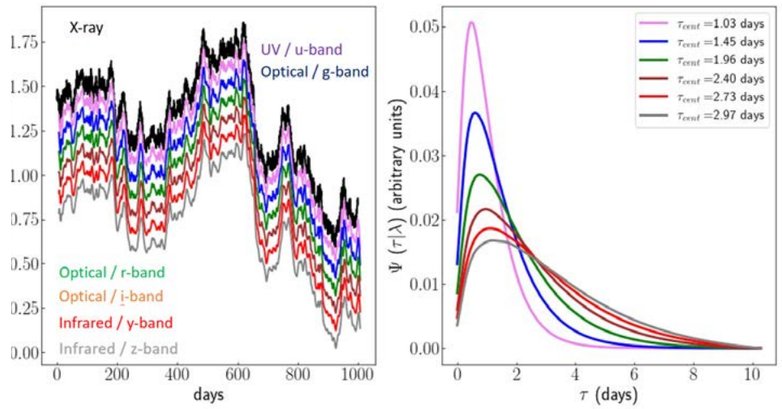


Figure 5: Convolved light curves (left) and AD transfer functions for LSST filters $ugriyz$ (magenta, blue, green, brown, red, grey). The centroid of the response function is marked with vertical dotted lines and is given in the observer frame. The convolved light curves include the host galaxy and nuclear reddening. The X-ray-simulated light curve is shown in black.

Exploring Self-Organizing Kohonen Maps

Astronomers need to work with data that come in large volumes and in a variety of modalities. A few years ago, we developed the Parallelized Rotation and Flipping Invariant Kohonen Maps (PINK) framework at HITS. Kohonen maps are also known as self-organizing maps (SOMs). PINK is an unsupervised machine learning method based on competitive learning that projects high-dimensional data onto a low-dimensional space. The framework has thus far been applied in galaxy-morphology problems with considerable success and is used in projects such as Lofar and SKA.

In 2021, we introduced an extension called UltraPINK, which is a web application for exploring SOMs created with PINK. This web application provides a visual frontend that visually assists the user in creating, inspecting, and managing SOMs (see Figure 6 for an example). This application fits directly in line with Licklider symbiosis between computers and humans and enables large amounts of data to be efficiently analyzed by using both computers to pre-sort and arrange objects based on morphological similarities and humans to provide a structural analysis.

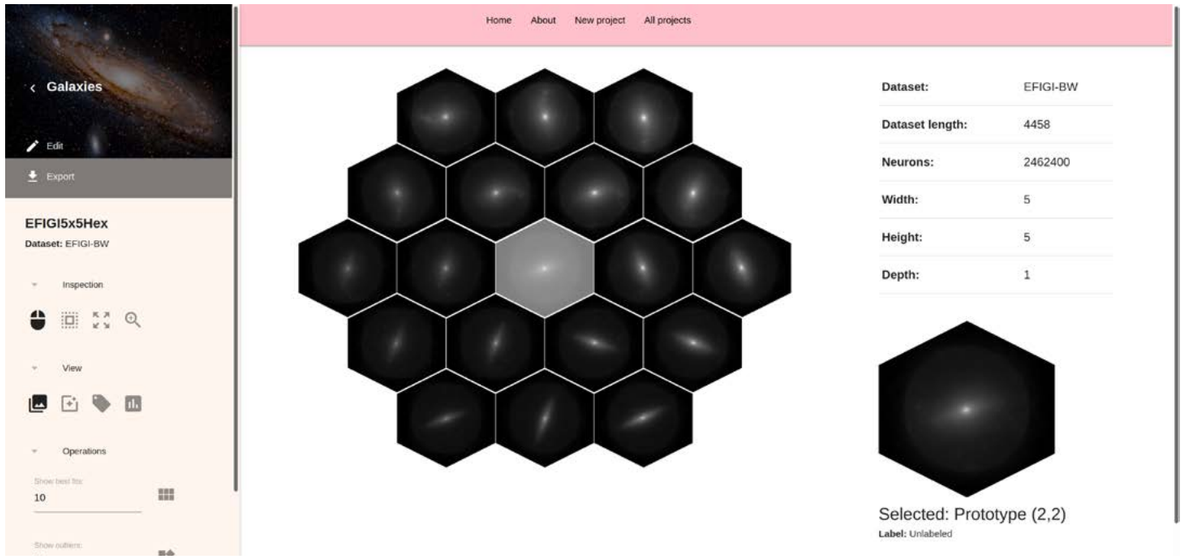


Figure 6: Inspecting a hexagonal self-organizing map trained on galaxy images using UltraPINK.

UltraPINK is a starting point and the first prototype of a more general framework of data analysis and exploration. The way UltraPINK allows its users to interact with data follows a generic implementation pattern. More specifically, the definition of interaction metaphors is independent of the data type, which allows a variety of modalities of astronomical data to be handled. In future works, we aim to advance this design to accommodate additional unsupervised machine learning algorithms other than PINK in order to offer the user a wider choice in how to explore the data at hand.

Computational Cardiology

Given our interest in galaxy morphology and in the analysis of massive datasets, we are also involved in a medical project that deals with morphological data at microscopic scales. The goal of the project is to classify cardiomyocytes according to the substance with which they have been previously treated (substances simulate medical conditions).
In our role in this project, we focus on extracting features that describe the morphological structure of each captured cell. Subsequently, these extracted features are used to build a robust and stable diagnostic tool for certain heart-related medical conditions.

In order to extract features for each cell, it is first necessary to segment the individual cells present in the image. However, this is a very challenging segmentation problem as the cells overlap considerably and assume highly variable forms. Current state-of-the-art segmentation approaches struggle with cardiomyocytes. In future plans, we would like to segment such overlapping cells with a Mask R-CNN network that is itself trained on human-provided segmentations. In order to collect such segmentations, we turn to Zooniverse – a crowd-sourcing platform that allows workflows for data-annotation tasks to be defined. We created a workflow (see Figure 7) for annotating cells in the captured images. Currently, the workflow is visible only to our expert medical partners. In the near future, we plan to make this workflow publicly visible so that more annotations can be sourced by volunteers.
In our present work, we resort to a more practical approach that avoids the difficult segmentation step. The images with which we work stem from fluorescence microscopy, which makes it easy for an algorithm to identify the nucleus of each cell in the image and to extract an image around this nucleus. In the current stage of the project, we simulate the diagnostic process by treating cells with eight distinct substances. In total, we have nine classes consisting of eight substances and untreated cells. To classify these items, we use a batch of cell patches instead of the whole image or segmented cells.

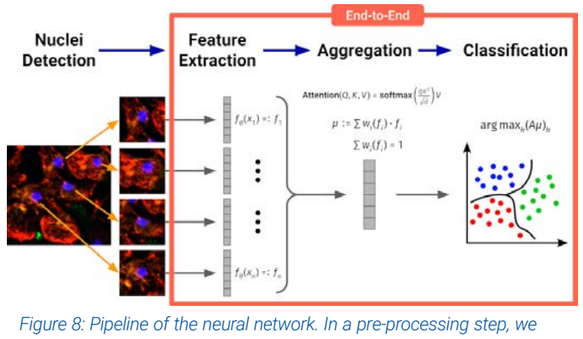


Figure 8: Pipeline of the neural network. In a pre-processing step, we extract all cell patches with respect to the centers of the cell nuclei. Our neural-network approach consists of three steps: First, for each input cell patch, features are extracted with the same parameters. In the next step, the extracted features are averaged using an attention mechanism. The final step consists of a projection on the logits of the possible classes.

For the classification task, we trained a neural network (see Figure 8) to match a batch of similarly treated cell patches to the used treatment. Then, we evaluated the neural net on a test dataset and achieved high accuracy. Consequently, our experiments show that we are able to predict the treatment of the cells by looking at only a few samples.

Future experiments will focus on interpretating the features extracted by the neural network. We also wish to make use of more accurate segmentations obtained with the Mask R-CNN neural network, which is itself trained on human-provided segmentations sourced from Zooniverse. We speculate that more easily interpretable features should be capable of being extracted when the available cell segmentations are more accurate.

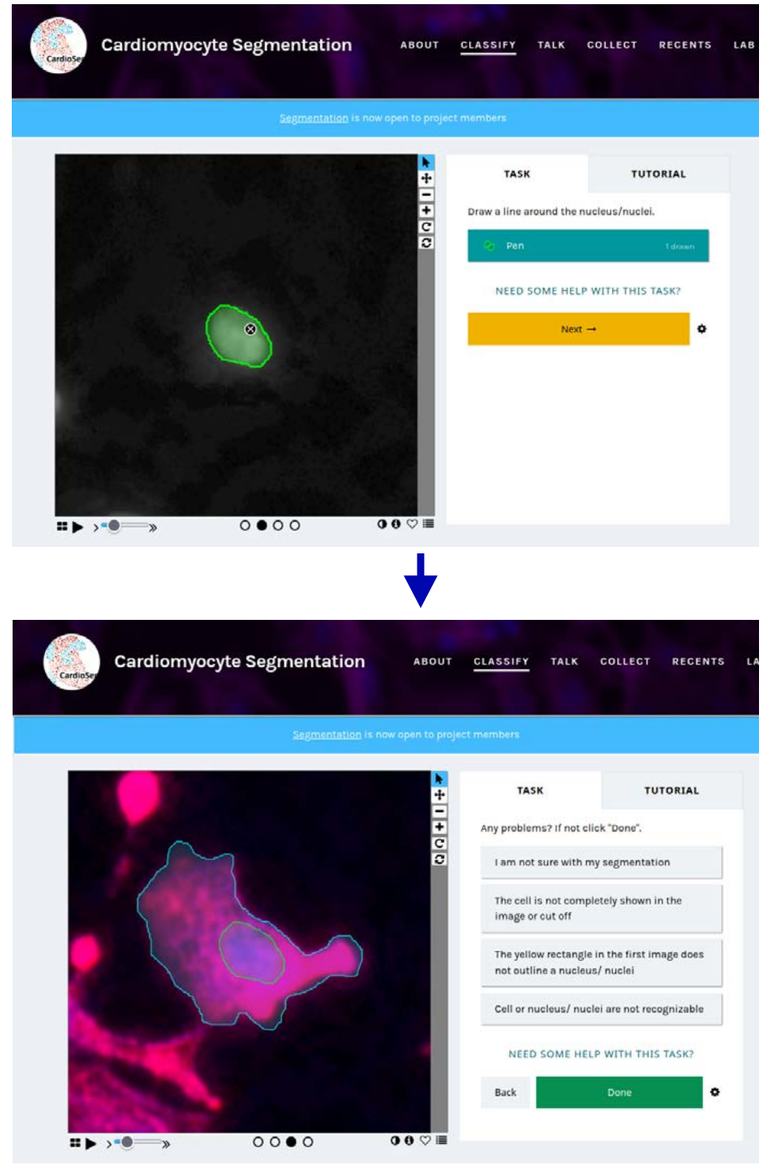
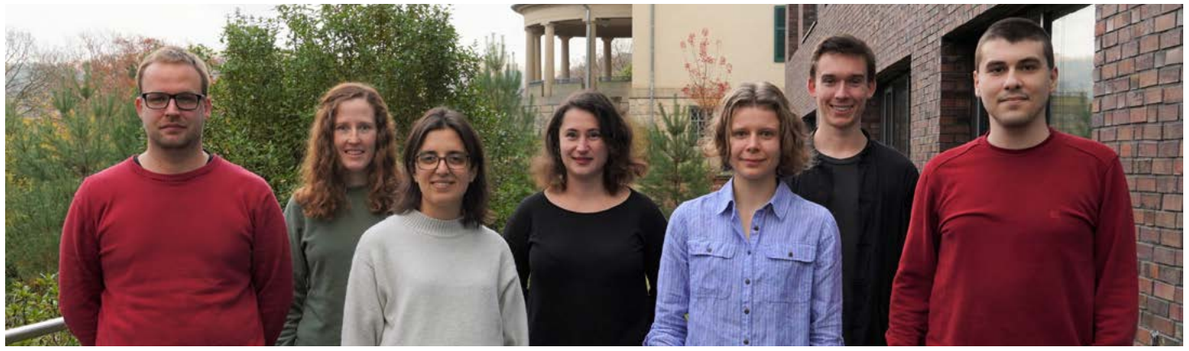


Figure 7: Workflow in Zooniverse, for which volunteers manually segment the cell nucleus and the associated cell body.

In den letzten Jahrzehnten hat der Einsatz von Computern die Astronomie stark beeinflusst. Der technologische Fortschritt ermöglichte den Bau neuer Detektoren und innovativer Instrumente sowie neuartiger Teleskope. Damit können Astronomen nun mehr Objekte als je zuvor mit bisher unerreichtem Detailreichtum, sowohl räumlich, spektral als auch zeitlich aufgelöst beobachten. Hinzu kommen neu Beobachtungsmöglichkeiten durch z.B. Astroteilchen sowie Gravitationswellen, die neben bisher nicht beobachtbaren Wellenlängenbereichen ein vollständigeres Bild des Universums bieten.
Die **Astroinformatik Gruppe (AIN)** beschäftigt sich mit den Herausforderungen die durch die Analyse und Verarbeitung dieser komplexen, heterogenen und großen Daten entstehen. In der Astronomie beschäftigen uns die Fragestellungen im Bereich der Galaxienentwicklung sowie die extremen physikalischen Vorgänge, wie man sie z.B. in der Umgebung von aktiven supermassereichen schwarzen Löchern in den Zentren von Galaxien findet. Auf diesen Fragestellungen basierend, entwickeln wir neue Methoden und Werkzeuge, die wir frei zur Verfügung stellen. In der Informatik liegt unser Interesse hierbei auf der Zeitreihenanalyse, dem Umgang mit spärlichen Daten, der morphologischen Klassifikation, der richtigen Auswertung und dem richtigen Training von Modellen sowie explorativen Forschungsumgebungen. Diese Werkzeuge und Methoden sind eminent wichtig für aktuelle und sich gerade in der Vorbereitung befindenden Projekten, wie SKA, Gaia, LSST und Euclid. Unser Ziel ist es, einen möglichst unvoreingenommenen Zugang zu dieser enormen Menge an Information zu gewährleisten.

2 Research

2.2 Computational Carbon Chemistry (CCC)



Group leader
Dr. Ganna Gryn'ova

Staff members
Dr. Christopher Ehlert
Stiv Lenga (since October 2021)
Anna Piras

Scholarship holder
Oğuzhan Kucur (until January 2021)

Visiting scientist
Dr. Michelle Ernst (SNSF Scholarship, since August 2020)

Project student
Juliette Schleicher (Heidelberg University; February–September 2021)

Master's students
Felix Trautner (Heidelberg University; November–December 2021)
Sebastian Pauly (Heidelberg University; May–June 2021)

computational carbon chemistry

Modern functional materials combine structural complexity with targeted performance and are utilized across many areas of industry and research ranging from nanoelectronics to large-scale production. Theoretical studies of these materials bring mechanistic underpinnings to light, facilitate the design and pre-screening of candidate architectures, and ultimately predict the physical and chemical properties of new systems.

The Computational Carbon Chemistry (CCC) group uses theoretical and computational chemistry to explore and exploit diverse functional organic and hybrid materials. In its 3rd year at HITS, the group proceeded to apply the previously established and benchmarked computational protocols to the design of novel materials, including improved graphene-based materials for environmental remediation and electrocatalysis, the design of which was based on the group's investigations into the fundamental mechanisms that underlie the performance of these materials. A new computational approach for predicting the structure of the host–guest complexes between organic frameworks and their molecular guests was also developed and tested.

The CCC group additionally began several fruitful scientific collaborations. Within the SFB1249 "N-Heteropolycycles as Functional Materials," the group explored structure–reactivity patterns in open-shell N-heteropolycycles using a combination of high-level quantum chemistry and machine learning. In close cooperation with the Greb group at Heidelberg University, key principles of the reactivity of main-group-element compounds were uncovered. These new research avenues deepen and expand the CCC group's research field by including new types of functional materials on the one hand and novel physical and chemical concepts on the other hand.

Chemisorption of dioxygen on calix[4]pyrrolato aluminate

Christopher Ehlert and Ganna Gryn'ova

Dioxygen (O_2) is an essential part of the cellular respiration cycle and is therefore vital to all aerobic organisms. Dioxygen makes up approximately 20% of the earth's atmosphere, which serves as a primary source of this essential chemical for humans. The fixation of dioxygen by other chemical compounds is a crucial step in metabolic cycles and can occur via different mechanisms. For example, the active site of hemoglobin – a protein located in human red blood cells – is an iron-porphyrin complex (shown in Figure 9A). Within this complex, dioxygen binds in the form of a superoxide (O_2^-) after an electron has been transferred from the iron center to the dioxygen (thereby oxidizing iron from +II to +III).

Within this project, which is performed in close cooperation with the Greb group (Heidelberg University), we perform in silico investigations into the mechanism of dioxygen fixation, which is assisted by calix[4]pyrrolato aluminate (AlCx) (shown in Figure 9B). While dioxygen may appear to be a rather simple molecule, its electronic structure can be challenging due to the triplet ground state and the two excited singlet states, which have strong multireference characters and might be involved in the fixation process. Furthermore, the final structure (dioxygen fixated by AlCx) is a singlet state, and crossing points between the potential energy surfaces of different spins thus occur during the attachment. Such a complex chemistry must be treated with electronic structure methods that go well beyond the standard "black-box" approaches. Using perturbation theory based on multireference self-consistent field, we were able to show that fixation occurs via the following steps:

I. An initial electron transfer from the Cx ligand to the dioxygen reduces the latter to the superoxide anion O_2^-, which subsequently attaches to the central aluminum atom. The total system remains in a triplet state, with a geometry shown in Figure 9C as $[^{Me}1\text{-}O_2]^-$. In this geometry, the triplet and singlet energies are degenerate due to a largely delocalized spin density in the Cx part of the complex.
II. As the oxygen moves toward a neighboring carbon atom, the degeneracy is lifted, and the singlet state constitutes the ground state. Therefore, an intersystem crossing occurs. The energy curves of the singlet and triplet energies are shown in the lower inset of Figure 9C.
III. On the singlet potential energy surface, oxygen relaxes via a small transition state to the final structure ($[^{Me}1^*\text{-}OO]^-$ in Figure 9C).

Overall, our computational results are well in line with experimental observations and provide a deeper understanding of the mechanisms that underlie the fixation of O_2 by AlCx.

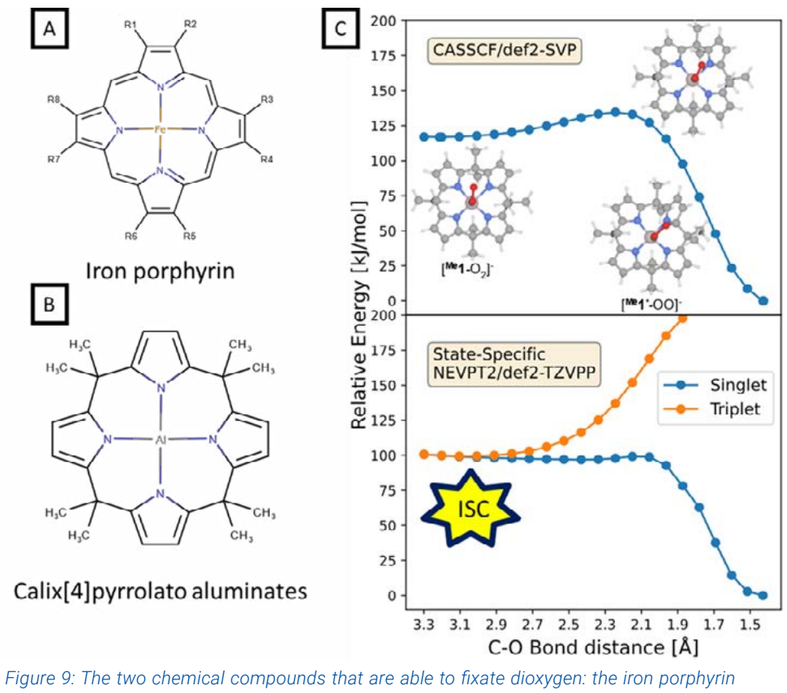


Figure 9: The two chemical compounds that are able to fixate dioxygen: the iron porphyrin complex (which is a part of the hemoglobin protein in human blood cells) (A) and the calix[4] pyrrolato aluminate (AlCx) complex, both of which are investigated in this project (B). (C) The results of multireference self-consistent field computations. In the upper inset, the energies of constraint optimizations are shown for the singlet state. The molecular structures correspond to initial, transition, and final states along the scanned degree of freedom. The lower inset shows the singlet and triplet energies in the optimized geometries. The degeneracy of both energies indicates the point at which an intersystem crossing is most probable and which explains the dioxygen fixation potential of AlCx.

Structure prediction of molecular targets encapsulated in organic frameworks

Stiv Llenga and Ganna Gryn'ova

Open-shell configurations of N-heteropolycyclic compounds are ubiquitous across the broad range of their chemical transformations, be it in organic synthesis or in molecular electronics applications. Charge transport in the latter context involves charge-carrier injection from the electrodes and a subsequent transfer of the holes and/or electrons formed in this process from one molecular fragment to the next (along a conductive polymer chain) or from one molecular core to the adjacent core (in 2- and 3-dimensional bulk organic semiconductors). The intermediary

species in these processes are therefore inherently deemed to bear unpaired electrons and are ion radicals by nature. Of no lesser interest are N-heteropolycycles, in which mono- and polyradical ground-state configurations are of sufficient thermodynamic stability and kinetic persistence to be isolable. Both low-spin (e.g., open-shell singlet diradicals) and high-spin (e.g., triplet diradicals) organic species are of potentially high utility due to their unique conducting-, optical, and magnetic properties provided that they are sufficiently air- and thermostable. In practice, open-shell N-heteropolycycles are used to construct novel organic light-emitting diodes as well as silicon- and germanium-free organic semiconductors, and they are also used as catalysts for various chemical transformations.

However, efficiently exploiting these properties necessitates an in-depth understanding of many diverse fundamental physico-chemical features of the N-heteropolycyclic radicals. In this project, we aim to explore the vast chemical space of these diverse open-shell compounds and to identify the structural and electronic factors that are pertinent to their stability and activity in the context of organic electronics applications.

Using automated molecular structure generators, more than 1,200 representatives of ortho-condensed rings were constructed by changing the number and position of nitrogen atoms within a prototypical skeleton (Figure 10A). Diverse quantum-chemical techniques were used to evaluate the fundamental physico-chemical properties of these systems, including frontier molecular orbital energy levels, ionization potentials, electron affinities (shown in Figure 10B), singlet–triplet gaps, electron paramagnetic resonance features, and radical stability properties (shown in Figure 10C). A range of dimensionality reduction techniques and statistical methods – including t-distributed stochastic neighbor embedding (t-SNE; shown in Figure 10D), uniform manifold approximation and projection (UMAP), and principal component analysis (PCA) – were used to visualize and analyze these data. In order to uncover the "hidden" relationships between molecular structures and properties, supervised and unsupervised machine learning techniques were employed. Our early results identified trends in the structure–property relationships of N-heteropolycyclic radicals that can be used to guide the targeted design of new compounds for practical applications in catalysis and organic electronics.

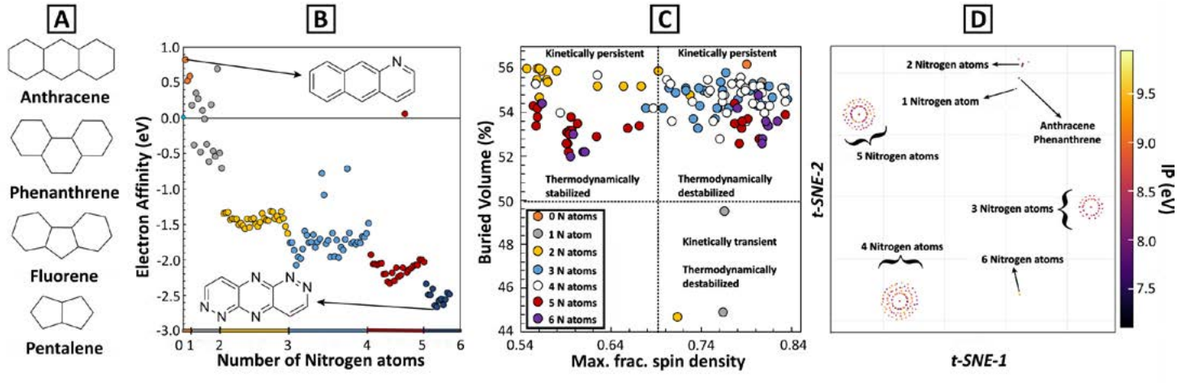


Figure 10: (A) Ortho-condensed ring skeletons. (B) Computed electron affinities of the studied N-heteropolycycles, arranged according to the number of the nitrogen atoms in the molecule. (C) Classification of the studied N-heteropolycycles according to the radical stability score. (D) t-SNE plot: The dimensionality reduction algorithm is performed on a database with 1,208 SMILES, each of which is linked to 1,700 molecular descriptors. The plot is color-coded by setting the IP values as a "target."

Tools for analyzing and visualizing non-covalent interactions in metal–organic frameworks

Michelle Ernst and Ganna Gryn'ova

Organic frameworks are porous crystalline materials that (a) consist of regularly connected nodes and linkers, (b) have high internal surface areas and low densities, and (c) can host small guest molecules. Depending on their composition, two principal types of frameworks exist: metal–organic frameworks (MOFs) – which were pioneered by Omar Yaghi in the late 1990s – and the more recently developed covalent-organic frameworks (COFs). Due to their highly tunable composition, their adaptable structures, and their ability both to absorb and store small guest molecules in their pores and to release these molecules depending on the chemical environment, MOFs and COFs are increasingly often utilized for gas storage and separation, drug delivery, and catalysis. Many of these applications involve environmental issues, and these materials – as well as the research on them – therefore have consequences that reach far beyond synthetic chemistry. The overall goal of this research avenue in the CCC is to shed light on the interactions

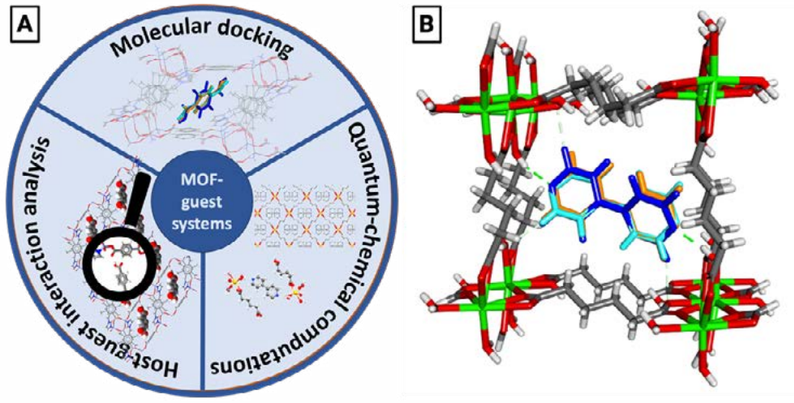


Figure 11: (A) New theoretical approach for investigating host–guest interactions in MOFs. (B) Comparison of experimental (orange), docked (light blue), and geometry-optimized (dark blue) positions of the bipyridine guest in the GW-MOF-7 pore.

between the frameworks and the guests (the host–guest interactions) using in silico methods.

While accurate structural data are key to unveiling the host–guest interactions in the COF and MOF complexes with their molecular targets, experimental structure determination remains rather challenging. Indeed, only a limited number of high-resolution structures have been reported in the literature. Moreover, these high-resolution structures generally do not simultaneously contain local atomistic-scale information on the adsorption site and the strength of the host–guest interaction. In order to address this problem in silico, molecular docking – a technique with a long tradition in the field of protein–ligand interactions – was adapted to identify adsorption sites within the MOF cavities (Figure 11A). Five MOF–guest systems for which the structures have been determined by single crystal X-ray diffraction as well as a newly synthesized MOF that contains two different halogenated organic guests were used to test a multiscale approach that combines molecular docking for structure prediction on the one hand with density functional theory (DFT) for geometry refinement and interaction energy computations on the other hand. Systems in which the guest forms only a few interactions with the framework or in which the pore size is too small for a given guest prove challenging for this simulation technique. However, in systems with well-defined host–guest interactions, molecular docking with subsequent refinement at the DFT level succeeds, as demonstrated by the good agreement between the theoretically predicted and experimentally measured structures (Figure 11B). This new approach is promising for making accurate predictions of the topology, strength, and nature of host–guest interactions between organic frameworks and molecular targets, which is key to their many practical applications.

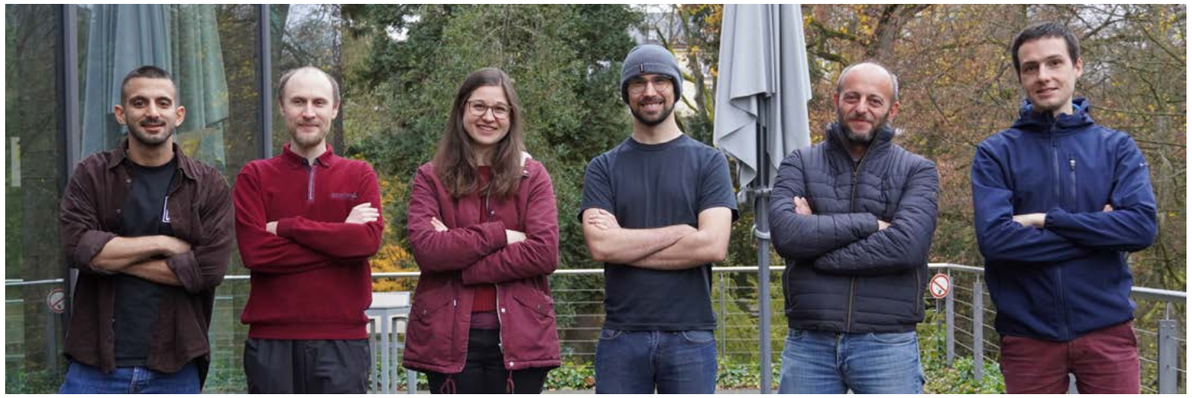
Moderne Funktionsmaterialien kombinieren strukturelle Komplexität mit zielgerichteter Performance und werden in verschiedenen Bereichen von Industrie und Forschung eingesetzt, von der Nanoelektronik bis hin zur Massenfertigung. Theoretische Studien dieser Materialien fördern mechanistische Grundlagen zutage, erleichtern das Design und Vorsortieren von Kandidaten und ermöglichen letztlich Vorhersagen zu physikalischen und chemischen Eigenschaften neu geschaffener Systeme.

Die Forschungsgruppe **Computational Carbon Chemistry (CCC)** nutzt theoretische und computergestützte Chemie, um verschiedene funktionale organische und Hybrid-Materialien zu untersuchen und auszuwerten. Im dritten Jahr ihres Bestehens am HITS wendet die Gruppe die zuvor etablierten und mit Benchmarks versehenen Berechnungsprotokolle auf die Entwicklung neuer Materialien an. Dazu gehören verbesserte Materialien auf Graphenbasis für die Umweltsanierung und Elektrokatalyse. Das Design dieser Materialien beruht auf der Erforschung der grundlegenden Mechanismen, die ihrer Perfomance zugrunde liegen. Außerdem wurde ein neuer rechnerischer Ansatz zur Vorhersage der Struktur von „Host-Guest"-Komplexen zwischen organischen Frameworks und ihren molekularen „Gästen" entwickelt und getestet.

Die CCC-Gruppe hat auch mehrere fruchtbare wissenschaftliche Kooperationen begonnen. Im Rahmen des SFB1249 „N-Heteropolyzyklen als Funktionsmaterialien" untersucht die Gruppe die Struktur-Reaktivitäts-Muster in offenschaligen N-Heteropolyzyklen, indem sie eine Kombination aus hochentwickelter Quantenchemie und maschinellem Lernen verwendet. In enger Zusammenarbeit mit der Greb-Gruppe an der Universität Heidelberg wurden Schlüsselprinzipien der Reaktivität von Verbindungen der Hauptgruppenelemente aufgedeckt. Diese neuen Forschungswege vertiefen und erweitern das Forschungsfeld der CCC-Gruppe in Richtung neuer Arten von Funktionsmaterialien einerseits und neuartiger physikalischer und chemischer Konzepte andererseits.

2 Research

2.3 Computational Molecular Evolution (CME)



Group leader
Prof. Dr. Alexandros Stamatakis

Staff members
Dr. Alexey Kozlov (staff scientist)
Benjamin Bettisworth
Dr. Benoit Morel
Lukas Hübner
Pierre Barbera (until March 2021)
Sarah Lutteropp (until August 2021)
Dimitri Höhler

Scholarship holder
Anastasis Togkousidis (since October 2021)

Students
Julia Haag
Xinyi Zhang (since June 2021)
Christoph Stelz (since September 2021)

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 analysis,
- supercomputing,
- quantifying biodiversity,
- next-generation sequence-data analysis, and
- scientific software quality & verification.

Secondary research interests include
- emerging parallel architectures,
- discrete algorithms on trees,
- ancient DNA analysis, and
- 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 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 happened in the lab in 2021?

In the winter of 2020/2021, Alexis, Benoit, Alexey, and Pierre taught the Introduction to Bioinformatics for Computer Scientists online 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/winter20_21.pdf).

During the summer semester of 2021, we again taught our main seminar, Hot Topics in Bioinformatics, as well as our master's-level Hands-on Bioinformatics programming practical. Our teaching activities continued to be heavily affected by the pandemic. Indeed, the seminar and internship practical in the summer were carried out entirely online. In addition, all oral exams for the class during the winter of 20/21 were also conducted online. Julia Haag successfully defended her master's thesis at the Department of Computer Science at KIT, the supervision of which was also conducted online to a very large extent. We are pleased to welcome Julia to the lab as a PhD student in 2022.

We are additionally delighted that Anastasis Togkousidis and former master's student Dimitri Höhler joined the lab as new PhD students in 2021. Another highlight was that Pierre and Benoit both successfully defended their PhD theses on 25 and 27 October, respectively.

Our recurring highlight – the summer school on Computational Molecular Evolution on Crete – was originally scheduled to take place in May 2020. It was then postponed to May and subsequently to October 2021 and ultimately unfortunately had to be cancelled due to the pandemic. The next iteration of the summer school will hopefully take place in Hinxton, UK, in the summer of 2022. Overall, our crisis management worked well as the decision to postpone and later cancel the event had been made sufficiently early and did not lead to almost any cancellation costs.
Alexis was listed on the Clarivate Analytics list of highly cited researchers for the sixth year in a row as well as for the fourth consecutive year in the new "cross-field" category, which comprises researchers with a focus on interdisciplinary research (see Chapter 9.5).

In April 2021, Alexis was officially appointed an Affiliated Scientist at the Paleogenomics and Evolutionary Genetics (PEG) lab within the Institute of Molecular Biology and Biotechnology at the Foundation for Research and Technology Hellas. The lab mainly focuses on ancient DNA research and is co-directed by Alexis, former CME postdoc Pavlos Pavlidis, and Nikos Poulakakis, the director of the Natural History Museum of Crete. The lab also includes senior postdoc and former CME visitor Nikos Psonis. The affiliation with the PEG will broaden the interdisciplinary breadth of the CME and foster scientific exchange between Germany and Greece when CME predoc Ben Bettisworth visits the PEG in early 2022 for a first collaborative project.

In sum, 2021 was again dominated by the pandemic. All lab members had already grown accustomed to working from home and online and had already become acquainted with online supervision. Nonetheless, everyone enjoyed meeting in person when possible and having dinner to celebrate the two successful PhD defenses.

Introduction

The term "computational molecular evolution" refers to computer-based methods of reconstructing evolutionary trees from DNA or – for example – from protein- or morphological data. The term also refers to the design of programs that estimate statistical properties of populations – that is, programs that disentangle evolutionary events within a single species.
The very first evolutionary trees were inferred manually by comparing the morphological characteristics (traits) of the species under study. Today, in the age of the molecular data avalanche, manually reconstructing trees is no longer feasible. Evolutionary biologists thus have to rely on computers and algorithms for phylogenetic and population-genetic analyses.
Since the introduction of so-called short-read sequencing machines (i.e., 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 now faces novel challenges.
One key problem that needs to be addressed is the fact that the volume 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 costs of sequencing a genome are decreasing at a faster rate than are the costs of computation, although the curve seems to have been flattening out in the last 3–4 years (see Figure 12, next page, and https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-

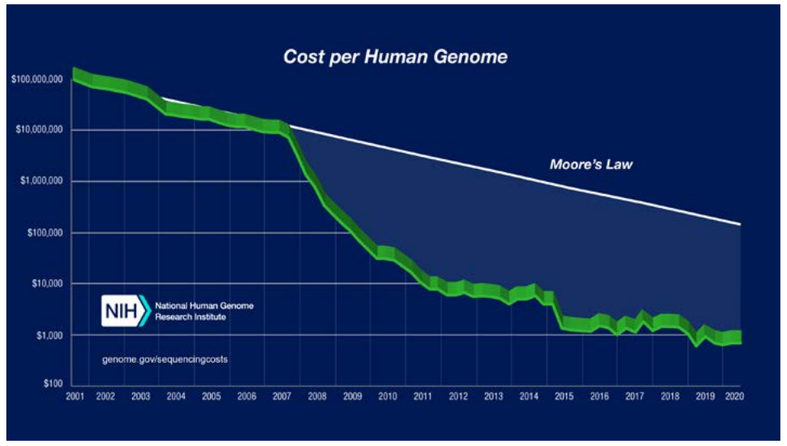


Figure 12: Cost of sequencing a human genome over time in comparison with the cost of computing according to Moore's law (source: National Human Genome Research Institute).

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

To address this scalability challenge, we have recently begun to investigate mechanisms for improving the fault tolerance (with respect to network- and processor failures) of large parallel scientific software tools using the example of RAxML-NG. At present, we are investigating methods to redundantly store and efficiently redistribute data from a parallel program following a core failure. Another novel line of research in this area is our new focus on making such large computational codes more energy efficient. Again, we conduct research in this domain using the example of RAxML-NG as it is the most widely used and most scalable bioinformatics tool that has been developed in our group. Hence, it also generates the largest CO_2 footprint.

Initial experiments have revealed that using fewer cores for the computations and reducing the clock frequency of these cores (as our computations are predominantly memory-bandwidth-bound – i.e., the cores waste cycles/energy by waiting to retrieve data from the main memory) can substantially decrease the total amount of energy-to-solution required.

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 aid in inferring 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, climate-, and archaeological data, for instance, 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, to analyze microbial diversity in the human gut, and to shed light on population movements during the Greek Dark Ages (ca. 1100–750 BCE) in pre-historic times.

Finally, phylogenies play an important role in analyzing the dynamics and evolution of the current SARS-CoV-2 pandemic and in conducting local contact tracing.

Using Phylogenies to Predict Football Tournaments

Forty years after Joe Felsenstein's seminal paper, which described an efficient dynamic programming algorithm for computing likelihood on phylogenetic trees, we finally managed to find out what the most important application of this algorithm is: It can be used to efficiently and precisely calculate win probabilities for knockout tournaments (i.e., on the tournament trees) given the pairwise win probabilities for all possible matches. Prior to the knockout stage of the FIFA Euro 2020 tournament, which took place in 2021 due to the pandemic, we released a preprint (https://www.biorxiv.org/content/10.1101/2021.06.24.449715v1.abstract) with the respective prediction and announced the availability of the prediction via social media and a press release. As expected, the media echo was loud, and Alexis gave several radio interviews as well as a TV interview on SKY Sports Germany (https://youtu.be/b0R_yPVLsjs).

The initial prediction proved to be incorrect as football knockout tournaments are extremely difficult to predict because chance plays a major role due to the low average number of goals per match. In addition, the tournament could unfortunately not be played, for example, 100 times with the same teams to enable more accurate assessments of the correctness of our calculated win probabilities. Nonetheless, we correctly predicted at least one of the two finalists.
The key contribution of our Phylourny open-source code for computing exact win probabilities is its speed compared with standard approximate methods for calculating win probabilities, which entail simulating a tournament tree 10,000 or – typically – 100,000 times. This increase in scalability allows for computing informative novel statistics – and particularly the variance of win probabilities via Markov Chain Monte Carlo sampling – and might thus serve as a useful tool for computing better predictions in the future. A refined version of our method that was also applied to the NCAA 2021 basketball tournament (also known as March Madness in the U.S.) was implemented in the journal version of the paper, which has been submitted.

Finally, it is worth noting that former HITS colloquium speaker Achim Zeileis substantially contributed to improving the final paper.

In Figure 13, we outline the calculated per-team tournament-winner probabilities and their variances as box plots for the Euro 2020 tournament using our revised version of Phylourny. Note the generally large variance and the small differences in win probabilities.

Improving Simulated Data Studies

A key problem that has bothered more or less every lab member in recent years is how to properly set up realistic and meaningful simulation experiments for phylogenetics. As the true evolutionary history of most species is not known, we do not have any ground truth for evaluating novel methods using empirical datasets. Hence, at least for program verification, the computational phylogenetics community routinely uses data-simulation studies. In these studies, an artificial true tree is generated, and subsequently, sequences are simulated along this tree using the established statistical models of sequence evolution. However, it is always unclear how best to generate a simulated true tree and how to set the model parameters or model-parameter ranges of the sequence evolution models such that they are realistic.
As a solution, we introduced the freely accessible RAxML Grove database (https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btab863/6486526). For more than a decade, the lab has been involved in operating one RAxML web server each at the Swiss Institute of Bioinformatics and the San Diego Supercomputer Center, where biologists can analyze their empirical datasets free of charge. Now, RAxML Grove regularly retrieves anonymized tree- and model data (but not the actual sequence data) from these two web servers such that users who want to conduct simulation studies can chose from currently about

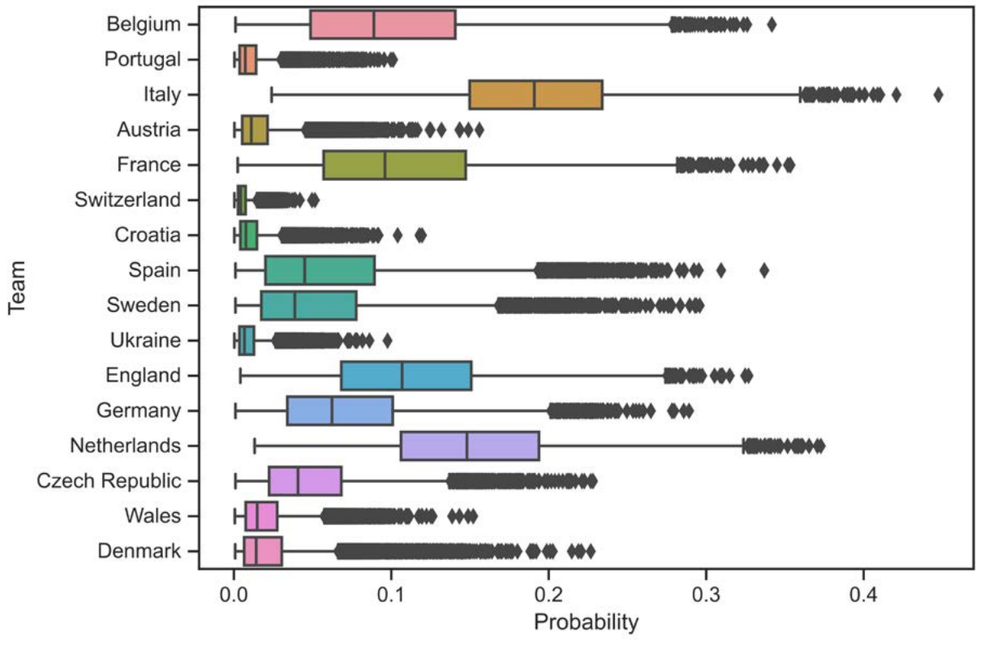


Figure 13: Box plot of probabilities of teams winning the Euro 2020 tournament.

60,000 empirical tree topologies that can be used as true trees for simulations, and these users can also draw realistic simulation model parameters from the respective histograms. Thus, by informing simulated data studies via empirical parameter distributions and tree shapes, we hope that more realistic simulated data studies can be conducted in the future that better reflect the intrinsic properties of the empirical datasets that are studied by the end users.

We believe that the RAxML Grove database will eventually constitute a gold mine for various follow-up simulation studies by our group. Two such studies are already underway.

Reconstructing Species Trees from Gene Trees

Substantial progress has also been made in the area of gene tree / species tree reconciliation (i.e., correcting gene trees using a given species tree or – in case of the new open-source SpeciesRax tool (https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab365/6503503) – using these gene trees to infer a species tree based on a given set of gene family trees). In 2021, we released the SpeciesRax tool – the first full-maximum-likelihood tool for inferring a rooted species trees from gene family trees under a statistical model that takes gene duplication-, gene loss-, and gene transfer events into account. In addition to inferring the species tree, the tool can also infer branch lengths for the species tree and introduces a new method for computing support values on the species tree. To calculate an initial "good" species tree (i.e., a reasonable, non-random starting tree with a "good" likelihood score), SpeciesRax integrates a novel Neighbor Joining algorithm that was developed within the framework of the insightful master's thesis of former master's student Paul Schade.

Using both empirical and simulated datasets, we showed that SpeciesRax is at least as accurate as the best competing methods while simultaneously being one order of magnitude faster on large datasets despite using a compute- and floating-point-intensive statistical model of reconciliation. We used SpeciesRax to infer a biologically plausible rooted phylogeny of vertebrates comprising 188 species from 31,612 gene families in just one hour using 40 cores.

A species tree that was inferred using SpeciesRax for a smaller empirical dataset that comprised 83 plant species is shown in Figure 15.

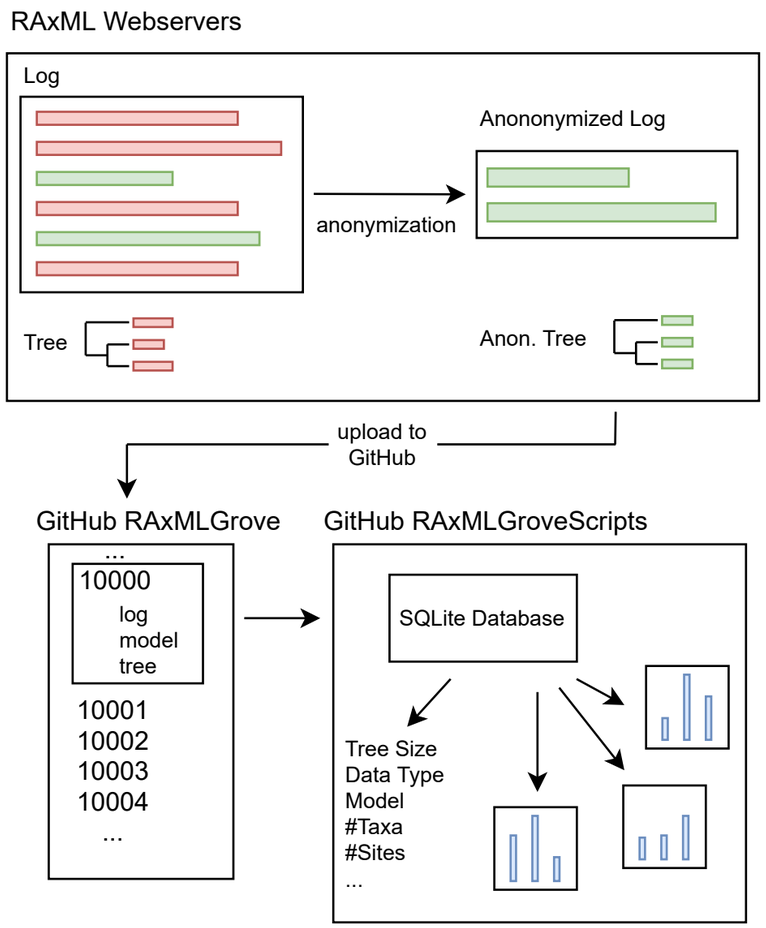


Figure 14: Schematic outline of the RaxML Grove data-processing pipeline and database.

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
- Analyse von Ancient DNA-Daten
- Methoden der Populationsgenetik

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

In der modernen Biologie ist die Evolution eine weithin akzeptierte Tatsache und kann heute anhand von DNA analysiert, beobachtet und verfolgt werden.
Ein berühmtes Zitat in diesem Zusammenhang stammt von Theodosius Dobzhansky: „Nichts in der Biologie ergibt Sinn, wenn es nicht im Licht der Evolution betrachtet wird".

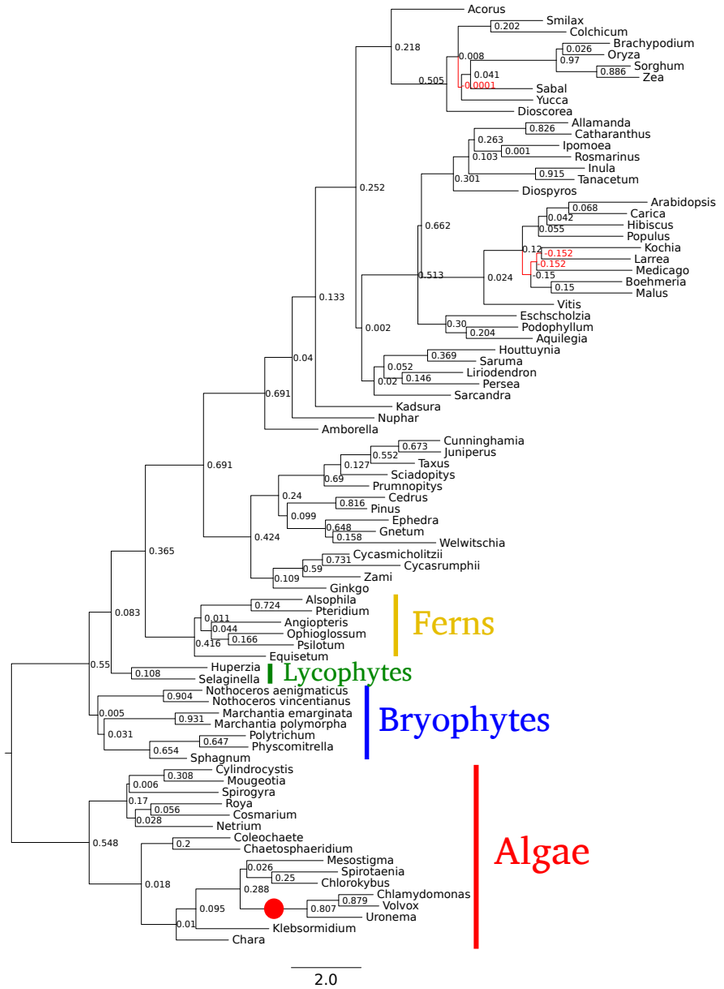
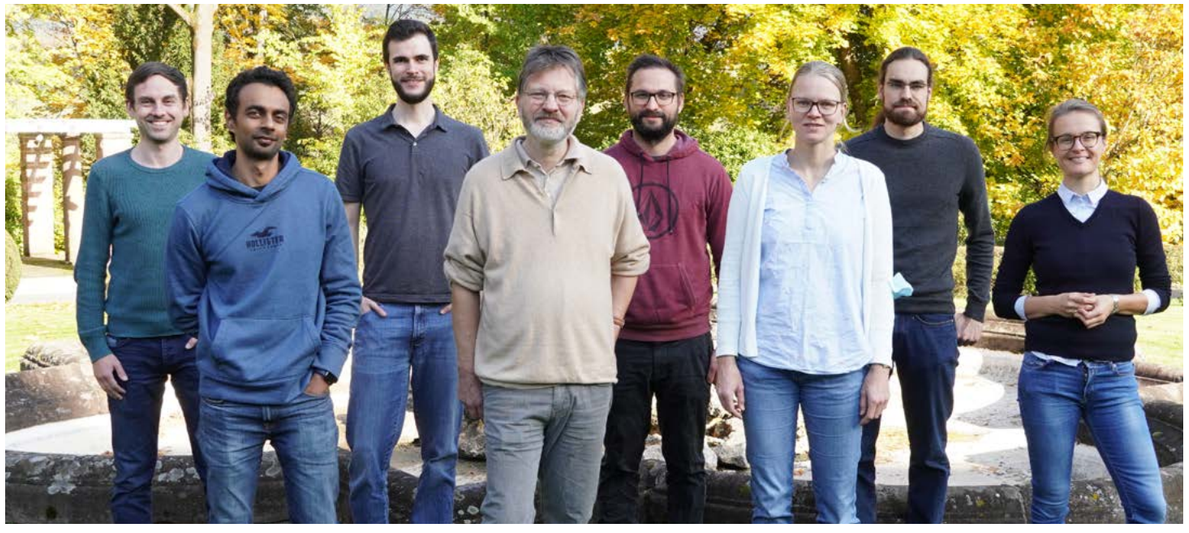


Figure 15: Species tree inferred with SpeciesRax from an empirical dataset comprising 83 plant species. The red circle on the tree indicates the inferred root position. The support values displayed at the nodes reflect the degree of agreement between the underlying gene family trees and the species tree shown here.

