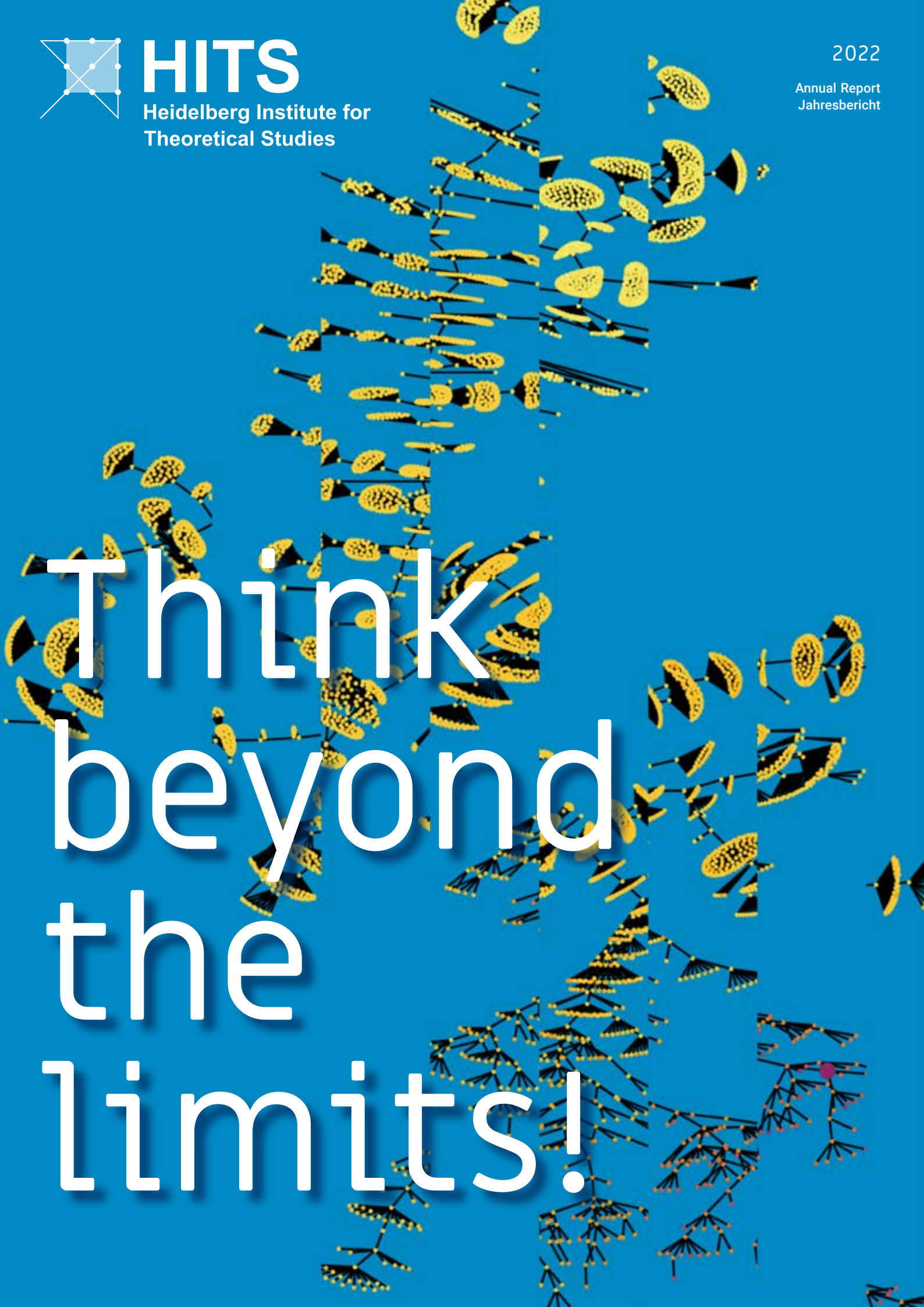
HITS
Heidelberg Institute for Theoretical Studies

2022

Annual Report
Jahresbericht



Think beyond the limits!

With the growing demand for alternative sustainable energy sources, electroactive organic molecules are becoming attractive electrode materials for rechargeable metal-ion batteries. However, the vast compound space of electroactive organic materials can only be sufficiently explored and rationally exploited in silico. Automated exploration that utilizes machine learning techniques can significantly facilitate property prediction and compound selection. In a project that is part of the SIMPLAIX strategic initiative (see chapter 7, p. 96), the Computational Carbon Chemistry group (CCC) has developed a methodology to meticulously explore the chemical space of redox-active organic molecules whilst rapidly and reliably predicting their redox potentials using a message-passing graph neural network. The image displays an exploration tree originating from methane (marked by a large dot with an arrow pointing towards it) and, through mutations, "growing" towards molecules with desired reduction potentials and synthetic accessibilities. (cf chapter 2.2 Computational Carbon Chemistry CCC, p. 16).

Die wachsende Nachfrage nach alternativen nachhaltigen Energiequellen lässt elektroaktive organische Moleküle als Material für wiederaufladbare Metall-Ionen-Batterien attraktiv werden. Die zahlreichen Verbindungen dieser organischen Materialien können jedoch nur am Computer ausreichend erforscht werden. Methoden maschinellen Lernens erleichtern hier die Vorhersage von Eigenschaften und die Auswahl von Verbindungen. In einem Teilprojekt der strategischen Initiative SIMPLAIX (siehe Kapitel 7, Seite 96) hat die Gruppe Computational Carbon Chemistry (CCC) hat eine Methode entwickelt, um den chemischen Raum redox-aktiver organischer Moleküle präzise zu erforschen und gleichzeitig deren Redoxpotenziale mithilfe eines Graph-Neuronalen Netzes schnell und zuverlässig vorherzusagen. Das Bild zeigt einen Explorationsbaum, der von Methan ausgeht (markiert durch einen großen Punkt mit einem Pfeil) und durch Mutationen zu Molekülen mit gewünschten Reduktionspotenzialen und synthetischen Zugangsmöglichkeiten "wächst". (Mehr dazu siehe Kapitel 2.2 Computational Carbon Chemistry CCC, S. 16).


CH4
1.0
0.8
0.6
0.4
0.2
0.0

1 Think beyond the limits!

Dr. Gesa Schönberger
(Managing Director / Geschäftsführerin)

Prof. Dr. Frauke Gräter
(Scientific Director / Wissenschaftliche Direktorin)

As a scientist recently said upon visiting HITS, "This place invites you to think!" It is thus no coincidence that Klaus Tschira and Andreas Reuter chose "this place" as the location for the Klaus Tschira Stiftung (Klaus Tschira Foundation) and later also for their scientific institute, HITS. Indeed, both men recognized the magic of the location and the park. Accordingly, in our HITS Strategy, we explicitly state that the Institute should be a place for creative thinking, for new ideas, and for lively scientific exchange.

To that end, we launched two new programs last year: the Klaus Tschira Guest Professorship Program and the HITS Independent Postdoc Program (see Chapter 6). The aim of both programs is to promote mutual scientific exchange, to give scientists the opportunity to advance their research during their time at HITS without any bureaucratic "red tape," and to enable networking with other HITSters in the process. Our hope is that the programs will spark new research initiatives and collaborations. In 2022, we welcomed Antonis Rokas from Vanderbilt University and Sarbani Basu from Yale University (both from the USA) as our first Klaus Tschira Guest Professors. Two further visits by guests from Oxford University (UK) and Carnegie Mellon University (USA) are planned for 2023. We were also able to recruit Rajika Kuruwita as the first HITS Independent Postdoc, who will work on a problem in astrophysics for 2–3 years in close collaboration with Fabian Schneider and his Stellar Evolution Theory group (see Chapter 2.15).

The great freedom of research that is available at our Institute was also one of the reasons that Jan Stühmer followed our call both to HITS and as a junior professor at KIT after having worked in industry for several years. Jan's group strengthens our activities in machine learning and thus also our pursuit of data-driven science (see Chapter 2.7). Last year, we were additionally forced to say both a happy and sad "goodbye" to Anna Wienhard and her Groups and Geometry group. We warmly congratulate Anna on her appointment as Max Planck Director in Leipzig. We will greatly miss her interdisciplinary perspective and her manifold scientific impulses, which stem from pure mathematics and are applied to computer simulations and machine learning methods.

We are also very pleased about the awards our scientists received in 2022, three of which we would like to highlight here. Ganna Gryn'ova – head of the Computational Carbon Chemistry group – was awarded an ERC Starting Grant, which she will use to break new ground in the development and optimization of organic functional materials (see Chapter 2.2). Moreover, Alexandros Stamatakis – head of the Computational Molecular Evolution group – won an ERA Chair from the EU, which will enable him to establish a group on biodiversity research at the Institute of Computer Science within the Foundation for Research & Technology, Hellas (ICS-FORTH), in Crete over the next five years (see Chapter 2.3). Yet again, Alexandros was recognized as a "highly cited researcher." These successes are an indication that

our Strategy – namely to establish HITS as a space for new ideas and creative minds and to fill this space with life – is gradually paying off.

Furthermore, our internal statistics also hint at our success: In 2022, individuals from 40 nations from around the world worked and conducted research at HITS. One group of people who were very important to us was represented by researchers from Ukraine, many of whom had been deprived of the ability to conduct their work at home due to the war. We thus established a study scholarship for refugee students who want to study at KIT or Heidelberg University in the fields in which HITS is active. Moreover, we also formulated an internal Code of Conduct for dealing with Russian and Russia-related institutions and researchers. Finally, the inhabitants of Heidelberg were able to get an impression of our space for creative thinking at our 2022 Open House. In addition to numerous lectures and hands-on stations, we also offered a guided tour of the park to mark its 100th anniversary (see Chapter 5.3).

In order to maintain HITS as a "magical" place for excellent science in the future and to visibly expand its basic research in the natural sciences, mathematics, and computer science internationally, the Institute completed an internal governance process in 2022. The goal of the process was to strengthen the Institute's scientific self-governance and the position of the Scientific Director. Since the summer of 2022, an amended "Gesellschaftsvertrag" (Shareholder Agreement) has been in force that will be used to provide us with an operational framework for further successful years.

It is our hope that our Annual Report 2022 will provide you with interesting insights into our research and activities from last year. We wish you lots of fun and creative thinking while reading!

„Dieser Ort lädt zum Denken ein!“ sagte eine Wissenschaftlerin kürzlich, als sie das HITS besuchte. Es kommt nicht von ungefähr, dass Klaus Tschira und Andreas Reuter diesen Ort zunächst für die Klaus Tschira Stiftung und dann für ihr wissenschaftliches Institut, das HITS, auserkoren haben. Sie hatten die Magie der Lage und des Parks erkannt. Entsprechend formulieren wir in unserer Strategie, dass das HITS explizit ein Ort für kreatives Denken sein soll – für neue Ideen und für einen regen wissenschaftlichen Austausch.

Dafür haben wir unter anderem zwei neue Programme ins Leben gerufen und im vergangenen Jahr zum ersten Mal mit Leben gefüllt: das Klaus Tschira Guest Professorship Program und das HITS Independent Postdoc Program (siehe Kapitel 6). Beide haben zum Ziel, den gegenseitigen wissenschaftlichen Austausch zu fördern, den Wissenschaftler*innen während ihrer Zeit am HITS die Möglichkeit zu geben, ihre Forschung ohne Wenn und Aber voranzubringen und sich dabei mit HITStern zu vernetzen. Unsere Hoffnung dabei ist, dass die Programme neue Forschungsrichtungen und Kollaborationen anstoßen. Als erste Klaus Tschira Gastprofessor*innen konnten wir 2022 Antonis Rokas von der Vanderbilt Universität und Sarbani Basu von der Yale Universität, beide aus den USA begrüßen. Zwei weitere Besuche von Gästen aus Oxford, Großbritannien, und der Carnegie Mellon Universität, USA, sind für das Jahr 2023 geplant. Als erste HITS Independent Postdoc konnten wir Rajika Kuruwita gewinnen, die für zwei bis drei Jahre an einem Problem der Astrophysik arbeiten wird, in engem Austausch mit Fabian Schneider und seiner Gruppe Stellar Evolution Theory.

Die große Forschungsfreiheit war auch einer der Gründe, warum Jan Stühmer unserem Ruf ans HITS und als Juniorprofessor am KIT gefolgt ist, nachdem er mehrere Jahre in der Industrie gearbeitet hat. Seine Gruppe verstärkt unsere Aktivitäten im Bereich des Maschinellen Lernens und damit unseren „Thread“ der datengetriebenen Wissenschaft (siehe Kapitel 2.7). Mit einem lachenden und einem weinenden Auge haben wir wiederum Abschied genommen von Anna Wienhard und ihrer Gruppe Groups and Geometry. Wir gratulieren ihr herzlich zu ihrem Ruf als Max Planck-Direktorin nach Leipzig. Ihren interdisziplinären Blick und ihre vielfältigen wissenschaftlichen Impulse werden wir sehr vermissen - aus der reinen Mathematik heraus in Computer-Simulationen und maschinelle Lernverfahren hinein.

Wir freuen uns ebenso sehr über einige Auszeichnungen unserer Wissenschaf-

ler*innen, von denen wir an dieser Stelle drei besonders hervorheben möchten. So ist Ganna Gryn'ova, Leiterin der Computational Carbon Chemistry-Gruppe ein ERC Starting Grant bewilligt worden, mit dem sie neue Wege in der Entwicklung und Optimierung organischer Funktionsmaterialien gehen will (siehe Kapitel 2.2). Alexandros Stamatakis, Leiter der Gruppe Computational Molecular Evolution hat einen ERA Chair der EU gewonnen, der es ihm erlaubt, in den nächsten fünf Jahren eine Gruppe zu Biodiversitätsforschung am Institut für Informatik der Foundation for Research & Technology, Hellas (ICS-FORTH) auf Kreta aufzubauen (siehe Kapitel 2.3). Alexandros wurde zudem zum wiederholten Male als „Highly Cited Researcher" ausgezeichnet. Diese Erfolge sind ein Hinweis darauf, dass unsere Strategie schrittweise aufgeht: nämlich HITS als Raum für neue Ideen und kreative Köpfe zu verstehen und mit Leben zu füllen.

Einen weiteren Hinweis liefert die interne Statistik: im Jahr 2022 haben Menschen aus 40 Nationen am HITS gearbeitet und geforscht. Eine uns wichtige Personengruppe waren Forschende aus der Ukraine, denen durch den Krieg vielfach die Arbeitsgrundlage genommen wurde. Für geflüchtete Studierende, die am KIT oder an der Universität Heidelberg in den am

HITS bearbeiteten Fächern studieren wollen, haben wir ein Studienstipendium aufgelegt. Wir haben darüber hinaus einen internen Code of Conduct formuliert, zum Umgang mit russischen und Russland-nahen Institutionen und Forschenden. Schließlich konnte sich die Heidelberger Bevölkerung am Tag der offenen Tür 2022 von diesem Ort für kreatives Denken ein Bild machen. Neben zahlreichen Vorträgen und Mitmach-Stationen haben wir, anlässlich seines 100-jährigen Bestehens, auch eine Führung durch den Park angeboten (siehe Kapitel 5.3).

Um HITS auch in Zukunft als ‚magischen' Ort für exzellente Wissenschaft zu erhalten und international sichtbar seine Grundlagenforschung in den Naturwissenschaften, der Mathematik und der Informatik auszubauen, hat HITS 2022 einen internen Governance-Prozess abgeschlossen. Ziel war es, die wissenschaftliche Selbstverwaltung und die Stellung des*der Wissenschaftlichen Direktors*in zu stärken. Seit Sommer 2022 ist nun ein geänderter Gesellschaftsvertrag in Kraft, der uns für weitere erfolgreiche Jahre einen Rahmen vorgeben wird.

Wir wünschen Ihnen viel Muße für kreatives Denken und beim Lesen des Jahresberichtes viel Vergnügen. Wir hoffen, dass wir Ihnen einen lebendigen Einblick in unsere Forschung und Aktivitäten des Jahres 2022 geben können.

2 Research
2.1 Astroinformatics (AIN)

Group leader
Dr. Kai Polsterer

Team
Dr. Nikos Gianniotis
Max Kahl (student assistant; since April 2022)
Fenja Kollasch (master's student)
Dr. Jan Plier (until February 2022)
Dr. Franciso Pozo Nuñez
Johanna Riedel (bachelor's student; since October 2022)
Solomiya Serkiz (scholarship holder; since April 2022)

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 volume of information in a bias-free manner.

Inferring the physical parameters of supermassive black holes from time series datas

Observations and theory suggest that galaxies harbor super-massive black holes in their centers. By accreting material, so-called active galactic nuclei (AGN) can outshine their host galaxy. As AGN are so bright, distance is not a limiting factor for our observations, which enables us to observe them even at extremely high redshifts (z ~ 7.0) and therefore also to infer some physical parameters and characteristics that enable us to study how the Universe itself has evolved.

We aim to estimate physical properties like the accretion rate or black hole mass of AGNs by inferring them through the delay estimation between light curves observed at different bandwidths.

A common approach to the problem of delay estimation between two general signals is cross-correlation. For a number of candidate delays, cross-correlation shifts one signal with respect to another in time. For each considered delay, cross-correlation calculates how much the two signals overlap. The estimated delay is then the one that corresponds to the best overlap. The very same approach has been adopted in astronomy for estimating the delay between two light curves. While the approach is intuitive, practical, and fast to calculate, it suffers from a number of disadvantages. For example, (a) it does not allow for a posterior distribution of delays, (b) it cannot jointly estimate the delays between more than two light curves, and (c) it does not produce out-of-sample predictions that are conditioned on a given delay.

In order to deal with the aforementioned issues, we formulated a probabilistic cross-correlation that is based on Gaussian processes. Our approach allows us to estimate a posterior distribution for the candidate delays, to work jointly with more than two light curves, and to make predictions for out-of-sample data. Roughly speaking, our approach puts forward a model that postulates a latent signal that underlies the light curves observed at different bandwidths. Each observed light curve is considered a scaled, offset (in flux), and shifted-in-time (i.e., delayed) version of the latent signal. This approach introduces a scaling parameter, an offset parameter, and a delay parameter for each considered light curve. Additionally, there is also a parameter that controls the length scale of the Gaussian process kernel function.

Given that our model views each observed light curve as an affine transformation of the latent signal, the resulting model is a multi-output Gaussian process with as many outputs as bandwidths. Our model allows us to easily marginalize the latent function values of the latent Gaussian process as well as the offset parameters. Regarding the scaling and length scale parameters, we resort to point estimates obtained by maximizing the model marginal likelihood. Unfortunately, we cannot analytically compute a posterior distribution for the delay parameters. Thus, we resort to computing the non-normalized posterior of the

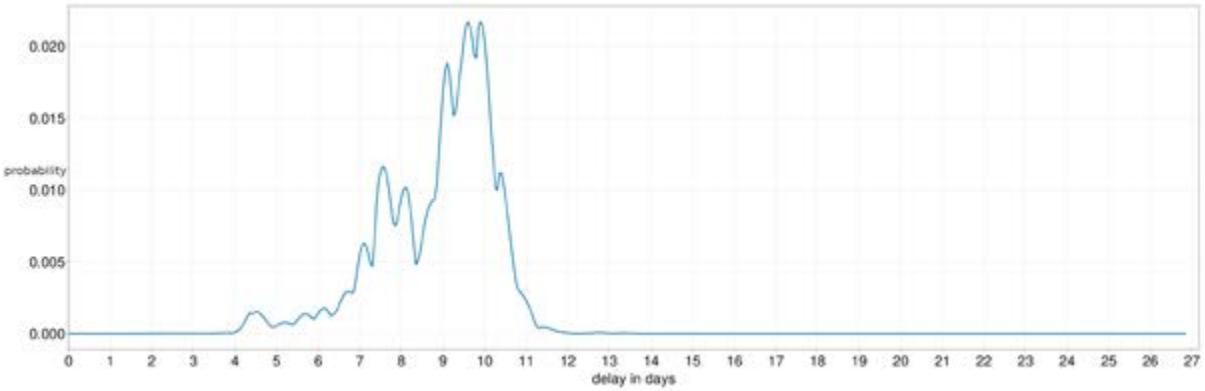


Figure 1: Posterior delay distribution for Mrk6. Our method reveals a highly multimodal distribution that is challenging to infer.

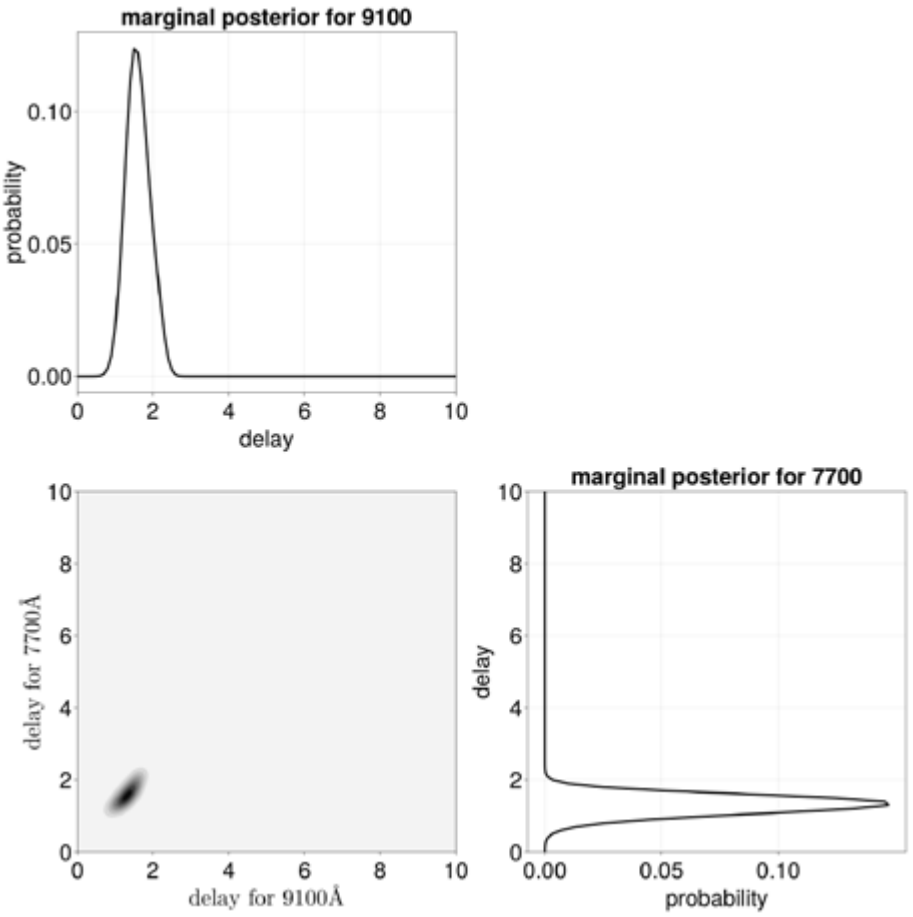


Figure 2: Joint delay posterior for MCG+08-11-011.

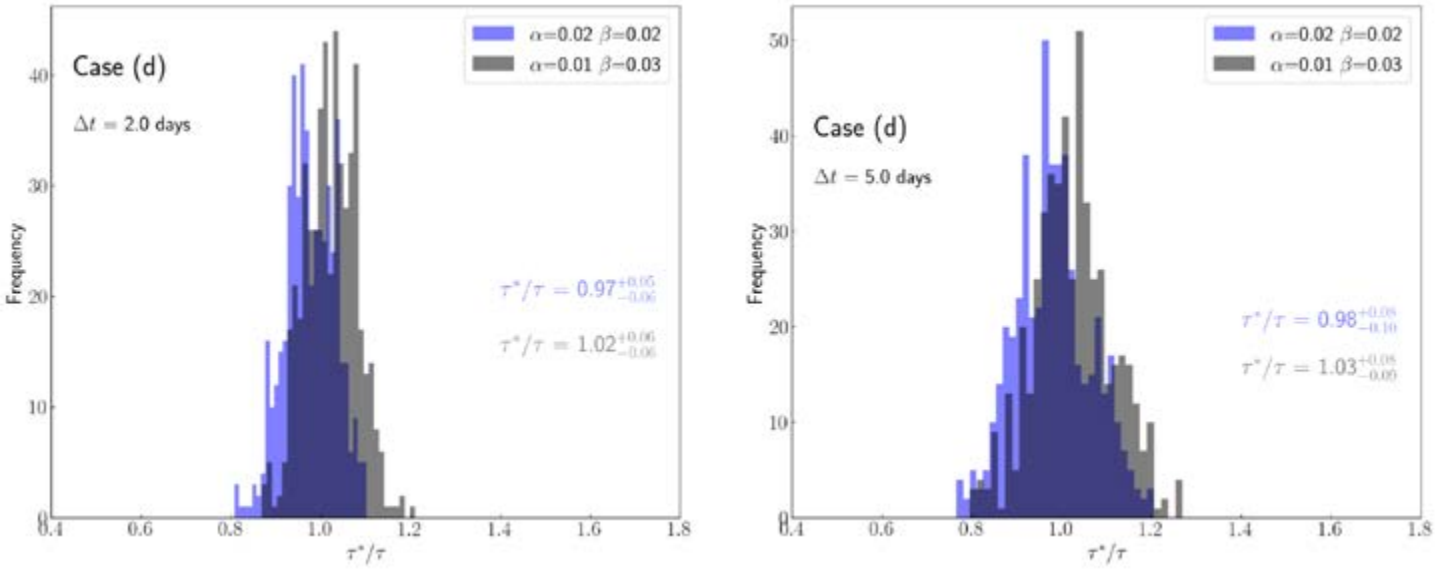


Figure 3: Recovered distributions of delays obtained for 1,000 mock light curves and specific α and β contributions from the BLR. The numbers indicate the median and the central 68% confidence (1σ) intervals of the distributions. Case(d) corresponds to quasars at redshift 1.5 < z < 2.0.

delays over a finite regular grid of delay combinations, and we then normalize these values into a multinomial distribution.

Figure 1 (previous page) displays the posterior distribution of candidate delays for an AGN object – called Mrk6 – as calculated by our approach. The posterior distribution shows a highly multimodal shape that is challenging to infer. The fact that our method provides out-of-sample predictions allows us to subject it to cross-validation for the purpose of comparing it against alternative, competitive models or for tuning certain design choices. In Figure 2 (previous page), we present an example in which we use our method to jointly estimate delays between three light curves. The object in question is called MCG+08-11-011, and we use the light curves at bands 5.100, 7.700, and 9.100 angstroms. The orientation of the joint posterior distribution reveals a positive correlation that signifies a trade-off between the two delays.

As quasar physics opens up the possibility to test cosmological models, our contribution to observational cosmology is represented by time series observations together with astrophysical models of the accretion disk (AD). Determining the size of the accretion disc is one of our major focusses and can be accomplished using photometric reverberation mapping to measure the time delays between light curves that are observed in different continuum bands. In order to recover the AD size–luminosity relationship and to potentially use quasars as standard candles, it is critical to quantify the constraints on the efficiency and accuracy of the delay measurements.

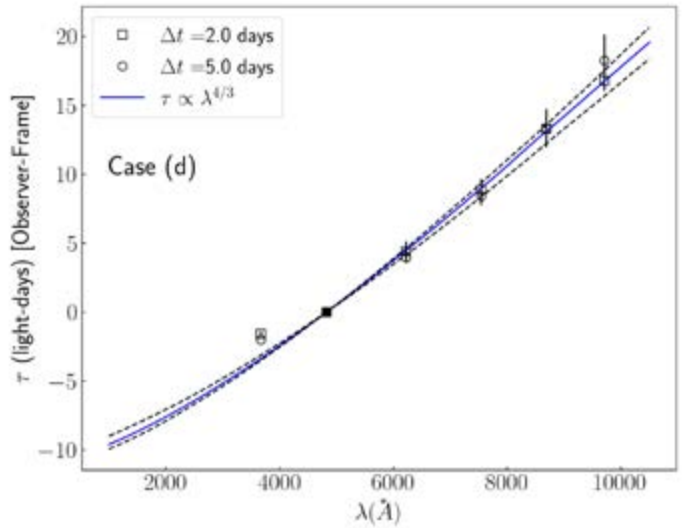


Figure 4: D time-delay spectrum (blue line) as predicted from physically motivated response functions. Filled squares and open circles denote the recovered delays at samplings of 2 and 5 days, respectively. The dotted lines show the delay spectrum obtained for a black hole mass with 30% uncertainty. Case(d) corresponds to quasars at redshift 1.5 < z < 2.0.

The forthcoming Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory is of particular interest. By the end of its ten-year operating lifespan, the LSST will have observed several thousand quasars with the Deep Drilling Fields and up to 10 million quasars with the main survey by using six broadband filters.

We developed extensive simulations that account for both the LSST survey characteristics and the intrinsic properties of the quasars. Light curves are characterized using the simulations, and the sizes of ADs are then calculated using various approaches. We find that for a particular redshift of 1.5 < z < 2.0, the size of quasar accretion discs can be recovered with an accuracy of 5 and 15% for light curves with a time sampling of 2 and 5 days, respectively. Given the actual LSST cadence, this redshift range performs best.

The redshift of the source and the relative contribution of broad line region (BLR) emissions to the bandpasses have a strong influence on the results.

Figure 3 presents an example of the

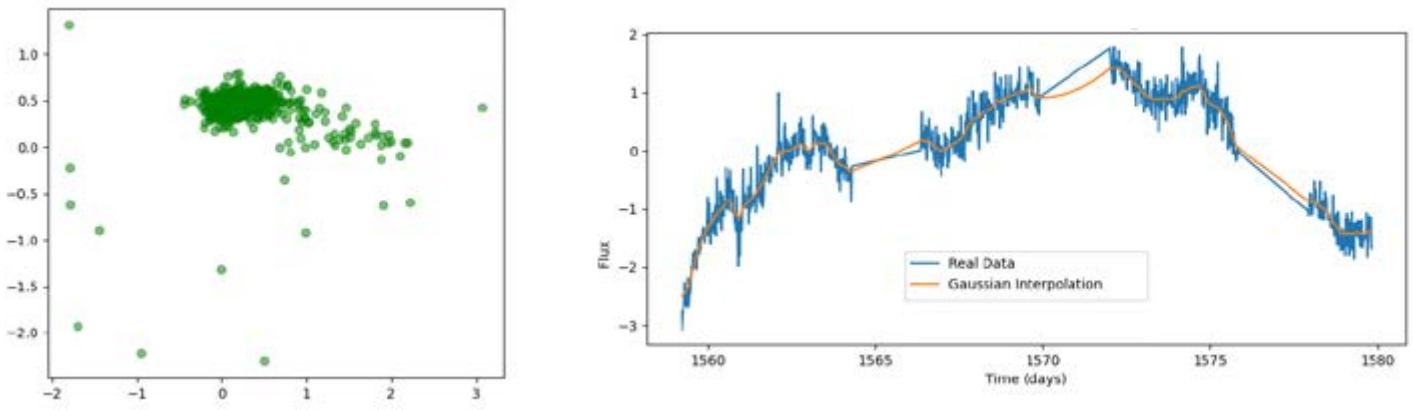


Figure 5: A set of 1,000 light curves as a point cloud (left); example of interpolating a light curve with a Gaussian process (right).

recovered distributions of delays obtained for 1,000 mock light curves from our catalogue and specific contributions from the BLR emission.
The recovered time-delay spectrum is consistent with black hole masses that are calculated with 30% uncertainty, assuming an optically thick and geometrically thin AD (see Figure 4).
More information on these simulations can be found in [Pozo Nunez et al, 2022].

Interactive exploration of astronomical data

As astronomical data come in various forms (e.g., spectra, time series, images), developing generalized and versatile solutions for data analysis is a challenging task. Our goal is to develop a toolkit that combines machine learning with an interactive exploration of pre-processed data by humans and thereby to realize Licklider's idea of human–machine symbiosis.
In recent years, we introduced UltraPINK, which is a web application for exploring the visualizations generated by self-organizing Kohonen maps (SOMs). SOMs reveal common patterns in the dataset, which makes them well-suited for tasks such as morphological classification. By relying on an abstract design pattern, UltraPINK represents the first step toward implementing an exploration pipeline.
We have used a similar approach to handle large datasets of light curves. A particular challenge with this type of data is its irregular sampling over time. Aside from non-trivial behaviors, observations are not made at regular intervals, which leads to gaps in the sequence. Irregular sampling precludes vectorizing the light curves, which is a common way of handling time series. We address observation gaps by modeling light curves with Gaussian processes, which allows us to interpolate the light curves in a plausible manner (i.e., we can avoid introducing artefacts) so that they can be re-sampled on a common, regular grid of time steps. Once all the light curves have been re-sampled, a second problem still remains: namely that the time series may have variable lengths simply because they were originally observed over different timespans.
In order to produce a visualization for the variable-length light curves, we are currently training a recurrent neural network on the entire dataset of light curves. Our approach is unique in that the recurrent neural network uses the same internal (hidden) weights for all light curves but learns a separate output weight vector for each individual light curve. This approach leads to the formation of a coherent state space and output vectors that uniquely characterize each light curve. Reducing the dimensionality of these output vectors causes the light curves to appear as a point cloud in which spatial properties express similarity (see Figure 5). We created a prototype that allows basic user interaction with the point cloud, and this prototype demonstrates the possibilities that the explorative analysis of time series has to offer.

Computational cardiology

Given our interests in galaxy morphology and the analysis of massive datasets, we are also involved in a medical project that deals with morphological data at microscopic scales. The ultimate goal of the project is to create a tool that aids in diagnosing heart diseases. To that end, cardiomyocytes are treated in vitro with plasma from potential patients. After this treatment, a plethora of microscopic images are collected for analysis and classification with respect to the present – or ideally, non-present – disease. In the current stages of the project, medical conditions are still simulated by eight different pharmacological substances, which give rise – with non-treated cells – to nine distinct classes. Our role in this project is to pave the way from a massive image dataset for use in making class predictions.

The beginning of the project was characterized by improving the quality of the datasets. Normalizing images, correcting background illumination, and removing out-of-focus artefacts are just some components of the pre-processing pipeline, which has now been completed. In so doing, we diverted our attention away from the previously prioritized task of cell segmentation and toward the task of nucleus detection.

Based on cell patches around the detected nuclei, the current classification pipeline comprises a feature extractor – that is, a VisionTransformer or ResNet18, feature aggregation via a transformer model, and classification via a shallow network, as shown in Figure 6.

First experiments appeared promising, with test accuracies well above 95% for each of the nine classes, but a second look revealed that generalization to new laboratory trials is still unattainable. Indeed, testing on a withheld trial resulted in a significant drop in performance. This drop can be traced back to distributional changes between trials that violated the common i.i.d. assumption for training data. In our case, at least three different sources of variation were responsible for these changes: the inherent heterogeneity of cardiomyocytes, technical variations, and the fact that each trial had been conducted with different genetic material.

Empirical results based on the visualization of hand-crafted and learned features from both supervised and self-supervised settings support this hypothesis (see Figure 7).

By interpreting each trial as its own domain, we embedded our problem into a domain adaption or generalization framework, where the difference lay within the degree of knowledge of the (unlabeled) test domain. After only minor performance improvements using linear methods such as TVN or CORAL, Bayesian approaches such as COMBAT, and even learning-based approaches such as DANN, Mixup, and Deep CORAL, recent results that revolved around the idea of learning features that are relative to non-treated cells turned out to be promising.

Future work will mainly address the development of this idea, which could require carrying out more trials because it remains necessary to assess whether the current number of trials is sufficient to essentially learn a distribution of distributions.

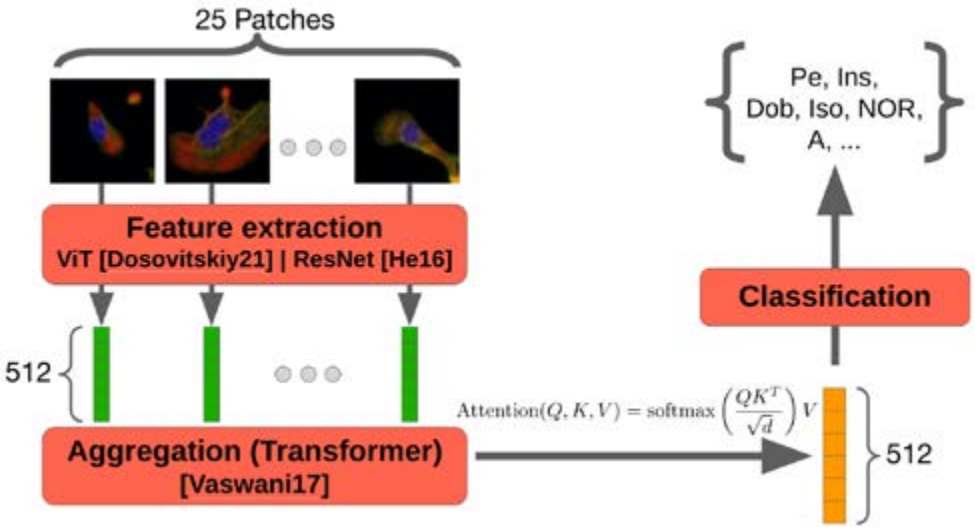


Figure 6: Summary of our current classification pipeline.

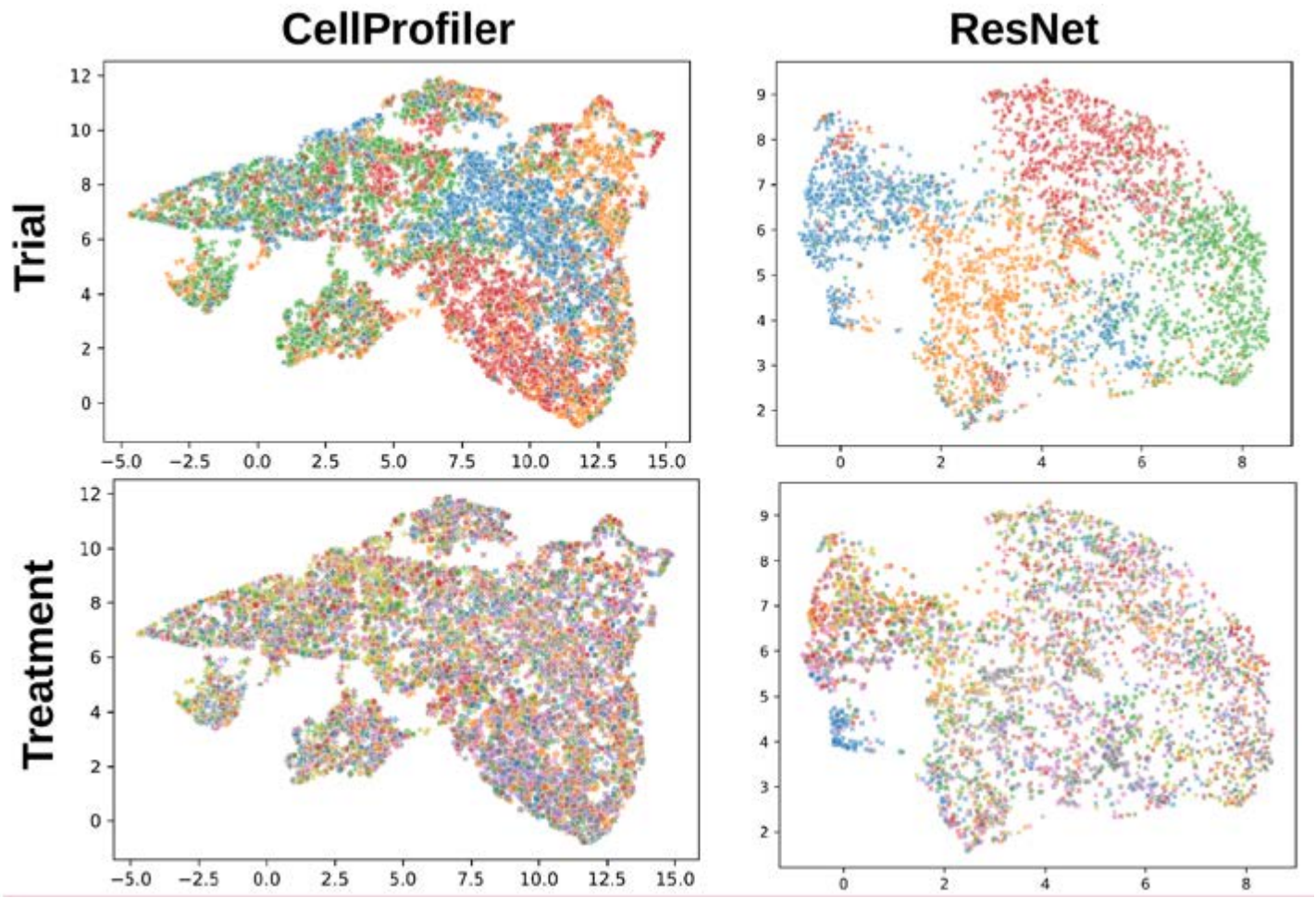


Figure 7: UMAP projections of high-dimensional handcrafted CellProfiler features in the left column and features that were learned through our pipeline in the right columns. All plots clearly show structure solely with respect to different trials and not with respect to different treatments, as is desired. Similar results are obtained if feature extraction is performed in a self-supervised fashion, for example, by SimCLR.

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 zum Beispiel Astroteilchen sowie Gravitationswellen, die neben bisher nicht beobachtbaren Wellenlängenbereichen ein vollständigeres Bild des Universums bieten.

Die **Astroinformatik** Gruppe befasst 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 beispielsweise 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 Informationsmenge zu gewährleisten.

2 Research

2.2 Computational Carbon Chemistry (CCC)

Group leader

Dr. Ganna Gryn'ova

Team

Sophia Ber (project student; Heidelberg University; March–April 2022)
Dr. Christopher Ehlert
Dr. Michelle Ernst
Dr. Abderrezak Torche (since October 2022)
Rostislav Fedorov (since June 2022)
Stiv Llenga
Anastasiia Nihei (visiting scientist; MSc student, V. N. Karazin Kharkiv National University, Ukraine)
Owen Paine (project student; Heidelberg University; October–November 2022)
Anna Piras
Wojtek Treyde (project student with Frauke Gräter and Robert Paton; Max Planck School "Matter to Life"; January–February 2022)

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 4th year at HITS, the group continued developing computational workflows to simulate complex materials, uncovered the fundamental mechanisms behind the interactions of these materials with small molecular targets, and used this knowledge to develop better sensors, catalysts, and nanocarriers. Within the ERC-funded project PATTERNCHEM "Shape and Topology as Descriptors of Chemical and Physical Properties in Functional Organic Materials," a new tool for structure decomposition and the structure–property analysis of covalent organic frameworks is currently being constructed. Within the SFB1249 „N-Heteropolyzyklen als funktionale Materialien," the largest database to date of N-heteropolycycles and their computed properties has already been assembled.

Machine learning became an increasingly important topic in the group last year. New molecular representations and similarity metrics were developed and benchmarked. Within the SIMPLAIX strategic initiative (see Chapter 7), a message-passing graph neural network is currently being developed to accurately predict the redox properties of electroactive molecules. Finally, fruitful collaborations with the Frauke Gräter (HITS) and Lutz Greb (Heidelberg University) groups continued, and new collaborations with the Peter Smillie (Heidelberg University), Ulrich Paetzold (KIT), and Bernd Schmidt (HHU Düsseldorf) groups also began in 2022.

Scale-bridging predictions of adsorption energies on graphene

Anna Piras

Nitrogen-containing aromatic compounds (NACs) such as 2,4-dinitrotoluene (DNT) and 2,4,6- trinitrotoluene (TNT) are omnipresent pollutants that can be found in explosive testing grounds, oil deposits, dye manufacture sites, and ore mines. These pollutants persist in soil and water for decades, resisting biodegradation and posing a serious risk to human health due to their facile absorption through the skin and their high bioaccumulation rates.

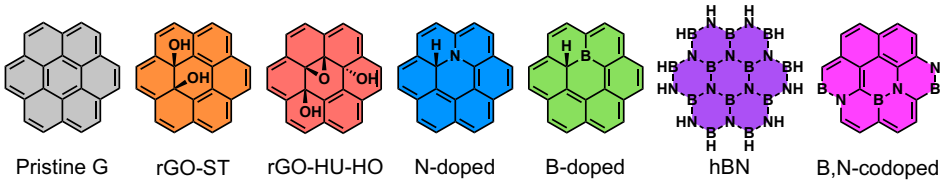


Figure 8: Coronene-sized models of graphene-based materials.

Consequently, detecting minute amounts of NACs is of great importance to public health and safety, forensics, and anti-terrorism operations. Among existing detection methods, electrochemical sensors that detect the reduction of nitro-groups to amino-groups enable real-time on-site analysis with the added benefits of low limits of detection, a large linear range, and a relatively low apparatus cost. The exceptional electrochemical and mechanical properties of various graphene-based materials (GBMs) make them particularly attractive for producing cheap, robust, and highly sensitive sensors. Over the past decade, a variety of metal-free graphene-based materials – such as electrochemically exfoliated graphene, hydrogenated graphene, reduced graphene oxide, and N-doped graphene – have been used to electrochemically detect DNT and TNT. Despite these formidable experimental efforts, a direct comparison between classes of materials is often obstructed by the disparate experimental conditions employed; thus, clear design principles that are grounded in a systematic understanding of NAC sensor chemistry are still lacking.

In order to establish these design guidelines, a combination of semi-empirical tight-binding quantum chemistry, meta-dynamics, density functional theory, and symmetry-adapted perturbation theory in conjunction with curated data from experimental literature were employed to investigate the physisorption of DNT and TNT on a series of functionalized graphene derivatives (Figure 8). To arrive at accurate estimates of adsorption energies on infinite (periodic) graphene, we employed our recent discovery – namely a way of extrapolating adsorption energy on infinite graphene from a series of adsorption energies that are computed for graphene nanoflakes of increasing sizes. This approach allows for estimating properties of periodic systems at a level of accuracy that is typically achievable for finite systems only. Our results reveal that the strongest adsorption of nitroaromatics occurs on B- and N-containing GBMs and – in particular – on graphene's fully doped analogue, hexagonal boron nitride (see Figure 9). In order to harness these design principles, we considered a series of boron and nitrogen (co-)doped two-dimensional materials. The system that features a chain of B–N–C units was found to adsorb nitroaromatic molecules more strongly than the pristine graphene itself. Interestingly, the slope of the computed extrapolation curves is a somewhat distinct property of a given material, and we are currently further investigating this tantalizing result. Whatever the physical origin of slope variation, it opens the door to size-selective graphene-based adsorbents. Finally, we also found that despite the highly polar nature of the analytes and sensing materials, their non-covalent interactions are largely driven by dispersion forces. These findings form the basis for the design principles of sensing materials and illustrate the utility of relatively low-cost in silico procedures for testing the viability of designed graphene-based sensors.

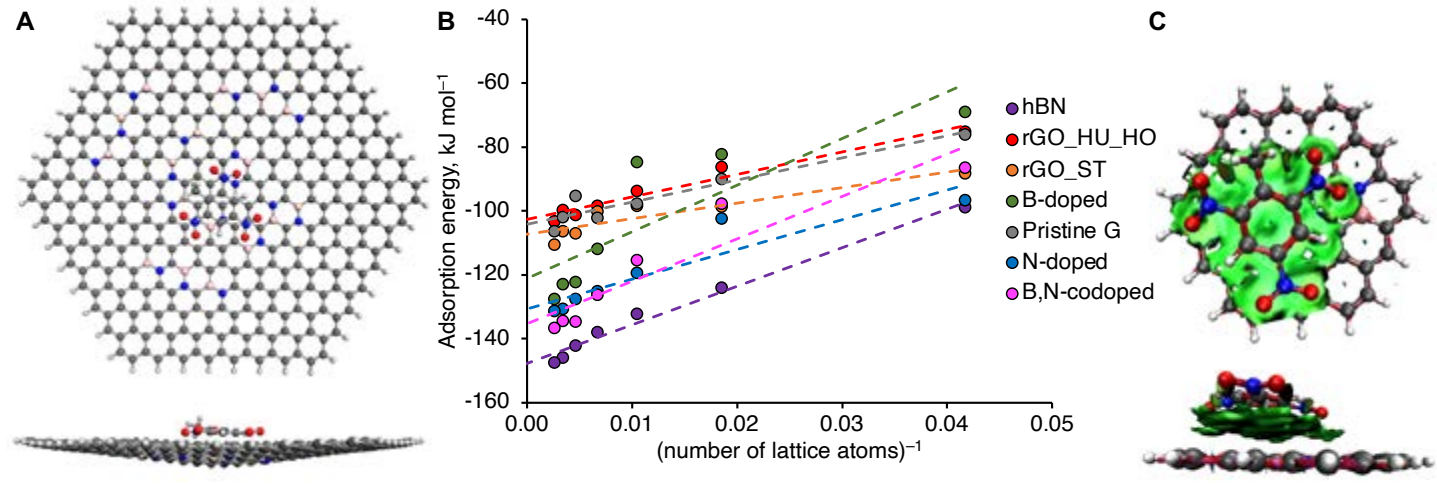


Figure 9: (A) Structure of the TNT adsorption complex on the largest modeled nanoflake of B,N-codoped graphene (top and side views). (B) Computed extrapolations for the adsorption of TNT on various GBMs. (C) Computed density overlap regions indicator (DORI) isosurfaces for TNT on a circumcoronene-sized B,N-codoped graphene model.

Learning redox properties of small molecules with graph neural networks

Rostislav Fedorov and Anastasiia Nihei

With the growing demand for alternative sustainable energy sources, diverse electroactive organic molecules are becoming attractive electrode materials for rechargeable metal-ion batteries. High capacity, facile synthesis, biocompatibility, structural flexibility, and tunable properties are among the many advantages of these molecules compared with their inorganic counterparts. However, the vast compound space of electroactive organic materials (i.e., chemical nature – e.g., quinones, carboxylates, nitroxyls – in conjunction with further functionalization via substitution and doping) can only be sufficiently explored and rationally exploited in silico. Beyond ab initio approaches to the individual compounds and properties, automated exploration that utilizes machine learning techniques can significantly facilitate property prediction and compound selection.

In order to overcome this challenge, we meticulously compiled an extensive database of over 120 experimentally measured reduction potentials, which were then benchmarked against four levels of density functional theory (PBE0, B3LYP, wB97X, and M06-2x) and two solvation models (CPCM and SMD). These results allowed us to identify the most reliable computational approach to estimating redox potentials, with a high coefficient of determination (R^2) of 0.94 and a low mean absolute error (MAE) of 0.21 eV. Next, we adopted a message-passing graph neural network (MPNN) for predicting redox potentials on a large scale. Unprecedented accuracy on a database of organic molecules OMEAD was achieved (R^2 = 0.92 and MAE = 0.20 eV) that surpassed the performance of the best approach based on natural language processing that has been reported in the literature. Finally, in an effort to find new chemically synthesizable molecules with desired redox properties, an evolutionary algorithm Evomol was combined with the MPNN to navigate the chemical space. The scoring function of this algorithm combines the literature synthesizability score with the value of the redox potential as predicted by the neural network. We can specify the desired redox potential to suit a given practical application, and we can specify the desired chemical nature to suit synthetic needs. This approach has shown promising results not only in discovering new battery materials, but also – and more broadly – in finding molecules with target redox properties (see Figure 10).

This project is part of the SIMPLAIX strategic initiative on bridging scales from molecules to molecular materials via multiscale simulation and machine learning.

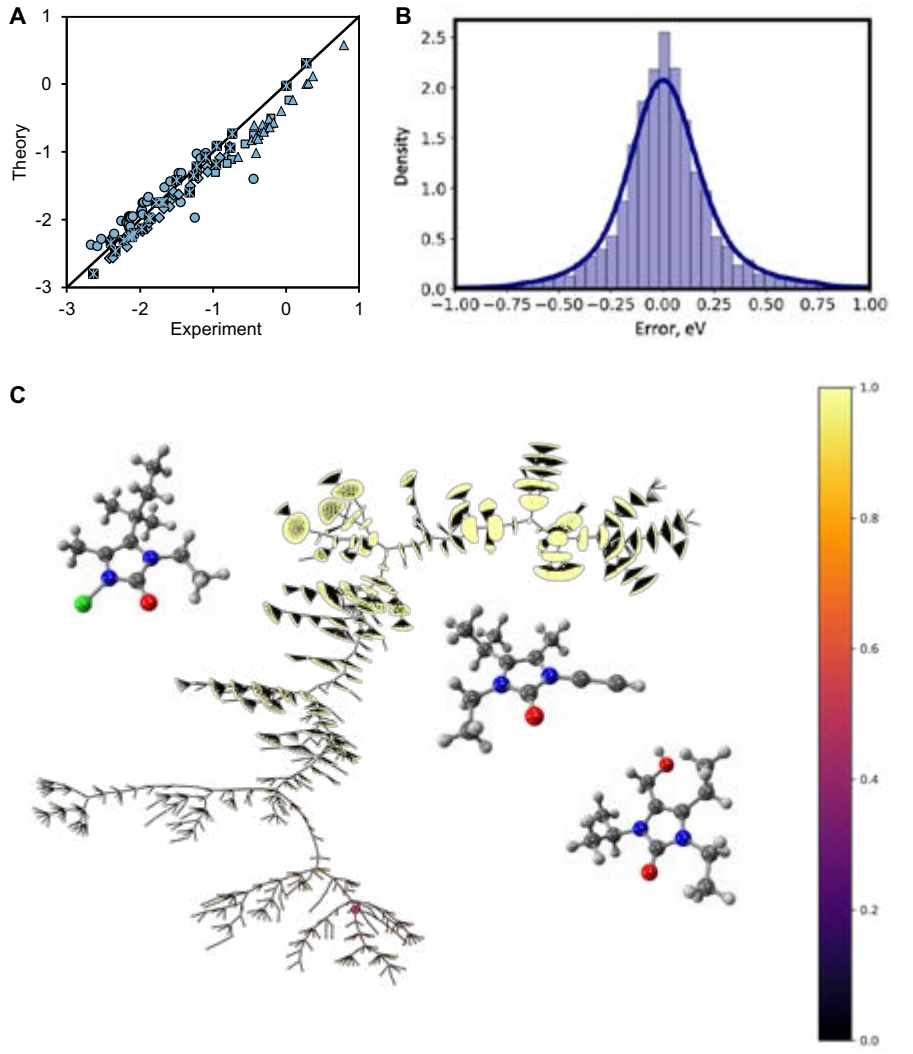


Figure 10: (A) Experimentally measured vs. computed reduction potentials of various organic electroactive molecules. Black line is x = y; symbols correspond to the source of the experimental data. Computations were performed with the M06-2X density functional and the SMD solvent model for acetonitrile. (B) Error distribution for MPNN on OMEAD database. (C) Exploration tree of a QED optimization run after 400 steps. The starting point is methane (purple circle). Edges represent mutations that lead to an improvement in the population. Solutions are colored according to the value of the scoring function. Molecular structures are representative top candidates for a small organic molecule with a synthesizability score of 3 and a reduction potential 0f −1.3 eV.

Quantum-inspired atomic and molecular representation

Stiv Llenga

Molecular (also often called chemical or quantum) machine learning (ML) has rapidly entered the domain of chemical science and facilitates both the development of simulation techniques and the design of better reagents, catalysts, functional materials, etc. Molecular representations are the prerequisite for machine-learning the chemical properties because these representations uniquely encode information about molecular composition and

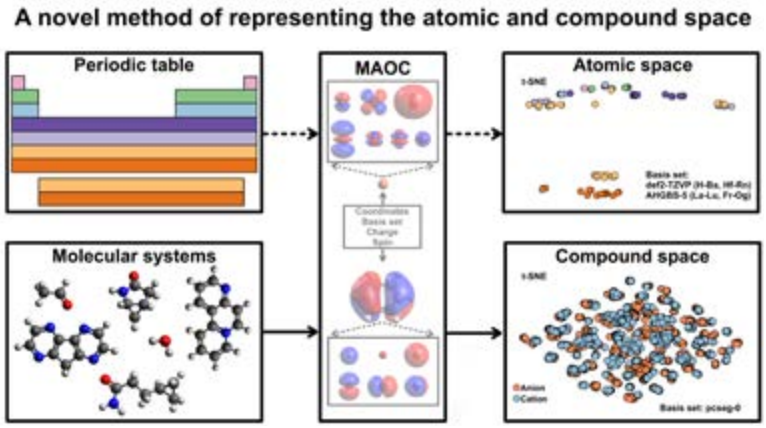


Figure 11: Capability of MAOC to represent atoms and molecules by projecting delocalized orbitals onto a pre-defined set of orthogonalized atomic orbitals. Anionic and cationic open-shell compounds from the N-HPC-1 dataset are used to map the compound space.

structure into a numerical format. Unfortunately, popular coordinate-based representations only consider the type of the chemical element (nuclear charge) and the position of the nuclei in space, assuming charge neutrality. As a result, these representations violate the injectivity requirement for the machine learning representations – that is, they are unable to distinguish between compounds with distinct electronic configurations but with identical atomic compositions and geometries. Therefore, these representations are not suitable for machine-learning the vertical or redox properties of rigid molecules.

In order to address this problem, we developed a new quantum-inspired molecular and atomic representation that contains both structural (composition and geometry) and electronic (charge and spin multiplicity) information (Figure 11). The matrix of orthogonalized atomic orbital coefficients (MAOC) is based on a cost-effective meta-Löwdin localization scheme that represents localized orbitals via a pre-defined set of atomic orbitals. These atomic orbitals can be constructed from such small atom-centered basis sets as pcseg-0 and STO-3G in conjunction with guess (non-optimized) electronic configuration of the molecule. Importantly, MAOC is suitable for representing monatomic, molecular, and periodic systems and can distinguish between compounds with identical compositions and geometries but with distinct charges and spin multiplicities. The performance of MAOC and several other coordinate- and Hamiltonian-based molecular representations was tested in conjunction with a kernel ridge regression machine learning model for predicting frontier molecular orbital energy levels and ground state single-point energies for chemically diverse neutral and charged, closed-, and open-shell molecules from an extended QM7b dataset as well as with two new datasets: N-HPC-1 (N-heteropolycycles) and REDOX (nitroxyl and phenoxyl radicals, carbonyl, and cyano compounds; Figure 12). MAOC affords accuracy that is either comparable or superior to other representations for a range of chemical properties and systems.This project is part of SFB1249 "N-Heteropolycycles as Functional Materials."

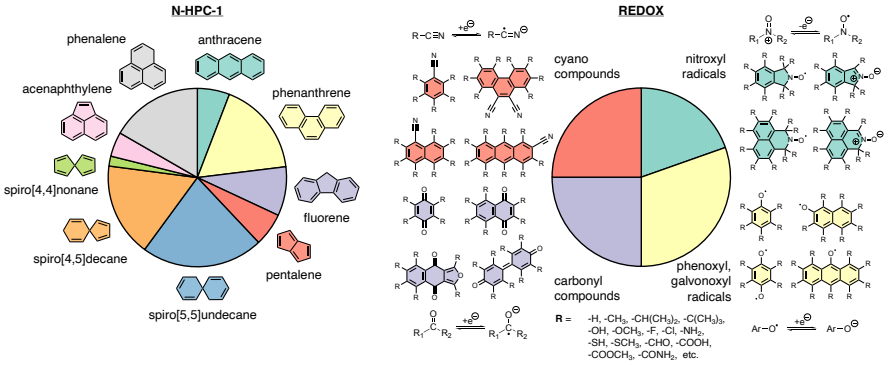


Figure 12: Composition of new datasets. N-HPC-1: Composition of the dataset according to the polycyclic skeleton, doped with nitrogen atoms. REDOX: Composition of the dataset according to both the type of redox-active molecules and the schemes of the corresponding redox reactions.

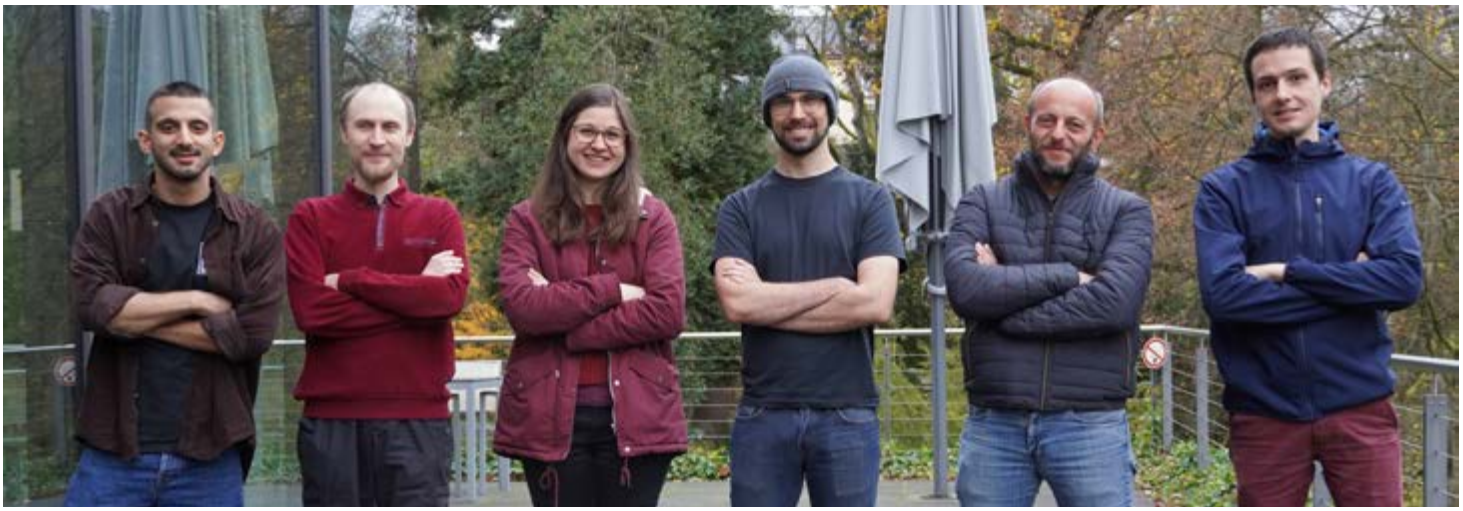
Moderne Funktionsmaterialien kombinieren strukturelle Komplexität mit zielgerichteter Performance und werden in verschiedenen Bereichen von Industrie und Forschung eingesetzt, von der Nanoelektronik bis hin zur Massenfertigung. Theoretische Studien dieser Materialien fördern mechanistische Grundlagen 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. In ihrem vierten Jahr am HITS entwickelte die Gruppe die Berechnungsabläufe zur Simulation komplexer Materialien weiter, deckte die grundlegenden Mechanismen hinter den Wechselwirkungen dieser Materialien mit kleinen Gasmolekülen auf und nutzte dieses Wissen zur Entwicklung besserer Sensoren, Katalysatoren und Nanocarrier. Die Forschenden arbeiten derzeit - im Rahmen des ERC-geförderten Projekts PATTERNCHEM („Form und Topologie als Deskriptoren chemischer und physikalischer Eigenschaften in funktionellen organischen Materialien") - an einem neuen Tool zur Gliederung von Strukturen und zur Analyse der Struktur-Eigenschaften kovalenter organischer Gerüstverbindungen. Innerhalb des SFB1249 „N-Heteropolyzyklen als funktionale Materialien" wurde bereits die bisher größte Datenbank von N-Heteropolyzyklen und ihren berechneten Eigenschaften aufgebaut.

Das maschinelle Lernen wurde im vergangenen Jahr für die Arbeit der Gruppe immer wichtiger. Neue molekulare Darstellungen und Ähnlichkeitskennzahlen wurden entwickelt und miteinander verglichen. Im Rahmen der SIMPLAIX-Initiative (siehe Kapitel 7) entwickeln die Wissenschaftler*innen derzeit ein Nachrichten austauschendes „message-passing graph neural network", um die Redoxeigenschaften von elektroaktiven Molekülen so genau wie möglich vorherzusagen. Und schließlich setzte die Gruppe ihre erfolgreiche Zusammenarbeit mit den Gruppen von Frauke Gräter (HITS) und Lutz Greb (Universität Heidelberg) fort. Im Jahr 2022 wurden außerdem neue Kooperationen mit den Gruppen von Peter Smillie (Universität Heidelberg), Ulrich Paetzold (KIT Karlsruhe) und Bernd Schmidt (HHU Düsseldorf) aufgenommen.

2 Research

2.3 Computational Molecular Evolution (CME)



Group leader

Prof. Dr. Alexandros Stamatakis

Team

Benjamin Bettisworth (PhD student)
Julia Haag (PhD student; HITS Scholarship since March 2022)
Luise Häuser (master's student)
Dimitri Höhler (PhD student)
Lukas Hübner (visiting scientist from KIT)
Dr. Alexey Kozlov (staff scientist)
Dr. Benoit Morel (postdoc)
Ioannis Reppas (student assistant)
Prof. Dr. Antonis Rokas (Klaus Tschira Guest Professor, Vanderbilt University, USA; June–September 2022)
Christoph Stelz (student assistant)
Jan Strehmel (bachelor's student)
Anastasis Togkousidis (PhD student; HITS Scholarship)
Qihao Yuan (student assistant)
Xinyi Zhang (master's student)

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,
- biodiversity quantification,
- 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.

Below, 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 these items available to evolutionary biologists.

In other words, we strive to support research. One aim of evolutionary biology is to infer evolutionary relationships between species on the one hand and the properties of individuals within populations of the same species on the other hand. 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 2022?

In the winter of 2021/2022, Alexis, Benoit, Alexey, and Lukas taught the Introduction to Bioinformatics for Computer Scientists online class at the Karlsruhe Institute of Technology (KIT).

During the summer semester of 2022, we again taught our main seminar, Hot Topics in Bioinformatics.

Our teaching activities in the winter term continued to be heavily affected by the pandemic. All oral exams for the class during the winter of 21/22 were thus also conducted online.

Julia Haag – our master's student from the Department of Computer Science at KIT – joined the lab as a PhD student in 2022.

Moreover, our recurring highlight – the summer school on Computational Molecular Evolution – finally took place again in 2022, this time in Hinxton, UK, after plans to hold the school on Crete had been postponed several times in 2020 and 2021 before finally being canceled. Ben and Benoit – who supported the summer school as teaching assistants – and Alexis – who served as a lecturer and co-organizer – greatly enjoyed the first in-person meeting with colleagues and students in July 2022 after a long pause of several years. Our next summer school is planned for May 2023, this time on Crete again.

Moreover, in 2022, Alexis was listed on the Clarivate Analytics list of highly cited researchers for the seventh year in a row (see Chapter 10).

We were additionally delighted to host our long-term collaborator Antonis Rokas, a professor at Vanderbilt University in the US, as the first Klaus Tschira guest professor at CME (see Chapter 7).

In 2022, Alexis received a 2.4-million EUR ERA chair grant from the EU to set up an additional research group at the Institute of Computer Science within the Foundation for Research & Technology – Hellas in Heraklion, Crete, Greece. The new group will be called the Biodiversity Computing Group (BCG) and will closely collaborate with both the CME group at HITS and the Computer Science Department at KIT in order to foster brain circulation. In addition, the research group will closely collaborate with local biodiversity research centers on Crete: namely the Natural History Museum of Crete and the Hellenic Center for Marine Research. Greece – and especially Crete, with its high abundance of endemic species – is a European Biodiversity hotspot. Therefore, the key goal of the group will be to develop energy-efficient and scalable open-source software for analyzing biodiversity research data. Beyond this, another goal of the ERA chair (A. Stamatakis) will be to foster institutional reforms in Greece and to work toward both reverting brain drain and fostering brain gain in the European periphery. In its initial phase, the BCG will hire 4 PhD students and 2 postdocs. The brain circulation between Greece and Germany has already been established, and plans are currently in place for CME PhD student Julia and master's student Luise to visit Crete in order to work on projects with local researchers that pertain both to uncertainty quantification in PCA analyses and to the inference of evolutionary trees of natural languages. ICS PhD student Angeliki Papadopoulou will in turn visit the CME to conduct technical work on missing data imputation using likelihood-based methods.

Last year also had a stronger focus on public outreach, particularly with the presentation of the new educational program for primary schools entitled "The Aegean Archipelago: A living laboratory of evolutionary biology," which Alexis tested in two remote Cretan primary schools with his colleagues from the Natural History Museum of Crete.

In sum, 2022 was dominated by the pandemic to a much lesser extent than the preceding two years had been, and we look forward to contributing toward establishing our BCG sister lab on Crete, which will open a new chapter in the history of the lab.

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, to 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 has faced novel challenges. One key problem that needs to be addressed is the fact that the volume of molecular data that are 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

flattened out somewhat in the last 3–4 years (see Figure 13, and https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost).

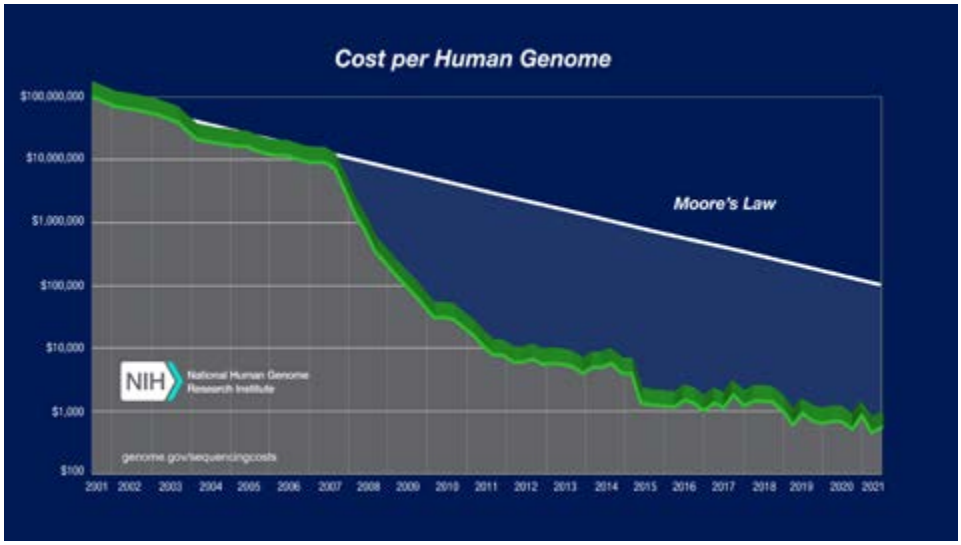


Figure 13: Cost of sequencing a human genome over time in comparison with the cost of computing according to Moore's law (source: National Human Genome Research Institute).

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.

In order 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 by using the example of RAxML-NG. In 2022, we also completed some initial work on methods of redundantly storing and efficiently redistributing 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 initially conducted research in this domain using the example of RAxML-NG because it is the most widely used and most scalable bioinformatics tool that is being developed and maintained in our group. Hence, RAxML-NG also generates the largest CO2 footprint. More extensive recent experiments that have used a broader and more representative set of scientific software tools – as contained, for instance, in the SPEC (Standard Performance Evaluation Corporation) benchmark suite – have revealed that adapting the CPU clock frequency as a function of the real-time contribution of renewable energy sources to overall electricity production can reduce costs (i.e., when more renewable energy is produced, electricity becomes cheaper) as well as energy-to-solution. More specifically, we have found that by slowing down computations by 10% via CPU clock frequency adaptation, we can achieve 20% savings in terms of energy and costs.

Overall, phylogenetic trees (i.e., evolutionary histories of species) – as well as the application of evolutionary concepts in general – are important to 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 (i.e., 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) of prehistoric times. Phylogenies can also be used to disentangle the evolution of natural languages in linguistics. We will begin exploring and understanding this application area of phylogenetic inference with our new collaborator Elena Anagnostopoulou, a theoretical linguist at the University of Crete.

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

Predicting the difficulty of a phylogenetic analysis

Just like every other lab does in order to maintain its "street credibility" in terms of conducting cutting-edge research, we now also apply machine learning methods to problems from evolutionary biology. However, our goal is not to conduct research as "art for art's sake," but rather to contribute something that is useful and usable. To that end, we developed a tool that relies on machine learning methods to predict the degree of difficulty of a phylogenetic analysis for a given input dataset. To achieve this goal, we initially developed a highly compute-intensive measure of phylogenetic difficulty that essentially quantifies the strength of the phylogenetic signal in a given dataset in order to generate ground-truth labels. Subsequently, we trained a machine learning model to predict this difficulty value in a substantially more compute-efficient way. Our newly introduced difficulty values range from 0 (easy) to 1 (hopeless). These values can be used to adjust the prior expectations of the end user with respect to the stability of the phylogeny to be inferred, and they thereby also enable us to appropriately adapt the analysis and

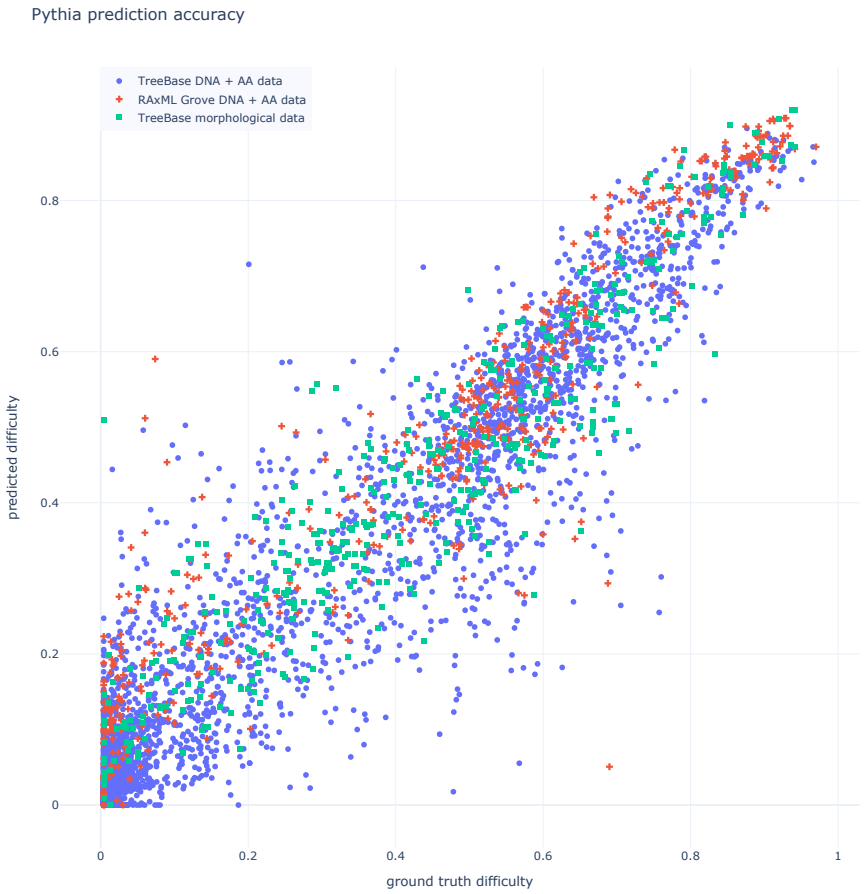


Figure 14: Correlation between predicted (y-axis) and ground truth (x-axis) tree-inference difficulty scores for molecular data from the RAxML-Grove database (red crosses) and the TreeBase database (blue circles), as well as morphological binary datasets (green rectangles).

post-analysis pipeline. For instance, easy datasets do not require compute- or CO2-intensive calculations because the signal is very strong. A classic example of difficult datasets is represented by SARS-CoV-2 datasets because they comprise huge numbers of sequences with relatively few mutations. For instance, the comparatively small (by current standards) SARS-CoV-2 dataset that we analyzed two years ago in order to highlight the difficulties of inferring large virus phylogenies has a difficulty score of 0.84. Apart from informing the user expectations and the analysis as well as the post-analysis setup, the difficulty-prediction score can also be used to automatically adapt the required degree of thoroughness of a phylogenetic tree-search algorithm with the goal of substantially speeding up the algorithm for easy datasets, which are far from uncommon. Initial results with some adaptive tree-search heuristics that use the predicted difficulty score as input are highly promising, and we hope to be able to write more about them in next year's Annual Report.

Using difficulty prediction to improve simulated data studies

The difficulty prediction tool now also allows for a more representative analysis of the behavior of maximum likelihood tree-inference programs because it enables us to characterize the difficulty distributions of datasets in large-scale empirical databases that contain inferred trees and the corresponding input data – that is, the multiple sequence alignments (MSA) of the inferred trees. More specifically, both for our own RAxML-Grove database – which comprises over 60,000 trees that have been inferred by users of RAxML – and for the TreeBase database, we predicted the difficulty score for every dataset contained therein. Interestingly, the difficulty distributions between these two databases were rather different (see Figure 15 below).

We can now draw datasets from these empirical distributions when setting up a representative performance study (i.e., in order to better match the datasets that are typically analyzed by practitioners) instead of simply using ad hoc parameters or ad hoc collections of benchmark datasets. This process allows for a more realistic assessment of program performance and – more importantly – also enables us to study the performance of various ML search heuristics for datasets with distinct difficulty levels. The process hence facilitates not only the selection of datasets, but also the interpretation of the results. It turns out that all inference methods perform similarly and exhibit decreasing accuracy with increasing

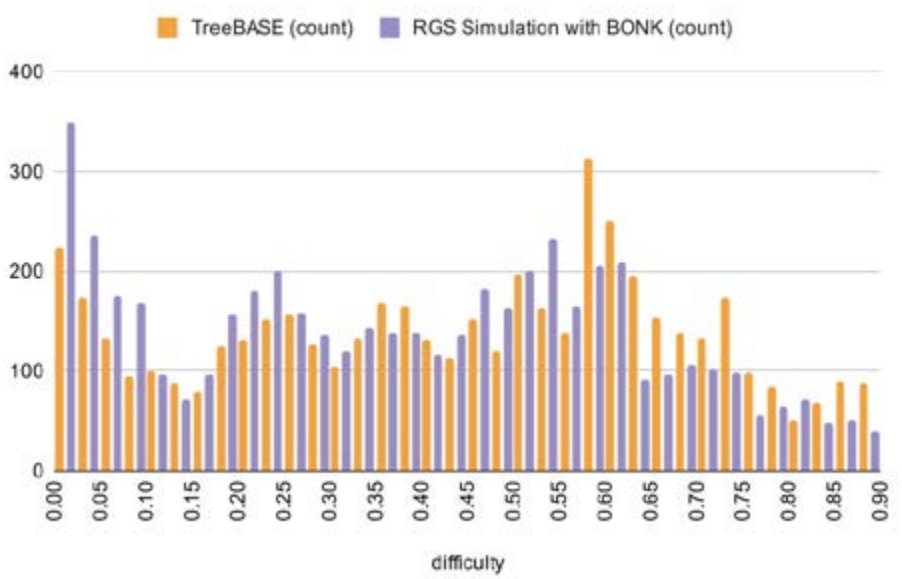


Figure 15: Difficulty distributions in the RAxML-Grove (RGS) and TreeBase databases.

difficulty. This very clear trend and correlation between increasing difficulty and decreasing accuracy also confirms that our definition of difficulty is meaningful.

Fault-tolerant computing: Rapid data recovery upon hardware faults

Our collaboration with Peter Sanders at KIT through our shared PhD student, Lukas, takes us back to basic computer science and high-performance computing. One key question in the area of fault-tolerant computing that strives to enable parallel programs to recover "on the fly" upon hardware failure involves how to recover the memory state that was stored in the RAM of a processor that has failed. In order to investigate this question, we developed the ReStore open-source code. When a processor fails, we need to efficiently redistribute the workload among the remaining intact processors, which also need to reload the lost data from the processor that has failed. ReStore is an algorithmic framework that is implemented as a C++ library for parallel MPI (Message Passing Interface) programs and that enables lost data to be recovered after one or more process failures by storing all required data in memory via an appropriate data distribution and data replication mechanism. Recovery from failure is therefore substantially faster than is recovery via standard checkpointing schemes that rely on a parallel file system.

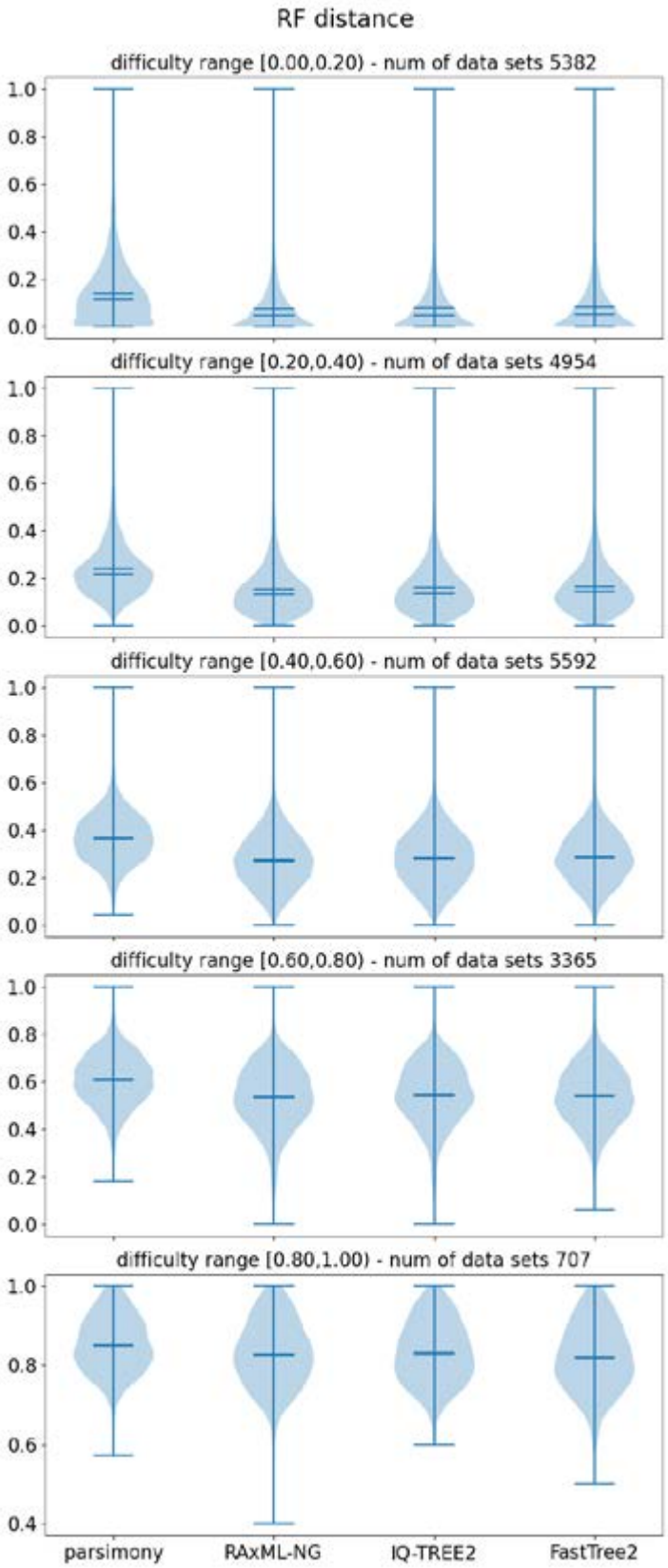


Figure 16: Topological distance in % (y-axis) to the true tree using simulated data in four distinct tree-inference methods (x-axis) and in 5 distinct difficulty ranges (increasing difficulty from top to bottom).

Because the application developer can explicitly specify which data to load, we also support shrinking recovery (i.e., continuing computations with fewer processors) instead of recovery using spare processors (i.e., continuing computations with the same number of processors by requesting/obtaining additional processors to replace any that have failed). We evaluated ReStore in both controlled and isolated environments as well as with real applications. Our experiments showed (1) loading times of lost input data in the range of milliseconds for up to 24,576 processors and (2) a substantial speedup in recovery times for the fault-tolerant version of RAxML-NG. In Figure 17, we provide a schematic representation of the ReStore data handling, replication, and recovery process.

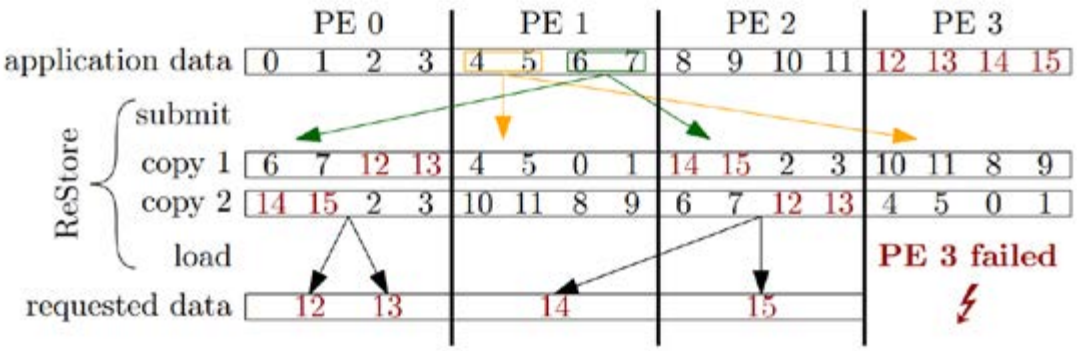


Figure 17: Example of the data submission and load operations as well as the data distribution of the redundant data copies with a random permutation for 4 PEs (processors), 16 data blocks, and 2 copies per datum. The first row shows the data submitted by the application (e.g., RAxML-NG). As an example, the orange and green arrows show the data that ReStore sends from PE 1 to the target PEs, which hold copies of the received data. When PE 3 fails, the application requests the data shown in the last row (dark red everywhere), which is provided by ReStore, as shown by the black arrows.

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:

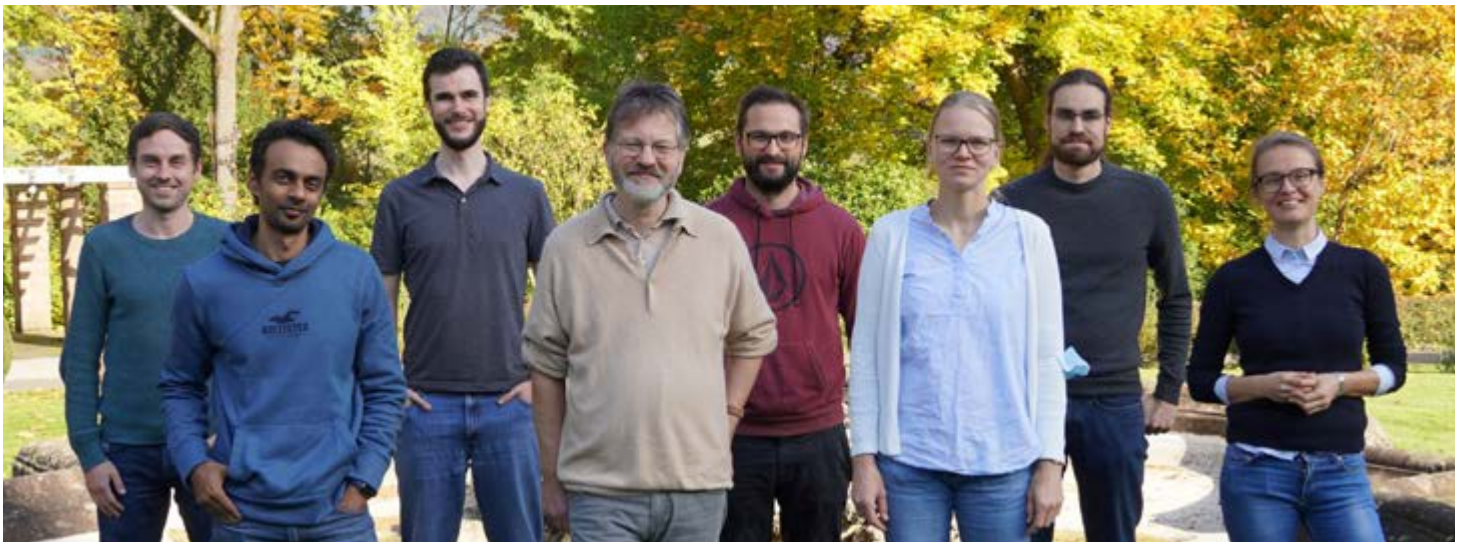
- Neue parallele Rechnerarchitekturen
- Diskrete Algorithmen auf Bäumen
- Analyse von Ancient DNA-Daten
- Methoden der Populationsgenetik.

Im Folgenden beschreiben wir unsere Forschungsaktivitäten. Unsere Forschung setzt an der Schnittstelle zwischen Informatik, Biologie und Bioinformatik an. Unser Ziel ist es, Evolutionsbiolog*innen neue Methoden, Algorithmen, Computerarchitekturen und frei zugängliche Werkzeuge für die Analyse molekularer Daten zur Verfügung zu stellen. Unser grundlegendes Ziel ist es, Forschung zu unterstützen. Die Evolutionsbiologie versucht die evolutionären Zusammenhänge zwischen Spezies sowie die Eigenschaften von Populationen innerhalb einer Spezies zu berechnen. In der modernen Biologie ist die Evolution eine weithin akzeptierte Tatsache und kann heute anhand von DNA analysiert, beobachtet und verfolgt werden.

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

2 Research

2.4 Computational Statistics (CST)



Group leader

Prof. Dr. Tilmann Gneiting

Team

Prof. Dr. Sándor Baran (visiting scientist, University of Debrecen, Hungary; July 2022)
Dr. Johannes Bracher (visiting scientist, Karlsruhe Institute of Technology, Germany)
Dr. Jonas Brehmer (until May 2022)
Jun. Prof. Dr. Timo Dimitriadis (visiting scientist, Heidelberg University, Germany)
Sebastian Gottheil (student; since December 2022)
Dr. Alexander I. Jordan (staff scientist)
Kristof Kraus (student; March–April 2022 & since October 2022)
Dr. Sebastian Lerch (visiting scientist, Karlsruhe Institute of Technology, Germany)
Marius Puke (visiting scientist, University of Hohenheim; since July 2022)
Dr. Ghulam Abdul Qadir
Johannes Resin
Prof. Dr. Melanie Schienle (visiting scientist, Karlsruhe Institute of Technology, Germany)
Evgeni Ulanov (student; since February 2022)
Eva-Maria Walz (visiting scientist, Karlsruhe Institute of Technology, Germany)
Daniel Wolffram
Prof. Dr. Johanna Ziegel (visiting scientist, 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 paradigm shift 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, creating the national COVID-19 Forecast and Nowcast Hubs, and contributing to similar efforts worldwide while placing methodological emphasis on generating and evaluating epidemiological ensemble forecasts.

Within the HITS Lab project "Emulation in simulation," we have joined forces with the MBM and PSO groups in an effort to develop surrogate model tools for astro- and biophysical applications.

General news

As in the previous year, 2022 continued to be affected by the pandemic. Hybrid group meetings became the new norm, and we intend to maintain them into the foreseeable future. Nevertheless, we were thrilled to take part in our first CST group excursion since the beginning of the pandemic. On a summer hike across the Königstuhl mountain to the town of Neckargemünd, everyone was eager to meet again as part of a larger group and to spend a hot day under the protective canopy of the Odenwald. In July, Sándor Baran (University of Debrecen, Hungary) visited us, and we had a group evaluation on 4 and 5 July, during which talks and poster presentations reported on the scientific output of our group over the past few years. On 20 July, we held a hybrid workshop on the HITS premises on the topic of post-processing in weather prediction (for more information, see Section 5.1.5). A further scientific highlight was a visit by Caroline Uhler (MIT, United States) in September, leading to many inspiring discussions.

In this year's scientific report, we describe our methodological work on the popular tool of the receiver operating characteristic (ROC) curve. When tools become exceptionally persuasive, there is often a tendency to apply them in cases for which they were not designed. One such example concerns ROC curves, which tend to be applied for evaluating the overall predictive performance of forecasts, although these curves were originally designed to measure only discrimination ability. In 2022, two projects on the characterization and generalization of ROC curves reached the stage of publication, and we report on some of our findings below.

Introduction to receiver operating characteristic (ROC) curves

Through all realms of science and society, assessing the predictive ability of scores or features for future binary outcomes is of critical importance. To give but a few examples, biomarkers are used to diagnose disease occurrence, weather forecasts serve to anticipate extreme precipitation events, judges need to assess recidivism in convicts, banks use customers' information to grant or deny credit, and email messages are identified as being either spam or legitimate. In these and a myriad of other, similar settings, ROC curves are key tools that are used to evaluate the predictive potential for binary outcomes.

For illustration, Figure 18 displays the initial levels of two biomedical markers – that is, serum albumin and serum bilirubin – in a Mayo Clinic trial on primary biliary cirrhosis, which is a chronic, fatal disease of the liver. Traditional ROC analysis mandates the outcome be a binary event, which we take here as survival beyond four years. Assuming that higher marker values are more indicative of survival, we can take any threshold value to predict survival if the marker exceeds this threshold; otherwise, the value is taken to predict non-survival. This type of binary classifier yields true positives, false positives (i.e., erroneous predictions of survival), true negatives, and false negatives (i.e., erroneous predictions of non-survival).

The ROC curve is the piecewise linear curve that results from plotting the true-positive rate versus the false-positive rate as the threshold for the classifier moves through all possible values. As combinations of high true-positive rate and low false-positive rate are desirable, a curve close to the top-left corner indicates good discrimination ability. Consequently, the Area Under the (ROC) Curve (AUC) has been popular as a numerical measure of classifier performance.

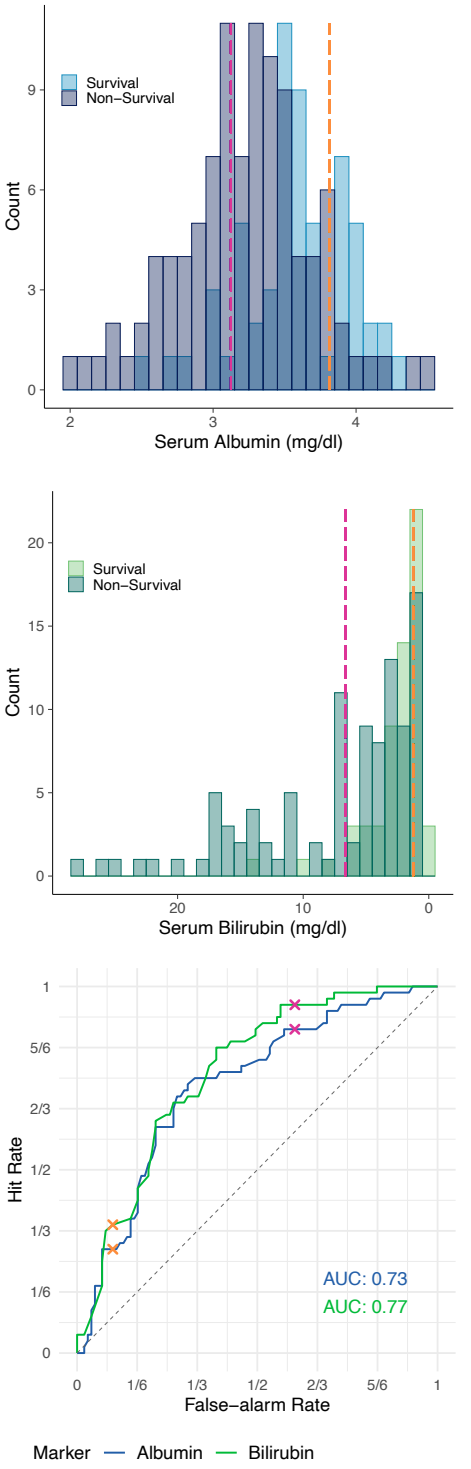


Figure 18: Traditional ROC curves for two biomedical markers – serum albumin and serum bilirubin – as predictors of patient survival beyond a threshold value of 1,462 days (i.e., four years) in a Mayo Clinic trial. Two histograms show the marker level counts for the survival and non-survival groups. The stronger shading results from overlap. For bilirubin, we reversed the orientation, as is customary in the biomedical literature.The third panel shows ROC curves and AUC values. The crosses correspond to binary classifiers at the feature thresholds indicated in the histograms.

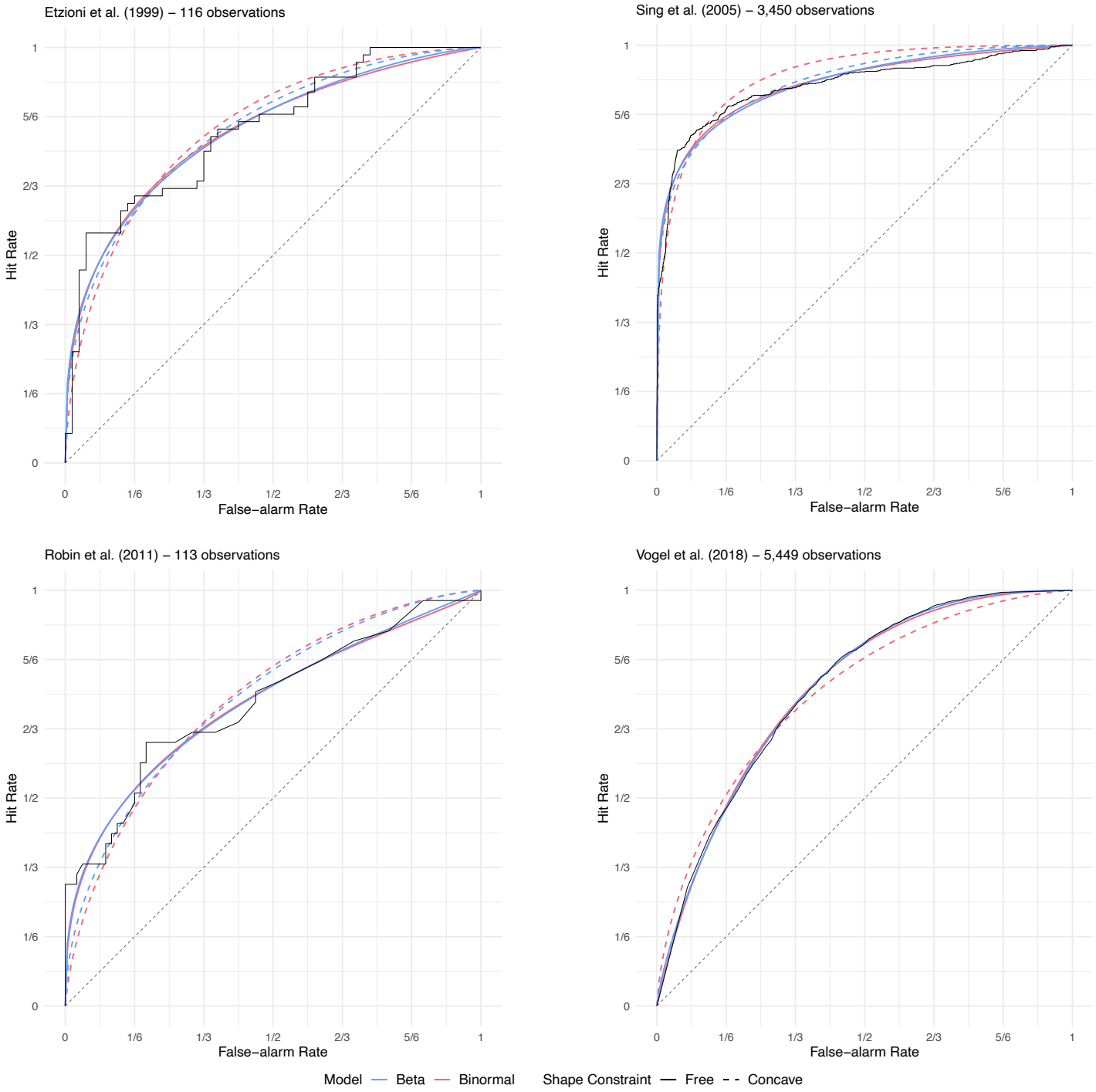


Figure 19: Empirical (black), fitted binormal (red), and fitted beta (blue) ROC curves in the unrestricted (solid) and concave (dashed) case for the datasets from the extant literature. For all datasets, we see a distinguishable deviation of the dashed red curve from the solid red curve, which indicates the lack of flexibility of the concave binormal specification.

Terminologies abound and differ markedly between communities. For example, the true-positive rate has also been referred to as the hit rate, sensitivity, or recall. Moreover, the false-positive rate is also known as the false-alarm rate or fall-out and equals one minus the true-negative rate, specificity, or selectivity.

A curve-fitting approach to ROC curves

In [Gneiting and Vogel, 2022], we demonstrated an equivalence between ROC curves and cumulative distribution functions (CDFs), introduced the flexible yet parsimonious two-parameter beta model for ROC curves, and discussed estimation and testing in this curve-fitting context.

Concavity plays a critical role in the interpretation and modeling of ROC curves. While the significance of this shape constraint is well known, the literature has long lacked a rigorous treatment that applies in general settings.

We note the equivalence of the following three conditions: (a) The ROC curve is concave, (b) the likelihood ratio is non-decreasing, and (c) the conditional event probability is non-decreasing as a function of the classifier level. In the context of Figure 18, the likelihood ratio is a positive value that depends on the classifier level, with a ratio greater than one meaning that the level is more likely to arise (and also indicating how much more likely to arise it is) in the survival group over the non-survival group. For ratios smaller than one, the opposite holds. The conditional event probability simply corresponds to the probability of survival given a particular classifier level. If the level of the classifier has a monotone association with survival outcome, as the biomedical literature suggests, then the likelihood ratio and the conditional event probability should be non-decreasing, and only the mathematical models that produce concave ROC curves would thus be reasonable.

The so-called binormal model is by far the most frequently used parametric model in the scientific literature. In the context of Figure 18, the model assumes that for both outcomes (i.e., survival and non-survival), the conditional distribution of the markers is Gaussian. As a result, the ROC curve can only be concave if the two conditional variances are equal, which is hardly ever the case in practice. In order to avoid these issues, we proposed a curve-fitting approach to the statistical modeling of ROC curves based on the two-parameter family of the CDFs of beta distributions. In this beta family, the condition for concavity is much less stringent than in the binormal family.

In order to illustrate the difference in flexibility between the binormal and beta families, Figure 19 displays binormal and beta ROC curves that have been fitted to empirical ROC curves both in the unrestricted case and under the constraint of concavity. In the unrestricted case, the binormal and beta fits are nearly indistinguishable visually. The fitted binormal ROC curves fail to be concave, and they change markedly when concavity is enforced. For the beta ROC curves, the differences between restricted and unrestricted fits are less pronounced, and in the example at bottom right, the unrestricted fit is concave. For this dataset, our newly developed goodness-of-fit test rejects both the unrestricted and the concave binormal models but does not reject the beta model. Thus, the use of the more flexible beta family is of great relevance and import. Generally, in the constrained case, the improvement in the fit under the more flexible beta model as compared with the classical binormal model is substantial.

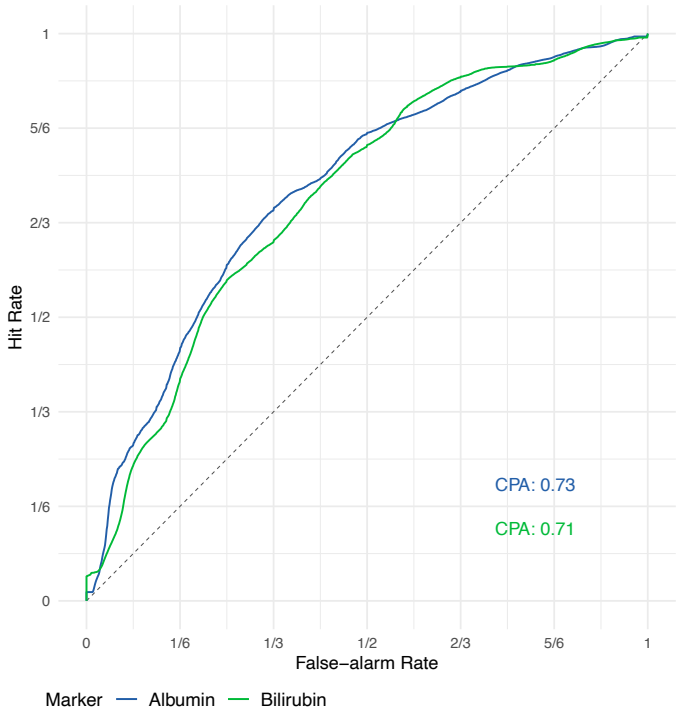


Figure 20: UROC curves and CPA for two biomedical markers – serum albumin and serum bilirubin – as predictors of patient survival (in days) in a Mayo Clinic trial. While UROC curves show a static average picture, ROC movies are animated and show the traditional ROC curves for binary events that correspond to patient survival beyond successively higher thresholds. For ROC movies, see the arXiv version of the paper at https://arxiv.org/abs/1912.01956.

ROC movies, universal ROC curves, and the coefficient of predictive ability

Despite its popularity, ROC analysis has been forced to deal with a fundamental shortcoming: namely its restriction to binary outcomes. Real-valued outcomes are ubiquitous in scientific practice, and investigators have had to artificially make these outcomes binary if the tools of ROC analysis are to be applied. As a result, researchers have long been seeking generalizations of ROC analysis that apply to any type of ordinal or real-valued outcomes in natural ways. Still, decades of scientific endeavors notwithstanding, a fully satisfactory generalization has proven elusive.

In [Gneiting and Walz, 2022], we proposed a powerful generalization of ROC analysis that overcomes extant shortcomings by introducing data science tools in the form of the ROC movie, the universal ROC (UROC) curve, and an associated, rank-based coefficient of (potential) predictive ability (CPA), all of which are tools that apply to any linearly ordered outcome, including to binary, ordinal, mixed discrete-continuous, and continuous variables.

The ROC movie comprises the sequence of the traditional, static ROC curves because the linearly ordered outcome is converted to a binary variable at successively higher thresholds (e.g., survival beyond successively longer periods of time). The UROC curve is a weighted

average of the individual ROC curves that constitute the ROC movie, where the weights depend on the class configuration – as induced by the unique values of the outcome – in judiciously predicated, well-defined ways. CPA is a weighted average of the individual AUC values in the same way that the UROC curve is a weighted average of the individual ROC curves that constitute the ROC movie. Hence, CPA equals the area under the UROC curve.

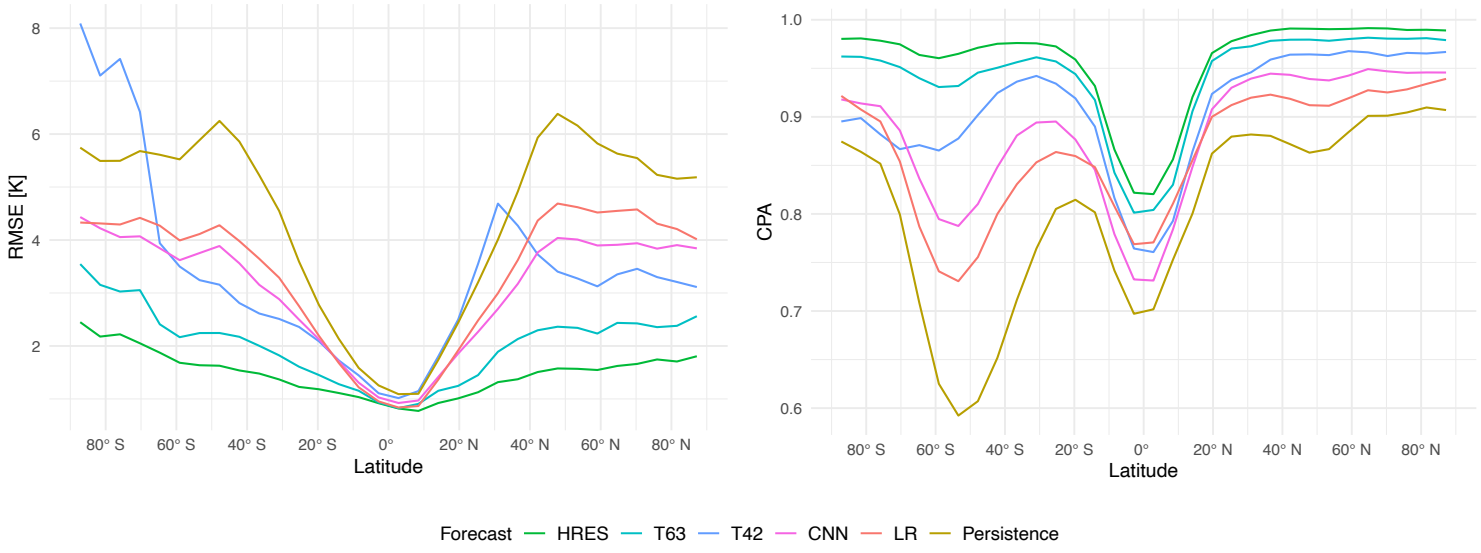
This set of generalized tools reduces to the standard ROC curve and AUC when applied to binary outcomes. Moreover, key properties and relations from conventional ROC theory extend to ROC movies, UROC curves, and CPA in meaningful ways and result in a coherent toolbox that properly extends the standard ROC concept. Equipped with this new set of tools, we no longer need to transform survival time into a specific dichotomous outcome. Figure 20 (previous page) illustrates UROC curves and CPA for the survival dataset.

The AUC value – and thereby also CPA – inherits a particular property of the ROC curve: When the classifier values are transformed by a strictly increasing function, the ROC curve remains the same. Moving beyond biomedical markers and toward actual forecasts, the values of the predictor become more relevant since systematic over- or underpredictions are undesirable. These shortcomings cannot be detected using AUC or CPA, and we additionally require proper scoring rules or consistent scoring functions in order to make a complete assessment. For example, a consistent scoring function for the mean functional encourages an honest assessment of the expected outcome. One such function is the squared difference of prediction and outcome, and the root-mean-squared-error (RMSE) is a corresponding summary measure of predictive performance.