2.4 Computational Statistics (CST)



Group leader
Prof. Dr. Tilmann Gneiting

Group members
Dr. Jonas Brehmer
Dr. Alexander I. Jordan (staff scientist)
Dr. Ghulam Abdul Qadir (since August 2021)
Johannes Resin
Daniel Wolffram

Visiting scientists
Dr. Johannes Bracher
Dr. Sebastian Lerch
Prof. Dr. Melanie Schienle (since September 2021)
Eva-Maria Walz
(Karlsruhe Institute of Technology, Germany)
Jun.-Prof. Dr. Timo Dimitriadis
(Heidelberg University, Germany)
Prof. Dr. Johanna Ziegel (since July 2021;
University of Bern, Switzerland)

The Computational Statistics group at HITS was established in November 2013, when Tilmann Gneiting was appointed both group leader and 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, over the past several decades, we have borne witness to 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 methodologies in statistics and machine learning, notably in connection with applications.
While weather forecasting and collaborative research with meteorologists continue to represent prime examples of our work, we have also addressed challenges raised by the pandemic by establishing collaborative relationships with epidemiologists, by creating the national COVID-19 Forecast and Nowcast Hubs, and by contributing to similar efforts worldwide while placing methodological emphasis on generating and evaluating epidemiological ensemble forecasts.

General news

We are happy to report on an exceptionally productive year for our group. Indeed, in 2021, we maintained our thematic focus on theory, methodology, and applications of forecasting. Much of this work was performed jointly with our newly appointed guest scientists, Melanie Schienle (at our home university, the Karlsruhe Institute of Technology (KIT)) and Johanna Ziegel (at the University of Bern in Switzerland), along with their students and postdocs.

Together with Melanie Schienle and her group, we continued work on the German–Polish COVID-19 Forecast Hub regarding numbers of cases and deaths, and we developed the German Nowcast Hub for hospitalization incidence. Our postdoc Johannes Bracher led these efforts and contributed to related projects worldwide. He is also a co-recipient of the 2021 Arthur-Linder Prize from the Austro-Swiss region of the International Biometric Society. In parallel, we maintained an intense and fruitful interdisciplinary collaboration and exchange both with meteorologists at KIT and within the collaborative research group Waves to Weather, which focuses on weather prediction.

In August 2021, postdoc Ghulam Qadir joined our team in a HITS Lab project to collaborate with the Molecular Biomechanics (MBM) and Physics of Stellar Objects (PSO) groups and to develop emulators – that is, smart and computationally efficient surrogate models that complement more complex, potentially prohibitively computationally expensive physics-based computer simulations. Ghulam is a recipient of the 2021 Al-Kindi Statistics Student Research Award for his doctoral research on multivariate spatial statistical models at King Abdullah University of Science and Technology (KAUST) in Saudi Arabia. Here at HITS, Jonas Brehmer received a 2021 thesis award for his PhD work at the University of Mannheim. Among other topics, his work addresses critical needs in evaluating earthquake forecasts.

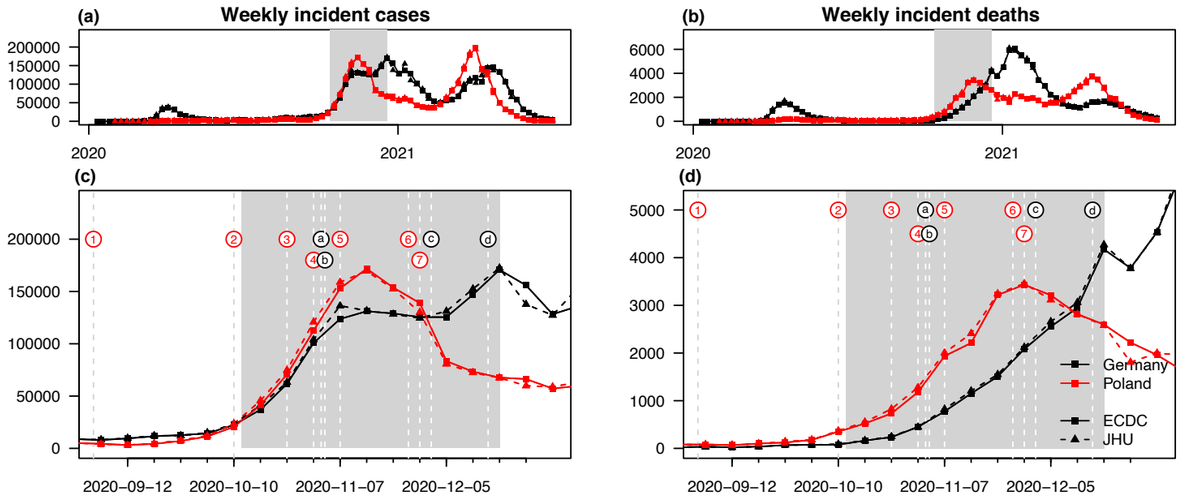
As in the previous year, 2021 was also shaped by a pandemic that has changed our lives and that continues to pose unprecedented challenges. Our internal group meetings, workshops, and informal exchange with colleagues around the world took place predominantly online in 2021. Among the few exceptions, we were able to hold a number of group meetings and other events in hybrid format. On 21–22 January, we held a hybrid meeting on isotonic regression with Johanna Ziegel's group at the University of Bern. Another hybrid workshop at Bern that focused on spatial stochastics and that was also organized by Johanna Ziegel brought together researchers from the University of Bern, HITS, and KIT on 11 October. In mid-October, our collaborator Warner Marzocchi – a geophysicist from the University of Naples Federico II in Italy – visited HITS to present a colloquium lecture and discuss ongoing joint work on the evaluation of operational earthquake forecasts (see Section 5.2). A bit later, on 25 October, we held a hybrid workshop on the topic of forecast calibration on the HITS premises. We report on this workshop in detail in Section 5.1.2.

Below, we describe the continued development of the German–Polish COVID-19 Forecast and Nowcast Hubs at KIT and HITS. Afterward, we move to methodological research on evaluating probability forecasts for binary events and to the development of isotonic distributional regression (IDR) – a powerful non-parametric technique for generating probabilistic forecasts.

COVID-19 Forecast and Nowcast Hubs

As detailed in last year's report, a highlight of our work in 2020 was the establishment of the German and Polish COVID-19 Forecast Hub (https://kitmetricslab.github.io/forecasthub/) in a joint endeavor with Melanie Schienle's group at the Chair of Econometrics and Statistics within the Department of Economics at KIT. In the Forecast Hub, we collaborate with modeling teams from Germany, Poland, Switzerland, the United Kingdom, and the United States that provide weekly forecasts in real time of confirmed cases and deaths from COVID-19 at prediction horizons of up to four weeks in advance. In addition, we produce and provide a range of baseline forecasts ourselves, and we aggregate the various individual forecasts into a simple ensemble forecast.

All forecasts are presented in a standardized format that consists of predictive quantiles at a total of 23 distinct levels. Equivalently, this format specifies the predictive median and central prediction intervals. The comparative evaluation of forecasts in this specific format poses new challenges, which we have addressed from both a theoretical and an applied perspective [Brehmer and Gneiting, 2021; Bracher et al., 2021a]. In particular, we recommend using the so-called weighted-interval score (WIS), which provides a negatively oriented (the smaller, the better) summary measure



Measures in Germany
a – semi–lockdown (restaurants closed, work f. home encouraged, no mass gatherings, inter–househ. contacts restr.; 2020–11–02)
b – new testing strategy announced (stricter criteria on symptoms; 2020–11–03)
c – reinforced restrictions on inter–househ. contacts (2020–12–01)
d – start of full lockdown (schools and non–essential shops; closed; 2020–12–16)

Measures in Poland
1 – stricter testing criteria (four symptoms needed; 2020–09–03)
2 – entire country classified yellow zone (mask wearing and limitations on gatherings; 2020–10–10)
3 – entire country classified red zone (partial school closure, restaurants closed, gatherings restricted; 2020–10–24)
4 – eased testing criteria (one symptom sufficient; 2020–10–31)
5 – reinforced restrictions (extension of school closure, shopping malls closed; 2020–11–07)
6 – bulk reporting of 22,000 cases (2020–11–24)
7 – shopping malls re–opened (2020–11–28)

Figure 16: Weekly incident cases (a, c) and deaths (b, d) from COVID-19 in Germany and Poland based on data from the European Centre for Disease Prevention and Control (ECDC) and the Centre for Systems Science and Engineering at Johns Hopkins University (JHU), along with landmarks in interventions and testing. The study period – from early October to mid-December 2020 – is highlighted in gray. Source: Figure 1 in [Bracher et al., 2021a].

of predictive performance and can be decomposed into a measure of sharpness and penalties for overprediction and underprediction, respectively. The WIS is a theoretically coherent measure that reduces to the mean absolute error (AE) in the case of a deterministic model that issues single-valued point forecasts only.

In [Bracher et al., 2021b], we report on the pre-registered short-term forecasting study that we conducted during the second wave of COVID-19 in early October through mid-December 2020. Figure 16 illustrates the development of the pandemic in Germany and Poland during the study period, and Figure 17 summarizes forecast performance for incident cases and incident deaths 1–4 weeks in advance by forecast horizon. The forecast models are identified by type, including simplistic baseline approaches, compartmental models, microsimulation models, approaches based on the renewal equation, a human-judgement project, and our Forecast Hub ensemble forecast, which aggregates information from the submitted models in real time. Cases are a more immediate measure of COVID-19 activity than deaths, and hardly any approach has succeeded in outperforming the Forecast Hub baseline beyond a prediction horizon of two weeks, with the sole exception being a microsimulation model for Poland. For deaths, both the ensemble forecast and various individual models outperformed the baseline at prediction horizons of up to four weeks in advance. Collaborative forecasting is generally beneficial, and our ensemble forecast performs better than the vast majority of individual models.

With increasing vaccination coverage in the year 2021, public attention shifted from case numbers to hospitalizations due to COVID-19. In Germany, the federal government made the decision to base interventions on seven-day hospitalization incidence – that is, the number of subsequently hospitalized cases of COVID-19 (normalized per 100,000 population) for which the date of infection – as recorded by local health authorities – lies within the seven preceding days. Evidently, the date of infection and the

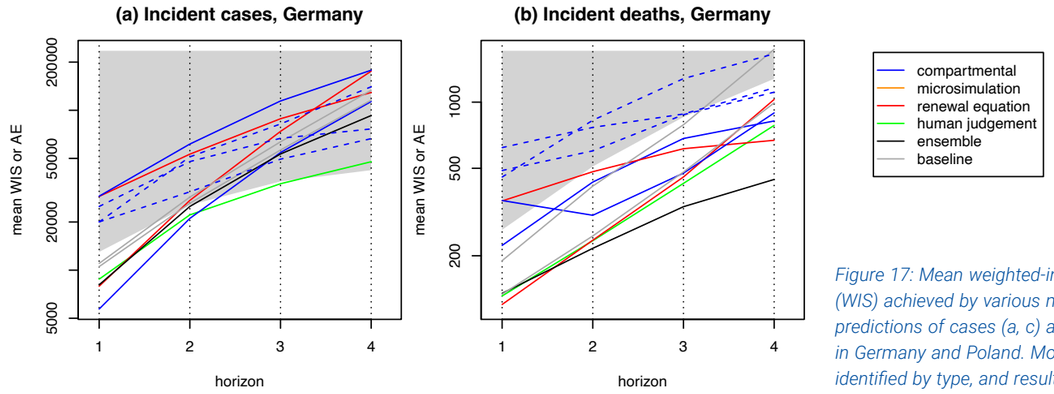


Figure 17: Mean weighted-interval score (WIS) achieved by various models for predictions of cases (a, c) and deaths (b, d) in Germany and Poland. Models are identified by type, and results are stratified by prediction horizon. For deterministic models, the mean absolute error (AE) is shown. The lower end of the gray area corresponds to the performance of the designated, simplistic Forecast Hub baseline. Line segments within the white area thus indicate that a model outperforms the baseline. Source: Supplementary Figure 9 in [Bracher et al., 2021a].

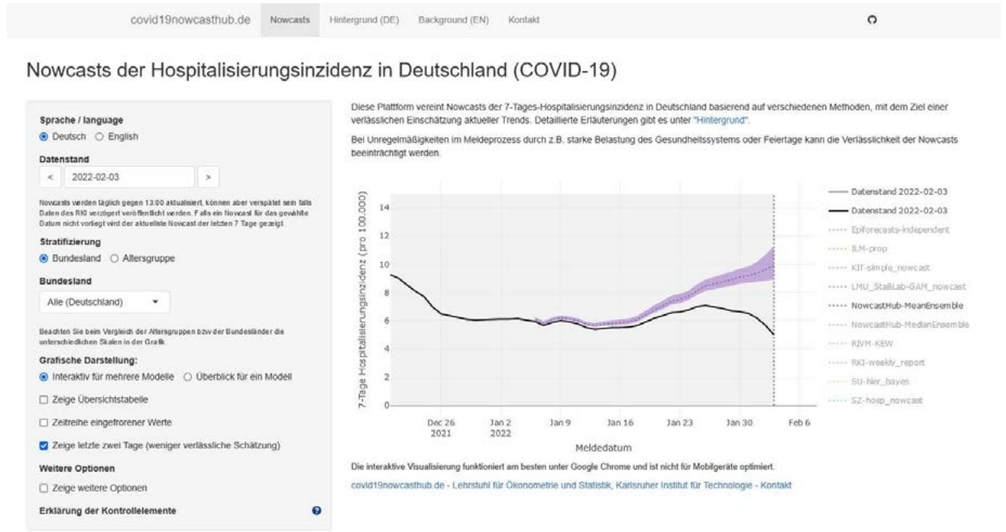


Figure 18: Screenshot from the Nowcast Hub from 3 February 2022. The thick black line shows the official seven-day hospitalization incidence as of this date. The broken black line shows our ensemble nowcast along with a 90% prediction interval.

date when hospitalization first occurs in official statistics can be several days or even weeks apart. Thus, daily values of hospitalization incidence are continuously corrected upward for days and weeks to come, and these values considerably underestimate the true number of hospitalizations caused by COVID-19. In particular, preliminary data can create the impression of decreasing hospitalization incidence even if this incidence is actually on the rise.

For these reasons, we developed the German COVID-19 Nowcast Hub (https://covid19nowcasthub.de/) in collaboration with – and under the leadership of – Melanie Schienle's group as well as with cooperating groups across Germany and beyond. A screenshot from the Nowcast Hub is shown in Figure 18 (see previous page). The main goal of this project is to reliably estimate the seven-day hospitalization incidence for Germany and its states and to assess recent trends using incomplete data. Nowcasts provide decision-makers and the general public with a better assessment of the current epidemic situation and have been reported by leading national media outlets. At the same time, our scientific interest lies in comparing different nowcasting methods and assessing the potential of combined ensemble nowcasts.

Stable reliability diagrams for probabilistic classifiers

Probabilistic classifiers assign predictive probabilities to binary events, such as rainfall tomorrow, a recession, a personal health outcome, or an epidemic. A key requirement for such a system is that it be reliable or calibrated. In other words, when looking back at a series of extant forecasts, the predictive probabilities should match the observed frequencies. For example, if we consider all cases with a predictive probability of about .80, the observed event frequency should also be about .80.

In practice, calibration is assessed graphically in reliability diagrams, but extant approaches to making this assessment are frequently hampered by ad-hoc implementation decisions, a lack of reproducibility, and inefficiency. While researchers and practitioners have been checking calibration for many decades, the topic has been subject to a surge of recent interest in machine learning spurred on by the recent recognition that modern neural networks tend to be uncalibrated. In [Dimitriadis et al., 2021], we introduced a new type of reliability diagram that addresses and resolves these issues.

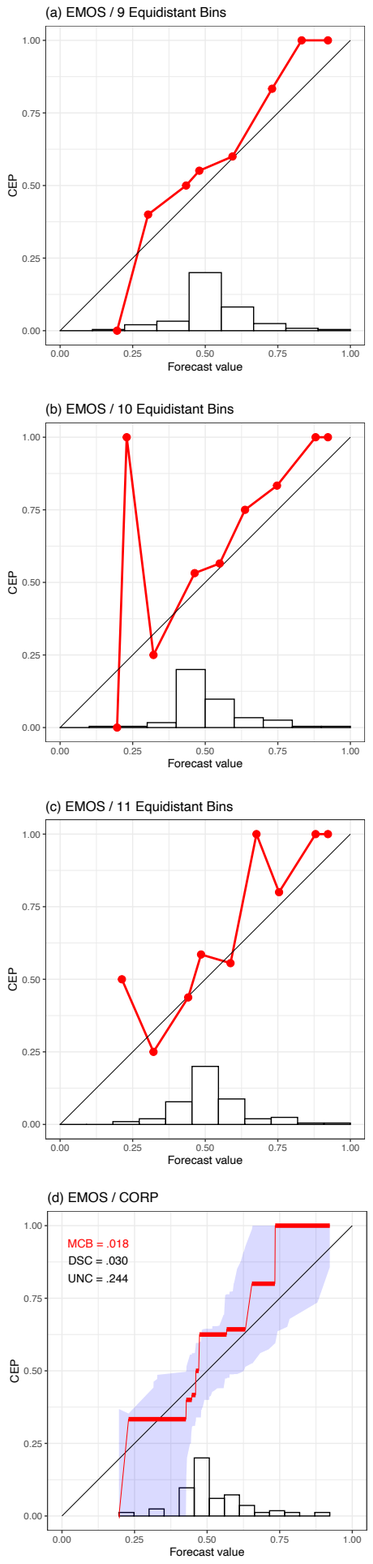


Figure 19: Reliability diagrams for the probability of precipitation forecasts over Niamey, Niger, between July and September 2016 made with the EMOS method using the binning-and-counting approach with a choice of (a) 9, (b) 10, and (c) 11 equidistant bins, together with (d) our newly developed CORP reliability diagram, for which we provide uncertainty quantification through 90% consistency bands. Source: Figure 2 in [Dimitriadis et al., 2021].

As the key diagnostic tool, a reliability diagram plots an observed event frequency against the predictive probability of the event. This process is straightforward in discrete settings in which there are only a few distinct predictive probabilities and in which each predictive probability occurs sufficiently many times. However, statistical and machine-learning approaches to binary classification generate continuous predictive probabilities that can take any value between 0 and 1, and the forecast values are typically pairwise distinct. In these settings, researchers have relied on the binning-and-counting approach, which begins by dividing the range from 0 to 1 into a typically arbitrary number of bins for the forecast values. Then, for each bin, the respective conditional event frequency versus the midpoint or average forecast value in the bin is plotted. The two quantities should match for calibrated forecasts, and the plotted points should thus lie on the diagonal.

The visual appearance of these classical reliability diagrams is highly sensitive to the specification of the bins. We provide an example in Figure 19 for the probability of precipitation forecasts over Niamey, Niger, between July and September 2016. These forecasts were made with the ensemble-model-output-statistics (EMOS) technique and were studied in our work on probabilistic weather forecasts [Vogel et al., 2021]. Choosing 9, 10, or 11 equidistant bins yields drastically distinct reliability diagrams. This situation is disconcerting for a widely used tool of data analysis and stands in opposition to well-argued recent pleas for reproducibility and stability.

In our paper, we introduced a new approach that resolves these issues in a theoretically optimal way that is subject to regularization with isotonicity. In other words, an increase in the forecast value cannot coincide with a decrease in the conditional event frequency. In a nutshell, we perform nonparametric isotonic regression using the pool-adjacent-violators (PAV) algorithm to estimate conditional event probabilities, which yields a fully automated choice of bins that adapts to both discrete and continuous settings without any need for tuning parameters. We equip the diagram with the new quantitative measures of (mis) calibration (MCB), discrimination ability (DSC), and uncertainty (UNC), as displayed in panel (d) of Figure 19. We call this stable new approach "CORP," which is an acronym for the powerful properties of "consistency" in the classical statistical sense of convergence to population characteristics, "optimality" in that no other choice of bins generates more skillful (re) calibrated forecasts under the isotonicity constraint, "reproducibility" by not requiring any tuning parameters, and is based on the "PAV algorithm". Our implementation of CORP is available as free and open-source software for the R language and environment for statistical computing and graphics.

Die **Computational Statistics Gruppe (CST)** 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 die Form von Wahrscheinlichkeitsverteilungen über zukünftige Ereignisse und Größen annehmen. Dementsprechend erleben wir seit nunmehr einigen Jahrzehnten einen transdisziplinären Paradigmenwechsel von deterministischen oder Punktvorhersagen hin zu probabilistischen Vorhersagen. Ziel der CST Gruppe ist es, diese Entwicklungen nachhaltig zu unterstützen, indem sie theoretische Grundlagen für wissenschaftlich fundierte Vorhersagen entwickelt, eine Vorreiterrolle in der Entwicklung entsprechender Methoden der Statistik und des maschinellen Lernens einnimmt und diese in wichtigen Anwendungsproblemen, wie etwa in der Wettervorhersage, zum Einsatz bringt.

In diesem Zusammenhang pflegen wir intensive Kontakte und Kooperationen mit Meteorolog/-innen zu Wettervorhersagen. Über kollaborative Projekte mit Epidemiolog/-innen, den Aufbau der nationalen COVID-19 Forecast und Nowcast Hubs und die Unterstützung von ähnlichen Projekten weltweit stellen wir uns durch die Pandemie ausgelösten neuen Herausforderungen. Unsere besondere Aufmerksamkeit gilt dabei der Erzeugung und Bewertung von epidemiologischen Ensemblevorhersagen.

Isotonic distributional regression

Isotonicity is a natural constraint in estimation- and prediction problems. Consider a postprocessing technique in weather forecasting in which the respective future weather quantity is modeled given the output of numerical-weather-prediction (NWP) models: Intuitively, if the NWP model output indicates more precipitation accumulation, the associated regression functions should also be larger. Isotonic relationships of this type hold in a plethora of applied settings. In fact, standard linear-regression analysis rests on the assumption of isotonicity.

There is also an emerging consensus in the transdisciplinary literature that regression analysis should be distributional and have a goal of modeling the conditional distribution of a response variable given the values of one or more explanatory variables. Distributional regression marks a clear break from the classical view of regression, which has focused on estimating the conditional mean of the response variable. Later extensions have considered other functionals of the conditional distributions, such as quantiles or expectiles. However, reducing a conditional distribution to a single-valued functional results in a tremendous loss of information.

In [Henzi et al., 2021], we developed a powerful new approach to nonparametric distributional regression under isotonicity constraints that we call isotonic distributional regression (IDR). The proposed distributional-regression technique generates a mapping from a covariate value to a probability measure, which serves to model the conditional distribution. For a covariate space with a partial order in the formal mathematical sense, this mapping is isotonic if a larger covariate value in the given partial order implies a conditional distribution that is larger in the stochastic order in the sense used in probability theory.

Useful comparisons of predictive distributions are in terms of proper scoring rules, with the continuous ranked probability score (CRPS) serving as a prominent and relevant example. In the paper, we prove that there is a unique isotonic distributional regression that is optimal with respect to the CPRS, and we refer to this regression as the IDR solution. As it turns out, the IDR solution is universal in that the estimate is optimal with respect to a broad class of proper scoring rules. Simultaneously, IDR avoids pitfalls commonly associated with nonparametric distributional regression, such as suboptimal partitions of the covariate space and level crossing.

Figure 20 shows IDR conditional cumulative distribution functions (CDFs) estimated on a training set. IDR is capable of estimating both the strongly right-skewed conditional distributions for lower values of the covariate X and the more symmetric distributions as X increases. The estimated conditional CDFs are piecewise constant, and they never cross one another. The computational cost of IDR may become prohibitive for large training sets. However, IDR can be usefully combined with subsample aggregation – much in the spirit of random forests – with the benefits of yielding reduced computational costs under large training samples, smoother regression functions, and (frequently) improved predictive performance.

We contend that IDR provides an attractive and widely applicable competitive benchmark in probabilistic forecasting problems. The use of benchmark techniques has been called for across application domains, and suitable methods should be competitive in terms of predictive performance while avoiding implementation decisions that may vary from user to user. IDR is well suited to this purpose as it is entirely generic; it does not involve any implementation decisions other than the choice of the partial order; it applies to all types of real-valued outcomes with discrete, continuous, or mixed discrete–continuous distributions; and it accommodates general types of covariate spaces.

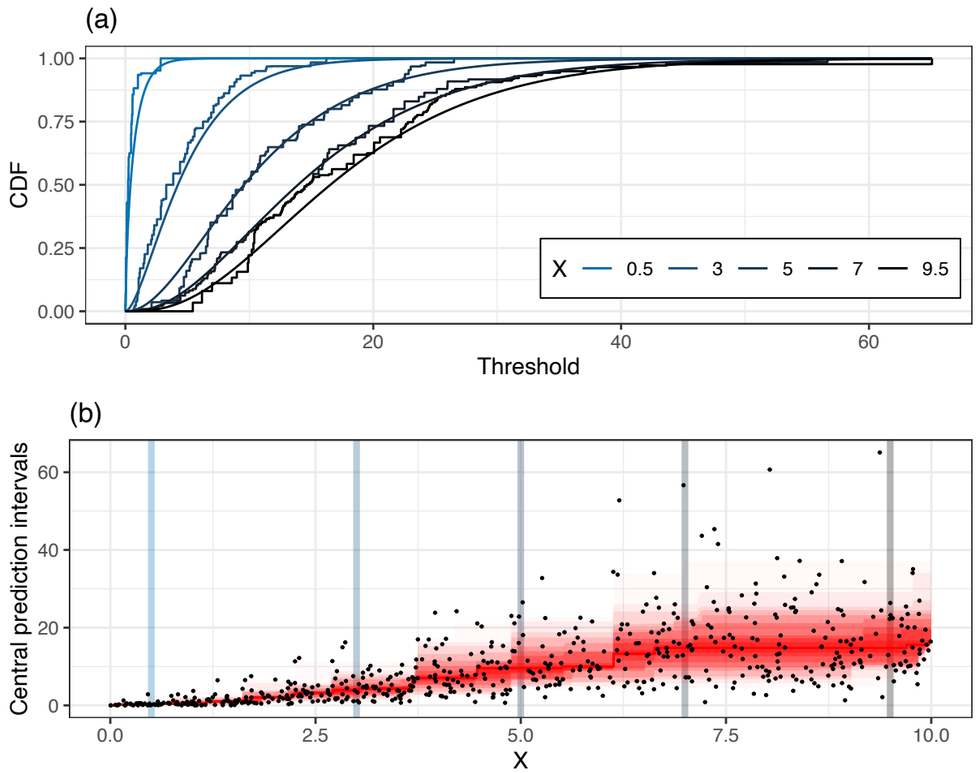


Figure 20: Simulation example based on a training sample of size 600: (a) True conditional CDFs (smooth) and IDR estimates (step functions) for selected values of the covariate. (b) IDR-estimated conditional distributions. The shaded bands correspond to probability masses of 0.10 each, with the darkest shade marking the central interval. Vertical strips indicate the cross-sections corresponding to the values of the covariate in panel (a). Source: Figure 1 in [Henzi et al., 2021].

2.5 Data Mining and Uncertainty Quantification (DMQ)

Group leader
Prof. Dr. Vincent Heuveline

Staff members
Dr. Philipp Gerstner
Philipp Lösel
Dr. Chen Song (until February 2021)
Jonas Roller (since April 2021)
Valentin Schmid (since September 2021)

Visiting scientists
Aksel Alpay (Heidelberg University)
Saskia Haupt (Heidelberg University)
Alejandra Jayme (Heidelberg University)
Elaine Zaunseder (Heidelberg University)
Marcus Buchwald (Heidelberg University)

Students
Jacob Jonas Relle
Jonas Roller (until April 2021)

The Data Mining and Uncertainty Quantification (DMQ) group – headed by Vincent Heuveline, Professor of Scientific Computing at Heidelberg University – began its research in May 2013. The group works in close collaboration with the Engineering Mathematics and Computing Lab (EMCL) – which is also headed by Vincent Heuveline – at the Interdisciplinary Center for Scientific Computing (IWR) at Heidelberg University. The DMQ group's research focus lies in gaining knowledge from extremely large and complex datasets through data-based modeling and data-mining technologies. Reliability considerations with respect to these datasets are addressed via methods of uncertainty quantification. 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 2021, the DMQ group focused on research activities in the areas of uncertainty quantification, machine learning, and numerical simulation for biomedical applications.

Using data mining and pattern recognition to support newborn screening

Newborn screening (NBS) programs aim at early – ideally pre-symptomatic – identification of treatable rare diseases that carry significant health burdens in order to reduce morbidity and mortality rates. For the diagnostics, blood samples taken from newborns within a few days after birth are analyzed to identify inherited metabolic diseases. Accurately and efficiently diagnosing these diseases is important but challenging due to their low prevalence. A quick diagnosis can lead to efficient therapies and treatments that can positively change the outcome and severity of the disease for these newborns. NBS programs demand both a very high sensitivity (ideally 100%) in order to avoid false negatives and a very high specificity (at least 99.5%) in order to keep the number of false positives low. This is especially challenging in NBS because the birth prevalence of the target diseases is very low (ranging from 1:10,000 to <1:1,000,000). Traditional cut-off-based approaches in NBS integrate only a fraction of the available information. Moreover, these approaches focus on the primary variables of the metabolic pathway that is affected in a particular metabolic disease and cannot deal with complex relationships among metabolites.

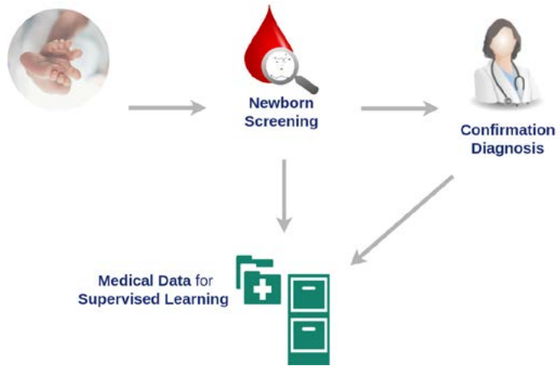


Figure 21: Medical data from newborn screening and confirmation diagnosis enable supervised machine learning methods to be applied.

In a collaborative initiative with the Center for Pediatric and Adolescent Medicine at University Hospital Heidelberg, we are currently working on a pilot project involving data from more than 2.2 million newborns that aims to improve the diagnosis of isovaleric acidemia. We therefore developed mathematical-based methods to exploit the complete information of NBS test results in order to improve the specificity and positive prediction of NBS. We applied these methods to examine large datasets with high-dimensional feature spaces by implementing a machine learning pipeline for NBS. By applying pattern recognition and dimensionality reduction techniques, such as t-distributed neighbor embedding and linear discriminant analysis, we were able to detect previously unknown metabolic patterns within the data. This enabled us to classify the patients by applying supervised machine learning methods, such as logistic regression analysis and support vector machines, thereby leading to reduced false positive classification while maintaining 100% sensitivity. These results are now being further analyzed in terms of how they can improve newborn screening programs.

Figure 22: Analyzed samples from newborns are taken shortly after birth (source: https://unsplash.com/s/photos/newborn).

Using a Bayesian deep learning framework to analyze cancer with radiogenomics data

Radiogenomics is based on the idea that entities at different scales – such as those of molecules, cells, and tissues – are linked to one another and can therefore be modeled together. Specifically, the field studies correlations between genome-wide molecular data and quantitative image features extracted from radiology images. There are three areas in which radiogenomics can make a contribution. The first is in the study of heterogeneity because intra- and inter-individual heterogeneities exist in complex diseases. However, the technology needed to assess this heterogeneity is not readily available.

Radiogenomic data fusion can have significant contributions by uncovering insights into how data at different scales are linked to one another. The second area in which radiogenomics can make a contribution is in finding non-invasive biomarkers. Currently, most clinically used biomarkers require a piece of tissue from surgical sampling;

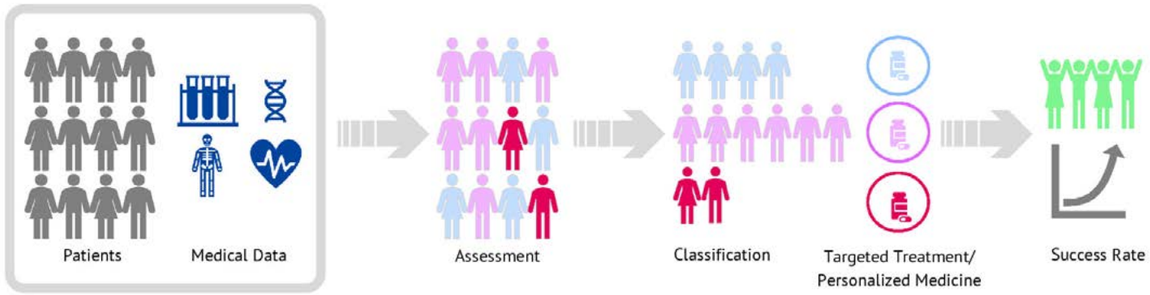


Figure 23: Leveraging advances in technology (e.g., data science, deep learning, hardware) and the increased availability of medical data for diseases prevention, early detection, targeted treatment, and personalized medicine.

however, this process carries added risk to patients and may not even be possible in some situations. Quantitative imaging is a great candidate for investigating whether an image feature or a combination of image features can serve as a biomarker.

The third and final area in which radiogenomics can make a contribution is in providing support for biomedical decisions. Radiogenomics can aid in predicting diagnosis, prognosis, and treatment in the context of precision medicine. Moreover, molecular biomarkers have already shown promise for changing treatment schemes, and quantitative imaging analyses have been reported to be capable of predicting prognosis and treatment. Combining both efforts might improve the accuracy of predictions of patient outcomes. The two general types of applications of radiogenomics modeling lie in finding associations between image- and molecular phenotypes (i.e., radiogenomic maps) and in supporting biomedical decisions using radiogenomic data. Most of these applications are in the field of oncology, for which radiogenomic biomedical data are routinely collected.

In our work, we propose using a deep learning framework to radiogenomically analyze cancer (see Figure 24). The first step in the framework is feature extraction (i.e., extracting image features from radiology data and extracting molecular markers from genomic data). Next, feature integration is accomplished through a perception component that is implemented via a Bayesian formulation of a deep neural network. The results are leveraged in a task-specific component – which can incorporate human knowledge or other predictive models – in order to determine relationships or to predict decisions. The task-specific component is a probabilistic graphical model that expresses the conditional dependence between observed and latent variables. Unifying the perception and task-specific components under a principled probabilistic framework is referred to as Bayesian deep learning. This approach provides a way for uncertainty in the data and in the model to be quantified and can provide insights into how confident the model is in its predictions, which can then be used to accept or reject the results.

We demonstrate the functionality of the framework by predicting the survival rate of glioblastoma multiforme patients and by exposing the significant feature correlations in these predictions. The used data collections stem from the Cancer Imaging Archive and from the Genomic Data Commons. In particular, we consider datasets from The Cancer Genome Atlas program.

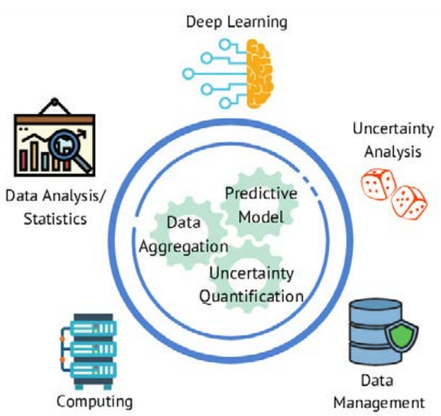


Figure 24: The framework is a data-driven predictive model that offers multi-modal data aggregation and uncertainty quantification.

Modeling colon cancer development at different scales

Cancer development is a multi-scale process: Alterations on the DNA level may lead to functional changes on the cellular level, which can alter tissue behavior and thus lead to cancer (Figure 25). Different alterations may lead to different molecular pathways of cancer development: In other words, there are different possibilities of how cancer can develop. There are thus also a variety of different types of cancer with immediate implications for clinical treatment- and prevention strategies at the population level. Therefore, in order to better understand cancer development and to improve clinical decision-making, the interactions between different scales of cancer development must be investigated. However, many underlying processes can hardly be observed in vivo. Thanks to the increasing volume of genetic and molecular data, mathematical modeling approaches can be used to unravel this multi-scale process.

This is exactly the aim of our collaborative initiative with the Department of Applied Tumor Biology (ATB) at University Hospital Heidelberg, which has led to a three-year research project entitled "Mathematics in Oncology – Towards optimal prevention and treatment in patients with inherited cancer syndrome" funded by the Klaus Tschira Foundation.

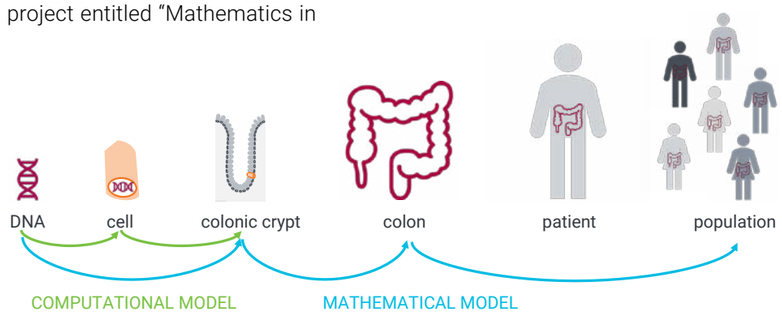


Figure 25: Modeling Lynch syndrome colorectal cancer development at different scales. The computational model describes processes at the cellular level within a single crypt. The mathematical model analyzes the impact of the alterations at the DNA level on the tissue level.

We developed one mathematical model each that describes cancer development (a) at the tissue level and (b) at the cell level using the example of colorectal cancer in Lynch syndrome, which is the most common form of inherited colorectal cancer predisposition syndrome.

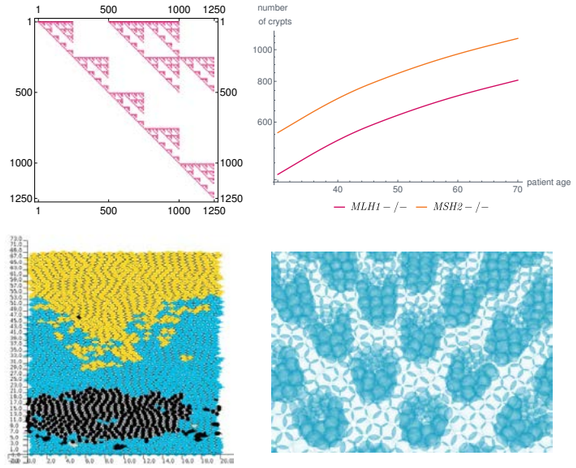


Figure 26: Modeling approaches and simulation results.
Top: A mathematical model of the tissue level. Left: The chosen matrix structure of the differential equation model allows for a systematic analysis of the individual components. Right: By extracting components of the model solution vector, we are able to simulate the evolution of specific precursors of colon cancer in Lynch syndrome. Results are in line with current medical data.

Bottom: A computational model of the cellular level. Left: A 2D simulation of a specific alteration (in black) in a single cell that is spreading throughout the crypt (simplified rectangular geometry). Right: A first 3D simulation of multiple crypts that aims to reproduce the morphological features of the crypts as accurately as possible.

First, at the tissue level, the mathematical model [Haupt et al, 2021a] uses differential equations to describe the evolution of different pathways of and precursor states during colorectal cancer development based on the mutations involved at the DNA level (Figure 26, top right). The model consists of different components that are all based on recent experimental data. The structure of the model matrix allows for precise mathematical analysis and medical interpretation (Figure 26, top left).

Second, using a computational model [Haupt et al, 2021b], we can more closely examine the cellular dynamics of individual colonic crypts – which are the building blocks of the colon wall – during Lynch syndrome colorectal cancer development. We simulated the effect of different alterations at the DNA level on the entire crypt (Figure 26, bottom left). Furthermore, we quantified how alterations in single cells within a crypt can take over the crypt such that they are present in every cell of the crypt. We are currently extending

this model to a three-dimensional setting that can visualize the interaction between multiple crypts in order to bridge the gap between the cellular and the tissue level (Figure 26, bottom right).

In summary, using these models, we obtained simulation results that are in line with clinical observations. Our models represent first mathematical approaches to providing a detailed multi-scale understanding of Lynch syndrome colorectal cancer development and thus to providing more effective clinical strategies for Lynch syndrome carriers.

Heat transfer in multi-physics flow simulation

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

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

Figure 27 illustrates the considered geometry, in which the fluid is contained inside the gap between two concentric cylinders. The inner cylinder is heated, whereas the outer cylinder is cooled. In addition, an electric field is applied between the two cylinders.

While conducting simulations in a vertical cylinder with distinct temperature differences between the inner and outer walls, we were able to observe the emergence of axially oriented vortices both in experimental and in numerical simulation. Due to these vortices, radial heat transfer is enhanced, as shown by the increase in the corresponding Nusselt number.

In the horizontal cylinder scenario, the fluid behavior is very much unlike that in the vertical case and can be characterized by periods of simultaneously

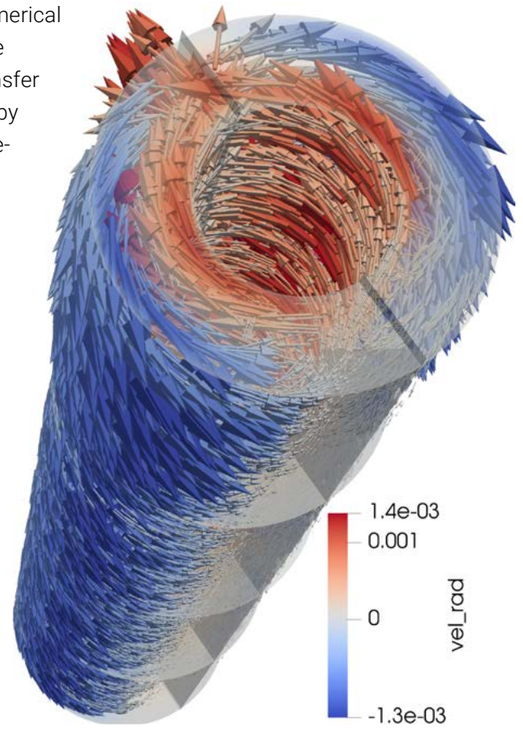


Figure 27: Radial velocity in a horizontal setting with a temperature differential of 10K.

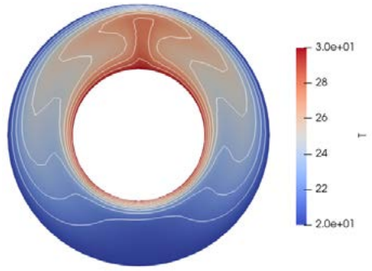


Figure 28: Temperature at the middle of the horizontal cylinder.

rising temperatures at the top of the cylinder and falling temperatures at the bottom. A stationary solution is only obtained for very low temperature differentials and applied voltages.

Figure 28 provides a snapshot of the fluid temperature in a situation in which the inner cylinder is 10K warmer than the outer cylinder, with an applied voltage of 7kV.

Finite element simulation of electric fields around molecules

When simulating molecular dynamics, for example, to investigate the effects of drugs or to better understand biochemical processes, it is crucial to compute the electric field that is induced by a certain molecule.

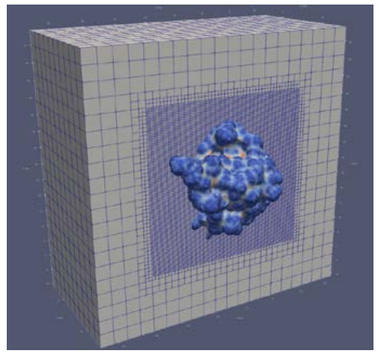


Figure 29: Solvent-excluded surface (SES) that correspond to the molecule Ubiquitin, and the surrounding computational mesh.

Typically, these molecules of interest are immersed in a solvent, such as water, and the electric field is created by the ions inside the molecule according to Gauss's law of electrostatics. In this joint project with the MCM group, we aim to develop a computationally efficient method for solving the associated Poisson–Boltzmann equations (PBE). The core of our method is formed by the so-called cut finite element method (CutFEM) in combination with a geometric multigrid solver (GMG).

CutFEM is used to tackle one of the main challenges when solving PBE: the presence of a jump in the electric field across the molecule's surface, which is caused by different dielectric properties of the molecule and of the surrounding solvent. With CutFEM, the computationally expensive generation of an interface-matching finite element mesh can be avoided. Instead, this method allows for using very simple structured grids that do not have to align with the molecule's geometry, which might be highly complex. Structured grids have several beneficial properties that can be used to speed up the overall simulation process.

First, they are well suited to be combined with GMG as linear solvers. GMG is considered to be among the most efficient methods for solving linear systems of equations. Indeed, GMG has O(N) complexity, where N denotes the number of grid points. Furthermore, this method exhibits good parallel scalability, thereby rendering it a favorable choice when using high-performance computing (HPC) systems. Another benefit of structured grids is the fact that they allow key linear algebra operations – such as matrix–vector products – to be performed in a stencil fashion. In this way, it is possible to reduce data traffic between the main memory and the processor, which is usually the bottleneck for numerical simulations. In our future work, our goal is to further exploit this property by integrating graphics processing units (GPU) into our numerical code.

Figure 29 illustrates how the molecule is immersed in a structured, hexahedral mesh. The visualized structure is the so-called solvent-excluded surface (SES), which denotes the part of the molecule that is accessible by the outer solvent. Figure 30 illustrates the electric potential, u, on the SES (right) and on several of its iso-surfaces (left).

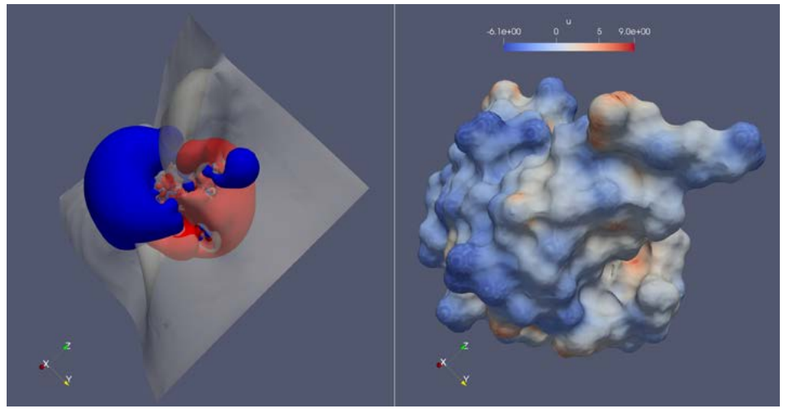


Figure 30: Isosurfaces of the electric potential (left) and of the electric potential on the SES (right).

Die Forschungsgruppe **Data Mining and Uncertainty Quantification (DMQ)** unter der Leitung von Vincent Heuveline, Professor für Wissenschaftliches Rechnen an der Universität Heidelberg, besteht seit Mai 2013. Sie arbeitet eng mit dem „Engineering Mathematics and Computing Lab" (EMCL) 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.
2021 wurde dazu in der Gruppe schwerpunktmäßig in folgenden Anwendungsbereichen gearbeitet: Uncertainty Quantification, maschinelles Lernen und numerische Simulation für biomedizinische Anwendungen.

2 Research

2.6 Groups and Geometry (GRG)

Group leader
Prof. Dr. Anna Wienhard

Staff members
Fernando Camacho Cadena
Dr. Brice Loustau
Dr. Mareike Pfeil (until December 2021)

Visiting scientists
Dr. Giulio Belletti (until November 2021)
Dr. Xian Dai
Colin Davalo
Dr. Nguyen-Thi Dang
Dr. Valentina Disarlo
Jun. Prof. Dr. James Farre (until July 2021)
Dr. Pengfei Huang
Dr. Mitul Islam
Dr. Dani Kaufman
Dr. Georgios Kydonakis
Marta Magnani
Arnaud Maret
Merik Niemeyer
Jun. Prof. Dr. Beatrice Pozzetti
Dr. Anja Randecker
Dr. Max Riestenberg
Evgenii Rogozinnikov (until March 2021)
Dr. Carmen Rovi (until July 2021)
Dr. Anna Schilling
Dr. Diaaeldin Taha
Dr. Gabriele Viaggi

Students
Levin Maier
Menelaos Zikidis (until July 2021)

The Groups and Geometry research group works closely with the Geometry & Dynamics Research Station at Heidelberg University. Both groups are headed by 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 it unchanged. They can be composed – that is, applied one after the other – to form a mathematical structure 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 group of transformations. In short: Geometry is symmetry.
This concept unified classical Euclidean geometry, the newly discovered field of hyperbolic geometry, and projective geometry, which has its origins in the study of perspective in art and is not based on the measurement of distances, but rather on incidence relations. Klein's concept fundamentally changed our view of geometry in mathematics and theoretical physics and continues to influence these fields to this day.