We illustrate how these measures complement one another by using the WeatherBench dataset [Rasp et al. 2020, "WeatherBench: A benchmark dataset for data-driven weather forecasting," Journal of Advances in Modeling Earth Systems, 12, e2020MS002203], which contains predictions that are made three days in advance for temperature at the 850-hPa pressure level, which exists at an altitude of around 1.5 km. The predictive performance is quantified over a test period (2017–2018) in terms of RMSE (in degrees Kelvin) and CPA (dimensionless). We compare six types of forecasts, which include three data-driven forecasts: (1) the simplistic persistence forecast, which assumes no change in temperature from day to day, (2) the statistical approach of linear regression (LR), and (3) the machine learning approach of a convolutional neural network (CNN). The three other competitors are numerical weather prediction models based on physics and that are represented on a grid across the globe. The resolution of the grid decreases from the highest (HRES) version to successively coarser versions (T63 and T42) of the numerical model.

The panels in Figure 21 display the performance measures as functions of latitude bands from the South Pole at 90° S to the equator at 0° and the North Pole at 90° N. The measures are initially computed grid-cell-by-grid-cell and then averaged across the cells in a latitude band, which is compatible with the latitude-based weighting that is employed in WeatherBench. Note that

Figure 21: Predictive ability of forecasts of 850-hPa temperature made three days in advance in 2017 and 2018 in the WeatherBench dataset. Performance is measured in RMSE (left) and CPA (right) and is shown against and averaged by latitude. Three numerical weather prediction models (i.e., high resolution (HRES), T63, and T42) at decreasing grid resolutions are compared with three data-driven models (i.e., convolutional neural network (CNN), linear regression (LR), and persistence). Note that RMSE is negatively oriented (the smaller, the better), whereas CPA is positively oriented.



RMSE is negatively oriented (i.e., the smaller, the better), whereas CPA is positively oriented (i.e., the closer to the ideal value of 1, the better).

As we can see, the difficulty of producing accurate forecasts is geographically dependent. While the results in terms of RMSE and CPA tend to agree in the forecast ranking over the latitude bands (i.e., HRES performs best, and persistence performs worst), these results strongly disagree on forecast difficulty for certain latitudes. This disagreement is most prominent in the equatorial region. Here, all models perform their best in terms of RMSE, whereas all (except for "persistence" and LR) perform their worst in terms of CPA. This finding can be explained by the small changes in temperature from day to day, which makes it easier to achieve small RMSE values. However, with a difficult-to-predict direction of change, reaching the same CPA values as in other latitudes proves impossible.

Of similar interest are the mid-latitude storm-track regions around 50° S and 50° N. Here, day-to-day changes are large and difficult to predict, which particularly affects the data-driven models. Peaks can be seen for RMSE, whereas CPA displays drop, with a striking CPA reduction in the Furious Fifties of the Southern Hemisphere. This area is almost entirely oceanic and houses mostly mobile low-pressure systems, whose dynamic behavior seems to be more easily captured by physics-based numerical weather prediction models.

In conclusion, RMSE and CPA bring orthogonal facets of predictive performance to researchers' attention, and we encourage the joint use of the two measures in forecast evaluation. Similar recommendations apply in many practical settings in which predictions of a real-valued outcome are evaluated. In the special case of probabilistic classifiers for binary outcomes, this recommendation corresponds to reporting both the Brier mean-squared-error measure and AUC.

Die **Computational Statistics Gruppe** am HITS besteht seit November 2013, als Tilmann Gneiting seine Tätigkeit als Gruppenleiter sowie Professor für Computational Statistics am Karlsruher Institut für Technologie (KIT) aufnahm. Der Schwerpunkt der Forschung der Gruppe liegt in der Theorie und Praxis der Vorhersage.

Im Angesicht unvermeidbarer Unsicherheiten sollten Vorhersagen 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. Im HITS Lab Projekt "Emulation in Simulation" entwickeln wir gemeinsam mit den MBM und PSO Gruppen Surrogatmodelle für astro- and biophysikalische Anwendungen.

2 Research

2.5 Data Mining and Uncertainty Quantification (DMQ)

Group leader

Prof. Dr. Vincent Heuveline

Team

Aksel Alpay (visiting scientist; Heidelberg University)
Marcus Buchwald (visiting scientist; Heidelberg University)
Ayse Erozan (visiting scientist; Heidelberg University)
Dr. Philipp Gerstner (until June 2022)
Saskia Haupt (visiting scientist; Heidelberg University)
Alejandra Jayme (visiting scientist; Heidelberg University)
Dr. Philipp Lösel (visiting scientist; Heidelberg University)
Stefan Machmeier (visiting scientist; Heidelberg University)
Jacob Jonas Relle (student)
Jonas Roller (until July 2022, since August: visiting scientist; Heidelberg University)
Valentin Schmid (visiting scientist; Heidelberg University)
Elaine Zaunseder (visiting scientist; Heidelberg University)
Alexander Zeilmann (since July 2022)
Yaroslav Zharov (visiting scientist; Heidelberg University)

The Data Mining and Uncertainty Quantification (DMQ) group – headed by Vincent Heuveline – began its research in May 2013. The group works in close collaboration with the Engineering Mathematics and Computing Lab (EMCL) at the Interdisciplinary Center for Scientific Computing (IWR) at Heidelberg University, which is also headed by Vincent Heuveline.
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 2022, the DMQ group focused on research activities in the areas of uncertainty quantification, machine learning, and numerical simulation for biomedical applications.

Inferring counterfactuals from medical data using causal deep learning

Estimating counterfactuals and investigating causal relationships remains one of the most profound challenges for research, especially in the field medical data analysis. Questions such as "What would have been the outcome for the patient if Drug B had been administered instead of Drug A?" and "How would the clinical phenotype of a disease change if the patient had a different gene variant?" are complex to answer as they require a deep understanding of the underlying mechanisms. Inferring causal dependencies from data is a challenging task because apparent causal relationships can be misleading due to potential confounders. Regarding the first question above, traditional causal inference methods such as randomized control trials are considered the gold standard for treatment effect estimation, but they are often limited by practical or ethical considerations.

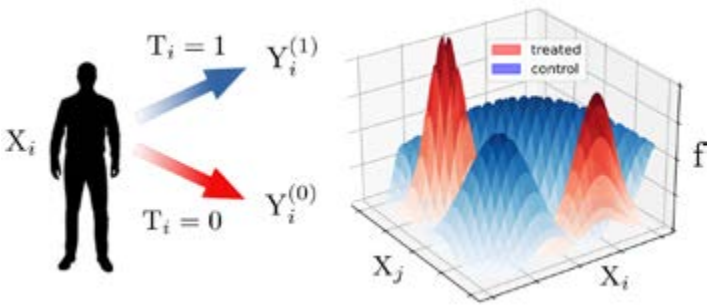


Figure 22: For ITE, both potential outcomes must be estimated, although only one is observed for observational data. Additionally, the covariate distributions differ.

Since 2022, we have been working in collaboration with the Mannheim Institute for Intelligent Systems in Medicine (MIISM) both on developing deep learning methods for individual treatment effect (ITE) estimation, such as graph neural networks when prior information about the data structure is available, and on improving existing architectures by using automated hyperparameter optimization. Beyond the challenge of potential confounding factors outside the observed features, these methods must also address the distributional covariate shift between treatment groups that often arises due to a treatment selection bias in observational data (see Figure 22).

Furthermore, our research has focused on counterfactual inferences in medical images using generative deep learning methods in structural causal models. The aim is to predict the effect of intervening on certain patient characteristics, such as on MRI images of the patient. In particular, the potential of de-noising diffusion probabilistic models – which have recently shown promising performance results in conditional image generation – is being further investigated for this counterfactual reasoning. One major challenge lies in the fact that validation in terms of performance as ground truth information about the non-factual scenario of interest is rarely available. Beyond relying on medical expertise about the medically plausible results of the model, our methods are currently being analyzed regarding the time-point prediction of longitudinal data as well as regarding accuracy when intervening in features that are directly extractable from the image, such as brain and ventricular volume. Future work will concentrate on maximizing predictive accuracy as these models not only allow medical expertise to be re-evaluated and extended, but they can also be used for modeling disease progression.

Automatic segmentation of single cells

In order to understand the mechanisms of disease and the development of specific disorder treatments, it is critical to analyze three-dimensional biological cell samples. Although many cutting-edge imaging technologies exist, soft X-ray tomography (SXT) (Figure 23, next page) is a unique form of technology that can image whole intact cells under normal and pathological conditions at high throughput and spatial resolution without labeling or fixation. Image segmentation is currently a primarily manual task, which is tedious, time-consuming, and prone to human error, and it remains a bottleneck. In our project, we aim to tackle this major barrier in the SXT data analysis pipeline by implementing the semi-automatic segmentation application Biomedisa and developing a fully automatic segmentation pipeline based on deep learning.

In a collaborative initiative with the Centre for Organismal Studies (COS) at Heidelberg University, SiriusXT Limited in Dublin, and the Department of Infectious Diseases, Molecular Virology at Heidelberg University Hospital, we are currently working on exactly this aim. For three years, this research has been funded by the framework of the Excellence Strategy of the Federal (BMBF) and the State Governments of Germany's "Flagship Initiative Engineering Molecular Systems" and the European Union Research and Innovation Act, project "CoCID."

For this purpose, we used the generated 3D labels along with 3D tomograms to train the U-Net framework for the automatic segmentation of cell cytoplasm. We used the U-Net implementation within Biomedisa. For training, we used a large number of SXT tomograms consisting of human T lymphocyte cells. The trained network is depicted as ACSeg-Net for the adaptive cytoplasm segmentation model. Because the cells are so diverse in size and morphology, it is typically difficult to apply a segmentation network to other cells if it is trained on one specific type of cell. In order to determine whether ACSeg-Net would have higher accuracy by generalizing training datasets, we tested ACSeg-Net

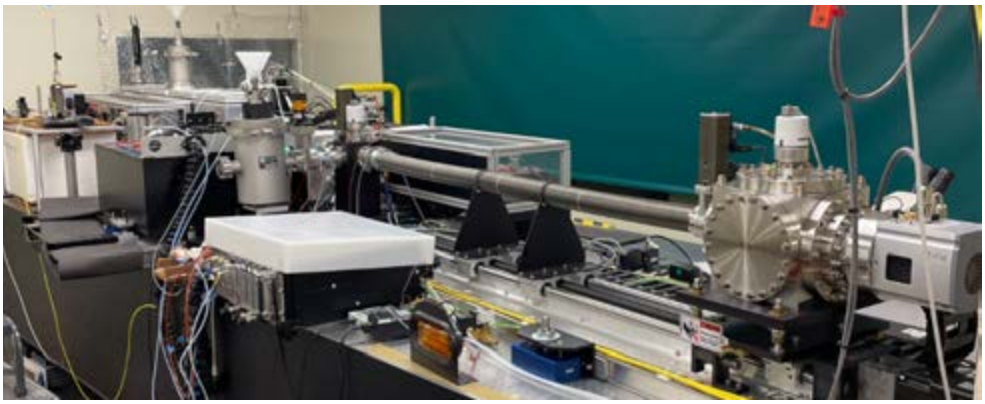


Figure 23: Soft X-ray tomography XM-2.

on SXT datasets that included divergent cell types. The algorithm has the potential to be expanded for semi-automatic segmentation of all organelles in any cell type.
We expect to provide high segmentation accuracies and to significantly accelerate the segmentation of large and complex data of single-cell data by creating an automatic segmentation platform for SXT data. We expect the morphological changes within individual cells in pathologic conditions – such as viral infection – to reveal new principles of virus-induced alteration, and we additionally hope to provide important information via quantitative automatic analysis for the development of antiviral therapies.

Predicting ventricular heart pressure from cardiac MRIs with deep neural networks

Heart failure is the leading cause of death in nearly all industrial countries. Heart failure occurs when the heart muscle is no longer able to pump enough blood to supply all organs with energy. Imaging techniques such as magnetic resonance imaging and echocardiography are often sufficient to diagnose heart failure. However, to diagnose heart failure with preserved ejection fraction, a characterization of the diastolic heart function is usually required. The gold-standard procedure for accessing the diastolic function is cardiac catheterization. In this procedure, a catheter is inserted through the groin into the ventricle in order to measure the ventricular pressure. Although cardiac catheterization is widely used, it still carries an interventional risk.
In order to nullify interventional complications, we developed an alternative non-invasive method of accessing the ventricular pressure. We proposed artificial intelligence agents for predicting the ventricular pressure from cardiac MRIs (see Figure 24). The main component of this approach is a deep convolutional neural network.
The composition of suitable training data for the AI was challenging. First, potential study patients required cardiac catheterization and MRI in a short time frame, and furthermore, any patients who showed evidence that pressure was not comparable at these two time points had to be excluded.
The final model outperformed human MRI experts and current state-of-the-art methods from Doppler echocardiography when it came to detecting elevated ventricular pressure, and this finding may potentially impact future diagnosis and treatment strategies. The patient collective for the model development stemmed from Heidelberg University Hospital. Furthermore, we validated the results in a multicenter study across Germany.
In order to further understand how the AI model is capable of ventricular pressure prediction from MRIs, we utilized attention mapping to visualize the most relevant regions in the MRI. The AI learned to focus on the mitral valve area and on blood flow, which is consistent with established methods from echocardiography.
In summary, neural networks were found to be capable of predicting increased filling pressure from cardiac MRIs, which could impact future diagnostic and treatment strategies for patients.

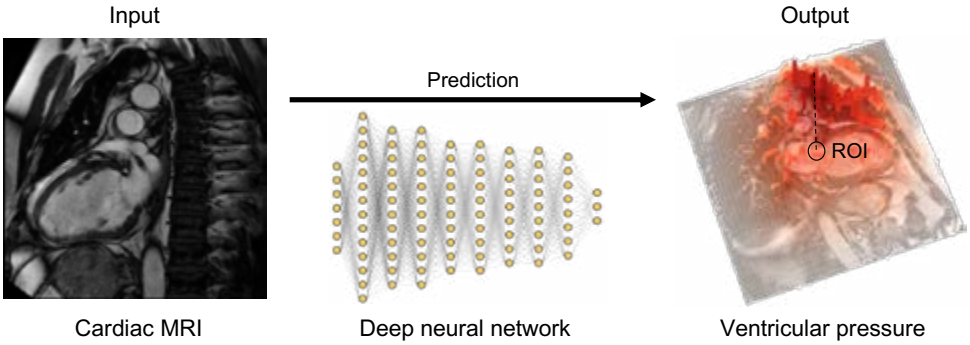


Figure 24: Predicting heart chamber pressure from cardiac MRIs that include the AI region of interest.

Toward simulating early colorectal cancer development

Colorectal cancer is the third most common cancer worldwide and arises from the abnormal growth of cells in the colon or large intestine. Understanding the development of colorectal cancer is crucial when it comes to improving early diagnosis and treatment. In recent years, mathematical modeling has emerged as a powerful tool for investigating the biological processes involved in the development of colorectal cancer (see Figure 25). As part of our long-term

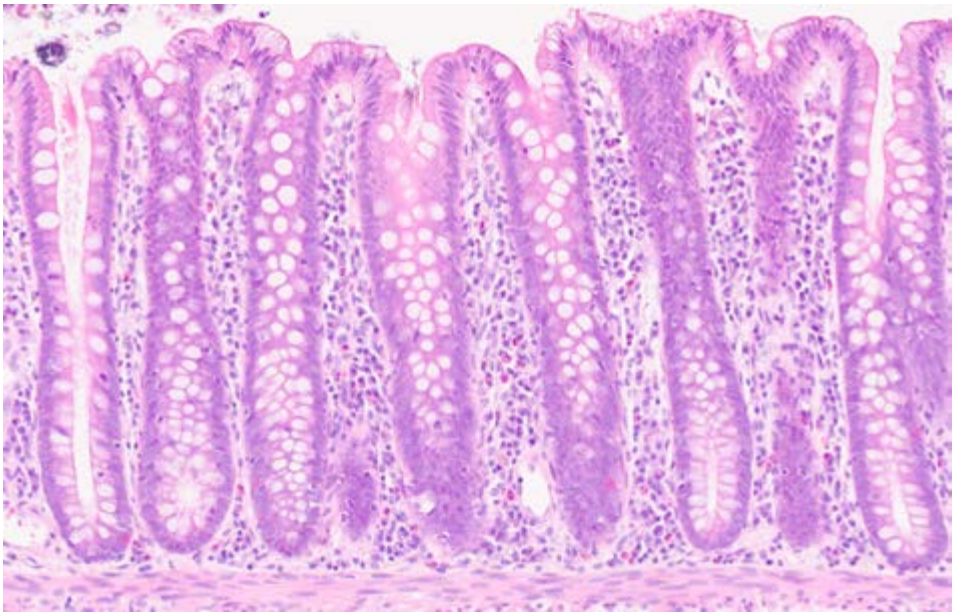


Figure 25: A microscopy image of colonic crypts, which was acquired by means of neutron imaging. From left to right, the image depicts reconstruction with filtered back projection, the same reconstruction after de-noising with the proposed adaptation of the Noise2Noise method, and the TV-TGV iterative reconstruction used in the original paper. The deep learning de-noising method provides more spatial details and works significantly faster.

research project "Mathematics in Oncology – Towards optimal prevention and treatment in patients with inherited cancer syndrome," we are currently developing a three-dimensional simulation for colorectal cancer development that uses partial differential equations (PDEs).

We consider processes that include cell proliferation and differentiation, apoptosis (programmed cell death), and the release of growth factors. To describe these processes, we propose using a set of three-dimensional nonlinear PDEs, which are a type of mathematical equation that can describe how a quantity changes over space and time.

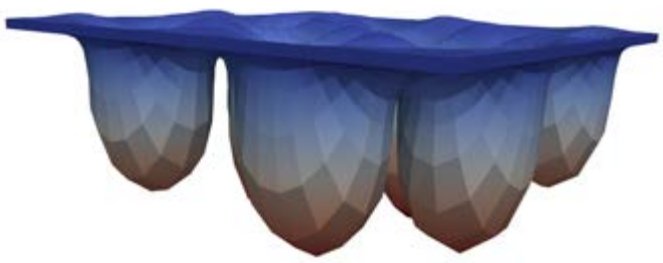


Figure 26: Three-dimensional numerical simulation of healthy crypts.

The finite element method (FEM) is a numerical technique that is well-suited for solving PDEs by representing the solution of mathematical equations as a combination of simple functions and by approximating the actual solution (see Figure 26). In our case, the flexible and efficient approach of the FEM can be used to handle the spatial geometries and complex nonlinear mechanisms that occur in the cancer-describing equations.

For healthy tissue, a set of parameters characterizes cell signaling and controlled cell growth. In cancer formation, these parameters change, which leads to disturbed cell signaling or uncontrolled growth. By keeping track of these parameter changes over time, we aim to study the impact of different factors on the development of colorectal cancer. In the future, we will investigate how well parameters can be adapted to specific characteristics of individual patients. In conclusion, mathematical modeling that uses partial differential equations and the finite element method offers a promising approach for studying the development of early colorectal cancer.

Employing self-supervised deep learning for synchrotron-based CT analysis

Synchrotron-based computed tomography (CT) is a non-invasive imaging technique that uses high-energy X-rays generated by a synchrotron light source to produce 3D images of objects or materials. The high energy and brightness of the synchrotron light source enable high-resolution high-throughput imaging, which makes synchrotron-based CT an important tool in fields such as material science, biology, and engineering.

As the size of the dataset grows with increasing imaging resolution and the number of samples that can be processed, the classical manual or semi-automated analysis methods begin to create a project bottleneck. That's where deep learning is typically employed. However, even for a classical deep learning approach, an extensively labelled dataset is necessary in order to train a model. To mitigate this situation, self-supervised training and self-training are becoming increasingly potent training techniques that can help minimize the time and resources required for dataset annotation and can also improve model performance without additional markup.

In self-supervised training, the algorithm is trained on the dataset without explicit human annotations by using techniques such as contrastive learning or generative models. This approach allows the model to learn features and representations that are relevant to the specific data, thereby reducing the amount of human annotation required.

Self-training, on the other hand, involves

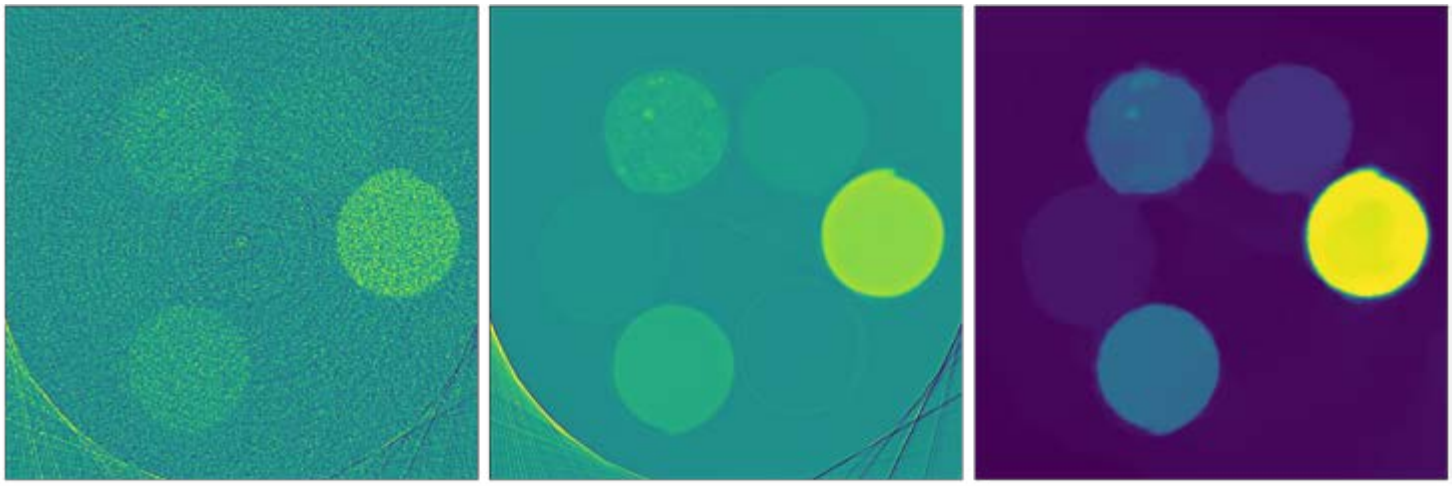


Figure 27: Tomographic slice of 5 different metal powder containers of healthy crypts.

training a model on a small annotated dataset and then using the model's predictions on the rest of the data as pseudo-labels to train a new model. The second model demonstrates improved performance without additional human annotation.

The use of self-supervised training and self-training can be particularly beneficial for synchrotron facilities that work with large datasets comprised of hundreds of samples. In our projects, the use of self-supervised training and self-training has shown promising results for computed tomography data analysis. One significant improvement has been observed in the quality of segmentation through the use of self-supervised pre-training and self-training. Another practical application of self-supervised training is the de-noising of multichannel tomography and cineradiography data (see Figure 27). By using self-supervised training based on the idea that adjacent channels share most of the information, the models can effectively de-noise the data, thereby leading to cleaner images and also improving the accuracy of further analysis.

In addition to these results, we are also working in the direction of automated artefact removal and the optimization of labelling procedures. These efforts aim to further reduce the time and resources required for manual annotation, thereby further improving the efficiency and accuracy of the analysis process.

Program once for all!

Modern supercomputers draw a large portion of their compute power not from their main processors (CPUs), but rather from specialized accelerators, such as graphics processors (GPUs). But how do we program such heterogeneous architectures? This is a pressing question, especially in light of the increasing diversity of hardware: In the past, accelerated computing was largely synonymous with GPUs from NVIDIA. As a result, GPU-accelerated programs were typically written in NVIDIA's proprietary programming model CUDA. However, more recently, other accelerators – such as GPUs from other vendors like AMD or Intel – have gained traction.

Consequently, portable, vendor-independent programming models are needed. To address this need, we have been developing Open SYCL (formerly known as hipSYCL; see github.com/OpenSYCL/OpenSYCL), which is an implementation of the SYCL standard, which defines a programming model for accelerated hardware architectures based on the widely used programming language C++. The Open SYCL software consists of a compiler – that is, an engine for translating source code to executable machine code – and a runtime system that orchestrates the execution of work. Open SYCL supports offloading work to CPUs as well as Intel, NVIDIA, and AMD GPUs (see Figure 28). This is the first SYCL implementation with robust support for GPUs from NVIDIA and AMD. In our work, we have shown that the software performance typically closely matches the performance of vendor-supported programming models such as CUDA on NVIDIA GPUs. The development of Open SYCL is supported by Intel as an Intel oneAPI Center of Excellence. Our software is also used in production by complex applications, such as the molecular dynamics simulation code Gromacs, which relies on Open SYCL to target AMD GPUs.

In order to further improve portability, we are interested in the question of how binaries can be generated with minimal compile times that can then be deployed across a wide variety of heterogeneous hardware. To that end, we have implemented the very first single-pass SYCL compiler using Open SYCL. This means that by analyzing the source code one single time, Open SYCL can generate a binary that runs on CPUs as well as on all supported GPUs. The binary can then dispatch work at runtime to whatever hardware it finds on the system.

This ability to adapt at runtime to the discovered hardware is connected to a wider range of questions that we are interested in: How – and to what extent – can performance be improved by

modifying the code at runtime based on the hardware, for example, by taking into account hardware characteristics or hard-wiring program arguments that are only known at runtime? Additionally, what are the consequences of such an adaptive approach to the design of a work scheduler? These are the challenges that we are currently addressing.

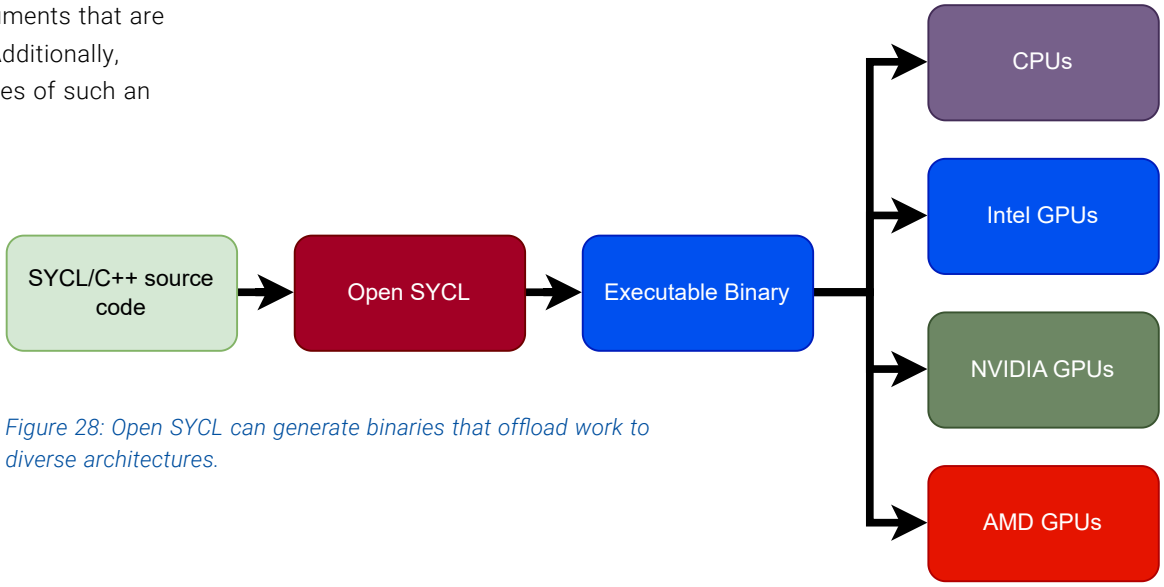


Figure 28: Open SYCL can generate binaries that offload work to diverse architectures.

Die Forschungsgruppe **Data Mining and Uncertainty Quantification (DMQ)** unter der Leitung von Vincent Heuveline 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.

2022 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 (until October 2022)

Team

Fernando Camacho Cadena (PhD student)
Dr. Nguyen-Thi Dang (until September 2022)
Dr. Valentina Disarlo (visiting scientist)
Dr. Brice Lousteau (until March 2022)
Marta Magnani (vising scientist)
Levin Maier (student; until March 2022)
Jun. Prof. Dr. Beatrice Pozzetti (visiting scientist)
Dr. Anja Randecker (visiting scientist)
Dr. Anna Schilling (visiting scientist)
Jiajun Shi (PhD student)
Jun. Prof. Dr. Peter Smillie (visiting scientist; since October 2022)
Dr. Gabriele Viaggi (visiting scientist)

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 the object unchanged. Moreover, these symmetries 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 unifies 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 based on incidence relations rather than on the measurement of distances. Klein's concept fundamentally changed our view of geometry in mathematics and theoretical physics and continues to influence both 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.

In November 2022, Anna Wienhard was appointed director of the Max Planck Institute for Mathematics in the Sciences in Leipzig. The group continues to collaborate with other HITS research groups via the HITS Lab.

Dynamics on character varieties

Riemann surfaces play an important role in different areas of mathematics and theoretical physics and are also central to the research of the GRG group. The area of higher Teichmüller theory deals with representations of the fundamental group of a surface – called S – with negative Euler characteristic (e.g., a surface without boundary and with at least 2 holes, as in Figure 29) in a Lie group, G. Spaces of such representations up to a natural notion of equivalence are called character varieties. In some special cases, components of the character variety have geometric meaning. For example, in the case when G is PSL(2,R) (i.e., the group of 2 x 2 matrices with real entries and determinant 1 up to sign), two components of the character variety – called Teichmüller space – parameterize hyperbolic structures on S. When G is PSL(3,R), components of the character variety parameterize convex projective structures on S. More generally, special components of the character variety encode a rich variety of geometric, dynamical, and algebro-geometric information. Below, we provide a glimpse into the symplectic and dynamical aspects of the theory.

Before delving into the specifics of higher Teichmüller theory, some motivation and background information on symplectic geometry and dynamics are needed. Formally, a symplectic structure on a manifold M allows for measuring 2-dimensional areas in M. Although the symplectic structure cannot be used to measure the length of curves (since curves are 1-dimensional), it gives rise to interesting flows on M. Take a "potential" on M, which is a function from M to the real numbers. Using the symplectic form, a unique flow on M called the Hamiltonian flow is defined. One property of Hamiltonian flows is that they preserve the value of the potential along flow lines. The main motivation of symplectic geometry stems from classical mechanics in the study of the motion of bodies such as planets or stars. In this case, the manifold M is the "phase space" and is given both by the position of the particle in 3-dimensional space and by the body's momentum in each direction. After endowing M with a symplectic form, the potential on M which measures the energy of the body is considered (which can be done because M also carries information about the body's momentum). The Hamiltonian flow then precisely describes the body's motion in space. As the body follows its trajectory, its energy is preserved.

With Hamiltonian flows at hand – or more generally, group actions on M – it is possible to study the dynamics of these flows. One way of doing so involves understanding how "chaotic" these flows are. A dynamical system is said to be ergodic if the only potentials on M that are preserved by the flow (or action) are the potentials that are constant on all of M. Ergodic systems are interpreted to be chaotic because orbits of the system equidistribute – that is, over long periods of time, each point in M has the same probability of being reached by the flow. An important example in this area is Boltzmann's ergodic hypothesis in statistical mechanics, which roughly states that in a mechanical system, particles will reach every point in phase space over long periods of time. In other words, the orbit will spread out over all of the phase space. On the opposite spectrum of chaotic systems, we have group actions that are proper. Formally, this means that given a small region in M, the orbit of any point in the small region will return to this region only a finite number of times. The notion of properness is therefore opposite to the notion of ergodicity, since in ergodic systems, orbits return infinitely many times to small regions.

We now return to character varieties. These spaces have a natural symplectic structure and hence admit a plethora of Hamiltonian flows and dynamical systems. In the examples of when

Figure 29: A 4-holed surface.

subsets of the character variety parameterize geometric structures on S, the flows can be interpreted as deformations of a starting geometric structure. For example, in Teichmüller space, a Hamiltonian deformation is described as follows: Take a curve alpha, as in Figure 30. Cut the surface along it, and then glue it back with a twist. The green curve will be deformed to be twisted around alpha and is therefore called a twist flow. Performing this operation along more complicated systems of curves leads to the earthquake flow (which was named because it resembles an earthquake on S in which the fault lines lie along the curve system).

This flow is a Hamiltonian flow, and its potential is given by measuring the length of the system of curves in a given hyperbolic structure. Twist flows can be generalized to higher Teichmüller spaces and give a rich source of dynamical systems on character varieties.

Another dynamical system on character varieties is given by the action of the mapping class group MCG(S) of the surface S, which is the group of diffeomorphisms of S up to isotopy. This group can be viewed as the group whose elements re-label the closed curves on S. The mapping class group naturally acts on character varieties and preserves the symplectic form. W. Goldman conjectured that a dichotomy exists regarding the action: namely that it is proper on components of the character variety that carry "geometric meaning" (e.g., Teichmüller space, or the space of convex projective structures) and is ergodic on the other components. On Teichmüller space, the action is known to be proper (i.e., not chaotic). Since the action is well-behaved, it is possible to take the quotient of Teichmüller space by the MCG(S) to obtain an orbifold with finite volume. The earthquake flow described above descends to the quotient, and as Maryam Mirzakhani proved in 2008, this flow is ergodic (i.e., chaotic). These are some of the few cases in which the action is well understood, and they represent an active area of research in higher Teichmüller theory.

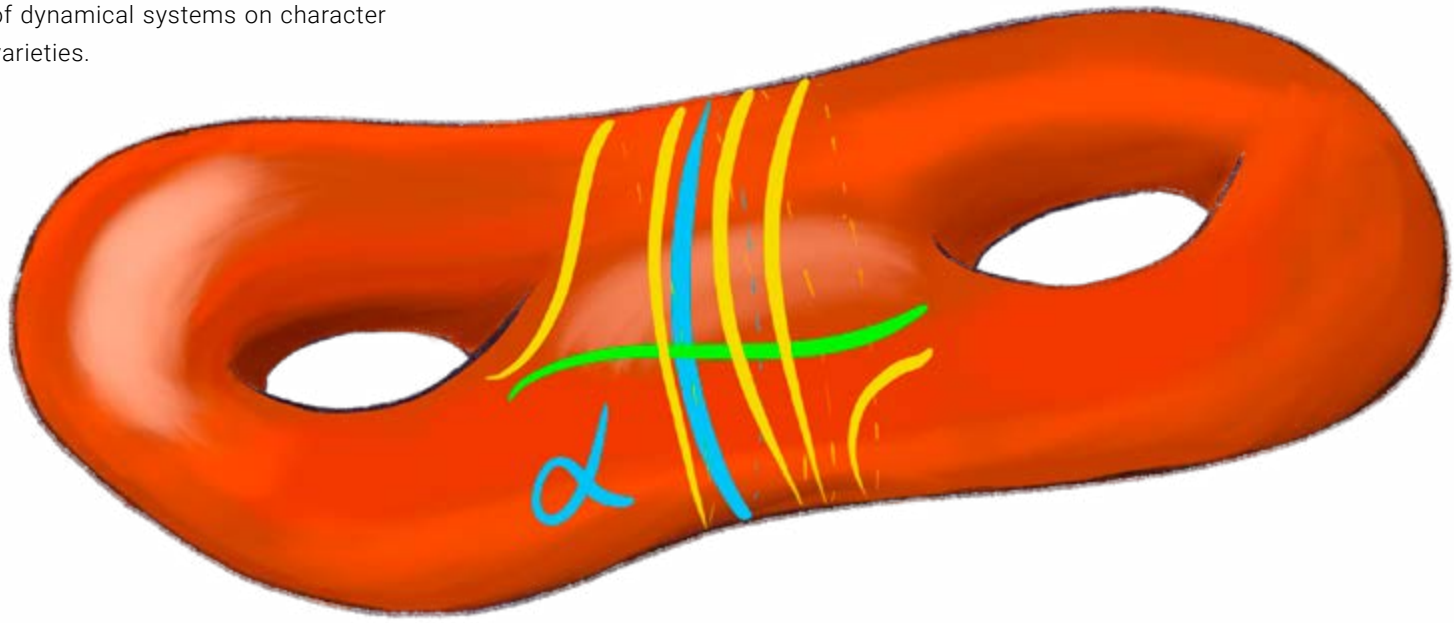


Figure 30: Twist flow around the curve alpha.

Die Arbeitsgruppe **Gruppen und Geometrie** 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 genannten, 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 Räume, 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.

Im November 2022 wechselte Anna Wienhard als Direktorin ans Max-Planck-Institut für Mathematik in den Naturwissenschaften in Leipzig. Ihre Gruppe arbeitet weiter mit anderen HITS-Forschungsgruppen in Projekten innerhalb des HITS Lab zusammen.

2 Research

2.7 Machine Learning and Artificial Intelligence (MLI)

Group leader
Jun.-Prof. Dr. Jan Stühmer

The MLI group – which was established at HITS in September 2022 – works on novel algorithms and models for data-efficient learning, interpretability, and geometric deep learning.
The data-efficient learning methods that are explored in the group enable us to take an existing model (e.g., a model that was trained on a big standard dataset) and adapt it to a novel application. The dataset of the new domain that is used for fine-tuning the model can then be much smaller than it would have to be without pre-training [Hu et al., 2022]. This smaller size reduces the time and resources needed for collecting training data and enables the use of machine learning approaches in application areas that have thus far not been able to benefit from this technology.
Another research focus is on learning interpretable representations with the goal of making models interpretable and therefore also enabling a better understanding of the models' underlying principles. Some of these methods enable us to reconstruct underlying latent factors and even causal relationships from observed data with exciting applications, especially in the natural sciences. To do so, the group applies and extends methods from the field of variational inference with statistical methods for independent component analysis.
Since joining HITS in September, several collaborations with other research groups have been initiated, including the Molecular Biomechanics (MBM) and the Computational Carbon Chemistry (CCC) groups. Within these collaborations, the MLI group contributes its expertise in geometric deep learning – a novel approach for understanding deep neural networks with mathematical tools from geometry and group theory. The resulting methods can be used for de novo protein design as well as for predicting molecular properties.

Kaum ein Schlagwort ist in den letzten Jahren so sehr Bestandteil des menschlichen Alltags geworden wie „Künstliche Intelligenz“ – von der Sprachassistenz über autonomes Fahren bis hin zu generativen Modellen, die ausgehend von einer Textbeschreibung beeindruckende Bilder unterschiedlicher künstlerischer Stilrichtungen erstellen können. Die Wissenschaft indes beschäftigt sich schon seit langem mit den Methoden, die hinter dem Schlagwort stecken. Daran arbeitet die **Machine Learning and Artificial Intelligence (MLI)** Gruppe, die im September 2022 am HITS gegründet wurde.

Die Gruppe beschäftigt sich mit der Entwicklung von neuartigen Algorithmen und Verfahren des maschinellen Lernens. Besondere Schwerpunkte sind hierbei dateneffiziente Lernverfahren, Interpretierbarkeit und Geometrisches Deep Learning. Die entwickelten Methoden des dateneffizienten Lernens erlauben es, ein an einem großen Datensatz vortrainiertes Modell an eine neuartige Anwendung anzupassen. Der Datensatz, der für dieses sogenannte „Fine-Tuning“ verwendet wird, kann dabei deutlich kleiner sein, als es ohne das Vortrainieren des Modells möglich wäre [Hu et al., 2022]. Dadurch reduziert sich der Zeit- und Kostenaufwand zum Erstellen der Trainingsdaten, und maschinelle Lernverfahren können in dafür bisher unzugänglichen Anwendungsfeldern verwendet werden.

Ein weiterer Schwerpunkt liegt in interpretierbaren Repräsentationen, mit dem Ziel, die Modelle interpretierbar und somit besser verständlich zu machen, und um grundlegende Zusammenhänge in Daten zu veranschaulichen. So lassen sich mit entsprechenden Lernalgorithmen den Daten zugrunde liegende Faktoren und teilweise sogar die kausalen Zusammenhänge von beobachteten Daten ableiten. Insbesondere in den Naturwissenschaften ergeben sich dadurch interessante Anwendungen. Zur Anwendung kommen hierbei Methoden der variationellen Inferenz und statistische Verfahren der Faktoranalyse.
Am HITS wurden bereits vielfältige Forschungskollaborationen mit anderen Gruppen begründet, unter anderem mit den Forschungsgruppen Molekulare Biomechanik (MBM) und Computational Carbon Chemistry (CCC). Hier steuert die MLI-Gruppe ihre Expertise im Bereich des Geometrischen Deep Learning bei, einem neuartigen Bereich des Maschinellen Lernens, der es erlaubt, Konzepte für die Struktur der Modelle aus den mathematischen Teilgebieten der Geometrie und Gruppentheorie herzuleiten.

Image created with Dall-E 2 and the HITS image generator

2 Research

2.8 Molecular Biomechanics (MBM)

Group leader

Prof. Dr. Frauke Gräter

Team

Atanas Aleksandrov (student; January–March 2022)
Helman Amaya (visiting scientist; Universidad de los Andes, Colombia)
Dr. Camilo Aponte-Santamaria (staff scientist and acting group leader)
Saber Boushehri
Matthias Brosz
Johanna Buck (student; June–July 2022)
Jannik Buhr
Svenja de Buhr
Mikaela Farrugia (visiting scientist, University of Notre Dame, Indiana, USA, until September 2022)
Eric Hartmann
Adel Iusupov (student; January–March 2022)
Denis Christian Kiesewetter (student; September–November 2022)
Dr. Markus Kurth
Fabian Kutzki visiting scientist, KIT Karlsruhe, until July 2022)
Isabel Martin (until May 2022)
Dr. Nicholas Michelarakis (until June 2022)
Juan Orjuela (visiting scientist; Universidad de los Andes, Colombia)
Benedikt Rennekamp
Kai Riedmiller
Boris Schüp (October–December 2022)
Andrea Sassoli (since September 2022)
Leif Seute (since September 2022)
Salome Steinke (until February 2022)
Daniel Sucerquia (since July 2022)
Giulia Tonon (student; February–April 2022)
Wojtek Treyde (student; until March 2022)
Evgeni Ulanov (student; since February 2022)
Aysecan Ünal

Proteins are the working horses of living systems. In recent decades, we have learned a great deal about how proteins look, move, and work. One rather newly discovered aspect of proteins is that many show very surprising properties once mechanical force acts on them, which has important consequences for the function of these proteins in the biological cell. The aim of the Molecular Biomechanics group is to decipher the mechanical function of proteins. We use molecular dynamics simulations at different resolutions as well as experiments to address this question on multiple time and length scales.

For many years, we have contributed to the developing understanding of how the mechanical force that is present in flowing blood regulates and triggers the action of proteins that are involved in blood coagulation. In so doing, our focus has been on von Willebrand factor. In 2022, we deciphered the mode of action of two more of the many structural components of this huge multimeric protein, which is part of our blood. The so-called C6 and D'D3 domains show unexpected changes in activity when perturbed by mutations or different redox states. We were able to resolve the underlying reason for these changes using molecular dynamics simulations.

As proteins are polymers, polymers can teach us a lot about how proteins respond to forces. Within the Excellence Cluster 3DMM2O, we were able to demonstrate that polymers as robust as a special, fully conjugated class of polymer chains can be mechanically broken. In 2022, we took the first step toward describing such processes across scales using simulations.

Our activities within the ERC Consolidator grant RADICOL on mechanoradicals in collagen in 2022 included the inauguration of our small wet lab at the Center of Advanced Materials at Heidelberg University. We can now build on our results both on radical energies in proteins that have been achieved within our HITS Lab project and on bond scission mechanisms in polymers, both of which are presented further below. Stay tuned for upcoming insights into the "how's," "when's," and consequences of collagen and radicals throughout 2023 and beyond!

Structure and dynamics of the C6 von Willebrand factor domain and the force response of the D'D3 VWF assembly complex determined by experiments and simulations

Von Willebrand factor (VWF) is a giant extracellular blood protein that has a key adhesive function during primary hemostasis. When activated by the shear of flowing blood, VWF recruits platelets at sites of vascular injury in order to build a plug that prevents bleeding. VWF malfunction is linked to a variety of bleeding disorders that range from acute bleeding to thrombosis. VWF is a large multimeric protein in which each multimer is composed of several protein domains. The interaction of VWF with platelets and with the surface of the injured site – as well as the mechanism of downregulating VWF activity – occurs at the so-called A domains (A1, A2, and A3). These three domains have been extensively studied; however, we are only now beginning to gain a molecular understanding of the function of the other regions of VWF: namely the C and D domains.

By taking a multidisciplinary approach that combines molecular dynamics (MD) simulations, nuclear magnetic resonance (NMR), and functional assays, we were able to determine the structure and dynamics of one such less-studied domain: namely C6. The NMR structure revealed that this domain features a two-hinge sub-domain structural fold that is reminiscent of other VWF C domains (Figure 31, A). C6 displays significant hinge flexibility. It undergoes conformational changes between extended and bent conformations. The extended conformations were the more commonly observed

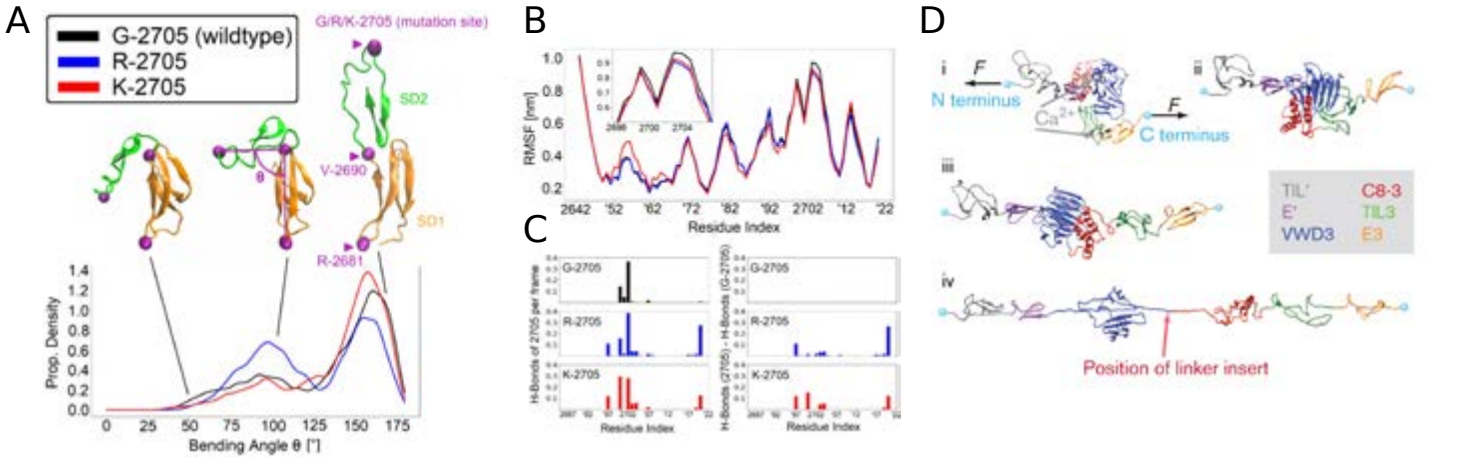


Figure 31: Structure and dynamics of the C6 von Willebrand factor (VWF) domain and force response of the D'D3 VWF assembly complex determined by experiments and simulations. (A) The structure and dynamics of the C6 domain were determined by nuclear magnetic resonance and molecular dynamics (MD) simulations. Two domains (SD1 (orange) and SD2 (green)) were observed to have remarkable hinge flexibility. Extended and bent conformations were observed, with a preference having been found for the extended conformations both for the native protein (black) and for mutants that replaced the clinically relevant G-2705 residue with positive residues (blue and red). (B–C) Simulations demonstrated local changes in the root mean square fluctuation (RMSF) near G-2705 upon mutation (B) as well as an increased formation of hydrogen bonds with nearby amino acids when this residue was replaced by arginine (R) or lysine (K) (C). (D) Force-probe MD simulations that mimicked optical tweezer experiments showed a sequential opening of the different sub-domains of the D'D3 VWF assembly complex (colored according to the legend) upon the application of a force F (Steps I–IV).

ones (Figure 31 A). The disease-related gain-of-function mutation G2705R is located in this domain. We observed that this mutation alters the structure and dynamics of C6 both globally (Figure 31 A) and locally (Figure 31 B-C) (see further information in ref. [Chen et al, 2022]).
Switching to the other side of the protein, we found the D'D3 assembly complex. This complex is an essential region of VWF because it contains the cross-linking site for multimerization and the binding site for factor VIII, which is a key protein during later stages of hemostasis.
Additionally, in a multidisciplinary effort that combined a diverse set of biophysical techniques (including optical tweezers), we demonstrated that this domain is destabilized by changes in pH, such as the changes that this protein encounters upon its release into the extracellular medium.
Our simulations also demonstrated a sequential opening of the D'D3 sub-domains that resulted in exposure of the multimerization of the cross-linking sites. This mode of opening is consistent with the elongation measured by optical tweezers (Figure 31, D) (see further information in ref. [Gruber et al, 2022]).
Overall, our studies contributed to the functional understanding of VWF by uncovering molecular details of less-studied regions of VWF beyond the canonical A domains.

Spaghetti in a bulk Martini 3 coarse-grained force field of poly(para-phenylene–ethynylene)s

Poly(para-phenylene–ethynylene)s (PPEs) are a class of conjugated and semi-conductive polymers with a strong delocalized π electron system that extends over the entire backbone.
The backbone of PPEs is composed of aromatic rings and stiff triple bonds, which are linked by single carbon bonds (see Figure 32 A). Based on the alternation of single and multiple bonds, the orbitals of the carbon atoms overlap, and the π electrons are delocalized from one end of the polymer to the other.
The delocalization of the π electrons is the reason for both the planar structure of monomers and the linear backbone together with its increased mechanical bending stiffness. Therefore, PPEs can be classified as semi-flexible polymers, and their bending stiffness is characterized by their persistence length. The persistence length of PPEs is in the range of the total polymer chain length and its bending stiffness and is comparable to a piece of spaghetti that has been cooked "al dente."
In order to better understand the large-scale assembly of semi-flexible polymers, we developed a coarse-grained (CG) model for PPEs using the Martini 3 force field to simulate the dynamics of larger systems on longer timescales.
The Martini 3 force field is one of the most commonly applied CG force fields in the field of biomolecular and material sciences. Martini 3 offers the advantage of being able to capture large spatio-temporal scales of molecular systems while adjusting the level of the resolution of the individual molecular components. Hence, the force field is suitable for simulating large bulks of PPEs while still taking into account the essential chemical details of the aromatic rings and triple bonds, thereby giving rise to enhanced chain stiffness. For PPEs, not only did we follow the classical way of developing coarse-grained models with Martini 3, but we also used a geometrical model in order to consider the shape of the conjugated backbone in greater detail (see Figure 32 A). Additionally, we employed a harmonic bond angle potential in order to tune the bending stiffness of a single chain toward our experiments. Our developed CG model now opens the door to analyzing PPE assemblies at larger length and time scales.
Figure 32 B presents snapshots of a bulk system that contains 500 PPEs that consists of 20, 40, and 60 monomers each. It is evident that short chains mostly align in parallel (comparable to uncooked pieces of spaghetti), whereas this alignment is lost across larger distances when the chain length grows. In order to quantify this nanometer-scale alignment of polymers within the network, we calculated the nematic correlation function (NCF), which describes the decay in structural order with increasing radial distance from a reference monomer. We observed a steady decrease in ordering with increasing polymer length across the whole range of radial distances. PPEs with a chain length in the range of their persistence length (20 monomers) showed a higher degree of alignment in comparison with longer-chain PPEs (40 or 60 monomers). The latter PPEs were less parallel and were instead assembled via entanglement as opposed to alignment. In this case, parallel alignment was maintained on a length scale of up to ~5 nm, beyond which a nano-domain formed that consisted of PPE chains aligned along a different direction.
The decrease in alignment with increasing polymer chain length was attributed to the competition between entropic and enthalpic effects. The influence of the entropic effects on the polymer dynamics increased with increasing polymer length, and the long-chain PPEs thus exhibited a random coil-like behavior with less ordering. Similar behavior can be observed in daily life when cooking spaghetti because initially rigid noodles become more entangled the longer they are cooked.
Taken together, the loss of interchain alignment was more pronounced for long-chain PPEs that had formed nano-domains on the ~5 nm scale and that were independent of the system size. Longer chains showed more chain fluctuations and thus entanglement, which impeded the preferred interchain alignment observed for PPE chains with shorter chain lengths – that is, lengths in the range of their persistence length.

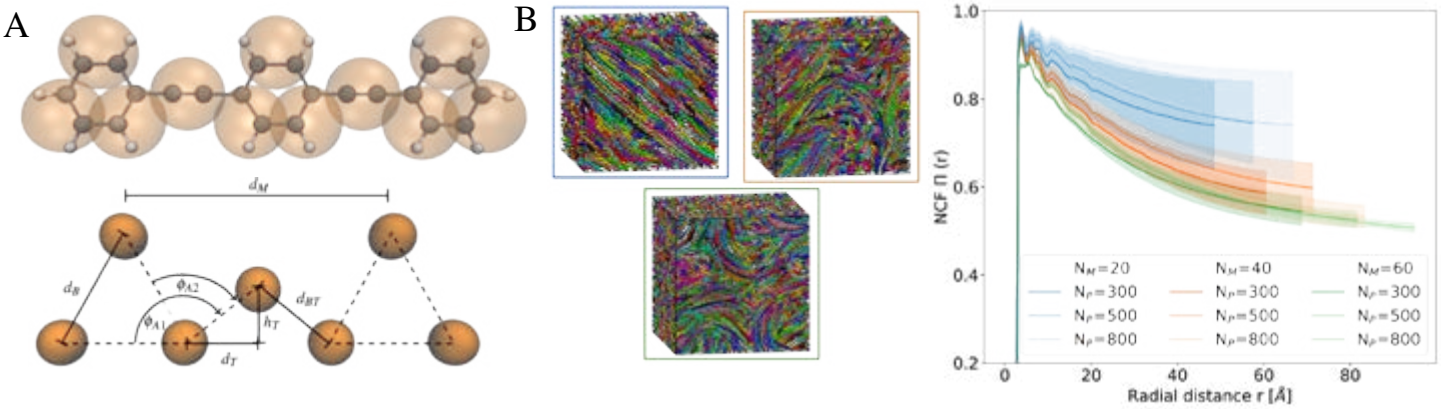


Figure 32. Mapping scheme and geometrical modeling of poly(para-phenylene–ethynylene)s (PPEs) and the alignment of large-scale PPE bulk systems. (A) Mapping from AA (black) to CG (orange) resolution. The dotted lines represent the bonded potentials of the CG model. The geometric model takes the shape of the PPE backbone into account. PPEs are represented as a linear chain of regular triangles with side length dB and spacing dM. The remaining equilibrium values are calculated via trigonometrical relations. (B) Snapshots of bulk systems of 500 PPEs with a length of 20, 40, or 60 monomers. Comparison of the NCF for PPE bulk systems of various numbers of monomers NM and polymers NP. Global ordering of PPEs decreases with increasing chain length.