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 symmetries of these spaces. We also apply the study of groups and geometry to other sciences, such as mathematical physics, data science, and machine learning.

Experimental mathematics

Contrary to widespread belief, mathematics is far from a purely deductive science. Like all scientists, mathematicians test their intuitions and theories in the real world over and over again. This is especially true of geometers, who often represent the objects that they study and measure various quantities. In fact, the word geometry refers to measurement. To take an illustrious example, the Pythagorean theorem was surely observed experimentally before it was proved deductively within a mathematical system, such as Euclid's Elements.

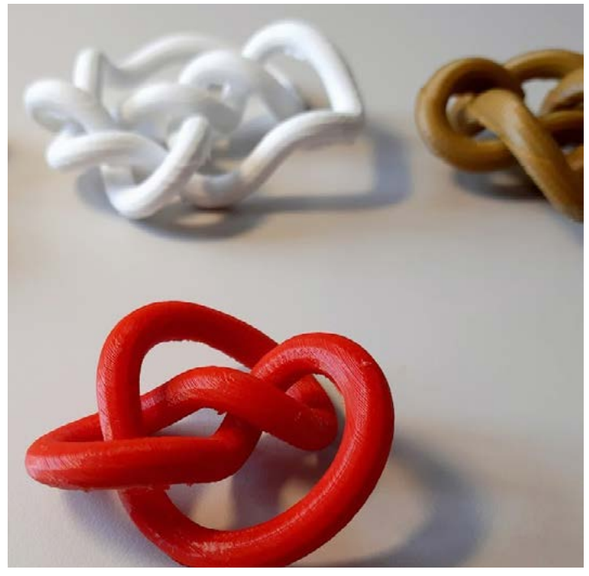
Creating representations of mathematical objects can be as simple as drawing a picture on a blackboard or tying tying a knot in a rope, but it can also involve more advanced technology. Understandably, the rise of computers has had a tremendous effect on mathematical research by providing computing- and visualization capacities that would have been inconceivable to previous generations.

Admittedly, mathematics is too often taught as a perfectly deductive and dry science, which is a far cry from its true nature as it is known to any working mathematician. Thankfully, an increasing number of initiatives now promote a more experimental, fun, and collaborative approach to mathematics at all levels.

In this year's HITS report, we showcase the experimental approach to mathematics that Anna Wienhard is spearheading at HITS as well as at Heidelberg University, where she is the director of the Geometry & Dynamics Research Station. Within this structure, the Heidelberg Experimental Geometry Lab (HEGL) was created in 2021 in collaboration with HITS and has had an amazingly positive response from students and members.

The Heidelberg Experimental Geometry Lab (HEGL)

Prelude: Experimental math at HITS (since 2016)
The GRG group has been conducting experimental mathematics research and collaborating with other HITS groups since its creation in 2015. In 2016, we started working on math visualization and 3D printing of



Figures 31a and 31b: 3D section of a Calabi–Yau quintic and knots. Printed at HITS in 2020.

complex mathematical objects, such as hyperbolic knots, toric fibrations, and Calabi–Yau manifolds. Our student worker Menelaos Zikidis (now a PhD student at the University of Sheffield) was responsible for many beautiful creations, some of which can be admired on his web page (see Figures 31a and 31b, previous page).

Conception and creation of HEGL (2019–2021)

Figure 32: In the Lab room, Anna Wienhard (right) is seen talking with Peter Albers (center) and Anna Schilling (left).

The idea to create a geometry lab in Heidelberg was born out of Anna Wienhard´s vision in a context where similar initiatives were being developed. In the United States, geometry labs have been [created in several universities under the umbrella of the Geometry-Labs-United (GLU) network (see [AMS article]), of which HEGL is now the first European node. In France and in Germany, a flourishing community of math visualization also exists. The philosophy of the geometry labs relates – on a much smaller scale – to the Fab Lab Network, an open community of digital fabrication labs around the world that aims to democratize access to technical tools.

The creation of HEGL became possible thanks to the support of both Heidelberg University and HITS. Planning began in late 2020 with the creation of a HEGL team, a room layout, and IT services; the selection and ordering of equipment; and the organization of seminars, communication, etc. The Lab room became available in spring 2021, and the official opening took place in July.

The HEGL Community and Philosophy

Figure 33: HEGL members in conversation.

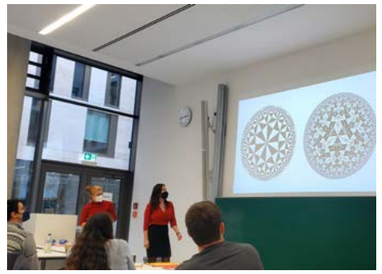
The goal of HEGL is to bring together a community of students and researchers at all levels both to learn from and exchange information and ideas with one another. HEGL promotes an inclusive, collaborative, and experimental approach to mathematics research:

- All are welcome: university students, members of the Research Station, HITS members, and all other math enthusiasts.
- Projects, seminars, and outreach events serve as catalysts for collaboration and community engagement.
- The Lab provides equipment and technical support for carrying out experimental and computational mathematical investigations.

Figure 34: Students working in the Lab.

In order to run the Lab and coordinate its activities, a dedicated team is in place that comprises:

- A Lab Manager: Dia Taha (since Sept. 2021)
- Members of the Research Station: Brice Loustau (the author of this report), Anja Randecker, and many others



Figures 35a and 35b: Student presentations.

- Student workers (staffing open hours and offering technical support for students)
- Hiring Peter Smillie as Junior Professor for “Geometry” who will join in fall 2022.

Figures 36a and 36b: Sierpinski Christmas tree, which was built in a joint effort by the community.

Lab Room and Equipment

The Lab room is located in the Mathematikon (INF 205) at Heidelberg University. It is a beautiful room of approx. 60m² with a glass wall that allows visitors look inside.
The Lab is nicely furnished and equipped with cutting-edge technology, including

- High-performance computers and 4k monitors
- Two 3D printers (Figure 39, next page)
- A laser engraver
- Virtual Reality equipment (see Figure 38, next page)
- A sofa, a conference table, a large TV, a glass display case, books, recreational math toys, games, a coffee machine. (see Figure 34).

Activities

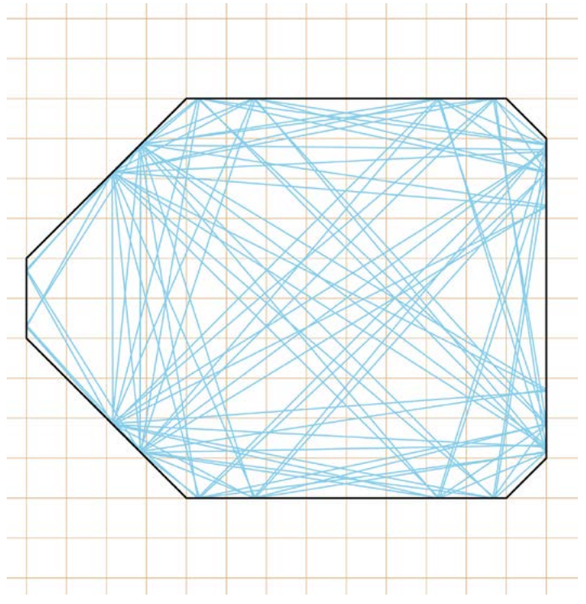
Seminars

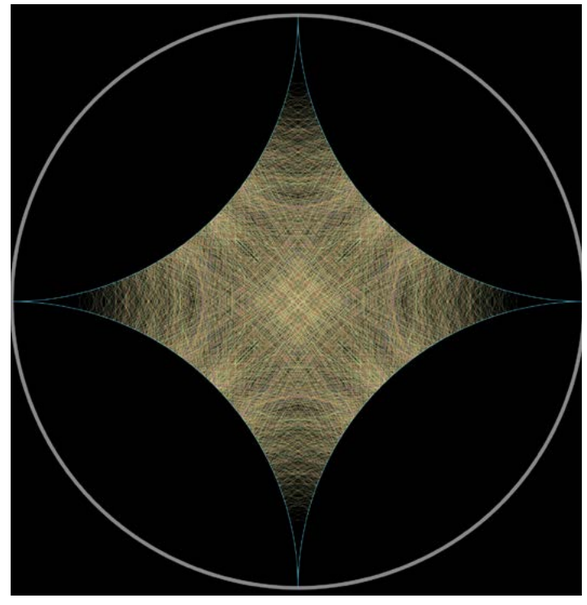
There are two weekly seminars: the HEGL Proseminar/Seminar (for students) and the HEGL community seminar (for everyone).

External activities and outreach

HEGL aims to increase its outreach to a wider audience via various external activities and programs, starting with:

- Creating a “Sierpinski Christmas tree” (see Figure 36)
- Participating in Girls’ Day
- Offering the Geometric Deep Learning Hackathon (Feb. 2022).





Figures 37a and 37b: Periodic orbits in a symplectic billiard and in an ideal hyperbolic billiard.

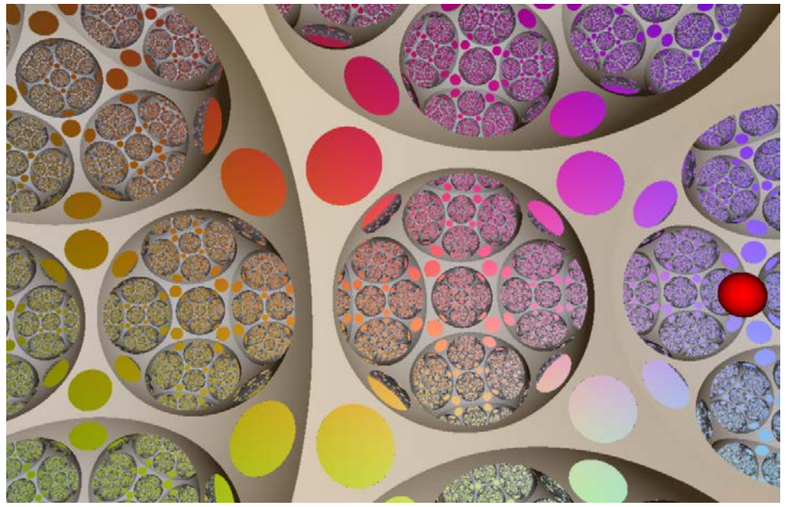
Math visualization

Computers and other technology offer fantastic tools for exploring mathematical objects, especially (but not only) in geometry. Below, we showcase a few examples that were developed at HEGL in 2021 [

Non-Euclidean billiards

Mathematical billiards are famous dynamical systems: A ball bounces off the walls of a billiard, which can have any shape. They are also quite complex: It is not easy to answer questions such as "Are there any periodic trajectories?" At HEGL, two student projects studied this question for symplectic and hyperbolic billiards, which are variations of standard (Euclidean) billiards. Both projects were able to produce beautiful computer-generated pictures (see Figure 37, previous page).

Virtual Reality (VR)

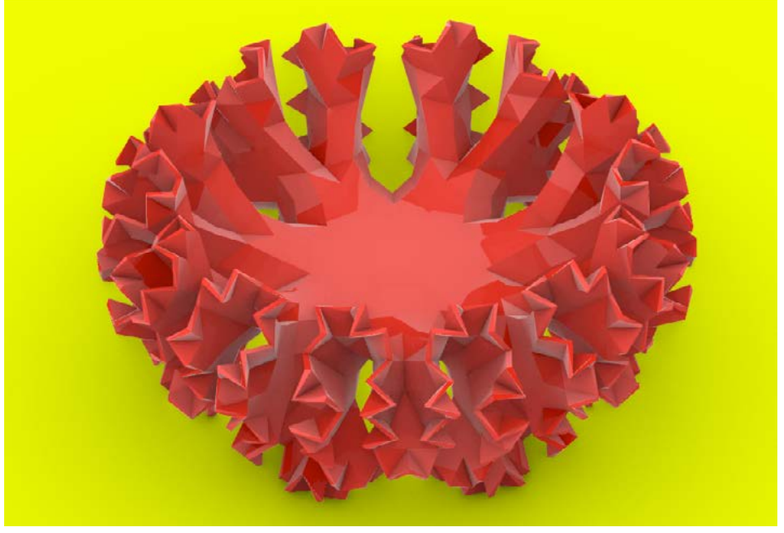


Figures 38a and 38b: Navigating in virtual reality.

Virtual reality is a fantastic technology for apprehending non-Euclidean geometries via an immersive experience. HEGL offers high-resolution VR equipment (see Figures 38a and 38b).

3D Printing

3D printers are becoming increasingly capable of producing complex shapes that can help us understand intricate mathematical objects. Figures 39a and 39b show a 3D model and 3D prints of hyperbolic corals, which represent a partial embedding of the hyperbolic plane in 3-dimensional space.



Figures 39a and 39b: A 3D model and two 3D prints of hyperbolic corals.

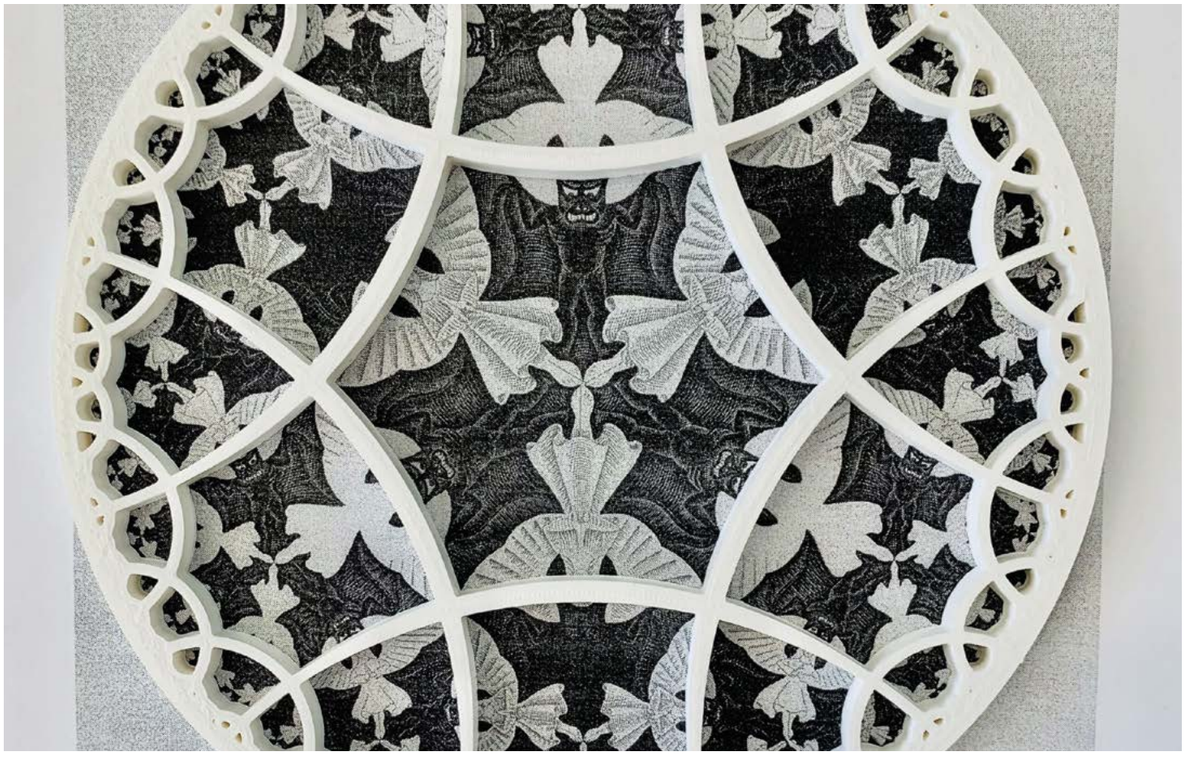
Student projects

One of the main benefits of the experimental approach to mathematics is to show that original research is possible at all levels. It is also a fun way for students to learn mathematics or, at the very least, to complement their theoretical courses. Below, we showcase three student projects that were conducted at HEGL in the 2021–2022 winter semester. More details can be found on HEGL's website: https://hegl.mathi.uni-heidelberg.de/blog/.

3D Printing Escher

M. C. Escher was a Dutch artist famous for his mathematical artwork. Some of his most famous pieces are based on tessellations of the hyperbolic plane. The goal of this student project was to recreate 3D prints of such tessellations (see Figures 40a and 40b).

Figures 40a and 40b: 3D prints of hyperbolic tessellations.



Gromov Hyperbolicity of the Poincaré Disk
Gromov introduced a very general notion of hyperbolicity: A triangle is called δ-slim if each of its sides is contained in the δ-neighborhood of the two others, and a space is δ-hyperbolic if all its triangles are δ-slim. This notion proved to be very effective in geometric group theory. In this project, the students checked experimentally that the Poincaré disk model of the hyperbolic plane is δ-hyperbolic for δ = 0.88137... (see Figure 41).

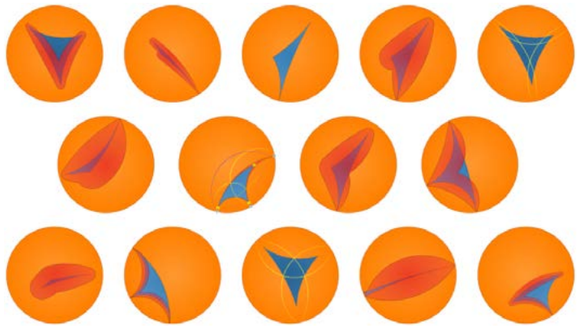


Figure 41: δ-slim triangles in the Poincaré disk.

Hyperbolic Ping-Pong
In this project, the students programmed an interactive computer game for two players to play ping-pong in the hyperbolic plane, see Figure 42.

Learn more

We invite curious readers to visit HEGL's website to learn more about current news and events, ongoing projects, galleries, blog posts, web apps, outreach, and beyond! HEGL is just getting started.

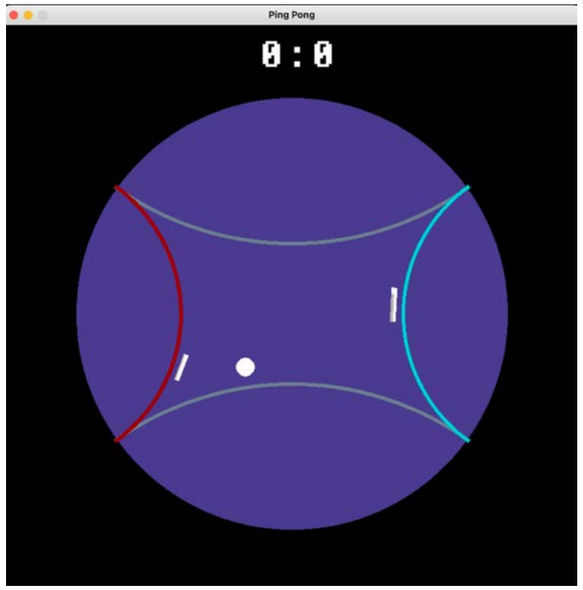


Figure 42: Playing ping-pong in the hyperbolic plane.

Die Arbeitsgruppe **Gruppen und Geometrie (GRG)** arbeitet eng mit der Research Station „Geometry & Dynamics" an der Universität Heidelberg zusammen. Beide Arbeitsgruppen werden von Anna Wienhard geleitet.

Symmetrien spielen eine zentrale Rolle in der Mathematik als auch in vielen Naturwissenschaften. In der Mathematik verstehen wir unter Symmetrien die Transformationen eines Objektes, die dieses invariant lassen. Solche Transformationen lassen sich verknüpfen, d.h. hintereinander ausführen und bilden so die mathematische Struktur einer, so genannte, Gruppe. Im 19. Jh. entwickelte der Mathematiker Felix Klein einen neuen Begriff der Geometrie: Geometrie ist das Studium der Eigenschaften eines Raumes, die invariant sind unter einer gegebenen Gruppe von Transformationen. Kurz gesagt: Geometrie ist Symmetrie.

Mit diesem Konzept vereinheitlichte Klein die klassische Euklidische Geometrie, die damals gerade neu entdeckte hyperbolische Geometrie als auch die projektive Geometrie, die aus dem Studium der perspektivischen Kunst erwuchs und die nicht auf dem Messen von Abständen, sondern auf Inzidenzrelationen beruht. Noch wichtiger ist, dass Felix Kleins Konzept unser Verständnis von Geometrie in der Mathematik und der theoretischen Physik grundlegend verändert hat und bis heute prägt.

Unsere Arbeitsgruppe beschäftigt sich mit verschiedenen mathematischen Forschungsfragen auf dem Gebiet der Geometrie, Topologie und der dynamischen Systeme, die das Zusammenspiel zwischen Räumen, wie zum Beispiel Mannigfaltigkeiten und metrische Räumen, und Gruppen, die als Symmetrien auf diese Räume wirken, einbeziehen. Außerdem wir beschäftigen uns mit den Anwendungen der Gruppentheorie und Geometrie in andere Disziplinen wie mathematische Physik, Datenwissenschaft und maschinelles Rechnen.

2.7 Molecular Biomechanics (MBM)

Group leader
Prof. Dr. Frauke Gräter

Staff members
Dr. Camilo Aponte-Santamaria (staff scientist)
Florian Franz (until September 2021)
Isabel Martin
Dr. Nicholas Michelarakis
Kai Riedmiller (HITS Lab)
Christopher Zapp (until May 2021)

Visiting scientists:
Saber Boushehri (KIT Karlsruhe)
Matthias Brosz (Heidelberg University)
Jannik Buhr (Heidelberg University; since June 2021)
Svenja de Buhr (Heidelberg University)
Eric Hartmann (Heidelberg University; since September 2021)
Dr. Fan Jin (Heidelberg University)
Dr. Markus Kurth (Heidelberg University)
Fabian Kutzki (Heidelberg University)
Benedikt Rennekamp (Heidelberg University)

Students
Elizaveta Bobkova (January–September 2021)
Thomas Ehret (February–April 2021)
Christoph Karfusehr (May–December 2021)
Rita Roessner (July–August 2021)
Rajlaxmi Saha (October–December 2021)
Salome Steinke (since November 2021)
Wojtek Treyde (since September 2021)
Aysecan Ünal (April–October 2021)
Dennis Wagner (until April 2021)

Life at the microscopic scale is in some respects not very different from life in the macroscopic world. Indeed, a polymer- or protein chain elongates and finally ruptures when pulled from its ends, just like a rubber band. Polymer chains align and assemble in a sheared solution, just like macroscopic objects. One difference, however, is that molecules are tiny and complex. Mechanical force can lead to specific changes in the shape of a protein, to the complete unfolding of its chain, or even to the rupture of chemical bonds. In this way, force can selectively switch enzymes on and off or cause biochemical reactions. Understanding the response of molecules such as proteins or polymers to stretching- or shearing forces requires computational models and simulations across time- and length scales in combination with experiments that test these models.

The aim of the Molecular Biomechanics group is to decipher the consequences of mechanical force on the dynamics, reactivity, and function of proteins, protein materials, and synthetic polymers. To that end, we use and further develop various computational techniques ranging from quantum chemical calculation to Molecular Dynamics simulations. In March 2021, we launched our own small wet lab at the Center of Advanced Materials at Heidelberg University. We are very excited about this new endeavor, which involves experiments that test our computational predictions, and we expect to have our first results in the coming year.

Proteins of particular interest to us in 2021 were – inter alia – the mechano-sensing adaptor protein Integrin-linked Kinase in focal adhesions and the large, disordered protein INCENP, which is involved in cell division. We were able to shed light on the role of these two important biological players as molecular connectors within cells and on how this function is regulated by mechanical force or other modifications. Understanding how force modulates structures is critical not only in the study of biological systems, but also in technical applications. Through our long-standing collaboration with Toyota Europe, we were able to delve into the self-assembly of polymers such as Nafion using multi-scale simulations. According to our study, the forces present in flowing polymer solutions cause the polymer chains to become ordered into lamellae that influence the final function of the fuel-cell membrane.

Another highlight of 2021 was the launch of our webservice, Colbuilder. The Colbuilder server offers access to a large set of atomistic models of collagen – the major force-bearing protein in our bodies – and will aid in future studies of collagen function in health and disease. Colbuilder therefore also provides a structural basis for our ERC grant, RADICOL, which was also kick-started in 2021 and allows us to more intensively advance our work on collagen radicals and on their role in ageing.

A small molecule with big effects: ATP stabilizes integrin-linked kinase to ensure cellular-force generation

Isabel Martin, Michele Nava, Sara Wickström, and Frauke Gräter

Cells respond to a broad range of signals of biochemical and mechanical nature from their surrounding environment, including the extracellular matrix to which the cells adhere. Signal transduction between the cell and the matrix is mediated by large protein complexes, which are named focal adhesions. These focal adhesions are essential to maintaining cellular homeostasis by regulating a broad range of cellular processes, such as adhesion, migration, growth, and differentiation. One crucial protein of focal adhesions is integrin-linked kinase (ILK). ILK is actually a pseudokinase, which implies that it is not capable of biochemical catalysis. The precise molecular functions – and especially the mechanosensory properties – of ILK remain unclear. Interestingly, ILK retains its ability to bind ATP even though it does not phosphorylate substrates; however, the significance of retained ATP binding for ILK function has not been well studied.

By combining molecular dynamics simulations with cell biology and traction force microscopy, we investigated the role of retained ATP binding in terms of ILK's function as a mechanosensor. We employed molecular dynamics simulations of human ILK in complex with its obligatory binding partner – parvin – and probed the kinase dynamics, which had been altered as a result of ATP removal. We found that ATP promotes the structural stability of ILK. Furthermore, using force distribution analysis, we determined that ATP allosterically strengthens the interaction of ILK with parvin over two key – and previously unrecognized – saltbridges between ILK and parvin. Moving beyond the time- and length scales of the simulations, we also studied the large-scale cellular effects of retained ATP binding to ILK in collaboration with Sara Wickström's group at the University of Helsinki. Experimental disruption of this ATP-dependent allosteric network via mutation of the saltbridges disrupts ILK:parvin binding and impedes accurate focal adhesion turnover and especially disassembly.

In addition to the stabilizing effect of ATP on the pseudokinase in equilibrium

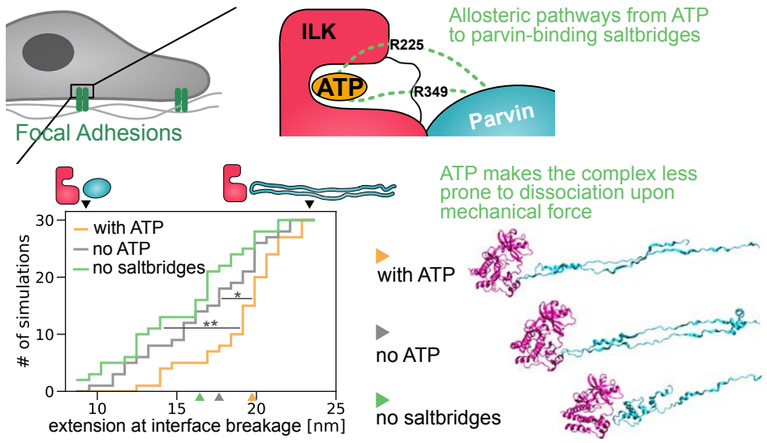


Figure 43: Upper panel: At focal adhesions, integrin-linked kinase (ILK) is structurally stabilized by ATP through an allosteric pathway that involves parvin-binding saltbridges. Lower panel: Under mechanical force, ATP makes the ILK:parvin complex less prone to dissociation, as seen from the cumulative histogram of the extension of the complex at the time of interface rupture.

conditions, we also observed that ATP conveys mechanical stabilization of the ILK:parvin complex under mechanical force. Using force-probe molecular dynamics simulations, we discovered that under tension, ATP shifts the balance from the rupture of the complex to protein unfolding. This stabilization of the ILK:parvin interface is at least partly mediated by ATP-influenced saltbridges. Experimental quantification of cellular traction stresses revealed that disruption of ATP binding or of ATP-dependent saltbridges in ILK indeed leads to reduced force generation and to cell migration. This finding supports the notion that on the cellular level, the stability of the ILK:parvin interaction determines the ability of focal adhesions to bear loads.

Thus, we propose that ILK functions as a pseudo-mechanokinase, with ATP serving a mechanotransducing role as an obligatory binding partner for the structural and mechanical integrity of the ILK:parvin complex in order to ensure proper force generation and cell migration. Furthermore, our study generally establishes a new role for retained ATP binding to pseudokinases and thereby contributes to our understanding of pseudokinase function and evolution. In general, mechanotransducing (pseudo)kinases such as ILK are uniquely suited to integrating mechanical signals into other signaling pathways, and more members of this intriguing class of proteins likely remain to be uncovered [Martin et al, 2021], Figure 43.

Shifting the cloud: Disordered proteins and their response to phosphorylation

Isabel Martin, Fan Jin, Camilo Aponte-Santamaria, and Frauke Gräter

Intrinsically disordered proteins (IDPs) are a class of proteins that lack both secondary and tertiary structures and instead explore a broad conformational ensemble. Their functions – which range from transcriptional regulation to signal transmission – are tightly regulated. IDPs are subject to extensive reversible post-translational modifications (PTMs), such as phosphorylation, methylation, and glycosylation. Among these PTMs, phosphorylation is one of the most common and important. However, the mechanism by which phosphorylation affects the conformations and functions of IDPs remains unclear. To solve this issue, we performed extensive all-atom molecular dynamics simulations for four representative IDPs – Ash1, E-Cadherin, CTD2', and p130Cas – in their unphosphorylated and phosphorylated forms. Our results revealed that all IDPs undergo a mild change upon multi-site phosphorylation, which is V-shaped when monitoring the effect of phosphorylation along different protein net charges: Phosphorylation moderately expands neutral or overall negatively charged IDPs and shrinks positively charged IDPs. Thus, electrostatic interactions often dominate and can lead to a shift in the conformational cloud of such disordered proteins. More importantly, in two of these IDPs, only two biologically relevant phosphorylation sites are needed to expose the adjacent negatively charged active site to the environment significantly more, which implies a higher probability of interacting with other binding partners (see [Jin et al, 2021]).

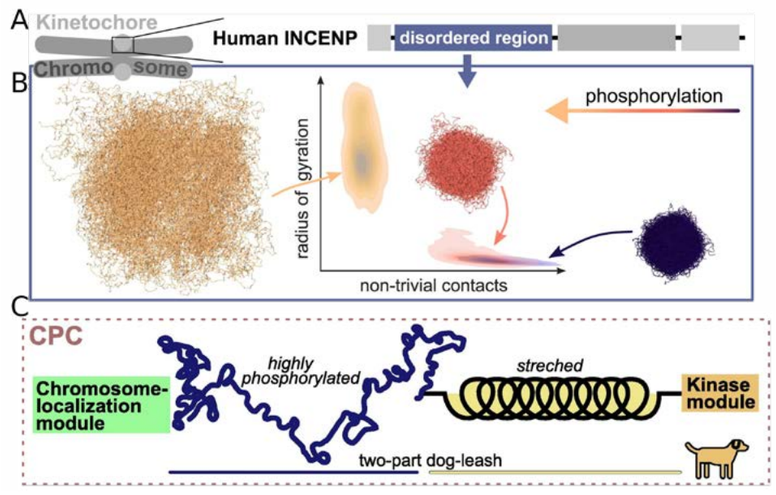


Figure 44: Phosphorylation tunes the elongation of the intrinsically disordered region of the inner-centromere protein (INCENP). A. INCENP is located at the interface between chromosomes and microtubules (i.e., the kinetochores) and contains an intrinsically disordered region of about 440 amino acids. B. Driven by the effect of phosphorylation, INCENP's disordered region transitions between collapsed conformations (i.e., a small radius of gyration and many internal non-trivial contacts) and a stretched conformation (i.e., a high radius of gyration and few contacts). C. INCENP is proposed to take part in the dog leash that connects the chromosome-localization module and the kinase modules, thereby allowing the latter to reach its targets on the kinetochore side.

Mitosis is an example of when phosphorylation plays an important role in regulating an IDP, and we embarked on deciphering this role via simulations. Mitosis is a key process in biology in which cells divide to form new cells. During this process, chromosomes are positioned in pairs at junctions with microtubules and are stabilized by large protein assemblies called kinetochores. Proper positioning of the chromosomes at these junctions is carefully fined-tuned, and chromosomes that are not properly positioned must be disengaged. An essential player in this control mechanism is the inner-centromere protein (INCENP), which is a large protein that contains a 440-amino-acid-long highly phosphorylated and intrinsically disordered region (Figure 44A). By using computer simulations both on the atomic scale and on a coarser level, we demonstrated that phosphorylation modifies the size of this disordered region, thereby switching it from a collapsed conformation to a stretched conformation (Figure 44BC). Our results suggest that the disordered

In so mancher Hinsicht unterscheidet sich das Leben auf der mikroskopischen Skala nur wenig vom Leben der makroskopischen Welt. Eine Polymer- oder Proteinkette verlängert sich und reißt schließlich, wenn sie an ihren Enden gezogen wird, genau wie ein Gummiband. Polymerketten richten sich in einer gescherten Lösung aus und fügen sich zusammen, ähnlich wie makroskopische Objekte. Ein wichtiger Unterschied ist natürlich, dass Moleküle winzig klein und komplex sind. Mechanische Krafteinwirkung kann zu gezielten Formänderungen eines Proteins oder zur vollständigen Entfaltung seiner Polymerkette bis hin zum Brechen chemischer Bindungen führen. Auf diese Weise kann Kraft gezielt Enzyme an- und ausschalten oder biochemische Reaktionen hervorrufen. Um die Reaktion von Molekülen auf Dehnungs- oder Scherkräfte zu verstehen, sind Computermodelle und Simulationen über Zeit- und Längenskalen hinweg erforderlich.

Ziel unserer Arbeitsgruppe **Molecular Biomechanics (MBM)** ist es, die Folgen mechanischer Krafteinwirkung auf Dynamik, Reaktivität und Funktion von Proteinen, Proteinmaterialien oder synthetischen Polymeren zu entschlüsseln. Dazu nutzen und entwickeln wir verschiedene Computertechniken, von quantenchemischen Berechnungen bis hin zu molekulardynamischen Simulationen. Seit März 2021 haben wir am Center of Advanced Materials der Universität Heidelberg ein eigenes, kleines Wet-Lab eröffnet und freuen uns sehr über diesen neuen experimentellen Ansatz, unsere rechnerischen Vorhersagen zu testen. Erste Ergebnisse hierzu erwarten wir im kommenden Jahr 2022.

Im Jahr 2021 waren für uns unter anderem folgende Proteine von besonderem Interesse: die auf mechanische Einflüsse reagierende Fokale Adhäsionskinase, die als Adapterprotein fungiert, und das große, ungeordnete Protein INCENP, das an der Zellteilung beteiligt ist. Wir konnten Aufschluss darüber geben, welche Rolle diese beiden wichtigen biologischen Akteure als molekulare Verbindungsglieder innerhalb von Zellen spielen und wie diese Funktion durch mechanische Kraft oder andere Modifikationen reguliert wird. Wie die Einwirkung von Kraft die Struktur verändert, ist nicht nur für biologische Systeme, sondern auch für technische molekulare Systeme entscheidend. Im Rahmen unserer langjährigen Zusammenarbeit mit Toyota Europe haben wir uns in Mehrskalensimulationen mit der Selbstorganisation von Polymeren wie Nafion beschäftigt. Die Kräfte in fließenden Polymerlösungen bewirken eine Anordnung der Polymerketten in Schichten, ein Ergebnis, das helfen kann, die Funktion von Brennstoffzellmembran zu verbessern.

Ein weiteres Highlight 2021 war die Einführung unseres Webservices Colbuilder. Er bietet Zugang zu einem großen Satz atomistischer Modelle von Kollagen, dem wichtigsten krafttragenden Protein unseres Körpers, und soll zukünftige Studien zum Einfluss von Kollagen auf Gesundheit und Krankheit unterstützen. Der Webservice bildet somit auch eine strukturelle Basis für unseren ebenfalls 2021 gestarteten ERC Grant RADICOL und ermöglicht uns, unsere Arbeiten zu Kollagenradikalen und ihre Rolle beim Altern intensiver voranzutreiben.

region of INCENP acts as a phosphorylation-regulated and length-variable component within the previously defined "dog-leash" model. This region thereby regulates the kinase module and enables this module to reach its targets for proper chromosome segregation during mitosis (Figure 44C) (see Martin, Aponte-Santamaría, et al. "Phosphorylation tunes elongation propensity and cohesiveness of INCENP's intrinsically disordered region". J Mol Biol. 434: 167387 (in print)).

Build your collagen fibril!

Agnieszka Obarska-Kosinska, Benedikt Rennekamp, Aysecan Ünal, and Frauke Gräter

Type I collagen is the main structural component of many tissues in the human body. It provides excellent mechanical properties to connective tissue and acts as a protein interaction hub. There is thus wide interest in understanding the properties and diverse functions of type I collagen at the molecular level.

A major bottleneck in examining the structural changes of collagen under load stems from the fact that only low-resolution structures of larger collagen fibril assemblies are available. Molecular Dynamics (MD) simulations have therefore been restricted to small models, which lack many important structural features, such as helix–helix packing and the details of crosslinking. In this project, we embarked on realistic multi-million-atom MD simulations of collagen using the required atomistic structures, which we built in our group and to which we added all structural knowledge available to date. To that end, we built full-atom models of crosslinked collagen fibrils by integrating the low-resolution structure of collagen fibrils – available from X-ray fiber diffraction – with high-resolution structures of short collagen-like peptides – available from X-ray crystallography and mass spectrometry data.

The obtained models are made available through an interactive web server called ColBuilder, which allows users to choose from among 20 different species, 8 different crosslink types, and crosslink composition (i.e., there are individual choices among crosslink types at different collagen sites within the same model). On the one hand, users can download PDB files that contain models of the complete 300nm single triple helix. These files include the symmetry record that allows for collagen fibril bundles of any size to be generated. On the other hand, we also provide models of predefined bundles of collagen fibril fragments, which consist of 41 triple helices that span one central overlap and one gap region (the latter in two parts), as shown in Figure 45. These models of fibril bundles are ready for use in MD simulations, for example, using GROMACS. In order to easily enable simulations, we provide parameters of the modeled crosslinks for an AMBER force field.

Figure 45: Exemplary atomistic model of a collagen type-1 fibril fragment as obtained from ColBuilder. The model comprises 41 triple helices spanning one central overlap and one gap region (in two parts: left and right). Triple helices are colored according to the staggering distance, and crosslinks are shown in blue.

To validate our models, we performed MD simulations of three representative collagen species with both divalent and trivalent crosslinks – a set of models that we expected to be representative. Taken together, the atomistic collagen models dynamically explore a conformational space that is close to the X-ray fiber-diffraction data and also in line with other experimental observations (gap/overlap ratio and heavy-atom labeling).

The repository of collagen models is available at https://colbuilder.h-its.org. We expect the repository to facilitate the structural investigation of a wide range of collagen aspects, such as molecular interactions with other proteins, the localization of other crosslinks, and disease mutants. In the future, we will update — and invite the community to contribute to — the repository with new models and force fields (see [Obarska-Kosinska et al, 2021]).

Toward a Greener Future: Understanding Shear Thinning in Nafion

Nicholas Michelarakis, Florian Franz, Konstantinos Gkagkas, and Frauke Gräter

Fuel cells offer a promising energy-generation alternative for a wide range of technologies thanks to their unparalleled efficiency and ecological friendliness. Proton exchange membrane (PEM) fuel cells are electrochemical cells that convert chemical energy to electrical energy. In these cells, hydrogen is fed into an anode, where it is oxidized. The protons generated by this process are conducted through an electronic non-conductive (proton-exchange) membrane to the cathode, and the electrons are re-routed through a bypass, thereby generating an electric current. The only byproducts generated through this process are water and heat.

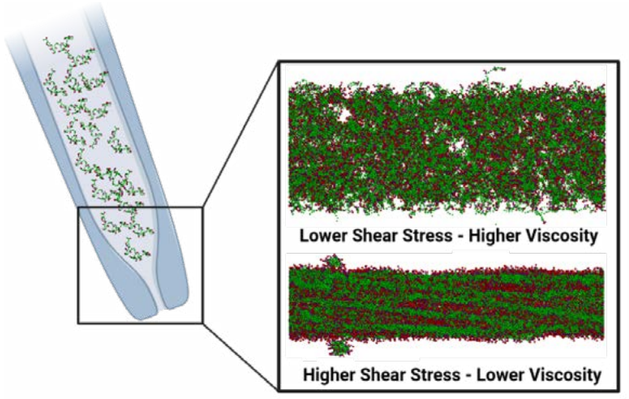


Figure 46: New insights into the mechanism governing the shear-thinning effects observed in Nafion solutions through the use of non-equilibrium coarse-grained molecular dynamic simulations. Higher shear stress leads to the formation of band-like Nafion structures that are oriented along the flow direction, which leads to a decrease in resistance to the flow and thereby results in lower shear viscosity (shear thinning). Lower shear stresses are not strong enough to break the interconnected structures of the Nafion strands, thereby resulting in a higher shear viscosity.

At the heart of a fuel cell lies the PEM, which is the cell's most important component and the area in which the proton transfer occurs. Currently, creating PEMs involves a variety of ink-based manufacturing processes by which the PEM solution (ink) is deposited (printed) on a suitable surface. The printing process often involves shearing the material. PEM inks typically consist of a variety of solvents and other materials, with perfluorosulfonic-acid-based ionomers – and more specifically, Nafion – being the most commonly used. The basic rheological properties of the PEM inks – such as their shear viscosity – notably affect the printing process and thus also the properties of the resulting PEM and the performance of the fuel cell. It therefore comes as no surprise that an in-depth understanding of Nafion PEM inks is of paramount importance to the commercialization process of PEMFCs. Nafion inks are known to exhibit a shear-thinning effect by which their shear viscosity is reduced as shear stress is applied; however, a direct relationship between ink viscosity and the underlying molecular structure of the Nafion solution has remained elusive. In our work (Michelarakis et al. 2021), we employ coarse-grained molecular dynamics to shed light on the connection between nano-scale ionomer interactions and their effect on the macro-scale rheological properties of Nafion inks. We simulate a Couette flow, which is analogous to the flow observed in the ink-jet printers employed in the manufacturing of PEM membranes. With our simulation setup, we are able to recover the shear-thinning effects of Nafion solutions by observing reduced shear viscosity when we increase the shear stress applied by the flow. We attribute this effect to the transformation of the rod-like structures of Nafion under equilibrium into band-like structures that are oriented and aligned along the flow direction. This leads to a decrease in resistance to shearing stresses, which results in a decrease in shear viscosity (see Figure 46).

Our study uncovered the molecular mechanism behind shear thinning in Nafion and offers new insights into the interplay between the structural and rheological properties of Nafion structures in solution under shear flow. Furthermore, the study paves the way for other large-scale computational studies with the aim of informing the manufacturing processes of PEM membranes (see Michelarakis N, Franz F, Gkagkas K, Gräter F. Longitudinal strand ordering leads to shear thinning in Nafion. Phys Chem Chem Phys. 2021 Nov;23(45) 25901-25910. doi:10.1039/d1cp02024b.).

2 Research

2.8 Molecular and Cellular Modeling (MCM)

Group leader
Prof. Dr. Rebecca Wade

Staff members
Christina Athanasiou
Dr. Giulia D'Arrigo (since April 2021)
Madhura De (April–December 2021)
Manuel Glaser
Dr. Daria Kokh (until January 2021)
Dr. Goutam Mukherjee (March–October 2021)
Abraham Muniz Chicharro
Dr. Giulia Paiardi (since April 2021)
Dr. Stefan Richter
Alexandros Tsengenes

Visiting scientists
Madhura De (DKFZ; until March 2021)
Marco Destro (January–February 2021)
Paolo Giacco (January–February 2021)
Anna Mazurek (May–July 2021)
Marcel Meyer (Heidelberg University; since October 2021)
Dr. Goutam Mukherjee (Heidelberg University; until February 2021)
Dr. Ariane Nunes-Alves (Capes-Humboldt fellowship; until June 2021)

Students
Ainara Claveras Cabezudo (since September 2021)
Sara Dakovic (July–August 2021)
Pierre De Marinis (April–July 2021)
Sungho Bosco Han (until April 2021)
Konstantinos Mavridakis (until October 2021)
Lena Meßner (September–December 2021)
Luise Nottmeyer (January–March 2021)
Rafael Salazar Claros (July–August 2021)
Jonathan Teuffel (since June 2021)

Molecular recognition, binding, and catalysis are fundamental processes to cell function. The ability to understand how macromolecules interact with their binding partners and participate in complex cellular networks is critical both to predicting macromolecular function and to applications such as protein engineering, systems biology, 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? 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? One of our aims is to gain a mechanistic molecular-level understanding of drug interactions along the process that extends from drug delivery to drug–target binding and further to drug metabolism.
We take an interdisciplinary approach that entails collaboration with experimentalists and that makes concerted use of computational approaches based on physics and bio-/chemo-informatics. The broad spectrum of techniques developed and employed ranges from interactive, web-based visualization tools and machine-learning methods to atomic-detail molecular simulations.
In this report, we outline some of the results achieved in 2021. Following a general overview of what was new in the group last year, we describe projects on (i) drug–target binding kinetics, (ii) macromolecular complexes, and (iii) structure-based drug design against SARS-CoV-2.

What was new in 2021?

In its second year, the coronavirus pandemic continued to challenge us to rethink how we work. We organized Biological Diffusion and Brownian Dynamics Brainstorm 5 (BDBDB5: https://bdbdb.h-its.org/) in March (see Chapter 5.1.1). This fifth installment was a shorter, fully online event that was very well attended and allowed us to connect with scientists around the world.
During the year, we launched three new SARS-CoV-2-related research activities. Within the EU-funded Human Brain Project (HBP), we are a partner in a new project, BRAVE, which was selected in a competitive call on "COVID-19 and its impact on the brain and mental health" aimed at encouraging the use of the HBP and EBRAINS infrastructure for COVID-19-related research. BRAVE began in December 2021 and aims at tackling COVID-19 brain inflammation via the computer-aided design of anti-inflammatory compounds. The project is coordinated by Francesca Spyrakis, University of Turin (Italy), and involves other collaborators at the University of Pavia (Italy) and the Forschungszentrum Jülich (Jülich Research Centre) (Germany). In 2021, we were also awarded funding from the German Research Foundation's (DFG) COVID-19 Focus Funding scheme for our project "meCocan – Towards a mechanistic understanding of the interaction of SARS-CoV-2 spike glycoprotein and host heparan sulphate proteoglycans." After having visited the MCM group in 2020, Giulia Paiardi completed her doctorate in Italy and then returned to HITS in April to work on this project, which is being pursued in collaboration with Marco Rusnati at the University of Brescia (Italy). In December, after many delays due to pandemic travel restrictions, Matheus Ferraz joined us from Roberto Lins's laboratory at the Federal University of Pernambuco, Recife (Brazil), for a 1-year stay as a CAPES-DAAD pre-doctoral scholar to work on applying computational techniques developed at HITS to the design of nanobodies for diagnosing and treating COVID-19.

SIMPLAIX – a new 3-way collaboration between HITS, Heidelberg University, and KIT – got underway towards the end of 2021, with Rebecca Wade (MCM) and Frauke Gräter (MBM) serving as spokespersons (see also Chapters 1 and 6). The scientific focus of SIMPLAIX is on bridging scales from molecules to molecular materials via multiscale simulation and machine learning. SIMPLAIX is funded by the Klaus Tschira Foundation and supported in kind by the two universities. The MCM group participates in two SIMPLAIX projects: One is a project on the use of invertible neural networks for predicting molecular interactions, on which Marcel Meyer began work with Ullrich Köthe (IWR, Heidelberg University), and the other will get underway in 2022 and involves a collaboration with Marcus Elstner (KIT) on a combined quantum mechanics and molecular mechanics study of inter-protein electron transfer.

In September, Madhura De defended her doctorate on the effects of linker DNA on linker-histone positioning on mononucleosomes and trinucleosomes, which she studied by carrying out biophysical experiments with Katalin Toth (DKFZ, Heidelberg) as well as computational modeling at HITS. Madhura will begin a postdoc at Harvard Medical School in March 2022. Sungho Bosco Han completed his master's thesis in Molecular Biosciences and has gone on to doctoral studies in Manchester (UK). Pierre De Marinis and Lena Messner completed their bachelor's theses in Molecular Biotechnology. Luise Nottmeyer (Molecular Biotechnology), Sara Dakovic, and Rafael Salazar Claros (Molecular Biosciences) completed internships in the group as part of their master's studies at Heidelberg University.

In 2021, there were also comings and goings at the postdoctoral level. Giulia D'Arrigo joined us from Turin to work on the HBP. After more than twelve very fruitful years at HITS, Daria Kokh left the Institute to take up a new position in health-data science. Goutam Mukherjee also left to take up a senior-scientist position at PharmCADD in Busan (South Korea). Following the completion of her CAPES-Humboldt postdoc scholarship, Ariane Nunes-Alves moved to the Institute of Chemistry at the Technical University of Berlin to set up a junior research group within the UniSysCat Cluster of Excellence.

Drug–target binding kinetics

The rates of association and dissociation of drug–target complexes are being increasingly recognized to have an influence on drug efficacy. Therefore, computational methods for predicting drug–target binding kinetics are required to guide the optimization of the kinetics during the drug-design process. In 2021, we reported advances in methods of computing drug–target binding kinetics as well as applications of these methods that provided insights into the determinants of drug-target binding kinetics.

We also developed a multiscale approach for computing the association rate constant for a drug-like compound that binds to a gated protein-binding site [Sadiq, 2021] (see Figure 48). The first step is to perform molecular dynamics simulations of the unbound protein and to analyze the trajectories in order to identify the main conformations adopted, their population and ligand accessibility, and their interchange kinetics. In the next step, Brownian dynamics simulations of the diffusion of the molecules to open protein structures are performed to compute association rates, which are then adjusted by the gating factor computed for the unbound protein. Applying this approach to the binding of inhibitors to HIV-1 protease revealed that conformational gating only modestly slows inhibitor binding despite the presence of two flaps that close over the binding site and thus also indicated that other factors may limit association rates. For systems modulated by conformational gating, the approach can be scaled computationally efficiently to screen association kinetics for many ligands.

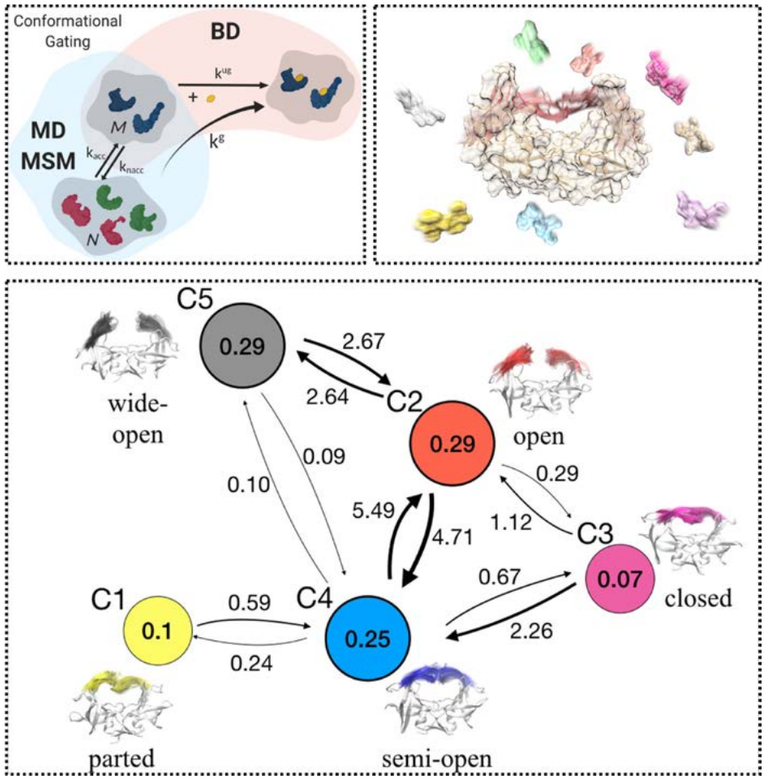


Figure 48: A multiscale approach for computing the association rate constant for gated protein–ligand binding from molecular dynamics- and Brownian dynamics simulations. The dynamics of the unbound protein are first characterized by molecular dynamics (MD) sampling and the computation of a Markov State Model (MSM). In this work, five states of the HIV-1 protease with different positions of the flaps over the ligand-binding site were identified, and the rates of interchange between the states as well as a gating factor for closing the binding site to drug-like molecules were computed. In a second step, Brownian dynamics (BD) simulations of the diffusional association of the drugs to the open conformation of the protein are performed to compute the diffusional association rate constant. This rate constant is then multiplied by the gating factor. Adapted from [Sadiq, 2021], with permission from from ©2021 American Chemical Society.

We previously developed the τRAMD (tau-Random Acceleration Molecular Dynamics) procedure to compute relative drug–target dissociation rates (and residence times **(τ);τ** is inversely related to the dissociation rate). We applied τRAMD to various soluble drug targets (see, e.g., [Berger, 2021]). Together with Marc Bianciotto and colleagues at Sanofi R&D (France), we assessed the scaled molecular dynamics and τRAMD methods to estimate residence times and introduced new techniques for analyzing simulated trajectories. The results indicate how these approaches could be used for lead optimization or virtual screening in a drug-design project [Bianciotto, 2021]. We also assessed the performance of τRAMD for a system that has been exceptionally well characterized by biophysical experiments and previously studied using many simulation techniques by applying τRAMD to the dissociation of small molecules from artificial cavities in mutants of T4 lysozyme [Nunes-Alves, 2021] (Figure 50). The RAMD simulations efficiently provided computed residence times that were in good agreement with experiment. Results revealed ligand unbinding routes consistent with previous studies and provided mechanistic insights: For example, the presence of a greater number of metastable states along egress pathways was found to lead to slower protein–ligand dissociation.

Many drug targets are membrane-bound proteins, and we therefore

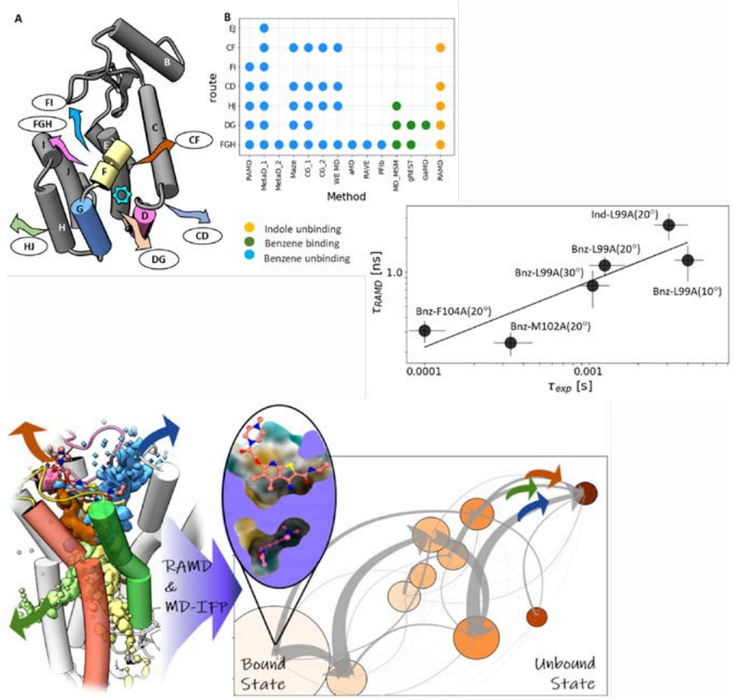
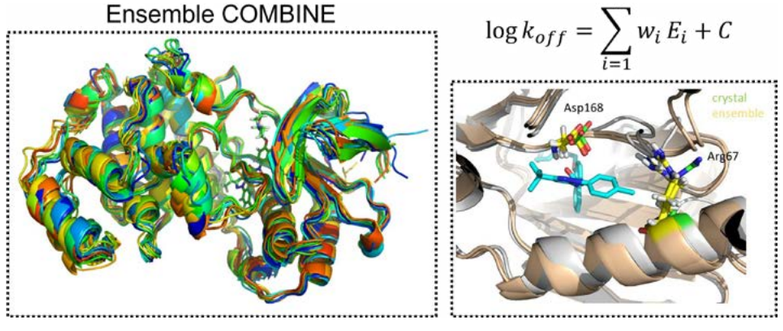


Figure 49: Validation and application of τRAMD- and MD-IFP methods for computing dissociation rates. Upper: RAMD simulations reveal unbinding mechanisms and dissociation rates for compounds from small cavities in T4 lysozyme mutants. Egress routes observed for benzene and indole on dissociation from the L99A mutant are indicated by arrows, and the plot shows which routes are detected by the computational methods that have been applied to this system, including RAMD. Not only does RAMD provide efficient sampling of egress routes, but it also provides relative residence times that are in good agreement with experimental measurements performed at three temperatures for different compounds and mutants. From [Nunes-Alves, 2021]. Lower: For two transmembrane G protein-coupled receptor proteins that are well-established drug targets, τRAMD- and MD-IFP analysis yielded insights into the effects of the binding of a ligand at a surface-exposed allosteric binding site on the rate and mechanism of dissociation of a drug that binds at the orthosteric site in the protein interior. The three main routes taken by the drug from the orthosteric site are shown in 3D along with a map of the intermediate states (defined by analyzing ligand–protein contacts) along these routes (blue, red, and green arrows). Reproduced from [Kokh, 2021], with permission from ©2021 American Chemical Society.

applied the τRAMD method to G protein-coupled receptors, which are important drug targets that are transmembrane proteins. We found that τRAMD reproduced relative residence times and – when used together with MD-IFP (molecular dynamics-interaction fingerprint) analysis – revealed the dissociation mechanisms and the effects of allosteric modulation for two G protein-coupled receptors [Kokh, 2021] (see Figure 49).

Comparative Binding Energy (COMBINE) analysis can be used to derive protein-structure-based quantitative structure-kinetic relationships (QSKR) (for a practical description, see [Ganotra, 2021]). We extended the method to account for the flexibility of proteins by considering an ensemble of structures instead of one structure for each protein–ligand complex [Nunes-Alves, 2021]. For the datasets studied thus far, considering an ensemble of structures did not improve predictive ability but did enable further protein–ligand contacts that affect residence times to be identified.



In vivo, drugs must diffuse through the crowded cellular environment in order to reach their specific targets. We are currently investigating how the presence of proteins and confining surfaces can affect the diffusion of drug molecules and thus also their ability to interact with their targets. For this purpose, we are studying defined drug–protein crowder combinations via simulation while our collaborators – Debrata Dey and Gideon Schreiber at the Weizmann Institute (Israel) – study the same systems via FRAP (fluorescence recovery after photobleaching). We first performed Brownian-dynamics simulations with a simple spherical particle model to help validate the Line-FRAP methodology for measuring molecular diffusion rates [Dey, 2021]. We also improved our models for treating long-range electrostatics and hydrodynamic interactions in Brownian dynamics simulations of protein

Figure 50: Ensemble COMBINE analysis uses a set of protein structures generated via molecular dynamics simulation and docking to predict quantitative structure-kinetic relationships (QSKR). The approach was evaluated for its ability to predict the dissociation rates (koff) of a set of inhibitors of p38 mitogen-activated protein, which has a flexible binding site. By considering the flexibility of the protein–ligand complexes, we were able to identify additional ligand–protein interactions that lead to longer residence times, such as interactions with residues Arg67 and Asp168, which are close to the ligand (cyan) in many crystal structures. Adapted from [Nunes-Alves, 2021, with permission from ©2021 American Chemical Society.

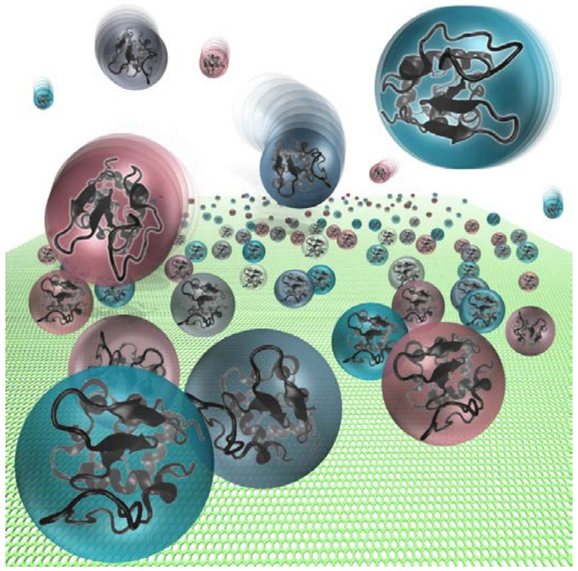


Figure 51: Illustration of the model for Brownian dynamics simulations of the diffusion of proteins in the presence of surfaces. We extended the model implemented in the Simulation-of-Diffusional-Association (SDA) software package to treat long-range electrostatic and mean-field hydrodynamic interactions between proteins and surfaces. The model combines the use of atomic-detail 3D structures of proteins and surfaces for computing shorter-range forces with the treatment of proteins as spheres and of surfaces as planes for computing longer-range forces. The model can be used to study the adsorption of proteins to surfaces. From [Reinhardt, 2021], with permission from ©2021 American Chemical Society.

diffusion in the presence of a surface [Reinhardt, 2021] (see Figure 51). By applying these computational and experimental methods to drug–protein crowder systems, we found that diffusive behavior differs according to the physicochemical properties of the different drugs and protein crowders [Dey, Nunes-Alves et al., submitted].

Macromolecular complexes

We apply modeling- and simulation approaches to a range of macromolecular complexes (see [Moreau, 2021], [Mukherjee, 2021], [Aslam, 2021], and [Rusnati,2021]). Below, we highlight two such applications.

A long-standing focus of the group has been on understanding how linker histone proteins bind to nucleosomes and how this binding affects the higher-order structuring of chromatin, which is key to the tight packing of DNA in cell nuclei and to the regulation of transcription processes. In 2021, we reported a study using single-pair FRET (Fluorescence/Förster Resonance Energy Transfer) spectroscopy – performed by Madhura De with Katalin Toth at DKFZ – and computational modeling to reveal how certain DNA sequences can orient the linker histone on the nucleosome [De, 2021]. In this work, a new mechanism by which linker histones can recognize stretches of contiguous adenines (so-called adenine- or A-tracts) on DNA using conserved arginine residues was identified. These results aid in our understanding of the linker histone's preferences for adenine tracts, which are abundant in certain parts of the genome – such as heterochromatin – and have been found in oncogenes. In subsequent work, we applied a similar approach to study how linker-histone orientation affects the structure of trinucleosomes [De, in preparation].

Within the Informatics4Life consortium, we collaborate with Patrick Most (Heidelberg University Hospital) to study the S100A1 protein. Its C-terminal peptide – S100A1ct – has been found to have potential in treating acute heart failure because it can improve the contractile performance of failing cardiac and skeletal muscle without arrhythmic side effects. However, its mechanism of action is unclear. In a study of the conformational landscape of S100A1ct in an aqueous environment, we found that the peptide can undergo conformational switching, which could be important for its positive inotropic action [Glaser, 2021] (see Figure 52). In ongoing studies, we are currently investigating the effects of mutations and the environment on the conformational and interaction properties of S100A1ct with the aim of understanding its mechanism of action and of providing a basis for designing peptidomimetics with improved properties.

Structure-based drug design against SARS-CoV-2

We are currently applying modeling- and simulation approaches to pursue several strategies for tackling the SARS-CoV-2 virus: virtual screening for small-molecule inhibitors of SARS-CoV-2 virus targets [Gossen, 2021], the design of protein inhibitors, and the design of polysaccharide heparin analogues. The latter approach is based on mechanistic insights gleaned from molecular dynamics simulations of SARS-CoV-2 spike glycoprotein–heparin complexes [Paiardi, 2022] (see Figure 53). These simulations – which were run using PRACE supercomputing resources – have revealed three ways that heparin can exert its antiviral activity on SARS-CoV-2 and have also provided insights into how host-cell heparan-sulfate proteoglycans may interact with SARS-CoV-2 and thereby affect host-cell infection and susceptibility. This research could impact the treatment of the viral infection because heparin is already used in treating COVID-19 patients due to its anticoagulant properties and because recent studies indicate the value of inhaled heparin for treating lung diseases.

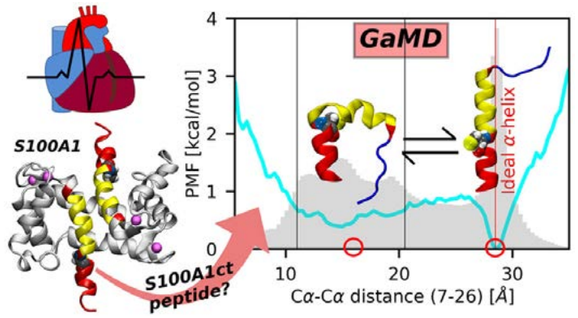
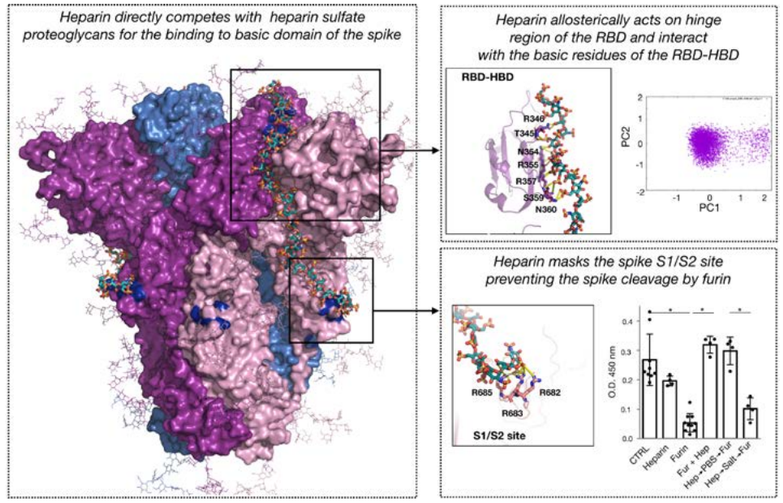


Figure 52: Molecular dynamics simulations of the positive inotropic peptide S100A1ct in an aqueous environment suggest that its activity may be related to its ability to act as a conformational switch. The peptide consists of the C-terminal 20 amino-acid residues (red, yellow) of the S100A1 protein fused to an N-terminal 6-residue hydrophilic tag (blue). To investigate the peptide's molecular mechanism of action, we studied its conformational landscape using conventional and Gaussian accelerated molecular dynamics (GaMD) simulation. Results revealed that although the helical conformation of the peptide – as observed in the S100A1 protein – is the most energetically stable, the peptide can adopt a range of kinked conformations. From [Glaser, 2021], with permission from ©2021 American Chemical Society.

Figure 53: Molecular dynamics simulations reveal three mechanisms by which heparin can exert its antiviral activity against SARS-CoV-2 by binding to the homotrimeric spike glycoprotein. Experiments (lower right) confirmed the ability of heparin to prevent spike cleavage by the furin protease. Adapted from [Paiardi, 2022] (Paiardi G, Richter S, Oreste P, Urbinati C, Rusnati M, Wade RC (2022). The binding of heparin to spike glycoprotein inhibits SARS-CoV-2 infection by three mechanisms, Journal of Biological Chemistry 298(2):101507).



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, Systembiologie und strukturbasierte Wirkstoffentwicklung.

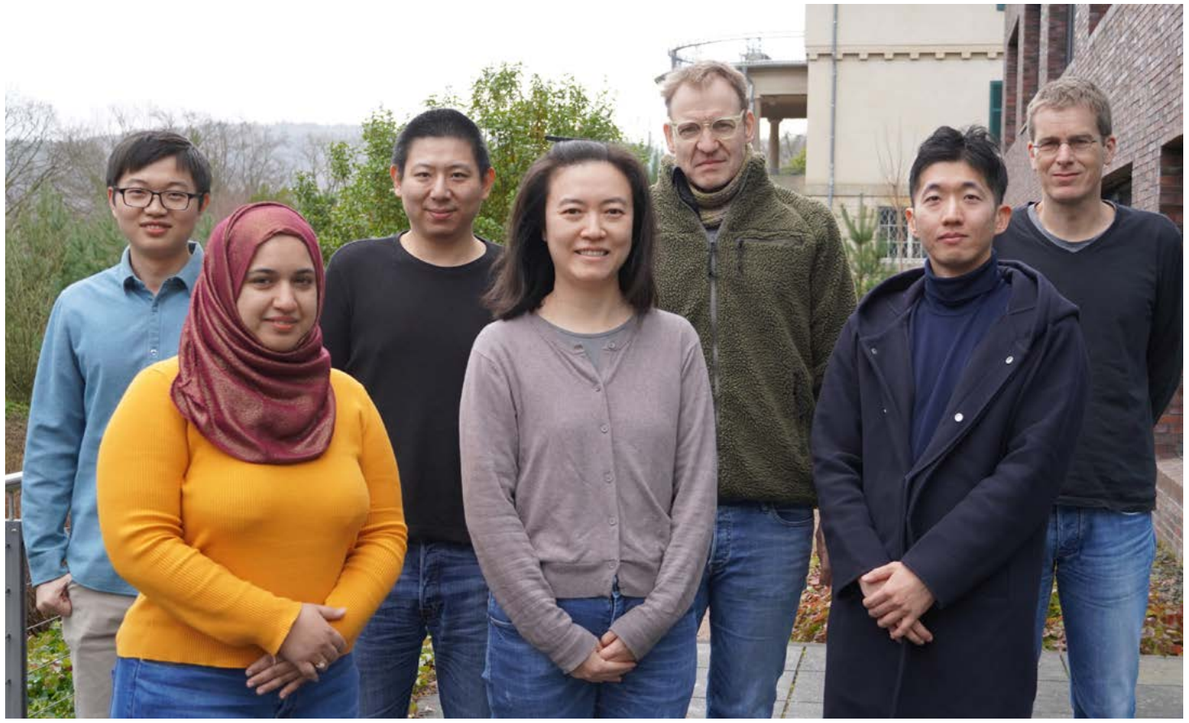
In der **Molecular and Cellular Modeling Gruppe (MCM)** sind wir in erster Linie daran interessiert zu verstehen, wie Moleküle interagieren. Was bestimmt die spezifische und selektive Wirkung beim Zusammenspiel von Wirkstoff und Rezeptor? Wie werden Proteinkomplexe gebildet und welche Formen können sie 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?

Eines unserer Ziele besteht darin, die Mechanismen besser zu verstehen, die bei Wechselwirkung von Medikamenten auf der molekularen Ebene ablaufen, von der Freisetzung des Wirkstoffs über die Bindung zum Rezeptor bis hin zum Metabolismus des Medikaments.

In einem interdisziplinären Ansatz kooperieren wir mit experimentell arbeitenden Forschenden und verwenden gemeinsam rechnerische Methoden aus den Bereichen der Physik-, Bio- und Chemoinformatik. Das breite Spektrum der Techniken, die wir entwickeln und einsetzen, reicht dabei von interaktiven web-basierten Visualisierungswerkzeugen bis hin zu Molekularsimulationen auf atomarer Ebene.

In diesem Bericht beschreiben wir einige der Ergebnisse aus dem Jahr 2021. Nach einem allgemeinen Überblick über Neuigkeiten in der Gruppe konzentriert sich der Bericht auf Projekte zu (i) Wirkstoff-Protein-Bindungskinetik, (ii) makromolekularen Komplexen und (iii) strukturbasierter Wirkstoffentwicklung gegen SARS-CoV-2.

2.9 Natural Language Processing (NLP)



Group leader
Prof. Dr. Michael Strube

Staff members
Dr. Mark-Christoph Müller
Wei Zhao (since July 2021)

Scholarship holders
Haixia Chai (HITS Scholarship)
Wei Liu (HITS Scholarship, since July 2021)
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, delegated, Heidelberg University)

Students
Tobias Martiné

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 automatically analyzing natural language. The group focuses on discourse processing and related applications, such as automatic summarization and readability assessment.

In 2021, the collaboration between the GRG and NLP groups within the HITS Lab project Geometry and Representation Learning produced exciting results. Federico López (NLP) implemented deep learning models for solving NLP tasks based on the mathematical ideas provided by Beatrice Pozzetti (Heidelberg University), Steve Trettel (Stanford University), and Anna Wienhard (GRG and Heidelberg University). This led to two publications at the two top international conferences in machine learning: ICML and NeurIPS. The NeurIPS paper was even highlighted as a spotlight presentation at the conference, which is a reflection of the high quality of our work. Federico managed not only to work on the HITS Lab project, but also to submit his PhD thesis at the end of 2021. He will begin a new job in industry in early 2022. Wei Zhao joined the NLP group in summer 2021 to continue to work on the project.

Alex Judea, who left the NLP group about two years ago, graduated with a PhD in summer 2021 and now works in industry as a data science engineer. Nafise Moosavi, who graduated in 2018 and has been a postdoc at TU Darmstadt since then, accepted a position as a lecturer at the University of Sheffield, which she will begin in early 2022. Congratulations!

Michael Strube was co-chair of the Second Workshop on Computational Approaches to Discourse, which took place at EMNLP 2021. This was the first conference to actually take place during the coronavirus pandemic. The conference and workshop were held in a hybrid format, which made organization even more difficult than it would have been for an entirely virtual event. The event consisted of invited talks, paper presentations, and the shared task on Anaphora, Bridging, and Discourse Deixis in Dialogue, which was also co-organized by Michael Strube. The third iteration of the workshop and the second one of the shared task are planned to be held at the COLING conference in October 2022.

Geometric Deep Learning (Federico López)

Learning faithful graph representations as sets of vertex embeddings has become a fundamental intermediary step in a wide range of machine learning applications. We propose that symmetric spaces be used systematically in representation learning, which is a class that encompasses many of the previously used embedding targets. Symmetric spaces are Riemannian manifolds with rich symmetry groups, which renders them algorithmically tractable. These spaces have a compound geometry that simultaneously contains both Euclidean and hyperbolic or spherical subspaces, thereby allowing them to automatically adapt to dissimilar features in the graph. In our group, we work on developing a general framework for choosing a Riemannian symmetric space and implementing the mathematical tools required to learn graph embeddings. This process enables us to introduce both the use of the vector-valued distance for computing distances and extracting geometric information from the manifolds and the use of Finsler metrics integrated in a Riemannian optimization scheme, which better adapts to dissimilar structures in the graph. We also develop tools for analyzing the embeddings and inferring structural properties of the datasets (Figure 54).

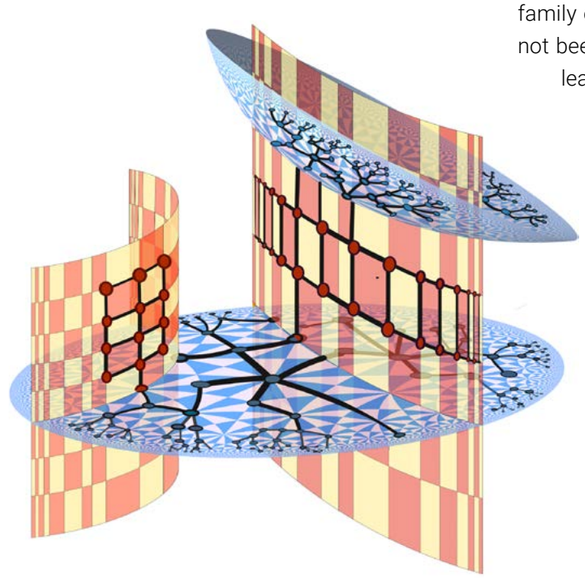


Figure 54: Symmetric spaces have a rich structure of entirely geodesic subspaces, including flat subspaces (orange) and hyperbolic planes (blue). This compound yet computationally tractable geometry allows isometric embeddings of many graphs, including those with subgraphs of dissimilar geometry. For example, the graph embedded in the picture has both trees and grids as subgraphs.

Siegel spaces provide a good opportunity for demonstrating a concrete implementation of our general framework. Siegel spaces are a family of symmetric spaces that has not been explored in geometric deep learning despite the fact that they are among the most versatile symmetric spaces of non-positive curvature. As we verify in our experiments, Siegel spaces with Finsler metrics are an excellent device for embedding complex networks without a priori knowledge of their internal structure. Our approach outperforms competitive baselines for graph reconstruction tasks on various synthetic and real-world datasets. We further demonstrate its applicability in downstream tasks, such as recommender systems and node classification.

We also work with symmetric positive definite (SPD) matrices and develop gyrovector calculus by constructing analogs of vector space operations in this curved space. Our implementation yields closed form expressions of arithmetic operations, such as addition, scalar multiplication, and matrix scaling. This provides the means to translate previously implemented ideas in different metric spaces into their analog notions in SPD. These arithmetic operations are also useful in adapting neural-network architectures to SPD manifolds. We showcase the versatility of SPD in the tasks of knowledge-graph completion, item recommendation, and question answering. In experiments, the SPD models outperform their equivalents in Euclidean and hyperbolic space. The vector-valued distance allows us to visualize embeddings, thereby showing that the models learn to disentangle representations of positive samples from negative ones.

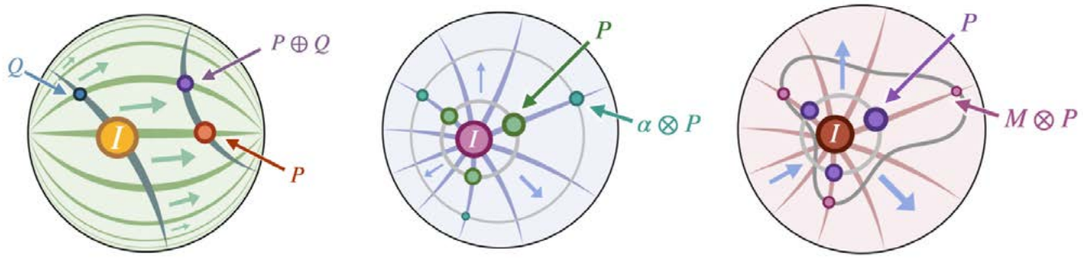


Figure 55: Gyro-addition (left), gyro-scalar multiplication (center), and matrix scaling (right).

NLP for Database Curation in the Biomedical Domain (Mark-Christoph Müller)

DeepCurate is a joint research effort of the SDBV and NLP groups at HITS and was initiated in 2020. The goal of DeepCurate is to develop an integrated, machine-learning-based system for semi-automatic database curation in the biomedical domain. The biocuration use case is provided by the SDBV group's SABIO-RK database and consists of (1) selecting potential source research papers for curation, (2) identifying paper sections that contain relevant, curatable content, and (3) extracting, normalizing, and inserting this content into the database.
Since the DeepCurate system is intended to support rather than replace human curators, care has to be taken to design it in a user-centric way that minimizes user strain and maximizes usability and user acceptance. In this respect, recognizing the importance of printed scientific documents for manual database curation is an important – if somewhat unexpected – process that resulted from discussions with fellow researchers and database-curation practitioners, for example, in the context of conference presentations.

Accordingly, one focus of our work in 2021 was on further integrating paper- and electronic documents. In [Müller et al, 2021], we presented and evaluated a system of creating word-level alignments of scanned scientific paper documents (from the SABIO-RK archive) and their machine-readable full-text versions from PubMed Central®. In the context of manual curation, the motivation for this work is to allow curators to continue using printed paper copies for close reading and markup while simultaneously sparing them the laborious and error-prone step of re-keying data from the paper into the database.
The alignment-data model – which is based on MMAX2 stand-off annotation – is depicted in Figure 56
Apart from the actual text, the model also stores different types of meta-data, depending on the data source: For converted full-text data, it provides typesetting information, such as sub- and superscript markup, which is used for text rendering in MMAX2 (upper part of Figure 56), while for the OCR-generated document text data, it provides the image coordinates of each word (as a "bounding box") and each word's recognition confidence (lower part of Figure 56).

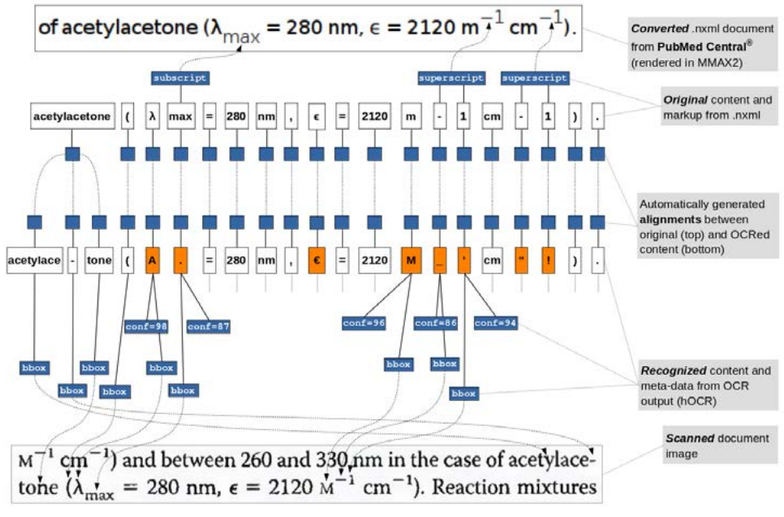


Figure 56: Representing word-level alignment of machine-readable full text (top) and OCRed document images (bottom) in MMAX2 stand-off annotation. (To avoid clutter, not all meta-data are provided in the image.)

Figure 57 below illustrates how the alignment information can be utilized to extract and correct manually highlighted text from the scanned document image. In the example, the expression in the black box at the top (Sub-Figure 1) was incorrectly recognized by OCR (Sub-Figure 2, Relation A). Based on the word-level alignment of the falsely recognized expression and its context (Relation B), it can be mapped to its correct expression (Sub-Figure 3). In a final step, the correct expression can be associated with the section containing the original expression in the image (relation C).

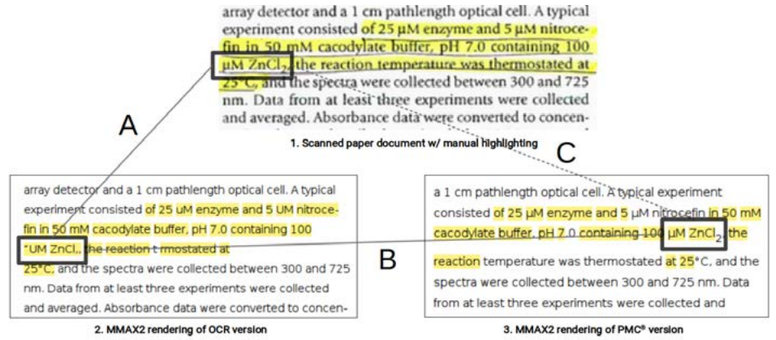


Figure 57: Using alignment for text extraction and correction. Relation A is established using image coordinates yielded by OCR, while Relation B is the result of word-level alignment. Relation C represents the actual extraction and correction as well as results from the combination of Relations A and B.

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.

Die Zusammenarbeit der GRG- und der NLP-Gruppen innerhalb des HITS Lab Projekts „Geometry and Representation Learning" führte 2021 zu faszinierenden Ergebnissen. Federico López implementierte Deep Learning Modelle zur Lösung von NLP Aufgaben, indem er mathematischen Ideen realisierte, die von Betrice Pozzetti (Universität Heidelberg), Steve Trettel (Stanford University) und Anna Wienhard (GRG und Universität Heidelberg) erarbeitet wurden. Das führte zu zwei Publikationen bei den beiden Top-Konferenzen im Bereich Machine Learning, ICML und NeurIPS. Das NeurIPS Papier wurde sogar als sogenannte Spotlight Presentation ausgezeichnet, was die Qualität unserer Zusammenarbeit unterstreicht. Neben der Arbeit im HITS Lab Projekt schaffte Federico es auch, Ende 2021 seine Dissertation einzureichen. Anfang 2022 wird er eine neue Stelle in der Industrie antreten. Wei Zhao, der seit Sommer 2021 zur NLP Gruppe gehört, wird die Arbeit im HITS Lab Projekt fortsetzen.

Alex Judea, der die NLP Gruppe schon vor mehr als zwei Jahren verlassen hatte, konnte im Sommer 2021 endlich promovieren. Er arbeitet schon seit einiger Zeit als Data Science Spezialist in der Industrie. Nafise Moosavi, die 2018 in der NLP Gruppe promovierte und seitdem als Postdoc an der TU Darmstadt arbeitete, nahm ein Angebot als Lecturer an der University of Sheffield an, wo sie Anfang 2022 anfangen wird. Herzlichen Glückwunsch!

Michael Strube war Program Co-Chair des „Second Workshop on Computational Approaches to Discourse", der im Rahmen der EMNLP 2021 stattfand. Dies war die erste Konferenz, die während der Covid-Pandemie im hybriden Modus stattfand, was die Organisation des Workshops noch schwieriger als das Organisieren des Online Workshops 2020 machte. Der Workshop umfasste eingeladene Vorträge, Vorstellungen regulärer Forschungsbeiträge und die Shared Task über „Anaphora, Bridging, and Discourse Deixis in Dialogue", die auch von Michael Strube mitorganisiert wurde. Der dritte Workshop der Reihe ist schon in der Vorbereitung und wird zusammen mit der Shared Task im Rahmen der COLING 2022 stattfinden.

The system was evaluated using the standard measures precision (P), recall (R), and F-measure (F). The alignment algorithm has several pre- and post-processing steps that influence the alignment results. Figure 58 presents the results using different combinations of pre- and post-processing. The left-hand side of Figure 58 (DOIxml–DOIconv) uses document images that were not scanned, but converted (i.e., directly rendered from PDF to images), while the right-hand side (DOIxml–DOIscan) uses actual scanned paper documents. It is clear that the pre- and post-processing steps consistently improve recall (R) at the expense of precision (P) but that overall F-measure also consistently increases. Moreover, it is interesting to note that the absolute differences between the converted and scanned documents are negligible.

Pre-/Post-Processing	$DOI_{xml} - DOI_{conv}$			$DOI_{xml} - DOI_{scan}$		
	P	R	F	P	R	F
–	**95.04**	76.90	85.01	**93.59**	75.29	83.45
dehyp	94.91	77.47	85.31	93.48	75.96	83.81
pre	**95.04**	77.40	85.32	93.57	75.83	83.77
dehyp + pre	94.90	77.97	85.61	93.47	76.52	84.15
post_force_align	95.03	78.57	86.02	93.57	76.99	84.48
dehyp + post_force_align	94.91	79.17	86.32	93.47	77.69	84.86
pre + post_force_align	95.02	79.08	86.32	93.56	77.55	84.81
dehyp + pre + post_force_align	94.90	**79.68**	**86.63**	93.47	**78.27**	**85.20**

Figure 58: Ablation results of the effects of different pre- and post-processing steps (n=68 document pairs).

Cross-Lingual Summarization (Mehwish Fatima)

Cross-lingual summarization (CLS) generates a concise and meaningful summary of a given document in which the input- and target languages are different. Due to the absence of real cross-lingual datasets, recent CLS studies have been conducted on existing monolingual news datasets and on off-the-shelf machine-translation (MT) systems. The MT systems introduce inherent noise to the data. In addition, these summarization models are trained on news datasets, which means that their performance in other domains – such as with scientific texts – is questionable. We addressed these issues by developing a summarization dataset containing English-German-language scientific texts from two resources: Spektrum der Wissenschaft (SPEKTRUM) and Wikipedia Science Portal (WIKIPEDIA). We explored the CLS task using English-language scientific documents in order to generate German-language summaries for German-language readers. Our primary dataset – SPEKTRUM – consisted of 1,510 English-language science articles with human-written German-language summaries. The SPEKTRUM data-collection process is described in Figure 59. Figure 60 displays an example of a German-language SPEKTRUM summary paired with its English-language article. To complement the SPEKTRUM dataset, we harvested our second dataset – containing 51,312 English- and German-language science articles – from WIKIPEDIA. The WIKIPEDIA data-collection process is shown in Figure 61. The WIKIPEDIA dataset can also be used for monolingual summarization (MS), which distinguishes it from existing datasets. We performed a detailed statistical analysis of the dataset that highlighted its distinguishing characteristics and conducted an empirical evaluation with several extractive baselines and abstractive-summarization models to validate the usability of our dataset for MS and CLS. We evaluated linguistic quality on a subset of the output summaries of the MS and CLS experiments with human judges (Fatima and Strube, 2021).

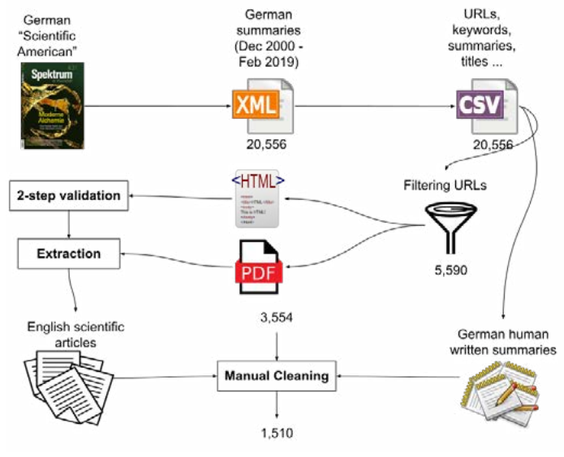


Figure 59: SPEKTRUM data-collection process

We further analyzed the SPEKTRUM dataset because journalists convert the complex English-language scientific papers into simple German-language stories for publishing. As a result, the transformation process involves summarization and translation (cross-lingually) as well as simplification of the text. Our analysis identified differences between the English-language texts and German-language summaries in terms of lexical richness, readability, and parts-of-speech (POS)-based scores. Lexical richness measures the quality of vocabulary in a language sample. We selected three different scores to measure lexical richness: HDD, MTLD, and MATTR. Figure 62 shows the distribution of these scores. The scores suggest that German-language summaries have higher lexical richness than English-language texts. A readability score is an objective measure of the complexity of a text. We chose three different formulas for readability: the Coleman Liau Index, the Linsear writing metric, and the percentage of difficult words. Figure 63 displays the readability scores. These scores indicate that German-language summaries are easier to read than English-language texts. The POS-based scores use grammatical structures to determine how information is carried in a text. We focused on five primary classes for analyzing our texts: nouns, pronouns, verbs, adverbs, and adjectives. A higher percentage of nouns in a text indicates that the text is difficult to understand, whereas the higher number of verbs indicates that it is easy to understand. These scores also suggest that German-language summaries are easier to read and understand than English-language texts. The German-language summaries contain fewer noun phrases while maintaining a higher percentage of verbs, adverbs, adjectives, and pronouns. The distribution of these scores is shown in Figure 64. This in-depth analysis provided us with a jumping-off point for text simplification by focusing on English-language text nominalizations.

English (En) Article

the exceptionally diverse species flocks of cichlid fishes in east africa are prime examples of parallel adaptive radiations . about 80 % of east africa 's more than 1 800 endemic cichlid species , and all species of the flocks of lakes victoria and malawi , belong to a particularly rapidly evolving lineage , the haplochromines . one characteristic feature of the haplochromines is their possession of egg-dummies on the males ' anal fins . these egg-spots mimic real eggs and play an important role in the mating system of these maternal mouthbrooding fish . here , we show that the egg-spots of haplochromines are made up of yellow pigment cells , xanthophores , and that a gene coding for a type iii receptor tyrosine kinase , colony-stimulating factor 1 receptor a (csf1ra) , is expressed in egg-spot tissue . [...]

German (De) summary

das farb - gen csf1ra sorgt in männlichen barschen für charakteristische pigmentflecken , die für die erfolgreiche befruchtung der eier notwendig sind . walter salzburger und ingo braasch klärten an der universität konstanz damit die genetische herkunft dieser optischen ei - attrappen auf , die im paarungsritus der fische die weibchen anlocken . die neunzehn barscharten , die die forscher für ihre untersuchungen auswählten , gehörten zu gruppe der haplochrominen . [...]

Figure 60: SPEKTRUM data example

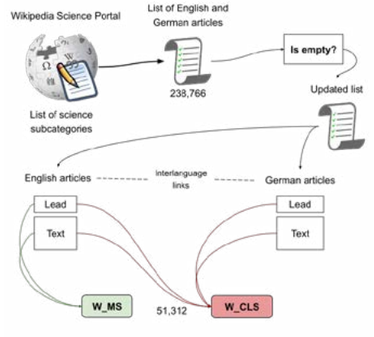


Figure 61: WIKIPEDIA data-collection process

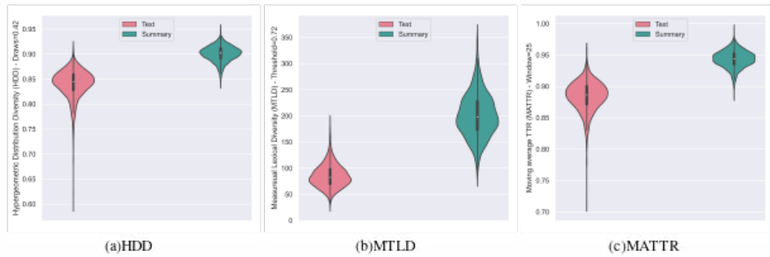


Figure 62: Distribution of lexical-richness scores in the SPEKTRUM dataset

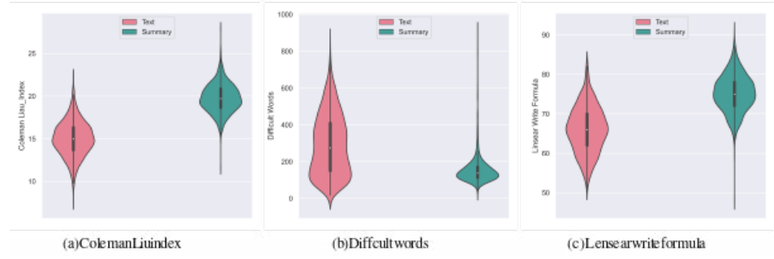


Figure 63: Distribution of readability scores in the SPEKTRUM dataset

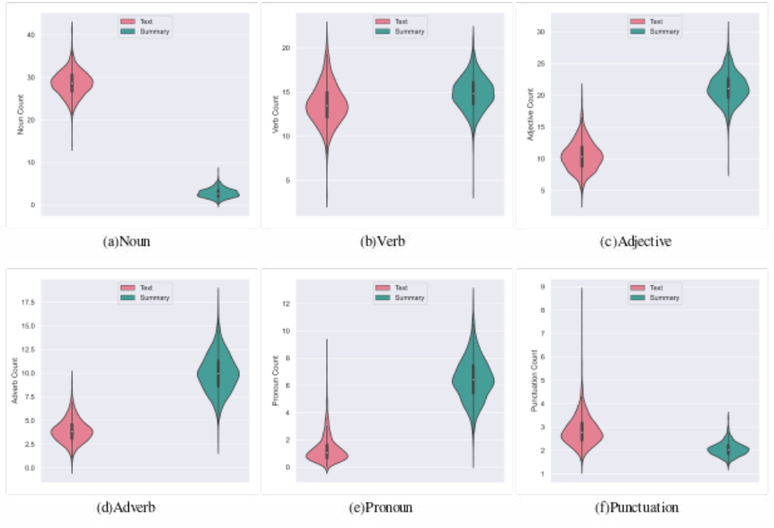
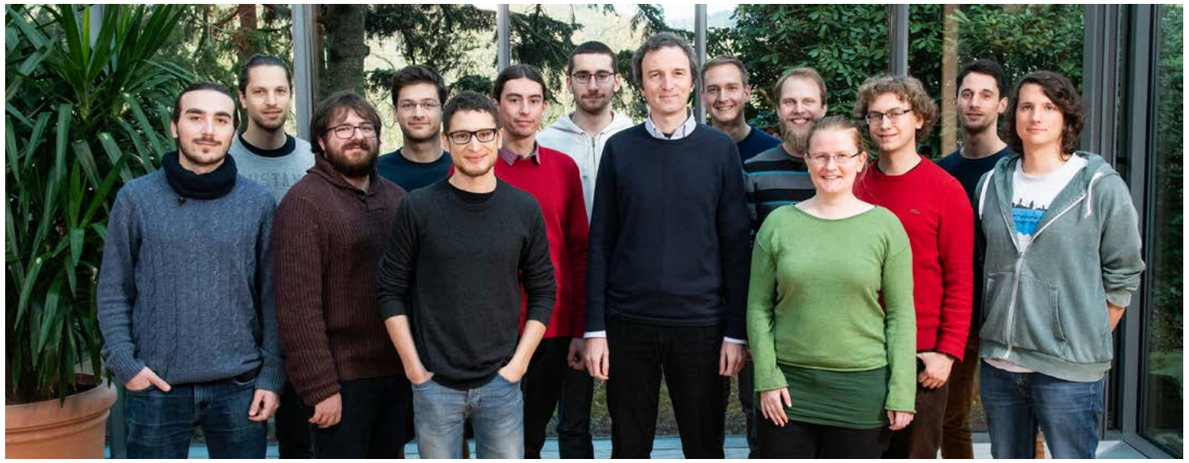


Figure 64: Distribution of normalized POS-based scores in the SPEKTRUM dataset

2 Research

2.10 Physics of Stellar Objects (PSO)



Group leader
Prof. Dr. Friedrich Röpke

Staff members
Dr. Róbert Andrássy
Dr. Johann Higl
Javier Morán Fraile
Kiril Maltsev

Scholarship holder
Christian Sand (HITS Scholarship)

Visiting scientists
Sabrina Gronow (Heidelberg University; until March 2021)
Alexander Holas (Heidelberg University; since October 2021)
Florian Lach
Giovanni Leidi
Theodoros Soultanis (IMPRS PhD student at MPIA Heidelberg)

Students
Thomas Baumann (until July 2021)
Marco Vetter (since July 2021)
Freyja Walberg (since July 2021)
Gabriel Wiest (since July 2021)

"We are stardust." The matter from which we are made stems from the primordial material formed during the Big Bang, which is processed in various places throughout the Universe. All heavier elements originate from nucleosynthesis in stars and in gigantic stellar explosions. How this material forms and how it is distributed throughout the Universe are fundamental concerns for astrophysicists. At the same time, stellar objects make the Universe accessible to us by way of astronomical observations. Stars shine – inter alia – in optical parts of the electromagnetic spectrum and are the fundamental building blocks both of galaxies and of larger cosmological structures.
With the help of extensive numerical simulations, the Physics of Stellar Objects research group seeks to understand the processes that take place in stars and in stellar explosions. Newly developed numerical techniques and the ever-increasing power of supercomputers facilitate the modeling of stellar objects in unprecedented detail and with unparalleled precision.

One of our group's primary goals is to model the thermonuclear explosions of white dwarf stars, which 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, which led 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. Such simulations 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 that 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 via 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.

Can we trust our simulations?

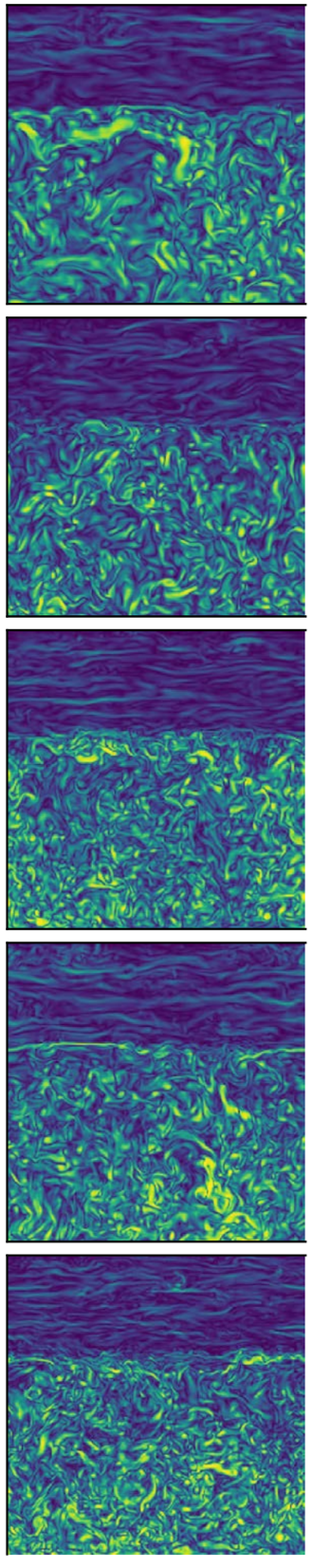
How stars live and die depends on a large number of physical processes that occur deep in their opaque stellar interiors. Although the recent boom in asteroseismology has revealed certain properties of the interiors of some stars, our understanding of how stars work is mostly based on computer simulations and is guided by observations of the surface properties of different types of stars.

Because stars live very long lives, models of stellar evolution generally resort to a drastic oversimplification by representing a star as a spherically symmetrical one-dimensional object. Multi-dimensional physical processes – such as convection or rotation – are included via simplistic parametric prescriptions, which are usually calibrated to reproduce a selected set of observations and are boldly applied to a wide range of stars. Simulations in two and three spatial dimensions can be performed on time scales much shorter than the stellar lifetime. These simulations reveal a complex spectrum of three-dimensional hydrodynamic flows and instabilities and allow us to develop and calibrate new simplified prescriptions. More importantly, the simulations give us the chance to learn more about how fundamental processes of fluid dynamics work in stars.

Such simulations usually include all the processes that we find relevant in any particular stellar environment. But how can we know if the results are correct? Laboratory experiments cannot even begin to replicate the extreme conditions that prevail in stellar interiors. However, it is possible to compare results obtained using different methods and simulation codes. Such comparisons are occasionally performed in astrophysics. However, their absence in the subfield of convective mixing processes (which take place deep in the stellar interior) motivated us to launch a new code comparison study in the summer of 2019.