Bond dissociation energies in amino acids

The radical chemistry of proteins is of fundamental importance in biochemistry. Radicals are involved in many enzymatic reactions. Moreover, oxidative damage to proteins is generally initiated by radicals. A common reaction of radicals in proteins is hydrogen atom transfer (HAT) – that is, the exchange of a hydrogen and a radical:

$$X\text{-}H + Y^{\cdot} \rightarrow X^{\cdot} + Y\text{-}H$$

In order to calculate the reaction energy of such HAT reactions, the energy necessary to break and build these bonds – namely the bond dissociation energy (BDE) – must be known. BDE is a key chemical quantity that provides insights into the kinetics and mechanisms of radical reactions.
We computed and published a comprehensive dataset of BDEs for every hydrogen in every natural amino acid by utilizing density functional theory. The accuracy of our dataset was further improved by employing isodesmic reactions, which allowed for error correction via the use of chemically similar experimental references, thereby bringing our accuracy close to experimental accuracy.
Our dataset provides a reliable base for making predictions using machine learning or calculations of thermodynamic and kinetic quantities.

Proteine sind die Arbeitspferde lebender Systeme. In den vergangenen Jahrzehnten haben wir viel darüber gelernt, wie Proteine aussehen, sich bewegen und funktionieren. Ein ziemlich neuer Aspekt ist, dass viele Proteine sehr überraschende Eigenschaften zeigen, sobald eine mechanische Kraft auf sie einwirkt, mit wichtigen Konsequenzen für ihre Funktion in der biologischen Zelle. Ziel der Arbeitsgruppe **Molekulare Biomechanik** ist es, die mechanische Funktion von Proteinen zu entschlüsseln. Wir verwenden molekulardynamische Simulationen verschiedener Auflösung sowie auch Experimente, um diese Frage auf mehreren Zeit- und Längenskalen zu beantworten.

Seit vielen Jahren tragen wir zum Verständnis bei, wie die im fließenden Blut vorhandene mechanische Kraft die Wirkung von Proteinen reguliert, die an der Blutgerinnung beteiligt sind. Unser Fokus ist hierbei der von-Willebrand-Faktor. 2022 konnten wir die Wirkungsweise von zwei weiteren der vielen Strukturbausteine dieses riesigen multimeren Proteins, das Teil unseres Blutes ist, entschlüsseln. Die sogenannte C6-Domäne und die D'D3-Domänen zeigen unerwartete Aktivitätsänderungen, wenn sie gestört werden, durch Mutationen oder unterschiedliche Redoxzustände. Wir waren in der Lage, die zugrunde liegende Ursache durch Molekulardynamik-Simulationen zu lösen.

Proteine sind am Ende des Tages Polymere, und man kann von Polymeren viel darüber lernen, wie Proteine auf Kräfte reagieren. Im Exzellenzcluster 3DMM2O zeigen wir, dass man sehr robuste Polymere wie zum Beispiel eine spezielle, vollkonjugierte Klasse von Polymerketten, mechanisch brechen kann. Im Jahr 2022 konnten wir einen ersten Schritt zur skalenübergreifenden Beschreibung solcher Prozesse mithilfe von Simulationen machen.

Unsere Aktivitäten im Rahmen des ERC Consolidator Grants RADICOL zu Mechanoradikalen in Kollagen sind 2022 auf Hochtouren gekommen. Wir können nun auf unseren Ergebnissen zu Radikalenergien in Proteinen (im Rahmen des HITS Lab Projektes „Emulation in Simulation") sowie Mechanismen der Bindungsspaltung in Polymeren aufbauen. Bleiben Sie dran für die kommenden Erkenntnisse über das Wie, Wann und die Folgen von Kollagen und Radikalen im Jahr 2023 und darüber hinaus!

2 Research

2.9 Molecular and Cellular Modeling (MCM)

Group leader

Prof. Dr. Rebecca Wade

Team

Christina Athanasiou
Ainara Claveras Cabezudo (until June 2022)
Dr. Giulia D'Arrigo
Michelle Emmert (April–September 2022)
Matheus Ferraz (visiting scientist; Federal University of Pernambuco, Recife, Brazil; DAAD scholarship)
Manuel Glaser
Dorothee Gross (January–March 2022)
Anton Hanke (until September 2022)
Vera Heesch (March–May 2022)
Melanie Käser (since May 2022)
Marcel Meyer (visiting scientist; Heidelberg University)
Dr. Karolina Mitusinska (visiting scientist; Silesian University of Technology, Gliwice, Poland; September–December 2022)
Emanuele Monaci (since November 2022)
Abraham Muniz Chicharro
Jakob Niessner (since October 2022)
Dr. Giulia Paiardi
Dr. Stefan Richter
Marco Rizzi (visiting scientist; University of Genoa, Italy; since September 2022)
Jonathan Teuffel
Lorenz Thielbeer (February–September 2022)
Alexandros Tsengenes
Congcong Xu (since November 2022)

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

In the MCM group, we are primarily interested in understanding how biomolecules interact. What determines the specificity and selectivity of a drug–receptor interaction? How can proteins assemble to form a complex? 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 makes concerted use of computational

approaches based on physics and bio-/chem-informatics. The broad spectrum of techniques developed and employed ranges from interactive, web-based visualization tools to machine-learning methods and atomic-detail molecular simulations.
In this report, we outline some of the results achieved in 2022. Following a general overview of what was new in the group last year, we describe our results in three research areas: (i) structure-based drug design, (ii) drugs in complex macromolecular environments, and (iii) the simulation of neurosignalling.

What happened in 2022?

In the third year of the coronavirus pandemic, we gradually transitioned to a hybrid way of working that now seems set to prevail. We aim to find the optimal balance between videoconferences and in-person meetings as well as between home-office and working in-person at HITS. This transition is exemplified by the meetings we organized in 2022: the first as a virtual meeting, the second as a streamed in-person event, and the third as a fully in-person event.
Together with partners in the Human Brain Project (HBP), Giulia D'Arrigo and Rebecca Wade organized the HBPMolSim Human Brain Project Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades as a virtual event in March (for details, see Section 5.1.1). This 4-day event was attended by 68 participants from 18 countries, and all participants had the opportunity to put the computational tools developed in the HBP into practice through interactive hands-on sessions. We presented the computation of binding kinetics using our τRAMD approach and SDA software.
The SIMPLAIX inaugural symposium – chaired by Rebecca Wade – was held in April in Studio Villa Bosch. For details on this and other SIMPLAIX activities, see chapter 7.
In September, the 23rd European Symposium on Quantitative Structure–Activity Relationships (EuroQSAR) was held in Heidelberg (see Chapter 5.1.6). Rebecca Wade was the chair of the symposium, and most of the MCM group was involved in helping the meeting to run smoothly in addition to presenting their own work. Christina Athanasiou won one of the poster prizes, and Giulia D'Arrigo and Giulia Paiardi organized a very well-received workshop entitled "Orienting your career compass," that took place during the meeting.
We welcomed two visiting scientists to the group in the autumn. Karolina Mitusinska (Silesian University of Technology, Gliwice, Poland) joined us from Artur Gora's laboratory for three months to learn about Brownian dynamics simulations with SDA and to show us how to use her AQUADUCT code to analyze water motion in molecular dynamics simulations of protein systems. Marco Rizzi (University of Genoa, Italy) joined us in September from Michelle Tonelli's group to perform computer-aided docking to potential enzyme targets of anti-parasitic compounds that he has synthesized during his doctoral studies.
Ainara Claveras Cabezudo and Anton Hanke completed their master's theses in Molecular Biotechnology and went on to doctoral studies in Frankfurt and Geneva, respectively. Jonathan Teuffel completed his master's thesis in Biochemistry and chose to remain in the group to work as a PhD student on a SIMPLAIX project on interprotein electron transfer. Michelle Emmert completed her bachelor's thesis in Molecular Biotechnology. Furthermore, Dorothee Gross (Molecular Biotechnology), Vera Heesch (Biochemistry), and Emanuele Monaci (Molecular Biosciences) carried out internships in the group as part of their master's studies at Heidelberg University.
Two new collaborative projects began during the year in the Heidelberg University Flagship Initiative "Engineering Molecular Systems": one in which our simulation work on the mechanism of heparin-SARS-CoV-2 spike interactions is combined with experimental investigations, and one that applies our simulation techniques to investigate the evolution of antibodies that are able to bind malarial antigenic peptides. Giulia Paiardi won a postdoctoral fellowship in the AI Health Innovation Cluster, an initiative of the Health+Life Science Alliance Heidelberg Mannheim, and a Joachim Herz Foundation "Add-on Fellowship for Interdisciplinary Life Science 2022" to work on a collaboration with Dr. Elke Burgermeister (University Hospital Mannheim) on the project "PepAISim: Combining AI and molecular simulation for anticancer peptide and peptidomimetic design."

Structure-based drug design

We apply structure-based drug design (SBDD) methods in multidisciplinary projects that aim at developing new therapeutic agents or diagnostics for cancer- [Costantino et al., 2022] [Rusnati et al., 2022], coronavirus- [Toral-Lopez et al., 2022] [Paiardi et al., 2022], cardiac, neurodegenerative [Rogdakis et al., 2022], bacterial [Eisenberg et al., 2022], and parasitic [Schmidt et al., 2022, Panecka-Hofman et al., 2022, Pöhner et al., 2022] diseases.

In the EU-supported NMTrypI project, we focused on drugs against neglected tropical diseases caused by trypanosomal parasites. These diseases result in a significant health and economic burden in developing countries. There are few effective and accessible treatments for these diseases, and existing therapies suffer from problems such as parasite resistance and side effects. In our research, we focus on targeting the folate pathway, and we recently reviewed the

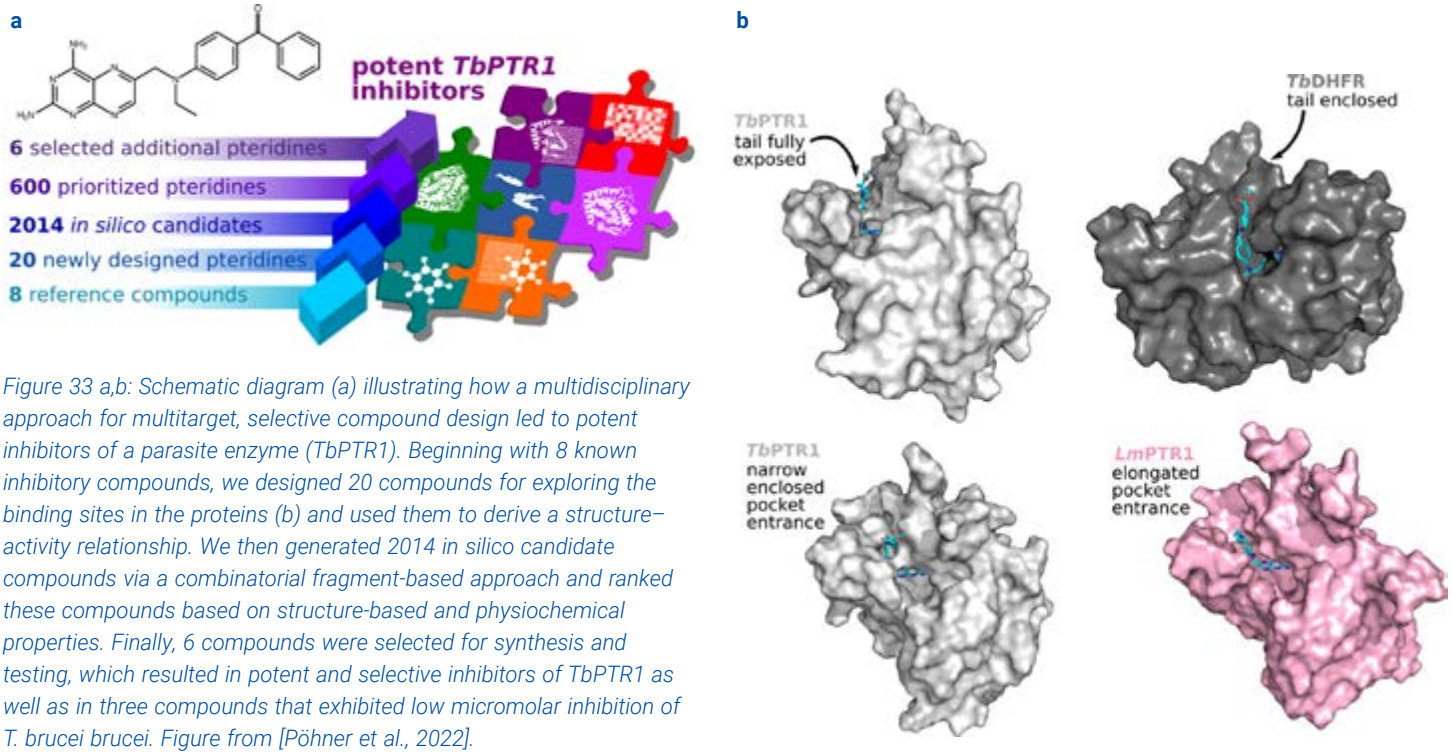


Figure 33 a,b: Schematic diagram (a) illustrating how a multidisciplinary approach for multitarget, selective compound design led to potent inhibitors of a parasite enzyme (TbPTR1). Beginning with 8 known inhibitory compounds, we designed 20 compounds for exploring the binding sites in the proteins (b) and used them to derive a structure–activity relationship. We then generated 2014 in silico candidate compounds via a combinatorial fragment-based approach and ranked these compounds based on structure-based and physiochemical properties. Finally, 6 compounds were selected for synthesis and testing, which resulted in potent and selective inhibitors of TbPTR1 as well as in three compounds that exhibited low micromolar inhibition of T. brucei brucei. Figure from [Pöhner et al., 2022].

application of SBDD for this task, highlighting the need for multiparameter optimization for discovering selective anti-parasitic drugs [Panecka-Hofman et al., 2022]. In 2022, we reported the discovery of potent inhibitors of kinetoplastid pteridine reductase 1 (PTR1) by using a systematic, multidisciplinary approach to developing multitarget, selective compounds (see Figure 33, from [Pöhner et al., 2022], Computational fragment-based design of novel pteridine derivatives along with iterations of crystallographic structure determination allowed for the derivation of a structure–activity relationship for multitarget inhibition. The approach yielded compounds that showed apparent picomolar inhibition of *Trypanosoma brucei* PTR1, nanomolar inhibition of *Leishmania major* PTR1, and selective submicromolar inhibition of parasite dihydrofolate reductase (DHFR) versus human DHFR. Moreover, by combining designing for polypharmacology with property-based on-parasite optimization, we found three compounds that exhibited micromolar EC50 values against *T. brucei brucei* while retaining their target inhibition. Our results provide a basis for the further development of pteridine-based compounds, and we expect our multitarget approach to be generally applicable to the design and optimization of anti-infective agents.

The rapid rise in antibiotic resistance indicates the urgent need to develop new antibiotics that specifically target pathogenic bacteria. The human pathogen *Streptococcus pyogenes* is an important pathogen that causes infections of varying severity, ranging from self-limiting suppurative infections to life-threatening diseases such as necrotizing fasciitis and

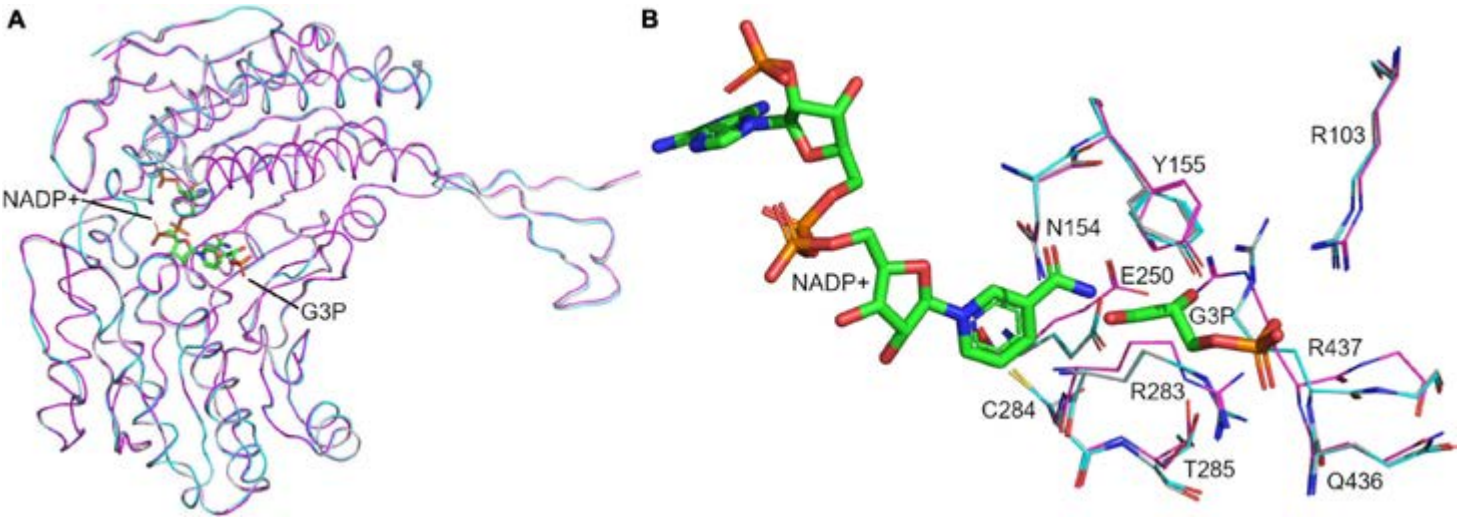


Figure 34: Modeled structure of holo GapN from Streptococcus pyogenes. Superimposition of the crystal structure of the apo GapN (C284S mutant) (magenta), the homology-modeled holo GapN from S. pyogenes (cyan), and the modeling template crystal structure from S. mutans (gray). (A) Aligned backbones of the subunits, showing their very high similarity. (B) Aligned binding-site residues, showing some differences in side-chain positions. The cofactors are shown in stick representation, with carbons colored green. Figure from [Eisenberg et al., 2022].

streptococcal toxic shock syndrome. Together with Tomas Fiedler (Rostock University, Germany), we are currently investigating the non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase GapN as a potential new drug target in *S. pyogenes*. Experiments show that GapN is an essential enzyme for *S. pyogenes* as it provides NADPH, which is otherwise missing in *S. pyogenes* and other streptococci. Modeling of *S. pyogenes* GapN yielded a structure very similar to the crystal structure subsequently determined by Hermann Schindelin and colleagues (University of Würzburg, Germany) (see Figure 34). Furthermore, molecular docking enabled us to correctly predict the competitive inhibition of *S. pyogenes* GapN by erythrose 4-phosphate [Eisenberg et al., 2022]. We therefore proceeded to perform virtual screening of large compound libraries against GapN structures and are currently selecting compounds to test for inhibitory activity.

Drugs in complex macromolecular environments

Crowded environments – such as those found in the cell – are known to affect the diffusion of macromolecules, but the effects of these environments on the diffusion of small molecules remain largely uncharacterized. In a collaboration with Debrata Dey and Gideon Schreiber at the Weizmann Institute (Israel), we investigated how three protein crowders – bovine serum albumin, hen egg-white lysozyme, and myoglobin – influence the diffusion rates and interactions of four small molecules: the dye fluorescein and the three drugs doxorubicin, glycogen synthase kinase-3 inhibitor SB216763, and quinacrine. Using Line-FRAP (fluorescence recovery after photobleaching) measurements, Brownian dynamics simulations (reviewed in [Muniz Chicharro et al, 2022]), and molecular docking, we found that the diffusion rates of the small molecules are highly affected by self-aggregation, by interactions with the proteins, and by surface adsorption (Figure 35). The diffusion of fluorescein decreased due to its interactions with protein crowders and their surface adsorption. Protein crowders increased the diffusion rates of doxorubicin and SB216763 by reducing surface interactions and self-aggregation, respectively. Quinacrine diffusion was not affected by protein crowders. The results indicate how compounds could be optimized for higher mobility in complex macromolecular environments.

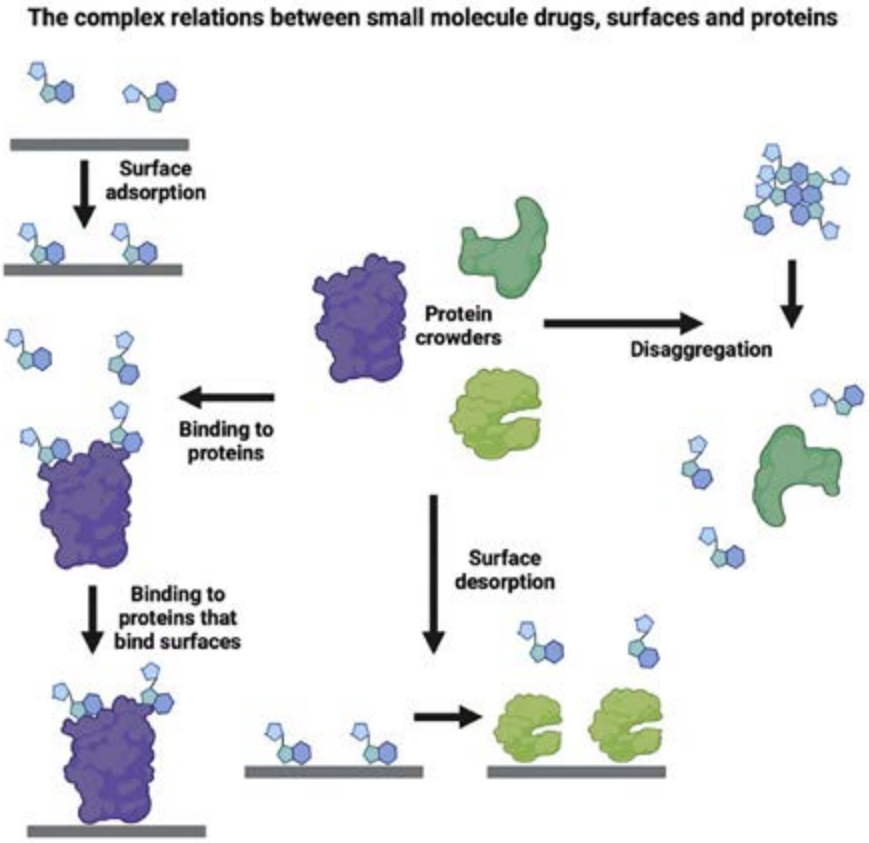


Figure 35: The effects of protein crowders on the diffusion of small molecule drugs go beyond the slower diffusion caused by excluded volume. We found that diffusion rates can be decreased due to surface adsorption or binding to proteins that are adsorbed on a surface. On the other hand, diffusion rates can be increased by surface desorption or via disaggregation of the small molecules. The specific effect of a protein crowder on the diffusion of a given small molecule depends on the physicochemical properties of both the crowder and the small molecule, which influence the protein-small molecule, protein-surface and small-molecule-surface interactions, and self-interactions. Figure from [Dey et al., 2022].

Simulation of neurosignalling

Together with partners in the EU-supported Human Brain Project, we are currently developing approaches for the multiscale molecular simulation of signaling networks [Keulen et al., 2023] and have highlighted the need for combining hypothesis- and data-driven modeling approaches in FAIR (Findable, Accessible, Interoperable, Reusable) workflows for this purpose [Eriksson et al., 2022].

The EU-supported EuroNeurotrophin international training network came to the end of its funding period at the beginning of 2022. The project focused on the discovery of neurotrophin small molecule mimetics as potential therapeutic agents for neurodegeneration and neuroinflammation, and the first paper revealing a compound with promising neuroprotective properties was published in 2022 [Rogdakis et al, 2022] (see Figure 36, next page).

We focused on modeling and simulating neurotrophin receptors in order to understand how they bind neurotrophins and neurotrophin mimetics and how this binding leads to signal transmission through the cell membrane. Neurotrophin receptors are large transmembrane proteins that form dimers, and neurotrophins bind to the globular extracellular part of the receptors. We built and simulated models of the complete receptors, but their simulation is highly computationally intensive due to their large size. Therefore, we also simulated the component parts of the receptors in order to study their dynamics in greater detail.

The transmembrane part of each neurotrophin receptor is a single alpha-helix. In the receptor dimer, the helices of the two receptor subunits interact, and this interaction is important for conveying signals across the membrane. Although the transmembrane helix dimer is only a small part of the receptor, sampling all the possible ways in which the helices interact – and how these interactions depend on the amino acid sequence – is challenging in all-atom molecular dynamics simulations. Therefore, we carried out simulations using the Martini 3 coarse-grained force field, which has shown promising results for proteins in phospholipid bilayers. However, simulating other lipid environments – such as the detergent micelles used in NMR studies of helix dimers – presented challenges due to the absence of validated parameters for their constituent molecules. We therefore derived parameters for the micelle-forming surfactant dodecylphosphocholine (DPC). These parameters resulted in micelle assembly with aggregation numbers that were in agreement with experimental values. However, we identified a lack of hydrophobic interactions between transmembrane helix protein dimers and the tails of DPC molecules, which prevented the insertion and stabilization of the protein in the micelles. This problem was also observed for protein insertion by self-assembling phospholipid bilayers. We found that a reduction of the non-bonded interactions between protein and water beads by 10% provided a simple and effective solution to this problem that enabled protein encapsulation in phospholipid micelles and bilayers without altering protein dimerization or bilayer structure (see Figure 37). We then used this approach to simulate neurotrophin receptor transmembrane helix dimers and to investigate the active and inactive arrangements of these dimers.

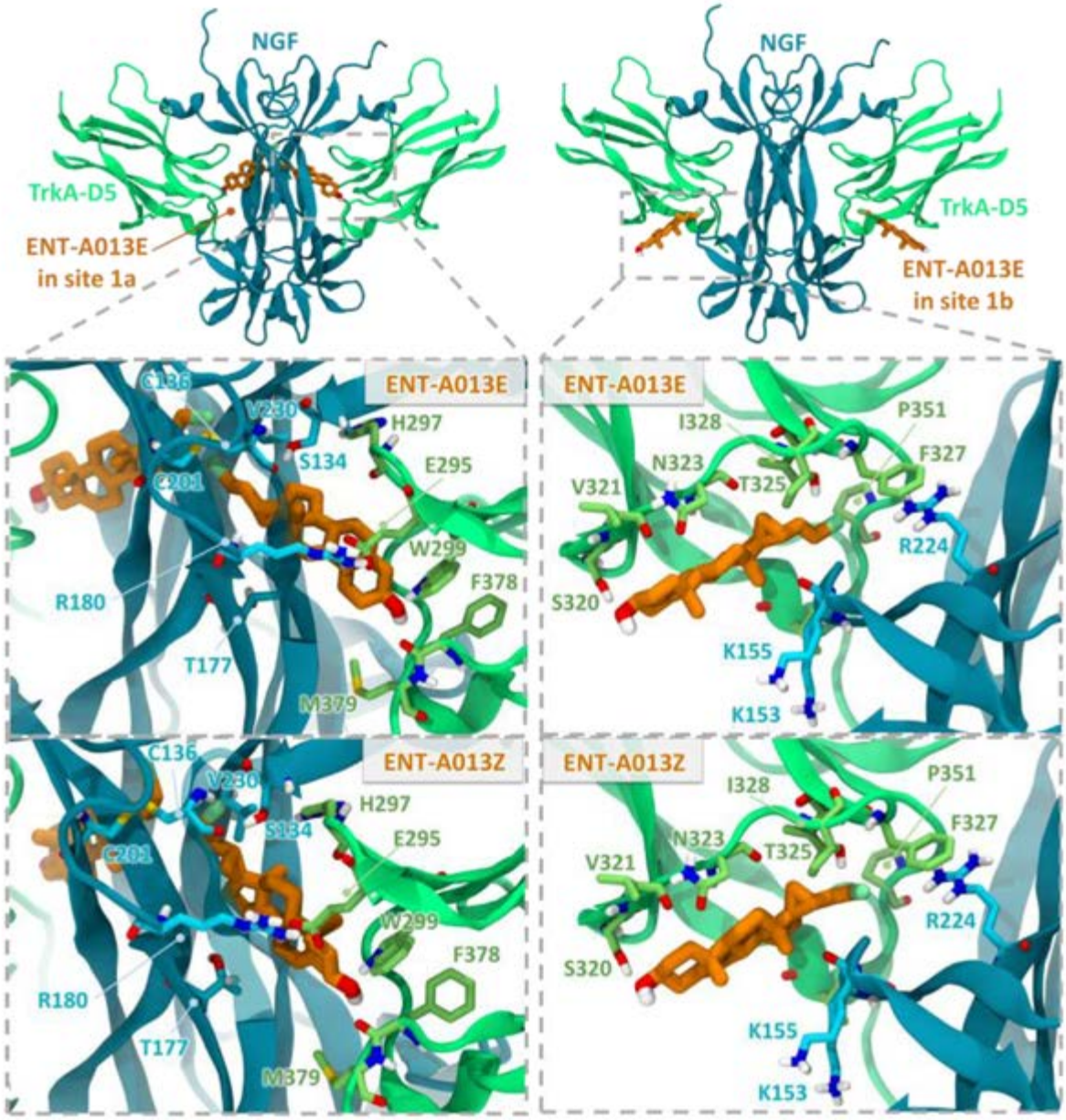


Figure 36: Docking poses of two isomers of the small molecule neurotrophin mimetic, ENT-A013 (orange), at symmetric interfacial sites in the TrkA receptor D5 domain (green)—NGF (blue) neurotrophin complex. The close-up views show residues lining the two sites. Figure from [Rogdakis et al., 2022].

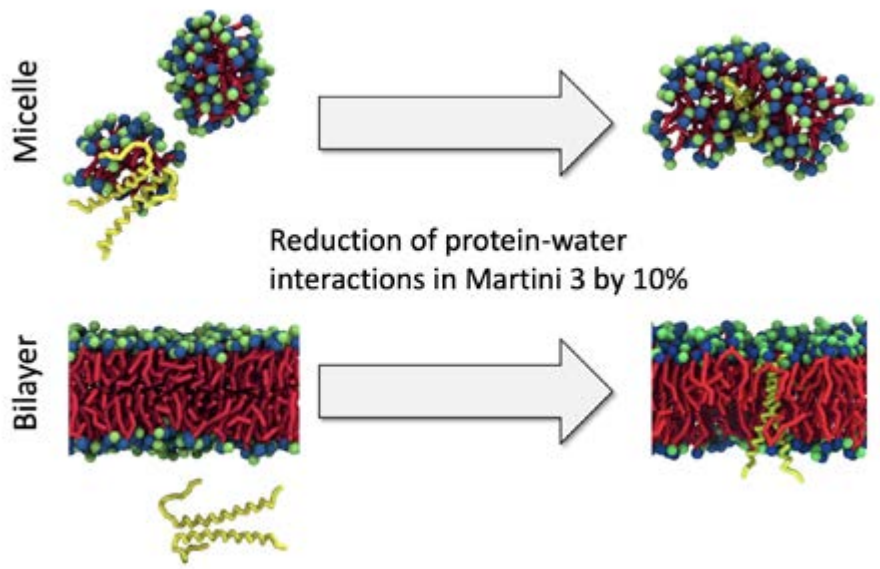


Figure 37: Simulation of transmembrane helix dimers in phospholipid micelles and bilayers. We found that down-scaling the protein–water interactions in the Martini 3 coarse-grained force field enables self-assembly of transmembrane helix dimers in phospholipid micelles and bilayers, which can then be simulated with (micelle) or without (bilayer) scaling of the protein–water interactions. Figure from Claveras et al.: Scaling Protein–Water Interactions in the Martini 3 Coarse-Grained Force Field to Simulate Transmembrane Helix Dimers in Different Lipid Environments, J. Chem. Theory Comput. 2023, 19, 7, 2109–2119.

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.

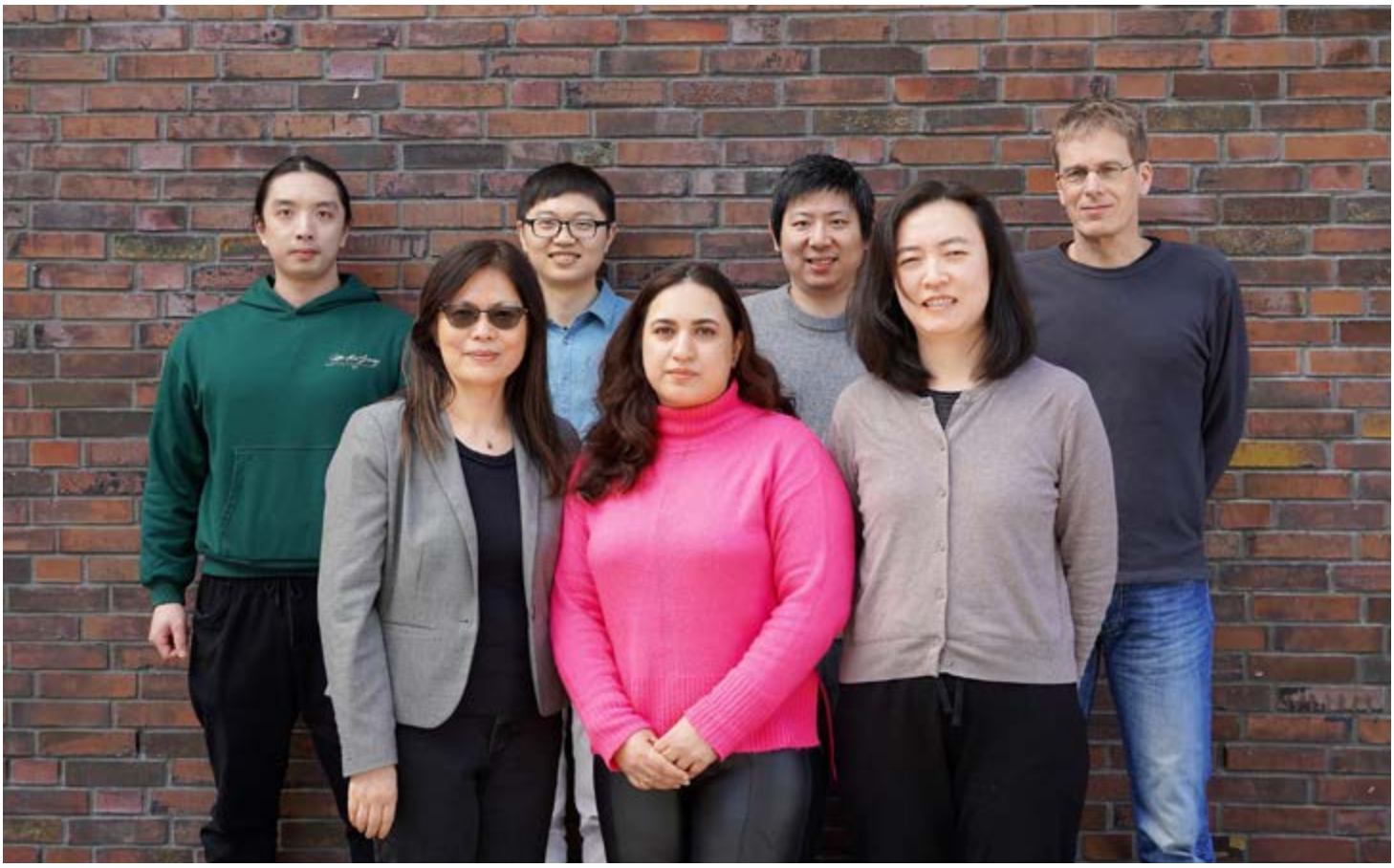
In der **Molecular and Cellular Modeling Gruppe (MCM)** sind wir in erster Linie daran interessiert zu verstehen, wie Moleküle interagieren. Was bestimmt die spezifische und selektive Wirkung beim Zusammenspiel von Wirkstoff und Rezeptor? Wie werden Proteinkomplexe gebildet und welche Formen können sie 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 Cheminformatik. 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 2022. Nach einem allgemeinen Überblick über Neuigkeiten in der Gruppe konzentriert sich der Bericht auf Projekte zu (i) struktur-basiertem Wirkstoffdesign, (ii) Wirkstoffmolekülen in komplexen makromolekularen Umgebungen, und (iii) zur Simulation von Neurosignalling.

2 Research

2.10 Natural Language Processing (NLP)



Group leader

Prof. Dr. Michael Strube

Team

Haixia Chai (HITS Scholarship holder)
Yi Fan (since November 2022)
Mehwish Fatima (visiting scientist; HEC-DAAD Scholarship)
Sungho Jeon (HITS Scholarship holder; until May 2022)
Tim Kolber (student; since September 2022)
Wei Liu (HITS Scholarship holder)
Yue Liu (student; since December 2022)
Xianghe Ma (student; since November 2022)
Tobias Martiné (student; until October 2022)
Dr. Mark-Christoph Müller (until May 2022)
Dr. Shimei Pan (visiting scientist; Fulbright Award; since October 2022)
Dr. des. Wei Zhao

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.

The NLP group has proudly hosted Shimei Pan as a guest scientist since October 2022. Shimei is a professor in the Information Systems Department at the University of Maryland, Baltimore County, USA. In recent years, her research has focused on biases in artificial intelligence and natural language processing as well as on extracting information from social media. Shimei's stay at HITS was made possible by a US Fulbright Award. In addition to interacting with members of the NLP group, Shimei has also co-taught a seminar at the Institute for Computational Linguistics at Heidelberg University together with NLP group leader Michael Strube, and she

regularly interacts with researchers at the university.
At the end of 2021, Federico López submitted his PhD thesis and left HITS to take up a job in industry. He finally defended his dissertation in 2022 with distinction. Congratulations! Federico's research stemmed from the highly successful collaboration between the NLP and GRG groups within the framework of the HITS Lab project on geometric deep learning. We decided to continue the collaboration, with postdoc Wei Zhao taking the lead.
Also in 2022, Mark-Christoph Müller concluded his work on the BMBF-funded project DeepCurate (together with the SDBV group). He has since left HITS and now works at the Leibniz Institute for the German Language in Mannheim in the field of linguistic annotation and annotation infrastructure. Kevin Mathews left HITS at the end of 2021 to begin working in industry. In addition to publishing a paper at the most prestigious NLP conference, ACL 2022, Sungho Jeon also began not only one, but two internships in industry. The first internship brought him to Seattle, Washington, USA, where he joined Amazon for three months, and the second internship began in early 2023 with Meta in Mountain View, California, USA. Finally, the NLP group was joined by new PhD student Yi Fan, who had recently obtained his master's degree in machine learning from UCL London.
Michael Strube was co-chair of the "Third Workshop on Computational Approaches to Discourse," which took place at COLING 2022 in Gyeongju, South Korea. Due to the still-ongoing COVID-19 pandemic, the workshop had to be organized as a hybrid event, which made it even more difficult to plan than a purely online workshop. (In fact, one of the invited speakers had planned to give his talk from the US but miscalculated the time difference to South Korea and showed up one day late!) The event also comprised a shared task on "Anaphora, Bridging, and Discourse Deixis in Dialogue," which was also co-organized by Michael Strube. The fourth iteration of the workshop is planned to be held at the ACL conference in July 2023 in Toronto, Canada.

Geometric Deep Learning

Wei Zhao

In recent years, a class of graph neural networks has become a popular choice for learning the numerical representation of graph data in various domains, such as social networks, biology, and molecular modeling, and these networks typically use Euclidean space to represent graph data. Nevertheless, Euclidean space has a grid-like structure that expands polynomially as more points are added, resulting in an inefficient use of space. Therefore, Euclidean space is often implemented in very many dimensions, but that can be computationally expensive and involve an excessive amount of model parameters. Furthermore, Euclidean space has been well-known to geometers for being inappropriate for modeling graphs with hierarchical structures, which are omnipresent in complex, real-world graph data. In 2022, the HITS Lab collaboration between the GRG and the NLP groups developed a new approach to modeling complex graphs in high-rank, non-compact symmetric spaces. These spaces were shown to be able to accommodate compound geometry with flat and negatively curved subspaces, which makes them particularly well-suited for modeling complex graphs. Within this

Figure 38: Evaluation of five graph neural networks in three spaces on real-world datasets: Citeseer (left) and Cora (right). Gridlines with circular shapes show the classification accuracy on a scale of 0–100. Results reveal that using SPD space to build graph neural networks leads to improved accuracy.

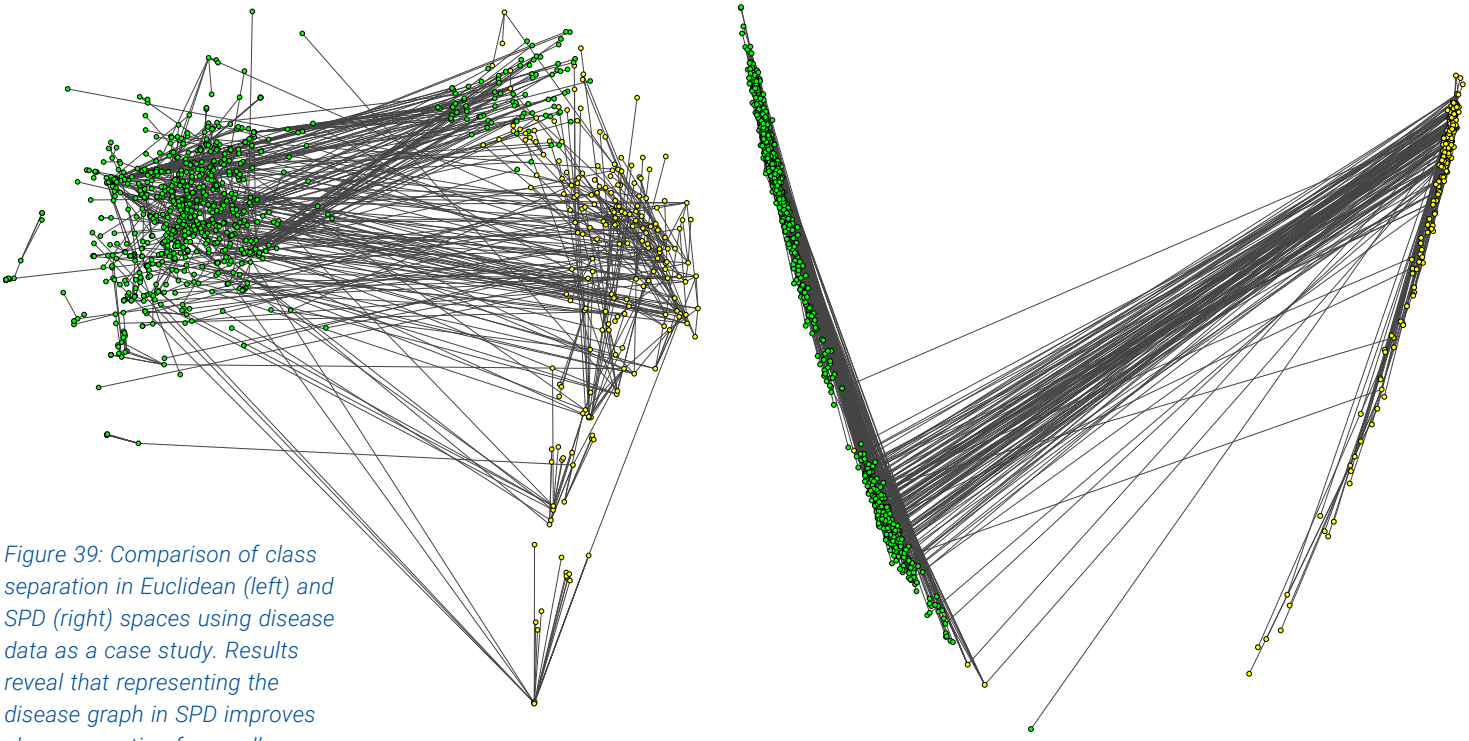


Figure 39: Comparison of class separation in Euclidean (left) and SPD (right) spaces using disease data as a case study. Results reveal that representing the disease graph in SPD improves class separation from yellow (right) to green (left).

project, we chose the space of symmetric positive definite (SPD) matrices as the embedding space for building graph neural networks as these matrices are considered particularly efficient at computing in the family of symmetric spaces.

Our large-scale empirical study revealed that using SPD space to build graph neural networks is more space-efficient than is using Euclidean and hyperbolic spaces. This finding has led to beneficial outcomes on machine learning tasks in terms of performance in low-dimensional spaces. We are currently in the process of making a submission to the International Conference on Machine Learning (ICML 2023).

Incorporating Centering Theory into Neural Coreference Resolution

Haixia Chai

Coreference resolution involves finding all expressions in a text that refer to the same entity. Coreferent mentions can occur anywhere in the discourse. In recent years, many transformer-based coreference resolution systems have achieved remarkable improvements on the CoNLL data, which is a commonly used benchmark for evaluating coreference resolution systems. However, the way in which knowledge about discourse structure can benefit coreference resolution has been less often explored in the neural NLP era.

Coreference plays an essential role in discourse coherence. A referring expression that uses a reduced linguistic form (e.g., a pronoun) indicates a referential relation to its antecedent in previous utterances. The referring expression connects utterances and contributes to discourse coherence. On the other hand, coreference resolution can benefit from coherent discourse. The notion that coherence structure can impose constraints on referential accessibility has long been acknowledged from a linguistic perspective. Centering theory is a method of formally describing discourse coherence by using attentional state (i.e., the focus of attention of the participants at each utterance in the discourse). Figure 40 displays how the coherence structure of an example text is built by means of tracking the changes in the local attentional state.

In this work, we propose incorporating centering transitions derived from centering theory in the form of a graph into a neural coreference model. First, we capture the most salient mentions of each sentence as centers in order to compute the local centering transition relationships in accordance with centering theory. We then extend the coherence structure globally in the form of a graph, which makes the centering transitions available between any two sentences. Lastly, we fuse the novel discourse

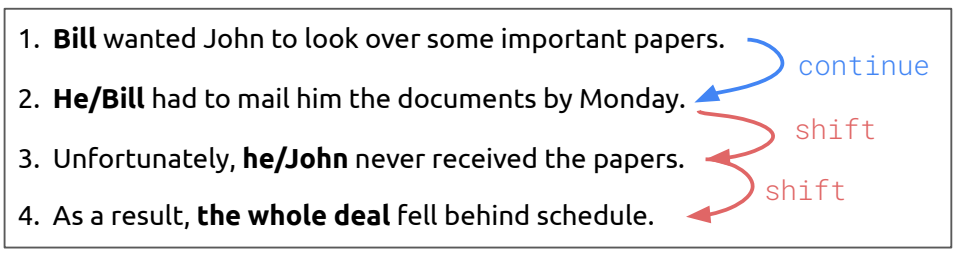


Figure 40: An example text shows how foci change sentence-by-sentence. The words in bold are the focus of each sentence. The arrows indicate centering transitions with two different transition types: continue and shift.

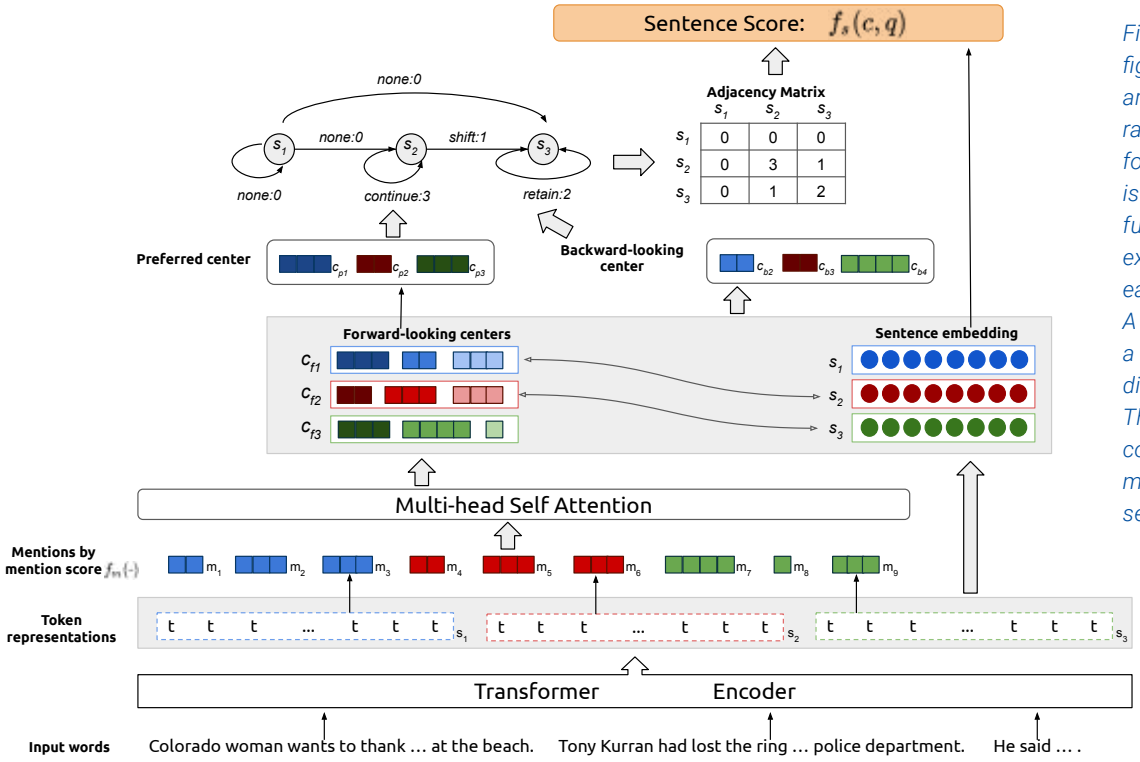


Figure 41.: Caption: The figure shows our model architecture, which incorporates centering transitions for the sentence score that is part of the final scoring function. There are three example sentences: one each in blue, red, and green. A string of squares refers to a mention that comprises a different number of tokens. The mention with a darker color indicates that it is a more salient center in a sentence.

structure into a neural coreference model. Figure 41 presents our model architecture.

Our proposed method improves the SOTA models up to a score of 80.9 F1. Our extensive analysis reveals that our approach performs better on pronoun resolution in long documents, formal well-structured text such as magazine and newswire genres, and documents with scattered mentions of clusters. Overall, we have observed that incorporating discourse structure derived from centering theory can benefit coreference resolution.

Cross-Lingual Summarization

Mehwish Fatima

A real-world example of cross-lingual science journalism is Spektrum der Wissenschaft. Spektrum is a popular German science magazine in Germany and represents an acclaimed bridge between local readers and the latest scientific research. Spektrum has one section (both in print and online) where journalists can read complex English-language scientific articles and convert them into popular science stories in German that should be comprehensible to local non-expert readers. These stories are summarized versions of original articles and are written in straightforward terms in a local language. Spektrum der Wissenschaft asked us to automate the process of their journalist's work. Therefore, we defined the job as the fusion of two high-level Natural Language Processing (NLP) tasks: text simplification and cross-lingual scientific summarization.

In order to investigate cross-lingual science journalism, we proposed using a Multi-Task Learning (MTL)-based model that is trained in two tasks: simplification and cross-lingual summarization. MTL is an approach to deep learning that improves generalization by learning different noise patterns from data that are related to different tasks. Our proposed model jointly trains SIMplification and Cross-lingual SUMmarization (SIMCSUM) in order to improve

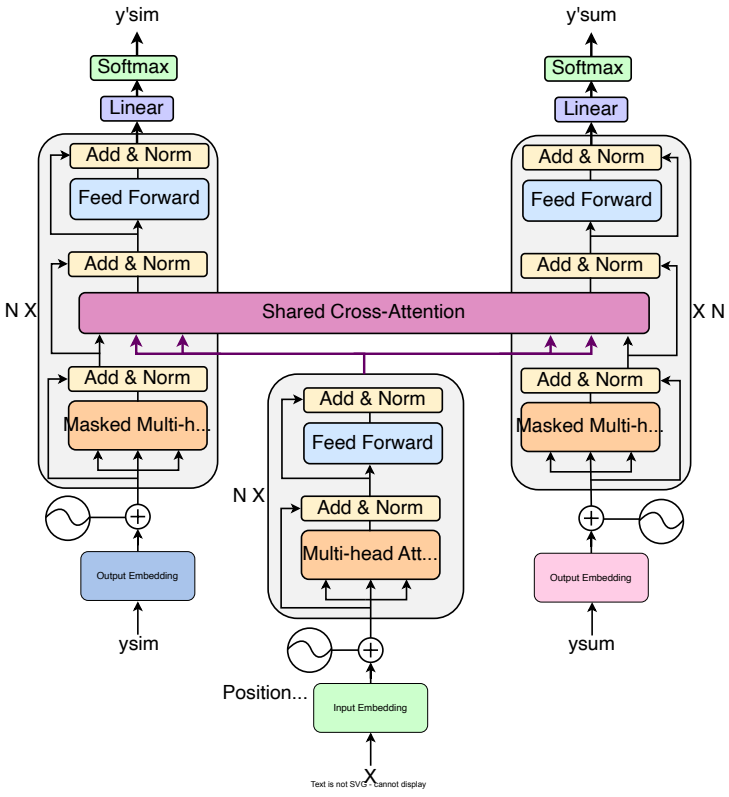


Figure 42: SIMCSUM architecture consists of one shared encoder with two decoding sides for simplification and cross-lingual summarization, which share cross-attentions.

the quality of cross-lingual popular science summaries. As shown in Figure 42 (previous page), SIMCSUM consists of one shared encoder and two independent decoders for each task based on a transformer architecture. The cross-attention between the decoders was concatenated. We combined each task's loss in order to calculate our model's total loss for updating the learnable parameters of the model. We introduced a variable λ with the task-specific loss in order to control the impact of each task during training. Finally, we considered cross-lingual summarization to be our main task and simplification to be our auxiliary task. Since the Spektrum dataset is too small to be used for training a neural network, we collected a second dataset from the Wikipedia Science Portal. Furthermore, we conducted our experiments using both cross-lingual scientific summarization datasets. We fine-tuned several summarization baselines with the Wikipedia dataset, and we also defined a pipeline-based baseline: Simplify-Then-Summarize (KIS-MB). We trained SIMCSUM on the Wikipedia dataset (for summarization) and its synthetic version for simplification, and we evaluated all models with three metrics: ROUGE (R-1, L) is a standard metric for summarization, BERT-score (BS) is a recent metric for summarization and simplification, and SARI and Flesch Kincaid Grade Level (FKGL) are the most frequently used metrics for English-language text simplification. We decided to use a variation of the Flesch Kincaid score for the German language – namely Flesch Kincaid Reading Ease (FRE) – because our output language was German.

Figure 43 displays the results of the experiments with three evaluation metrics. Gold represents the FRE score of reference summaries of the Spektrum dataset. The Figure shows prominent SIMCSUM performance over the baselines and a similar FRE score of SIMCSUM and Gold summaries. We also performed a human evaluation on a random sample set of SIMCSUM and mBART. The human evaluation additionally suggested that the SIMCSUM outputs are better than the mBART outputs. From these results, we inferred that SIMCSUM impacts the lexical and syntactic properties of the generated summaries to improve their readability.