For the study, members of the PSO group who develop our in-house SLH hydrodynamics code teamed up with the developers of other major codes that are used in the field to create a 21-member collaboration involving researchers from Germany, the US, the UK, Canada, and Australia. Together, we managed to involve five different simulation codes: FLASH, MUSIC, PPMSTAR, PROMPI, and SLH. All of these codes – with the exception of FLASH – were set up and run by researchers who were directly involved in the codes' development. This approach also helped eliminate any user issues.

The usual code testing strategy is to solve a number of simple problems that focus on specific flow types that are known to challenge numerical schemes in different ways. All of our codes passed such tests during their development. We decided to go one step further and compare the codes' predictions for a single complex test problem involving turbulent convection, convective boundary mixing (which is currently a hot topic in the field), and the generation of internal gravity waves – all in a single simulation box. However, we took care in formulating the technical details of the problem in order to make it both easily implementable in most existing codes and available to anyone who might want to use it in the future. We further facilitated the reproducibility of our study by providing the results of our



simulations and the whole data processing pipeline in the form of a public Jupyter Hub, which can be found at https://www.ppmstar.org/coco. The study was recently accepted for publication in Astronomy & Astrophysics [Andrassy et al. 2022, preprint available at https://arxiv.org/abs/2111.01165].

Snapshots from some of our simulations are shown in Figure 65. Although they appear different, it is necessary to bear in mind that turbulence produces pseudo-random flow patterns and that only some statistical properties of the flow field are meaningfully comparable. We performed such statistical analyses and found the five codes to be in excellent statistical agreement.

The highlight of our study was a comparison of convective boundary mixing rates at the interface between the convective and stable layers. This quantity – which is critically important to drawing correct conclusions about stellar evolution – turned out to agree within a few percent, even on easily affordable numerical grids. All in all, the results of the study bolstered our confidence in the predictions we make using our codes on a daily basis.

Figure 65: Snapshots of vorticity produced by the five codes run on the same numerical grid of 256³ cells and taken at the same simulation time. Gravity points downward in the plane shown. The lower half of the simulation box is convective and reaches large vorticity values, whereas the overlying field of internal gravity waves is dominated by nearly horizontal structures of somewhat lower amplitude.

The birth of planetary nebulae from binary-star interaction

The introduction of telescopes into astronomy allowed early astronomers to observe never-before-seen objects. Some round and blurry shapes reminded them of planets such as Jupiter, and the name "planetary nebulae" was therefore coined for these objects – probably by the famous German-born British astronomer William Herschel in the 1790s. However, this term turned out to be a misnomer: Indeed, modern observations – most prominently the iconic images taken by the Hubble space telescope – show that the objects are actually gaseous structures that extend across ca. one lightyear, which is roughly the size of the entire solar system. This finding suggests that such objects are related to stars and to their evolution rather than to planets.

However, the physical mechanism responsible for the formation of these multifaceted and colorful structures has remained unclear. Gas from the envelope of a star is ejected, thereby exposing its hot inner core, which emits energetic ultraviolet radiation. This radiation is absorbed by the ejected gas and re-emitted in visible wavelengths, thus giving rise to the vibrantly colored objects that are observed as planetary nebulae.

But what causes the gas to be ejected in the first place? A common morphology of planetary nebulae takes a bipolar shape that is sometimes very prominent (i.e., as two jet-like structures shooting out from the center of the object) and sometimes broader and blurrier. The shapes are reminiscent of butterflies

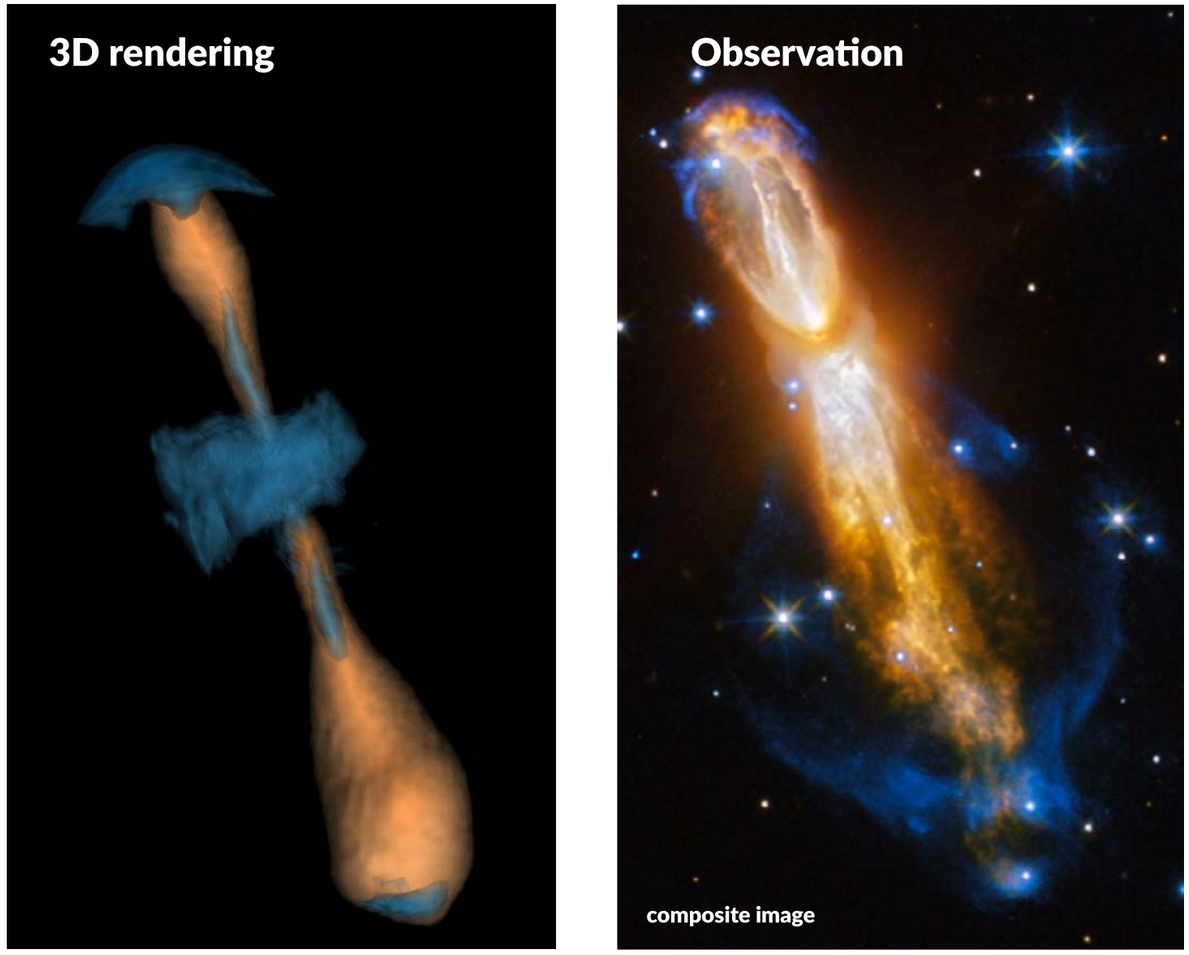


Figure 66: Qualitative comparison of the morphology of gas ejected in a simulation of common-envelope interaction in a binary stellar system (left panel) with the Calabash Nebula (OH 231.84 +4.22, right panel; composite optical image taken by the NASA/ESA Hubble Space Telescope).

or antennae and gave rise to nicknames for some popular planetary nebulae. An example is the Calabash Nebula, which is shown in the right-side panel of Figure 66. More detailed observations have revealed that the ejected gas consists of two components: one that is slowly expanding at velocities of about 10 km/s, and another fast component that is characterized by velocities of 100 km/s. These bipolar planetary nebulae have long been speculated to originate from gas ejected by a binary star, and indeed, in some objects, a central binary star has been detected.

When one of the stars in a binary stellar system evolves faster and enters its giant phase, it may engulf its companion within its envelope, thereby giving rise to the enigmatic "common-envelope" phase of binary stellar evolution, which has posed a challenge to theory for decades. The two stellar cores spiral in and come closer, thus losing their orbital energy to the envelope through drag (see Panels a, b, f, and g in Figure 67, next page). Ultimately, the common envelope is ejected, and a tight binary stellar system emerges. The ejected envelope may form the rather spherical component of the

gas in a planetary nebula. In extensive three-dimensional hydrodynamic simulations performed by the PSO group, this ejected envelope has been found to have typical velocities of ca. 10 km/s. It is ejected into a toroidal structure in the direction of the orbital plane of the core binary system, and a funnel forms perpendicular to it – along the axis of rotation – that contains little gas (see Panel h in Figure 67). This funnel structure had been speculated to act as a nozzle in shaping the bipolar gas outflow at high velocities. But how is this fast outflow launched?

With their pioneering magnetohydrodynamic simulations (Ondratschek, Astronomy & Astrophysics, (2022), in print) of a common-envelope system, members of the PSO group were able to show that magnetic fields are strongly amplified in this phase. The researchers discovered that in late phases of the evolution, when large parts of the envelope have been ejected, the remaining stellar cores come close enough such that mass is transferred from one core to another (see Panels e and j in Figure 67). Instabilities in the mass overflow again boost the strength of the magnetic field, which launches a high-velocity jet-like outflow along the rotation axis – that is, perpendicular to the orbital plane and along the funnels in the ejected common-envelope gas (barely visible in Panel i in Figure 67). The high-velocity gas drives bow shocks into the envelope and is subject to hydrodynamic instabilities. The resulting structure has a morphology that closely resembles young planetary nebulae (see Figure 66, previous page, for a comparison of the structure obtained from the simulation with the observation of the Calabash Nebula). With this mechanism, the PSO group and its collaborators were able to explain the formation of bipolar planetary nebulae for the first time with a self-consistent model.

Patrick A. Ondratschek et al.: Bipolar planetary nebulae from common envelope evolution of binary stars

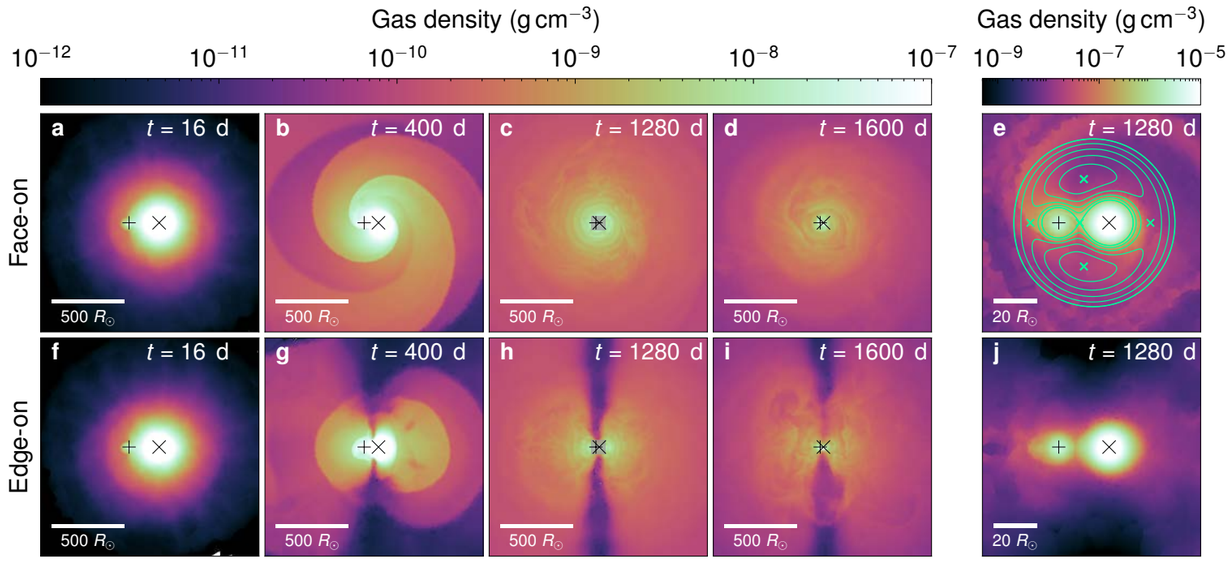


Figure 67: Slices through the density in the orbital plane (top row) and the plane perpendicular to it (bottom row) at different times in the simulation of the common-envelope interaction. The inspiral and the formation of a funnel along the rotation axis are visible. In snapshot (i), a jet-like outflow propagates through the funnel and gives rise to hydrodynamic instabilities (barely visible in the lower hemisphere). Panels (e) and (j) show close-ups of the central binary system, in which mass transfer at this late stage gives rise to a strong amplification of magnetic fields, which drives the bipolar outflow.

„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 Löcher. 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 (PSO)** 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.

2 Research

2.11 Scientific Databases and Visualization (SDBV)

Group leader

PD Dr. Wolfgang Müller

Staff members

Dr. Alain Becam (since January 2021)
Helen Desmond (until June 2021)
Dr. Dorotea Dudas
Dr. Sucheta Ghosh
Martin Golebiewski
Bettina Heinlein (since February 2021)
Xiaoming Hu
Vivien Louise Junker (until April 2021)
Dr. Olga Krebs
Michael Lieser (until May 2021)
Lukrécia Mertová (since April 2021)
Ghadeer Mobasher
Dr. Maja Rey
Dr. Natalia Simous
Dr. Andreas Weidemann
Dr. Ulrike Wittig

Students

Alexander Christiansen (until September 2021)
Laura Fisch (until July 2021)
Vincent Fischer (until October 2021)
Alexander Goitom (until April 2021)
Anton Hanke (since January 2021)
Jan Koß
Jana Krieg (since May 2021)
Fabian Springer

The SDBV group at HITS has maintained a steady mission: to provide FAIR data to scientists. FAIR is an acronym that refers to the notion that data should be Findable, Accessible, Interoperable, and Reusable.

Our group has been working on biological data for more than 20 years and in so doing has followed and contributed to exciting developments. Fifteen years ago, funding for research infrastructure was severely lacking. Now, 15 years later, we are coming to the end of 7 years of BMBF funding for the de.NBI infrastructure – funding that was instrumental in building the German ELIXIR Node using European bioinformatics infrastructure.

Additionally, we have become part of the NFDI (National Research Data Infrastructure), which is also intended to be a long-term endeavor. Funders now fully embrace the notion that just as infrastructure made of concrete and steel is necessary for society to function, research also needs an information infrastructure to function.

Despite all these changes, the central questions of our group have remained relevant and exciting:

- How can we best integrate FAIR data and tools and make them more useful?
- How can we make FAIR data more findable (as well as more accessible, interoperable, and reusable)?
- How can we help others to make data FAIR?

The present contribution to the 2021 annual report provides new answers to two of these questions. The RDMKit offers a lightweight method of bringing tools together, and the SABIO-VIS project focuses on visualizing SABIO-RK reaction-kinetics data with the aim of more intuitively and precisely representing data.

And the third question? Yes, in addition to large-scale collaborative projects, such as LiSyM, de.NBI, and NFDI, we are also involved in smaller projects that investigate how to facilitate curation with the goal of developing infrastructure. The NLP-related projects DeepCurate and PoLiMeR are now entering their third year. Moreover, we are also now using our HITS funding for two more small projects, which we look forward to examining more closely next year.

ELIXIR Research Data Management Kit (RDMkit)

ELIXIR (https://elixir-europe.org/) is a European intergovernmental organization with the goal of building a bioinformatics infrastructure to manage the increasing complexity of data and to make it easier to find the right tools and training. To that end, ELIXIR builds on pre-existing and new pieces of bioinformatics infrastructure. In 2020, the EU-funded project ELIXIR CONVERGE was launched with partners from 23 ELIXIR nodes. The focus of the project is to standardize research data management in the life sciences across Europe. Key objectives include establishing a European expert network of data stewards, providing comprehensive training and capacity building, and developing a common research data management toolkit. The ELIXIR CONVERGE project consists of 9 work packages, 3 of which have been added since the project launched in order to cover the increasing need for managing and sharing SARS-CoV-2/COVID-19-related data caused by the coronavirus pandemic.

The ELIXIR Research Data Management Kit (RDMkit) (https://rdmkit.elixir-europe.org/) was launched in March 2021. It is an online guide for research data management in the life sciences that includes best practices and examples as well as links to data management tools, resources, and training. RDMkit supports the use of FAIR (Findable, Accessible, Interoperable, and Reusable) data from the beginning of a research project through to the final steps of depositing data in public archives via data management planning. RDMkit thus plays a central role in data management and is important to all steps of the research data life cycle during a research project, which includes planning, collecting, processing, analyzing, preserving, sharing, and reusing data.

RDMkit was developed for users who play different roles within a research project. For researchers and project consortia, RDMkit offers best-practice guidelines for daily work with research data in all steps of the research data life cycle and fosters the creation of project data management plans. Data stewards can find recommendations and resources for improving their expertise, while funding agencies and policymakers are supported by guidelines on research data management policies. Links to corresponding training material and events as well as cross-references to other ELIXIR resources are included based on availability.

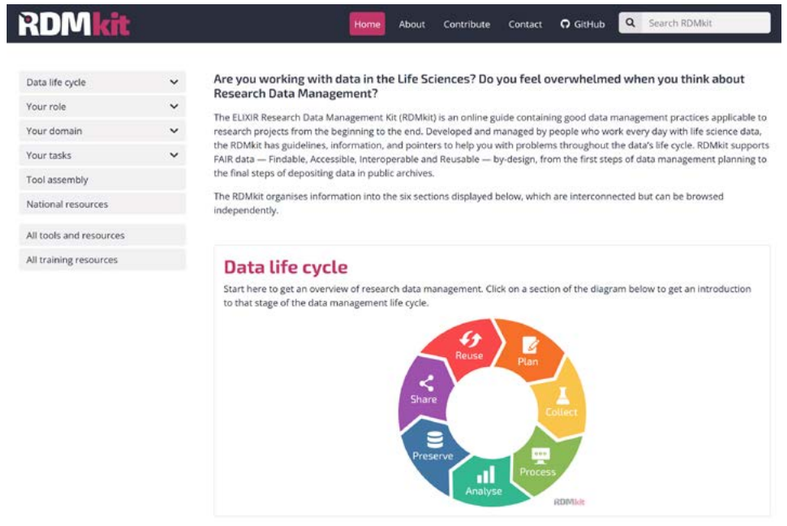


Figure 68: Screenshot of the RDMkit front page (https://rdmkit.elixir-europe.org/), including the research data life cycle (as of January 2022).

The information in RDMkit is organized in different sections based on (i) the stages of the research data life cycle, (ii) the role of the user within a project, (iii) the specific scientific domain, and (iv) the specific task or problem with which the user is confronted. Additionally, country-specific information is collected in national-resource pages for special requirements and tools within each ELIXIR node. A list of tool assemblies provides examples of the domain- or country-specific integration of different data management tools and resources and highlights the interoperability of these items.

One main data management tool for collecting, storing, and sharing research data and that is currently highlighted in four RDMkit tool assemblies is the data management platform FAIRDOM-SEEK (https://fairdomseek.org/). FAIRDOM-SEEK was developed by the SDBV at HITS together with collaboration partners in FAIRDOM (https://fair-dom.org/) and is listed as an ELIXIR service. The integration of this open-source tool into various domains and data management workflows demonstrates its flexibility and high degree of acceptance within the research community.

RDMkit is an open and ongoing community effort of all ELIXIR nodes and of partners from outside ELIXIR, who contribute content to the different webpages. Authors can directly contribute through GitHub, which provides a simple and sustainable platform. An editorial board of 15 members reviews and manages the content and drives further developments. Ulrike Wittig from the SDBV group represents HITS and Germany on RDMkit's editorial board. As of January 2022, there are 116 contributors who have edited 53 RDMkit pages containing 351 tools and resources. Each contributor is acknowledged on the corresponding RDMkit page. To coordinate the contribution to specific domains or tasks, experts meet in focus groups – or "content-athons" – to define content and create pages together.

ELIXIR's node-specific information, resources, and tools – which include recommendations and policies for research data from national funding agencies, institutions, universities, etc. – are collected on national-resource pages in RDMkit.

As confirmation of the importance of the work performed on RDMkit, the European Research Council included RDMkit in the guidelines for Open Research Data and Data Management Plans and stated that "the ELIXIR Research Data Management Kit (RDMkit) is a useful resource for grantees in the life sciences."

SABIO-VIS

SABIO-VIS is a BMBF-funded project designed to help improve the visual exploration and validation of data in SABIO-RK (http://sabiork.h-its.org/) with the goal of supporting modelers, database curators, and experimenters. SABIO-RK is a manually curated database for biochemical reactions and their kinetic properties that has been being developed by SDBV at HITS for over than 15 years. The data are mainly manually extracted from scientific literature published from the 1960s to the present. The growing volume of data in the database requires a better overview of the database's content and especially of the distribution of kinetic parameters for specific scientific questions. The main goal of the project is to develop a visualization of the data in SABIO-RK, thereby supporting a better understanding of the database's content and facilitating the detection of possible inaccuracies in kinetic parameters from the literature. Data clustering and grouping (e.g., kinetic parameters, EC numbers, environmental conditions) are implemented according to the needs of SABIO-RK users and curators.

In the first version of SABIO-VIS, data on reactions, proteins/enzymes, organisms, tissues, and the experimental conditions of pH and temperature were included. Three different visualization concepts are available: a heat-map overview, parallel coordinates, and a scatter-plot matrix with histograms. In this first version, which is already accessible online in the SABIO-RK user interface, the kinetic parameters (e.g., Km, Vmax, kcat) were excluded. These visualizations of the general data in SABIO-RK are indexed to database entries.

For its second version, SABIO-VIS was extended to show kinetic parameters – that is, the kinetic data of the entries. Since each entry may contain several kinetic parameters that include types, values, units, and associated species, these parameters are dealt with separately, thereby allowing for better

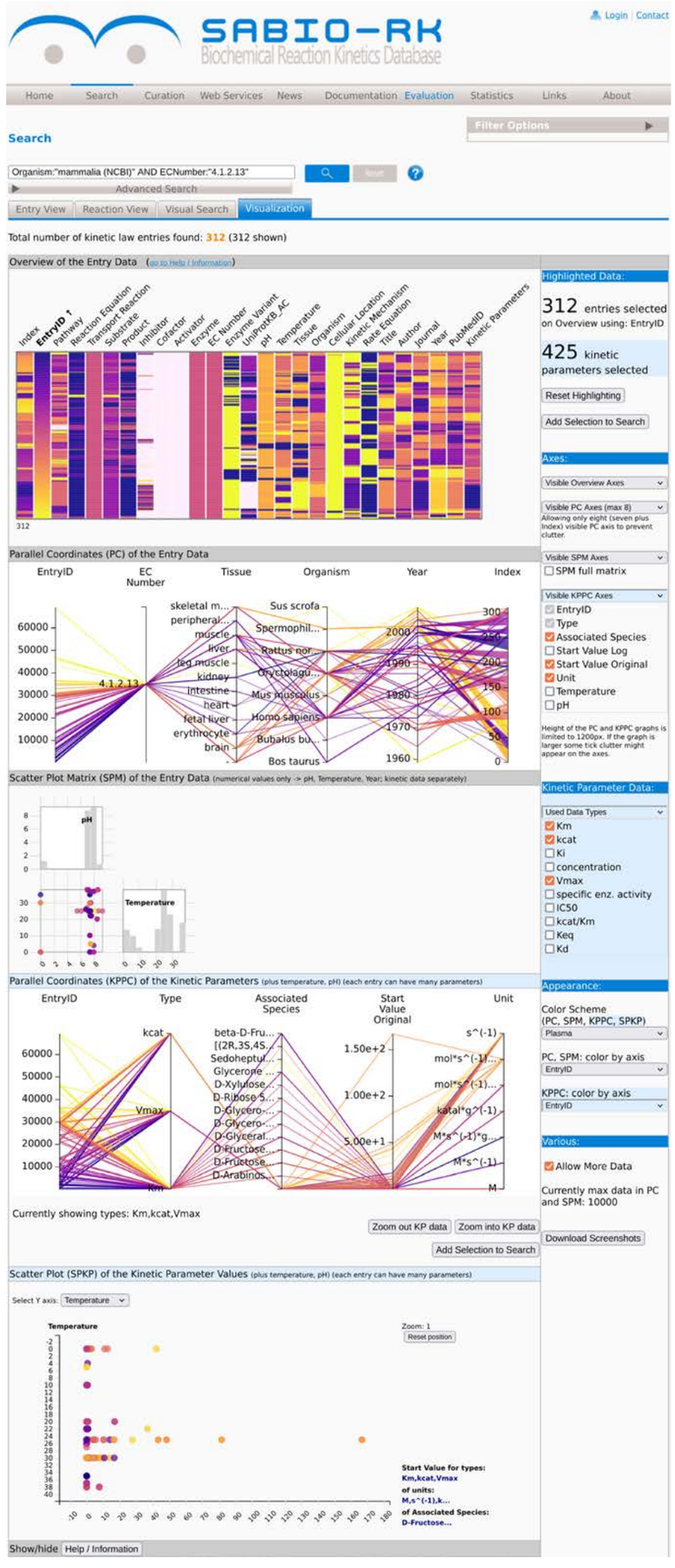


Figure 69: Screenshot of the visualization of data in SABIO-RK based on a search for the enzyme Fructose-bisphosphate aldolase (EC 4.1.2.13) in mammals.

exploration of the kinetic data and the connections of these data to the rest of the data in SABIO-RK. After overcoming challenges with handling kinetic data in the database, two additional representations that deal with kinetic parameters were added to the visualization.

Kinetic parameters are shown on an additional parallel-coordinates plot indexed by the parameters themselves. The user can explore the parameter space – similarly to the entry-data space – by using the kinetic-parameter data keys (unit, type, value, associated species), which are represented as axes of the parallel coordinates. The new additional parallel-coordinates plot is connected to the previous entry-based plots, and selecting data there means that the corresponding entries (heat-map overview, parallel coordinates, and the scatter-plot matrix) are highlighted on the entry-based plots. Before searching the whole database for the narrowed selection of kinetic data (note that searches provide full entries all of the kinetic parameters as a result with), the user can locally zoom in and out of the kinetic-parameter data space, which allows for a more precise composition of the desired search. A pre-selection of the kinetic-parameter data types shown in the graph is then made, and only the most prominent types are shown in the beginning (if any are found). However, in the GUI, the user can choose which types should be shown.

An additional scatter plot is introduced to show the connection of the kinetic-parameter values and a chosen numerical value (which can be temperature, pH value, or simply the entry ID). This graph corresponds to a selection on the parallel-coordinates plot of the kinetic parameters, from which users can select their desired types, units, and associated species from the data.

By combining the two kinetic-data graphs, it is possible to gain deeper insights into the kinetic data at hand. This process can also be used by curators to detect errors as well as by users to identify clusters or outliers of kinetic-parameter values.

In addition to the possibility to explore the kinetic data, SABIO-VIS has had additional functionality added to the GUI, and its speed and usability have also been improved.

Figure 69 (see previous page) shows a screenshot of the second version of SABIO-VIS, which includes kinetic parameters. This version will soon be available online for users to test. Extensive user testing and feedback collection – planned for 2022 – will help to improve the usability of SABIO-VIS and to further adapt its functionality to user requirements.

Die Mission der **Scientific Databases and Visualization Gruppe (SDBV)** am HITS bleibt gleich: FAIRe Daten für Wissenschaftler/-innen. FAIR ist eine Abkürzung, die bedeutet, dass Daten auffindbar (findable), zugreifbar (accessible), interoperabel und wiederverwendbar (reusable) sein sollen.

Seit mehr als 20 Jahren arbeitet unsere Gruppe an Bio-Daten und ist Teil (zuschauend und beitragend) von großen Entwicklungen. Vor 15 Jahren war es schwer, Arbeit an Forschungsinfrastruktur zu finanzieren. 15 Jahre später können wir auf 7 Jahre BMBF-Finanzierung der verteilten de.NBI-Infrastruktur zurückblicken. de.NBI bildet den deutschen Knoten von ELIXIR in der europäischen Bioinformatik-Infrastruktur.

Wir sind auch Teil der NFDI, der nationalen Forschungsdateninfrastruktur, die auf längere Zeit ausgerichtet ist. Die Forschung fördernden Organisationen haben jetzt voll verinnerlicht, dass neben Forschungsinfrastruktur aus Beton und Stahl auch Dateninfrastruktur wichtig ist.

Bei all diesen Veränderungen bleiben die Hauptfragen relevant und spannend:
- Wie integrieren wir FAIRe Daten und Tools, damit sie möglichst nützlich werden?
- Wie können wir FAIRe Daten möglichst gut auffindbar, zugreifbar, interoperabel und nutzbar machen?
- Wie können wir Menschen dabei unterstützen, ihre Daten FAIR zu machen?

Dieser Jahresbericht beschreibt neue Antworten für zwei dieser Fragen. Das RDMkit zeigt einen leichtgewichtigen Weg, Werkzeuge zusammenzubringen. Und SABIO-VIS ist ein Projekt zur Visualisierung von SABIO-RK Daten zur Reaktionskinetik. Hierbei wird eine intuitive aber präzise Repräsentation der Daten angestrebt.

Und die dritte Frage? Ja, zusätzlich zu sehr großen Verbundprojekten wie LiSyM, de.NBI und NFDI sind wir Teil kleinerer Projekte, die Forschung betreiben, wie man Kuratierung verbessern kann. Dies hat spätere Infrastrukturentwicklung zum Ziel. Die Projekte mit NLP-Bezug DeepCurate und PoLiMeR kommen in ihr drittes Jahr, und wir nutzen unser HITS-Funding für zwei weitere, kleine Projekte. Wir freuen uns darauf, im nächsten Jahr von ihnen berichten zu können.

2 Research

2.12 Stellar Evolution Theory (SET)

Group leader
Dr. Fabian Schneider (since January 2021)

Staff members
Vincent Bronner (since October 2021)
Jan Henneco (since February 2021)
Dr. Eva Laplace (since November 2021)
Duresa Temaj (since September 2021)

Visiting scientist
Dr. Dandan Wei (CSC China – DAAD; since October 2021)

Student
Lorenz Thielbeer (January–October 2021)

Stars are the basic building blocks of the visible Universe and produce almost all chemical elements heavier than helium. Understanding how stars have transformed the primitive Universe into the one we live in today lies at the heart of astrophysics research.
Massive stars are cosmic powerhouses. They can be several million times more luminous than the Sun, have strong stellar winds, and explode as powerful supernovae. Thanks to the energetic feedback they provide, massive stars have helped to re-illuminate the Universe since the Cosmic Dark Ages, to drive the evolution of galaxies, and to lay the foundation for life as we know it.
At the end of their lives, massive stars leave behind some of the most exotic forms of matter: neutron stars and black holes. By observing these remnants, we can study matter under conditions that are unavailable on Earth. Mergers of neutron stars and black holes are now routinely observed using gravitational-wave observatories, which have opened a new window to the Universe.
Today, we know that most massive stars are born in binaries and higher-order multiples — triples, quadruples, and so on. This situation has interesting consequences. As stars age, they grow and may eventually become giants with radii measuring up to ~1,000 times that of our Sun. Stars in binaries can reach a stage in which their outer layers are transferred to their companion. In about 25% of massive stars, this mass-transfer phase is unstable and leads to the merger of both binary components. Mass-exchange episodes and the even-more drastic merger events profoundly change the evolution of stars and their ultimate fate. For example, if a star loses its envelope in a mass-transfer phase, it may explode as a supernova and produce a neutron star instead of collapsing into a black hole at the end of its life.
The Stellar Evolution Theory (SET) group investigates the turbulent and explosive lives of massive stars. Currently, the group focuses on massive binary stars and their intricate merging process. Mergers produce strong magnetic fields, and the products of these mergers may forge highly magnetized neutron stars in the resulting supernova explosions. These magnetic neutron stars – known as "magnetars" – are the strongest magnets in the Universe.

A new group

The Stellar Evolution Theory (SET) Group was founded at HITS in January 2021 thanks to an ERC Starting Grant. The first member was then-bachelor's student Lorenz Thielbeer, who joined immediately, followed by PhD student Jan Henneco in February. It took a few months before master's students Duresa Temaj and Vincent Bronner joined in September and October, respectively, followed by postdoctoral researchers Dandan Wei and Eva Laplace in October and November, respectively. The latest team member is bachelor's student Tina Neumann, who joined the group in December. The team has grown steadily, and together, we are now investigating the intricate and complex lives, deaths, and afterlives of massive stars.
While about half of all solar-like stars have a companion, virtually every massive star is born with a companion. Often, even three or four stars are found orbiting one another. But why does that matter so much? The answer lies in the fact that stars expand in size as they age. It is therefore possible for a star to reach such a large radius that its outermost layers are gravitationally more strongly attracted by its companion than by the star itself, and a phase of mass transfer begins (Figure 70).

Mass transfer can lead to various pathways and final outcomes depending mostly on the initial masses of the two stars and their initial separation. Most commonly, a stable phase of mass exchange sets in, during which the envelope of the initially more massive star is transferred to the companion (Figure 70, middle). The former donor star is now said to have become a "stripped-envelope" star, which soon ends its life in a supernova explosion, thereby giving birth to a neutron star (NS) or a black hole (BH). The binary system can undergo further mass-transfer episodes (but now, mass is transferred to the NS or BH) until the other star also potentially explodes as a supernova, thereby forming a second NS or BH. Thanks to gravitational-wave emission, the binary orbit shrinks until the two compact objects merge to form one black hole. The latter merger phase has been routinely observed since 2016 thanks to new gravitational-wave observatories and has already resulted in many breakthrough discoveries. Most notably, such discoveries were recognized with the 2017 Nobel Prize in Physics. During mass transfer, both stars can also evolve into a state of physical contact (Figure 70, left). This phase remains largely unexplored and may lead to stellar mergers, which result in spectacular transients. Indeed, the Great Eruption of Eta Carinae in the 1840s might just have been such an event (see below). During the merger, strong magnetic fields are produced (Schneider et al. 2019, Nature, 574, 211). These magnetic fields may hold one clue to understanding the origin of strong surface magnetic fields, which are observed in about 10% of massive stars. While still speculative, the idea is that such

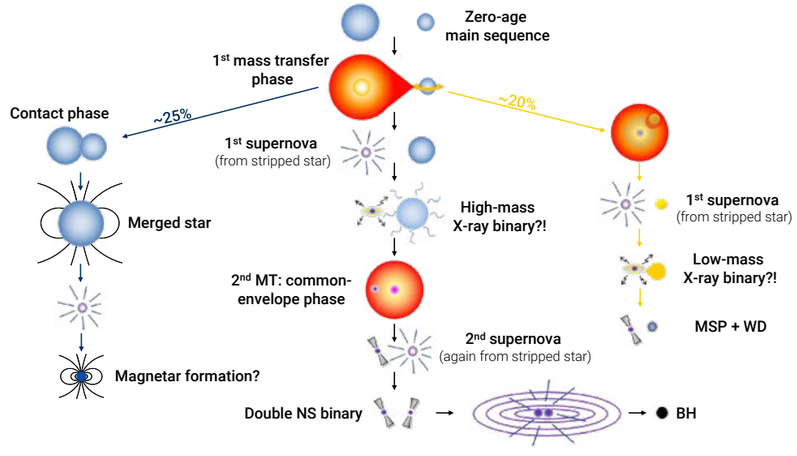


Figure 70: Schematic evolutionary paths of binary stars evolving through various phases of mass transfer (MT). The starting points are so-called zero-age main-sequence stars, in whose cores nuclear fusion has just begun. The figure shows paths leading to a contact configuration with subsequent merger and magnetar formation (left), a double-neutron-star (NS) binary with a gravitational-wave-merger event forming a black hole (BH; middle), and a millisecond-pulsar white-dwarf binary (MSP+WD; right). Image credits: Thomas Tauris.

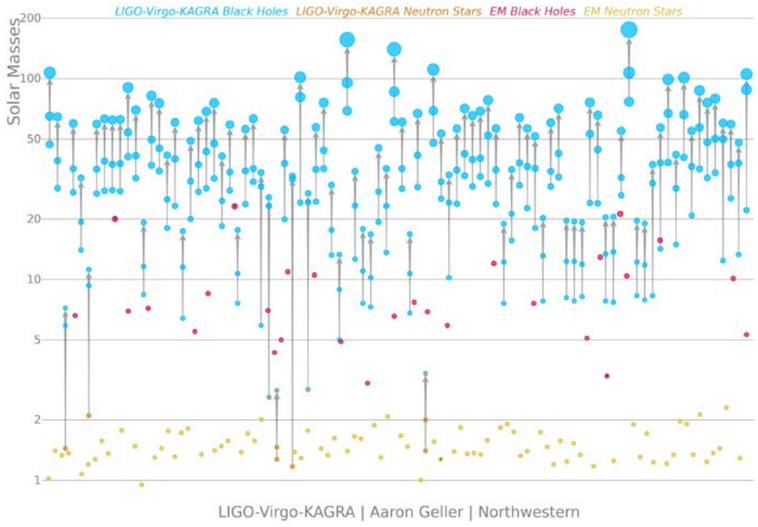


Figure 71: Masses in the stellar graveyard (in units of solar mass). The figure shows inferred gravitational masses of neutron stars and black holes from electromagnetic (EM) and gravitational-wave observations (LIGO-Virgo-KAGRA). Arrows connect two merging compact objects and their merged remnant as seen by gravitational-wave emissions. Visualization credits: LIGO-Virgo-KAGRA / Aaron Geller / Northwestern.

magnetic stars could be the progenitors of the strongest magnets in the Universe: highly magnetic NSs, which are called "magnetars."

When mass transfer becomes unstable (Figure 70, right), one star can grow to such a large radius that it completely engulfs its companion, and a common-envelope phase begins. The companion feels a strong drag force (just like a raindrop falling through Earth's atmosphere), decelerates, and quickly spirals deeper into the envelope. This inspiral releases energy, which might be sufficient to eject the entire envelope. As a result, a much closer binary is formed. Following a supernova explosion, further mass transfer is possible, during which mass accretion onto the NS remnant emits X-ray radiation (X-ray binary). When the secondary star also ends its life, a binary consisting of an NS and a white dwarf (WD) might be formed.

Neutron-star- and black-hole formation in binary stars

Almost all masses of NSs and BHs are measured in binary-star systems (see "stellar graveyard" in Figure 71). Such binaries evolve analogously to the exemplary pathways shown in Figure 70 – that is, the progenitor systems all undergo some sort of mass exchange. The masses of NSs come mostly from double NSs, NSs+WDs, or X-ray binaries, while the masses of local BHs come from nearby X-ray binaries (the "electromagnetic" [EM] BHs in Figure 71). Most BH masses today are inferred from gravitational-wave-merger events. In particular, the first-born compact object in binary stars with an NS and/or a BH always stems from stripped-envelope stars (see also Figure 70).

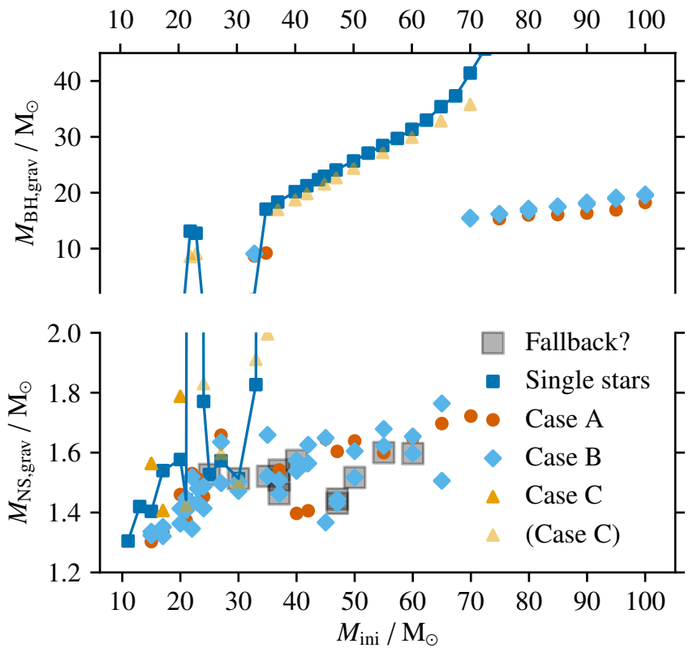


Figure 72: Neutron-star- and black-hole masses formed by stars that have lost their outer, hydrogen-rich envelopes in mass-transfer phases as a function of the initial mass of the stars. Results for unperturbed single stars are also shown. Cases A, B, and C correspond to envelope stripping in different evolutionary phases. Adapted from Schneider et al. 2021, A&A, 645, 5.

A few interesting features can be immediately derived from Figure 71. First, there appears to be a dearth or gap at 2–5 solar masses: the so-called NS–BH mass gap. Second, no nearby BH has been observed that is more massive than about 20 solar masses, yet most BHs seen from gravitational-wave-merger events appear to be more massive.

In Schneider et al. 2021, A&A, 645, 5, we investigated – inter alia – how the mass of NSs and BHs is affected by envelope stripping in mass-transfer episodes of binary stars and found that this envelope stripping indeed greatly modifies the interior structure of stars at the pre-supernova stage. Stripped-envelope stars produce NSs over a much-wider initial-mass range than do single stars that have not lost their envelopes (Figure 72). Neutron stars are formed for initial masses of <35 solar masses in single stars and for initial masses of <70 solar masses in stripped stars. Moreover, the BH masses of stripped stars are naturally reduced due to the envelope loss.

As a direct consequence, we predicted a maximum BH mass for Galactic X-ray binaries of about 20 solar masses, which corresponds nicely to the above-mentioned maximum EM BH mass found in the high-mass X-ray binary Cygnus X-1. We also showed that the NS–BH mass gap is set not by stellar evolution, but by details of successful and failed supernova explosions.

The Great Eruption of Eta Carinae

In the 1840s, the star Eta Carinae erupted and – for a few years – became the second-brightest star in the night sky. This Great Eruption created the intricate Homunculus Nebula (Figure 73). To date, however, what exactly caused this eruption remains unknown.

Today, Eta Carinae – with a mass of about 100 solar masses – is one of the most massive stars in the Milky Way Galaxy. It is located some 7,500 lightyears away and has a ~30-solar-mass binary companion in a wide, eccentric orbit of about 5.5 years. The Homunculus Nebula itself is also very massive (10–20 solar masses) and expands with a velocity of several 100 km/s, which implies a kinetic energy comparable to that of supernova explosions.

Together with collaborators from Australia, the UK, and the US, we combined a series of detailed simulations to show that a stellar-merger scenario in an initial triple-star system can explain many of the observable features of Eta Carinae (Hirai et al. 2021, MNRAS, 503, 4276; Figure 74, next page). In this scenario, an inner binary with stars of 60 and 40 solar masses is orbited by an outer tertiary star of 50 solar masses. If the outer orbit is about 5–10 times larger than the inner orbit, triple systems are known to be stable and are called "hierarchical triples." Once the

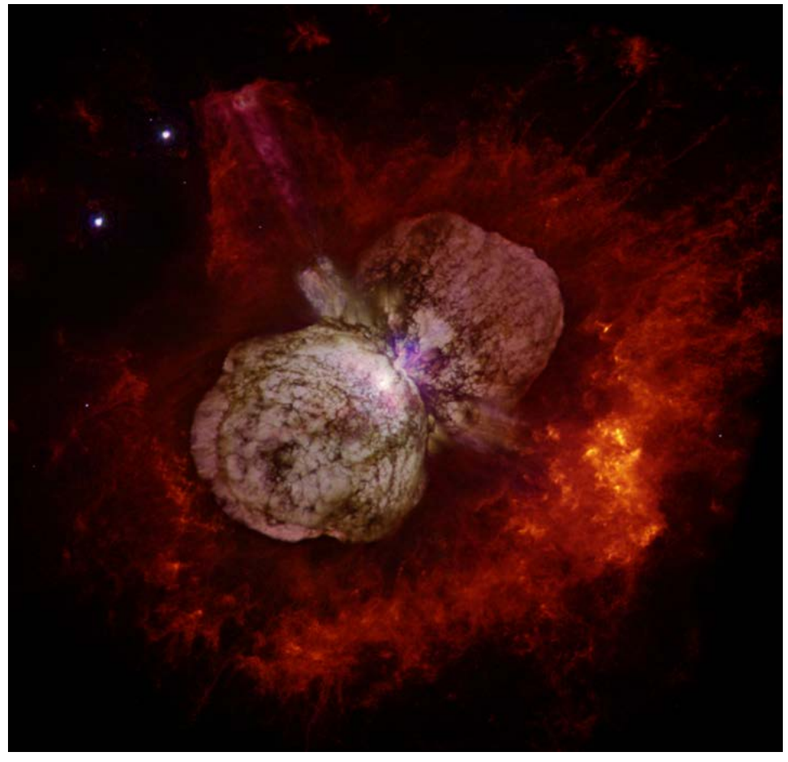


Figure 73: Hubble Space Telescope image of the Homunculus Nebula with the very massive star Eta Carinae hiding in the center. Credits: Nathan Smith (University of California, Berkeley) and NASA.

60-solar-mass star begins to transfer mass to its 40-solar-mass companion, the orbit of the inner binary widens and renders the triple system unstable. Chaotic dynamics set in with close flybys and partner swaps until one of the encounters is so close that the two stars merge. In the merger, an enormous amount of energy is released. Indeed, it was just such an energy release that created the Homunculus Nebula, in which the merged star itself became very massive. This star has not yet fully recovered from the merger event. In fact, it is still radiating away some of the merger energy at an increased rate and drives a very strong stellar wind, as revealed by observations. This wind is responsible for sweeping up merger ejecta into the hollow Homunculus Nebula, which we see today.

Our model can also explain a large set of other observations, such as the very high velocities of some of the ejecta (up to 10,000 km/s); "sprays" surrounding the nebula that originate from close encounters during the time when the triple was evolving chaotically; the dense, equatorial "skirt" of the nebula; and possibly also the X-ray emissions produced when fast ejecta collide with previously lost material. Future observations will further put our model to the test and yield new insights.

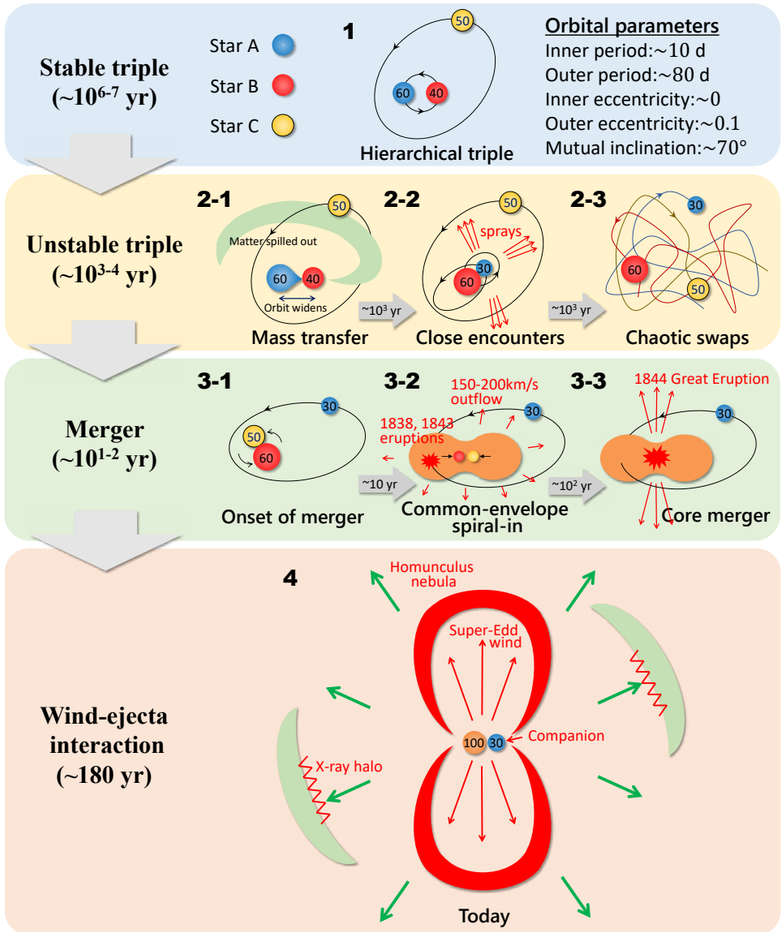


Figure 74: Schematic picture of the formation of the Homunculus Nebula from a stellar merger in an original triple-star system that evolved into an unstable orbital configuration. Adapted from Hirai et al. 2021, MNRAS, 503, 4276.

Sterne sind die elementaren Bausteine des sichtbaren Universums und produzieren fast alle chemischen Elemente, die schwerer als Helium sind. Seit jeher beschäftigt sich die Astrophysik mit der Frage, wie sich unser Universum seit dem Urknall in seine heutige Gestalt verwandelt hat.

Dabei spielen massereiche Sterne eine besondere Rolle, da sie kosmische Kraftwerke sind. Sie können teilweise mehrere Millionen Mal heller sein als die Sonne, haben starke Sternwinde und explodieren in gewaltigen Supernovae. Dank dieser Eigenschaften haben massereiche Sterne dazu beigetragen, nach den kosmischen „Dark Ages", das Licht ins Universum zurückzubringen, die Evolution von Galaxien voranzutreiben und den Grundstein für das Leben zu legen, wie wir es heute kennen.

Am Ende ihres Lebens hinterlassen massereiche Sterne einige der exotischsten Formen von Materie: Neutronensterne und Schwarze Löcher. Die Untersuchung dieser Überbleibsel ermöglicht Einblicke in Materieformen, die so auf der Erde nicht verfügbar sind. Die Verschmelzungen von Neutronensternen und Schwarzen Löchern werden mittlerweile routinemäßig von Gravitationswellenobservatorien beobachtet und bieten neue Einblicke in unser Universum.

Heute wissen wir, dass die meisten massereichen Sterne mit einem oder sogar mehreren Begleitern in Doppelstern- bzw. Mehrfachsystemen geboren werden, was zu interessanten Konsequenzen führt. Wenn Sterne altern, werden sie größer und können schließlich zu Riesen mit Radien von bis zum 1000-fachen unserer Sonne anwachsen. Doppelsterne können dadurch ein Stadium erreichen, in dem ihre äußeren Schichten auf ihren Begleiter übertragen werden. Bei etwa 25% der massereichen Sterne wird dieser Massenaustausch instabil und führt zu einer Verschmelzung beider Sterne. Der Massenaustausch im Allgemeinen und Sternverschmelzungen im Speziellen haben einen grundlegenden Einfluss auf die Entwicklung der Sterne sowie ihr letztendliches Schicksal. Wenn beispielsweise ein Stern bei der Massenübertragung seine Hülle verliert, kann er in einer Supernova explodieren und einen Neutronenstern produzieren, anstatt in ein Schwarzes Loch zu kollabieren.