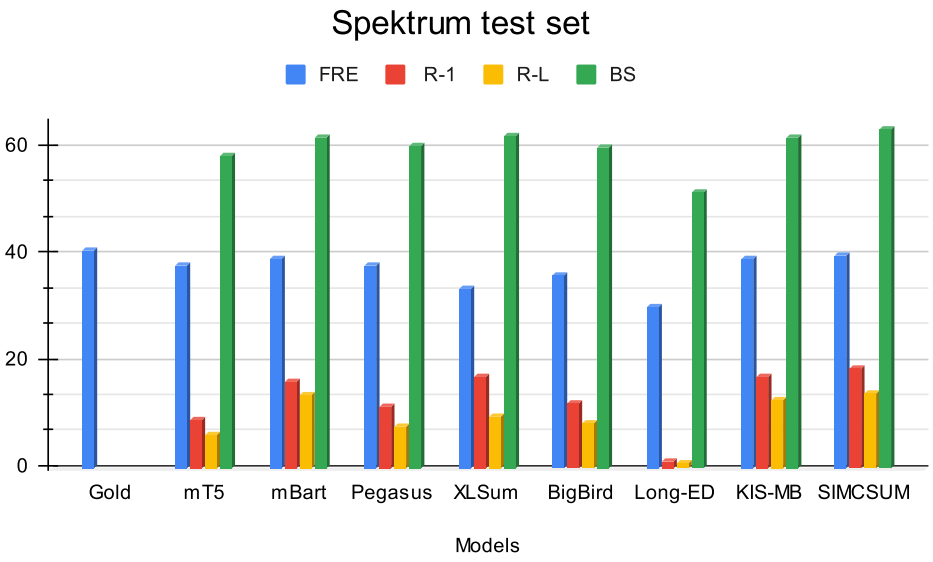


Figure 43: Evaluation results on the Spektrum test set.

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.

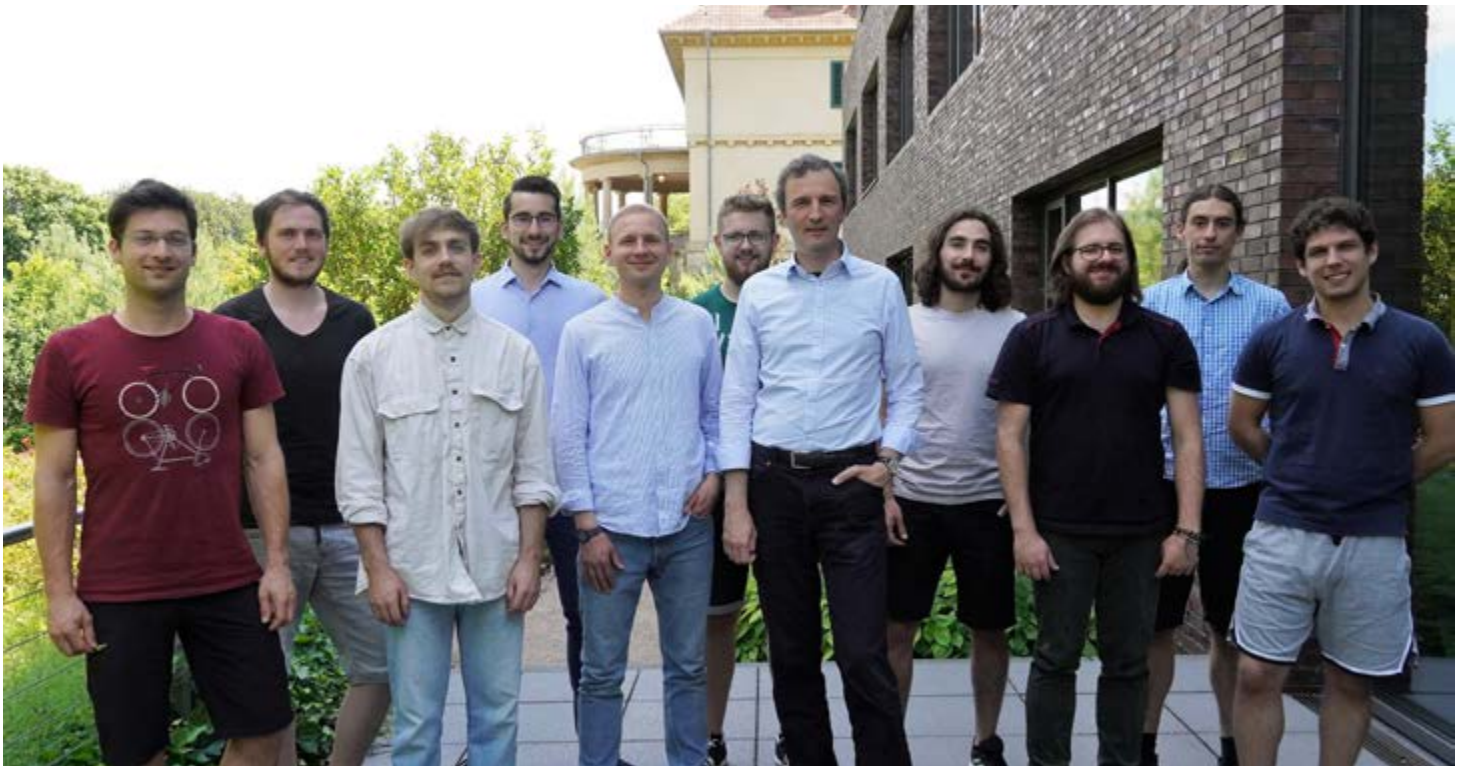
Die NLP Gruppe begrüßte im Oktober 2022 Shimei Pan als Gastwissenschaftlerin. Shimei ist Professorin im Information Systems Department der University of Maryland, Baltimore County, USA. Ihre Forschung konzentriert sich auf Bias (Voreingenommenheit) in Künstlicher Intelligenz und Natural Language Processing und auf das Extrahieren von Information aus sozialen Medien. Ihr Aufenthalt am HITS wird durch einen US Fulbright Award ermöglicht. Sie arbeitet aber nicht nur mit Wissenschaftler*innen am HITS, sondern auch am Institut für Computerlinguistik der Universität Heidelberg. Dort unterrichtet sie auch zusammen mit Michael Strube ein Seminar über Ethik in NLP.

Federico López reichte seine Doktorarbeit Ende 2021 ein und verließ HITS, um in der Industrie zu arbeiten. 2022 verteidigte er seine Dissertation mit Auszeichnung. Herzlichen Glückwunsch! Seine Forschung entstammte der höchst erfolgreichen Zusammenarbeit der NLP und der GRG-Gruppe im Rahmen des HITS Lab-Projekts Geometric Deep Learning. Wir entschieden, die Zusammenarbeit unter der Führung des Postdoktoranden Wei Zhao fortzusetzen.

Mark-Christoph Müller beendete die Forschung im vom BMBF geförderten Projekt DeepCurate. Er verließ HITS und arbeitet jetzt am Leibniz-Institut für Deutsche Sprache in Mannheim auf dem Gebiet der linguistischen Annotation und der Entwicklung von Annotationsinfrastruktur. Kevin Mathews verließ HITS Ende 2021, um in der Industrie zu arbeiten. Sungho Jeon publizierte nicht nur ein Papier bei der international angesehensten NLP-Konferenz, ACL 2022, sondern verließ HITS vorübergehend auch, um nicht nur eines sondern zwei Industriepraktika zu machen. Das erste führte ihn nach Seattle, Washington (USA), wo er bei Amazon arbeitete. Das zweite begann Anfang 2023 bei Meta in Mountain View, California. Schließlich begrüßte die NLP Gruppe im November den neuen Doktoranden Yi Fan, der gerade seinen Masterabschluss in Machine Learning vom University College London erworben hatte.

Michael Strube war Programm Co-Chair des "Third Workshop on Computational Approaches to Discourse", der im Rahmen der COLING 2022 in Gyeongju, Südkorea, stattfand. Wegen der immer noch grassierenden Covid-Pandemie fand der Workshop im hybriden Modus statt, was das Organisieren nicht einfacher machte. Nur als Anekdote: Einer der eingeladenen Sprecher trug über Videokonferenz aus den USA vor. Er verrechnete sich mit der Zeitdifferenz und war einen Tag zu spät. Der Workshop umfasste auch eine „Shared Task“ über „Anaphora, Bridging, and Discourse Deixis in Dialogue“, die von Michael Strube mitorganisiert wurde. Der vierte Workshop der Reihe ist schon in der Vorbereitung und wird im Rahmen der ACL 2022 im Juli 2023 in Toronto, Kanada, stattfinden.

2 Research
2.11 Physics of Stellar Objects (PSO)



Group leader
Prof. Dr. Friedrich Röpke

Team
Dr. Róbert Andrássy
Ferdinand Berwig
Dr. Johann Higl
Javier Morán Fraile
Alexander Holas (visiting scientist; Heidelberg University)
Florian Lach
Giovanni Leidi
Kiril Maltsev
Christian Sand (HITS Scholarship holder)
Theodoros Soultanis (IMPRS PhD student at MPIA Heidelberg)
Marco Vetter
Freyja Walberg
Gabriel Wiest

"We are stardust." Indeed, the very matter we are made of is largely the result of processing the primordial material that formed during the Big Bang, while heavier elements originate from nucleosynthesis in stars and in gigantic stellar explosions. Discovering how this material formed 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 in optical and other parts of the electromagnetic spectrum and are the fundamental building blocks of galaxies and 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 stellar explosions. Newly developed numerical techniques and the ever-increasing power of supercomputers facilitate the modeling of stellar objects in unprecedented detail and with unparalleled precision.
One of our group's primary goals is to model the thermo-nuclear explosions of white dwarf stars that lead to the astronomical phenomenon known as Type Ia supernovae. These supernovae are the main source of iron in the Universe and have been instrumental as distance indicators in cosmology, which has 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 and are also applied in the PSO group to other kinds of cosmic explosions. Classical astrophysical theory describes stars as one-dimensional objects in hydrostatic equilibrium – an approach that has proven extremely successful and 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.

A numerical tool for simulating stellar dynamos

Spectroscopic and photometric observations of active stars have revealed the presence of strong magnetic fields (up to several kG) on the surfaces of these stars. The closest example of stellar magnetism is our Sun, whose magnetic fields give rise to a whole variety of structures and magneto-hydrodynamic (MHD) processes that can be observed in great detail, including sunspots, coronal mass ejections, and the famous 22-year solar cycle.

Although magnetic fields are now accepted to be capable of strongly affecting the outer layers of certain stars, the origin of such fields is still under debate. In some cases, magnetic fields are likely created by the action of a dynamo, which converts the kinetic energy of turbulent stellar flows into magnetic energy. In some other cases, however, the observed magnetic field is likely inherited from earlier stages during the star's formation. To make things even more complicated, magnetic fields are often generated in the deep interiors of stars, for which we do not have direct observational constraints.

In order to understand how these fields are generated, we must rely on MHD simulations. A common tool used in astrophysical simulations is finite-volume discretization, which allows the conservation properties of a magnetized fluid to be retained (e.g., the conservation of total mass and energy). However, a certain degree of numerical dissipation must be added to the system in order to keep it stable and to prevent numerical discretization errors from excessively growing over time. Because most MHD codes are developed to model shocks and supersonic regimes, the timescale on which the dissipation acts is comparable to the sound crossing time over a grid cell. This constraint makes conventional codes unfeasible for modeling flows in stellar interiors, where shocks are absent and the typical velocities V are much smaller than the speed of sound c, which means that the Mach number M=V/c is very low. Under such conditions, the resulting flows would be completely dominated by numerical dissipation, which is undesirable.

In the past two years, members of the PSO group have dedicated a great deal of effort to overcome such numerical limitations. A new MHD method was successfully developed by Giovanni Leidi in his doctoral project and that is now part of the Seven-League-Hydro code, which was also built in-house. Several verification tests were performed in order to prove the method's reliability at simulating low-Mach-number MHD flows. Figure 44 presents a comparison between this improved MHD scheme (called LHLLD) and a more conventional method (HLLD) in simulations of a magnetized Kelvin–Helmholtz instability.

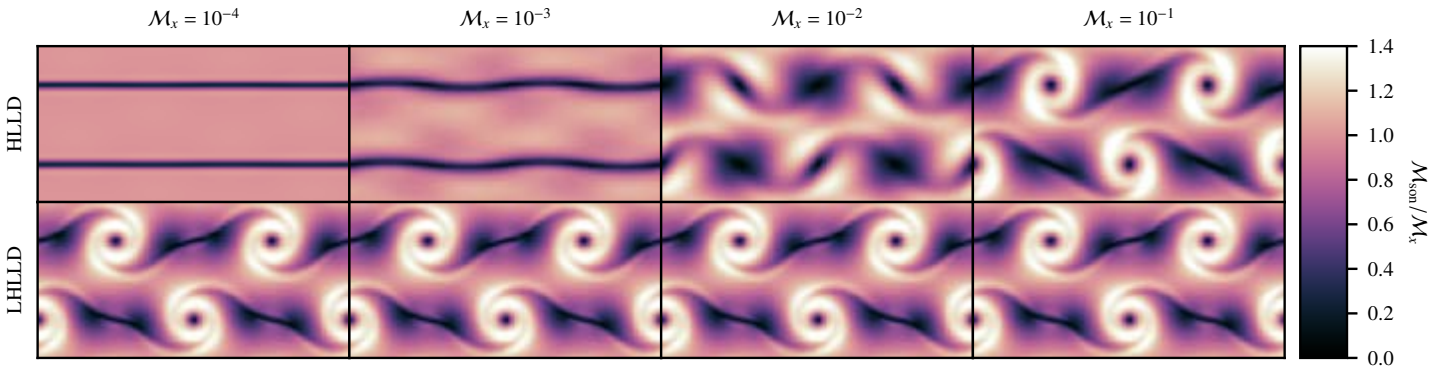


Figure 44: Snapshots of Mach number in simulations of a Kelvin–Helmholtz instability that were obtained with the HLLD (top panels) and LHLLD (bottom panels) solvers for different Mach numbers of the initial shearing flows. Figure from Leidi et al, 2022.

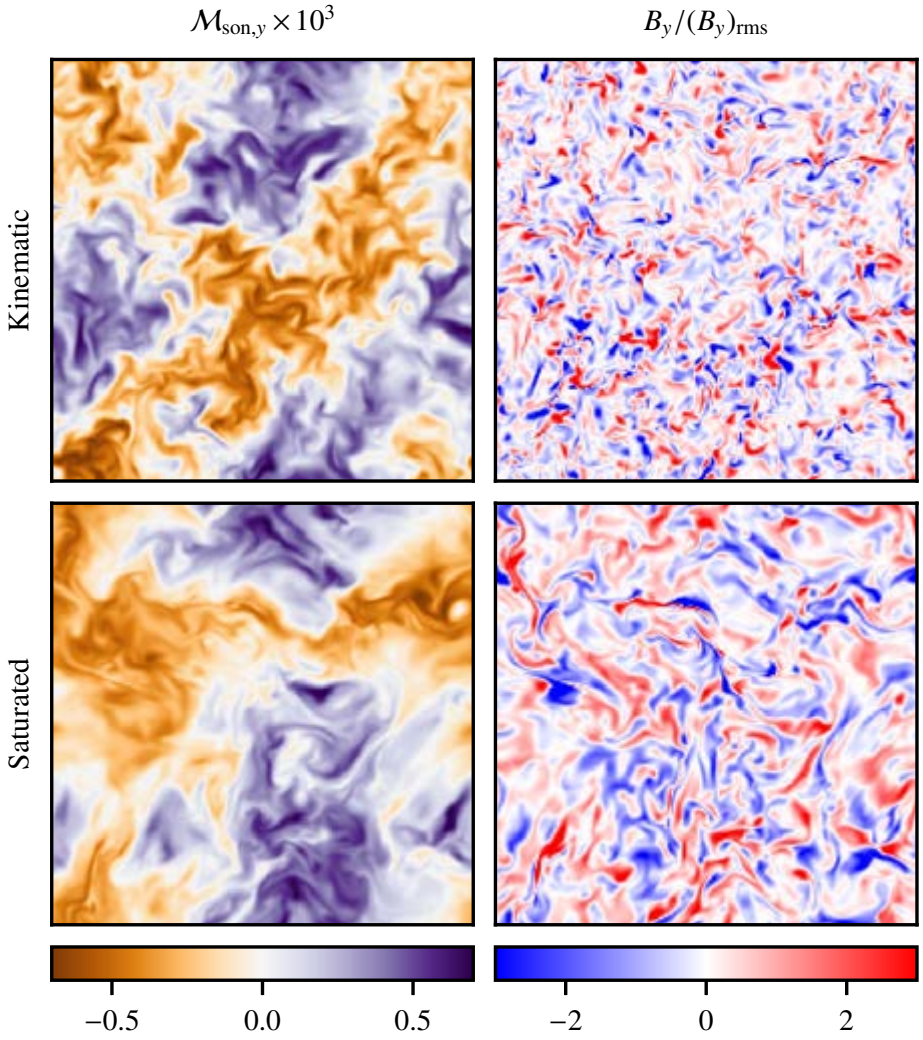


Figure 45: Horizontal slices taken in the midplane of a turbulent box with dynamo amplification showing the vertical Mach number on the left and vertical magnetic field on the right. The upper plots show these two quantities in the kinematic phase, during which the magnetic field is amplified by dynamo stretching, while the bottom plots show the final saturation stage, during which the Lorentz force is strong enough to damp the small-scale structures in the turbulent flow. Figure from [Leidi et al, 2022].

This comparison demonstrates how HLLD becomes progressively more diffusive as the Mach number of the shearing flows is reduced. In contrast, LHLLD performs well even in very low-velocity regimes.
As the ultimate goal of the new code is to test dynamo mechanisms in stars, Giovanni Leidi also set up a more challenging problem that involved convection and dynamo amplification as is found in the deep region of a massive star during a late phase of its life when oxygen is burned. The magnetic field is initially very weak, but it is exponentially amplified by dynamo and stretches up to the point at which Lorentz forces begin to act on the plasma. This feedback mechanism quickly damps the small-scale structures that are typical for turbulent flows. Such effects can be observed in Figure 45, in which snapshots of the magnetic and velocity field during the exponential rise and the final saturation stage are shown. This result may shed some light on the origin of the magnetic fields in stars. In fact, the fields that are generated by dynamo action can buoyantly rise toward the surface, where they can potentially be observed, thereby offering a unique glimpse into MHD processes that take place in very deep layers of stars.
All of these tests were published alongside a description of the newly implemented MHD method (see Leidi et al., 2022).

A special class of supernovae from explosions of Chandrasekhar-mass white dwarf stars

Type Ia supernovae have been instrumental as cosmic distance indicators. Indeed, they are bright and can be observed throughout large parts of the Universe. What is even more remarkable is that these supernovae seem to be rather homogeneous, which allows their peak brightnesses to be calibrated and enables distance measurements to be taken. One spectacular outcome of such observations was the discovery of the accelerated expansion of the Universe, for which the 2011 Nobel Prize in Physics was awarded.
Despite their astrophysical importance, the explosion mechanism behind Type Ia supernovae remains a puzzle. Consensus exists among researchers that these events arise from thermonuclear explosions of white dwarf stars that consist of carbon and oxygen. However, both the way in which such an object triggers explosive thermonuclear burning and its physical state at the onset of the explosion remain unknown. Although Type Ia supernovae are bright and well-observed, their progenitors are faint and have not yet been directly observed.

A convincing scenario is that the star accretes material from a companion until it reaches the limit of its stability: namely the famous Chandrasekhar limit of 1.4 solar masses. For a long time, this was the “textbook explanation” for Type Ia supernovae. There are, however, several arguments that can be made

against this scenario accounting for the majority of Type Ia supernovae. The PSO group at HITS has worked on alternative explosion models, (see [Collins, Gronow et al 2022]; [Pakmor, Callan et al. 2022]; [Battino et al 2022]), and we have therefore been forced to consider whether we should abandon the Chandrasekhar-mass explosion scenario. Over the past decade, it has become clear that Type Ia supernovae are not as homogeneous in their properties as was initially thought. Indeed, several peculiar sub-classes have been identified, some of which occur rather frequently.
This finding raises the question as to where the explosions of Chandrasekhar-mass white dwarfs should be situated in the landscape of all supernovae. Do they occur in nature at all? Can they be distinguished from other supernova classes? What are their specific observational imprints?

These questions are explored by the PSO group at HITS. With their specialized and highly efficient LEAFS code, the researchers from this group perform three-dimensional hydrodynamic simulations of the explosion stage on high-performance parallel computers. One particular challenge is the scale problem: While the width of thermonuclear combustion fronts is only on the order of millimeters to centimeters, a Chandrasekhar-mass white dwarf star has a radius of about 2,000 kilometers and expands during the explosion. Another challenge arises from the fact that subsonic burning in the form of a thermonuclear deflagration (which resembles chemical flames in many respects) is subjected to instabilities and interacts with turbulent motions. This process accelerates the star's combustion and can give rise to an explosion. Over the past years, the PSO group and its collaborators have developed numerical schemes that meet these challenges.

As part of his doctoral project, Florian Lach from the PSO group performed simulations of thermonuclear explosions in Chandrasekhar-mass white dwarf stars with the goal of exploring the faint and low-energy end of the distribution. A specific class of observed Type Ia supernovae – the so-called Type Iax objects – were suggested to arise from turbulent thermonuclear deflagrations in Chandrasekhar-mass carbon–oxygen white dwarfs. While previous models of the PSO group reproduced the properties of the brighter events of this sub-class, the question to be answered was whether the fainter end of this observational class could be explained with the same theoretical model. To that end, Florian Lach performed simulations of asymmetric explosions that ignite

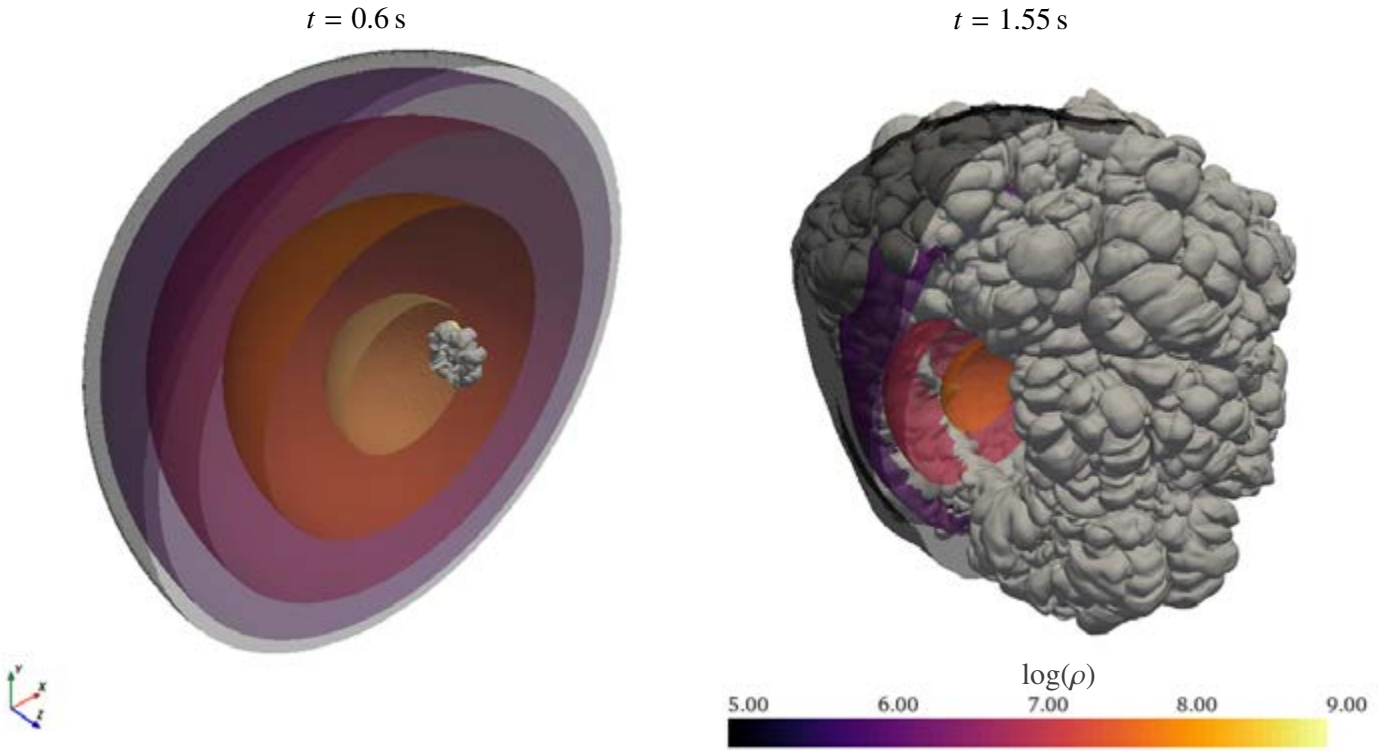


Figure 46: Thermonuclear explosion in a Chandrasekhar-mass white dwarf: The mass density is color-coded. The subsonic deflagration flame (visualized as a gray surface) is ignited off-center and rises toward the surface (left panel) of the star. The burnt material then expands and encloses a bound remnant (right panel). Figure from [Lach et al., 2022b].

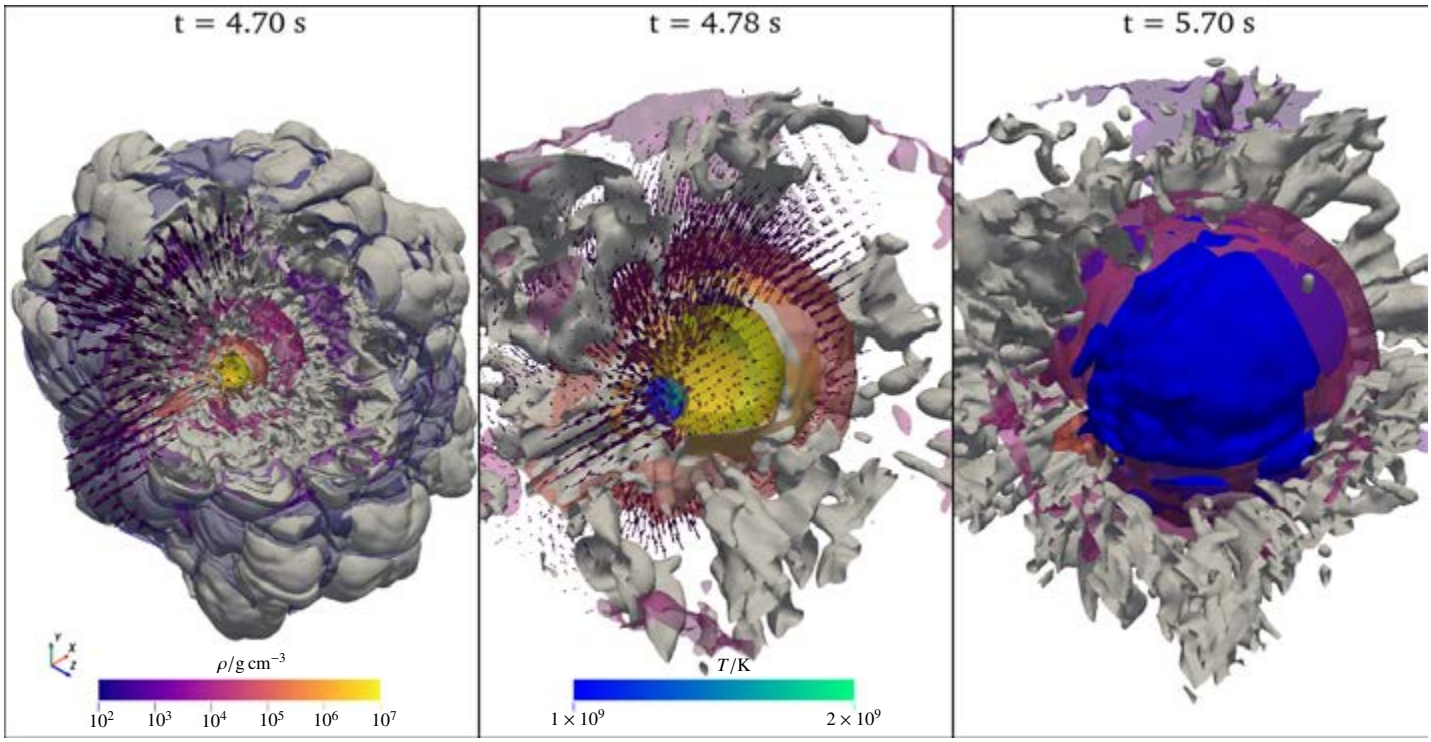


Figure 47: Thermonuclear explosion in a Chandrasekhar-mass white dwarf: The mass density (left panel), temperature (central panel), and combustion mode (right panel) are color-coded. As in Fig.3, the subsonic deflagration flame (visualized as a gray surface) is ignited off-center. A supersonic detonation (shown as a blue surface in the right panel) – which triggers with some delay – consumes the remaining fuel material and leads to a powerful explosion. Figure from [Lach et al, 2022b].

off-center in the white dwarf star (see Figure 46, previous page). These models predict a weak and incomplete explosion of the white dwarf star. A bound remnant is expected to survive the explosion, with only part of the material being ejected. In order to compare this outcome with observations, collaborators at Queen's University in Belfast, UK, performed radiative transfer simulations based on the explosion model of the PSO group in Heidelberg. The predicted observables match well with some of the properties of observed Type Iax supernovae, but some discrepancies persist, particularly at the very faint end of the observed sample.

In a follow-up publication (see [Lach et al., 2022b]), Florian Lach explored the question of how the characteristics of the modeled events would change if a supersonic detonation – which is a combustion model driven by shock waves – were triggered at a later stage of the explosion after initial subsonic deflagration burning (see Figure 47). Such a detonation leads to a much brighter and more powerful explosion that incinerates the entire star and does not leave behind a bound remnant. Again, the collaborators from Queen's University Belfast simulated the radiative transfer in the ejecta. The predicted observables could match another peculiar sub-class of Type Ia supernovae: the so-called 91T-like events.

„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** 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

2.12 Scientific Databases and Visualization (SDBV)

Group leader

PD Dr. Wolfgang Müller

Team

Dr. Haitham Abaza (since July 2022)
Dr. Alain Becam
Dr. Ina Biermayer (since September 2022)
Dr. Dorotea Dudas
Dr. Susan Eckerle (since January 2022)
René Geci (student; since April 2022)
Dr. Sucheta Ghosh
Martin Golebiewski
Anton Hanke (student; until September 2022)
Bettina Heinlein (until July 2022)
Xiaoming Hu
Jan Koß (student)
Dr. Olga Krebs
Jana Krieg (student; until August 2022)
Lukrécia Mertová
Ghadeer Mobasher
Dr. Maja Rey
Maria Paula Schröder (student; since February 2022)
Dr. Natalia Simous (until April 2022)
Fabian Springer (student)
Dr. Andreas Weidemann
Dr. Ulrike Wittig
Yueyang Xie (student; since November 2022)

Over its more than 20 years of existence – first as part of EML, then as part of EML Research, and finally as part of HITS – the SDBV group has not changed its mission much. Indeed, the mission remains to create tools that provide information to scientists.
Each tool needs a data model that can provide different data with a common – and somewhat standardized – structure. But how should biological data be standardized? In addition to developing and running tools, SDBV is part of community and worldwide standards.
Providing data with a consistent structure currently requires human work. But with the vast amount of data out there, how can we enable as few humans as possible to enter as much of this data in as high a quality as possible into our systems? SDBV works on improving the

curation process. The work presented as highlights from 2022 shared the goal of facilitating data extraction and matching compound names.
In addition to the professionally curated and maintained data collection of SABIO-RK, we also run and maintain the FAIRDOMHub and other FAIRDOM SEEK instances with the aim of enabling scientists to contribute to preparing their data to be shared FAIRly. FAIR stands for Findable, Accessible, Interoperable, Reusable and describes the goals that the worldwide community of scientists who manage data should strive for.

Making data FAIR is a challenging task that involves users and data managers. In large projects, data managers should be close to the program management in order to be of greatest use to the project. Therefore, since 2022, HITS has been the host organization of the Program Director of the BMBF LiSyM Cancer network, Beat Müllhaupt, who comes from Zürich. Beat leads the consortium as a contractor and is the head of the program management run by Susan Eckerle and Ina Biermayer at SDBV.

A long path toward smart curation

Below, we outline two articles that provide a glimpse into our work that aims at smart curation. Alongside the DeepCurate project, which has been mentioned and featured in a previous HITS Annual Report, Ghadeer Mobasher – funded by the EU Innovative Training Network PoLiMeR (Grant No. 821616) – and Lukrécia Mertová are currently working on finding compound names in texts as well as on linking these names to ontologies. Ontologies are machine-readable knowledge representations. The future use of this work is twofold: First, the long-term aim is to enable the automatic extraction of useful data from the literature, and second, the short-term aim is to help human curators to curate more quickly and with less effort.

WeLT: Improving biomedical pre-trained language models with cost-sensitive learning

BioNER (biomedical named entity recognition) refers to the recognition of words that are names of things in texts, such as protein names or chemical compound names. BioNER is the basis of information extraction and information retrieval, which are key tasks for navigating what is called the "data deluge," or the overabundance of potentially useful information. A key challenge involves robustness – that is, methods that work with a large variety of texts. Having robust NER models can assist researchers in finding and identifying relevant research, thereby accelerating the process of scientific discovery. In recent years, deep learning has become the main research direction of many natural language processing (NLP) tasks due to the development of effective transformer pre-trained language models (PLMs) [Casola, Silvia et al. Pre-trained transformers: an empirical comparison. Machine Learning with Applications, 2022 9, 100334]. One of the main powerful training techniques that PLMs offer is fine-tuning, which involves using and refining models that are trained for general tasks for use with specific datasets for a particular task.
For the training, NER is broken down into a classification task: A piece must be defined as either uninteresting or the beginning/middle of a named entity. Despite this powerful technique, naïvely fine-tuning a model on targeted datasets without considering the class distribution – that is, the frequency of named entities and their parts – can be problematic, especially when dealing with imbalanced datasets like most of the biomedical gold-standard datasets, as depicted in Figure 48. A strong class imbalance means that there are many more instances of some classes than others. In this context, the classes with the largest proportion of data are called the majority classes, while the smaller proportion examples are the minority classes. Imbalanced classifiers tend to be biased and perfectly predict the majority classes while struggling to classify the minority classes. Thus, they tend to achieve high accuracy, which can be misleading.
For instance, the Linnaeus text corpus consists of 100 full-text documents. Within this corpus, all mentions of species terms were manually annotated and normalized to the NCBI taxonomy IDs [Gerner, Martin et al. LINNAEUS: a species name identification system for biomedical literature, 2010 BMC Bioinformatics, 11(1), 1–17] in order to provide ground truth for training and testing. As shown in Figure 49 (next page), the

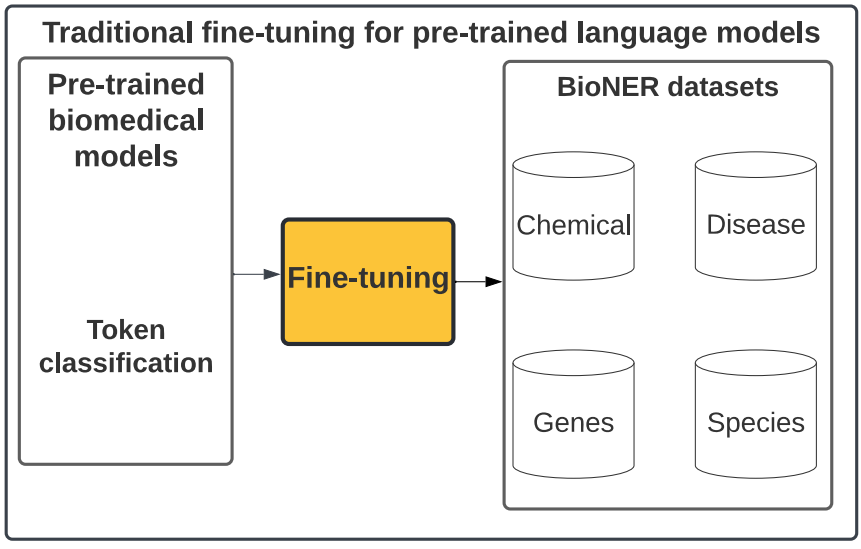


Figure 48: Traditional fine-tuning for pre-trained biomedical models.

corpus is a highly skewed dataset. The inside–outside–beginning (IOB) tagging scheme is commonly used for token classification. This IOB tagging scheme contains three classes: (a) a B-prefix tag denotes the beginning of an entity, (b) an I-prefix tag denotes that the token is inside the chunk of the recognized entity when split into multiple tokens, and (c) an O-prefix tag denotes other tokens that do not belong to an entity [Sang, Erik F. Tjong Kim, & de Meulder, Fien Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition, 2003 Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, 142–147]. In order to address the class imbalance problem, we propose a weighted loss trainer (WeLT) for yielding a class-balanced loss for the task of biomedical NER. We present a cost-sensitive approach for handling the class imbalance before fine-tuning PLMs on targeted datasets, as illustrated in Figure 50. The new re-scaled class weights are based on the WeLT approach, which is applied fairly to the three tags of the IOB tag schemes. The main motive behind WeLT is to introduce new coefficients to the trainer's loss function that give more attention to the minor classes and that penalize the major class. Each class coefficient is calculated using the normalized inverse ratio of this class distribution over the total class distributions of the training datasets, as shown in Figure 51. Our goal is to have a higher predictive performance of WeLT in comparison with the traditional fine-tuned models, which do not address the class imbalance by having higher recall while also maintaining high precision.

We test the impact of WeLT on mixed-domain and domain-specific BioPLMs and compare the results with traditional fine-tuning approaches. In our experiments, we evaluated WeLT on different transformer architectures as follows: (a) BioBERT [Lee, Jinhyuk et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining, 2020 Bioinformatics, 36(4), 1234–1240], (b) PubMedBERT [Gu, YU et al. Domain-specific language model pretraining for biomedical natural language processing, 2021 ACM Transactions on Computing for Healthcare (HEALTH), 3(1), 1–23, 2021], (c) BlueBERT [Peng, Yifan et al. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing, 2019 ACM Transactions on Computing for Healthcare (HEALTH), 3(1), 1-23], (d) SciBERT [Beltagy, Iz et al. SciBERT: A Pretrained Language Model for Scientific Text, 2019 In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3,615–3,620, Hong Kong, China. Association for Computational Linguistics], and (e) BioELECTRA [Kanakarajan, Kamal Raj et al. BioELECTRA: Pretrained Biomedical text Encoder using Discriminators, 2021 n Proceedings of the 20th Workshop on Biomedical Language Processing, pages 143–154, Online. Association for Computational Linguistics]. For a FAIR comparison, we used the same hyper-parameters for both the traditional and the WeLT fine-tuning experiments, and we provide the code for reproducing our experimental results

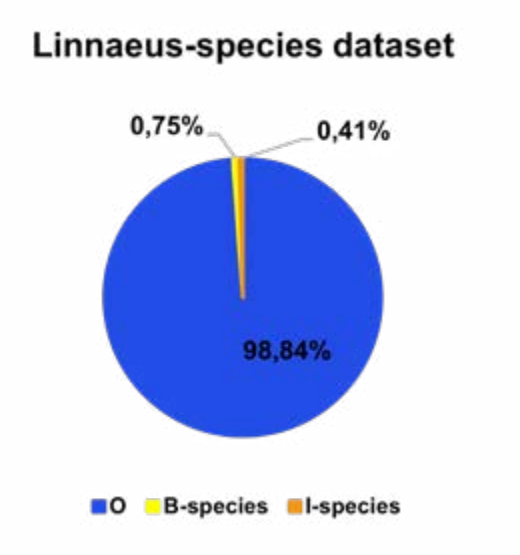


Figure 49: Class distribution percentage for Linnaeus dataset. The majority class is represented by the "O" tags, and the minority classes are represented by the "B" and "I" tags.

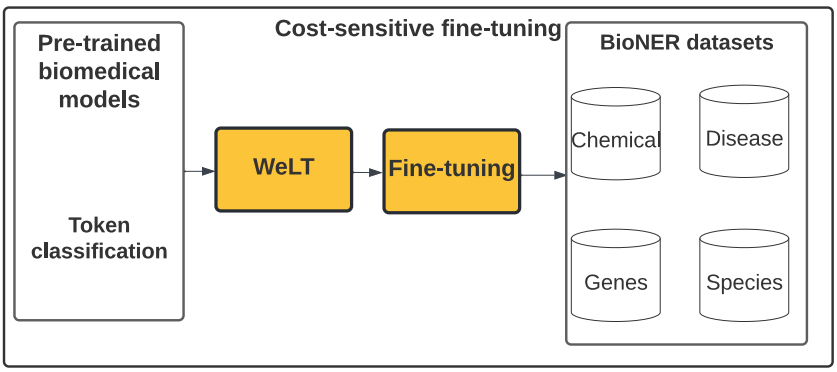


Figure 50: WeLT fine-tuning approach, which addresses the class imbalance of the targeted datasets.

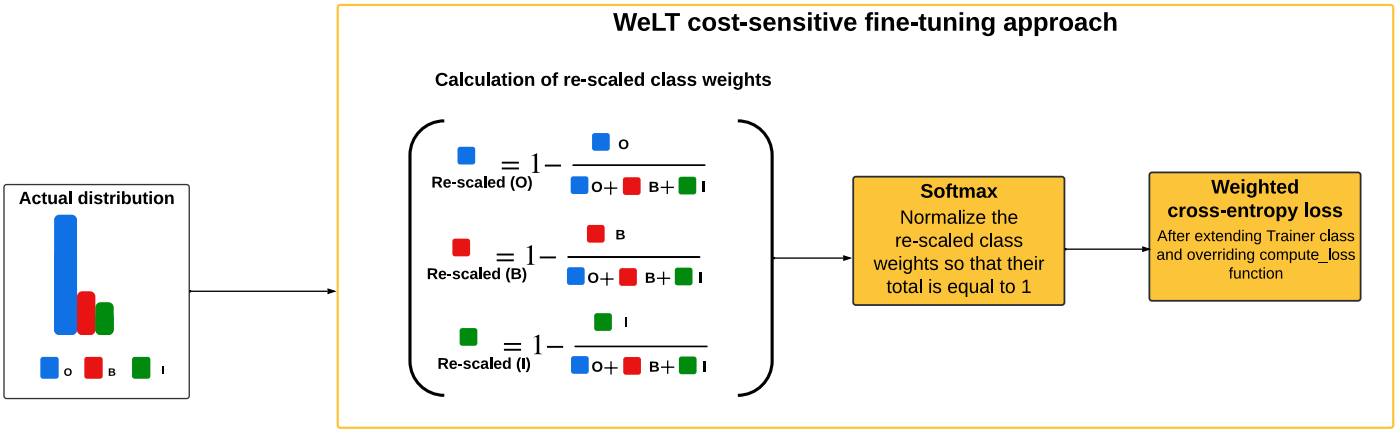


Figure 51: Calculation of re-scaled class weights, which passed to the weighted cross-entropy loss function.

from scratch. At the time of writing this contribution to the Annual Report 2022, the experiments have been completed, and the manuscript is close to submission.

Figure 52 presents the Linnaeus NER results. The models that used the WeLT surpassed their corresponding original trainer.

In our work, we use WeLT, which is a simple yet effective approach that handles class imbalance before fine-tuning BERT models. We present a generic way of calculating the re-scaled class weights, which are passed to the weighted cross-entropy loss function of WeLT. We assessed the performance of WeLT against its corresponding original trainer, which does not handle the class imbalance on BioNER. In so doing, we focused on the biomedical domain and applied WeLT to various BERT and ELECTRA models; however, we hypothesized that our approach could be transferable to other non-biomedical domains using various PLM architectures, such as RoBERTa [Zhuang, Liu et al. A Robustly Optimized BERT Pre-training Approach with Post-training, 2019 In Proceedings of the 20th Chinese National Conference on Computational Linguistics, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China] and T5 [Raffel, Colin et al. Exploring the limits of transfer learning with a unified text-to-text transformer, 2020 The Journal of Machine Learning Research, 21(1), 5485-5551].

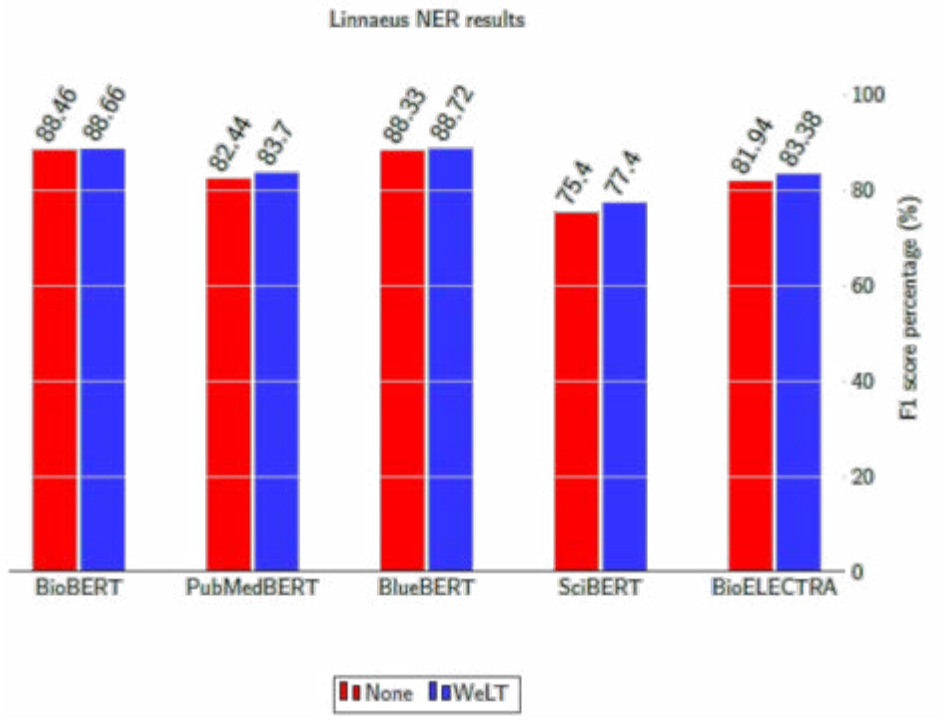


Figure 52: Linnaeus NER F1 scores of WeLT (blue bar), and the traditional scores, which do not address the class imbalance (red bar).

This project is supported by European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement PoLiMeR, No 812616.

Database approximate search and entity linking in the biochemical domain

Entity linking (EL) links textual mentions of named entities to knowledge base entries. Applied in the biochemical domain, EL connects entities in the scientific literature (or models) to chemical databases. Due to the nature of the chemical nomenclature, EL must be specifically tailored for chemical terms that are constructed using established systematic rules mixed with a diverse habit of writing trivial or semi-trivial names. Naturally, this process introduces significant ambiguity.

We propose a proximity search for linking names to the chemical databases based on their chemical and linguistic similarities. We use traditional string similarity functions – such as Levenshtein, trigrams, and Soundex – together with novel chemical similarity functions. Furthermore, we utilize internal tokenization based on chemical semantics, and we combine rule-based and dictionary-based approches (see Figure 53). The initialization of the internal database is the first step in our pipeline. Tokenized chemical names from the source database – which can be an arbitrary chemical database (such as PubChem [Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2023). PubChem 2023 update. Nucleic Acids Res., 51(D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956] or ChEBI [Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. (2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res.]) – are reproduced and stored as a short token in the local database schema. These short parts of the chemical names contain the base structural and functional information in the form of prefixes, suffixes, and name cores. We generate synonyms based on the transformation of the modifiers from each token in the internal database.

The semantic analysis in our solution uses the knowledge and rules of organic chemistry according to IUPAC and allows for the differentiation of each part of the compound names into tokens, stereodescriptors, locants, etc. Thus, we can separate – or tokenize – the individual components of each chemical and assign these components different priorities according to their position and type. This process allows us to design a more precise ranking system and a better tokenization approach compared with typical tokenizers of natural languages.

The search process consists of two steps. The first step – which is faster and less precise – involves using a fuzzy index search for all candidates. The result is a list of candidates that will be examined more closely. The second step – which is fine-tuned and more complex – uses a similarity evaluation function designed for chemical nomenclature, which results in a score value (difference) of two chemical names. Similarity evaluation (or ranking) works with the structural and syntactic properties obtained from a chemical compound in the semantic analysis step and deals with chemical errors and typos. The entities with a minimal distance from the query are our results.

Our approach achieves promising precision while dealing with chemical synonyms, structural similarities, abbreviations, and typos. We address time complexity with an adapted database structure and indexing tailored to the needs of the similarity algorithm.

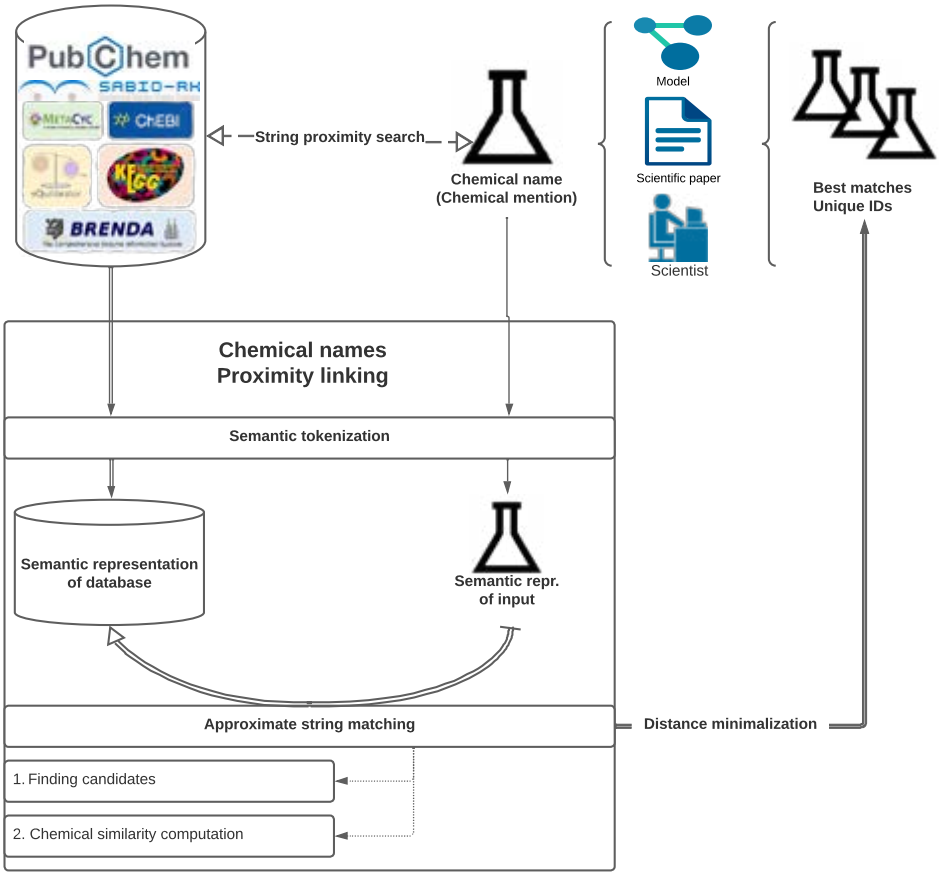


Figure 53: The proximity search scheme for linking names to the chemical databases, based on their chemical and linguistic similarities.

In den mehr als 20 Jahren ihres Bestehens am EML European Media Laboratory, dann am EML Research und schließlich am HITS hat die **Scientific Databases and Visualization (SDBV)** Gruppe ihren Auftrag nicht wesentlich verändert. Die Aufgabe besteht darin, Werkzeuge zu entwickeln, die Forschenden Informationen liefern.

Jedes Werkzeug braucht ein Datenmodell, das unterschiedliche Daten in eine gemeinsame, ja sogar in einem gewissen Sinne standardisierte Struktur bringt. Wie sollte man biologische Daten standardisieren? Die Gruppe entwickelt und betreibt nicht nur Werkzeuge, sondern ist auch Teil der Community und der weltweiten Standards.

Um Daten in eine einheitliche Struktur zu bringen, bedarf es derzeit menschlicher Arbeit. Es gibt so viele Daten da draußen, wie können wir es einigen wenigen Menschen ermöglichen, möglichst viele dieser Daten in hoher Qualität in unsere Systeme einzugeben? Die SDBV-Gruppe arbeitet an der Verbesserung des Kuratierungsprozesses. Die in diesem Jahr als Highlights vorgestellten Arbeiten zielen darauf ab, die Datenextraktion und den Abgleich der Namen von Verbindungen zu erleichtern.

Neben der professionell kuratierten und gepflegten Datensammlung von SABIO-RK betreiben und pflegen wir den FAIRDOMHub und andere FAIRDOM SEEK-Instanzen. Ziel ist es, Wissenschaftler*innen in die Lage zu versetzen, ihre Daten so aufzubereiten, dass sie FAIR geteilt werden können. FAIR steht für Findable, Accessible, Interoperable, Reusable (auffindbar, zugänglich, interoperabel, wiederverwendbar) und beschreibt, welche Ziele eine weltweite Gemeinschaft von Wissenschaftlern, die Daten verwalten, anstreben sollte.

Daten FAIR zu machen ist eine anspruchsvolle Aufgabe, an der Nutzer*innen und Datenmanager*innen beteiligt sind. Bei großen Projekten sollten die Datenverwalter*innen mit der Programmleitung eng zusammenarbeiten, um für das Projekt von größtem Nutzen zu sein. Deshalb fungiert das HITS seit 2022 als Gastorganisation des Programmdirektors des BMBF LiSyM Cancer-Netzwerks, Beat Müllhaupt aus Zürich. Der Medizinprofessor leitet das Konsortium als Auftragnehmer und ist Leiter des Programmmanagements, das von den SDBV-Mitarbeiterinnen Susan Eckerle und Ina Biermayer geführt wird.

2 Research
2.13 Stellar Evolution Theory (SET)

Group leader
Dr. Fabian Schneider

Team
Vincent Bronner
Jan Henneco
Dr. Rajika Kuruwita (HITS Independent Postdoc, since October 2022)
Dr. Eva Laplace
Tina Neumann (student, until May 2022)
Julian Leon Saling (student))
Duresa Temaj
Dr. Dandan Wei (visiting scientist; CSC China – DAAD)

Stars are the basic building blocks of the visible Universe and produce almost all chemical elements that are heavier than helium. Understanding how stars have transformed the pristine 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 enormous feedback they provide, massive stars helped to re-illuminate the Universe after the Cosmic Dark Ages. Moreover, they drive the evolution of galaxies and lay the foundation for life as we know it.

At the end of their lifespan, 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 to us on Earth. Mergers of neutron stars and black holes are now routinely observed thanks to gravitational-wave observatories, thereby opening a new window into the Universe.

Today, we know that most massive stars are born in binary and higher-order multiples, including triples, quadruples, and so on. This fact 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 onto their companion. In about 25% of massive stars, this mass-transfer phase is unstable and leads to a merger of both binary components. Mass-exchange episodes and the even-more-drastic merger events profoundly change both 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 rather than 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, on the question of which stars form black holes, and on the intricate merging process of stars. Mergers produce strong magnetic fields, and the products of these mergers may forge highly magnetized neutron stars in their terminal supernova explosions. These magnetic neutron stars – known as magnetars – are the strongest known magnets in the Universe.