Die **Stellar Evolution Theory Gruppe (SET)** untersucht das turbulente und explosive Leben massereicher Sterne. Derzeit konzentriert sich die Gruppe auf massereiche Doppelsternsysteme und deren Verschmelzungsprozesse. Diese Verschmelzungen erzeugen starke Magnetfelder und können zu stark magnetisierten Neutronensternen führen. Diese sind auch unter dem Namen Magnetar bekannt und besitzen die stärksten Magnetfelder im ganzen Universum.

2 Research

2.13 Theory and Observations of Stars (TOS)

Group leader
Prof. Dr. Ir. Saskia Hekker

Staff members
Lynn Buchele (since October 2021)
Quentin Coppee (since October 2021)
Daria Mokrytska
Dr. Nathalie Themeßl (SFB 881; since February 2021)

MSc students
Julian Schlecker
Teresa Braun (since May 2021)
Alba Covelo Paz

Internship student
Jacob Kosowski

Stars are an important source of electromagnetic radiation in the Universe that allow us to study many phenomena ranging from distant galaxies to the interstellar medium and to extra-solar planets. However, due to their opacity, Sir Arthur Eddington once said that "at first sight it would seem that the deep interior of the Sun and stars is less accessible to scientific investigation than any other region of the universe" (Arthur S. Eddington: "The Internal Constitution of the Stars," 1926). Now, through modern mathematical techniques and high-quality data, it has become possible to directly probe and study the internal stellar structure through global stellar oscillation – a method known as "asteroseismology." Asteroseismology uses similar techniques to helioseismology – which is carried out on our closest star, the Sun – to study the structure of other stars. The properties of waves are used to trace a star's internal conditions. Oscillations that impact upon the whole star reveal information that is hidden by the star's opaque surface. This asteroseismic information from the CoRoT, Kepler, K2, TESS, SONG, and Plato space observatories – combined with astrometric observations from Gaia, spectroscopic data from SDSS-V APOGEE, interferometry, photometry, and state-of-the-art stellar models, such as MESA – provides insights into the stellar structure and the physical processes that take place in stars.
Understanding the physical processes within stars and how these processes change as a function of stellar evolution is the ultimate goal of the Theory and Observations of Stars (TOS) group at HITS, which was established in 2020. We focus on – but do not limit ourselves to – low-mass main-sequence stars, subgiants, and red giants. These stars are interesting as they go through a series of changes to their internal structure. Furthermore, they are potential hosts of planets and are standard candles for Galactic studies (e.g., studies on core-helium-burning red giants). As a result, exoplanet studies as well as Galactic archaeology also benefit from an increased understanding of these stars.

Background

In the TOS group, we focus on stars with oscillations similar to those of the Sun. These so-called "solar-like oscillations" have a low amplitude and are stochastically excited through turbulence in the near-surface convection layer of a star. The oscillations are sound waves and are expected to be present in all stars with convective outer layers. A convective envelope is typically present in low-mass main-sequence stars, in subgiants, and in red giants with surface temperatures below ~6,700 °K.

The stellar structure is imprinted in the global oscillation modes of a star. An oscillation mode is uniquely determined by the properties of the matter through which it travels and is described by its frequency (or period) and mode identification – that is, its radial order (the number of nodal lines in the radial direction), its spherical degree (the number of nodal lines on the surface), and its azimuthal order (the number of nodal lines that cross the spin axis). In evolved, so-called "red giants," the dipole (spherical degree = 1) modes are sensitive to both the deep interior and the outer layers – that is, the oscillations resonate in an inner (gravity) and an outer (acoustic) cavity, which are separated by an evanescent zone (the area between the cavities where oscillations cannot propagate and decay exponentially; see Figure 75). The coupling between the two oscillating cavities and the phases of the waves in each cavity can be derived from the resulting mixed pressure–gravity oscillations and can provide information on the physical conditions in the evanescent region. Furthermore, the difference in period between pure gravity dipole modes with consecutive radial orders (so-called "period spacing," which can be extracted from mixed dipole modes) provides a measure of the extent of the gravity-mode cavity and thus of the properties of the stellar core. Determining these values and understanding the physical processes in these deep parts of stars is one of the main aims of the TOS group.

Ongoing work

Extracting oscillation features
In order to extract the oscillations, we work in the Fourier space, in which stellar oscillations reveal themselves as peaks. Due to their stochastic driving mechanism, solar-like oscillators have peaks with a width that depends on their lifetimes. Identifying these peaks from the noise is therefore not trivial. We developed a peak-detection method based on a Mexican-hat wavelet that automatically and reliably determines and characterizes the oscillation signals in terms of frequencies, amplitudes, and lifetimes (see Garcia Saravia Ortiz de Montellano, Hekker, Themeßl, "Automated asteroseismic peak detections," MNRAS 476, 1470, 2018).
We also continue to develop ways of characterizing the modes in terms of their radial order, their spherical degree, and their azimuthal order. A full code that we plan to release – TACO (Tools for the Automated Characterisation of Oscillations) – is currently in its final stages of development.

Rotation
Given that the interior and exterior of stars rotate, the mixed dipole modes always carry a signature of this rotation. These dipole modes are split into several components depending on the orientation of the star. If the rotation rate is high enough, this splitting can be observed. A significant part of the rotationally split mixed dipole modes – and even of the most pressure-dominated ones – is sensitive to the core (Beck et al., "Fast core rotation in red giants as revealed by gravity-dominated mixed modes," Nature 481, 55, 2012). Therefore, the core rotation rate can be determined via the mixed dipole modes (Mosser et al., "Spin down of the core rotation in red giants," A&A 548, A10, 2012, and Gehan et al., "Core rotation braking on the red giant branch for various mass ranges," A&A 616, A24, 2018). However, in order to obtain the radial rotation profile, the envelope rotation rate is an essential input. Currently, the envelope rotation rate has only been derived for about 20 red giants (Aerts, Mathis, Rogers, "Angular momentum transport in stellar interiors," ARAA 57, 35, 2019, and the references therein).

In order to confirm and update results in the literature, we determined asteroseismic rotation rates for red-giant-branch stars. Parallel to this effort, we began investigating the accuracy and precision of the rotation rates obtained from rotation inversions (publication in preparation), and together with collaborators, we proposed an update to these rotational inversion techniques in order to be able to obtain not only core rotation rates, but also surface rotation rates. These rates will prove crucial in furthering our understanding of angular momentum transport (Ahlborn et al. submitted).

Red-giant bump

We aim to further our understanding of what happens during the red-giant bump, which is a short phase in stellar evolution during which changes in surface temperature and brightness retract before continuing along their paths. This feature is well known in models as well as in observations; however, the physical mechanism behind it remains unknown, and the location of the bump is not the same in models and observations. In 2020, we published a paper (Hekker et al. "Mirror principle and the red-giant bump: the battle of entropy in low-mass stars," MNRAS 492, 5940, 2020) in which we proposed a physical mechanism that could cause the red-giant bump. We now aim to check whether the developed toy model indeed works and whether it can lead to a better understanding of the physics behind the event.

SFB 881: "The Milky Way System"

Since early 2021, the TOS group has participated in SFB 881: "The Milky Way System," in which we use asteroseismology to obtain accurate masses, radii, ages, and distances as well as cluster membership of individual cluster stars. Using these data, we improve the accuracy of the determined age and distance of the cluster when seen as one entity. We additionally use the clusters to update the reference values of the so-called "scaling relations" for stars in different evolutionary phases (Themeßl et al., in preparation).

Moreover, we are currently working on extracting asteroseismic time series data by analyzing the so-called "superstamp data" that Kepler took from two open clusters (NGC6791 and NGC6819), and we are performing aperture photometry on targets selected from observations made by the Gaia mission.

ERC Consolidator grant "DipolarSound"

The ERC Consolidator grant "DipolarSound" began in October 2021. Within this grant, we study stars that show different morphologies in their Fourier spectra (see Figure 75) and aim to answer the following questions:

- What are the physical differences in the structures of / conditions in red giants that lead to different oscillation spectra?
- What is the cause of the different structures / conditions in these stars?

The aim of the DipolarSound proposal is to unravel the physical conditions and physical processes at play in red giants using mixed dipole oscillation modes and to understand the underlying physical origins of the different oscillation spectra observed in red giants.

Within this framework, several work packages have been launched. First, we have begun to investigate "messy" stars by focusing on a set of stars for which period spacings can still be determined but are smaller than expected. This new class of stars will be announced in a forthcoming publication (Elsworth, Braun, Hekker, in preparation). Within the group, we are currently investigating the cause of these smaller-than-expected period spacings, which tentatively seem to be related to the coupling between resonance cavities in the star.

Furthermore, we have begun work on suppressed dipole mode stars in which we first verify whether the dipole modes are mixed modes or whether they are completely damped in the gravity-mode cavity. We will subsequently investigate this physical process in terms of high magnetic fields (Fuller J et al.: Asteroseismology Can Reveal Strong Internal Magnetic Fields in Red Giant Stars, Science (2015), 350, 423-426) or binarity (Themeßl et al.: Presence of mixed modes in red giants in binary systems, EPJ Web of Conferences 160, 05009 (2017)), the latter of which could potentially cause the lack of amplitude in these modes.

In order to further investigate the internal stellar structure, we are also developing inversion techniques that enable us to investigate what the stellar structure should be in light of the measured (oscillation) parameters. This technique has been applied and developed for the Sun with great success and has been extended to solar-like stars in recent years. We now aim to bridge the research gap by adapting the technique in order to be able to invert more evolved stars.

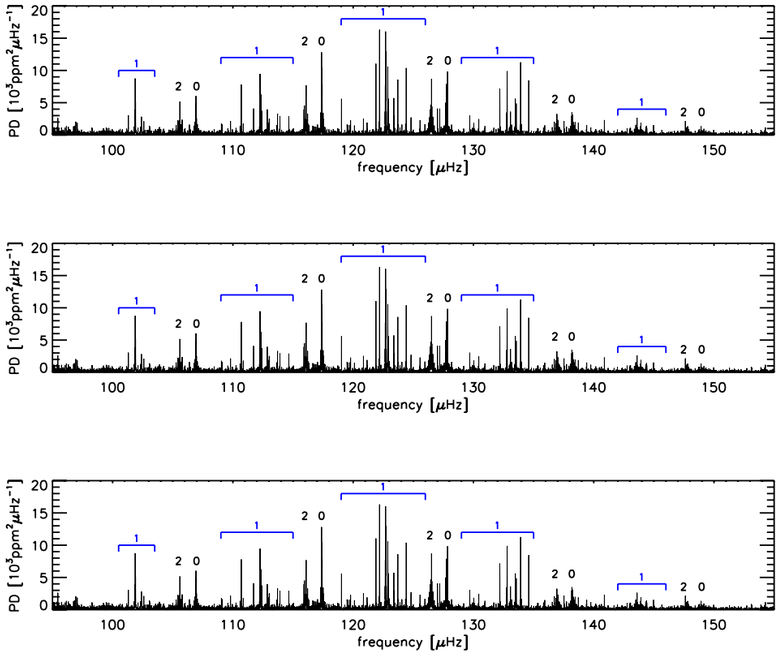


Figure 75: Power-density spectra of three red giants observed by the Kepler space mission in which dipole (degree = 1) modes are suppressed (top), have expected amplitudes and regularities (middle), and are "messy" (bottom). The numbers in each panel indicate the degree (i.e., the number of nodal lines on the surface) of the oscillation modes.

Sterne sind eine wichtige Quelle elektromagnetischer Strahlung im Universum, mit der viele Phänomene untersucht werden können, von fernen Galaxien über das interstellare Medium bis hin zu Exoplaneten. Aufgrund ihrer Undurchsichtigkeit hat Arthur Eddington einmal gesagt, dass „auf den ersten Blick das tiefe Innere der Sonne und der Sterne für wissenschaftliche Untersuchungen weniger zugänglich zu sein scheint als jede andere Region des Universums" (Arthur S. Eddington: "Der innere Aufbau der Sterne", 1926, dt. Ausgabe 1928). Durch moderne mathematische Methoden und die Menge und Qualität verfügbarer Daten ist es nun jedoch möglich geworden, die innere Sternstruktur direkt durch Sternschwingungen zu erforschen: eine Methode, die als Asteroseismologie bekannt ist.

Die Asteroseismologie verwendet ähnliche Techniken wie die Helioseismologie, die an unserem nächstgelegenen Stern, der Sonne, durchgeführt wird, um die Struktur anderer Sterne zu untersuchen. Hierzu werden die Eigenschaften von Wellen verwendet, um Rückschlüsse auf die innere Beschaffenheit von Sternen zu ziehen. Schwingungen, die auf den ganzen Stern einwirken, enthüllen so Informationen, die durch die undurchsichtige Oberfläche normalerweise verborgen sind. Diese asteroseismischen Informationen der Weltraumobservatorien wie CoRoT, Kepler, K2, TESS, SONG und Plato kombiniert mit astrometrischen Beobachtungen von Gaia, spektroskopischen Daten von SDSS-V APOGEE, Interferometrie, Photometrie und hochmodernen Sternmodellen wie MESA, geben Einblicke in die Sternstruktur und die physikalischen Prozesse, die in Sternen ablaufen.

Das Ziel der **Theory and Observations of Stars (TOS)** Forschungsgruppe am HITS, die 2020 ihre Arbeit aufnahm, ist die Untersuchung dieser physikalischen Prozesse, die in Sternen ablaufen, und wie sich diese in Abhängigkeit von der Sternentwicklung verändern. Die Gruppe konzentriert sich hierbei unter anderem auf sogenannte Hauptreihen-Sterne geringer Masse, „Unterriesen" und rote Riesensterne. Diese Sterne sind deshalb interessant, weil sich ihre innere Struktur schnell ändert. Da sie potenziell von Planeten umgeben und kosmologische „Standardkerzen" für Galaxienstudien sind, können sowohl die Exoplanetenforschung als auch die Galaktische Archäologie vom wachsenden Verständnis dieser Sterne profitieren.

3 Centralized Services

3.1 Administrative Services

The HITS administration serves the Institute's groups in almost all necessary administrative processes. It takes care of HR support, operating offices and buildings, making purchases, and settling invoices, and it also supports the Communications team in organizing events. Moreover, the administration additionally ensures that legal issues are resolved and that all processes at the Institute comply with legal requirements.

HITS grew significantly in 2021, as is particularly evident in the group of doctoral students, which increased from 25 individuals at the end of 2020 to 45 at the end of 2021. The reasons for this growth were (1) that two new groups were established – TOS and SET – and (2) that HITS was highly successful in attracting external funding. Indeed, several large projects were launched in 2021, such as Saskia Hekker's ERC Consolidator Grant (see Chapter 2.13), Frauke Gräter's ERC Consolidator Grant (see Chapter 2.8), the AIN group's work with the support of an ERC Synergy Grant (see chapter 2.1), and the SDBV group's work with both the National Research Data Infrastructure for Personal Health Data (NFDI4Health) and the BMBF LiSyM-Cancer project (see Chapter 2.11).

Group leader

Dr. Gesa Schönberger

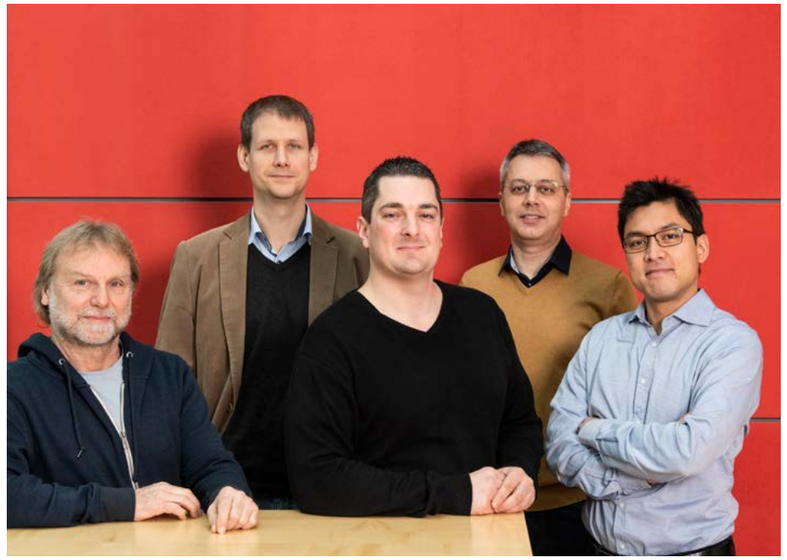
Staff members

Yashasvini Balachandra (Controlling; since November 2021)
Christina Blach (Office)
Frauke Bley (Human Resources)
Christina Bölk-Krosta (Controlling)
Benedicta Frech (Office)
Silvia Galbusera (Human Resources; since January 2021)
Ingrid Kräling (Controlling)
Dr. Barbara Port (Scientific Manager; since June 2021)
Thomas Rasem (Controlling)
Rebekka Riehl (Human Resources and Assistant to the Managing Director)
Stefanie Szymorek (Human Resources; until March 2021)
Irina Zaichenko (Accounting)

The new SIMPLAIX project (see Chapters 2.8 and 6) – which involves cooperation between KIT, Heidelberg University, and HITS – also began in 2021. With SIMPLAIX, not only are three HITS groups involved, but HITS also coordinates the project. For the administration, more groups and more third-party funding mean more HITSters to supervise and a larger budget to manage. However, the administration has only changed slightly in terms of personnel: To relieve the workload of the Scientific Director, we were able to recruit a Scientific Manager. There was also a change in the HR team at the beginning of the year, and a new colleague joined the Controlling team at the end of 2021. As a rule, more employees also require more workplaces. However, in the second year of the pandemic, this was not the case. On the contrary: In line with state- and federal requirements, we mostly had to manage with half-staffed offices. However, productivity remained unchanged.

A key project involved converting from SAP R3 to SAP S4HANA at our billing company KTA e. K. The conversion required close cooperation – especially with HITS Controlling – in order to meet the extensive requirements of project- and third-party funding. The new software will make many things easier that had previously required "manual work" and will significantly change administrative processes: In addition to controlling and accounting, this change will also affect reporting and planning, travel requests, and accounting as well as management for inventory, hardware, and software. The changeover took place on 1 January 2022.

Along with these fundamental administrative reforms, other changes are currently in the pipeline, including plans to overhaul the ordering process. Overall, we aim to achieve greater effectiveness both by achieving an even higher degree of digitization – which will also facilitate work under pandemic conditions – and by streamlining internal processes.



Group leader

Dr. Ion Bogdan Costescu

Staff members

Julian Aris | Student
Dr. Bernd Doser | Senior Software Developer
Dr. Simon Kreuzer | System Administrator (since February 2021)
Norbert Rabes | System Administrator
Andreas Ulrich | System Administrator
Taufan Zimmer | System Administrator

3.2 IT Infrastructure and Network

In many respects, 2021 was a continuation of the previous year. The ongoing COVID-19 pandemic strongly influenced the way everyone at HITS worked, with videoconferencing and remote work now being the norm. Similar to the previous year, the ITS group members combined on-site and remote work in order to maintain the hardware infrastructure, while also observing mandated distancing and hygiene measures.

During most of the year, we continued to refine the setup of the HPC cluster and storage system, which were installed in 2020. This process was interrupted in the autumn by a failure of the older storage system, when a hardware defect led to file system corruption. The broken device could not be properly replaced until the end of the year due to the worldwide manufacturing and delivery difficulties. In the meantime, we set up an alternative device and worked with the company developing BeeGFS to bring the file system back in order and to minimize the amount of lost scientific data. This was no easy task given that the storage system holds more than 2.5 PB of data. In the end, only around 200 files were lost or corrupted out of a total of more than 300 million files.

Since 2013, HITS has used the KIT archive system for long-term storage of scientific data. Due to significant changes to the system at KIT, we decided to begin using the archive system at Heidelberg University's Computing Centre. This latter system employs the same software used previously at KIT and is "closer" to HITS in network topology, thereby leading to higher data transfer rates. The archives from KIT had been transferred back to HITS during the previous year. In the second half of 2021, we began testing and finally transferring the data to the new system.

The number of HITSters continued to grow, in part due to the launch of two new scientific groups (TOS and SET) at the end of 2020 and the beginning of 2021. This led – inter alia – to an increased usage of the IT systems that support remote work, some of which required capacity adjustments. Additionally, several videoconferencing systems were purchased in order to enhance online meetings – especially hybrid setups with multiple people on-site. Despite these and further improvements to the resources we use for working remotely, we are very much hoping that 2022 will bring with it an end to the pandemic and allow us all to meet in person again.

4 Communication and Outreach

HITS Communications team in 2021(f.l.t.r.): Olexandr Golovin, Adrian Stoll, Isabel Lacurie, Peter Saueressig, Angela Michel.

Head of Communications
Dr. Peter Saueressig

Staff members
Isabel Lacurie
Angela Michel

Students
Olexandr Golovin (until June 2021)
Adrian Stoll (since July 2021)

The HITS Communications team is the Institute's central hub for external and internal communications. Its main tasks are to raise the profile of HITS by coordinating media relations, digital and social media communications, and the Institute's publications, design, and branding as well as by organizing events for the scientific community, such as conferences and workshops. Moreover, we strive to spark enthusiasm for science among school students and the general public alike through our outreach activities. Our 2021 report illustrates how we have adapted to the special situation caused by the pandemic, which has continued to challenge us.

Science, awards, and the media

A research institute´s communication is highly dependent on its researchers and their scientific success, without which, communicators do not have much to say. In 2021, the Communications team was again pleased to announce success stories to the public.
To begin, group leader Anna Wienhard (GRG, see Chapter 2.6) was awarded an ERC Advanced Grant by the European Research Council – a great success and a clearly visible sign of the Institute´s close cooperation with Heidelberg University. Anna´s research topic plays a central role in many fields of mathematics and is an important tool in theoretical physics. With this new award, a total of four ERC grantees were working at HITS by the end of 2021.

Another HITS ERC grantee – Astrophysicist Fabian Schneider, head of the SET group (see Chapter 2.12), who had earned his award in 2020 – received the renowned Ludwig Biermann Award of the Astronomische Gesellschaft (AG) in 2021. The Ludwig Biermann Award is granted in recognition of outstanding young astronomers. The AG stated that Fabian "is an internationally recognized expert in his field."

Last but not least, according to this year's Highly Cited Researchers list from Clarivate, computer scientist and CME group leader Alexandros Stamatakis (see Chapter 2.3) was named one of the most-cited researchers worldwide for the sixth year in a row. The ranking is an important indicator for the impact of a researcher's scientific publications.

HITS researchers were also successful in developing new methods, tools, and projects. The CST group developed a new approach that enriches the toolbox of forecast evaluation. In a PNAS paper, the researchers introduced the so-called "CORP" diagram, which yields improved graphical tools that are more stable and efficient than the classical type of diagram, which has been used for decades (see Chapter 2.4).

Together with colleagues from the KIT, the CST team also refined the German–Polish COVID-19 Forecast Hub and evaluated different forecast models in a systematic study, which was published in Nature Communications. The result: Combining different models leads to better predictions (see Chapter 2.4).

Together with tumor biologists from Heidelberg University, Heidelberg University Hospital, and the German Cancer Research Center (DKFZ), the DMQ group began a new interdisciplinary project that uses computer simulations to model the biological processes behind the initial steps of colon cancer development. The aim is to understand tumor initiation, evolution, and immunology by means of mathematical modeling. The scientists work on approaches that could help prevent hereditary cancer in the future. The project is funded by the Klaus Tschira Foundation (see Chapter 2.5).

Moreover, HITSters also delivered radio and TV interviews on different topics. Sometimes, they even presented unusual topics, as with the CME group, which predicted the outcome of the European Football Championship via a phylogenetic algorithm (see Chapter 2.3). The forecast resonated strongly with the media in the form of newspapers, radio, and TV stations, and CME group leader Alexandros Stamatakis had the chance to explain the algorithm and the scientific background to a non-scientific audience in a live interview on the Sky TV channel.

(Virtual) events, productions, and pop-up science

2021 was again not conducive to on-site events. In fact, HITS saw only one workshop in early October (see Chapter 5.1.2) and two hybrid events in the Studio Villa Bosch, each with only a small number of participants present. All other activities had to be prepared and organized virtually, including the long-awaited HITS Alumni Meeting, which was held on a newly installed interactive platform (see Chapter 5.3), and the internal HITS Fest, which included one dozen different round tables (see Chapter 5.4). For 2001's International Summer Science School (ISH) in early August, HITS offered an online workshop on "When News Goes Fake," prepared and held by Angela Michel in two different sessions (depending on the time zone) and with 17 participants from three continents.

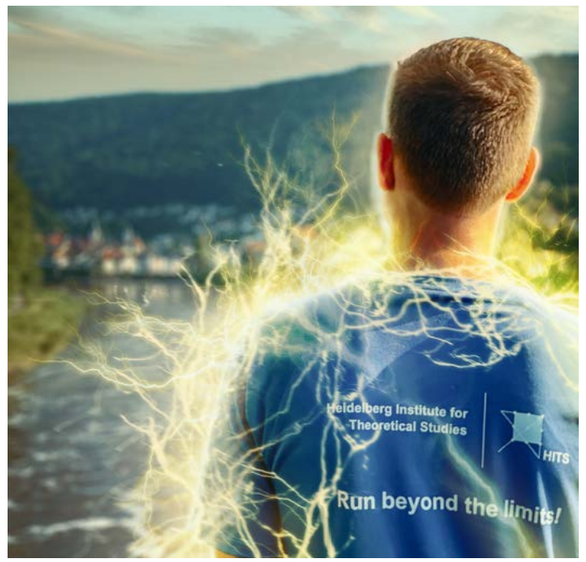
Kai Polsterer and Angela Michel hosted the virtual lab visit of the Humboldt Foundation. The event was broadcast directly from the Klaus Tschira Library.

However, the fact that other organizations also had to switch to the digital sphere offered new opportunities for HITS to participate, as with the Annual Meeting of the Humboldt Foundation, at which more than 1,000 participants from nearly 80 countries gathered on 23 and 24 June in a virtual space under the motto “Diversity of Ideas.” Whereas former in-person meetings had been organized in Berlin, in 2021, the Humboldt Foundation offered virtual lab visits to 13 institutions across Germany, which included HITS. Kai Polsterer (AIN) and moderator Angela Michel welcomed 27 “Humboldtians” via live stream from the Klaus Tschira library on the HITS premises. To provide the visitors from all over the world with a comprehensive introduction to the Institute, we produced a video trailer that comprised a virtual tour of the HITS campus and a portrait of Kai Polsterer and his research. Speaking of video productions: We organized the video shoots of two movies with young researcher Svenja de Buhr (MBM) for the Cluster of Excellence “3D Matter Made to Order” (3DMM2O), which was made on the HITS campus. Moreover, during the summer, we shot several scenes for a new HITS image video that is scheduled to be finished in 2022.

The annual NCT Run – an established charity run organized by the National Center for Tumor Diseases (NCT) in Heidelberg that gathers donations for cancer research – was again organized as a virtual three-day event in early July. A 29-member HITS running team again participated in the run, which attracted more than 7,000 runners worldwide for its 10th anniversary under the motto “Alone. Together! 10 Years of NCT RUNning against cancer.” The team “The HITSters” completed almost 800 km in four different countries, thereby contributing its fair share to the total of 125,000 kilometers. The combined efforts resulted in €50,000 in donations for cancer research at the NCT in Heidelberg. The HITS runners came from 10 research groups, from the Administration, and from the complete Communications team and also included alumni, family, and friends.

To document the event, all runners were asked to send a picture of themselves in their HITS running shirt while striking a superhero pose. The best pictures were professionally edited by a photographer and fantasy artist and put into the HITS Heroes Hall of Fame.

Shortly before the end of the year, HITS was part of a planned astronomy outreach event: The city of Heidelberg rented the vacant rooms of a former gallery in the historical center of town for temporary use. Under the motto “Science in the City,” several research institutes and initiatives contributed to this pop-up science exhibition, with the “Haus der Astronomie” (“House of Astronomy”) being responsible for the organization and coordination.
In addition to the history of astronomy in Heidelberg, the event is set to include many posters, exhibits, and videos that present the participating institutes and their research. The exhibition opening had originally been planned for early December but had to be postponed due to the current situation. Visitors could, however, check it out from the outside and watch the prepared astronomy videos, which include the birth of a magnet star by the SET group.



The HITS Heroes Hall of Fame, with some of the participants in the 2021 NCT run pictured (edited by Nathalie Le Fay).

The outlook: A silver lining of hope

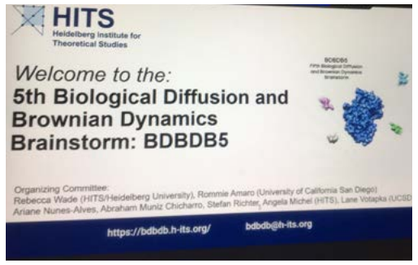
We hope in 2022 to again have the chance to organize in-person events– above all the Open House event in July – and to again welcome guests on site. The first such guest will be Carl Smith from Sydney (Australia), our 10th Journalist in Residence, who will arrive at HITS in January and stay until June. The "HITS Journalist in Residence" program offers science journalists a paid sojourn at the Institute. During their stay, these journalists can learn more about data-driven science and get to know researchers and new research topics without the pressure of the “daily grind.” We firmly believe that it is crucial to support independent science journalism and to establish reliable and sustainable journalistic contacts as an important prerequisite for successful science communication.

To date, journalists from India, the U.S., Canada, Spain and Germany have participated in the HITS fellowship — which started in 2012 with the award-winning German science journalist Volker Stollorz. In the meantime, Stollorz became chief editor of the German Science Media Center in Cologne, a non-profit organization run by journalists that helps journalists with their reporting.

5 Events

5.1 Conferences, Workshops & Courses

5.1.1 BDBDB5 Virtual Online Meeting



24–25 March 2021

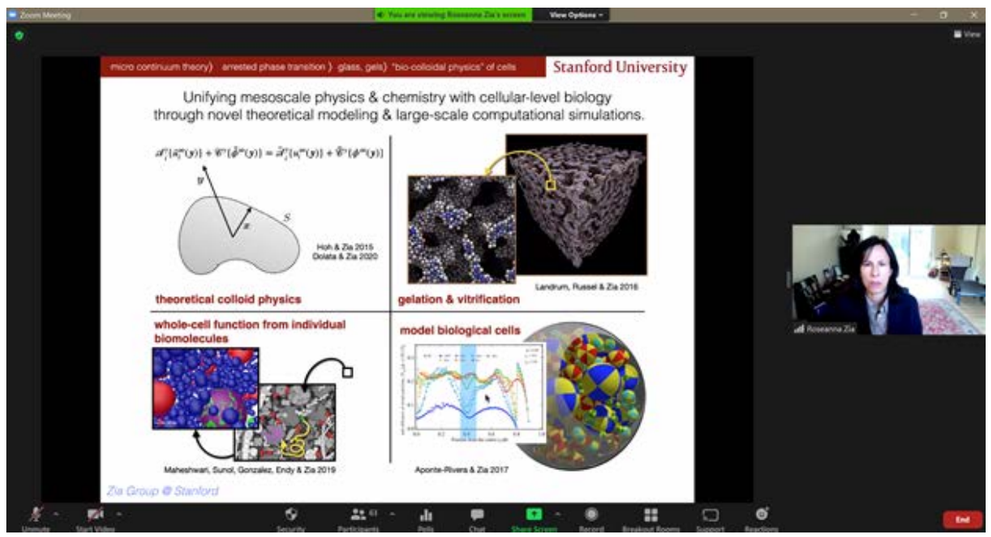
The goal of BDBDB5 (Biological Diffusion and Brownian Dynamics Brainstorm 5) was to provide a forum for presentations and informal discussions about the current state of experimental and theoretical studies of biological diffusion, with a focus on the Brownian Dynamics method for simulating biological macromolecules.

In contrast to the previous meetings in the BDBDB series that were held in in-person and hybrid formats, we decided to hold a shorter, fully virtual meeting that took place in two evening (in Heidelberg) sessions timed to enable participants from around the world to connect and participate. We had four very inspiring keynote lectures from our invited speakers: Aleksei Aksimentiev (University of Illinois at Urbana-Champaign, USA), Gerhard Hummer (Max Planck Institute for Biophysics, Frankfurt, Germany), Syma Khalid (University of Southampton, UK) and Roseanna Zia (Stanford University, USA).

In addition, there were contributed talks and discussion sessions on Brownian Dynamics and related molecular simulation methodologies, along with recent theoretical and experimental advances in studying diffusion from molecular to cellular levels. The workshop attracted over 120 registrations from theoreticians and experimentalists from around the world with 50-60 participants present at each session.

For more details, see: https://bdbdb.h-its.org/

Organizers:
Rebecca Wade (HITS/Heidelberg Univ), Rommie Amaro (University of California San Diego), Stefan Richter (HITS), Lane Votapka (University of California, San Diego, USA), Ariane Nunes-Alves (HITS), Abraham Muniz Chicharro (HITS)



Roseanna Zia (Stanford University, USA) delivering her virtual keynote speech.

5.1.2 Workshop on Forecast Calibration

HITS, Heidelberg, 25 October 2021

Probabilistic forecasts are issued in the form of probability distributions for future quantities or events. A forecast of this type is calibrated (or, equivalently, reliable) if the outcomes are statistically consistent with the predicted probability distributions.

The workshop participants during a break in the HITS garden.

This seemingly simple notion of forecast calibration entails major challenges both in its mathematical formalization and in the development of diagnostic tools for calibration checks. For instance, practitioners have used reliability diagrams to check the calibration of probability forecasts for binary events. Extant approaches to plotting reliability diagrams rely on binning and counting and have been hampered by ad-hoc implementation decisions, a lack of reproducibility, and inefficiency. With the new and recently published “CORP” approach, HITS researchers have introduced a way to better determine the reliability of forecasting methods (see Chapter 2.4).

This was one of the topics discussed in a mini workshop on forecast calibration organized by the Computational Statistics group (CST) on 25 October 2021 at the HITS campus. The workshop consisted of both a hybrid-format part and an in-person part.

Discussed topics included newly developed notions of calibration in terms of statistical functionals in both univariate and multivariate settings, the CORP approach, calibration tests, and case studies. The speakers were Alexander Henzi (University of Bern, Switzerland), Marc-Oliver Pohle (Goethe University Frankfurt, Germany), Johannes Resin, and Alexander Jordan (both HITS).

5.2 HITS Colloquia

Siobhan Roberts

Free journalist and author, Journalist in residence at HITS 2020/21
20 January 2021: Embracing the Uncertainties (Online)

Prof. Dr. Erin R. Johnson

Department of Chemistry, Department of Physics & Atmospheric Science, Dalhousie University, Halifax, Canada
25 January 2021: Dispersion Interactions in Density-Functional Theory and Application to Molecular Crystal-Structure Prediction (Online)

Prof. Dr. Lillian T. Chong

Department of Chemistry, University of Pittsburgh, USA
26 April 2021: Weighted ensemble simulations of long-timescale dynamics: From chemical reactions to SARS-CoV-2 (Online)

Alina Schadwinkel

Managing Editor Online of "Spektrum der Wissenschaft", Heidelberg / Germany
18 May 2021: Science journalism in Corona times (Online)

Prof. Dr. Alán Aspuru-Guzik

Department of Chemistry and Department of Computer Science, University of Toronto, Canada
28 June 2021: There is no time for science as usual: Materials acceleration platforms (Online)

Prof. Dr. Daniel Müller

Professor of Biophysics and current Head of the Department of Biosystems Science and Engineering at ETH Zürich, Switzerland
19 July 2021: Mechanically quantifying and guiding biological processes (Online)

Prof. Dr. Warner Marzocchi

Professor of Professor of Geophysics and Natural Hazard Forecasting at the University of Naples, Federico II, and at the Scuola Superiore Meridionale, Italy
18 October 2021: Model validation in natural hazard forecasting – a scientific perspective. Studio Villa Bosch, Heidelberg (hybrid)

Prof. Dr. Yuji Sugita

Chief Scientist, Theoretical Molecular Science Laboratory, RIKEN, Saitama, Japan
22 November 2021: Parallel computing algorithms in molecular dynamics simulations for extremely large-scale biological systems (Online)

5.3 HITS Virtual Alumni Meeting

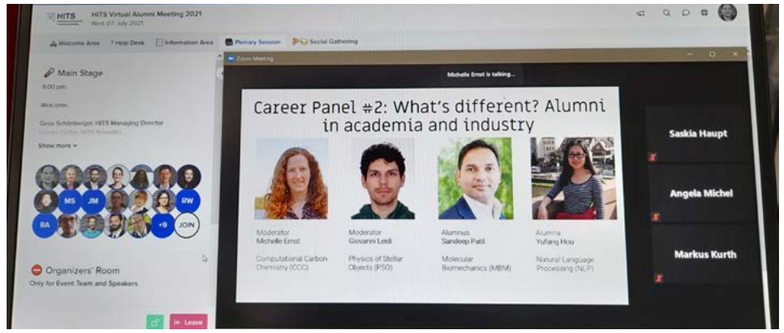
7 July 2021 (online)

Although the summer of 2021 appeared promising and to be better suited for in-person meetings, the decision was made to hold the HITS Alumni Meeting online both due to travel restrictions that prevented alumni from visiting Heidelberg and in order to avoid the risk of spreading the virus in a crowd of people meeting on the campus.

However, the online event also proved to have some benefits: Alumni who were not able to travel to Heidelberg could participate and meet former colleagues, regardless of the time zone in which they were living at the time. It turned out that 95 alumni and current HITSters had registered for the event on 7 July 2021.

To give the participants the chance to chat with current and former colleagues in different ways, the communications department provided an interactive platform, tailored it to the event, and adjusted the sessions and breaks to the digital format.

The meeting began in the "virtual coffee bar" with chats and interactive warm-up rounds. Then, Managing Director Gesa Schönberger and Scientific Director Frauke Gräter welcomed everyone, reported on new developments at HITS, and introduced the two new group leaders: Saskia Hekker (TOS) and Fabian Schneider (SET). After a break, two career panels took place with a focus on the topic of "What´s different?" Subsequently, Antonia Höhn (nee Stank; now Roche Penzberg, Germany, former MCM), Freeke van de Voort (Cardiff University, former TAP), Yufang Hou (now Dublin, former NLP), and Sandeep Patil (RWTH Aachen, Germany, former MBM) discussed the pros and cons of working in academia and in industry and talked about their respective careers in an event moderated by current HITSters Michelle Ernst (CCC), Giovanni Leidi (PSO), Saskia Haupt (DMQ), and Markus Kurth (MBM). The meeting concluded with a social gathering in several virtual rooms, which continued into the summer night.



The "main stage" of the virtual Alumni Meeting during the career panels.

The feedback on the first HITS virtual Alumni Meeting was quite positive. We all hope, however, that the next bi-annual meeting in 2023 will be able to take place on the HITS campus again!

5.4 Virtual HITS Fest

19 July 2021 (online)

The HITS Fest is an internal event for all HITSters. It is designed to bring together people from various groups and to support the interdisciplinary idea of the Institute. The program therefore includes both a scientific and a social element and is organized by a committee consisting of members from different groups and departments. The first installment of this event took place in the summer of 2018 and was held both in the Studio Villa Bosch and on the HITS campus. After several efforts for an on-site event had had to be postponed, the 2021 HITS Fest was finally held as a virtual event on 19 July. The program included a scientific colloquium led by Daniel J. Müller (ETH Zürich, Switzerland), an icebreaker session filled with wicked riddles, round-table discussions on various topics, and finally, a social gathering that included a movie night, a game night, as well as a pub quiz.

To wrap up the event, the round-table discussions as well as the slides that had been presented during the sessions were documented, as planned. The proceedings were edited and put on the HITS intranet by Isabel Lacurie (Communications) in order to allow all HITSters – including those who could not participate – to be informed of the results of the 2021 HITS Fest.

The scientific colloquium on mechanobiology held by Daniel J. Müller, Professor of Biophysics at ETH Zürich, Switzerland. Daniel Müller is also a member of the board of trustees of the Klaus Tschira Foundation.

6 Collaborations

Heidelberg Laureate Forum

The Heidelberg Laureate Forum (HLF) is a networking conference at which 200 carefully selected young researchers in mathematics and computer science spend a week interacting with laureates from these disciplines, including recipients of the Abel Prize, the ACM A.M. Turing Award, the ACM Prize in Computing, the Fields Medal, and the Nevanlinna Prize. Established in 2013, the HLF is held annually by the Heidelberg Laureate Forum Foundation (HLFF). Having initially served as a co-initiator of the event, HITS has been a scientific partner of the HLF since 2016. Moreover, HITS group leader Anna Wienhard has been Scientific Chairperson of the HLFF since 2020.

Anna Wienhard delivering a speech during the HLF. (© Heidelberg Laureate Forum Foundation)

The 8th HLF: Digital, dynamic, interactive

From 20–23 September 2021, the digital 8th HLF offered a diverse blend of traditional and novel components for the selected young researchers and the other participants alike. An effort was made to construct a program that unmistakably resembled the HLF yet that had also been undeniably optimized for an online format. The principal motivation in tailoring each element was to provide a space for effective interaction. Certain elements had been designed to simulate an in-person event, including laureate lectures and the Hot Topic hybrid session on "Mathematics of Disease: The Science of Epidemic Modeling," which was moderated by science journalist Martin Enserink (Science Magazine) in the Neue Aula (New Assembly Hall) of Heidelberg University, with two of the speakers participating on site.

At the virtual opening ceremony of the HLF in the Neue Aula of Heidelberg University (© Heidelberg Laureate Forum Foundation).

Other components were introduced, such as the laureate debate and speed-networking. Many elements had been optimized for direct interplay, and the entire program was held on a highly usable platform. HITS had set up a virtual booth from which Deputy Scientific Director Tilmann Gneiting welcomed interested guests. Leading up to the 8th HLF, a few workshops were designed to provide the young researchers with direct and tangible benefits for their careers.

Though the digital 8th HLF looked and felt very different from an in-person event, the fundamental principle remained the same: Some of the brightest minds in mathematics and computer science came together for interdisciplinary exchange. Video material from most of the sessions is available on the HLF YouTube channel.

SIMPLAIX members during the first meeting in the Studio Villa Bosch in October 2021.

SIMPLAIX

Traditionally, the investigation of molecular mechanisms and the rational design of molecules and materials for targeted applications have been guided by physics-based modeling and simulation. While this process has revolutionized modern science and technology, it remains necessary to map, explore, and analyze the infinitely complex and variable structure-property space of molecular systems across scales. This is where data-driven and machine-learning methodologies offer a promising approach. But the major obstacles to applying data-driven methods to atomistic systems – for example, dealing with complex three-dimensional structures and integration within multiscale simulation algorithms, to name just two – must first be resolved.

SIMPLAIX is a new 3-way inter-institutional collaboration between HITS, Heidelberg University, and the Karlsruhe Institute of Technology (KIT). It aims to pool the expertise of the three partners to address the challenge of bridging scales from molecules to molecular materials using multiscale simulation and machine learning. In SIMPLAIX, these methods are being developed and employed to study a set of challenging problems in biomolecules and molecular materials within 8 multidisciplinary, inter-institutional research projects. Examples of the topics studied include discovering the connection between collagen-related diseases and defenses against radical damage, introducing machine learning for supported organic electrode materials, speeding up classical simulations or quantum chemical computations via predictions from machine learning, and predicting where inorganic and organic materials will break under force.

SIMPLAIX is coordinated by HITS group leaders Rebecca Wade (MCM) and Frauke Gräter (MBM), who – together with group leader Ganna Gryn'ova (CCC) – are among its 8 Principal Investigators. The initiative was launched in October 2021 with an internal kick-off meeting. The official inauguration event will take place in April 2022 at the Studio Villa Bosch in Heidelberg. SIMPLAIX is funded by the Klaus Tschira Foundation and supported by in-kind contributions from KIT and Heidelberg University. During the initiative, 8 young researchers will work on SIMPLAIX projects.

Why is SIMPLAIX called SIMPLAIX?

A "simplex" is a generalized triangle, and this shape reflects the triangular (3-way, trilateral) collaboration between the 3 participating institutions.

SIM stands for SIMulation.
AI stands for Artificial Intelligence.
X represents the interdisciplinary eXchange between simulation, AI, molecules, and materials.