Thanks to an ERC Starting Grant, the Stellar Evolution Theory (SET) group was founded at HITS in January 2021. The group follows three main research themes that are centered around the evolution of binary star systems. First, we study (1) which binary systems are expected to lead to the coalescence of both stars, (2) exactly what happens in such stellar mergers (e.g., how magnetic fields are amplified), and (3) how merged stars can be identified in observational campaigns. Second, we aim to understand (4) which stars form black holes and (5) which stars explode as supernovae. Given that most massive stars are found in binary-star- or other multiple-star systems, mass exchange between individual components can significantly affect the evolution and ultimate fate of these stars (see below). This topic is particularly relevant in the era of gravitational-wave astronomy, during which, mergers of neutron stars and black holes are routinely observed. Third, giant stars can swallow companions, thereby initializing a phase of so-called common-envelope evolution. While the PSO group (see Chapter 2.11) conducts three-dimensional (magneto-)hydrodynamic simulations of this phase, we work on one-dimensional hydrodynamic models that are calibrated against these computationally much more expensive models. Moreover, the evolution that immediately follows a common-envelope event is of great interest when it comes to better understanding the many merging neutron star and black hole binaries that are observed via their gravitational-wave emissions. For example, we found that circumbinary disks that are left over after the common-envelope phase can severely shrink the binary orbit of a star and thereby affect the further evolution toward gravitational-wave merger events. In 2023, we will add a fourth research topic: data-driven and machine-learning-powered stellar astrophysics. In this area, we will take advantage of our involvement in the ongoing HITS Lab emulator project.
In 2022, our bachelor's student Tina Neumann finished her research project and went to Leiden University, where she is now working on a Master of Science degree in astrophysics. Moreover, both of our master's students successfully completed their projects: Duresa Temaj is currently writing a paper about her findings while staying a few more months at HITS, and Vincent Bronner recently began working as a PhD student in the group. In October, the first HITS independent postdoc, Rajika Kuruwita, began her work and formally joined the SET group (see also chapter 2.15). We are very happy to welcome Rajika to HITS and look forward to fruitful discussions, new research ideas, and cross-disciplinary collaborations with the many groups at HITS.

Finding the first X-ray-quiet, stellar-mass black hole outside the Milky Way Galaxy

As their name suggests, black holes are dark and cannot be easily seen. The first object commonly accepted as a black hole was Cygnus X-1, which is one of the strongest X-ray sources detectable from Earth. Cygnus X-1 was the subject of a famous bet in 1974 between physicists Stephen Hawking and Kip Thorne (with the latter having won the 2017 Physics Nobel Prize for the observation of gravitational waves from merging black holes). Hawking bet that there was no black hole in Cygnus X-1. His hope was that if he were proven wrong and his past years of research had thus been shown to have been wasted, he would at least win a subscription to the magazine Private Eye. Later, Hawking conceded his bet, and today, we know that Cygnus X-1 harbors the most massive stellar-mass black hole known in the Milky Way, with a mass of about 21 times that of our Sun. Its enormous X-ray radiation stems from matter that is accreted from the black hole's massive companion star. The accreted matter falls onto the black hole and forms an accretion disk that heats up so much that X-rays are emitted.

In addition to such X-ray-bright black holes, we have also predicted the existence of many more X-ray-quiet black holes (Langer N et al: Properties of OB star–black hole systems derived from detailed binary evolution models, A&A 638, A39 (2020)). These black holes are predicted to be in wider orbits around companion stars, where they cannot accrete so much mass and hence do not emit detectable X-rays. Ultimately, a small percentage of massive stars may have an unseen black-hole companion. But how should we go about finding them?
One possibility is to conduct large spectroscopic surveys. Each star emits a characteristic spectrum that is comprised of atomic emission- and absorption lines. If two stars orbit each other, both of their spectra shift periodically due to the Doppler effect. While one star recedes away from us, the other approaches us. This causes the two spectra to move relative to each other such that – with enough observations – the spectra can be disentangled, and it becomes possible to determine the characteristic set of spectral lines of each star. If one of the orbiting stars is dark and does not emit light, then the periodically moving spectrum of only one star is available.
After analyzing data from six years of observations with the European Southern Observatory's (ESO) Very Large Telescope (VLT), we finally found the invisible needle in the stellar haystack of the Tarantula Nebula in our neighboring Galaxy the Large Magellanic Cloud [Shenar et al., 2022]. Given the orbital velocities of the stars that were measured using the Doppler-shifted spectrum of the binary star system, it was concluded that the visible star of 25 solar masses must orbit around a hidden object with at least 9 times the mass of our Sun. Thanks to sufficient spectra taken across the 10.4-day orbit, the composite spectrum of the source could be disentangled to reveal the contribution of both stars (Figure 54, next page). Surprisingly, the 9-solar-mass companion does not emit any appreciable light, and its spectrum is consistent with being constant across wavelengths.

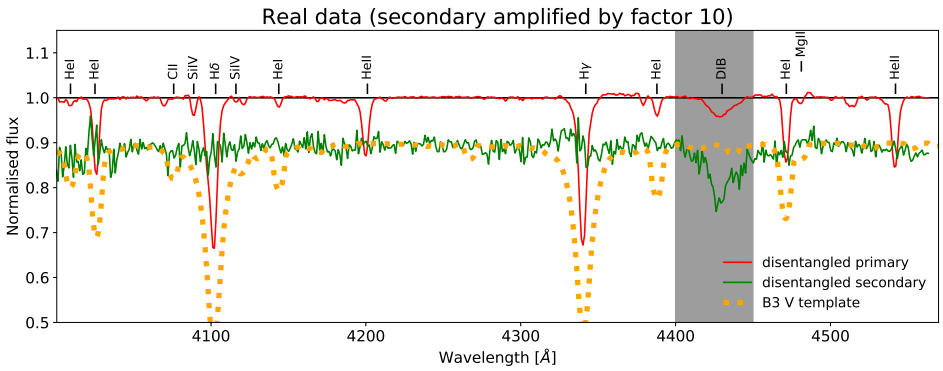


Figure 54: Disentangled spectra of the binary star VFTS 243. The atomic absorption lines in the spectrum of the visible primary component in VFTS 243 (red) could be separated from the spectrum of the secondary companion (green; amplified by a factor of 10). For comparison, an example spectrum of a B3 V star is also shown. Except for one absorption feature (DIB) that originated from our own Earth's atmosphere, the secondary component in VFTS 243 does not show any absorption- or emission lines. The wavelengths in the figure are in units of angstroms (1 Å=10-10 m).

In a hare-and-hound exercise, we produced composite spectra and mock observations in order to determine what kind of 9-solar-mass object could be hiding in VFTS 243. Mock observations made by members of the team in the US were sent to Europe for analysis without the European team knowing what its US counterparts had produced. As companions, we tried lower-mass main-sequence dwarfs, helium stars, and even another binary star system. In all cases, we were able to correctly identify the hidden companions. Having rejected all these scenarios, we were left with the conclusion that the hidden companion must be a stellar-mass black hole of at least 9 solar masses (Figure 55). When massive stars explode as supernovae, their cores collapse and can forge neutron stars – a form of ultracompact matter. In most cases, these neutron stars receive a kick of several hundreds of kilometers per second and are shot away from their explosion site into interstellar space. The black hole in VFTS 243 did not receive such a kick, which suggests that its progenitor star directly collapsed into a black hole without signs of a strong supernova explosion. This finding should also help in understanding the formation histories of the many black-hole mergers that are now observed via gravitational-wave astronomy.

Which stars form black holes?

In the classic picture of stellar evolution, the formation of black holes is linked solely to the mass of their progenitor stars. Moreover, as the mass of a star increases, its fate changes. Low-mass stars do not reach the necessary conditions to fuse carbon in their cores and thus form white dwarfs at the end of their lives. Higher-mass stars can fuse carbon and heavier elements in their cores until they produce iron. No energy can be extracted from nuclear fusion of the iron core, and the core eventually collapses, thereby triggering a supernova explosion that leaves behind a neutron star. For stars with the largest masses, the collapse of the core cannot be stopped, which leads to the formation of a black hole. However, detailed models of the ultimate fate of stars have revealed a more complex picture. Some massive stars explode, while some intermediate-mass stars form black holes. But what determines the fate of these massive stars, if not their total mass? Moreover, exactly which stars form black holes?

In order to answer these questions, we simulated the evolution of massive stars of between 13 and 70 solar masses and applied a supernova model to predict which stars will form black holes and which stars will explode. We found that the stars that will form black holes have more compact cores than the stars that will explode, which confirmed the findings of other groups. We further investigated the origin of this "compactness" and found that it mainly depends on the amount of carbon left after core helium burning as well as on the core mass, which determines the temperature and density conditions in the core. Depending on these properties, the core can transition from being dominated by the energy generated by nuclear fusion to losing energy due to the production of neutrinos that escape from the star. During these transitions, the core contracts in order to counter the energy loss and becomes so compact that the star inevitably forms a black hole. With these findings, we can predict the "death landscape" of stars – which is depicted in Figure 56 – and compare it with observations.

Observationally determining which stars will create black holes is difficult. While supernova explosions are extremely bright cosmic signposts that point to the location of a star that has reached the end of its life, stars that form black holes do so more quietly. In most cases, these stars are expected to simply vanish as they collapse, potentially ejecting small amounts of material in the process. In order to catch a star that will form a black hole, it is necessary to repeatedly monitor large numbers of stars in an effort to find those that disappear. This task is complicated by the extreme variability in the brightness of stars. To date, a single credible disappearing star has been found, and it has not yet reappeared (see Figure 57). Interestingly, comparing the brightness and temperature of this star with our models (i.e., the gray

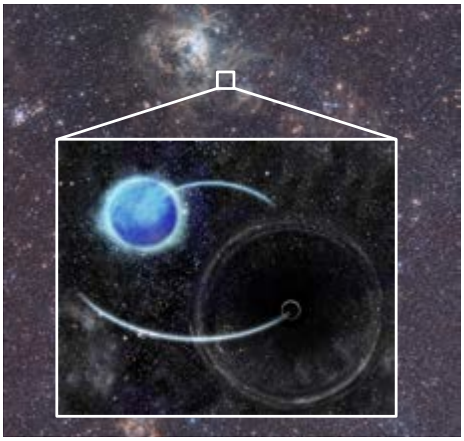


Figure 55: Artist's impression of the VFTS 243 binary star. The 25-solar-mass primary star is in a 10.4-day orbit with a >9-solar-mass black hole. The background is an image of the Tarantula Nebula in the Large Magellanic Cloud that was produced by the Visible and Infrared Survey Telescope for Astronomy (VISTA). Background image credit: ESO/M.-R. Cioni/VISTA Magellanic Cloud survey. Acknowledgement: Cambridge Astronomical Survey Unit. Visualization credit: Isca Mayo / Sara Pinilla.

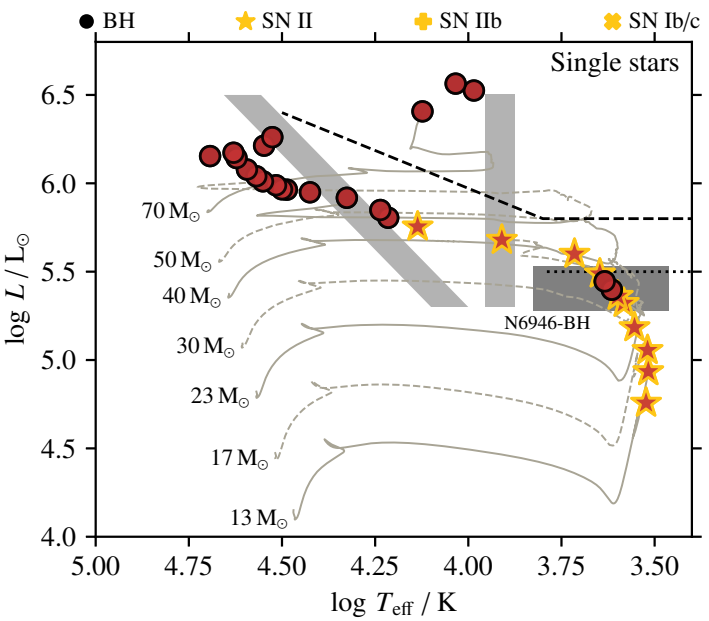


Figure 56: Death landscape of massive stars in an effective temperature-vs.-brightness diagram. Black circles indicate stars that will produce black holes, while star symbols indicate stars that will explode and leave behind a neutron star.

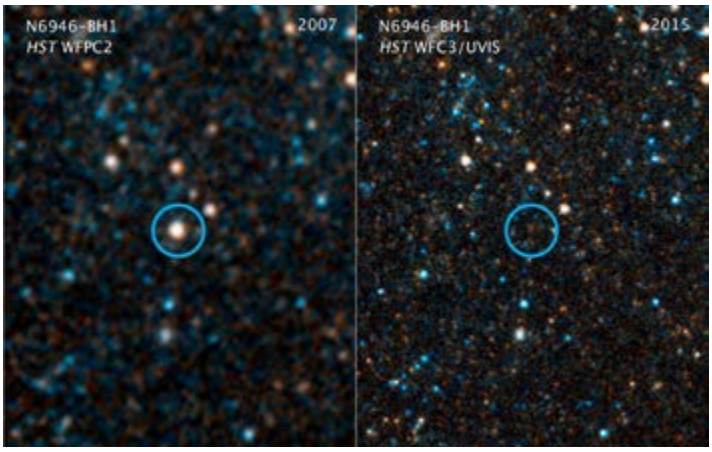


Figure 57: Giant star N9646-BH observed in 2007 with NASA's Hubble Space Telescope (left) that went missing in 2009 after a transient phase. Observations of the same location in 2015 (right) show that the star indeed disappeared without exploding in a supernova. It most likely collapsed into a BH. Image credit: NASA/ESA/C. Kochanek (OSU).

box labelled "N6946-BH" in Figure 56), we can see that it lies precisely at the location where our models predict the formation of black holes.

Our models are affected by several uncertainties. For example, it is unclear how interior mixing affects the compactness of stellar cores and therefore also their ability to explode or to instead form black holes. In her master's thesis, Duresa Temaj addressed this question by computing and analyzing models of stars with different degrees of core boundary mixing and determining the ultimate fate of these stars using our supernova model. The results indicate that enhanced mixing increases the core masses of stars and changes their compactness. However, this finding does not change the observed luminosity of black hole progenitors because they all have a similar core mass, which determines the ultimate luminosity of these stars. That means that despite uncertainties in the interior mixing of stars, we are able to reliably predict the properties of black hole progenitors.

Predicting the formation of black holes has several implications. In particular, we can use our findings to infer the mass distribution of binary black-hole mergers, which can now be detected using gravitational-wave astronomy.

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.

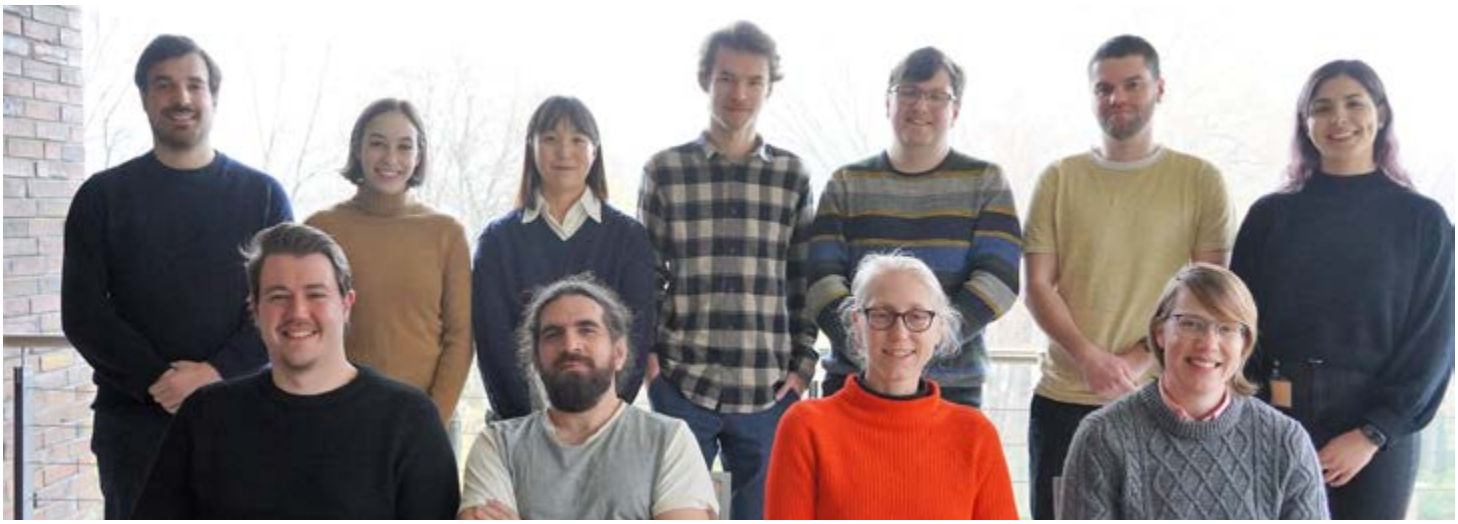
Am Ende ihres Lebens hinterlassen massereiche Sterne einige der exotischsten Formen von Materie: Neutronensterne und Schwarze Löcher. Die Untersuchung dieser Überbleibsel ermöglicht Einblicke in Materieformen, die so auf der Erde nicht verfügbar sind. Die Verschmelzungen von Neutronensternen und Schwarzen 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 (SET)** Gruppe untersucht das turbulente und explosive Leben massereicher Sterne. Derzeit konzentriert sich die Gruppe auf massereiche Doppelsternsysteme, deren Verschmelzungsprozesse und die Frage, welche Sterne als schwarze Löcher enden. Sternverschmelzungen erzeugen starke Magnetfelder und können zu stark magnetisierten Neutronensternen führen. Diese als Magnetare bekannten magnetischen Neutronensterne sind die stärksten Magnete im Universum.

2 Research

2.14 Theory and Observations of Stars (TOS)



Group leader

Prof. Dr. Ir. Saskia Hekker

Team

Dr. Felix Ahlborn (since October 2022)
Prof. Dr. Sarbani Basu (Klaus Tschira Guest Professor; from September to November 2022)
Dr. Michaël Bazot (since February 2022)
Beatriz Bordadagua (since October 2022)
Teresa Andrea Maria Braun (student; until June 2022)
Lynn Buchele
Jeong Yun Choi (since November 2022)
Quentin Coppèe
Francisca Espinoza (since October 2022)
Jan Henneco (partly in TOS and partly in SET)
Daria Mokrytska (until April 2022)
Jonas Müller (since November 2022)
Dr. Anthony Noll (since February 2022)
Alba Covelo Paz (student; until September 2022)
Julian Schlecker (student; until mid-February 2022)
Dr. Nathalie Themessl (visiting scientist; Heidelberg University)

Stars are an important source of electromagnetic radiation in the Universe and allow for the study of many phenomena ranging from distant galaxies to the interstellar medium and extra-solar planets. However, due to their opacity, Sir Arthur Eddington once stated 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" (1926). Now, through modern mathematical techniques and high-quality data, it has become possible to directly probe and study the internal structure of stars through global stellar oscillations – a method known as asteroseismology.

Asteroseismology uses similar techniques to helioseismology (as carried out on our closest star, the Sun) to study the structure of other stars. The properties of waves are used to trace internal stellar conditions. Oscillations (waves) that propagate through the whole star reveal information that is otherwise hidden by the star's opaque surface. This asteroseismic information collected by the CoRoT, Kepler, K2, TESS, SONG, and Plato space observatories – combined with interferometry, photometry, astrometric observations from Gaia, spectroscopic data from the SDSS-V APOGEE, 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 these physical processes and how they change as a function of stellar evolution is the ultimate goal of the Theory and Observations of Stars (TOS) group. In this group, 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 also potential hosts of planets and serve as standard candles (e.g., core-helium-burning red giant stars) for galactic studies. Exoplanet studies as well as galactic archaeology can hence benefit from an increased understanding of these stars.

Background

In the TOS group, we focus on stars with oscillations similar to those present in our Sun. These so-called solar-like oscillations are low-amplitude oscillations that are stochastically excited through turbulence in the near-surface convection layer of a star. The oscillations are sound waves that are expected to be present in all stars with convective outer layers. A convective envelope is typically present in low-mass main-sequence stars, subgiants, and red giant stars 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, by its radial order (i.e., the number of nodal lines in the radial direction), its spherical degree (i.e., the number of nodal lines on the surface), and its azimuthal order (i.e., the number of nodal lines that cross the spin axis).

In evolved, so-called red giant stars, the dipole modes (i.e., those with a spherical degree of 1) are sensitive to both the deep interior and the outer layers of the stars. In other words, the oscillations resonate in an inner (gravity) and an outer (acoustic) cavity, which are separated by an evanescent zone (i.e., the area between the cavities, where oscillations cannot propagate and thus decay exponentially). From the resulting mixed pressure–gravity oscillations, the coupling between the two oscillating cavities and the phases of the waves in each cavity can be derived and 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 also 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.

Scientific highlights

Subdwarf B progenitors?

Within the framework of the ERC consolidator grant "DipolarSound," in 2022, we began to investigate "messy" stars (see Figure 58 for the power-density spectrum of a "messy" star). As part of our work, we focus on a set of stars for which the period spacings can still be determined but are smaller than expected (see Figure 59). This new class of stars will be announced in a publication that has been submitted to MNRAS (Elsworth, Braun & Hekker "Low-period spacing core-helium burning giants I: A new class of star?", submitted to MRNAS, 24 November 2022)

We subsequently investigated the potential cause of these small period spacings. We computed models with the MESA stellar evolution code (Paxton et al.: Modules for Experiments in Stellar Astrophysics (MESA): Binaries, Pulsations, and Explosions; ApJS, 220, 15 (2015)). If we assume a canonical evolution of stars in the observed mass range, we can find models with these small period spacings either only for very short periods of time in the evolution of the stars – which makes them very unlikely to be observed – or for when the stars are well beyond the age of the Universe. Therefore, we concluded that some mechanisms that are not included in the canonical evolution are at play. We found that a fast mass-loss episode in stars with masses above 2 solar masses alters the evolution of

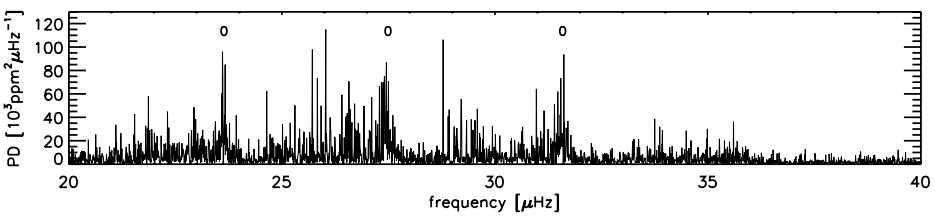


Figure 58: Typical example of a "messy" star. The peaks identified with a "0" are radial modes. For the other peaks, it is much more difficult to identify their degree (degree = the number of nodal lines on the surface).

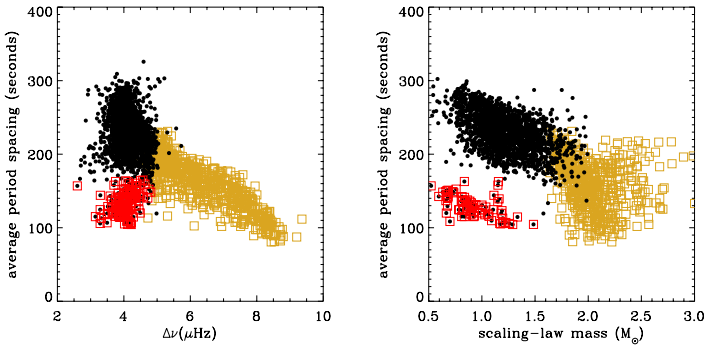


Figure 59: Left: the measured average period spacing as a function of large frequency separation (Δν). Right: the measured average period spacing as a function of the stellar mass computed via scaling relations. The black dots represent stars that have their onset of core-He-burning in a He flash, while the golden squares represent core-He-burning stars with higher masses in which the onset of He-burning is more gradual. The red squares indicate the set of stars that are in a theoretically unexpected place in these diagrams and that we thus investigated. Figure taken from Elsworth, Braun & Hekker "Low-period spacing core-helium burning giants I: A new class of star?" , submitted to MRNAS, 24 November 2022.

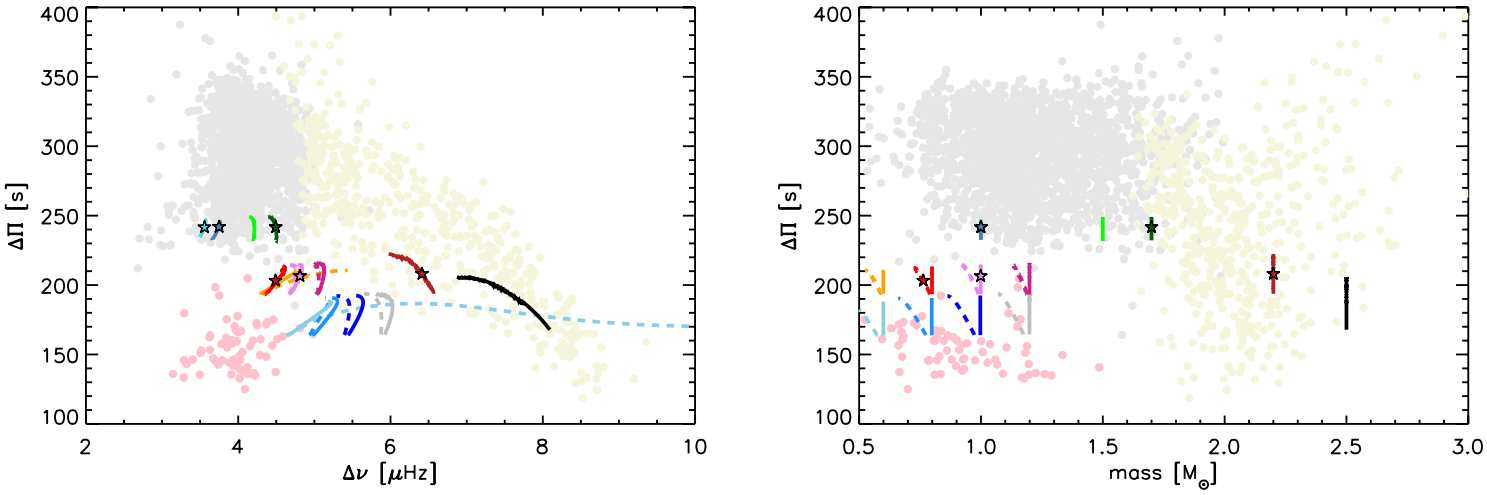


Figure 60: Same as Figure 59, with the observations of Figure 59 now in light colors and the results of the models in dark colors. The cyan, light green, dark green, maroon, and black dots represent results for canonical models of 1.0, 1.5, 1.7, 2.2, and 2.5 solar masses, respectively. The dark blue models represent the results of mass loss for the models with an initial mass of 1.7 solar masses and an end mass of 1.0 solar masses. The magenta, pink, bright red, and orange models all have an initial mass of 2.2 solar masses and an end mass of 1.2, 1.0, 0.8, and 0.6 solar masses, respectively. The gray, royal blue, blue, and light blue models all have an initial mass of 2.5 solar masses and an end mass of 1.2, 1.0, 0.8, and 0.6 solar masses, respectively. The models that continue to lose mass are indicated by the dashed lines. Note that the vertical offset between the observed data and models is a known deficiency of the models. Figure taken from Hekker, Elsworth, Braun & Basu "Low-period spacing core-helium burning giants II: subdwarf progenitors?", submitted to MRNAS, 24 November 2022.

the models such that we can qualitatively reproduce the observations (see Figure 60). Essentially, the core of the star does not have enough time to adjust to the lost mass and acts as the core of a star with the initial mass, while the outer layers from which the mass has been lost act as the less massive stars. This places the stars in the thus-far unexplored part of the period spacing vs. mass space.

To put this result into context, one of the scenarios in which Subdwarf B stars are formed is essentially a fast mass-loss event that occurs shortly before the onset of the core-He-burning phase, in which the envelope is essentially stripped. Although this mass-loss scenario is implemented in an ad hoc manner in our work, it is based on prior knowledge. Therefore, we also investigated whether these stars could potentially evolve into Subdwarf B stars if we evolved the models further. Indeed, if the stars continue to lose some mass during the core-He-burning phase, the stellar models increase their surface temperature, while the luminosity remains relatively low. These stars may thus be regarded as Subdwarf B progenitors.

A view into Jupiter's interior

Gaseous giant planets make up a significant fraction of all confirmed and candidate exoplanets. In our Solar System, these gaseous giants are represented by Jupiter and Saturn. Understanding their formation and evolution is critical to obtaining a clear picture of planetary systems. Furthermore, theoretical models of stars and giant planets share some common principles and have structural similarities. Natural bridges therefore exist between the stellar and planetary sciences, especially in the context of Plato, a forthcoming ESA mission.

A core team composed of M. Bazot (HITS), T. Guillot (Observatoire de la Côte d'Azur), and Y. Miguel (Leiden Observatory) undertook the task of porting computational methods in Bayesian statistics that had already been developed for Sun-like stars to the case of gaseous planetary giants. This was done within the framework of the NASA mission Juno. We applied the aforementioned methods in order to constrain the interior of Jupiter using Juno's extremely precise measurements of the planet's gravitational field.

We first investigated the possibility that the distribution of metals in the interior of Jupiter could be inhomogeneous, and we obtained a positive answer [Miguel, Bazot, Guillot et al. 2022]. This result pertains to the issue of Jupiter's formation, for which a long-standing problem involves understanding how the protoplanet captured metals while accreting gas from the protoplanetary disk. Metals can be captured via the accretion of planetesimals, kilometer-long solid bodies, or smaller pebbles. Planetesimal accretion occurs over longer time scales and leads to inhomogeneous metal distribution in the planetary interior, whereas pebble accretion is

shorter-lived and produces homogeneous interiors. Our findings strongly hint at the likelihood of the planetesimal scenario. This process is believed to also be at play in the formation of exoplanets, which broadens the relevance of our study.

In 2022, we further studied the impact of (1) changes in the equation of state used to model Jupiter and (2) assumptions made regarding the compactness of Jupiter's core (Howard, Guillot, Bazot et al., submitted). We concluded that the formulation of the equation of state has a strong impact on our final estimates. Our best models display dilute cores, which are structures in which the innermost metallicity is slowly transported into the surrounding envelope.

Finally, we used our approach to estimate masses and metallicities for 37 giant exoplanets (Bloot, Miguel, Bazot, & Howard, in preparation). In so doing, we revealed that it is possible to obtain good constraints on these quantities by using the limited measurements available for exoplanets. Our results are in agreement with those of previous studies on giant planets in the Solar System.

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 wurde jedoch 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“ (Sir Arthur Eddington, 1926). 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 eingerichtet wurde, 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 Galaxien-Archäologie vom wachsenden Verständnis dieser Sterne profitieren.

2 Research

2.15 HITS Independent Postdoc Research

HITS Independent Postdoc

Dr. Rajika Kuruwita (since October 2022)

The HITS Independent Postdoc Program offers a great opportunity for highly talented young scientists wanting to transition from PhD student to junior group leader. It supports young scientists in exploring their own ideas and testing new hypotheses. High-risk, high-gain projects are encouraged. Selected postdocs will collaborate with group leaders at HITS while developing and pursuing their independent research projects (see chapter 6.2).
The Fellowship is awarded for two years, with an option for a one-year extension after positive evaluation. It offers a vibrant research community and a highly interdisciplinary and international working environment, with close links to the HITS shareholders Heidelberg University and the Karlsruhe Institute of Technology (KIT). In addition, successful candidates benefit from outstanding computing resources and various courses offered at HITS.
The astrophysicist Rajika Kuruwita is the first HITS Independent Postdoc. Born in Sri Lanka, she completed her PhD at the Australian National University, was a fellow at the University of Copenhagen, and joined HITS in September 2022. She is collaborating closely with the SET group (see chapter 2.13).

The formation of multiple star systems

Rajika Kuruwita's work focuses on star formation, particularly the formation of binary and multiple stars and the possible implications that these stars have for planet formation. This is an important question to investigate because observations have revealed that most stars are born with at least one companion. This finding is shown in Figure 61, which reveals that in the very earliest stages of star formation (i.e., the Class 0 stage), the multiplicity frequency (i.e., the fraction of stars with a companion) is approximately

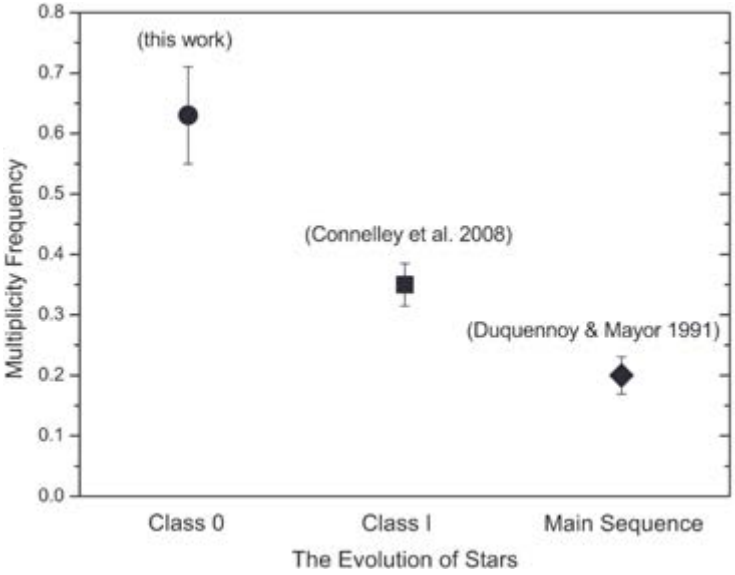


Figure 61: Multiplicity frequency as protostellar evolutionary classes and hydrogen-burning stars on the main sequence - a specific phase of a star's evolution, namely when stars fuse hydrogen into helium in their cores. (Credit: Chen et al. 2013).

two-thirds. Many of the stars in these young multiple-star systems interact, and some systems may disintegrate, all while the stars host circumstellar discs that will be the site of future planet formation.

In the age of the Kepler Space Telescope and the more recent Transiting Exoplanet Survey Satellite (TESS), we have discovered approximately 5,000 exoplanets (Figure 62). Around 700 of these exoplanets have more than one sun, and 30 are in circumbinary orbits. Many stars that end up as single stars likely were born with a companion or experienced stellar interactions during their formation that may have affected the resulting stellar systems.
At HITS, Rajika continues to investigate binary and multiple star formation using magnetohydrodynamical simulations. She is currently investigating whether the binary star formation pathway inherently produces rapidly rotating stars, which would have implications for the lifetime and evolution of stars. Rajika is also examining the origins of FU Orionis-type outbursts, which have very sharp increases in accretion rate that are then maintained for decades. These outbursts can significantly affect the disc structure and chemistry of the places where planets form. While her work focuses on star formation, it may also yield insights into the formation of progenitor systems that undergo interactions later in their lifetimes.

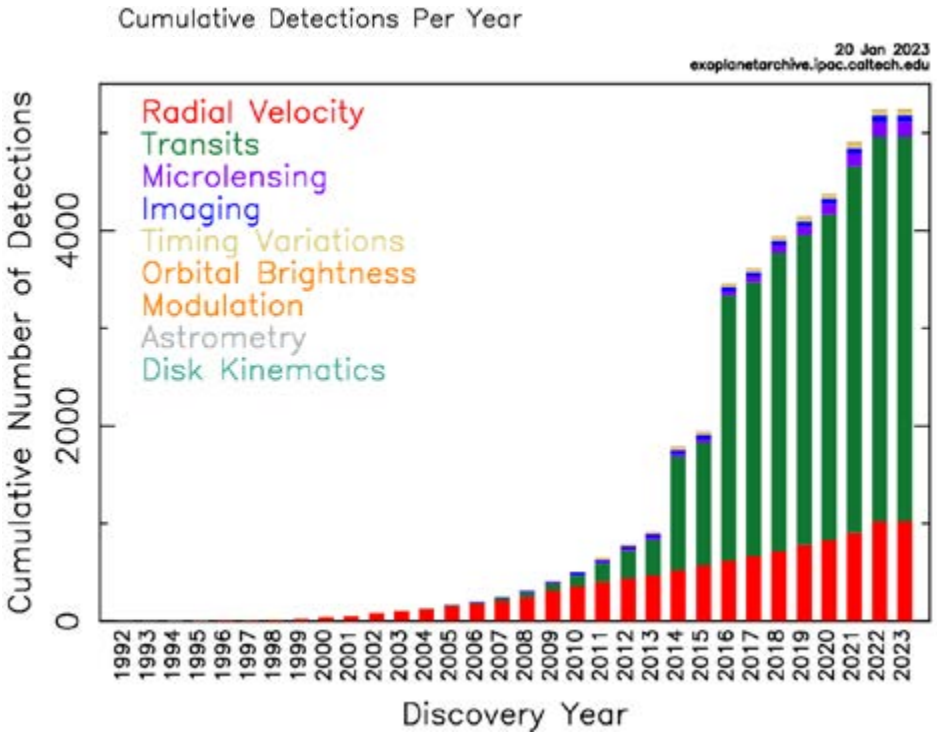


Figure 62: Cumulative number of detected exoplanets over time. The different detection methods are highlighted. Credit: NASA Exoplanet Archive.

Das **HITS Independent Postdoc Programm** bietet Doktorandinnen und Doktoranden eine großartige Chance beim Übergang zum Nachwuchsgruppenleiter bzw. zur Nachwuchsgruppenleiterin. Es unterstützt junge Wissenschaftler*innen dabei, eigene ambitionierte Ideen zu erforschen und neue Hypothesen zu testen. Projekte mit hohem Risiko und hohem Gewinn sind willkommen. Die ausgewählten Postdocs arbeiten mit Gruppenleiterinnen und Gruppenleitern am HITS zusammen, während sie ihre unabhängigen Forschungsprojekte entwickeln und verfolgen.

Das Programm bietet einen Anstellungsvertrag für zwei Jahre, mit der Option auf eine einjährige Verlängerung nach positiver Evaluation. Es bietet eine lebendige Forschungsgemeinschaft und ein stark interdisziplinäres und internationales Arbeitsumfeld mit engen Verbindungen zur Universität Heidelberg und dem Karlsruher Institut für Technologie (KIT). Darüber hinaus profitieren erfolgreiche Kandidat*innen von den herausragenden IT-Ressourcen und wissenschaftlichen Seminar- und Lehrangeboten am HITS.

Die Astrophysikerin Rajika Kuruwita ist die erste Wissenschaftlerin im „HITS Independent Postdoc“ Programm. Sie stammt aus Sri Lanka, promovierte an der Australian National University und war danach Fellow an der Universität Kopenhagen. Sie kam im September 2022 ans HITS und arbeitet eng mit der SET-Gruppe (siehe Kapitel 2.13) zusammen.

3 Centralized Services

Group leader
Dr. Gesa Schönberger

Team
Yashasvini Balachandra
Christina Blach
Frauke Bley
Christina Bölk-Krosta
Benedicta Frech
Silvia Galbusera
Ingrid Kräling
Dr. Barbara Port
Thomas Rasem
Rebekka Riehl
Irina Zaichenko

3.1 Administrative Services

The HITS administration serves the Institute's groups in almost all necessary administrative processes. It takes care of HR support, operates offices and buildings, makes purchases, and settles invoices in addition to supporting the Communications team in organizing events. Moreover, the administration ensures that legal issues are resolved and that all processes at the Institute comply with legal requirements.

After 2.5 years under the working conditions of the pandemic, we finally (!) got back to a sense of normalcy over the course of 2022. All of the hurdles we had been forced to go through in terms of gaining access to enclosed spaces and showing documentation under 3G rules; the time-consuming checking of the constantly changing specifications made by the city, the state, and the federal government; the permanent need to adapt to and pass on in-house hygiene rules – everything had been virtually eliminated by the end of the year.

The research groups cautiously resumed their travel activities and began to meet again in person at conferences and workshops. Offices began to be allowed to be used without restrictions, which everyone was more than happy to take advantage of. Even an open-house day for the residents of Heidelberg (see Chapter 5.3) and a company party on the HITS premises were again possible. For the administration, this also meant a return to processes that had been in little demand in the previous two years.

However, the fact that the Institute was not yet in full operation compared with the time before the pandemic was to our advantage because the digitization of many processes (see the 2021 Annual Report) relating to invoices and accounting, planning and reporting, as well as travel requests and invoicing had been newly introduced and required time to get used to.
We also used the time to work in various committees on topics such as family friendliness and diversity. The Equality Plan required by the EU and other third-party funding bodies was developed, as was our voluntary commitment to Good Scientific Practice, which is currently undergoing final clarifications.

Last but not least, we accompanied the launch of the MLI group with Jan Stühmer as group leader (see Chapter 2.7), and we recruited numerous researchers for externally funded projects, including an ERC Starting Grant for Ganna (Anya) Gryn'ova of the CCC group (see Chapter 2.2).

3.2 IT Infrastructure and Network

The COVID-19 pandemic continued to influence activities at HITS in 2022, with remote work and videoconferencing becoming well-established. In the second half of the year, after the distancing and hygiene mandates had been canceled, ITS group members took advantage of the institute-wide policy that allows part of the work to be performed remotely. Still, all of them warmly welcomed the freedom to work closer together and meet other HITSters in person again.

One of the prevailing themes in the IT world of 2022 was the disruption of supply chains. In order to circumvent long delays and even uncertain deliveries for new orders, ITS chose to prolong the service for several existing infrastructure components. Most of the new purchases made at HITS focused on storage, which appeared to be less affected by the disruptions.

At the beginning of the year, the small storage system that held the home directories for our HPC environment was replaced with a more powerful NetApp storage appliance, which also increased the reliability of the system.

Our large storage system at Heidelberg University's Computing Centre (URZ) – initially set up in 2016 and containing around 2.5PB of scientific data – became too small, and its components began to show their age. A new system – with a capacity of around 4.3PB – was ordered in summer, but delivered at the very end of the year. The initial transfer of data from the old storage system to the new one took place over the quieter Christmas holiday, with the final synchronization and the switch to production mode planned for the end of January 2023.

Group leader

Dr. Ion Bogdan Costescu

Team

Julian Aris (student, until August 2022)
Dr. Bernd Doser (Senior Software Developer)
Dr. Simon Kreuzer (System Administrator)
Norbert Rabes (System Administrator)
Andreas Ulrich (System Administrator)
Taufan Zimmer (System Administrator)

A further improvement in the area of scientific data storage was the addition of two backup servers, each of which is shared by two groups with similar data management needs. For long-term storage of scientific data, HITS began to use the URZ archive system.

The backup infrastructure for servers, workstations, and notebooks was completely overhauled together with our colleagues from the KTA IT. Now, HITS operates a separate backup system for physical (as opposed to virtual) machines both for our own computers and for those of our sister company HeiGIT.

The network connection between the two parts of our HPC environment, located at HITS and URZ, was upgraded from 40Gbit/s to 100Gbit/s, and the switches were replaced with more powerful models. This new setup provides better access to remote storage systems, an increased reliability, and ample reserves for future network expansions.

4 Communication and Outreach

Head of Communications

Dr. Peter Saueressig

Team

Isabel Lacurie (until May 2022)
Angela Michel
Anna Cap (student)
Hanna Rabes (student; from July to September 2022)

The HITS Communications team is the Institute's central hub for external and internal communications. We strive 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 work on sparking enthusiasm for science among school students and the general public alike through our outreach activities. In 2022, we were challenged by several developments and incidents ranging from the return of live events to a personnel change. Our Annual Report 2022 illustrates how we coped with these challenges.

Successful research and public interest

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 2022, the Communications team was again pleased to announce success stories to the public.

At the beginning of the year, junior group leader Ganna (Anya) Gryn'ova (CCC) received an ERC Starting Grant of approximately €1.5 million for her project PATTERNCHEM (see Chapters 1 and 2.2.). Ganna's proposal was one of only 10% of projects that were selected for funding. These prestigious grants reflect the high quality of our research. In 2022, researchers from six of our thirteen groups (including Ganna's group) worked on or participated as a beneficiary in an ERC Grant.

Group leader Alexandros Stamatakis (Computational Molecular Evolution) successfully applied for an ERA (European Research Area) Chair and will receive funding of €2.4 million from the European Commission to establish computational biodiversity research in Crete, Greece, while working closely with local institutions on the island (see Chapters 1 and 2.3). Moreover, according to this year's Highly Cited Researchers list from Clarivate, Alexandros was named one of the most-cited researchers worldwide for the seventh year in a row. This ranking is an important indicator of the impact of a researcher's scientific publications.

HITS researchers were also very busy developing projects and publishing papers in 2022. For example, the new "SIMPLAIX" collaboration gained momentum, and we had the pleasure of issuing an international press release to announce the "magic triangle" of cooperation between HITS, Heidelberg University, and Karlsruhe Institute of Technology (KIT) with the aim of addressing challenges in the simulation of biomolecules and molecular materials by pooling the institutes' expertise in multiscale computer simulation and machine learning (see Chapter 7).

Moreover, in 2022, an international research team with the participation of members of the SET group discovered a "dormant" black hole in a binary star system outside our Galaxy, and the findings were published in "Nature Astronomy." The researchers literally found an "invisible needle in a stellar haystack" because these black holes do not emit any X-ray radiation (see Chapter 2.13).

Together with colleagues from different institutions in Heidelberg, scientists from the DMQ group also established an international initiative for unraveling the role of a certain type of antigen as a risk modifier in individuals with a genetic predisposition to cancer. In 2022, the scientists published a paper in the International Journal of Cancer that emphasized the relevance of data analysis and mathematical modeling. The project is being pursued within the framework of "Mathematics in Oncology" and is funded by the Klaus Tschira Foundation (see Chapter 2.5).

Events: Back to real life

After three years of struggling with mostly online meetings, we were happy to organize in-person events again, mainly for scientific workshops and conferences, but also for outreach purposes.

The first such event was the inaugural symposium of SIMPLAIX on 12 April (https://www.h-its.org/2022/04/28/simplaix-symposium/) (see Chapter 7).

Another highlight was the "VFTS & friends" astrophysics meeting organized by Fabian Schneider (SET) in late June, at which scientists from seven countries and a number of institutes met at Studio Villa Bosch in Heidelberg to analyze and

Making sure things work like they should for our hands-on station at Explore Science – Christopher Ehlert (left) and Jonathan Teuffel on the HITS terrace a few days before the show.

Vincent Heuveline (DMQ) talking about "KI und der Komponist aus den Fugen" ("AI and the Composer Gone Haywire") during the "Klangforum Heidelberg" ("Sound Forum Heidelberg") concert at the Providence Church in Heidelberg.

initiate new developments regarding massive stars, black holes, and binary star systems (see Chapter 5.1.3) (https://www.h-its.org/2022/07/12/vfts-meeting/).
Only a couple of days later, HITS again participated in the "Explore Science" show at Luisenpark Mannheim (22–26 June). The topic for 2022 was "Digitale Welten" ("Digital Worlds"), for which scientists from 5 groups (CCC, CST, MBM, MCM, PSO) ran hands-on stations on stars and magnets with the support of the Communications and Administration teams.
Furthermore, after four years, we finally opened our doors to the public again. As part of the broad theme of "Digital Worlds 20.22," the program included science talks in English and German, presentations, and hands-on stations, all of which showcased the Institute's research (see Chapter 5.3).
Shortly afterward, the Communications team helped the GRG group organize a large three-day workshop on "Geometry and Machine Learning" in the Studio Villa Bosch that brought together around 90 mathematicians and machine learning experts (see Chapter 5.1.4).
Moreover, the Heidelberg Laureate Forum (see chapter 7) was re-launched: On 21 September, we were happy to host around 25 young researchers from five continents at the Studio Villa Bosch, with presentations by Rebecca Wade (MCM), Vincent Heuveline (DMQ), and Saskia Hekker (TOS) as well as with a poster session by researchers from the CST, GRG, and DMQ groups.

Last but not least, Vincent Heuveline participated in a musical project dedicated to artificial intelligence in a fruitful collaboration with the "Klangforum Heidelberg" ("Sound Forum Heidelberg"), for which he gave talks about AI and the art of composing music to a lay audience at one concert each in Heidelberg and Karlsruhe.
By the end of 2022, the event team (Communications plus Christina Blach and Benedicta Frech from the HITS Administration) had to keep track of an increasing number of in-person scientific events, including the NFDI4Health consortium meeting in October (see Chapter 5.1.7) and the Astrophysics workshop in December (see Chapter 5.1.8), not to mention the 11 colloquium talks that needed to be organized as hybrid events, streamed, recorded, and broadcast on the HITS YouTube channel.

These growing demands had to be covered by a team that shrank by one-third over the course of the year: After nine years at HITS, Isabel Lacurie left the Institute in May 2022 to head for a science communications position in London, UK. For the rest of the year, the remaining two-thirds of the Communications team managed the balance between organization and support on the one hand and outreach and press activities on the other hand.
Angela Michel succeeded in applying for a poster session at the European Open Science Forum in Leiden (Netherlands) together with Fabian Schneider from the SET group. Angela presented the poster – entitled "Universe inside" – at the conference in mid-July. Furthermore, Peter Saueressig was invited to present the HITS "Journalist in Residence" program in an online talk at the "Europäische Akademie" ("European Academy") in Berlin.

A meeting to remember: Carl Smith (HITS Journalist in Residence 2022, right) together with Volker Stollorz (HITS Journalist in Residence 2012) at the Science Media Center Germany, Cologne (photo: Science Media Center Germany).

A success story continued: The "HITS Journalist in Residence" program

We firmly believe that an important prerequisite for successful science communication involves fostering reliable and sustainable journalistic contacts. Since 2012, HITS has refined its "Journalist in Residence" program, which is geared toward experienced science journalists and offers these individuals a paid sojourn at HITS. During their stay, the journalists can learn more about data-driven science and get to know researchers as well as new research topics without the pressure of the "daily grind."
Our 10th "Journalist in Residence" was Carl Smith (Sydney), who works for the Australian Broadcasting Corporation (ABC). Carl stayed for six months, gave an internal seminar on science journalism for kids, and delivered a public talk on how different countries had handled the COVID-19 pandemic. Moreover, Carl also used his stay to conduct research at several scientific locations and to visit the "Science Media Center Germany" (SMC) in Cologne. While there, Carl met the SMC team, which is led by managing director Volker Stollorz, who was the first HITS Journalist in Residence.
Carl Smith also joined HITS Head of Communications Peter Saueressig in a virtual conference organized by the European Research Council (ERC). As the ERC had planned to establish a new project modeled after the HITS residency (called "Science Journalism Initiative"; https://erc.europa.eu/apply-grant/science-journalism-initiative), both Carl and Peter were invited to talk about the program and their experiences with it.

In the summer, HITS announced its next call for applications, with candidates from six continents applying. A committee of science journalists and scientists selected science journalist and trained electronics and computer engineer Anil Ananthaswamy (USA/India) as the next HITS Journalist in Residence. Anil will come to Heidelberg in April 2023 for a six-month stay.

5 Events

5.1 Conferences, Workshops & Courses

5.1.1 HBPMolSim Human Brain Project Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades

7-10 March 2022 (online)

One of the goals of the EU-supported Human Brain Project (HBP) is to develop computational approaches for the multiscale simulation of brain processes, and to make the corresponding data, software, and workflows available in the EBRAINS infrastructure (https://ebrains.eu). HBP training workshops aim to provide practical training in the use of these computational tools to interested scientists ranging from master students to postdoctoral researchers. In 2022, we organized the first HBPMolSim training workshop on computational tools developed in HBP that enable brain simulation and modelling at the molecular and subcellular levels. This 4-day event was attended by 68 participants from 18 countries, and all participants had the opportunity to put the computational tools developed in the HBP into practice through interactive hands-on sessions. The training topics included OMICs and molecular data, molecular dynamics, Brownian dynamics and QM/MM simulations, and subcellular modeling and simulation. Members of the MCM group at HITS presented the computation of binding kinetics using the τRAMD approach and the SDA software.

For more details, see: https://www.humanbrainproject.eu/en/education/molecular-simulation-tools/

Organizers: Rebecca Wade, Giulia D'Arrigo (HITS), Giulia Rossetti (Forschungszentrum Jülich), Jeanette Hellgren Kotaleski (Royal Institute of Technology, Stockholm)

5.1.2 LiSyM-Cancer Annual Status Seminar

18-19 May 2022, Heidelberg

SPONSORED BY THE
Federal Ministry of Education and Research

LiSyM-Cancer (Liver systems medicine against cancer) is a multidisciplinary research network funded by the Federal Ministry of Education and Research (BMBF) within the Framework of the National Decade Against Cancer. The systems medicine network conducts research for the early detection and prevention of liver cancer. The first status seminar of the research network took place in Heidelberg at the German Cancer Research Center (DKFZ) and the BioQuant, Heidelberg University. About 70 clinical researchers, molecular and cell biologists and experts in mathematical modeling from all over Germany participated in the seminar, as well as representatives from the Projektträger Jülich (PtJ). The partners presented their projects and engaged in discussions about clinical, experimental and modeling aspects related to the investigation of the development of liver cancer from pre-existing conditions such as non-alcoholic fatty liver or liver cirrhosis. Members of the Scientific Advisory Board (SAB) from the USA, the Netherlands, Austria and Germany provided valuable input and feedback.
The event was organized in collaboration by the program directorate and data management team of the LiSyM-Cancer network, which is affiliated with the Scientific Databases and Visualization group (SDBV): Alain Becam, Susan Eckerle, Olga Krebs, and Wolfgang Müller, under the supervision of program director Beat Müllhaupt (see chapter 2.12). During the event, they were supported by Angela Michel from the HITS event team.
More information on LiSyM-Cancer, its partners and research: www.lisym-cancer.org

5.1.3 Massive stars, black holes, and binaries: VFTS & Friends meeting

20–22 June 2022, Studio Villa Bosch, Heidelberg

It all started with an “ESO Large Programme”: In 2008, the VLT-FLAMES Tarantula Survey (https://www.roe.ac.uk/~cje/tarantula/) began mapping massive stars with Chris Evans as PI (then in Edinburgh, UK). The program was extremely successful. After all the observational data had been analyzed and the report had been published, the involved researchers decided to continue collaborating. They adopted the title of the ESO program under the acronym "VFTS."

After two online workshops, the first meeting took place in person at Studio Villa Bosch in Heidelberg from 20–22 June 2022, coordinated and organized by HITS group leader Fabian Schneider (Stellar Evolution Theory Group; SET). Observers, data analysts, and modelers came together from seven countries and a number of institutes, including Armagh Observatory in Northern Ireland, the Astrophysical Institute of the Canary Islands in Spain, the Max Planck Institute for Astrophysics in Garching (Germany), and the University of Amsterdam (Netherlands). The event was designed as an open format for analyzing the current state of research and for initiating new developments. The meeting also served as a platform for young scientists and provided new insights into massive stars, black holes, and binary star systems.

The VLT-FLAMES Tarantula Survey is an ESO Large Programme that has provided a rich legacy dataset for studies on both resolved and integrated populations of massive stars. Initiated in 2008 (ESO Period 82), the Fibre Large Array Multi Element Spectrograph (FLAMES) has been used to observe more than 800 massive stars in the dramatic 30 Doradus star-forming region in the Large Magellanic Cloud. Chris Evans serves as the survey’s PI, and HITS researcher Fabian Schneider is one of the collaborators. A summary of the survey with the latest results was published in 2020 (https://doi.eso.org/10.18727/0722-6691/5207).

5.1.4 Workshop on Geometry and Machine Learning

11–13 July 2022, Studio Villa Bosch, Heidelberg

In recent years, there has been a remarkable increase in research into applying geometry to machine learning problems, which has been epitomized by the upward trend in the use of geometric deep learning architectures, tools, and publications. Geometry has successfully provided diverse methods for describing the structure of data, and researchers have suggested frameworks for analyzing, unifying, and generalizing machine learning techniques in new settings.

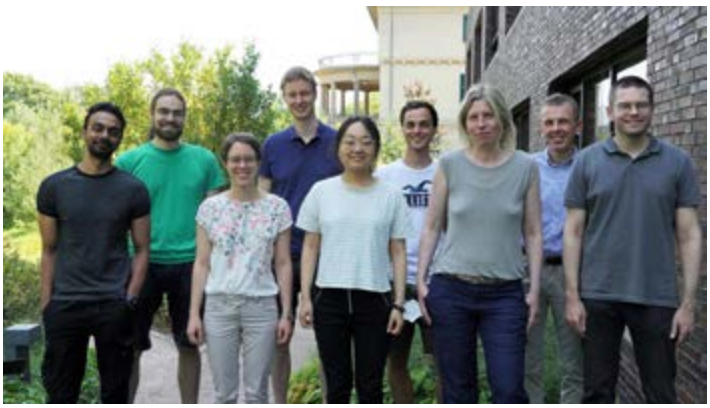
The three-day "Workshop on Geometry and Machine Learning" brought together researchers and practitioners from both fields to interact and exchange ideas. The workshop consisted of invited talks, including three keynotes by well-known experts. Moreover, there were two hands-on mini courses on implementing geometric deep learning methods in Python.

The workshop was organized by Valentina Disarlo, Diaaeldin Taha, and Anna Wienhard (Geometry and Groups, GRG).

Among the speakers was Beatrice Pozzetti (Heidelberg University), who discussed the HITS Lab project "Geometry and Representation Learning." Other speakers included Erik J Bekkers (University of Amsterdam, Netherlands), Maxim Kochurov (PYMC Labs), Maximilian Nickel (Facebook Research, New York, USA), Björn Ommer (LMU Munich, Germany), Emanuele Rodola (Sapienza University, Rome, Italy), Anastasis Kratsios (University of Basel, Switzerland), Xavier Pennec (INRIA, France), Nicolas Guigui (CNRS, France), and Pim de Haan (Qualcomm AI Research, Netherlands).

Anna Wienhard at the opening of the workshop

5.1.5 Workshop on post-processing



20 July 2022, HITS campus, Heidelberg

Numerical weather prediction models based on physical models of the atmosphere exhibit systematic errors between generated forecasts and observed outcomes. Post-processing aims to recalibrate the output of physical models in order to make them statistically consistent with actual outcomes. One fundamental prerequisite of modern forecasting is that predictions be probabilistic – that is, uncertainty must be quantified within the forecast and given in the form of a probability distribution.

Traditionally, these challenges have been addressed via the statistical technique of distributional regression, in which parameters of a family of distributions are estimated using proper scoring rules. The two main directions of active research include (1) providing new distributional regression approaches via techniques of machine learning and (2) improving the spatio-temporal and inter-variable coherence of forecast fields.

These were the main topics discussed during the hybrid-format Computational Staistics (CST) Workshop on post-processing, which covered applications to topical issues such as power generation and high-impact weather events.

Participants included long-term collaboration partners from the University of Debrecen, the Karlsruhe Institute of Technology, and the University of Bern as well as colleagues from the University of Hildesheim, the University of Bielefeld, the University of Zurich, and MeteoSwiss.

5.1.6 EuroQSAR: 23rd European Symposium on Quantitative Structure-Activity Relationships

26–30 September 2022, Heidelberg University

Co-funded by the European Union

The EuroQSAR symposia are the leading international symposia in the field of computational drug design and the study of Quantitative Structure-Activity Relationships (QSAR). The scope ranges from basic academic research into drug mechanisms and design to applications such as the monitoring of toxic substances by government agencies and the discovery of new pharmaceuticals and pesticides by industry. The EuroQSAR symposia have been held since 1973 and have their roots in the pioneering work of Hansch and Fujita in the 1960s who developed quantitative models relating the structures of chemical compounds to their activities. The scope of EuroQSAR has since expanded dramatically to broadly cover the different approaches to computational drug design such as structure-based drug design and virtual screening. The focus of the 23rd EuroQSAR was on "Integrative Data-Intensive Approaches to Drug Design" and included presentations on cutting-edge approaches, such as new techniques for drug discovery that combine molecular simulation and artificial intelligence, or that exploit the information in very large medical datasets by data mining.

The 23rd EuroQSAR was the first EuroQSAR symposium to be held in Germany since the year 2000. It was chaired by Molecular and Cellular Modeling (MCM) group leader Rebecca Wade, and organized on behalf of QCMS – the QSAR, chemoinformatics and modeling society (https://www.qsar.org/). QCMS is an international society open to scientists involved in QSAR, cheminformatics, and modeling for medicinal, agricultural, or environmental chemistry. Almost 300 participants working in academia, government and industry labs around the world attended the symposium, which provided an important venue for interaction between scientists across these areas. The symposium included 4 award lectures, 12 plenary lectures, 16 session lectures, 11 oral communications, several workshops, poster presentations and a commercial exhibition. The 23rd EuroQSAR was supported by the DFG (German Research Foundation), HITS, and Heidelberg University, was an EFMC (European Federation for Medicinal Chemistry and Chemical Biology) sponsored event and was sponsored by the Interdisciplinary Center for Scientific Computing (IWR) at Heidelberg University, Novartis, and other institutions, companies, and publishers.

For more details, see: https://www.euroqsar2022.org/

Participants at the 23rd EuroQSAR

5.1.7 NFDI4 Health Annual Meeting

20–21 October 2022, Studio Villa Bosch, Heidelberg

Deutsche Forschungsgemeinschaft
German Research Foundation

The aim of the National Research Data Infrastructure for Personal Health Data (NFDI4H) project is to create an infrastructure based on standards that enables harmonization, is expandable, and facilitates the retrieval and use of public health data.
The annual meeting of the consortium took place from 20–21 October 2022 at the Studio Villa Bosch in Heidelberg. For two days, partners and members of the international advisory board met to present and discuss the results of several work packages as well as other topics, such as the NFDI4Health Main Service, future services, community outreach, and data harmonization. Within both the plenary sessions and the small working groups, the 50 participants also developed a concrete action plan for the next steps of NFDI-4Health. The event was organized by members of the Scientific Databases and Visualization (SDBV) group and the HITS event team in cooperation with the NFDI4H project management team in Bonn.

HITS Managing Director Gesa Schönberger delivering the welcome address.

5.1.8 Astrophysics "Würzburg" workshop

12–14 December 2022, Studio Villa Bosch, Heidelberg

From 12–14 December 2022, the "16th Würzburg workshop" on stellar astrophysics took place in person again. The international group of about 40 participants included scientists from several institutions both from within Germany and abroad. The workshop included topics on supernova research, binary stellar evolution, neutron stars, and stellar hydrodynamics.
The plenary session featured invited talks by Philipp Podsiadlowski (University of Oxford, UK), Federico Rizzuti (Keele University, UK), Fionntan Callan (Queen's University Belfast, UK), Dandan Wei (HITS), Georgios Lioutas (GSI Darmstadt, Germany), Wasilij Barsukow (Université de Bordeaux, France), Rajika Kuruwita (HITS), and Vishnu Varma (Keele University, UK).
The plenary session was followed by topical sessions in which the participants presented their latest results and discussed collaborations projects.

Stairway to the sky: the Workshop group in the Studio Villa Bosch

5.2 HITS Colloquia

Eva Wolfangel

Freelance Science and Tech Journalist, Speaker, Stuttgart, Germany.
24 January 2022: Immersive Media for Science Journalists and Science Communication (Online)

Prof. Frank Noé

Department of Mathematics and Computer Science, Mathematical Modeling in the Life Sciences, Freie Universität Berlin, Germany.
21 February 2022: SIMPLAIX Joint Colloquium on Deep Learning for Molecular Physics and Chemistry (Online)

Prof. Ruth Nussinov

National Cancer Institute, Center for Cancer Research, USA.
25 April 2022: Unraveling Oncogenic Mechanisms and their Linkage to Neurodevelopmental Disorders (Online)

Prof. Dr.-Ing. Dr. h.c. Andreas Reuter

Former Managing Director of HITS, Germany.
16 May 2022: Habitual inclination towards scrutiny – A brief reflection on how HITS came about (Hybrid)

Carl Smith

Journalist in Residence 2022, Sydney, Australia.
24 May 2022: Fortress Australia and E-Narnia – Did unique societal adaptions help Australia and Estonia during the pandemic? (Hybrid)

Prof. Dr. Ing. Julio Saez-Rodriguez

Institute of Computational Biomedicine, Heidelberg University Hospital, Germany.
28 June 2022: Computational Models from Multi-omics Data for personalized Medicine (Hybrid)

Prof. Antonis Rokas

Klaus Tschira Guest Professor, Department of Biological Sciences, Vanderbilt University, Nashville/Tennessee, USA.
6 July 2022: Incongruence in the Tree of Life (Hybrid)

Prof. Eike Hermann Müller

Associate Professor, Department of Mathematical Sciences, University of Bath, UK.
18 July 2022: Efficient Fast Multiple Methods for (kinetic) Monte Carlo simulations of interacting particle systems (Hybrid)

Prof. Sarbani Basu

Klaus Tschira Guest Professor, Professor of Astronomy and Chair, Department of Astronomy, Yale University, USA.
4 October 2022: Efficient Learning Physics through Astronomy – The Sun and Stars as Laboratories (Hybrid)

David Dao

GainForest/ PhD candidate ETH Zürich, Switzerland.
24 October 2022: Gainforest – Using artificial intelligence to help restore the natural world (Hybrid)

Prof. Michele Ceriotti

Associate Professor, COSMO Lab, EPFL, STI SMX-GE, Lausanne, Switzerland.
21 November 2022: Atomic-scale Modeling in the Age of Machine Learning (Hybrid)

5.3 HITS Open House Event

9 July 2022

After a pause of four years, HITS finally opened its doors to the public again last summer. Under the overall theme of "Digital Worlds 20.22: In the beginning was the Code," the event included science talks in English and German, presentations, and hands-on stations, all of which showcased the Institute's research.

The official program on 9 July kicked off with one of this year's highlights: a tour through our beautiful garden under the expert guidance of our gardener Andrea Baumgärtner. The tours were offered in English and German and were fully booked in next to no time.

Shortly after noon, bioinformatician Alexandros Stamatakis (Computational Molecular Evolution, CME) was the first of three HITS scientists to present their research in a packed auditorium in the Studio Villa Bosch. His talk on how algorithms can be used to predict the outcome of soccer games was very well

Saskia Hekker during her talk.

The weather was perfect.

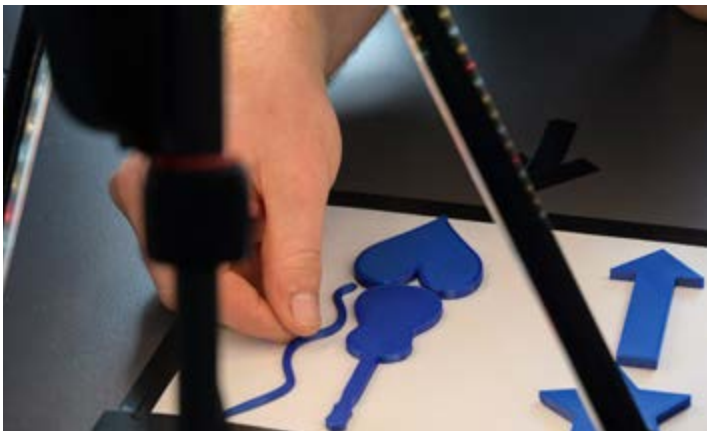
A walk in the park: the garden tour.

Go with the flow: testing the visitors' streamlining skills.

received by the predominantly lay audience. Alexandros's offering was followed by HITS astrophysicist Saskia Hekker's (Theory and Observations of Stars, TOS) presentation of the latest news from inside the stars and by statistician Johannes Bracher's (Computational Statistics, CST) introduction to why nowcasting is the new type of forecasting for predicting corona infections.

Throughout the day, four parallel hands-on stations invited visitors to interact with HITS researchers, for example, by testing their streamlining skills in a "Go-With-The-Flow Challenge," which members of the Data Mining and Uncertainty Quantification group (DMQ) had set up in the foyer of the main building. Outside the garage, visitors could demonstrate a different set of skills by dancing the "Black Hole" with the Astroinformatics group (AIN). The other two stations – "Molecules in Motion" by the Molecular and Cellular Modeling group (MCM) and "The Secrets of Stability" by the Molecular Biomechanics group (MBM) – were less arduous but equally as popular.

As in previous years, the food prepared by HITS chef Ralph Westermann and his team was also among the highlights of the event. The more than 400 visitors could choose between mushroom ragout, pulled turkey burgers, and many other tasty dishes on this beautiful summer's day while enjoying lunch with a view of the HITS garden.



Alexandros Stamatakis talking about algorithms and soccer.

Johannes Bracher on "nowcasting."

Presenting the "secrets of stability."

6 Special programs

6.1 Klaus Tschira Guest Professorship

In 2022, HITS introduced the Klaus Tschira Guest Professorship Program, which aims to enhance international exchange and scientific collaboration at the Institute in the fields of natural, mathematical, and computer science. To that end, HITS invites internationally renowned scientists for sabbaticals or extended research stays that range from three weeks to six months. Invited guest professors collaborate with scientists at the Institute and potentially develop joint research projects. Additionally, they are encouraged to engage with the wider scientific community both at HITS and in the region in the form of lectures, teaching, and scientific discussions.

This year, HITS welcomed two Klaus Tschira Guest Professors. Due to a small overlap, they even had the chance to meet each other and have a chat in the HITS coffee bar lounge.

The tree of life and the physics of the Sun

The first Klaus Tschira Guest Professor was biologist Antonis Rokas from Vanderbilt University, USA. He arrived in June 2022 and stayed at HITS until the end of September.

"The funding mechanism here is great. You are investing in people, not in projects. That's the best model."

(Antonis Rokas)

Antonis used his time to intensively collaborate with the Computational Molecular Evolution group (CME). Together with group leader Alexandros Stamatakis, he worked on a synthesis article about the challenges faced by evolutionary biology due to increasing datasets and the millions of genomes to be sequenced. Antonis also began a project with the CME group on the irreproducibility problem in computational biology, in which many factors must be considered, including hardware, software, and the data involved. Last but not least, he additionally gave a colloquium talk in the Studio Villa Bosch on the "Incongruence in the Tree of Life" (see Chapter 5.2).

Learning physics through astronomy: Sarbani Basu during her colloquium talk.

Science coffee: Antonis Rokas (left) and Sarbani Basu (right).

Antonis Rokas talking about "Incongruence in the Tree of Life" at the HITS colloquium.

The second Klaus Tschira Guest Professor, astrophysicist Sarbani Basu, arrived in mid-September. A specialist in research on the Sun and former chair of the Department of Astronomy at Yale University, Connecticut (USA), Sarbani spent more than two months at the Institute until late November 2022, during which time she collaborated mostly with the Theory and Observations of Stars group (TOS). Together, they worked on several projects involving the interior of – and rotations inside – stars.

"This program is an amazing opportunity to sit down and just concentrate on science."

(Sarbani Basu)

Moreover, Sarbani gave a HITS colloquium talk on the physics of the Sun in early October (see Chapter 5.2), and she delivered further talks at Heidelberg as well as at the joint colloquium series of the Max Planck Institute for Astrophysics (MPA), the Max Planck Institute for Extraterrestrial Physics (MPE), and the European Southern Observatory (ESO) in Garching. Sarbani additionally enjoyed the opportunity to engage in discussions with scientists from other disciplines at HITS, for example, in the coffee bar or on her daily commute in the "Science Bus" that connects the HITS campus to the city. The Klaus Tschira Guest Professorship program will be continued 2023 with two guests: an astrophysicist from the USA and a chemist from the United Kingdom.

6.2 HITS Independent Postdoc Program

The HITS Independent Postdoc Program offers an exciting opportunity for highly talented young scientists who wish to transition from PhD students to junior group leaders. It supports these young scientists in exploring their own ideas and testing new hypotheses. High-risk, high-gain projects are encouraged. Selected postdocs collaborate with group leaders at HITS while developing and pursuing their independent research projects.

The fellowship is awarded for two years, with an option for a one-year extension following a positive evaluation. It offers a vibrant research community and a highly interdisciplinary and international working environment, with close links to HITS shareholders Heidelberg University and the Karlsruhe Institute of Technology (KIT). In addition, successful candidates benefit from outstanding computing resources and various courses offered at HITS. Candidates for this program must hold a doctoral degree or an equivalent academic qualification at the beginning of the fellowship. Application is open to candidates for up to three years after the completion of their PhD by the time of the application deadline. This limit can be extended in the case of documented career breaks, for example, due to parental leave. Candidates must not have carried out research at HITS previously, except for brief visits, and the main thread of their research must not have been in collaboration with a HITS group leader.

The first HITS Independent Postdoc is astrophysicist Rajika Kuruwita. Born in Sri Lanka, Rajika completed her PhD at the Australian National University, was a fellow at the University of Copenhagen, and joined HITS in September 2022 (see Chapter 2.15).

In summer 2023, the second HITS Independent Postdoc is due to join the Institute. The next opening for the program will be in 2024.

7 Collaborations

SIMPLAIX

SIMPLAIX – a 3-way inter-institutional collaboration between HITS, Heidelberg University, and the Karlsruhe Institute of Technology (KIT) – aims to pool the expertise of the three partner institutes with the goal of addressing the challenge of bridging scales from molecules to molecular materials by using multiscale simulations and machine learning.
In SIMPLAIX, these methods are developed and employed to study a set of challenging problems in biomolecules and molecular materials within 8 multidisciplinary, inter-institutional research projects. 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. SIMPLAIX is funded by the Klaus Tschira Foundation and is supported by in-kind contributions from KIT and Heidelberg University.
On 12 April 2022, the "SIMPLAIX" collaboration was launched with an inaugural symposium at the Studio Villa Bosch in Heidelberg, with project members and representatives of all three institutions participating. In this hybrid meeting, 45 participants attended the event in person, and another 30 people joined via videoconference.
After the welcome addresses by the representatives of the three partner institutions and the Klaus Tschira Foundation, two members of the SIMPLAIX international scientific advisory board – both of whom are experts in machine learning approaches to studying molecular systems – gave scientific talks: Anatole von Lilienfeld (University of Vienna) spoke about "Quantum Machine Learning," and Jörg Behler (University of Göttingen) explained "High-dimensional neural network potentials for simulations of complex systems."
Part of the initiative are 8 positions for young researchers who work on SIMPLAIX projects. In the course of 2022, all these positions were filled. Two project meetings were held, one of which took place at the Studio Villa Bosch. Moreover, several joint HITS–SIMPLAIX colloquia and scientific talks were organized.

The (hybrid) inaugural symposium in April 2022.

The outlook for 2023 is also promising: In May, a three-day international workshop will be organized, and several joint colloquia and project meetings are also in the offing.

At the inaugural symposium, Anatole von Lilienfeld gave a keynote speech on Quantum Machine Learning.

Project meeting at the Studio Villa Bosch in September.

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 the same two 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). 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.

Tilmann Gneiting welcomed the young researchers in the Carl Bosch Auditorium.

Back to live again: The 9th HLF

After two years of digital events, the 9th HLF 2022 took place in person once again in Heidelberg. From September 18–23, a plethora of activities awaited the participants, including laureate lectures, panel discussions, and various interactive program elements. The "Hot Topic" centered on Deep Learning as well as on its applications and implications. During the week-long conference, young researchers and other participants had the opportunity to connect with scientific pioneers and to learn how the laureates had made it to the top of their fields.

During a break in the Villa Bosch garden.

HITSters meet young researchers

As in the years before the pandemic, HITS hosted a group of 25 young researchers ranging from undergraduate students to postdoctoral scientists. Deputy Scientific Director Tilmann Gneiting welcomed the group and introduced the presentations by HITS group leaders Saskia Hekker (TOS), Vincent Heuveline (DMQ), and Rebecca Wade (MCM). Moreover, members of the CST, DMQ, and GRG groups presented their current research topics and publications in a poster session. As the group was larger than in previous years, the event took place in the Studio Villa Bosch.

HLF poster session in the Studio Villa Bosch.

NFDI4Health

Deutsche Forschungsgemeinschaft
German Research Foundation

The aim of the NFDI4Health project is to create a standards-based infrastructure that enables harmonization, is expandable, and facilitates the retrieval and use of public health data, (dietary) exposure data, and clinical trial data, thereby facilitating structured combination and interoperability. The initiative consists of an interdisciplinary team of 18 partners. A total of 46 renowned institutions from the health sector have confirmed their participation, including major professional associations and epidemiological cohorts.

With its expertise in research data management and scientific databases as well as its experience in data standardization, the Scientific Databases and Visualization group (SDBV) serves as a partner in the project. The FAIRDOM-SEEK software will play a central role as the data management platform and will be adapted to meet the project-specific requirements. Relevant data standards will be adapted and harmonized in order to facilitate finding and comparing the collected personal health data. The project has been granted funding for five years by both the German Federal Government and German state governments (see also Chapter 5.1.6).

8 Publications

Ahadova A, Witt J, Haupt S, Gallon R, Hüneburg R, Nattermann J, Ten Broeke S, Bohaumilitzky L, Hernandez-Sanchez A, Santibanez-Koref M, Jackson MS, Ahtiainen M, Pylvänäinen K, Andini K, Grolmusz VK, Möslein G, Dominguez-Valentin M, Møller P, Fürst D, Sijmons R, Borthwick GM, Burn J, Mecklin JP, Heuveline V, von Knebel Doeberitz M, Seppälä T, Kloor M (2022). Is HLA type a possible cancer risk modifier in Lynch syndrome? Intl Journal of Cancer, ijc.34312

Amaro-Seoane P, Andrews J, Sedda MA, Askar A, Balasov R, Bartos I, Bavera SS, Bellovary J, Berry CPL, Berti E, Bianchi S, Blecha L, Blondin S, Bogdanović T, Boissier S, Bonetti M, Bonoli S, Bortolas E, Breivik K, Capelo PR, Caramete L, Catorini F, Charisi M, Chaty S, Chen X, Chruślińska M, Chua AJK, Church R, Colpi M, D'Orazio D, Danielski C, Davies MB, Dayal P, De Rosa A, Derdzinski A, Destounis K, Dotti M, Duţan I, Dvorkin I, Fabj G, Foglizzo T, Ford S, Fouvry J-B, Fragkos T, Fryer C, Gaspari M, Gerosa D, Graziani L, Groot PJ, Habouzit M, Haggard D, Haiman Z, Han W-B, Istrate A, Johansson PH, Khan FM, Kimpson T, Kokkotas K, Kong A, Korol V, Kremer K, Kupfer T, Lamberts A, Larson S, Lau M, Liu D, Lloyd-Ronning N, Lodato G, Lupi A, Ma C-P, Maccarone T, Mandel I, Mangiagli A, Mapelli M, Mathis S, Mayer L, McGee S, McKernan B, Miller MC, Mota DF, Mumpower M, Nasim SS, Nelemans G, Noble S, Pacucci F, Panessa F, Paschalidis V, Pfister H, Porquet D, Quenby J, Röpke F, Regan J, Rosswog S, Ruiter A, Ruiz M, Runnoe J, Schneider R, Schnittman J, Secunda A, Sesana A, Seto N, Shao L, Shapiro S, Sopuerta C, Stone N, Suvorov A, Tamanini N, Tamfal T, Tauris T, Temmink K, Tomsick J, Toonen S, Torres-Orjuela A, Toscani M, Tsokaros A, Unal C, Vázquez-Aceves V, Valiante R, van Putten M, van Roestel J, Vignali C, Volonteri M, Wu K, Younsi Z, Yu S, Zane S, Zwick L, Antonini F, Baibhav V, Barausse E, Rivera AB, Branchesi M, Branduardi-Raymont G, Burdge K, Chakraborty S, Cuadra J, Dage K, Davis B, de Mink SE, Decarli R, Doneva D, Escoffier S, Gandhi P, Haardt F, Lousto CO, Nissanke S, Nordhaus J, O'Shaughnessy R, Zwart SP, Pound A, Schussler F, Sergijenko O, Spallicci A, Vernieri D, Vigna-Gómez A (2022). Astrophysics with the Laser Interferometer Space Antenna. arXiv e-prints:arXiv:2203.06016

Andrassy R, Higl J, Mao H, Mocák M, Vlaykov DG, Arnett WD, Baraffe I, Campbell SW, Constantino T, Edelmann PVF, Goffrey T, Guillet T, Herwig F, Hirschi R, Horst L, Leidi G, Meakin C, Pratt J, Rizzuti F, Röpke FK, Woodward P (2022). Dynamics in a stellar convective layer and at its boundary: Comparison of five 3D hydrodynamics codes. A&A 659:A193

Barayeu U, Schilling D, Eid M, Xavier da Silva TN, Schlicker L, Mitreska N, Zapp C, Gräter F, Miller AK, Kappl R, Schulze A, Friedmann Angeli JP, Dick TP (2022). Hydropersulfides inhibit lipid peroxidation and ferroptosis by scavenging radicals. Nature Chemical Biology 19: 28-37

Battino U, Lederer-Woods C, Travaglio C, Röpke FK, Gibson B (2022). Slow White Dwarf Mergers as a New Galactic Source of Trans-Iron Elements. EPJ Web Conf. 260:06002

Bayer S, Dimitriadis T (2022). Regression-based expected shortfall backtesting. Journal of Financial Econometrics, 20(3):437-471

Bettisworth B, Smith SA, Stamatakis A (2022). Lagrange-NG: The next generation of Lagrange. bioRxiv 2022.04.19.488734

Boch T, Allen M, Bot C, Fernique P, Baumann M, Buga M, Bonnarel F, Durand D, Polsterer K (2022). Innovative Tools Fostered by the HiPS Ecosystem. vol. 532 of Astronomical Society of the Pacific Conference Series, p 463

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, Fiedler J, Fuhrmann J, Funk S, Gambin A, Gogolewski K, Heyder S, Hotz T, Kheifetz Y, Kirsten H, Krueger T, Krymova E, Leithäuser N, Li ML, Meinke JH, Miasojedow B, Michaud IJ, Mohring J, Nouvellet P, Nowosielski JM, Ozanski T, Radwan M, Rakowski F, Scholz M, Soni S, Srivastava A, Gneiting T, Schienle M (2022). National and subnational short-term forecasting of COVID-19 in Germany and Poland during early 2021. Communications Medicine, 2:136

Brands SA, de Koter A, Bestenlehner JM, Crowther PA, Sundqvist JO, Puls J, Caballero-Nieves SM, Abdul-Masih M, Driessen FA, García M, Geen S, Gräfener G, Hawcroft C, Kaper L, Keszthelyi Z, Langer N, Sana H, Schneider FRN, Shenar T, Vink JS (2022). The R136 star cluster dissected with Hubble Space Telescope/STIS. III. The most massive stars and their clumped winds. A&A 663:A36

Braud C, Hardmeier C, Li JJ, Loaiciga S, Strube M, Zeldes A (eds) (2022) Proceedings of the 3rd Workshop on Computational Approaches to Discourse. International Conference on Computational Linguistics, Gyeongju, Republic of Korea and Online

Bridgeman M, Pozzetti B, Sambarino A, Wienhard A (2022). Hessian of Hausdorff dimension on purely imaginary directions. In Bulletin of the London Mathematical Society 54, 3: 1027 - 1050

Brosz M, Michelarakis N, Bunz UHF, Aponte-Santamaría C, Gräter F (2022). Martini 3 coarse-grained force field for poly(para-phenylene ethynylene)s. Physical Chemistry Chemical Physics,24(17):9998-10010

de Buhr S, Gräter F (2022). Myristoyl's dual role in allosterically regulating and localizing Abl kinase. bioRxiv 2022.12.20.521177

Carvajales L, Dai X, Wienhard A (2022). Thurston's asymmetric metrics for Anosov representations. Preprint available online: arXiv:2210.05292

Chai H, Moosavi NS, Gurevych I, Strube M (2022). Evaluating coreference resolvers on community-based question answering: From rule-based to state of the art. In Proceedings of the Fifth Workshop on Computational Models of Reference, Anaphora and Coreference, Gyeongju, Republic of Korea, 16-17 October, pp. 61-73

Chen P, Kutzki F, Mojzisch A, Simon B, Xu E-R, Aponte-Santamaría C, Horny K, Jeffries C, Schneppenheim R, Wilmanns M, Brehm MA, Gräter F, Hennig J (2022). Structure and dynamics of the von Willebrand Factor C6 domain. Journal of Structural Biology,214(4):107923

Collin CB, Gebhardt T, Golebiewski M, Karaderi T, Hillemanns M, Khan FM, Salehzadeh-Yazdi A, Kirschner M, Krobitsch S, EU-STAND-S4PM consortium, Kuepfer L (2022). Computational Models for Clinical Applications in Personalized Medicine—Guidelines and Recommendations for Data Integration and Model Validation. J Pers Med. Jan 26;12(2):166

Collins CE, Gronow S, Sim SA, Röpke FK (2022). Double detonations: variations in Type Ia supernovae due to different core and He shell masses - II. Synthetic observables. Monthly Notices of the Royal Astronomical Society 517(4):5289-5302

Costantino L, Ferrari S, Santucci M, Salo-Ahen OM, Carosati E, Franchini S, Lauriola A, Pozzi C, Trande M, Gozzi G, Saxena P, Cannazza G, Losi L, Cardinale D, Venturelli A, Quotadamo A, Linciano P, Tagliazucchi L, Moschella MG, Guerrini R, Pacifico S, Luciani R, Genovese F, Henrich S, Alboni S, Santarem N, da Silva Cordeiro A, Giovannetti E, Peters GJ, Pinton P, Rimessi A, Cruciani G, Stroud RM, Wade RC, Mangani S, Marverti G, D'Arca D, Ponterini G, Costi MP (2022). Destabilizers of the thymidylate synthase homodimer accelerate its proteasomal degradation and inhibit cancer growth. Elife. Dec 7;11:e73862

Cramer EY, Huang Y, Wang Y, Ray EL, Cornell M, Bracher J, Brennen A, Rivadeneira AJC, Gerding A, House K, Jayawardena D, Kanji AH, Khandelwal A, Le K, Mody V, Mody V, Niemi J, Stark A, Shah A, Wattanchit N, Zorn MW, Reich NG, Gneiting T, Mühlemann A, Gu Y, Chen Y, Chintanippu K, Jivane V, Khurana A, Kumar A, Lakhani A, Mehrotra P, Pasumarty S, Shrivastav M, You J, Bannur N, Deva A, Jain S, Kulkarni M, Merugu S, Raval A, Shingi S, Tiwari A, White J, Adiga A, Hurt B, Lewis B, Marathe M, Peddireddy AS, Porebski P, Venkatramanan S, Wang L, Dahan M, Fox S, Gaither K, Lachmann M, Meyers LA, Scott JG, Tec M, Woody S, Srivastava A, Xu T, Cegan JC, Dettwiller ID, England WP, Farthing MW, George GE, Hunter RH, Lafferty B, Linkov I, Mayo ML, Parno MD, Rowland MA, Trump BD, Chen S, Faraone SV, Hess J, Morley CP, Salekin A, Wang D, Zhang-James Y, Baer TM, Corsetti SM, Eisenberg MC, Falb K, Huang Y, Martin ET, McCauley E, Myers RL, Schwarz T, Gibson GC, Sheldon D, Gao L, Ma Y, Wu D, Yu R, Jin X, Wang Y-X, Yan X, Chen Y, Guo L, Zhao Y, Chen J, Gu Q, Wang L, Xu P, Zhang W, Zou D, Chattopadhyay I, Huang Y, Lu G, Pfeiffer R, Sumner T, Wang D, Wang L, Zhang S, Zou Z, Biegel H, Lega J, Hussain F, Khan Z, Van Bussel F, McConnell S, Guertin SL, Hulme-Lowe C, Nagraj VP, Turner SD, Bejar B, Choirat C, Flahault A, Krymova E, Lee G, Manetti E, Namigai K, Obozinski G, Sun T, Thanou D, Ban X, Shi Y, Walraven R, Hong Q-J, van de Walle A, Ben-Nun M, Riley S, Riley P, Turtle J, Cao D, Galasso J, Cho JH, Jo A, DesRoches D, Forli P, Hamory B, Koyluoglu U, Kyriakides C, Leis H, Milliken J, Moloney M, Morgan J, Nirgudkar N, Ozcan G, Piwonka N, Ravi M, Schrader C, Shakhnovich E, Siegel D, Spatz R, Stiefeling C, Wilkinson B, Wong A, Cavany S, España G, Moore S, Oidtman R, Perkins A, Ivy JS, Mayorga ME, Mele J, Rosenstrom ET, Swann JL, Kraus A, Kraus D, Bian J, Cao W, Gao Z, Ferres JL, Li C, Liu T-Y, Xie X, Zhang S, Zheng S, Chinazzi M, Vespignani A, Xiong X, Davis JT, Mu K, Piontti AP y, Baek J, Farias V, Georgescu A, Levi R, Sinha D, Wilde J, Zheng A, Lami OS, Bennouna A, Ndong DN, Perakis G, Singhvi D, Spantidakis I, Thayaparan L, Tsiourvas A, Weisberg S, Jadbabaie A, Sarker A, Shah D, Celi LA, Penna ND, Sundar S, Berlin A, Gandhi PD, McAndrew T, Piriya M, Chen Y, Hlavacek W, Lin YT, Mallela A, Miller E, Neumann J, Posner R, Wolfinger R, Castro L, Fairchild G, Michaud I, Osthus D, Wolffram D, Karlen D, Panaggio MJ, Kinsey M, Mullany LC, Rainwater-Lovett K, Shin L, Tallaksen K, Wilson S, Brenner M, Coram M, Edwards JK, Joshi K, Klein E, Hulse JD, Grantz KH, Hill AL, Kaminsky K, Kaminsky J, Keegan LT, Lauer SA, Lee EC, Lemaitre JC, Lessler J, Meredith HR, Perez-Saez J, Shah S, Smith CP, Truelove SA, Wills J, Gardner L, Marshall M, Nixon K, Burant JC, Budzinski J, Chiang W-H, Mohler G, Gao J, Glass L, Qian C, Romberg J, Sharma R, Spaeder J, Sun J, Xiao C, Gao L, Gu Z, Kim M, Li X, Wang Y, Wang G, Wang L, Yu S, Jain C, Bhatia S, Nouvellet P, Barber R, Gaikedu E, Hay S, Lim S, Murray C, Pigott D, Reiner RC, Baccam P, Gurung HL, Stage SA, Suchoski BT, Fong C-Y, Yeung D-Y, Adhikari B, Cui J, Prakash BA, Rodríguez A, Tabassum A, Xie J, Asplund J, Baxter A, Keskinocak P, Oruc BE, Serban N, Arik SO, Dusenberry M, Epshteyn A, Kanal E, Le LT, Li C-L, Pfister T, Sinha R, Tsai T, Yoder N, Yoon J, Zhang L, Wilson D, Belov AA, Chow CC, Gerkin RC, Yogurtcu ON, Ibrahim M, Lacroix T, Le M, Liao J, Nickel M, Sagun L, Abbott S, Bosse NI, Funk S, Hellewell J, Meakin SR, Sherratt K, Kalantari R, Zhou M, Karimzadeh M, Lucas B, Ngo T, Zoraghein H, Vahedi B, Wang Z, Sen Pei, Shaman J, Yamana TK, Bertsimas D, Li ML, Soni S, Bouardi HT, Adee M, Ayer T, Chhatwal J, Dalgic OO, Ladd MA, Linas BP, Mueller P, Xiao J, Bosch J, Wilson A, Zimmerman P, Wang Q, Wang Y, Xie S, Zeng D, Bien J, Brooks L, Green A, Hu AJ, Jahja M, McDonald D, Narasimhan B, Politsch C, Rajanala S, Rumack A, Simon N, Tibshirani RJ, Tibshirani R, Ventura V, Wasserman L, Drake JM, O'Dea EB, Abu-Mostafa Y, Bathwal R, Chang NA, Chitta P, Erickson A, Goel S, Gowda J, Jin Q, Jo H, Kim J, Kulkarni P, Lushtak SM, Mann E, Popken M, Soohoo C, Tirumala K, Tseng A, Varadarajan V, Vytheeswaran J, Wang C, Yeluri A, Yurk D, Zhang M, Zlokapa A, Pagano R, Jain C, Tomar V, Ho L, Huynh H, Tran Q, Lopez VK, Walker JW, Slayton RB, Johansson MA, Biggerstaff M, Reich NG (2022). The United States COVID-19 Forecast Hub dataset. Scientific Data, 9:462

Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, Castro Rivadeneira AJ, Gerding A, Gneiting T, House KH, Huang Y, Jayawardena D, Kanji AH, Khandelwal A, Le K, Mühlemann A, Niemi J, Shah A, Stark A, Wang Y, Wattanachit N, Zorn MW, Gu Y, Jain S, Bannur N, Deva A, Kulkarni M, Merugu S, Raval A, Shingi S, Tiwari A, White J, Abernethy NF, Woody S, Dahan M, Fox S, Gaither K, Lachmann M, Meyers LA, Scott JG, Tec M, Srivastava A, George GE, Cegan JC, Dettwiller ID, England WP, Farthing MW, Hunter RH, Lafferty B, Linkov I, Mayo ML, Parno MD, Rowland MA, Trump BD, Zhang-James Y, Chen S, Faraone SV, Hess J, Morley CP, Salekin A, Wang D, Corsetti SM, Baer TM, Eisenberg MC, Falb K, Huang Y, Martin ET, McCauley E, Myers RL, Schwarz T, Sheldon D, Gibson GC, Yu R, Gao L, Ma Y, Wu D, Yan X, Jin X, Wang Y-X, Chen Y, Guo L, Zhao Y, Gu Q, Chen J, Wang L, Xu P, Zhang W, Zou D, Biegel H, Lega J, McConnell S, Nagraj VP, Guertin SL, Hulme-Lowe C, Turner SD, Shi Y, Ban X, Walraven R, Hong Q-J, Kong S, van de Walle A, Turtle JA, Ben-Nun M, Riley S, Riley P, Koyluoglu U, DesRoches D, Forli P, Hamory B, Kyriakides C, Leis H, Milliken J, Moloney M, Morgan J, Nirgudkar N, Ozcan G, Piwonka N, Ravi M, Schrader C, Shakhnovich E, Siegel D, Spatz R, Stiefeling C, Wilkinson B, Wong A, Cavany S, España G, Moore S, Oidtman R, Perkins A, Kraus D, Kraus A, Gao Z, Bian J, Cao W, Lavista Ferres J, Li C, Liu T-Y, Xie X, Zhang S, Zheng S, Vespignani A, Chinazzi M, Davis JT, Mu K, Pastore y Piontti A, Xiong X, Zheng A, Baek J, Farias V, Georgescu A, Levi R, Sinha D, Wilde J, Perakis G, Bennouna MA, Nze-Ndong D, Singhvi D, Spantidakis I, Thayaparan L, Tsiourvas A, Sarker A, Jadbabaie A, Shah D, Della Penna N, Celi LA, Sundar S, Wolfinger R, Osthus D, Castro L, Fairchild G, Michaud I, Karlen D, Kinsey M,

Mullany LC, Rainwater-Lovett K, Shin L, Tallaksen K, Wilson S, Lee EC, Dent J, Grantz KH, Hill AL, Kaminsky J, Kaminsky K, Keegan LT, Lauer SA, Lemaitre JC, Lessler J, Meredith HR, Perez-Saez J, Shah S, Smith CP, Truelove SA, Wills J, Marshall M, Gardner L, Nixon K, Burant JC, Wang L, Gao L, Gu Z, Kim M, Li X, Wang G, Wang Y, Yu S, Reiner RC, Barber R, Gakidou E, Hay SI, Lim S, Murray C, Pigott D, Gurung HL, Baccam P, Stage SA, Suchoski BT, Prakash BA, Adhikari B, Cui J, Rodríguez A, Tabassum A, Xie J, Keskinocak P, Asplund J, Baxter A, Oruc BE, Serban N, Arik SO, Dusenberry M, Epshteyn A, Kanal E, Le LT, Li C-L, Pfister T, Sava D, Sinha R, Tsai T, Yoder N, Yoon J, Zhang L, Abbott S, Bosse NI, Funk S, Hellewell J, Meakin SR, Sherratt K, Zhou M, Kalantari R, Yamana TK, Pei S, Shaman J, Li ML, Bertsimas D, Skali Lami O, Soni S, Tazi Bouardi H, Ayer T, Adee M, Chhatwal J, Dalgic OO, Ladd MA, Linas BP, Mueller P, Xiao J, Wang Y, Wang Q, Xie S, Zeng D, Green A, Bien J, Brooks L, Hu AJ, Jahja M, McDonald D, Narasimhan B, Politsch C, Rajanala S, Rumack A, Simon N, Tibshirani RJ, Tibshirani R, Ventura V, Wasserman L, O'Dea EB, Drake JM, Pagano R, Tran QT, Ho LST, Huynh H, Walker JW, Slayton RB, Johansson MA, Biggerstaff M, Reich NG (2022). Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. In Proceedings of the National Academy of Sciences, 119(15):e2113561119

Czech L, Stamatakis A, Dunthorn M, Barbera P (2022). Metagenomic Analysis Using Phylogenetic Placement—A Review of the First Decade. Front. Bioinform. 2,871393

Daday C, de Buhr S, Mercadante D, Gräter F (2022). Mechanical force can enhance c-Src kinase activity by impairing autoinhibition. Biophys J,121(5):684-691

Dey D, Nunes-Alves A, Wade RC, Schreiber G (2022). Diffusion of small molecule drugs is affected by surface interactions and crowder proteins. iScience 25(10):105088

Disarlo V, Randecker A, Tang R (2022). Rigidity of the saddle connection complex. In Journal of Topology 15, 1248-1310

Dupuy A, Santamaría CA, Yeheskel A, Gräter F, Hogg PJ, Passam FH, Chiu J (2022). Mechano-redox control of Mac-1 de-adhesion from ICAM-1 by protein disulfide isomerase promotes directional movement of neutrophils under flow. bioRxiv 2022.03.29.486223

Eisenberg P, Albert L, Teuffel J, Zitzow E, Michaelis C, Jarick J, Sehlke C, Große L, Bader N, Nunes-Alves A, Kreikemeyer B, Schindelin H, Wade RC, Fiedler T (2022). The Non-phosphorylating Glyceraldehyde-3-Phosphate Dehydrogenase GapN Is a Potential New Drug Target in Streptococcus pyogenes. Front. Microbiol. 13,802427

Eisenstein L, Schulz B, Qadir GA, Pinto JG, Knippertz P (2022). Identification of high-wind features within extratropical cyclones using a probabilistic random forest - Part 1: Method and case studies. Weather and Climate Dynamics, 3(4):1157-1182

Eriksson O, Bhalla US, Blackwell KT, Crook SM, Keller D, Kramer A, Linne M-L, Saudargienė A, Wade RC, Hellgren Kotaleski J (2022). Combining hypothesis- and data-driven neuroscience modeling in FAIR workflows. eLife 11,e69013

Ernst M, Gryn'ova G (2022). Strength and Nature of Host-Guest Interactions in Metal-Organic Frameworks from a Quantum-Chemical Perspective. ChemPhysChem 23(8)

Franz F, Tapia-Rojo R, Winograd-Katz S, Boujemaa-Paterski R, Li W, Unger T, Albeck S, Aponte-Santamaría C, Garcia-Manyes S, Medalia O, Geiger B, Gräter F (2022). How talin allosterically activates vinculin. bioRxiv 2022.08.01.502287

Gneiting T, Walz E-M (2022). Receiver operating characteristic (ROC) movies, universal ROC (UROC) curves, and coefficient of predictive ability (CPA). Machine Learning, 111:2769-2797

Gneiting T, Vogel P (2022). Receiver operating characteristic (ROC) curves: Equivalences, beta model, and minimum distance estimation. Machine Learning, 111:2147-2159

Golebiewski M (2022). ISO 20691:2022 Biotechnology — Requirements for data formatting and description in the life sciences. https://www.iso.org/standard/68848.html

Gruber S, Löf A, Hausch A, Kutzki F, Jöhr R, Obser T, König G, Schneppenheim R, Aponte-Santamaría C, Gräter F, Brehm MA, Benoit M, Lipfert J (2022). A Conformational Transition of the D'D3 Domain primes von Willebrand Factor for Multimerization. Blood Adv,6(17):5198-5209

Guastoni A, Nestola F, Zorzi F, Lanza A, Ernst M, Gentile P, Andò S, Lorenzetti A (2022). Marchettiite, (NH 4)C 5 H 3 N 4 O 3 , a new organic mineral from Mount Cervandone, Devero Valley, Western-Central Alps, Italy. MinMag 86(6):966-974

Gómez-Flores CL, Maag D, Kansari M, Vuong V-Q, Irle S, Gräter F, Kubař T, Elstner M (2022). Accurate Free Energies for Complex Condensed-Phase Reactions Using an Artificial Neural Network Corrected DFTB/MM Methodology. J Chem Theory Comput,18(2):1213-1226

Haag J, Hübner L, Kozlov AM, Stamatakis A (2022). The Free Lunch is not over yet - Systematic Exploration of Numerical Thresholds in Phylogenetic Inference. bioRxiv 2022.07.13.499893

Haag J, Höhler D, Bettisworth B, Stamatakis A (2022). From Easy to Hopeless - Predicting the Difficulty of Phylogenetic Analyses. bioRxiv 2022.06.20.496790

Hamenstädt U, Viaggi G (2022). Small eigenvalues of random 3-manifolds. In Transactions of the American Mathematical Society 375: 3795-3840

Hernández Hernández J, Hruvsak M, Morales I, Randecker A, Sedano M, Valez F (2022). Conjugacy classes of big mapping class groups. In Journal of the London Mathematical Society, Series 106, no.~2, 1131-1169

Holas A, Koch CY, Leibold J, Prendi A, Schlachta TP, Sophia Schmid A, Schmitt L (2022). On the energy consumption of online and on-site lectures. Environ. Res. Commun. 4(6):061002

Hu SX, Li D, Stühmer J, Kim M, Hospedales TM (2022). Pushing the Limits of Simple Pipelines for Few-Shot Learning: External Data and Fine-Tuning Make a Difference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9068-9077

Jeon S, Strube M (2022). Entity-based Neural Local Coherence Modeling. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Dublin, Ireland, pp 7787–7805

Jordan AI, Mühlemann A, Ziegel JF (2022). Characterizing the optimal solutions to the isotonic regression problem for identifiable functionals. Annals of the Institute of Statistical Mathematics, 74:489-514

De Keer L, Van Steenberge P, Reyniers M-F, Gryn'ova G, Aitken HM, Coote ML (2022). New mechanism for autoxidation of polyolefins: kinetic Monte Carlo modelling of the role of short-chain branches, molecular oxygen and unsaturated moieties. Polym. Chem.,10.1039. D1PY01659H

Koetsenruijter J, Wronski P, Ghosh S, Müller W, Wensing M (2022). The Effect of an Additional Structured Methods Presentation on Decision-Makers' Reading Time and Opinions on the Helpfulness of the Methods in a Quantitative Report: Nonrandomized Trial. JMIR Med Inform. Apr 12;10(4):e29813

Kollasch F, Polsterer K (2022). Interactive Exploration Framework for Big Data Sets. Astronomical Data Analysis Software and Systems XXX. ASP Conference Series, Vol. 532, In Proceedings of a virtual conference held 8-12 November 2020. p.11

Kydonakis G, Sun H, Zhao L (2022). Monodromy of rank 2 parabolic Hitchin systems. In Journal of Geometry and Physics 171: 104411

Kydonakis G (2022). From hyperbolic Dehn filling to surgeries in representation varieties. In In the Tradition of Thurston II: 201-260

Lach F, Callan FP, Bubeck D, Röpke FK, Sim SA, Schrauth M, Ohlmann ST, Kromer M (2022a). Type Iax supernovae from deflagrations in Chandrasekhar mass white dwarfs. A&A 658:A179

Lach F, Callan FP, Sim SA, Röpke FK (2022b). Models of pulsationally assisted gravitationally confined detonations with different ignition conditions. A&A 659:A27

Laplace E (2022). TULIPS: A Tool for Understanding the Lives, Interiors, and Physics of Stars. Astronomy and Computing 38:100516

Leidi G, Birke C, Andrassy R, Higl J, Edelmann PVF, Wiest G, Klingenberg C, Röpke FK (2022). A finite-volume scheme for modeling compressible magnetohydrodynamic flows at low Mach numbers in stellar interiors. A&A 668:A143

Lerch S, Polsterer K (2022). Convolutional autoencoders for spatially-informed ensemble post-processing. arXiv:2204.05102

Liu Z-W, Röpke FK, Zeng Y (2022). Signatures of a Surviving Helium-star Companion in Type Ia Supernovae and Constraints on the Progenitor Companion of SN 2011fe. ApJ 928(2):146

Lösel PD, Monchanin C, Lebrun R, Jayme A, Relle J, Devaud J-M, Heuveline V, Lihoreau M (2022). Natural variability in bee brain size and symmetry revealed by micro-CT imaging and deep learning. bioRxiv 2022.10.12.511944

Maloni S, Pozzetti B (2022). Geometric limits of cyclic subgroups of SO(1,k+1) and SU(1,k+1). In Algebraic & Geometric Topology 22: 1461-1495

Maret A (2022). Ergodicity of the mapping class group action on super-maximal representations. In Groups Geometry and Dynamics 16: 1341-1368

Martin IM, Nava MM, Wickström SA, Gräter F (2022). ATP allosterically stabilizes integrin-linked kinase for efficient force generation. PNAS,119 (11) e2106098119

Martin IM, Aponte-Santamaría C, Schmidt L, Hedtfeld M, Iusupov A, Musacchio A, Gräter F (2022). Phosphorylation tunes elongation propensity and cohesiveness of INCENP's intrinsically disordered region. J Mol Biol,434(1):167387

Masur H, Rafi K, Randecker A (2022). Expected covering radius of a translation surface. In International Mathematics Research Notices 10, 7967-8002

Mayer D, Lever F, Picconi D, Metje J, Alisauskas S, Calegari F, Düsterer S, Ehlert C, Feifel R, Niebuhr M, Manschwetus B, Kuhlmann M, Mazza T, Robinson MS, Squibb RJ, Trabattoni A, Wallner M, Saalfrank P, Wolf TJA, Gühr M (2022). Following excited-state chemical shifts in molecular ultrafast x-ray photoelectron spectroscopy. Nat Commun 13(1),198

Mendes MF de A, de Souza Bragatte M, Vianna P, de Freitas MV, Pöhner I, Richter S, Wade RC, Salzano FM, Vieira GF (2022). MatchTope: A tool to predict the cross reactivity of peptides complexed with Major Histocompatibility Complex I. Front. Immunol. 13,930590

Miller R, Ajello M, Auchettl K, Beacom J, Bloser P, Burrows A, Frebel A, Fryer C, Hartmann D, Hoeflich P, Hungerford A, Leising M, Lopez L, Milne P, Peplowski P, Roepke F, Scolnic D, Seitenzahl I, The L, Young C (2022). The Lunar Occultation eXplorer (LOX): MeV Gamma-Ray Astrophysics Across Space and Time. In AAS/High Energy Astrophysics Division, vol. 54 of AAS/High Energy Astrophysics Division, pp. 108.58

Miller R, Ajello M, Auchettl K, Beacom J, Bloser P, Burrows A, Frebel A, Fryer C, Hartmann D, Hoeflich P, Hungerford A, Leising M, Lopez L, Milne P, Peplowski P, Roepke F, Scolnic D, Seitenzahl I, The L-S, Young CA, LOX Team (2022). Ex Luna, Scientia: The Lunar Occultation eXplorer (LOX) and the Future of MeV γ-Ray Astrophysics. In APS April Meeting Abstracts, vol. of APS Meeting Abstracts, p S13.005

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 (2022). SPART: A versatile and standardized data exchange format for species partition information. Molecular Ecology Resources 22(1):430-438

Moreno MM, Schneider FRN, Röpke FK, Ohlmann ST, Pakmor R, Podsiadlowski P, Sand C (2022). From 3D hydrodynamic simulations of common-envelope interaction to gravitational-wave mergers. A&A 667:A72

Muñiz-Chicharro A, Votapka LW, Amaro RE, Wade RC (2022) Brownian dynamics simulations of biomolecular diffusional association processes. WIREs Computational Molecular Science n/a:e1649

Møller P, Seppälä T, Dowty JG, Haupt S, Dominguez-Valentin M, Sunde L, Bernstein I, Engel C, Aretz S, Nielsen M, Capella G, Evans DG, Burn J, Holinski-Feder E, Bertario L, Bonanni B, Lindblom A, Levi Z, Macrae F, Winship I, Plazzer J-P, Sijmons R, Laghi L, Valle AD, Heinimann K, Half E, Lopez-Koestner F, Alvarez-Valenzuela K, Scott RJ, Katz L, Laish I, Vainer E, Vaccaro CA, Carraro DM, Gluck N, Abu-Freha N, Stakelum A, Kennelly R, Winter D, Rossi BM, Greenblatt M, Bohorquez M, Sheth H, Tibiletti MG, Lino-Silva LS, Horisberger K, Portenkirchner C, Nascimento I, Rossi NT, da Silva LA, Thomas H, Zaránd A, Mecklin J-P, Pylvänäinen K, Renkonen-Sinisalo L, Lepisto A, Peltomäki P, Therkildsen C, Lindberg LJ, Thorlacius-Ussing O, von Knebel Doeberitz M, Loeffler M, Rahner N, Steinke-Lange V, Schmiegel W, Vangala D, Perne C, Hüneburg R, de Vargas AF, Latchford A, Gerdes A-M, Backman A-S, Guillén-Ponce C, Snyder C, Lautrup CK, Amor D, Palmero E, Stoffel E, Duijkers F, Hall MJ, Hampel H, Williams H, Okkels H, Lubiński J, Reece J, Ngeow J, Guillem JG, Arnold J, Wadt K, Monahan K, Senter L, Rasmussen LJ, van Hest LP, Ricciardiello L, Kohonen-Corish MRJ, Ligtenberg MJL, Southey M, Aronson M, Zahary MN, Samadder NJ, Poplawski N, Hoogerbrugge N, Morrison PJ, James P, Lee G, Chen-Shtoyerman R, Ankathil R, Pai R, Ward R, Parry S, Dębniak T, John T, van Overeem Hansen T, Caldés T, Yamaguchi T, Barca-Tierno V, Garre P, Cavestro GM, Weitz J, Redler S, Büttner R, Heuveline V, Hopper JL, Win AK, Lindor N, Gallinger S, Le Marchand L, Newcomb PA, Figueiredo J, Buchanan DD, Thibodeau SN, ten Broeke SW, Hovig E, Nakken S, Pineda M, Dueñas N, Brunet J, Green K, Lalloo F, Newton K, Crosbie EJ, Mints M, Tjandra D, Neffa F, Esperon P, Kariv R, Rosner G, Pavicic WH, Kalfayan P, Torrezan GT, Bassaneze T, Martin C, Moslein G, Ahadova A, Kloor M, Sampson JR, Jenkins MA (2022). Colorectal cancer incidences in Lynch syndrome: a comparison of results from the prospective lynch syndrome database and the international mismatch repair consortium. Hered Cancer Clin Pract 20(1),36

Weidemann A, Dudas D, Rey M, Wittig U, Müller W (2022). More findability, more interoperability for SABIO-RK, the curated database for biochemical reaction kinetics.