SIMPLAIX is pronounced /sɪm'pleɪ̯ks/ (IPA)

7 Publications

Addison BC, Wright DJ, Nicholson BA, Cale B, Mocnik T, Huber D, Plavchan P, Wittenmyer RA, Vanderburg A, Chaplin WJ, Chontos A, Clark JT, Eastman JD, Ziegler C, Brahm R, Carter BD, Clerte M, Espinoza N, Horner J, Bentley J, Jordán A, Kane SR, Kielkopf JF, Laychock E, Mengel MW, Okumura J, Stassun KG, Bedding TR, Bowler BP, Burnelis A, Blanco-Cuaresma S, Collins M, Crossfield I, Davis AB, Evensberget D, Heitzmann A, Howell SB, Law N, Mann AW, Marsden SC, Matson RA, O'Connor JH, Shporer A, Stevens C, Tinney CG, Tylor C, Wang S, Zhang H, Henning T, Kossakowski D, Ricker G, Sarkis P, Schlecker M, Torres P, Vanderspek R, Latham DW, Seager S, Winn JN, Jenkins JM, Mireles I, Rowden P, Pepper J, Daylan T, Schlieder JE, Collins KA, Collins KI, Tan T, Ball WH, Basu S, Buzasi DL, Campante TL, Corsaro E, González-Cuesta L, Davies GR, de Almeida L, do Nascimento J, García RA, Guo Z, Handberg R, Hekker S, Hey DR, Kallinger T, Kawaler SD, Kayhan C, S. Kuszlewicz J, Lund MN, Lyttle A, Mathur S, Miglio A, Mosser B, Nielsen MB, Serenelli AM, Aguirre VS, Themeßl N (2021). *TOI-257b (HD 19916b): a warm sub-saturn orbiting an evolved F-type star.* Monthly Notices of the Royal Astronomical Society 502(3):3704-3722

Aslam M, Kandasamy N, Ullah A, Paramasivam N, Öztürk MA, Naureen S, Arshad A, Badshah M, Khan K, Wajid M, Abbasi R, Ilyas M, Eils R, Schlesner M, Wade RC, Ahmad N, Engelhardt Jv (2021). *Putative second hit rare genetic variants in families with seemingly GBA-associated Parkinson's disease.* npj Genom. Med. 6(1),2

Banterle N, Nievergelt AP, Buhr Sd, Hatzopoulos GN, Brillard C, Andany S, Hübscher T, Sorgenfrei FA, Schwarz US, Gräter F, Fantner GE, Gönczy P (2021). *Kinetic and structural roles for the surface in guiding SAS-6 self-assembly to direct centriole architecture.* Nature Communications,12(1)

Baran Á, Lerch S, Ayari ME, Baran S (2021). *Machine learning for total cloud cover prediction.* Neural Computing and Applications, 33(7):2605–2620

Barbera P, Stamatakis A (2021). *Efficient Memory Management in Likelihood-based Phylogenetic Placement.* IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW),pp.218-227,IEEE

Bauswein A, Blacker S, Lioutas G, Soultanis T, Vijayan V, Stergioulas N (2021). *Systematics of prompt black-hole formation in neutron star mergers.* Phys. Rev. D 103(12),123004

Bellinger EP, Basu S, Hekker S, Christensen-Dalsgaard J, Ball WH (2021). *Asteroseismic Inference of the Central Structure in a Subgiant Star.* ApJ 915(2):100

Berger B, Amaral M, Kokh DB, Nunes-Alves A, Musil D, Heinrich T, Schröder M, Neil R, Wang J, Navratilova I, Bomke J, Elkins JM, Müller S, Frech M, Wade RC, Knapp S (2021). *Structure-kinetic relationship reveals the mechanism of selectivity of FAK inhibitors over PYK2.* Cell Chemical Biology 28(5):686-698.e7

Bettisworth B, Stamatakis A (2021). *Phylourny: Predicting the Knock-out-phase of Tournaments via Phylogenetic Methods by example of the UEFA EURO 2020.* biorxiv,.06.24.449715v1,[Preprint]

Bianciotto M, Gkeka P, Kokh DB, Wade RC, Minoux H (2021). *Contact Map Fingerprints of Protein–Ligand Unbinding Trajectories Reveal Mechanisms Determining Residence Times Computed from Scaled Molecular Dynamics.* J. Chem. Theory Comput. 17(10):6522-6535

Bodensteiner J, Sana H, Wang C, Langer N, Mahy L, Banyard G, Koter A, de Mink S, Evans C, Götberg Y, Patrick L, Schneider F, Tramper F (2021). *The young massive SMC cluster NGC 330 seen by MUSE II.* Multiplicity properties of the massive-star population. A&A 652:A70

Bowden J, Crowley D, Davis J, Friedl S, Rovi C, Tillmann S (2021). *Open Problems in the Topology of Manifolds.* 2019-20 MATRIX Annals 4:647-659, Springer International Publishing

Bowman DM, Burssens S, Simón-Díaz S, Edelmann PVF, Rogers TM, Horst L, Röpke F, Aerts C (2021). *Collective velocity broadening from gravity waves as a plausible mechanism for macroturbulence in massive stars.* In MOBSTER-1 virtual conference: Stellar Variability as a Probe of Magnetic Fields in Massive Stars, p. 15

Bracher J, Wolffram D, Gneiting T, Schienle M (2021). *Vorhersagen sind schwer, vor allem die Zukunft betreffend: Kurzzeitprognosen in der Pandemie.* Mitteilungen der Deutschen Mathematiker-Vereinigung, 29(4):186–190

Bracher J, Wolffram D, Deuschel J, Görgen K, Ketterer JL, Ullrich A, Abbott S, Barbarossa MV, Bertsimas D, Bhatia S, Bodych M, Bosse NI, Burgard JP, Castro L, Fairchild G, Fuhrmann J, Funk S, Gogolewski K, Gu Q, Heyder S, Hotz T, Kheifetz Y, Kirsten H, Krueger T, Krymova E, Li ML, Meinke JH, Michaud IJ, Niedzielewski K, Ożański T, Rakowski F, Scholz M, Soni S, Srivastava A, Zieliński J, Zou D, Gneiting T, Schienle M (2021). *A pre-registered short-term forecasting study of COVID-19 in Germany and Poland during the second wave.* Nature Communications, 12:5173

Bracher J, Ray EL, Gneiting T, Reich NG (2021). *Evaluating epidemic forecasts in an interval format.* PLoS Comput Biol 17(2):e1008618

Braun L, Nikas S, Song C, Heuveline V, Fröning H (2021). *A Simple Model for Portable and Fast Prediction of Execution Time and Power Consumption of GPU Kernels.* ACM Trans. Archit. Code Optim. 18(1):1-25

Brehmer J (2021). *A construction principle for proper scoring rules.* Proceedings of the American Mathematical Society, Series B 8(24):297–301

Brehmer JR, Gneiting T (2021). *Scoring interval forecasts: Equal-tailed, shortest, and modal interval.* Bernoulli 27(3):1993–2010

Darms J, Henke J, Hu X, Schmidt CO, Golebiewski M, Fluck J; NFDI4Health Task Force COVID-19 (2021). *Improving the FAIRness of Health Studies in Germany: The German Central Health Study Hub COVID-19.* Studies in health technology and informatics 287:78-82

De M, Öztürk MA, Isbaner S, Tóth K, Wade RC (2021). *DNA sequence-dependent positioning of the linker histone in a nucleosome: A single-pair FRET study.* Biophysical Journal 120(17):3747-3763

Dey D, Marciano S, Nunes-Alves A, Kiss V, Wade RC, Schreiber G (2021). *Line-FRAP, A Versatile Method to Measure Diffusion Rates In Vitro and In Vivo.* Journal of Molecular Biology 433(9):166898

Dimitriadis T, Liu X, Schnaitmann J (2021). *Encompassing tests for value at risk and expected shortfall multistep forecasts based on inference on the boundary.* Journal of Financial Econometrics nbab004. https://doi.org/10.1093/jjfinec/nbab004

Dimitriadis T, Schnaitmann J (2021). *Forecast encompassing tests for the expected shortfall.* International Journal of Forecasting, 37(2):604–621

Dimitriadis T, Gneiting T, Jordan AI (2021). *Stable reliability diagrams for probabilistic classifiers.* Proc Natl Acad Sci USA, 118(8):e2016191118

Edelmann PVF, Horst L, Berberich JP, Andrassy R, Higl J, Leidi G, Klingenberg C, Röpke FK (2021). *Well-balanced treatment of gravity in astrophysical fluid dynamics simulations at low Mach numbers.* A&A 652:A53

Ehlert C, Piras A, Gryn'ova G (2021). *Assessing in Silico Models and Methods for Non-Covalent Interactions between Graphene and Small Molecules.* Chemistry doi: 10.26434/chemrxiv.14188154.v1 [Preprint]

Ernst M, Gryn'ova G (2021). *Strength and Nature of Host-Guest Interactions in Metal-Organic Frameworks from a Quantum-Chemical Perspective.* ChemRxiv. doi:10.26434/chemrxiv.14363024.v1 [Preprint]

Fatima M, Strube M (2021). *A Novel Wikipedia based Dataset for Monolingual and Cross-lingual Summarization.* In Proceedings of the Third Workshop on New Frontiers in Summarization. Association for Computational Linguistics, Online and in Dominican Republic, pp 39–50

Ferrand G, Warren DC, Ono M, Nagataki S, Röpke FK, Seitenzahl IR, Lach F, Iwasaki H, Sato T (2021). *From Supernova to Supernova Remnant: Comparison of Thermonuclear Explosion Models.* The Astrophysical Journal 906(2):93

Fluck J, Lindstädt B, Ahrens W, Beyan O, Buchner B, Darms J, Depping R, Dierkes J, Neuhausen H, Müller W, Zeeb H, Golebiewski M, Löffler M, Löbe M, Meineke F, Klammt S, Fröhlich H, Hahn H, Schulze M, Pischon T, Nöthlings U, Sax U, Kusch H, Grabenhenrich L, Schmidt CO, Waltemath D, Semler S, Gehrke J, Kirsten T, Praßer F, Thun S, Wieler L, Pigeot I (2021). *NFDI4Health – Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten.* Bausteine Forschungsdatenmanagement 2:72-85

Frost AJ, Mahy L, Sana H, LeBouquin J-B, Wade G, Merand A, Schneider FRN, Shenar T, Barbá R, Barron J, Bowman DM, Fabry M, Farhang A, Morrell NI, Munoz M, Smoker JV (2021). *A massive binary system with a single magnetic star.* In MOBSTER-1 virtual conference: Stellar Variability as a Probe of Magnetic Fields in Massive Stars39

Frost AJ, Mahy L, Sana H, Barba RH, Barron J, Bowman DM, Fabry M, Bouquin J-B, Morrell NI, Marchant P, Merand A, Munoz M, Schneider FRN, Shenar T, Wade G (2021). *Observational evidence of coalescence as a viable cause of magnetism in massive stars.* In OBA Stars: Variability and Magnetic Fields19

Ganotra GK, Nunes-Alves A, Wade RC (2021). *A Protocol to Use Comparative Binding Energy Analysis to Estimate Drug-Target Residence Time.* Protein-Ligand Interactions and Drug Design, Methods in Molecular Biology (MIMB) 2266:171-186

Ghosh S, Wronski P, Koetsenruijter J, Müller W, Wensing M (2021). *Watching people decide: decision prediction using heat-maps of reading of a decision-support document.* Journal of Vision 21(9):2631

Gianniotis N, Nuñez FP, Polsterer KL (2021). *Disentangling the optical AGN and host-galaxy luminosity with a probabilistic flux variation gradient.* Astronomy & Astrophysics 657:A126

Glaser M, Bruce NJ, Han SB, Wade RC (2021). *Simulation of the Positive Inotropic Peptide S100A1ct in Aqueous Environment by Gaussian Accelerated Molecular Dynamics,* J. Phys. Chem. B 125(18):4654-4666 1

Gneiting T, Vogel P (2021). *Receiver operating characteristic (ROC) curves: Equivalences, beta model, and minimum distance estimation.* Machine Learning, published online. doi: 10.1007/s10994-021-06115-2

Gneiting T, Walz E (2021). *Receiver operating characteristic (ROC) movies, universal ROC (UROC) curves, and coefficient of predictive ability (CPA).* Machine Learning, published online. doi: 10.1007/s10994-021-06114-3

Gossen J, Albani S, Hanke A, Joseph BP, Bergh C, Kuzikov M, Costanzi E, Manelfi C, Storici P, Gribbon P, Beccari AR, Talarico C, Spyrakis F, Lindahl E, Zaliani A, Carloni P, Wade RC, Musiani F, Kokh DB, Rossetti G (2021). *A Blueprint for High Affinity SARS-CoV-2 Mpro Inhibitors from Activity-Based Compound Library Screening Guided by Analysis of Protein Dynamics.* ACS Pharmacology & Translational Science 4(3):1079-1095

Gronow S, Côté B, Lach F, Seitenzahl IR, Collins CE, Sim SA, Röpke FK (2021). *Metallicity-dependent nucleosynthetic yields of Type Ia supernovae originating from double detonations of sub-MCh white dwarfs.* Astronomy & Astrophysic 656:A94

Gronow S, Collins CE, Sim SA, Röpke FK (2021). *Double detonations of sub-MCh CO white dwarfs: variations in Type Ia supernovae due to different core and He shell masses.* Astronomy & Astrophysic 649:A155

Haupt S, Gleim N, Ahadova A, Bläker H, von Knebel Doeberitz M, Kloor M, Heuveline V (2021). *A computational model for investigating the evolution of colonic crypts during Lynch syndrome carcinogenesis.* Computational and Systems Oncology 1(2):e1020

Haupt S, Zeilmann A, Ahadova A, Bläker H, Doeberitz MvK, Kloor M, Heuveline V (2021). *Mathematical modeling of multiple pathways in colorectal carcinogenesis using dynamical systems with Kronecker structure.* PLoS computational biology 17(5):e1008970.

Henzi A, Ziegel JF, Gneiting T (2021). *Isotonic distributional regression.* Journal of the Royal Statistical Society: Series B (Statistical Methodology) 83(5):963–993

Higl J, Müller E, Weiss A (2021). *Calibrating core overshooting parameters with two-dimensional hydrodynamical simulations.* A&A 646:A133

Hirai R, Podsiadlowski P, Owocki SP, Schneider FR, Smith N (2021). *Simulating the formation of η Carinae's surrounding nebula through unstable triple evolution and stellar merger-induced eruption.* Monthly Notices of the Royal Astronomical Society 503(3):4276-4296

Hodcroft EB, Maio ND, Lanfear R, MacCannell DR, Minh BQ, Schmidt HA, Stamatakis A, Goldman N, Dessimoz C (2021). *Want to track pandemic variants faster? Fix the bioinformatics bottleneck.* Nature 591(7848):30-33

Hoga Y, Dimitriadis T (2021). On testing equal conditional predictive ability under measurement error. Journal of Business & Economic Statistics pp 1–13

Holub P, Wittner R, Mascia C, Frexia F, Müller H, Plass M, Allocca C, Betsou F, Burdett T, Cancio I, Chapman A, Chapman M, Courtot M, Curcin V, Eder J, Elliot M, Exter K, Fairweather E, Goble C, Golebiewski M, Kisler B, Kremer A, Lin-Gibson S, Marsano A, Mattavelli M, Moore J, Nakae H, Perseil I, Salman A, Sluka J, Soiland-Reyes, S, Strambio-De-Castillia C, Sussman M, Swedlow JR, Zatloukal K, Geiger J (2021). *Towards a Common Standard for Data and Specimen Provenance in Life Sciences.* Zenodo. Doi: 10.5281/zenodo.5093125. [Preprint]

Horst L, Hirschi R, Edelmann PVF, Andrássy R, Röpke FK (2021). *Multidimensional low-Mach number time-implicit hydrodynamic simulations of convective helium shell burning in a massive star.* A&A 653:A55

Höhler D, Pfeiffer W, Ioannidis V, Stockinger H, Stamatakis A (2021). *RAxML Grove: An empirical Phylogenetic Tree Database.* Bioinformatics, btab863

Hübner L, Kozlov AM, Hespe D, Sanders P, Stamatakis A (2021). *Exploring Parallel MPI Fault Tolerance Mechanisms for Phylogenetic Inference with RAxML-NG.* Bioinformatics, btab399

Jayme A, Lösel P, Fischer J, Heuveline V (2021). Comparison of Machine Learning Methods for Predicting Employee Absences. Preprint Series of the EMCL. No.2, 2021

Jeon S, Strube M (2021). *Countering the Influence of Essay Length in Neural Essay Scoring.* In Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing. Association for Computational Linguistics, Virtual, pp 32–38

Jin F, Gräter F (2021). *How multisite phosphorylation impacts the conformations of intrinsically disordered proteins.* PLoS computational biology 17(5):e1008939

Jordan AI, Mühlemann A, Ziegel JF (2021). *Characterizing the optimal solutions to the isotonic regression problem for identifiable functionals.* Annals of the Institute of Statistical Mathematics, published online. https://doi.org/10.1007/s10463-021-00808-0

Kallinowski F, Ludwig Y, Löffler T, Vollmer M, Lösel P, Voß S, Görich J, Heuveline V, Nessel R (2021). *Biomechanics applied to incisional hernia repair – Considering the critical and the gained resistance towards impacts related to pressure.* Clinical Biomechanics 82:105253

Kokh DB, Wade RC (2021). *G Protein-Coupled Receptor–Ligand Dissociation Rates and Mechanisms from τRAMD Simulations.* J. Chem. Theory Comput. 17(10):6610-6623

Krüger F, Lerch S, Thorarinsdottir T, Gneiting T (2021). *Predictive inference based on Markov chain Monte Carlo output.* International Statistical Review, 89(2):274–301

König M, Grzegorzewski J, Golebiewski M, Hermjakob H, Hucka M, Olivier B, Keating S, Nickerson D, Schreiber F, Sheriff R, Waltemath D (2021). *Ten Simple Rules for Sharing Experimental and Clinical Data with the Modeling Community.* 2021080303 doi: 10.20944/preprints202108.0303.v1. [Preprint]

Liu Z, Röpke FK, Zeng Y, Heger A (2021). *Observational signatures of the surviving donor star in the double-detonation model of Type Ia supernovae.* A&A 654:A103

Lutteropp S, Scornavacca C, Kozlov AM, Morel B, Stamatakis A (2021). *NetRAX: Accurate and Fast Maximum Likelihood Phylogenetic Network Inference ⋆.* biorxiv;.08.30.458194v1[Preprint]

López F, Pozzetti B, Trettel S, Strube M, Wienhard A (2021). *Vector-valued Distance and Gyrocalculus on the Space of Symmetric Positive Definite Matrices.* In Proceedings of the Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS), Online

López F, Pozzetti B, Trettel S, Strube M, Wienhard A (2021). *Symmetric Spaces for Graph Embeddings: A Finsler-Riemannian Approach.* In Proceedings of the 38th International Conference on Machine Learning. arXiv:2106.04941

Maag D, Putzu M, Gómez-Flores, C. L., Gräter F, Elstner M, Kubař T (2021). *Electrostatic interactions contribute to the control of intramolecular thiol–disulfide isomerization in a protein.* Physical Chemistry Chemical Physics 23(46):26366-26375

Mahé F, Czech L, Stamatakis A, Quince C, de Vargas, C, Dunthorn M, Rognes T (2021). *Swarm v3: towards tera-scale amplicon clustering.* Bioinformatics btab493

Martin I, Nava M, Wickström S, Gräter F (2021). *ATP allosterically stabilizes Integrin-linked kinase for efficient force generation.* bioRxiv doi:10.1101/2021.03.30.437490 [Preprint]

Matschinske J, Alcaraz N, Benis A, Golebiewski M, Grimm DG, Heumos L, Kacprowski T, Lazareva O, List M, Louadi Z, Pauling JK, Pfeifer N, Röttger R, Schwämmle V, Sturm G, Traverso A, Steen KV, Freitas MVd, Silva GCV, Wee L, Wenke NK, Zanin M, Zolotareva O, Baumbach J, Blumenthal DB (2021). *The AIMe registry for artificial intelligence in biomedical research.* Nature Methods 18(10):1128-1131

Mayer G, Müller W, Schork K, Uszkoreit J, Weidemann A, Wittig U, Rey M, Quast C, Felden J, Glöckner , F.O., Lange M, Arend D, Beier S, Junker A, Scholz U, Schuler D, Kestler HA, Wibberg D, Pühler A, Twardziok S, Eils J, Eils R, Hoffmann S, Eisenacher M, Turewicz M (2021). *Implementing FAIR data management within the German Network for Bioinformatics Infrastructure (de.NBI) exemplified by selected use cases.* Brief Bioinform. Sep 2;22(5). pii: 6135008

Michelarakis N, Franz F, Gkagkas K, Gräter F (2021). *Longitudinal strand ordering leads to shear thinning in Nafion.* Physical Chemistry Chemical Physics 23(45):25901-25910

Miralles A, Ducasse J, Brouillet S, Flouri T, Fujisawa T, Kapli P, Knowles LL, Kumari S, Stamatakis A, Sukumaran J, Lutteropp S, Vences M, Puillandre N (2021). *SPART, a versatile and standardized data exchange format for species partition information.* bioRxiv doi: 10.1111/1755-0998.13470 [Preprint]

Mobasher G, Mertová L, Ghosh S, Krebs O, Heinlein B, Müller W (2021). *Combining dictionary- and rule-based approximate entity linking with tuned BioBERT.* bioRxiv doi: 10.1101/2021.11.09.467905 [Preprint]

Moreau CA, Quadt KA, Piirainen H, Kumar H, Bhargav SP, Strauss L, Tolia NH, Wade RC, Spatz JP, Kursula I, Frischknecht F (2021). *A function of profilin in force generation during malaria parasite motility that is independent of actin binding.* J Cell Sci 134(5):jcs233775

Morel B, Schade P, Lutteropp S, Williams TA, Szöllősi GJ, Stamatakis A (2021). *SpeciesRax: A tool for maximum likelihood species tree inference from gene family trees under duplication, transfer, and loss.* bioRxiv doi: 10.1101/2021.03.29.437460 [Preprint]

Mostert RIJ, Duncan KJ, Röttgering HJA, Polsterer KL, Best PN, Brienza M, Brüggen M, Hardcastle MJ, Jurlin N, Mingo B, Morganti R, Shimwell T, Smith D, Williams WL (2021). *Unveiling the rarest morphologies of the LOFAR Two-metre Sky Survey radio source population with self-organised maps.* A&A 645:A89

Mukherjee G, Nandekar PP, Wade RC (2021). *An electron transfer competent structural ensemble of membrane-bound cytochrome P450 1A1 and cytochrome P450 oxidoreductase.* Commun Biol 4(1),55

Müller M, Ghosh S, Wittig U, Rey M (2021). *Word-Level Alignment of Paper Documents with their Electronic Full-Text Counterparts.* In Proceedings of the 20th SIGBioMed Workshop on Biomedical Language Processing. arXiv:2104.14925, doi: 10.18653/v1/2021.bionlp-1.19

Nunes-Alves A, Ormersbach F, Wade RC (2021). *Prediction of the Drug–Target Binding Kinetics for Flexible Proteins by Comparative Binding Energy Analysis.* J. Chem. Inf. Model. 61(7):3708-3721

Nunes-Alves A, Kokh DB, Wade RC (2021). *Ligand unbinding mechanisms and kinetics for T4 lysozyme mutants from τRAMD simulations.* Current Research in Structural Biology 3:106-111

Obarska-Kosinska A, Rennekamp B, Ünal A, Gräter F (2021). *ColBuilder: A server to build collagen fibril models.* Biophysical journal 120(17):3544-3549

Ostaszewski M, Niarakis A, Mazein A, Kuperstein I, Phair R, Orta-Resendiz A, Singh V, Aghamiri SS, Acencio ML, Glaab E, Ruepp A, Fobo G, Montrone C, Brauner B, Frishman G, Gómez LCM, Somers J, Hoch M, Gupta SK, Scheel J, Borlinghaus H, Czauderna T, Schreiber F, Montagud A, Leon MPd, Funahashi A, Hiki Y, Hiroi N, Yamada TG, Dräger A, Renz A, Naveez M, Bocskei Z, Messina F, Börnigen D, Fergusson L, Conti M, Rameil M, Nakonecnij V, Vanhoefer J, Schmiester L, Wang M, Ackerman EE, Shoemaker JE, Zucker J, Oxford K, Teuton J, Kocakaya E, Summak GY, Hanspers K, Kutmon M, Coort S, Eijssen L, Ehrhart F, Rex DAB, Slenter D, Martens M, Pham N, Haw R, Jassal B,

Matthews L, Orlic-Milacic M, Ribeiro AS, Rothfels K, Shamovsky V, Stephan R, Sevilla C, Varusai T, Ravel J, Fraser R, Ortseifen V, Marchesi S, Gawron P, Smula E, Heirendt L, Satagopam V, Wu G, Riutta A, Golebiewski M, Owen S, Goble C, Hu X, Overall RW, Maier D, Bauch A, Gyori BM, Bachman JA, Vega C, Grouès V, Vazquez M, Porras P, Licata L, Iannuccelli M, Sacco F, Nesterova A, Yuryev A, Waard Ad, Turei D, Luna A, Babur O, Soliman S, Valdeolivas A, Esteban-Medina M, Peña-Chilet M, Rian K, Helikar T, Puniya BL, Modos D, Treveil A, Olbei M, Meulder BD, Ballereau S, Dugourd A, Naldi A, Noël V, Calzone L, Sander C, Demir E, Korcsmaros T, Freeman TC, Augé F, Beckmann JS, Hasenauer J, Wolkenhauer O, Wilighagen EL, Pico AR, Evelo CT, Gillespie ME, Stein LD, Hermjakob H, D'Eustachio P, Saez-Rodriguez J, Dopazo J, Valencia A, Kitano H, Barillot E, Auffray C, Balling R, Schneider R (2021). *COVID19 Disease Map, a computational knowledge repository of virus–host interaction mechanisms.* Molecular systems biology 17(10):e10387

Paiardi G, Milanesi M, Wade RC, D'Ursi P, Rusnati M (2021). *A Bittersweet Computational Journey among Glycosaminoglycans.* Biomolecules 11(5):739

Piras A, Gryn'ova G (2021). *The Law of Attraction: Relating Computed Energetics of Physisorption with Performance of Graphene-Based Sensors for Nitroaromatic Contaminants.* Chemrxiv doi: 10.26434/chemrxiv.14364800.v2 [Preprint]

Piras A, Ehlert C, Gryn'ova G (2021). *Sensing and sensitivity: Computational chemistry of graphene-based sensors.* WIREs Comput Mol Sci 11(5)

Plier J., Savarino F., Kočvara M., Petra S. (2021) *First-Order Geometric Multilevel Optimization for Discrete Tomography.* In: Elmoataz A., Fadili J., Quéau Y., Rabin J., Simon L. (eds) Scale Space and Variational Methods in Computer Vision. SSVM 2021. Lecture Notes in Computer Science, vol 12679. Springer, Cham. doi: 10.1007/978-3-030-75549-2_16

Psonis N, Antoniou A, Karameta E, Darriba D, Stamatakis A, Lymberakis P, Poulakakis N (2021). *The wall lizards of the Balkan peninsula: Tackling questions at the interface of phylogenomics and population genomics.* Molecular Phylogenetics and Evolution 159:107121

Range J, Halupczok C, Lohmann J, Swainston N, Kettner C, Bergmann FT, Weidemann A, Wittig U, Schnell S, Pleiss J (2021). *EnzymeML - a data exchange format for biocatalysis and enzymology.* FEBS J. Dec 10. doi: 10.1111/febs.16318

Reinhardt M, Bruce NJ, Kokh DB, Wade RC (2021). *Brownian Dynamics Simulations of Proteins in the Presence of Surfaces: Long-Range Electrostatics and Mean-Field Hydrodynamics.* J. Chem. Theory Comput. 17(6):3510-3524

Rey M, Weidemann A, Wittig U, Dudas D, Ghosh S, Golebiewski M, Hu X, Müller W (2021). *FAST, ADAPTED, CURATED FAIR DATA FOR COVID-19 RESEARCH.* In: Data Analysis for the COVID-19 Research: Contributions of the German Network of Bioinformatics Infrastructure. doi: 10.4119/unibi/2950259

Rusnati M, Paiardi G, Tobia C, Urbinati C, Lodola A, D'Ursi P, Corrado M, Castelli R, Wade RC, Tognolini M, Chiodelli P (2021). *Cholenic acid derivative UniPR1331 impairs tumor angiogenesis via blockade of VEGF/VEGFR2 in addition to Eph/ephrin.* Cancer Gene Ther. doi: 10.1038/s41417-021-00379-5

Khosla S, Yu J, Manuvinakurike RR, Ng V, Poesio M, Strube M, Rosé CP (2021). *The CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue.* In: Proceedings of the CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue. doi: 10.18653/v1/2021.codi-sharedtask.1

Sadiq SK, Chicharro AM, Friedrich P, Wade RC (2021). *Multiscale Approach for Computing Gated Ligand Binding from Molecular Dynamics and Brownian Dynamics Simulations.* J. Chem. Theory Comput. 17(12):7912-7929

Salo-Ahen OMH, Alanko I, Bhadane R, Bonvin AMJJ, Honorato RV, Hossain S, Juffer AH, Kabedev A, Lahtela-Kakkonen M, Larsen AS, Lescrinier E, Marimuthu P, Mirza MU, Mustafa G, Nunes-Alves A, Pantsar T, Saadabadi A, Singaravelu K, Vanmeert M (2021). *Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development.* Processes 9(1):71

Schmidt P, Katzfuss M, Gneiting T (2021). *Interpretation of point forecasts with unknown directive.* Journal of Applied Econometrics 36(6):728–743

Schmidt CO, Fluck J, Golebiewski M, Grabenhenrich L, Hahn H, Kirsten T, Klammt S, Löbe M, Sax U, Thun S, Pigeot I, Ahrens W, Darms J, Henke J, Hu X, Klopfenstein S, Langnickel L, Lassen-Schmidt B, Pohlabeln H, Lieser M, Näher A, Scholz M, Vorisek C, Waltemath D, Wünsche H (2021). *COVID-19-Forschungsdaten leichter zugänglich machen – Aufbau einer bundesweiten Informationsinfrastruktur.* Bundesgesundheitsblatt 64 1084–1092

Schmidt CO, Darms J, Shutsko A, Löbe M, Nagrani R, Seifert B, Lindstädt B, Golebiewski M, Koleva S, Bender T, Bauer CR, Sax U, Hu X, Lieser M, Junker V, Klopfenstein S, Zeleke A, Waltemath D, Pigeot I, Fluck J (2021). *Facilitating Study and Item Level Browsing for Clinical and Epidemiological COVID-19 Studies.* Studies in health technology and informatics 281:794-798

Schmiester L, Schälte Y, Bergmann FT, Camba T, Dudkin E, Egert J, Fröhlich F, Fuhrmann L, Hauber AL, Kemmer S, Lakrisenko P, Loos C, Merkt S, Müller W, Pathirana D, Raimúndez E, Refisch L, Rosenblatt M, Stapor PL, Städter P, Wang D, Wieland FG, Banga JR, Timmer J, Villaverde AF, Sahle S, Kreutz C, Hasenauer J, Weindl D. *PEtab-Interoperable specification of parameter estimation problems in systems biology.* PLoS Comput Biol. 2021 Jan 26;17(1):e1008646

Schneider F (2021). *State-of-the-art evolution models for single, binary and magnetic massive stars.* In MOBSTER-1 virtual conference: Stellar Variability as a Probe of Magnetic Fields in Massive Stars p.30. doi: 10.5281/zenodo.5534518

Schneider F, Podsiadlowski P, Müller B (2021). *Pre-supernova evolution, compact-object masses, and explosion properties of stripped binary stars.* A&A 645:A5 doi: 10.1051/0004-6361/202039219

Schreiber F, Gleeson P, Golebiewski M, Gorochowski TE, Hucka M, Keating SM, König M, Myers CJ, Nickerson DP, Sommer B, Waltemath D (2021). *Specifications of standards in systems and synthetic biology: status and developments in 2021.* Journal of integrative bioinformatics, 18(3):20210026

Schulz B, Ayari ME, Lerch S, Baran S (2021). *Post-processing numerical weather prediction ensembles for probabilistic solar irradiance forecasting.* Solar Energy, 220:1016–1031

Schween N, Gerstner P, Meyer-Hübner N, Heuveline V (2021). *New Time Step Strategy for Multi-period Optimal Power Flow Problems.* In: Mathematical Modeling, Simulation and Optimization for Power Engineering and Management 34:199-217 Springer International Publishing doi: 10.1007/978-3-030-62732-4_9

Serenelli A, Weiss A, Aerts C, Angelou GC, Baroch D, Bastian N, Beck PG, Bergemann M, Bestenlehner JM, Czekala I, Elias-Rosa N, Escorza A, Van Eylen V, Feuillet DK, Gandolfi D, Gieles M, Girardi L, Lebreton Y, Lodieu N, Martig M, Bertolami MMM, Mombarg JSG, Morales JC, Moya A, Nsamba B, Pavlovski K, Pedersen MG, Ribas I, Schneider FRN, Aguirre VS, Stassun KG, Tolstoy E, Tremblay P, Zwintz K (2021). *Weighing stars from birth to death: mass determination methods across the HRD.* Astronomy & Astrophysics Review 29(1) 4

Shenar T, Sana H, Marchant P, Pablo B, Richardson N, Moffat AFJ, Van Reeth T, Barbá RH, Bowman DM, Broos P, Crowther PA, Clark JS, de Koter A, de Mink S, Dsilva K, Gräfener G, Howarth ID, Langer N, Mahy L, Apellániz JM, Pollock AMT, Schneider FRN, Townsley L, Vink JS (2021). *The Tarantula Massive Binary Monitoring.* V. R 144: a wind-eclipsing binary with a total mass ≳140 M_⊙. Astronomy & Astrophysics 650:A147

Sigmund LM, Ehlert C, Enders M, Graf J, Gryn'ova G, Greb L (2021). *Dioxygen Activation and Pyrrole α-Cleavage with Calix[4] pyrrolato Aluminates: Enzyme Model by Structural Constraint.* Angewandte Chemie International Edition60(28):15632-15640

Song C, Roller J, Ponce-Bobadilla AV, Palacio-Escat N, Saez-Rodriguez J, Heuveline V (2021). *Spatial Effect on Separatrix of Two-Cell System and Parameter Sensitivity Analysis.* Preprint Series of the EMCL, No. 1, 2021

Stephens D, Herwig F, Woodward P, Denissenkov P, Andrassy R, Mao H (2021). *3D1D hydro-nucleosynthesis simulations- I. Advective-reactive post-processing method and its application to H ingestion into He-shell flash convection in rapidly accreting white dwarfs.* Monthly Notices of the Royal Astronomical Society 504(1):744-760

Strube M (2021). *Computerwissenschaften und The Circle -- The Circle und Computerwissenschaften.* In Kempter, Klaus und Martina Engelbrecht (Eds.): Krise(n) der Moderne. Über Literatur und Zeitdiagnostik. Universitätsverlag Winter, Heidelberg, Germany. 451-460 (2021)

Vassella CC, Koch J, Henzi A, Jordan A, Waeber R, Iannaccone R, Charrière R (2021). *From spontaneous to strategic natural window ventilation: Improving indoor air quality in Swiss schools.* International Journal of Hygiene and Environmental Health 234:113746

Vogel P, Knippertz P, Gneiting T, Fink AH, Klar M, Schlueter A (2021). *Statistical forecasts for the occurrence of precipitation outperform global models over northern tropical Africa.* Geophysical Research Letters 48(3):e2020GL091022

Walz E, Maranan M, van der Linden R, Fink AH, Knippertz P (2021). *An IMERG-based optimal extended probabilistic climatology (EPC) as a benchmark ensemble forecast for precipitation in the tropics and subtropics.* Weather and Forecasting 36(4):1561–1573

Whan K, Zscheischler J, Jordan AI, Ziegel JF (2021). *Novel multivariate quantile mapping methods for ensemble post-processing of medium-range forecasts.* Weather and Climate Extremes 32:100310

Wronski P, Wensing M, Ghosh S, Gärttner L, Müller W, Koetsenruijter J (2021). *Use of a quantitative data report in a hypothetical decision scenario for health policymaking: a computer-assisted laboratory study.* BMC Medical Informatics and Decision Making 21(1),32

Yu J, Hekker S, Bedding TR, Stello D, Huber D, Gizon L, Khanna S, Bi S (2021). *Asteroseismology of luminous red giants with Kepler – II. Dependence of mass-loss on pulsations and radiation.* Monthly Notices of the Royal Astronomical Society 501(4):5135-5148

Zha J, Zhang Y, Xia K, Gräter F, Xia F (2021). *Coarse-Grained Simulation of Mechanical Properties of Single Microtubules With Micrometer Length.* Frontiers in Molecular Biosciences 7: 632122

8 Teaching

Degrees

Pierre Barbera:
"High-Performance approaches for Phylogenetic Placement, and its application to species and diversity quantification", Ph.D. thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2021).

Thomas Baumann:
"Simulating Common Envelope Interactions in Systems with Different Mass Ratios, Primary Star Masses and Evolutionary Stages", Master´s thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Friedrich Röpke (2021).

Elizaveta Bobkova:
"From Random Forests to PaiNN: Predicting Hydrogen Transfer Activation Energies for Kinetic Molecular Dynamics", Master's thesis, Department of Physics and Astronomy, Heidelberg University, HITS: Frauke Gräter and KIT: Pascal Friederich (2021).

Madhura De:
"The effects of the linker-DNA on linker histone positioning: a single-pair FRET study from mono to trichromatosomes", Ph.D. thesis, Combined Faculty for the Natural Sciences and Mathematics, Heidelberg University, Katalin Toth (DKFZ) and Rebecca C. Wade (HITS) (2021).

Jan-Niklas Dohrke:
"Molecular mechanisms of AMPAR modulation by auxiliary subunits", Master's thesis, Molecular Biotechnology, Faculty of Biosciences, Heidelberg University, and HITS: Ariane Nunes-Alves and Rebecca C. Wade (2021).

Florian Franz:
"Molecular Dynamics Of The Membrane-Talin-Vinculin Axis Under Force", Ph.D. thesis, Combined Faculty of Natural Sciences and Mathematics, Heidelberg University and HITS: Frauke Gräter (2021).

Sabrina Gronow:
"Contribution of Type Ia supernovae to the chemical enrichment of the Milky Way: explosions of sub-MCh white dwarfs", Ph.D. thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Friedrich Röpke (2021).

Julia Haag:
"Empirical Numerical Properties of Maximum Likelihood Phylogenetic Inference", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2021).

Sungho Bosco Han:
"Multi-Resolution Simulation Study of CYP 2B4 to Investigate Complex Formation with Human NADPH-Cytochrome P450 Oxidoreductase in Lipid Bilayer", Master's thesis, Molecular Biosciences, Faculty of Biosciences, Heidelberg University, and HITS: Goutam Mukherjee and Rebecca C. Wade (2021).

Jannis Heising:
"Algorithmic implementation of the solution to the world problem in Right Angled Artin groups", Bachelor´s thesis, Mathematics Institute, Heidelberg University: Beatrice Pozzetti (2021).

Leonhard Horst:
"Simulation of Hydrodynamic Phenomena in Stellar Interiors", Ph.D. thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Friedrich Röpke (2021).

Mitul Islam:
"Rank one phenomena in convex projective geometry", Ph.D. thesis, Department of Mathematics, University of Michigan: Ralf Spatzier (2021).

Victor Alexander Judea:
"Global Inference and Local Syntactic Representations for Event Extraction", Ph.D. thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, (2021).

Christoph Karfusehr:
"Where does collagen break? A QM, MD and KIMMDY study", Master´s thesis, Faculty of Engineering Sciences, Heidelberg University & Max Planck School Matter to Life (Franziska Thomas), and HITS: Frauke Gräter (2021).

Serife Kol:
"Konzeption und Entwicklung eines Prüfprozesses für die Prüfung der Informationssicherheit von Software- & Lizenzen im Rahmen einer Beschaffung", Master´s thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2021).

Kristof Kraus:
"Probabilistic energy forecasting using isotonic distributional regression", Bachelor's thesis, Faculty of Mathematics, Karlsruhe Institute of Technology, and HITS: Sebastian Lerch and Tilmann Gneiting (2021).

Siting Liang:
"Summarizing German Radiology Findings for Cancer Patients", Master´s thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, (2021).

Pierre De Marinis:
"Assessing ML-database augmentation approaches for Druggability Score calculations in TRAPP-pocket", Bachelor's thesis, Molecular Biotechnology, Faculty of Biosciences, Heidelberg University, and HITS: Stefan Richter and Rebecca C. Wade (2021).

Lena Meßner L:
"Comparison of ligand docking to the AMPAR-TARP interaction site between different isoforms", Bachelor's thesis, Molecular Biotechnology, Faculty of Biosciences, Heidelberg University, and HITS: Giulia D'Arrigo and Rebecca C. Wade (2021).

Joachim Meyer:
"Compiler-assisted optimizations for data-parallel paradigms in hipSYCL", Master´s thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2021).

Benoit Morel:
"Algorithms, load balancing strategies, and dynamic kernels for large-scale phylogenetic tree inference under Maximum Likelihood", Ph.D. thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2021).

Melvin Moreno:
"Magnetohydrodynamic Simulations of the Common Envelope Phase with a Red Supergiant Star", Master´s thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Friedrich Röpke (2021).

Patrick Ondratschek:
"Magnetohydrodynamic Simulations of the Common Envelope Phase in Stellar Binary Systems", Master´s thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Friedrich Röpke (2021).

Benedikt Pfahls:
"Universal constants in 3-dimensional hyperbolic manifolds", Ph.D. thesis, Mathematics Institute, Heidelberg University: Beatrice Pozzetti (2021).

Mareike Pfeil:
"Cataclysm deformations for Anosov representations", Ph.D. thesis, Mathematics Institute, Heidelberg University: Anna Wienhard and Beatrice Pozzetti (2021).

Jonas Robl:
"Balancing privacy and utility in differentially private generative models", Master´s thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2021).

Jonas Roller:
"Numerical Simulation and Uncertainty Quantification of a Spatially Heterogeneous Two-Cell System", Master´s thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2021).

Anna Schilling:
"The horofunction compactification of finite-dimensional normed vector spaces and of symmetric spaces", Ph.D. thesis, Mathematics Institute, Heidelberg University: Anna Wienhard (2021).

Valentin Schmid:
"Numerical Simulation of Nanosecond Pulsed Discharges", Master´s thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2021).

Balint Soproni:
"Task parallelism in hipSYCL ", Bachelor´s thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2021).

Lorenz Thielbeer:
"Semi-analytic common-envelope simulations", Bachelor´s thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Fabian Schneider (2021).

Aysecan Ünal:
"Structural analysis of collagen with cryoEM", Master´s thesis, Faculty of Engineering Sciences, Heidelberg University & Max Planck School Matter to Life and HITS: Frauke Gräter (2021).

Dennis Wagner:
"Intermolecular Interactions in SAS-6 from Coarse-Grained Molecular Dynamics Simulations and Sequence Analysis", Master´s thesis, Faculty of Physics, Heidelberg University (Ulrich Schwarz), and HITS: Frauke Gräter (2021).

Daniel Wolffram:
„Building and evaluating forecast ensembles for COVID-19 deaths", Master's thesis, Faculty of Mathematics, Karlsruhe Institute of Technology, and HITS: Johannes Bracher, Melanie Schienle, and Tilmann Gneiting (2021).

Holger Wünsche:
"SYCL 2020 work group parallel primitives: Optimized algorithms for GPUs and CPUs in hipSYCL", Bachelor´s thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2021).

Pavel Zwerschke:
"Deep nonparametric distributional regression for probabilistic energy prediction", Bachelor's thesis, Faculty of Informatics and Faculty of Mathematics, Karlsruhe Institute of Technology, and HITS: Sebastian Lerch and Tilmann Gneiting (2021).

Lectures, courses and seminars

Aksel Alpay:
Supervision of the Heidelberg University team for the ISC21 Student Cluster Competition, with: Rod Burns, Igor Vorobtsov, Ronan Keryell, Michael Steyer, and Gordon Brown.: *"A Hands-On Introduction To SYCL."* In International Workshop on OpenCL (IWOCL'21), 27 – 29 April 2021; with: Rod Burns, Gordon Brown, James Reinders, Michael Wong, Peter Zuzek, and Ronan Keryell. 2021: *"Heterogeneous Programming in Modern C++ with SYCL."* CppCon 2021.

Camilo Aponte-Santamaría and Frauke Gräter:
Course on "Data Science", Matter to Life Max Planck school, winter semester 2020/2021.

Alain Becam, Xiaoming Hu, Wolfgang Müller, Maja Rey, Ulrike Wittig:
Data Management Workshop at MPI for Evolutionary Biology Plön, online, 18 - 19 May 2021.

Giulio Belletti:
"3-manifolds topology", Heidelberg University, summer semester 2021.

Jannik Buhr:
Lecture Series on "Introduction to Data Analysis with R", Heidelberg University, winter semester 2021/2022.

Fernando Cadena:
"Random Walks on Reductive Groups", Reading Program, Yale, fall semester 2021.

Colin Davalo:
Tutor in "A first course in dynamical system and ergodic theory", Heidelberg University, winter semester 2021/2022.

Dorotea Dudas, Xiaoming Hu, Maja Rey, Andreas Weidemann, Ulrike Wittig:
de.NBI Course "Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOMHub/SEEK", online, 20 - 22 April 2021.

James Farre:
Integral calculus (MATH 115), Yale University, fall semester 2021. RTG lecture *"Aspects of Teichmüller-Thurston theory"*, Heidelberg, Summer semester 2021.

Tilmann Gneiting and Johannes Resin:
Course on "Forecasting: Theory and Practice II", Karlsruhe Institute of Technology, summer semester 2021. Seminar on "Statistical Forecasting and Classification", Karlsruhe Institute of Technology, winter semester 2021/22.

Frauke Gräter and Friedrich Röpke:
Lecture on "Fundamentals of Simulation Methods", Heidelberg University, winter semester 2020/2021.

Frauke Gräter, Rebecca Wade, Rob Russell and others:
Lecture on "Computational Biochemistry", Heidelberg University, winter semester 2020/2021.

Frauke Gräter, Rebecca Wade:
M.Sc. Seminar course "Machine Learning for the Biomolecular World", Heidelberg University, summer semester, 2021.

Ganna Gryn'ova, Michelle Ernst and Christopher Ehlert:
Special Lecture Course "Applied Computational Chemistry", Heidelberg University, summer semester 2021.

Saskia Hekker:
Oberseminar: "Theory and Observations of Stars" (summer semester 2021 and winter semester 2021/2022). Class: "Asteroseismology" (only winter semester 2021/2022).

Jan Henneco:
Tutorial for Lecture course "Fundamentals of Simulation Methods", Heidelberg University, winter semester 2021/2022; lecturers: Friedrich Röpke and Mario Flock.

Vincent Heuveline, Philipp Gerstner:
Lecture "Finite Elements", Heidelberg University, winter semester 2020/2021. Lecture "Computational Fluid Dynamics", Heidelberg University, summer semester 2021. Seminar "Computational Fluid Dynamics", Heidelberg University, winter semester 2021/2022.

Vincent Heuveline, Maximilian Hoecker:
Lecture "IT Sicherheit", Heidelberg University, winter semester 2020 / 2021. Seminar *"IT Sicherheit"*, Heidelberg University, summer semester 2021. Lecture *"IT Sicherheit"*, Heidelberg University, winter semester 2021 / 2022.

Dani Kaufman:
Mini Course on "Cluster Algebras", Heidelberg University, winter semester 2021/2022.

Olga Krebs
"FAIRDOMHub for storing, finding, sharing, and reusing of data", Online practical training for Innovative Training Network (ITN) PoLiMeR, 25 February 2021. *"Data management with FAIRDOMHub"*, online practical training for SCyCODE Project at the University of Duisburg-Essen, 31 May 2021. *"Data publishing and archival. A possible path to publishing with FAIRDOMHub and Zenodo"*, lecture and computer practical at the ELIXIR Luxembourg's training on "Best practises in research data management and stewardship", 14 - 17 June 2021. FAIRDOMHub practical training at the DAAD German-Ukrainian Autumn School *"Lessons in Biomedicine learnt from Nanotechnology and Artificial Intelligence"*, 29 September - 5 October 2021, Kiev/Kharkiv, Ukraine. *"FAIR data principles and Data publishing and archival"*, lecture at the ELIXIR Luxembourg online training on "Best practises in research data management and stewardship", 5 - 8 October 2021. *"FAIRDOMHub for storing, finding, sharing, and reusing of data, models, operations"*, online lecture and hands on training for Innovative Training Network (ITN) CC-TOP, 6 December 2021.

Markus Kurth:
3DMM2O EPR Spectrometer facility: Introduction and measurements for cluster members and guests Heidelberg University, Centre for Advanced Materials" (March-December 2021).

Giovanni Leidi:
Tutorial accompanying the lecture course "Fundamentals of Simulation Methods", Heidelberg University, winter semester 2020/2021.

Brice Loustau:
HEGL Proseminar "Graphs on hyperbolic spaces", Heidelberg University, summer semester 2021. HEGL Proseminar "Visualizations in hyperbolic spaces", Heidelberg University, winter semester 2021/2022.

Marta Magnani:
Tutor in "Geometric group theory", Heidelberg University, winter semester 2021/2022.

Kai Polsterer:
Antonio Picariello Lectures, UNINA, From Photometric Redshift to Improved Weather Forecasts, lecture on data-science, UNINA, Naples, Italy (online), 13 January 2021.

Beatrice Pozzetti:
Student seminar "Dynamics in one complex variable", Heidelberg University, summer semester 2021. "Geometrische Gruppentheorie", Heidelberg University, winter semester 2021/2022.

Anja Randecker:
Exercises and coordination for "Linear Algebra 2", Heidelberg University, summer semester 2021. HEGL Proseminar "Graphs on hyperbolic spaces", Heidelberg University, summer semester 2021. HEGL Proseminar "Visualizations in hyperbolic spaces", Heidelberg University, winter semester 2021/2022.

Maja Rey, Ulrike Wittig:
First ERNEST Training School "Computational Tools to Study GPCR Signaling: From the Genomic to the Systems Level", online, 15 - 16 June 2021.

Friedrich Röpke:
Research Seminar / Oberseminar "Physics of Stellar Objects", Heidelberg University, winter semester 2020/2022, summer semester 2021, and winter semester 2021/2022.
Lecture course "Fundamentals of Simulation Methods", Heidelberg University, winter semester 2021/2022 with Mario Flock.

Friedrich Röpke, Fabian Schneider:
Lecture course "The Stellar Cookbook: A practical guide to the theory of stars", Heidelberg University, winter semester 2020/2021, and winter semester 2021/2022.

Anna Schilling:
"Geometric Aspects from Analysis and Linear Algebra", Heidelberg University (online), summer 2021.

Fabian Schneider:
Lecture course "Stars Squared: Evolution of Binary Stars", summer semester 2021. Lecturer at IMPRS Summer School "Stellar Ecosystems", 2021.

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, 2020/2021.

Alexandros Stamatakis, Benoit Morel, Alexey Kozlov, Lukas Hübner:
Lecture "Introduction to Bioinformatics for Computer Scientists", computer science Master's program at Karlsruhe Institute of Technology, winter semester, 2021/2022.

Alexandros Stamatakis, Ben Bettisworth, Lukas Hübner:
Practical "Hands-on Bioinformatics Practical", computer science Master's program at Karlsruhe Institute of Technology, summer semester, 2021.

Alexandros Stamatakis, Benoit Morel, Alexey Kozlov:
Seminar "Hot Topics in Bioinformatics", computer science Master's program at Karlsruhe Institute of Technology, summer semester, 2021.

Michael Strube:
Ph.D. Colloquium, Department of Computational Linguistics, Heidelberg University (Winter Semester 2020/2021). Ph.D. *Colloquium*, Department of Computational Linguistics, Heidelberg University, summer semester 2021.

Diaaeldin Taha:
Calc III, American University in Cairo, spring semester 2021. Statistics for Business, spring semester 2021. Fundamentals of Data Science II, spring semester 2021.

Rebecca Wade:
Lecture contributions to the M.Sc. Molecular & Cellular Biology Modules 3 & 4, M. Sc. Biochemistry "Computational Biochemistry" course, and B.Sc. Molecular Biotechnology "Biophysical Chemistry" course.

Rebecca Wade, Stefan Richter, and Manuel Glaser:
Lectures and exercises on „Proteinstrukturen und -domänen: Analyse und Vorhersage" – BSc. Grundkurs Bioinformatik, Heidelberg University, winter semester 2020/2021.

Anna Wienhard:
Linear Algebra 1, Heidelberg University, winter semester 2020/2021.
Linear Algebra 2, Heidelberg University, summer semester 2021.

9 Miscellaneous

9.1 Guest Speaker Activities (invited only)

Giulio Belletti:
"The maximum volume of hyperbolic polyhedra", "Real and Complex Manifolds" conference, Pisa, Italy, June 2021; "Winter Braids XI", December 2021.

Johannes Bracher:
"Collaborative forecasting of COVID-19: Assembling, comparing and combining short-term predictions", BigData Technology Warsaw Summit, Warsaw, Poland (online), 23 February 2021; online seminar on COVID-19 econometrics, WHU – Otto Beisheim School of Management, Düsseldorf, Germany (online), 7 May 2021. *"Collaborative forecasting of COVID-19 cases and deaths in Germany and Poland"*, International Workshop on Forecasting for Social Good (online), 24 June 2021. *"A marginal moment matching approach for fitting endemic-epidemic models to underreported disease surveillance counts"*, Conference of the Austro-Swiss Region (ROeS) of the International Biometric Society, Salzburg, Austria, Conferral of Arthur-Linder Award, 8 September 2021. *"Nowcasting COVID-19 hospitalization incidences in Germany"*, COVID-19 Modeling and Forecasting Call, European Center for Disease Prevention and Control (online), 8 December 2021. *"Collaborative nowcasting and short-term forecasting during the COVID-19 pandemic"*, NordicMathCovid Research Meeting (online), 16 December 2021.

Jonas Brehmer:
"Using scoring functions to evaluate point process forecasts", Extreme Value Analysis 2021, Edinburgh, United Kingdom (online), 28 June – 2 July 2021.

Xian Dai:
"Introduction to Pressure Metric", "Workshop on Anosov Representations", Aussois, France, 29 August 2021. *"The pressure metric in the space of Riemannian metrics"*, "Mini-Workshop Anosov^3", Oberwolfach, 5 December 2021.

Madhura De:
"Single-molecule studies on mono- and trinucleosomes", Fragile Nucleosome Seminar https://generegulation.org/fragile-nucleosome/, on-line, March 10, 2021; Multiscale Genome Organisation Biophysical Society Subgroup Seminar, on-line, 10 November 2021.

Valentina Disarlo:
"The model theory of the curve graph", "Topics at the interface of low dimensional group actions and geometric structures", IMS Singapore, January 2021.

Christopher Ehlert:
"A Little Tour on Computational Chemistry", International Summer Science School Heidelberg, Germany, (online), 26 July – 9 August 2021.

Tilmann Gneiting:
"Isotonic distributional regression (IDR): Leveraging monotonicity, uniquely so!", Keynote Lecture, Nordstat 2021, Tromsø, Norway (online), 21 June 2021.

Martin Golebiewski:
"Standards for data acquisition and management" Joint Workshop of the European Commission and CEN/CENELEC "Putting Science into Standard - Organ-on-chip: Toward standardization", online, 28 - 29 April 2021. *"Standards und Werkzeuge für FAIRe Forschungsdaten in der Biomedizin"*, Workshop "Perspektiven des Forschungsdatenmanagements in den Lebenswissenschaften", 66th Annual Conference of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) and INFORMATIK 2021 – 51st Annual Conference of the German Informatics Society (GI), online, 30 September 2021.

Frauke Gräter:
"Hybrid simulations of protein mechanochemistry: jumping over learned barriers", Multiscale Modelling Conference, Manchester, 29 March 2021. *"Feeling the force: molecular mechano-sensors in biological systems"*, BioSoft Colloquium Forschungszentrum Jülich, 1 April 2021. *"A radical tale: Collagen as a buffer of mechanical and oxidative stress."*, Collagen Café, Dartmouth, 3 May 2021. *"Collagen: Rediscovering the most abundant protein of our body with HPC"*, CaSToRC HPC National Competence Center Seminar, Cyprus, 4 May 2021. *"Scale-bridging simulations of proteins to uncover mechano-sensing mechanisms"*, Workshop 'Engineering Molecular Systems', Heidelberg, 9 June 2021. Bio-molecular Structure and Mechanism retreat, Morschach, Switzerland 30 August 2021. *"Stretching collagen: Mechanoradicals and their biochemical consequences"*, 7th Collagen Symposium, Freiberg, 29 September 2021. Online-Symposium Atomistic Simulations, 26 October 2021. *"How functional disulfides regulate proteins in blood flow"*, Society for Redox Biology and Medicine's 28th Annual Conference, Savannah, GA, 19 November 2021. *"Feeling the force: molecular mechano-sensors at cellular interfaces"*, Pacifichem 2021, Honolulu, 16 December 2021. *"Mechanoradicals in collagen: biomaterials as mechano-sensors?"*, Pacifichem 2021, Honolulu, 17 December 2021.

Ganna Gryn'ova:
"Computational Carbon Chemistry", Elisabeth-Schiemann Kolleg (online), Max Planck Society, Germany, 30 June 2021. *"Research in the Computational Carbon Chemistry Group"*, Lieseberg Colloquium (online), Heidelberg University, Germany, 5 July 2021. *"Computational Graphene Chemistry"*, Joint Theoretical Chemistry Seminar (online), Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Germany, 23 November 2021.

Saskia Hekker:
"The power of asteroseismology: the internal structure of stars", ARI Colloquium Heidelberg, Germany (online), 14 January 2021. *"Mirror principle and the red-giant bump: the battle of entropy in low-mass stars"*, SAC seminar, Aarhus, Denmark (online), 21 January 2021. *"The power of asteroseismology"*, Keynote lecture workshop on precision spectroscopy, Sao Paolo, Brazil (online), 1 February 2021. *"Asteroseismology: What makes stars tick?"*, Königstuhl Colloquium, Heidelberg, Germany, (online), 24 September 2021.

Vincent Heuveline:
"Digital Twins in der Medizin: Herausforderungen und Perspektiven aus Sicht der KI und der Mathematik", 50. Jahrestagung der Deutschen Gesellschaft für Thorax-, Herz- und Gefäßchirurgie (DGTHG), online, 27 February 2021. *"Digital Twins in Medicine: Challenges and Perspectives from the point of view of AI and Mathematics"*, BioRN Lounge, Heidelberg, online, 27 April 2021. *"IT-Sicherheit in der Akademia: im Spannungsfeld von Handlungsfähigkeit und digitaler Selbstbestimmung"*, Keynote Digitales Ökosystem, Jahrestagung der Digitalen Hochschule NRW, online, 15 September 2021.

Johann Higl:
"Calibrating the Core Overshooting Parameter With Hydrodynamical Simulations", invited Astrophysics Seminar, University of Exeter, United Kingdom, 16 June 2021.

Pengfei Huang:
"Geometry of base manifold which parametrizes a family of Higgs bundles", "Geometry & Quantization (GEOQUANT 2021)", Albert-Ludwigs-Universität Freiburg, August 2021. *"An investigation of Dolbeault moduli spaces via Simpson-Mochizuki correspondence"*, "SFB CRC/TRR 191 Retreat 2021", Köln, September 2021.

Mitul Islam:
"Convex co-compact groups and relative hyperbolicity", "Young researcher's meeting of the GDR Platon", CIRM - Marseille Luminy, France, 17 November 2021.

Dani Kaufman:
"Scattering Amplitudes, Cluster Algebras and Positive Geometries", "Oberwolfach mini-workshop on Scattering Amplitudes, Cluster Algebras and Positive Geometries", Oberwolfach, Germany, 5-11 December 2021.

Olga Krebs:
"Introduction to FAIR data management: general concepts and principles", online talk at the DAAD German-Ukrainian Autumn School "Lessons in Biomedicine learnt from Nanotechnology and Artificial Intelligence", 29 September – 5 October 2021, Kiev/ Kharkiv Ukraine.

Sebastian Lerch:
"AI methods for post-processing ensemble predictions", EUMETNET Workshop on Verification, Extremes and Forecasting Value Chain for High Impact Weather (online), 28 October 2021; *"AI for probabilistic weather prediction"*, Future Lab on Artificial Intelligence, Potsdam Institute for Climate Impact Research (online), 3 August 2021.

Philipp Lösel:
"Large-scale Analysis of The Honey Bee Brain Using Micro-CT Imaging and Deep Learning", Stockholm University Brain Imaging Centre (SUBIC), Stockholm (online), 7 October 2021, on invitation by Dr. Tunhe Zhou.

Wolfgang Müller:
"FAIRDOM": EJP RD lecture series on Convergence of FAIR initiatives, online, 25 June 2021.

Wolfgang Müller, Olga Krebs:
"From SysMO-DB to FAIRDOM, data management for Life Science ERANets and beyond", Keynote at The Third BioTech Research and Innovation Hack 2021, 29 September 2021.

Giulia Paiardi:
"Molecular modeling in sugar-protein interactions: a bittersweet computational journey", scientific seminar, University of Brescia, Italy, 27 March 2021. "A bittersweet computational journey between glycosaminoglycans", Innogly meeting on "GlycosAmino-Glycans: What remains to be solved?", Heraklion, Greece, 27-29 September 2021.

Mareike Pfeil:
"Cataclysms for Anosov representations", the "Nearly Carbon Neutral Geometric Topology" conference, June 2021.

Anna Piras:
"How Can We Predict the Performance of Graphene-Based Sensors?", Hütter group seminar series, University of Zurich, Switzerland (online), 29 March 2021.

Kai Polsterer:
"Using proper scoring rules to derive well calibrate photometric redshift models", SAE2021 – BIG4small conference, 24 September 2021 (online).

Beatrice Pozzetti:
"Positive surface group representations in SO(p, q)", "Differentialgeometrie im Grossen", Oberwolfach, Germany, July 2021.

Anja Randecker:
"Trees everywhere (The saddle connection complex)", "Karlsruher Weihnachtsworkshop on Geometry and Number Theory", Karlsruhe, Germany, December 2021. *"Topological behaviour of conjugacy classes of big mapping class groups"*, "Swiss Knots"", Fribourg, Switzerland, July 2021.

Friedrich Röpke:
"A 3D view on stellar astrophysics", invited ARC seminar talk, Queen's University Belfast (online), 3 March 2021. *"Multidimensional simulations of thermonuclear supernovae"*, invited talk at the 16th Marcel Grossmann Meeting (online), 5 July 2021.

Carmen Rovi:
"Manifolds and K-theory: the legacy of Andrew Ranicki", "Memorial conference for Andrew Ranick", ICMS, June 2021.

Anna Schilling:
"The horofunction compactification of symmetric spaces", "Nearly Carbon Neutral Geometric Topology" conference, June 2021.

Fabian Schneider:
"Neutron stars and black hole formation in binary stars", Carnegie Observatories Colloquium, Carnegie Observatories, USA, 19 Jan 2021; Radboud Astrophysics Colloquium, Radboud University Nijmegen, The Netherlands, 30 Mar 2021; Astronomy Colloquium, University of Hertfordshire, United Kingdom, 26 May 2021.
"Turbulent Lives of Stars", Ludwig Biermann Award Lecture, Annual Meeting of the German Astronomical Society, 16 Sep 2021; National Astronomical Observatory of Japan Science Colloquium, Japan, 1 Dec 2021.
"The strongest magnets in the Universe", IvS Astrophysics Colloquium, KU Leuven, Belgium, 15 Jan 2021
"Bayesian inference in stellar astrophysics", STRUCTURES Jour Fixe, University of Heidelberg, Germany, 5 Feb 2021

Alexandros Stamatakis:
"Current and Future Research in Phylogenetics", on-line bioinformatics/genomics seminar by the Department of Computational Biology, Institut Pasteur, April 2021.

Rebecca C. Wade:
"Prediction of the Structure and Dynamics of Protein Complexes in Membranes", (Session chair) 65th Biophysical Society Annual Meeting, online, 22-26 February 2021. *"Computational Mapping of the Druggability of Cryptic Pockets in Proteins"*, ECBS, online, Frontiers in Medicinal Chemistry, GDCh Fachgruppe Medizinische Chemie, on-line, 8 -10 March 2021. *"Molecular dynamics simulations to investigate the interactions of heparin with the Sars-CoV2 spike glycoprotein"* (with G. Paiardi, M Rusnati), EuroHPC Summit Week: PRACEdays21, COVID-19 Day, online, 24 March 2021. *"Bridging timescales: Protein-ligand binding kinetics and beyond"*, 4th CCP-BioSim/CCP5 Multiscale Modelling Conference, online, 29-31 March 2021. *"Protein-Ligand Dissociation Rates and Mechanisms from τ-Random Acceleration Molecular Dynamics Simulation"*, American Chemical Society Spring Meeting, online, 5-16 April 2021. *"Combining molecular simulation and machine learning approaches for structure-based drug design"*, (Keynote), European Chemical Biology Symposium (ECBS), EU-OPENSCREEN and the EuChemS Division of "Chemistry in Life Sciences Division" online, 26-28 May 2021. *"Protein-Ligand Dissociation Rates and Mechanisms from τ-Random Acceleration Molecular Dynamics Simulation"*, MolSSI School on Open Source Software in Rare Event Path Sampling Strategies, Pittsburgh, USA, online, 8 June 2021 *"Beyond affinity: predicting protein-ligand binding kinetics"*, Affinity 2021, online, 22-24 June 2021. *"Zooming in on the dynamic interactions of proteins and drugs by computer simulation"*, CBT Seminar, St. Jude Children's Research Hospital, online, 23 September 2021. *"From molecular simulations towards antiviral therapeutics against COVID-19"*, 84th Annual Meeting of the German Physical Society (DPG), online, 27 September 2021. *"Towards the rational design of therapeutics against neurodegenerative diseases"*, Cecam Flagship Workshop on "Innovative strategies for neurodegenerative diseases: a perspective from molecular simulation, machine learning and experiment", Pisa, Italy, hybrid, September 29, 2021.

Anna Wienhard:
"Graph embeddings in symmetric spaces", Heidelberg University, 14 June 2021. *"Computational Aspects of Discrete Subgroups of Lie Groups"*, ICERM, USA. "Symmetric spaces and discrete structures: Geometry, Dynamics, and Applications", Colloquium Series, National University Singapore, 16 July 2021. *"Dynamics of Anosov representations - in the interior and at the boundary"*, the mini-workshop "Anosos^3", Mathematisches Forschungsinstitut Overwolfach, 9 December 2021.

Ulrike Wittig:
"Collecting, curating, interlinking, and sharing data with SABIO-RK and FAIRDOM-SEEK", ESAB Webinar "Synthesis Planning and Data Management in Biocatalysis", online, 22 February 2021. *"Woman Power in Data Curation and Data Management"*, Joint Bayer AG & de.NBI/ELIXIR Germany Event "Women in Data Science - Perspectives in Industry and Academia", online, 18 June 2021. *"Follow the Data Life Cycle - Research Data Management Best Practices"*, Core Technologies in Life Sciences CTLS 2021, online, 13 - 15 September 2021. *"Follow the Cycle - Research Data Management is more than Data Storage"*, de.NBI Summer School "Analysis and integration of mass spectrometry based omics data", online, 27 - 30 September 2021.

Adrian Zapletal and Dimitri Höhler:
"Das SoftWipe-Tool zur Analyse von Code-Qualität", CampusSource Herbsttagung 2021, University of Hagen, Germany, September 2021.

9.2 Presentations (contributed talks and posters)

Aksel Alpay:
"SYCL 2020 in hipSYCL: DPC++ features on AMD GPUs, NVIDIA GPUs and CPUs. Keynote." oneAPI Developer Summit at International Workshop on OpenCL (IWOCL'21), 27 – 29 April 2021. *"HipSYCL in 2021: Peculiarities, unique features and SYCL 2020."* International Workshop on OpenCL (IWOCL'21), 27 – 29 April 2021.

Róbert Andrássy:
"Comparison of five stellar-hydrodynamic codes on a turbulent convection problem", talk at Challenges and Innovations in Computational Astrophysics III, online workshop, 17 June 2021.

Camilo Aponte-Santamaría:
"Localization of Cholesterol around the metabotropic glutamate receptor 2". Workshop on computer simulation and theory of macromolecules. Virtual meeting, 23-24 April, 2021. *"Mechanical stress pattern of integrin αIIbβ3 upon disulfide bond cleavage revealed by molecular dynamics simulations and force distribution analysis"*. Functional Disulfides in Health & Disease. Virtual meeting. 14-16 June 2021.

Christina Athanasiou:
"Mechanistic studies of TrkA and TrkB neurotrophin receptors", International Conference on "Neurotrophic Factors and Neurodegenerative Disorders; Current Advances and Future Perspectives", Heraklion (Crete), Greece, 27-29 August 2021.

Giulia D'Arrigo:
"From the detection of cross-correlations till the identification of alternative druggable pockets in Nuclear Receptors", CDDD 7th Meeting, Virtual meeting, 25 June 2021.

Pierre Barbera:
"Efficient Memory Management in Likelihood-based Phylogenetic Placement", 2021 IEEE International Parallel and Distributed Processing Symposium Workshops, on-line presentation, May 2021.

Giulio Belletti:
"The maximum volume of hyperbolic polyhedra", the Geometry seminar, University of Toulouse, November 2021.

Jonas Brehmer:
"Using scoring functions to evaluate point process forecasts", German Probability & Statistics Days Mannheim, Mannheim, Germany (online), 27 September – 1 October 2021.

Matthias Brosz:
"Mechanics and structural organization of Poly (para-phenylene ethynylene)s from coarse-grained simulations". 3DMM2O Winter Cluster meeting, online, Germany, 23 February 2021.; Multiscale Mechanochemistry & Mechanobiology, 23-25 August 2021, online, Germany. *"Martini 3 Coarse-Grained Force Field: Extension to Poly (para-phenylene ethynylene)s"*. Martini Workshop, online, The Netherlands, 1-3 September 2021.

Svenja de Buhr:
"Mechanoactivation of c-Abl Kinase". Future 3D Additive Manufacturing - The 3DMM2O Conference: 3D Hybrid Organotypic Systems, online, 1-4 March 2021. *"c-Abl Kinase Binds to the Cellular Membrane and Becomes Mechanoactivated"*. EBSA 2021 - 13th European Biophysics Conference, online and Vienna, Austria, 24 - 28 July 2021.

Fernando Cadena:
"Exponential Multiple Mixing of Abelian Differentials", the Geometry and Topology Seminar, Yale, November 2021. Ibid, the Dynamics Seminar, University of Maryland, December 2021. *"Multiple Mixing for Semisimple Lie group actions"*, the Dynamics Reading Seminar, Yale, October 2021.

Xian Dai:
"Correlation theorem for Hitchin representations", the UCR Differential Geometry seminar, 2 April 2021. Ibid, the Geometry and topology seminar, Université Côte d'Azur, France, 14 October 2021.

Madhura De:
"Single-molecule studies on mono- and trinucleosomes", ZMBH ZAPPS seminar, Heidelberg University, online, 25 November 2021.

Timo Dimitriadis:
"Stable reliability diagrams for probabilistic classifiers", INFORMS Annual Meeting, Anaheim, California, United States (online), 27 October 2021.
"Forecast calibration, backtests, and score decompositions for Value at Risk forecasts", QAT Seminar, Credit Suisse, Switzerland (online), 28 October 2021.

Valentina Disarlo:
"The model theory of the curve graph", Ibid, "On-line Webinar Geometric Group Theory in East Asia (GGTea)", 26 March 2021; Séminaire Virtuel Francophone Groupes et Géométrie, 29 April 2021; Heriot-Watt University, Edinburgh, UK, 30 June 2021.

Christopher Ehlert:
"PSIXAS: A PSI4 Plugin for Efficient Simulations of X-Ray Absorption Spectra", Gühr group seminar series, University of Potsdam, Germany (online), 12 March 2021. "Quantum Chemical Assessment of Well-Defined Catalysts for the Oxygen Reduction Reaction", DPG Spring Meetings (online), 22 – 24 March 2021 (poster).

Michelle Ernst:
"Strength and Nature of Host-Guest Interactions in Metal-Organic Frameworks From a Quantum Chemical Perspective", XXV General Assembly and Congress of the International Union of Crystallography (IUCr 2021, online), 14 – 21 August 2021; American Chemical Society Fall Meeting (online), 22-26 August 2021. *"An Assessment of Quantum Chemical Tools for The Analysis of Host-Guest Interactions in MOFs"*, Second Discussion Meeting on Quantum Crystallography: Expectations and Reality, CECAM Flagship Workshop (online), 9 -12 September 2021 (poster); 4th European Conference on Metal Organic Frameworks and Porous Polymers (online), 13–15 September, 2021 (poster).

James Farre:
"Conjugating flows on the moduli of hyperbolic and flat surfaces", the Pacific Dynamics seminar, April 2021. *"Curvature bounds on least area fibers in hyperbolic 3-manifolds"*, the Geometry seminar, City University of New York, 19 October 2021. *"Long curves and random hyperbolic surfaces"*, the Geometry seminar, Wesleyan University, 27 October 2021. *"Hamiltonian flows on spaces of surfaces"*, the Geometry and Topology seminar, Yale University, 9 November 2021.

Philipp Gerstner:
"The Influence of Gravity on Thermal Electro-hydrodynamic Flow in Annular Geometry", video presentation at 21st International Couette-Taylor Workshop, online, 5–9 July, 2021.

Sucheta Ghosh:
"Beim Lesen beobachtet – Erkenntnisse aus einem Eye-Tracking-Labor zur Nutzung eines Versorgungsberichts in einem hypothetischen gesundheitspolitischen Entscheidungsszenario", 21st German Conference on Health Services Research 2021, online, 5-7 October 2021.

Sucheta Ghosh, Pamela Wronski, Jan Koetsenruijter, Wolfgang Müller, Michel Wensing:
"Watching people decide: decision prediction using heatmaps of reading of a decision-support document", Vision Sciences Society Annual Meeting 2021, online, 21-26 May 2021 (poster).

Nikos Gianniotis:
"Probabilistic flux variation gradient", Annual Meeting of the Astronomische Gesellschaft 2021, 13-17 September 2021.
"A Flux Variation Gradient based on probabilistic principal component analysis", Astroinformatics 2021, 15-19 November 2021.

Manuel Glaser:
"Combining experimental and computational approaches to unravel the biology of the inotropic peptide S100A1ct", Informatics for Life – 16th Regular Project Meeting, virtual, 1 June 2021.

Martin Golebiewski:
"ISO/TC 276/WG 5 Data Processing and Integration - Report of the Convenor", Committee Meetings of ISO/TC 276/WG 5, online, 27 January and 19 July 2021 and ISO/TC 276 Biotechnology - Plenary Meeting, online, 30 June 2021. *"Data and metadata standards and integration"*, NFDI4health TA2 & TA5 Meeting, online, 20 April 2021. *"CoreTrustSeal"*, Nfdi4health Task Force Covid-19 Meeting, online, 10 May 2021. "ISO/TC 276/WG 5 Data Processing and Integration", Committee Meetings of the German standardization committe "Arbeitsausschuss Biotechnologie", DIN, online, 18 May and 21 September 2021. *"Data sources and standards for predictions in personalized medicine"* (talk jointly with Niklas Blomberg), EU-STANDS4PM annual conference 2020, online, 1 - 2 July 2021. *"Standards for FAIR Data"* (Report from NFDI4Health task area TA2) and "TA2 Wrap-up", NFDI4Health General Assembly and SAB Meeting, Göttingen (Germany), 25 October 2021. *"Data and metadata standards and integration"* and "FAIR data sharing – community aspects", NFDI4Health Consortium Meeting, Göttingen (Germany), 26 October 2021. *"Standards for Health Data and Modelling"*, Joint EU-STANDS4PM - VPHi - Avicenna Alliance - In Silico World, online, 21 December 2021.

Ganna Gryn'ova:
"Nanographene Catalysts for the Oxygen Reduction Reaction", European Conference on Chemistry of Two-Dimensional Materials (chem2Dmat, online), 31 August – 3 September 2021.

Saskia Haupt:
"A computational model for investigating the evolution of colonic crypts during Lynch syndrome carcinogenesis", SMB2021, Oncology subgroup, online, 13–17 June 2021. *"Computation of confidence intervals in PLSD"*, EHTG Meeting 2021, online, 8–9 October 2021. *"Mathematically modeling Lynch syndrome colorectal carcinogenesis at different scales"*, EHTG Meeting 2021, online, 8–9 October 2021.

Johann Higl:
"Fully compressible low-Mach simulations and their usability in stellar evolution models", talk at XXth Workshop on Nuclear Astrophysics, Ringberg Castle, Tegernsee, Germany, 4 November 2021.

Pengfei Huang:
"What is nonabelian Hodge theory?" "What is… Seminar" (WiSe), University of Queensland, Australia, May 2021. *"On Simpson's oper stratum conjecture"*, the "Workshop on Moduli Space of Vector Bundles", Zhejiang University, Hangzhou, China, May 2021. *"Construction of Deligne–Hitchin twistor spaces via nonabelian Hodge correspondence"*, "IASM"", Harbin Institute of Technology, Harbin, China; November 2021.

Mitul Islam:
"Convex co-compactness, in rank one and beyond", the Colloquium, Indian Statistical Institute, Kolkata, India, 15 December 2021. *"Convex co-compact groups and relative hyperbolicity"*, the Topology and Geometric Group Theory Seminar, Ohio State University, USA, 9 November 2021. *"Convex co-compact representations of 3-manifold groups"*, the Geometry seminar, University of Virginia, USA, 3 March 2021. *"Preliminaries on dynamics on geometrically finite hyperbolic manifolds"*, "Arbeitsgemeinschaft: Thin Groups and Super-strong Approximation", Oberwolfach, Germany, 11 October 2021. *"Hyperbolic Geometry and Beyond: An algebraic and dynamical perspective"*, the webinar of the Department of Applied Mathematics, University of Calcutta, India, 21 December 2021.

Fan Jin:
"Molecular mechanisms of E-cadherin mediated mechano-sensing", SPP 1782 virtual International Meeting 2021, 4-5 October 2021.

Dani Kaufman:
"Scattering Amplitudes, Cluster Algebras and Positive Geometries", the Oberwolfach seminar on Knot Theory, 7 December 2021.

Fenja Kollasch:
"UltraPINK: Newest developments in visualizing and interacting with Self-Organizing Kohonen Maps", Annual Meeting of the Astronomische Gesellschaft 2021, 13-17 September 2021. *"UltraPINK - New possibilities to explore Self-Organizing Kohonen Maps"*, Astronomical Data Analysis Software and Systems (ADASS) XXXI, 24-28 October 2021.

Markus Kurth:
"A role for dihydroxyphenylalanine (DOPA) and its radical as marker for mechanical stress in collagen", 13th European Biophysics Conference (online), 24 - 28 July 2021; Multiscale Mechanochemistry & Mechanobiology (online), 23-25 August 2021.

Georgios Kydonakis:
"Toda equations, parabolic Higgs bundles, and related topics", the Geometry seminar, Waseda University, Tokyo, Japan, 5 October 2021. *"Surgeries in representations varieties"*, the Greek Mathematical Seminar, Greece, 6 October 2021.

Eva Laplace:
"TULIPS: a Tool for Understanding the Lives, Interiors, and Physics of Stars", Award ceremony of the Royal Holland Society of Sciences and Humanities, KHMW, The Netherlands, 29 November 2021.

Sebastian Lerch:
"Post-processing numerical weather prediction ensembles for probabilistic solar irradiance forecasting", European Geosciences Union General Assembly 2021 (online), 27 April 2021. *"Incorporating large-scale spatial information into ensemble post-processing via autoencoder neural networks"*, NOAA Workshop on Leveraging AI in Environmental Sciences, United States (online), 15 September 2021.

Philipp Lösel:
"Large-scale Analysis of The Honey Bee Brain Using Micro-CT Imaging and Deep Learning", HeKKSaGOn University Consortium, The 8th German – Japanese University Presidents' Conference, Tohoku University (online), 10 September 2021.

Brice Loustau:
"Harmonic maps and the Labourie conjecture", the Geometry seminar, Heidelberg University, April 2021. *"The sum of Lagrange numbers"*, the Differential Geometry Group Meeting, Heidelberg University, June 2021.

Kiril Maltsev:
"The Penrose 1965 singularity theorem in historical context of the Black Hole paradigm", talk at the 16th Marcel Grossmann Meeting (online), 8 July 2021.

Arnaud Maret:
"Remarkable surface group representations in genus zero", the Geometry and Topology Seminar, University of Bristol, 19 October 2021. *"Remarkable surface group representations in genus zero"*, the Geometry Seminar, ETH Zurich, 27 October 2021. *"Character varieties, mapping class group action and ergodicity"*, the Geometry Graduate Colloquium, ETH Zurich, 28 October 2021.

Isabel Martin:
"Mechanotransduction in the focal adhesion pseudokinase ILK". Life at the Periphery: Mechanobiology of the Cell Surface. Heidelberg, Germany (online), 2-3 March 2021; Workshop on computer simulation and theory of macromolecules. Hünfeld, Germany (online), 23-24 April 2021.

Micaela Menegaldo:
Astronomical Data Analysis Software and Systems (ADASS) XXXI, Unsupervised classification of simulated black hole shadows, 24-28 October 2021.

Ghadeer Mobasher, Lukrecia Metrova, Sucheta Ghosh, Olga Krebs, Bettina Heinlein, Wolfgang Müller:
"Combining dictionary-and rule-based approximate entity linking with tuned BioBERT", BioCreative VII Challenge Evaluation Workshop, online, 8-10 November 2021 (poster).

Javier Morán Fraile:
"Gravitational Wave emission from Dynamic Stellar interactions", talk at "Common Envelope Physics and Outcomes (CEPO 2021)" (online), 1 September 2021.

Melvin Moreno Maldonado:
"Common envelope interactions with a 10-solar-mass primary star", talk at "Common Envelope Physics and Outcomes (CEPO 2021)" (online), 1 August 2021.

Wolfgang Müller:
"Data management and sharing in LiSyM. Achievements and wrap up phase", LiSyM 2021 jamboree, online, 10-12 May 2021.

Abraham Muniz-Chicharro:
"A multiscale approach to compute conformation-gated ligand-receptor binding kinetics.", 5th Biological Diffusion and Brownian Dynamics Brainstorm (BDBDB5) Meeting, online, March 25, 2021. *"A multiscale approach to compute conformation-gated ligand-receptor binding kinetics."*, ZMBH PhD Student Conference, held online, 26 April 2021.

Ariane Nunes-Alves:
"Ligand unbinding mechanisms and kinetics for T4 lysozyme mutants", Webinar for the Red Latinoamericana de Fisicoquímica Teórica, online, 3 February 2021. *"Fluorescence recovery curves from Brownian dynamics simulations"*, 5th Biological Diffusion and Brownian Dynamics Brainstorm (BDBDB5) Meeting, online, 24-25 March 2021.

Giulia Paiardi:
"Molecular dynamics simulations to investigate the interactions of heparin with the SARS-CoV2 spike glycoprotein", American Chemical Society symposium: A Call to Action: The Many Roles of Computational Chemistry in Addressing COVID-19, Virtual conference, online, 5 April 2021.

Anna Piras:
"How Can We Simulate the Detection Of Nitroaromatic Contaminants by Graphene-Based Sensors?", European Conference on Chemistry of Two-Dimensional Materials (chem2Dmat, online), 31 August - 3 September 2021. *"How Do Nitro-Aromatic Compounds Interact with Graphene-Based Materials?"*, 18th European Symposium on Organic Reactivity (online), 21 – 23 September 2021.

Jan Plier:
"First-Order Geometric Multilevel Optimization for Discrete Tomography", 8. International Confernce on Scale Space and Variational Methods in Computer Vision, 17 May 2021.

Francisco Pozo Nunez:
"AGNs as an independent tool for H0 determination", Astroinformatics 2021, 15-19 November 2021.

Beatrice Pozzetti:
"On Θ-positive surface subgroups in PO(p,q)", IISc, 1 December 2021. Ibid, the Geometry seminar, Singapore, 27 October 2021. Ibid, IHES, 11 October 2021. *"Entropy and Hausdorff dimension for hyperconvex Anosov sub- groups"*, the Geometry seminar, Orsay, France, 14 October 2021. *"Orbit growth rate in Higher rank Teichmüller spaces"*, the Geometry seminar, Yale University, 12 April 2021. Ibid, the University of Münster, 20 May 2021.

Ghulam Qadir:
"Flexible modeling of variable asymmetries in cross-covariance functions for multivariate random fields", German Probability & Statistics Days Mannheim, Mannheim, Germany (online), 27 September – 1 October 2021.

Anja Randecker:
"Topological behaviour of conjugacy classes of big mapping class groups", the Geometry and Topology Seminar, Haifa, Israel, November 2021. *"Saddle connection complex: coarse and fine."*, the Hyperbolic Geometry Seminar, New York, USA, November 2021. Ibid, the Quasiworld Seminar, Online, June 2021. *"The Veech group of the golden ladder"*, the "Big Surf(aces)" Seminar, online, January 2021.

Benedikt Rennekamp:
"MULTISCALE SIMULATIONS OF COLLAGEN FAILURE", (public part of) virtual Matter-to-Life mid-term evaluation, 10 March 2021. *"Hybrid Kinetic Monte Carlo / Molecular Dynamics Simulations of Bond Scissions in Proteins"*, DPG spring meeting, Dresden (online), 22-24 March 2021. *"Collagen structure and rupture modes enable mechanical buffering"*, Virtual department seminar of Max-Planck-Institute for Biophysical Chemistry in Göttingen, 26 May 2021. *"Hybrid Simulations of Collagen as Mechanical and Chemical Buffer"*, Virtual Workshop COMPUTER SIMULATION AND THEORY OF MACROMOLECULES 2021, in Hünfeld (Germany), 23-24 April 2021. *"Collagen as a Buffer of Mechanical and Oxidative Stress",* Virtual Conference on Multiscale Mechanochemistry and Mechanobiology, 23-25 August 2021; Virtual Matter-to-Life Fall Days, 7-8 October 2021. *"Where does collagen break?"*, Virtual department seminar of Max-Planck-Institute for Biophysical Chemistry in Göttingen, 11 November 2021. *"Where does (tendon) collagen break?"*, 4th Canadian Collagen Cafe (online), Dalhousie University (Canada), 22 November 2021.

Maja Rey:
"SABIO-RK Introduction", de.NBI Course "Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOMHub/SEEK", online, 22 April 2021.

Max Riestenberg:
"Symmetric spaces (from the Riemannian Geometry perspective)", the BZR Seminar, UT Austin, 24 September 2021. *"Crash course on hyperbolic geometry"*, the Geometric Group Theory Learning Seminar, UT Austin, 15 October 2021. *"A quantified local-to-global principle for Anosov representations"*, the Geometry and Topology seminar, Université du Luxembourg, 2 November 2021. *"Conics inside and outside properly convex domains"*, the Differential Geometry Group Meeting, Heidelberg University, 1 December 2021.

Friedrich Röpke:
Session introduction at "Common Envelope Physics and Outcomes (CEPO 2021)" (online), 31 August 2021. *"Simulations of common-envelope interaction in binary stellar systems"*, talk at XXth Workshop on Nuclear Astrophysiks, Ringberg Castle, Tegernsee, Germany, 4 November 2021.

Carmen Rovi:
"Kontsevich's characteristic classes", the Münster topology reading seminar, January 2021. Ibid, the Topology Seminar, *MPIM Bonn, January 2021. Ibid, the* Colloquium at Loyola University Chicago, February 2021. *"Surgery theory"*, the MPIM Topology Seminar, April 2021. *"Cut and paste invariants and K-theory"*, the Differential Geometry Group meeting, Heidelberg University, July 2021. *"Cut and Paste invariants and TQFTs"*, the Structures CP7 lunch seminar, Heidelberg University, July 2021.

Fabian Schneider:
"Neutron stars and black hole formation in binary stars", Auckland Astronomy Seminar, The University of Auckland, New Zealand, 18 May 2021; High Energy Astrophysics Seminar, UNAM, Mexico, 18 November 2021.

Theodoros Soultanis:
"Semi-analytic model of the spectral properties of gravitational waves from neutron star merger remnants", talk at The 15th Hellenic Astronomical Conference, 5 July 2021.

Jonathan Teuffel:
"Combined quantum mechanics and molecular mechanics study of interprotein electron transfer". SIMPLAIX first meeting, Heidelberg, Germany, 15 October 2021.

Athanasios-Alexandros Tsengenes:
"Computational Design of Neurotrophin Mimetics", International Conference on Neurotrophic Factors and Neurodegenerative Disorders; Current Advances and Future Perspectives, Heraklion (Crete), Greece, 27-29 August 2021.

Rebecca Wade:
"Molecular Recognition in silico", Systems Immunology Workshop, Heidelberg University, online, 18 February 2021. *"Zooming in on the dynamic interactions of proteins and drugs by computer simulation"*, DKFZ-ZMBH Alliance Seminar, Heidelberg, online, 3 May 2021. *"Molecular and Subcellular Simulation"*, (with Rui Ribeiro (Forschungszentrum Jülich), EBRAINS Summit, online, 12-14 October 2021.

Andreas Weidemann:
"Upload of EnzymeML documents in SABIO-RK", 2nd EnzymeML Workshop, online, 19 May 2021.

Ulrike Wittig:
"Introduction to FAIRDOM-SEEK", de.NBI Course "Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOMHub/SEEK", online, 22 April 2021; Data Management Workshop at MPI for Evolutionary Biology Plön, online, 18 - 19 May 2021. *"Research Data Management"*, First ERNEST Training School "Computational Tools to Study GPCR Signaling: From the Genomic to the Systems Level", online, 15 - 16 June 2021. *"Curation of EnzymeML documents in SABIO-RK"*, 2nd EnzymeML Workshop, online, 19 May 2021. *"Research Data Management in Germany"*, ELIXIR CONVERGE WP1 Meeting, online, 28 June 2021. *"Research Data Management for Collaborative Projects"*, COST-EU "One Health Drug Discovery" Workshop, online, 21 July 2021. *"RDMkit - Research Data Management Best Practices"*, NFDI4Ing conference 2021, Award Nomination "Favorite FDM Tool", online, 27 September 2021.

9.3 Memberships

Camilo Aponte-Santamaría:
Referee Work: Biophysical Journal (2 assignments), Journal of physical chemistry (1 assignment), Journal of chemical information and modeling (1 assignment), Nature communications (2 assignments).

Tilmann Gneiting:
Affiliate Professor, Department of Statistics, University of Washington, Seattle, Washington, United States. Member, Steering Committee, Karlsruhe Institute of Technology, Center MathSEE: Mathematics in Sciences, Economics and Engineering. Member, Ensemble Advisory Board, United States COVID-19 Forecast-Hub. Member, Committee on Publications, Institute for Mathematical Statistics.

Martin Golebiewski:
Convenor (chair) of the ISO/TC 276 Biotechnology working group 5 "Data Processing and Integration", International Organization for Standardization (ISO), reelected 2021. 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). Member of the steering committee of the German National Research Data Infrastructure for Personal Health Data (NFDI4Health). 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 steering committee of the AIMe registry for artificial intelligence in biomedical research. Member of the IEC SEG 12 Bio-Digital Convergence, International Electrotechnical Commission (IEC).

Frauke Gräter:
Max Planck Fellow of the Max Planck School Matter to Life. Member of DFG Graduate school 'Tailored Scale-Bridging Approaches to Computational Nanoscience' at KIT (GRK 2450). Member of the Biophysical Society, the German Physical Society, the German Biophysical Society, the German Chemical Society. Faculty member, Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg. Associated faculty member, HGS MathComp Graduate School, University of Heidelberg. Faculty member, Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology (HBIGS), University of Heidelberg. Co-Speaker of the Biological Physics division within the Condensed Matter Section of the German Physical Society (2019 - 2021). Member of the coordinating committee of the excellence cluster "3D Matter Made to Order" (KIT and Heidelberg University). Member of the Executive Board, Flagship Engineering Molecular Systems, Heidelberg University, since 2020. Referee Work: Biophysical Journal, Journal of the American Chemical Society, PLoS Journals, Nature Journals, Proceedings of the National Academy of Sciences, eLife, Angewandte Chemie, ChemBioChem, Proteins, Nano Letters, J. Phys. Chem. B, German Research Society (DFG).

Wolfgang Müller:
Deputy Chairman of SIG 4 (Infrastructure & data management), German Network for Bioinformatics Infrastructure (de.NBI). Leadership Team of LiSyM research network Liver Systems Medicine.

Abraham Muñiz-Chicharro:
HGS Fellow Speaker, HGSMathComp graduate school, Heidelberg University.

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

Kai Polsterer:
International Astro Informatics Association (Vice President), Standing Committee on Science Priorities of the International Virtual Observatory Alliance, Deutsche Physikalische Gesellschaft (AKPIG), International Astrostatistics Association.

Friedrich Röpke:
Advisory Board: Sterne und Weltraum.

Alexandros Stamatakis:
Member of the steering committee of the Munich Supercomputing System HLRB at LRZ. 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. Affiliated Scientist, Paleogenomics and Evolutionary Genetics (PEG) lab, Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology Hellas, Greece.

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 (until June 2021).

Diaaeldin Taha:
Member of the hiring committee, American University in Cairo: MACT (Math and Actuarian Science). Chair of MACT institutional memory comittee.

Jonathan Teuffel:
Spokesperson, Heidelberg students' group of the German society for Biochemistry and Molecular Biology (Junior-GBM Heidelberg). Vice Head, Junior Society for Biochemistry and Molecular Biology in South-West Germany.

Rebecca Wade:
Associate Editor: Journal of Molecular Recognition, PLOS Computational Biology. Section Editor: Molecular Informatics, International Journal of Molecular Sciences.
Editorial Board: Advances and Applications in Bioinformatics and Chemistry; BBA General Subjects; Journal of Chemical Information and Modeling; Journal of Computer-aided Molecular Design; Biopolymers; Protein Engineering, Design and Selection. 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: HBIGS (Heidelberg Biosciences International Graduate School) faculty, HGS MathComp Graduate School faculty, Helmholtz Information and Data Science School for Health (HIDSS4Health) faculty, Interdisciplinary Center for Scientific Computing (IWR), DKFZ-ZMBH Alliance of the German Cancer Research Center and the Center for Molecular Biology at Heidelberg University; Steering committee: Thematic Research Network "Molecular Mechanisms in health and disease – from understanding to engineering (MINDS)". Member of Managing Board of Directors, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University. Member of Informatics4Life Consortium.

Anna Wienhard:
Fellow of the American Mathematical Society. Member of Advisory Board, Springer Lecture Notes in Mathematics. Member of Advisory Board, Mathematisches Forschungsinstitut Oberwolfach. Member of Scientific Committee Wissenschaftskommunikation.de. Member of Heidelberg Academy of Sciences. Member of Berlin-Brandenburg Academy of Sciences and Humanities. Member of Selection Committee, Heiner-Maier Leibnitz Preis der Deutschen Forschungsgemeinschaft. Member of the Stiftungsrat of the Oberwolfach Stiftung. Editor: Annales Henri Lebesgue; Annales scientifiques de l'École Normale Superieure; Geometric and Functional Analysis; Geometry & Topology; Proceedings of the London Mathematical Society.

Ulrike Wittig:
Member of the STRENDA Commission (Standards for Reporting Enzymology Data). Editorial board member of ELIXIR Research Data Management Kit (RDMkit).

9.4 Contributions to the Scientific Community

Program Committee Memberships

Sucheta Ghosh:
The 35th AAAI Conference on Artificial Intelligence (AAAI), Virtual Conference, 2-9 February 2021. (Program Committee). The 18th International Conference Applied Computing 2020, Virtual, 13-15 October 2021. (Program Committee). The 22st Annual Conference of the International Speech Communication Association- Interspeech, Brno, Czechia, 30 August-3 September 2021 (Technical Program Committee). The 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Bangkok, Thailand. 1-6 August 2021. The 43rd Annual Meeting of the Cognitive Science Society, Vienna, Austria, 26-29 July 2021.

Martin Golebiewski:
Joint Workshop of the European Commission and CEN/CENELEC "Putting Science into Standard - Organ-on-chip: Toward standardization", 28-29 April, 2021, online: Chair and moderator of the session "Standards for data acquisition and management".

Kai Polsterer:
Data Science in Astronomy at EAS 2021, Leiden.

Michael Strube:
Program Co-Chair of CODI 2021, The Second Workshop on Computational Approaches to Discourse at EMNLP 2021, Punta Cana, Dominican Republic, 10-11 November 2021. Program Co-Chair of CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue at CODI 2021, Punta Cana, Dominican Republic, 10 November 2021.

Workshop and Conference Organization

James Farre:
Organizer of the Geometry and Topology seminar, Yale University, USA, 9 November 2021.

Martin Golebiewski
Host and chair of Committee Meetings of ISO/TC 276 Biotechnology working group WG5 "Data Processing and Integration", online, 27 January - 8 February 2021 and 19 - 26 July 2021. Chair of the NFDI4Health TA2 Meeting "Standards for FAIR Data", Annual NFDI4Health Consortium Meeting, Göttingen (Germany), 26 October 2021.
Co-host of HARMONY 2021 - The COMBINE Hackathon on Resources for Modeling in Biology, online, 22-26 March 2021.
Co-host and session chair of the COMBINE 2021 Online Forum: 12th Computational Modeling in Biology Network Meeting, online, 11 - 15 October 2021.

Saskia Haupt and Vincent Heuveline:
Two-session minisymposium organization: "Mathematical Oncology: From methodological studies to clinical applications", MS11 & MS12 at SMB2021, Oncology subgroup, online, 13–17 June 2021.

Saskia Hekker and Friedrich Röpke:
SOC member, IMPRS Summer School on 'Ecosystems', Heidelberg, Germany (online), 13 – 17 September 2021.

Olga Krebs:
OC Member of ELIXIR LU online course" Best practices in research data management and stewardship" 14-17 June 2021. OC Member of ELIXIR LU online course" Best practices in research data management and stewardship" 5 - 8 October 2021.

Georgios Kydonakis:
Co-organizer for the seminar: Spectral Networks and the WKB method, University of Strasbourg, France, 7-11 June 2021.

Brice Loustau:
Co-organizer of the HEGL Community seminar, Heidelberg University, winter semester 2021-2022.

Marta Magnani:
Organizer of RTG seminar, RTG retreat, Heidelberg University, 7-8 November 2021.

Wolfgang Müller:
LiSyM Jamboree 2021, online, 10-12 May 2021.

Merik Niemeyer:
Co-organizer of the Junior Seminar Geometry and Dynamics, Heidelberg University, winter semester 2021-2022.

Ariane Nunes-Alves:
Co-Organizer of "LatinasenQuimica – Desayuno Global", virtual networking event, 9 February 2021.

Kai Polsterer:
Co-organizer of Astroinformatics 2021. Co-organizer of E-Science and E-Infrastructure Splinter at AG-Tagung 2021. Co-organizer of IVOA KDIG meeting during spring interop.

Beatrice Pozzetti:
Co-organizer of the mini-workshop "Anosov^3" at Oberwolfach, with B. Küster, C.Guillarmou and T. Weich, 5-11 December 2021.

Anja Randecker:
Co-organizer of the STRUCTURES CP7 lunch seminar, Heidelberg University, winter semester 2021-2022. Co-organizer of the HEGL Community seminar, Heidelberg University, winter semester 2021-2022.

Carmen Rovi:
Organizer of the TACOS ("Topology, Algebra, Combinatorics, and Operators") seminar at Loyola University Chicago, fall 2021.
Co-organizer of the reading group "Geometric Multivector Analysis", fall 2021, and the reading group "Understanding Diversity", Loyola University Chicago, summer 2021.

Fabian Schneider:
Main organizer of the Virtual VFTS Workshop, HITS, Heidelberg, Germany.

Alexandros Stamatakis:
Organizer of 2020 Computational Molecular Evolution Summer School, Heraklion, Crete, Greece. The summer school was postponed to 2021 and finally cancelled. Associate Editor, Joint Oxford & ISMB Journal: Advances in Bioinformatics, since 2021.

Rebecca Wade:
Reviewer, HCERES Evaluation, Department of Structural Biology and Chemistry, Institut Pasteur, Paris, France.

Rebecca Wade, Ariane Nunes-Alves, Abraham Muniz Chicharro and Stefan Richter:
Organizers, together with Rommie Amaro, Lane Votapka (University of California San Diego): Biological Diffusion and Brownian Dynamics Brainstorm 5 (BDBDB5), online, 24-25 March 2021.

Other contributions

Valentina Disarlo:
Project Leader in the DFG SPP 2026 “Geometry at Infinity” for Project 44: “Actions of mapping class group and its subgroups”. Deputy Speaker for YRC STRUCTURES. Co-organizer of the 47th Heidelberg Graduate Days. Organizer of the seminar “Hyperbolic Geometry & Data Science”.

Saskia Haupt:
“Modeling multiple pathways of carcinogenesis using the Kronecker structure: Behind the paper.”, The Mathematical Oncology Blog, 7 June 2021. “Using mathematics for deciphering molecular pathways of carcinogenesis in hereditary cancer syndromes: Original perspective.” The Mathematical Oncology Blog, 29 December 2021.

Saskia Hekker:
“Der Soundcheck der Sterne“, Article in „Pro Physik”, March 2021 https://www.pro-physik.de/physik-journal/maerz-2021. “The internal structure of stars”, Article in “Ruperto Carola”, research magazine of Heidelberg University, July 2021. https://heiup.uni-heidelberg.de/journals/index.php/rupertocarola/article/view/24368/18120

Pengfei Huang:
European Kovalevskaya Travel Grant for attending ICM 2022, Deutsche Mathematiker Vereinigung/European Mathematical Society.

Georgios Kydonakis:
Member of the RTG 2229: Asymptotic Invariants and Limits of Groups and Spaces. Co-organizer for the Heidelberg-Karlsruhe RTG Colloquium.

Brice Loustau:
Co-organized the creation of the Heidelberg Experimental Geometry Lab (HEGL).

Mareike Pfeil:
DFG SPP 2026 Geometry at Infinity in Project 28 : Rigidity, deformations and limits of maximal representations.

Kai Polsterer:
“From Photometric Redshifts to Improved Weather Forecasts”, talk at the virtual Alexander von Humboldt Lab Visit, HITS, Heidelberg (Germany), 24 June 2021 (online).

Beatrice Pozzetti:
Principal Investigator in the RTG 2229: “Asymptotic Invariants and Limits of Groups and Spaces”. Project Leader in the DFG SPP 2026 “Geometry at Infinity” for Project 28: “Rigidity, deformations and limits of maximal representations” and Project 71: “Rigidity, deformations and limits of maximal representations II”. Leader of the DFG Emmy Noether (Projektnummer 427903332).

Anja Randecker:
Principal Investigator in the RTG 2229: “Asymptotic Invariants and Limits of Groups and Spaces”. Project Leader in the DFG SPP 2026 “Geometry at Infinity” for Project 72: “Limits of invariants of translation surfaces”.

Anna Schilling:
Co-organizer of the “Girls Day online workshop” on the Euler characteristic, April 2021.

Anna Wienhard:
Co-Spokesperson Cluster of Excellence “STRUCTURES - a unifying approach to emergent phenomena in the physical world, mathematics, and complex data”. Co-Spokesperson DFG Research Training Group 2229 “Asymptotic invariants and limits of groups and spaces”, Heidelberg – Karlsruhe. Member of Executive Committee, DFG Priority Program 2026 “Geometry at Infinity”. Member of Executive Committee, NSF Research Network in the Mathematical Sciences “Representation varieties and geometric structures" (GEAR). Member of the Structure Committee of the IMU. Scientific Chair of the Heidelberg Laureate Forum Foundation. Member of the Board of the Collaborative Research Center 191 “Symplectic Structures in Geometry, Algebra and Dynamics”. Member of the Scientific Committee of the IMU Centennial Conference. Scientific Director, Research Station Geometry and Dynamics, Heidelberg University. Scientific Director, UPSTREAM Mentoring Network, Heidelberg University.

9.5 Awards

Giulio Belletti:
Humboldt Foundation fellowship, June-November 2021. Hadamard Foundation fellowship, December 2021.

Johannes Bracher:
Arthur-Linder Prize, Austro-Swiss region of the International Biometric Society, 2021.

Fernando Cadena:
HITS Isabel Rojas Travel Award. STRUCTURES Young Researchs Convent Travel Award. GEAR Graduate Internship Program.

Valentina Disarlo:
MOXOFF Brand Ambassador 2021.

Ganna Gryn’ova:
Selected Schiemann Fellow of the Max Planck Society.

Saskia Haupt:
Best presentation: 1st price for the best presentation Y-EHTG. Talk title: Mathematically modeling Lynch syndrome colorectal carcinogenesis at different scales. 5th European Hereditary Tumour Group (EHTG) Meeting 2021, 8 October 2021 – 9 October 2021.

Mitul Islam:
Rackham Pre-doctoral Fellowship, University of Michigan, USA, July 2020-June 2021.

Dani Kaufman:
Alexander Prize, University of Maryland, USA.

Abraham Muñiz-Chicharro:
Isabel Rojas Travel Award (October 2021) to visit University of California San Diego (Amaro Lab).

Ghulam Qadir:
Al-Kindi Statistics Student Research Award, King Abdullah University of Science and Technology (KAUST), 2021.

Carmen Rovi:
Summer research grant ($7000), Loyola University Chicago, December 2021.

Juliette Schleicher:
Dr. Sophie-Bernthsen Award, Faculty of Chemistry, Heidelberg University.

Fabian Schneider:
Ludwig Biermann Award of the German Astronomical Society.

Alexandros Stamatakis:
Highly Cited Researcher in Cross Field research, Clarivate Analytics, 2021. RECOMB 2021 conference test of time award for paper on "How Many Bootstrap Replicates are Necessary" (originally published in 2009), 2021.

Anna Wienhard:
Henriette-Herz-Scout, Alexander von Humboldt Foundation. ERC Advanced Grant 101018839 PosLieRep: Positivity in Lie groups and representation varieties.

10 Boards and Management

The HITS Scientific Advisory Board. From left to right: Wolfgang Müller (HITS Scientific Director 2020), Tony Hey (retired at the end of 2020), Frauke Gräter (HITS Scientific Director 2021), Alex Szalay, Victoria Stodden, Thomas Lengauer, Adele Goldberg, Barbara Wohlmuth, Dieter Kranzlmüller, Gert-Martin Greuel, Gesa Schönberger (HITS Managing Director), Jeffrey Brock (picture taken December 2019).

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 HITS research groups. It presents the results to the HITS management and makes recommendations regarding how to further improve the Institute's research performance. In 2021, 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 (Mathematical Research Institute of Oberwolfach), Germany
- **Prof. Dr. Dieter Kranzlmüller**, Ludwig Maximilians University, Munich, Director of the Leibniz Super Computing Center, Germany (Chair, SAB)
- **Prof. Dr. Thomas Lengauer**, Max Planck Institute for Computer Science, Saarbrücken, Germany
- **Prof. Dr. Alex Szalay**, Johns Hopkins University, USA
- **Prof. Dr. 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 (HITS Foundation)
Prof. Dr. Wilfried Juling
Member of the Board of Directors

Prof. Dr. Carsten Könneker
Member of the Board of Directors
© Mück/Klaus Tschira Stiftung

Heidelberg University
Prof. Dr. Jörg Pross
Vice-President of Research and Structure
© 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 Scientific Director is one of the group leaders appointed by the HITS shareholders for a period of two years and represents the Institute in all scientific matters vis-à-vis cooperation partners and the public.

Managing Director:
Dr. Gesa Schönberger

Scientific Director:
Prof. Dr. Frauke Gräter
(2021 – 2022)

Deputy Scientific Director:
Prof. Dr. Tilmann Gneiting
(2021 – 2022)

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 of 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, which is a subsidiary of the Klaus Tschira Foundation, 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 | © 2022 HITS gGmbH

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