Müller MC (2022). A proposal for explicit word formation annotation in discourse corpora. Book of Abstracts of the Symposium on Word Formation and Discourse Structure, Leipzig, Germany, May, pp14-15

Negri MM, Fortuin V, Stühmer J (2022). Meta-learning richer priors for VAEs. In Fourth Symposium on Advances in Approximate Bayesian Inference

Niarakis A, Ostaszewski M, Mazein A, Kuperstein I, Kutmon M, Gillespie ME, Funahashi A, Acencio ML, Hemedan A, Aichem M, Klein K, Czauderna T, Burtscher F, Yamada TG, Hiki Y, Hiroi NF, Hu F, Pham N, Ehrhart F, Willighagen EL, Valdeolivas A, Dugourd A, Messina F, Esteban-Medina M, Peña-Chilet M, Rian K, Soliman S, Aghamiri SS, Puniya BL, Naldi A, Helikar T, Singh V, Fernández MF, Bermudez V, Tsirvouli E, Montagud A, Noël V, de Leon MP, Maier D, Bauch A, Gyori BM, Bachman JA, Luna A, Pinero J, Furlong LI, Balaur I, Rougny A, Jarosz Y, Overall RW, Phair R, Perfetto L, Matthews L, Rex DAB, Orlic-Milacic M, Cristobal MGL, De Meulder B, Ravel JM, Jassal B, Satagopam V, Wu G, Golebiewski M, Gawron P, Calzone L, Beckmann JS, Evelo CT, D'Eustachio P, Schreiber F, Saez-Rodriguez J, Dopazo J, Kuiper M, Valencia A, Wolkenhauer O, Kitano H, Barillot E, Auffray C, Balling R, Schneider R (2022). A versatile and interoperable computational framework for the analysis and modeling of COVID-19 disease mechanisms. bioRxiv 2022.12.17.520865

Ondratschek PA, Röpke FK, Schneider FRN, Fendt C, Sand C, Ohlmann ST, Pakmor R, Springel V (2022). Bipolar planetary nebulae from common-envelope evolution of binary stars. A&A 660:L8

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

Pakmor R, Callan FP, Collins CE, de Mink SE, Holas A, Kerzendorf WE, Kromer M, Neunteufel PG, O'Brien JT, Röpke FK, Ruiter AJ, Seitenzahl IR, Shingles LJ, Sim SA, Taubenberger S (2022). On the fate of the secondary white dwarf in double-degenerate double-detonation Type Ia supernovae. Monthly Notices of the Royal Astronomical Society 517(4):5260-5271

Panecka-Hofman J, Poehner I, Wade RC (2022). Anti-trypanosomatid structure-based drug design – lessons learned from targeting the folate pathway. Expert Opinion on Drug Discovery 17(9):1029-1045

Plier J, Zisler M, Furkel J, Knoll M, Marx A, Fischer A, Polsterer K, Konstandin MH, Petra S (2022). Learning Features via Transformer Networks for Cardiomyocyte Profiling. Bildverarbeitung für die Medizin,pp.167-172,Springer Fachmedien Wiesbaden

Poręba T, Macchi P, Ernst M (2022). Pitfalls in the location of guest molecules in metal-organic frameworks. Nat Commun 13(1),5288

Poręba T, Świątkowski M, Ernst M, Confalonieri G (2022). Premelting Anomalies in Pyromellitic Dianhydride: Negative Thermal Expansion, Accelerated Radiation Damage, and Polymorphic Phase Transition. J. Phys. Chem. C 126(17):7648-7659

Pöhner I, Quotadamo A, Panecka-Hofman J, Luciani R, Santucci M, Linciano P, Landi G, Di Pisa F, Dello Iacono L, Pozzi C, Mangani S, Gul S, Witt G, Ellinger B, Kuzikov M, Santarem N, Cordeiro-da-Silva A, Costi MP, Venturelli A, Wade RC (2022). Multitarget, Selective Compound Design Yields Potent Inhibitors of a Kinetoplastid Pteridine Reductase 1. J. Med. Chem. 65(13):9011-9033

Rennekamp B, Karfusehr C, Kurth M, Ünal A, Riedmiller K, Gryn'ova G, Hudson DM, Gräter F (2022). Collagen breaks at weak sacrificial bonds taming its mechanoradicals. bioRxiv 2022.10.17.512491

Ghosh S, Mueller W, Wittig U, Rey M (2022). Towards Effortless Navigation of Scientific-Literature Screen Reading for Biocuration. In Proceedings of 1st UK-Local Biocuration Conference International Society for Biocuration. 10.1093/database/baac027. P.21

Rogdakis T, Charou D, Latorrata A, Papadimitriou E, Tsengenes A, Athanasiou C, Papadopoulou M, Chalikiopoulou C, Katsila T, Ramos I, Prousis KC, Wade RC, Sidiropoulou K, Calogeropoulou T, Gravanis A, Charalampopoulos I (2022). Development and Biological Characterization of a Novel Selective TrkA Agonist with Neuroprotective Properties against Amyloid Toxicity. Biomedicines 10(3):614

Rusnati M, Paiardi G, Tobia C, Urbinati C, Lodola A, D'Ursi P, Corrado M, Castelli R, Wade RC, Tognolini M, Chiodelli P (2022). Cholenic acid derivative UniPR1331 impairs tumor angiogenesis via blockade of VEGF/VEGFR2 in addition to Eph/ephrin. Cancer Gene Ther 29(7):908-917

Martins dos Santos V, Anton M, Szomolay B, Ostaszewski M, Arts I, Benfeitas R, Dominguez Del Angel V, Ferk P, Fey D, Goble C, Golebiewski M, Gruden K, Heil KF, Hermjakob H, Kahlem P, Klapa MI, Koehorst J, Kolodkin A, Kutmon M, Leskošek B, Moretti S, Müller W, Pagni M, Rezen T, Rocha M, Rozman D, Šafránek D, Sheriff RSM, Suarez Diez M, Van Steen K, Westerhoff HV, Wittig U, Wolstencroft K, Zupanic A, Evelo CT, Hancock JM (2022). Systems Biology in ELIXIR: modelling in the spotlight. F1000Res 11:1265

Schatz H, Becerril Reyes AD, Best A, Brown EF, Chatziioannou K, Chipps KA, Deibel CM, Ezzeddine R, Galloway DK, Hansen CJ, Herwig F, Ji AP, Lugaro M, Meisel Z, Norman D, Read JS, Roberts LF, Spyrou A, Tews I, Timmes FX, Travaglio C, Vassh N, Abia C, Adsley P, Agarwal S, Aliotta M, Aoki W, Arcones A, Aryan A, Bandyopadhyay A, Banu A, Bardayan DW, Barnes J, Bauswein A, Beers TC, Bishop J, Boztepe T, Côté B, Caplan ME, Champagne AE, Clark JA, Couder M, Couture A, de Mink SE, Debnath S, deBoer RJ, den Hartogh J, Denissenkov P, Dexheimer V, Dillmann I, Escher JE, Famiano MA, Farmer R, Fisher R, Fröhlich C, Frebel A, Fryer C, Fuller G, Ganguly AK, Ghosh S, Gibson BK, Gorda T, Gourgouliatos KN, Graber V, Gupta M, Haxton WC, Heger A, Hix WR, Ho WCG, Holmbeck EM, Hood AA, Huth S, Imbriani G, Izzard RG, Jain R, Jayatissa H, Johnston Z, Kajino T, Kankainen A, Kiss GG, Kwiatkowski A, La Cognata M, Laird AM, Lamia L, Landry P, Laplace E, Launey KD, Leahy D, Leckenby G, Lennarz A, Longfellow B, Lovell AE, Lynch WG, Lyons SM, Maeda K, Masha E, Matei C, Merc J, Messer B, Montes F, Mukherjee A, Mumpower MR, Neto D, Nevins B, Newton WG, Nguyen LQ, Nishikawa K, Nishimura N, Nunes FM, O'Connor E, O'Shea BW, Ong W-J, Pain SD, Pajkos MA, Pignatari M, Pizzone RG, Placco VM, Plewa T, Pritychenko B, Psaltis A, Puentes D, Qian Y-Z, Radice D, Rapagnani D, Rebeiro BM, Reifarth R, Richard AL, Rijal N, Roederer IU, Rojo JS, J S K, Saito Y, Schwenk A, Sergi ML, Sidhu RS, Simon A, Sivarani T, Skúladóttir Á, Smith MS, Spiridon A, Sprouse TM, Starrfield S, Steiner AW, Strieder F, Sultana I, Surman R, Szücs T, Tawfik A, Thielemann F, Trache L, Trappitsch R, Tsang MB, Tumino A, Upadhyayula S, Valle Martínez JO, Van der Swaelmen M, Viscasillas Vázquez C, Watts A, Wehmeyer B, Wiescher M, Wrede C, Yoon J, Zegers RGT, Zermane MA, Zingale M (2022). Horizons: nuclear astrophysics in the 2020s and beyond. J. Phys. G: Nucl. Part. Phys. 49(11):110502

Schmidt H, Mauer K, Glaser M, Dezfuli BS, Hellmann SL, Silva Gomes AL, Butter F, Wade RC, Hankeln T, Herlyn H (2022). Identification of antiparasitic drug targets using a multi-omics workflow in the acanthocephalan model. BMC Genomics 23(1),677

Schulz B, Lerch S (2022). Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison. Monthly Weather Review,150(1):235-257

Shenar T, Sana H, Mahy L, Maíz Apellániz J, Crowther PA, Gromadzki M, Herrero A, Langer N, Marchant P, Schneider FRN, Sen K, Soszyński I, Toonen S (2022). The Tarantula Massive Binary Monitoring. VI. Characterisation of hidden companions in 51 single-lined O-type binaries: A flat mass-ratio distribution and black-hole binary candidates. A&A 665:A148

Shenar T, Sana H, Mahy L, El-Badry K, Marchant P, Langer N, Hawcroft C, Fabry M, Sen K, Almeida LA, Abdul-Masih M, Bodensteiner J, Crowther PA, Gieles M, Gromadzki M, Hénault-Brunet V, Herrero A, Koter A de, Iwanek P, Kozłowski S, Lennon DJ, Apellániz JM, Mróz P, Moffat AFJ, Picco A, Pietrukowicz P, Poleski R, Rybicki K, Schneider FRN, Skowron DM, Skowron J, Soszyński I, Szymański MK, Toonen S, Udalski A, Ulaczyk K, Vink JS, Wrona M (2022). An X-ray-quiet black hole born with a negligible kick in a massive binary within the Large Magellanic Cloud. Nat Astron 6(9):1085-1092

Shingles LJ, Flörs A, Sim SA, Collins CE, Röpke FK, Seitenzahl IR, Shen KJ (2022). Modelling the ionization state of Type Ia supernovae in the nebular phase. Monthly Notices of the Royal Astronomical Society, 512(4):6150-6163

Sigmund LM, Ehlert C, Gryn'ova G, Greb L (2022). Stereoinversion of tetrahedral p-block element hydrides. J. Chem. Phys. 156(19):194113

Sirazitdinov A, Buchwald M, Hesser J, Heuveline V (2022). Review of Deep Learning Methods for Individual Treatment Effect Estimation with Automatic Hyperparameter Optimization. TechRxiv. Preprint.

Soultanis T, Bauswein A, Stergioulas N (2022). Analytic models of the spectral properties of gravitational waves from neutron star merger remnants. Phys. Rev. D 105(4),043020

Srivastava A, Rastogi A, Rao A, Shoeb AAM, Abid A, Fisch A, Brown AR, Santoro A, Gupta A, Garriga-Alonso A, Kluska A, Lewkowycz A, Agarwal A, Power A, Ray A, Warstadt A, Kocurek AW, Safaya A, Tazarv A, Xiang A, Parrish A, Nie A, Hussain A, Askell A, Dsouza A, Slone A, Rahane A, Iyer AS, Andreassen A, Madotto A, Santilli A, Stuhlmüller A, Dai A, La A, Lampinen A, Zou A, Jiang A, Chen A, Vuong A, Gupta A, Gottardi A, Norelli A, Venkatesh A, Gholamidavoodi A, Tabassum A, Menezes A, Kirubarajan A, Mullokandov A, Sabharwal A, Herrick A, Efrat A, Erdem A, Karakaş A, Roberts BR, Loe BS, Zoph B, Bojanowski B, Özyurt B, Hedayatnia B, Neyshabur B, Inden B, Stein B, Ekmekci B, Lin BY, Howald B, Diao C, Dour C, Stinson C, Argueta C, Ramírez CF, Singh C, Rathkopf C, Meng C, Baral C, Wu C, Callison-Burch C, Waites C, Voigt C, Manning CD, Potts C, Ramirez C, Rivera CE, Siro C, Raffel C, Ashcraft C, Garbacea C, Sileo D, Garrette D, Hendrycks D, Kilman D, Roth D, Freeman D, Khashabi D, Levy D, González DM, Perszyk D, Hernandez D, Chen D, Ippolito D, Gilboa D, Dohan D, Drakard D, Jurgens D, Datta D, Ganguli D, Emelin D, Kleyko D, Yuret D, Chen D, Tam D, Hupkes D, Misra D, Buzan D, Mollo DC, Yang D, Lee D-H, Shutova E, Cubuk ED, Segal E, Hagerman E, Barnes E, Donoway E, Pavlick E, Rodola E, Lam E, Chu E, Tang E, Erdem E, Chang E, Chi EA, Dyer E, Jerzak E, Kim E, Manyasi EE, Zheltonozhskii E, Xia F, Siar F, Martínez-Plumed F, Happé F, Chollet F, Rong F, Mishra G, Winata GI, de Melo G, Kruszewski G, Parascandolo G, Mariani G, Wang G, Jaimovitch-López G, Betz G, Gur-Ari G, Galijasevic H, Kim H, Rashkin H, Hajishirzi H, Mehta H, Bogar H, Shevlin H, Schütze H, Yakura H, Zhang H, Wong HM, Ng I, Noble I, Jumelet J, Geissinger J, Kernion J, Hilton J, Lee J, Fisac JF, Simon JB, Koppel J, Zheng J, Zou J, Kocoń J, Thompson J, Kaplan J, Radom J, Sohl-Dickstein J, Phang J, Wei J, Yosinski J, Novikova J, Bosscher J, Marsh J, Kim J, Taal J, Engel J, Alabi J, Xu J, Song J, Tang J, Waweru J, Burden J, Miller J, Balis JU, Berant J, Frohberg J, Rozen J, Hernandez-Orallo J, Boudeman J, Jones J, Tenenbaum JB, Rule JS, Chua J, Kanclerz K, Livescu K, Krauth K, Gopalakrishnan K, Ignatyeva K, Markert K, Dhole KD, Gimpel K, Omondi K, Mathewson K, Chiafullo K, Shkaruta K, Shridhar K,

McDonell K, Richardson K, Reynolds L, Gao L, Zhang L, Dugan L, Qin L, Contreras-Ochando L, Morency L-P, Moschella L, Lam L, Noble L, Schmidt L, He L, Colón LO, Metz L, Şenel LK, Bosma M, Sap M, ter Hoeve M, Farooqi M, Faruqui M, Mazeika M, Baturan M, Marelli M, Maru M, Quintana MJR, Tolkiehn M, Giulianelli M, Lewis M, Potthast M, Leavitt ML, Hagen M, Schubert M, Baitemirova MO, Arnaud M, McElrath M, Yee MA, Cohen M, Gu M, Ivanitskiy M, Starritt M, Strube M, Swędrowski M, Bevilacqua M, Yasunaga M, Kale M, Cain M, Xu M, Suzgun M, Tiwari M, Bansal M, Aminnaseri M, Geva M, Gheini M, T MV, Peng N, Chi N, Lee N, Krakover NG-A, Cameron N, Roberts N, Doiron N, Nangia N, Deckers N, Muennighoff N, Keskar NS, Iyer NS, Constant N, Fiedel N, Wen N, Zhang O, Agha O, Elbaghdadi O, Levy O, Evans O, Casares PAM, Doshi P, Fung P, Liang PP, Vicol P, Alipoormolabashi P, Liao P, Liang P, Chang P, Eckersley P, Htut PM, Hwang P, Miłkowski P, Patil P, Pezeshkpour P, Oli P, Mei Q, Lyu Q, Chen Q, Banjade R, Rudolph RE, Gabriel R, Habacker R, Delgado RR, Millière R, Garg R, Barnes R, Saurous RA, Arakawa R, Raymaekers R, Frank R, Sikand R, Novak R, Sitelew R, LeBras R, Liu R, Jacobs R, Zhang R, Salakhutdinov R, Chi R, Lee R, Stovall R, Teehan R, Yang R, Singh S, Mohammad SM, Anand S, Dillavou S, Shleifer S, Wiseman S, Gruetter S, Bowman SR, Schoenholz SS, Han S, Kwatra S, Rous SA, Ghazarian S, Ghosh S, Casey S, Bischoff S, Gehrmann S, Schuster S, Sadeghi S, Hamdan S, Zhou S, Srivastava S, Shi S, Singh S, Asaadi S, Gu SS, Pachchigar S, Toshniwal S, Upadhyay S, Shyamolima, Debnath, Shakeri S, Thormeyer S, Melzi S, Reddy S, Makini SP, Lee S-H, Torene S, Hatwar S, Dehaene S, Divic S, Ermon S, Biderman S, Lin S, Prasad S, Piantadosi ST, Shieber SM, Misherghi S, Kiritchenko S, Mishra S, Linzen T, Schuster T, Li T, Yu T, Ali T, Hashimoto T, Wu T-L, Desbordes T, Rothschild T, Phan T, Wang T, Nkinyili T, Schick T, Kornev T, Telleen-Lawton T, Tunduny T, Gerstenberg T, Chang T, Neeraj T, Khot T, Shultz T, Shaham U, Misra V, Demberg V, Nyamai V, Raunak V, Ramasesh V, Prabhu VU, Padmakumar V, Srikumar V, Fedus W, Saunders W, Zhang W, Vossen W, Ren X, Tong X, Zhao X, Wu X, Shen X, Yaghoobzadeh Y, Lakretz Y, Song Y, Bahri Y, Choi Y, Yang Y, Hao Y, Chen Y, Belinkov Y, Hou Y, Hou Y, Bai Y, Seid Z, Zhao Z, Wang Z, Wang ZJ, Wang Z, Wu Z (2022). Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. arXiv:2206.04615

Stegmann J, Antonini F, Schneider FRN, Tiwari V, Chattopadhyay D (2022). Binary black hole mergers from merged stars in the Galactic field. Phys. Rev. D 106(2),023014

Tetenoire A, Ehlert C, Juaristi JI, Saalfrank P, Alducin M (2022). Why Ultrafast Photoinduced CO Desorption Dominates over Oxidation on Ru(0001). J. Phys. Chem. Lett. 13(36):8516-8521

Toral-Lopez A, Kokh DB, Marin EG, Wade RC, Godoy A (2022). Graphene BioFET sensors for SARS-CoV-2 detection: a multiscale simulation approach. Nanoscale Adv. 4(14):3065-3072

Treyde W, Riedmiller K, Gräter F (2022). Bond dissociation energies of X–H bonds in proteins. RSC Advances,12(53):34557-34564

Vaduvescu O, Aznar Macias A, Wilson TG, Zegmott T, Pérez Toledo FM, Predatu M, Gherase R, Pinter V, Pozo Nunez F, Ulaczyk K, Soszyński I, Mróz P, Wrona M, Iwanek P, Szymanski M, Udalski A, Char F, Salas Olave H, Aravena-Rojas G, Vergara AC, Saez C, Unda-Sanzana E, Alcalde B, de Burgos A, Nespral D, Galera-Rosillo R, Amos NJ, Hibbert J, López-Comazzi A, Oey J, Serra-Ricart M, Licandro J, Popescu M (2022). The EURONEAR Lightcurve Survey of Near Earth Asteroids 2017-2020. Earth Moon Planets 126(2),6

Vargas Pallete F, Farre J (2022). Minimal area surfaces and fibered hyperbolic 3-manifolds. In Proceedings of the American Mathematical Society, 150: 4931-4946

Wielgórski P, Pietrzyński G, Pilecki B, Gieren W, Zgirski B, Górski M, Hajdu G, Narloch W, Karczmarek P, Smolec R, Kervella P, Storm J, Gallenne A, Breuval L, Lewis M, Kałuszyński M, Graczyk D, Pych W, Suchomska K, Taormina M, Rojas Garcia G, Kotek A, Chini R, Pozo Nũnez F, Noroozi S, Sobrino Figaredo C, Haas M, Hodapp K, Mikołajczyk P, Kotysz K, Moździerski D, Kołaczek-Szymański P (2022). An Absolute Calibration of the Near-infrared Period–Luminosity Relations of Type II Cepheids in the Milky Way and in the Large Magellanic Cloud. ApJ 927(1):89

Witt J, Haupt S, Ahadova A, Bohaumilitzky L, Fuchs V, Ballhausen A, Przybilla MJ, Jendrusch M, Seppälä TT, Fürst D, Walle T, Busch E, Haag GM, Hüneburg R, Nattermann J, von Knebel Doeberitz M, Heuveline V, Kloor M (2022). A simple approach for detecting HLA-A*02 alleles in archival formalin-fixed paraffin-embedded tissue samples and an application example for studying cancer immunoediting. HLA,101(1):24-33

Yu J, Khosla S, Manuvinakurike R, Levin L, Ng V, Poesio M, Strube M, Rosé C (eds) (2022). Proceedings of the CODI-CRAC 2022 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue. Association for Computational Linguistics, Gyeongju, Republic of Korea

Yu J, Khosla S, Manuvinakurike R, Levin L, Ng V, Poesio M, Strube M, Rosé C (2022). The CODI-CRAC 2022 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue. In Proceedings of the CODI-CRAC 2022 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue. Association for Computational Linguistics, Gyeongju, Republic of Korea, October, pp. 1–14

Zaunseder E, Haupt S, Mütze U, Garbade SF, Kölker S, Heuveline V (2022). Opportunities and challenges in machine learning-based newborn screening—A systematic literature review. JIMD Reports 63(3):250-261

Zeng Y, Liu Z-W, Heger A, McCully C, Röpke FK, Han Z (2022). Long-term Evolution of Postexplosion Helium-star Companions of Type Iax Supernovae. ApJ 933(1):65

Zhao W, Eger S (2022). Constrained density matching and modeling for cross-lingual alignment of contextualized representations. In Proceedings of Machine Learning Research, Asian Conference on Machine Learning, Hyderabad, India, 12-14 December

Zhao W, Mathews K, Chai H (2022). Improving coreference resolution with word formation. Book of Abstracts of the Symposium on Word Formation and Discourse Structure, Leipzig, Germany, May 2022, pp.16-17

9 Teaching

Degrees

Sohebullah Abdi:
"Erweiterung eines Intrusion-Detection-Systems mit Supervised-Machine-Learning-Algorithmen für die Netzwerksicherheit", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Leonie Boland:
"Template Matching using CNN and Hypernetwork Hybrid Model", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Teresa Braun:
"Red-giant stars with low period spacings" Master's thesis, Center for Astronomy (ZAH), Heidelberg University and HITS: Saskia Hekker (2022).

Simon Cello:
"Investigating biological radicals using SQUID magnetometry on the example of the DOPA radical", Bachelor's thesis, Department of Physics and Astronomy, Heidelberg University, and HITS: Frauke Gräter (2022).

Ainara Claveras Cabezudo:
"Structural Insights into Neurotrophin Receptor Dimerization using Coarse-grained Molecular Dynamics Simulations", Master's thesis, Molecular Biotechnology, Faculty of Biosciences, Heidelberg University and HITS: Christina Athanasiou and Rebecca C. Wade (2022).

Gül Hanife Durmus:
"Vergleich mathematischer Modelle zur SARS-CoV-2 Pandemieentwicklung", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Michelle Emmert:
"Generation and Testing of Boltzmann Generators Trained on Peptide Molecular Dynamics Simulations to Predict Peptide Conformational Behaviour", Bachelor's thesis, Molecular Biotechnology, Faculty of Bioscience, Heidelberg University and HITS: Manuel Glaser and Rebecca Wade (2022).

Kira Feldmann:
"Aspects of spatial postprocessing for global temperature forecasts", Ph.D. thesis, School of Business Informatics and Mathematics, University of Mannheim: Martin Schlather and HITS: Tilmann Gneiting (2022).

Manuel Hammer:
"Security Analyses of Vulnerable Hosts : Concepts ands Methods for Detecting and Mitigating Log4Shell", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Anton Hanke:
"P. falciparum circumsporozoite protein epitope/paratope interactions: Addressing the key role of molecular flexibility", Master's thesis, Molecular Biotechnology, Faculty of Biosciences, Heidelberg University and HITS: Giulia D'Arrigo and Rebecca C. Wade (2022).

Larry Harbrecht:
"Functionality and limitation of DPI circumvention software", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Tassia Heuser:
"Polynomielle Chaosentwicklung zur Schätzung unsicherer Parameter durch Bayessche Inferenz am Beispiel der Boussinesq-Approximation", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Rosa Huisinga:
"Improvement of the current methods of mechanoradical-measurements in rat tail tendons", Bachelor's thesis, Fakultät für Chemie und Geowissenschaften, Heidelberg University, and HITS: Frauke Gräter (2022).

Florian Lach:
"Chandrasekhar-mass explosions as a model for Type Ia supernovae and their contribution to cosmic nucleosynthesis" Ph.D. thesis, Fakultät für Physik und Astronomie, Heidelberg University and HITS: Friedrich Röpke (2022).

Philipp Lösel:
"GPU-basierte Verfahren zur Segmentierung biomedizinischer Bilddaten" Ph.D. thesis, Heidelberg University and HITS: Vincent Heuveline (2022).

Federico López:
"Learning Neural Graph Representations in Non-Eucledian Geometries", Ph.D. thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2022.

Stefan Machmeier:
"Honeypot Implementation in a Cloud Environment", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022)

Arnaud Maret:
"The symplectic geometry of surface group representations in genus zero", Ph.D. thesis, Mathematisches Institut, Heidelberg University: Anna Wienhard and Peter Albers (2022).

Isabel Martin:
"Exploring Mechanical Signaling at Cellular Force Transduction Hubs using Molecular Simulations", PhD thesis, Combined Faculty of Mathematics, Engineering and Natural Sciences, Heidelberg University and HITS: Frauke Gräter, 2022.

Tina Neumann:
"Probing merger methods in a 9 and 8 solar mass binary star system", Bachelor's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Fabian Schneider (2022).

Eric Ommert:
"Löwner-John ellipsoids", Bachelor's thesis, Mathematisches Institut, Heidelberg University: Maria Beatrice Pozzetti (2022).

Alba Covelo Paz:
"Light curve extraction of red giants in NGC 6791 and NGC 6819" Master's thesis , Center for Astronomy (ZAH), Heidelberg University and HITS: Saskia Hekker (2022).

Lea Reisinger:
"Using Deep Packet Inspection to Analyse and Reduce DDoS Attacks on Servers and Applications", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Julian Schlecker:
"Rotation of the core of red-giant branch stars measured through mixed oscillation modes" Master's thesis, Center for Astronomy (ZAH), Heidelberg University and HITS: Saskia Hekker (2022).

Janik Schmid:
"Deep Learning für Parameteridentifizierung von Konvektionsproblemen mit PINNs", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Marco Schröder:
"VTable-Hijacking", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Fenja Schweder:
"Probabilistische Modellierung heterogen gesampelter Zeitreihen für die explorative Suche in großen astronomischen Archiven", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Kai Polsterer, Vincent Heuveline (2022).

Antonia Seifert:
"CAT(0) spaces and Gromov's condition", Bachelor's thesis, Mathematisches Institut, Heidelberg University: Maria Beatrice Pozzetti (2022).

Theodoros Soultanis:
"Spectral properties of gravitational waves from neutron star merger remnants" Ph.D. thesis, Fakultät für Physik und Astronomie, Heidelberg University and HITS: Andreas Bauswein (2022).

Björn-Uwe Steinorth:
"Ermittlung und Auswertung von CPU-Nutzungswerten eines HPC-Clusters aus InfluxDB-Daten für die Blaue Engel Zertifizierung", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2022).

Christoph Stelz:
"Core-Count Independent Reproducible Reduce", Bachelor's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis, 2022.

Marcel Stoklasa:
"Limit sets for once punctured torus groups acting on the Riemann sphere" Bachelor's thesis, Mathematisches Institut. Heidelberg University: Maria Beatrice Pozzetti with Gabriele Viaggi (2022).

Jonathan Teuffel:
"Molecular Dynamics Simulation of a Mammalian Cytochrome P450 Protein and its Interaction with Substrates and Redox-Partners", Master's thesis, Biochemistry, Faculty of Biosciences and Faculty of Chemistry and Earth Sciences, Heidelberg University and HITS: Goutam Mukherjee and Rebecca C. Wade (2022).

Cornelius Zenkert:
"Dimensionen fraktaler Schwämme, Methode der Konstruktion, Berechnung und Visualisierung", Bachelor's thesis, Mathematisches Institut, Heidelberg University: Maria Beatrice Pozzetti (2022).

Xinyi Zhang:
"Investigating the relationship between tree distance and model distance", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis, (2022).

Lectures, courses and seminars

Camilo Aponte-Santamaria, Florian Heigwer, Frauke Gräter, Michael Boutros:
Lecture and Practicals on *"Data Science and Simulations"* within Matter to Life Master program, WS 2021/22.

Giulia D'Arrigo:
Lecture on *"Modeling and simulation of biomolecular interactions"*, M.Sc. Molecular & Cellular Biology (Module 4 - Special Topic Series), Faculty of Biosciences, Heidelberg University, 9 December 2022.

Alain Becam, Maja Rey, Andreas Weidemann, Ulrike Wittig:
de.NBI Course *"Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOMHub/SEEK"*, Heidelberg, Germany, 9-11 August 2022.

Alain Becam, Wolfgang Müller, Olga Krebs, Susan Eckerle:
"Data Management tools: hands on sessions (OpenBIS, SEEK, NextCloud)" at LiSyM-Cancer Young Scientists Retreat, Hofgeismar, Germany, 7-9 September 2022.

Johannes Bracher:
Seminar on *"Probabilistic Time Series Forecasting Challenge"*, Karlsruhe Institute of Technology, winter semester 2022/23. With N. Koster, F. Krüger, S. Lerch.

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

Fernando Camacho Cadena, Colin Davalo and Merik Niemeyer:
Reading seminar on *"AdS geometry and representations"*, Universität Heidelberg, summer semester 2022. Reading seminar on *"Affine buildings",* Universität Heidelberg, summer semester 2022.

Xian Dai:
Lecture on *"A first course in dynamical system and ergodic theory"*, Universität Heidelberg, winter semester 2021/2022.

Valentina Disarlo, Marta Magnani, Diaaeldin Taha:
Reading seminar on *"Geometric deep learning"*, Universität Heidelberg, summer semester 2022.

Javier Morán Fraile:
Tutorials on *"Computational Astrophysics"*, Heidelberg University, summer semester 2022.

Nikos Gianniotis:
"Bayesian inference in linear-Gaussian models", Astrohack week 2022, Haus der Astronomie, Heidelberg, Germany, 18 October 2022.

Tilmann Gneiting:
Lecture on *"Time Series Analysis"*, Karlsruhe Institute of Technology, summer semester 2022.
Lecture on *"Forecasting: Theory and Practice"*, Karlsruhe Institute of Technology, winter semester 2022/23.

Frauke Gräter, Camilo Aponte-Santamaria (MBM), Rebecca C. Wade, Stefan Richter, Abraham Muñiz Chicharro (MCM):
M.Sc. lecture and practical course on *"Computational Molecular Biophysics"*, Heidelberg University, summer semester, 2022.

Frauke Gräter, Eric Hartmann (MBM), Rebecca C. Wade, Jonathan Teuffel (MCM):
M.Sc. seminar course on *"Machine Learning for the Biomolecular World"*, Heidelberg University, winter semester, 2022/23.

Frauke Gräter, Rebecca Wade, Camilo Aponte-Santamaría, Svenja De Buhr, Matthias Brosz, Abraham Muniz Chicharro, Stefan Richter:
Lecture on *"Computational Molecular Biophysics"*, Heidelberg University, summer semester 2022.

Frauke Gräter, Rebecca Wade (and Rob Russell, Heidelberg University):
Lecture on *"Computational Biochemistry"*, Heidelberg University, winter semester 2021/22

Ganna Gryn'ova, Michelle Ernst and Christopher Ehlert:
Special Lecture Course *"Applied Computational Chemistry"*, Heidelberg University, summer semester 2022.

Saskia Hekker:
Lecture on *"Asteroseismology"*, winter semester 2021/22, winter semester 2022/23. Seminar series on *"Applications of asteroseismology"*, summer semester 2022.

Alexander Holas:
Tutorial on *"Physikalisches Praktikum für Studierende der Medizin und Zahnmedizin"*, Heidelberg University, winter semester 2021/2022.

Pengfei Huang:
Lecture on *"Geometric invariant theory and nonabelian Hodge correspondence"*, Universität Heidelberg, summer semester 2022.

Olga Krebs, Maja Rey:
1st Training School for ITN CC-TOP *"Open Science/Open Access/ Open Data", "Writing DMP", "Writing SOP", "Standardisation, Standard based data storage templates", "FAIRDOM-SEEK hands-on"*, Dubrovnik/Cavtat, Croatia, 22-28 April 2022.

Olga Krebs:
Training in Writing Data Management Plan at LiSyM-Cancer Young Scientists Retreat, Hofgeismar, Germany, 7-9 September 2022.

Kiril Maltsev:
Tutorial in *"Analytical Mechanics and Thermodynamics"*, Heidelberg University, summer semester 2022.

Maria Beatrice Pozzetti:
Lecture on *"Einführung in die Geometry"*, Universität Heidelberg, summer semester 2022. Lecture on "Geometrische Gruppentheorie", Universität Heidelberg, winter semester 2021/2022.

Anja Randecker:
Lecture on *"Einführung in die Geometry"*, Universität Heidelberg, summer semester 2022. Student seminar "HEGL Proseminar/ Seminar: Visualizations in hyperbolic space", Universität Heidelberg, winter semester 2021/2022.

Friedrich Röpke:
Lecture course *"Fundamentals of Simulation Methods"* (with Mario Flock), Heidelberg University, winter semester 2021/2022. Lecture course *"Computational Astrophysics"*, Heidelberg University, summer semester 2022. Seminar on *"Physics of Stellar Objects"*, Heidelberg University, winter semester 2021/2022, summer semester 2022, and winter semester 2022/2023. Seminar on *"Physics of Stellar Objects"*, Heidelberg University, winter semester 2021/2022, summer semester 2022, and winter semester 2022/2023.

Friedrich Röpke, Fabian Schneider:
Lecture course *"The Stellar Cookbook: A practical guide to the theory of stars"*, Heidelberg University, winter semester 2021/2022, and winter semester 2022/2023.

Anna Schilling:
Lecture on *"Fun Facts aus Analysis und linearer Algebra"*, Universität Heidelberg, winter semester 2021/2022.

Fabian Schneider:
Lecture course *"Stars Squared: Evolution of Binary Stars"*, summer semester 2022.

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, 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, 2022/2023.

Alexandros Stamatakis, Benoit Morel, Alexey Kozlov:
Seminar *"Hot Topics in Bioinformatics"*, computer science Master's program at Karlsruhe Institute of Technology, summer semester, 2022.

Alexandros Stamatakis, Ben Bettisworth, Benoit Morel:
Summer School *"Computational Molecular Evolution"*, Welcome Trust Genome Campus, Hinxton, UK, July 2022.

Michael Strube:
PhD Colloquium, Department of Computational Linguistics, Heidelberg University (winter semester 2021/2022). Seminar *"Discourse Processing"*, Department of Computational Linguistics, Heidelberg University (winter semester 2021/2022). PhD Colloquium, Department of Computational Linguistics, Heidelberg University (summer semester 2022).

Jan Stühmer:
Geometric Deep Learning, Karlsruhe Institute of Technology, winter semester 2022/2023.

Diaaeldin Taha:
Pro-seminar on *"Experimental Geometry and Math Visualization (HEGL)"*, Universität Heidelberg, summer semester 2022.

Rebecca Wade, Stefan Richter, and Manuel Glaser:
BSc. lecture and practical course on Bioinformatics, Heidelberg University, winter semester 2021/22.

Rebecca Wade:
Lecture contributions to the M.Sc. Molecular & Cellular Biology Module 3 on *"Protein Modelling"*, M. Sc. Biochemistry *"Computational Biochemistry"* course, and Molecular Biotechnology *"Mobi-4all"* lecture series.

10 Miscellaneous

10.1 Guest Speaker Activities (invited talks):

Michaël Bazot:
"An asteroseismic butterfly diagram", Stars meeting at Landessternwarte Heidelberg, 7 June 2022.

Michelle Ernst:
"Host-guest interactions in MOFs and COFs", Hutter group seminar, University of Zurich, Zurich, Switzerland, 12 December 2022.

Sucheta Ghosh:
"Humanity Towards Opulence using AI", International Seminar on *"Role of Technology in the New Normal: Social, Ethical and Psychological Challenges"* IQAC & Department of Philosophy, RKS Mission Vivekananada Vidyabhavan & Scottish Church College, Calcutta, India, 5-6 December 2022.

Martin Golebiewski:
"Standards for in-silico modeling in health research and medicine", SimCardioTest Workshop on Verification & validation of in-silico models, online, 8 February 2022. *"Standards und Werkzeuge für FAIRe Forschungs-*

daten in der Biomedizinischen Forschung", TMF Arbeitsgruppe IT-Infrastruktur und Qualitätsmanagement, online, 23 March 2022. *"Standardization Activities of ISO/TC 276/WG 5 Data Processing and Integration",* BBMRI.QM Newsroom of the European research infrastructure for biobanking, online, 29 March 2022. *"Standards for integrating heterogenous data and metadata from organ-on-chip technologies"* (talk invited by the European Commission), 1st MICROPHYSIOLOGICAL SYSTEMS WORLD SUMMIT (MPS World Summit 2022), New Orleans, Louisiana (USA), May 30 – June 3, 2022. *"Standards for integrating heterogenous data and metadata in the life sciences"* (talk invited by the European Commission), 1+ Million Genomes declaration special group, online, 20 June 2022. *"Standards for Data Integration and Model Validation: From Community to ISO Standards",* Webinar of the European Commission *"Towards broadly applicable standards for in silico approaches in personalized medicine"*, 14 September 2022.

Frauke Gräter:

"Stressed soft matter: How tendons handle tension and free electrons", Physikalisches Kolloquium Uni Leipzig, 11 January 2022 (online); FOR5099 Freiburg, 19 January 2022 (online). *"How to feel the force: proteins translating mechanics into biochemistry"*, King's College London, 22 February 2022. *"Shifting clouds: How charges impact IDP dynamics"*, IDP Conversations seminar, New Equilibrium Bio, Cambridge/MA, 9 February 2022. *"Shifting clouds: How phosphorylation impacts IDP dynamics"*, CECAM Workshop, From disordered biomolecular complexes to biological coacervates, 8-11 March 2022. *"HOW COLLAGEN DYNAMICALLY TAMES ITS RADICALS"*, Les Houches Protein Dynamics Workshop, Aussois, France, 22-27 February 2022. Seminar Talk, Weizmann Institut, Rehovot, Israel, 15 June 2022. *"Stressed soft matter: How biopolymers in our tissues handle tension and free electrons"*, Seminar Series, Max Planck Institute for Polymer Research, Mainz, 24 June 2022. Conference *"Mechanobiology in Evolution"*, Internationales Wissenschaftsforum, Heidelberg, 4-6 July 2022; Gordon Research Conference on Mechanochemistry and Mechanobiology, Ventura, CA, 31 July – 5 August 2022; German Biophysical Society Meeting, Konstanz, 25-28 September 2022. *"Collagen: merely a strong material or rather a huge enzyme?"*: Matter to Life – Fall Days 2022, Aachen, 29-30 September 2022. *"Protein function studied by machine learning scale-bridging simulations"*: MINDS PI Meeting, 24 October 2022.

Ganna Gryn'ova:

"Computational Chemistry of Functional Organic Materials", IWR Evaluation, Heidelberg, Germany, 16-17 May 2022. *"Structural and Reactivity Patterns in N-Heteropolycyclic Radicals"*, 2nd Symposium *"Materials for Organic Electronics: Synthesis, Spectroscopy and Theory"*, Heidelberg, Germany, 10-11 June 2022. *"Application-Oriented Chemistry of Functional Organic Materials"*, 12th Triennial Congress of the World Association of Theoretical and Computational Chemists (WATOC 2020), Vancouver, Canada, 3-8 July 2022. *"Noncovalent Interactions in Functional Organic Materials"*, 2nd International Conference on Noncovalent Interactions (ICNI), Strasbourg, France, 18-22 July 2022. *"Breaking Down the Structures and Building Up the Properties in Functional Organic Materials"*, Aslla Symposium *"Accelerate Chemical Science using AI"*, Gangneung, Korea, 25-28 September 2022. *"Computational Chemistry and Machine Learning of Functional Organic Materials"* (online), WCPM Seminar Series, Warwick Centre for Predictive Modelling, University of Warwick, UK, 7 November 2022. *"Non-Covalent Interactions in Framework Materials"*, Institute of Organic and Macromolecular Chemistry, Heinrich-Heine-University, Düsseldorf, Germany, 24 November 2022.

Julia Haag:

"Predicting the Difficulty of a Phylogenetic Analysis", EVOLCYP Workshop on Biodiversity Genomics, University of Cyprus, Cyprus, September 2022.

Saskia Haupt:

"Unraveling the black boxes in early cancer development using mathematical modeling at different scales", Joint Mathematics Meetings (JMM) 2022, AMS Special Session on Current Advances in Computational Biomedicine, online, 6 April 2022 – 9 April 2022. *"Why the combination of modeling and machine learning could be the future direction in mathematical oncology"*, EKFZ Lecture Series Focus on AI, EKFZ Digital Health Dresden, Germany, 25 August 2022.

Saskia Hekker:

"Asteroseismology of red-giant stars", Invited review at the annual spring DPG meeting, Heidelberg, 23 March 2022. *"The power of asteroseismology: stellar structure revealed by global oscillations",* Institute for Astronomy, Göttingen, Germany, 12 May 2022; Munich Physics Colloquium, LMU, Munich, Germany, 25 July 2022. *"The power of Asteroseismology" Probing stellar interiors"* (inaugural lecture), faculty of Physics and Astronomy, Heidelberg University, 8 July 2022. *"Asteroseismology of solar-like oscillators",* Invited review at TASC6/KASC13 Conference, Leuven, Belgium, 14 July 2022. *"Asteroseismology"*, Invited review at COSPAR, Athens, Greece, 21 July 2022; invited talk at Stars 2020, to mark the 80th birthday of Peter Eggleton, Cambridge UK, 14-20 August 2022.

Vincent Heuveline:

"Der KI-Komponist aus den Fugen: im Spannungsfeld zwischen mathematischer Unsicherheit und Schönheit." Talk in the *„Klangforum"* series *„Künstliche Intelligenz und das Seufzen der Kultur? Von Geist und Transhumanismus"*, Providenzkirche Heidelberg, 17 September 2022; ZKM Karlsruhe, 18 September 2022.

Alexey Kozlov:

"EcoFreq: compute with cleaner energy via carbon-aware power scaling", GSI ITEE-Palaver, Darmstadt, Germany, May 2022.

Philipp Lösel:

"Large-scale analysis of the honey bee brain using micro-CT imaging and the online segmentation platform Biomedisa", BMIT user workshop on X-ray micro-CT data processing, Canadian Light Source (CLS), Saskatoon, Canada (online), 6 October 2022.

Benoit Morel:

"Inferring gene trees and species trees in presence of gene duplication, loss, and transfer", EVOLCYP Workshop on Biodiversity Genomics, University of Cyprus, Cyprus, September 2022.

Wolfgang Müller:

"50.... - 4 Ways to FAIR your data": Kiwi-biolab talk, online/TU Berlin, 10 May 2022.

Kai Polsterer:

"From Photometric Redshifts to Improved Weather Forecasts: an interdisciplinary view on machine learning in astronomy", opening symposium key note at the Center for Data and Computing in Natural Sciences (CDCS), Hamburg, Germany, 16 April 2022. *"Review of supervised and unsupervised learning techniques for astronomy"*, Machine Learning for Astrophysics, Catania, Italy, 30 May 2022. *"From Supervised to Unsupervised ML: lessons learned from learning machines",* colloquium, Universitat Autonoma de Barcelona (UAB) Institut de Fisica d'Altes Energies (IFAE), 8 September 2022.

Friedrich Röpke:
"Multidimensional simulations of thermonuclear supernovae" ChETEC-INFRA 2nd General Assembly, Padova, Italy, 1 June 2022.

Fabian Schneider:
"Turbulent Lives of Stars", Munich Physics Colloquium, LUM and TUM joint colloquium, Munich, Germany, 10 January 2022; Leiden Observatory Colloquium, Leiden University, The Netherlands, 7 July 2022. *"The strongest magnets in the Universe"*, Virtual Nordic Dynamo Seminar, NORDITA, Stockholm, Sweden, 22 February 2022. *"Magnetogenesis in stellar mergers and common-envelope events"*, Virtual Delaware Astronomy & Space Seminar, University of Delaware, USA, 4 October 2022.

Michael Strube:
"Generalizability and Robustness in Coreference Resolution", CLUNCH at Penn NLP, University of Pennsylvania, Philadelphia, PA, USA., 12 September 2022.

Jonathan Teuffel:
"Effects of conformational transitions and redox protein binding on the catalytic properties of CYP17A1 and CYP2B4 revealed by ligand egress patterns", 22nd International Conference on Cytochrome P450, Washington DC, USA, 21 July 2022.

Rebecca C. Wade:
"Zooming in on the dynamic interactions of proteins and drugs by computer simulation", SFB 1381 Young Researchers' Retreat, Freiburg University, Germany, 2-3 June 2022. *"Computational challenges for biomolecular simulation approaches to drug discovery"*, International Conference on Computer Sciences – ICCS 2022, Brunel University, London, UK, 21-23 June 2022. *"Bridging timescales to predict protein-ligand binding kinetics"*, Workshop on *"Multiscale modelling in materials science, chemistry, and biology"*, KIT, Karlsruhe, Germany, 11-12 July 2022. *"Computation of Drug-Target Binding Kinetics"*, Gordon Conference on Computational Chemistry, Barcelona, Spain, 17-22 July 2022. *"Mapping dynamic protein binding sites for the design of selective anti-infective agents"*, 3rd Molecules Medicinal Chemistry Symposium (MMCS 2022), Rome, Italy, 27-29 July 2022. *"Brownian dynamics simulation of protein and drug diffusion in the presence of surfaces"*, CECAM Flagship Workshop on *"Immobilizing peptides and proteins: Interplay between theoretical and experimental approaches"*, Institut de Biologie Physico-Chimique, Paris, France, 4-7 October 2022. *"Computer simulation of the interactions of drug molecules with polymer excipients, protein crowders and target receptors"*, NIPER-PHARMACON 2022 International Symposium on Recent Trends and Future Opportunities in Pharmaceuticals. Chandigarh, India, 9-11 November 2022. *"Bridging timescales to predict protein-ligand binding kinetics"*, Taller de technicas avanazadas de modelado de Biomoleculas, UNAM, Mexico (online), 30 November 2022. *"Computing the Brain"*, Moderator, HBP Brain Matters webinar series #13 (online), 13 December 2022.

Ulrike Wittig:
"Woman Power in Data Curation and Data Management", Joint Bayer AG & de.NBI/ELIXIR Germany & Bioinfo4Women Event *"Women in Data Science - Perspectives in Industry and Academia"*, online, 18 May 2022. *"User support and training activities in FAIRDOM"*, 2nd ELIXIR-CONVERGE Community of Practice Event, online, 21 June 2022.

10.2 Presentations

Talks (Contributed talks)

Róbert Andrássy:
"Convective penetration in stars", poster at the IAU Symposium 361, Ballyconnell, Ireland, 8-13 May 2022. *"Convective penetration in massive stars"*, talk at the 16th Würzburg Winter Workshop, Heidelberg, Germany, 13 December 2022.

Camilo Aponte-Santamaría:
"Energetics and permeation of photo-resists used for 3D-laser printing across biological lipid bilayers" Talk at 3DMM2O Winter Cluster Meeting, Neustadt a.d. Weinstraße (Germany), 19-20 April 2022. *"Computational studies of von Willebrand factor"*. (Seminar), University of Siegen (Siegen, Germany), 27 August 2022.

Christina Athanasiou:
"Mechanistic Studies of the TrkA and TrkB Neurotrophin Receptors", Heidelberg University Center for Molecular Biology (ZMBH) ZAPP seminar, online, 2 June 2022.

Giulia D'Arrigo:
Lecture and Practical on *"τRAMD calculations of protein-ligand dissociation rates"*, HBPMolSIM Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades", online, 7-10 March 2022. *"Estimation of Protein-Protein Dissociation Rates from τRAMD Simulations"*, ISQBP President's meeting 2022. Innsbruck, Austria. 10-14 July 2022. *"Computation of Biomolecular Dissociation Rates and Mechanisms from τRAMD Simulations"*, 23rd EuroQSAR, Heidelberg, Germany, 26-30 September 2022. Lecture and Practical on *"Estimation of residence time and determination of unbinding pathways through τ-Random Acceleration Molecular Dynamics (τ-RAMD)"*, *"Simulate with EBRAINS"* Training Workshop, online, 7-11 November 2022.

Michaël Bazot:
"An asteroseismic butterfly diagram", Cool stars meeting, Toulouse, France, 5 July 2022.

Matthias Brosz:
"Martini 3 coarse-grained force field for poly(para-phenylene ethynylene)s", Winter Cluster Meeting 3DMM2O (Poster, Neustadt a.d. Weinstraße), 19-20 April 2022; *"From polymers to proteins with Coarse-Grained Molecular Dynamics"*, 3DMM2O Success and Challenges (Talk, Heidelberg), 21 June 2022; *"Ballmilling Poly(para-phenylene ethynylene)s – Unlikely Radicals via Mechanical Stress"*, Summer Cluster Meeting 3DMM2O (Poster, Heidelberg) 10-11 October 2022.

Susan Eckerle:
"General Introduction to LiSyM-Cancer Network Structure, Programme Management and Activities", LiSyM-Cancer Status Seminar, Heidelberg, 18-19 May 2022.

Christopher Ehlert:
"Quantum Chemical Assessment of Well-Defined Catalysts for the Oxygen Reduction Reaction" (online), International Conference on Catalysis and Chemical Engineering, San Francisco, USA, 22-24 February 2022.

Michelle Ernst:
"Insights into the Nature of Host-Guest Interactions in MOFs And COFs", European Young Chemists' Meeting (online), Fribourg, Switzerland, 19-21 January 2022; 30th Annual Meeting of the German Crystallographic Society (DGK-30, online), Munich, Germany, 14-17 March 2022.

Rostislav Fedorov:
"Simulate with EBRAINS" Training Workshop, online, 7-11 November 2022. *"Recent advances in MPNN for Battery Properties Prediction"*, Scientific Seminar @ AiMat, KiT, Karlsruhe, Germany, 24 November 2022.

Matheus Ferraz:
Mini-course on *"Molecular modelling of antigens and antibodies"*, Brazilian National Congress of Immunology, online, 11 April 2022. Mini-course on *"Molecular Modelling in Biotechnology"*, FIOCRUZ Workshop on Bioscience and Biotechnology in Health, Oswaldo Cruz Foundation, Recife, Brazil, online, 8-9 November 2022.

Nikos Gianniotis:
"Probabilistic Cross-Correlation for Delay Estimation", AG-Tagung, Bremen, Germany, 15 September 2022.

Manuel Glaser:
"SP2: Combining experimental and structure-based computational approaches to unravel the molecular mechanisms of the regulatory hub protein S100A1 and its C-terminus in the heart", Informatics for Life – 29th Regular Project Meeting, Heidelberg, 6 December 2022.

Sucheta Ghosh:
"Towards Effortless Navigation of Scientific-Literature Screen Reading for Biocuration", Software, Applications and Systems in Biocuration Session II. 1st UK-Local Biocuration Conference, 5-6 May 2022, Hinxton, Cambridge, UK.

Martin Golebiewski:
"Standards for in-silico modelling in personalized medicine", Meeting International Consortium for Personalised Medicine (ICPerMed) and EU-STANDS4PM, online, 21 January 2022. *"ISO/TC 276/WG 5 Data Processing and Integration"*, Committee Meetings of the German standardization committe *"Arbeitsausschuss Biotechnologie"*, DIN, online, 8 February, 27 April and 19 October 2022. *"Standards for Human Digital Twins and In-silico Modeling in Health Research and Medicine"*, IEC SEG 12 Bio-Digital Convergence, International Electrotechnical Commission (IEC), online, 17 March 2022. *"Wie mache ich Health-Daten FAIR: Benutzung der Forschungsdateninfrastrukturen von NFDI4Health und FAIR4Health"*, DMEA - Satellitenveranstaltung 2022 von GMDS und BVMI, Berlin (Germany), 25 April 2022. *"ISO standards relevant for personalized medicine"* and *"Development of community-based guidelines and normative documents: Recommendations and requirements for predictive computational models in personalised medicine"*, EU-STANDS4PM annual meeting, online, 18-19 May 2022. *"Standards in Action: Implementierung des Covid-19-Metadatenmodells und die NFDI4Health Standardisierungs-Roadmap"*, NFDI4Health Task Force COVID-19-Symposium, Fraunhofer-Institut für Digitale Medizin MEVIS, Bremen (Germany), 13 June 2022. *"Standardization needs of the NFDI4Health use cases"*, Annual NFDI4Health Consortium Meeting, Studio Villa Bosch, Heidelberg (Germany), 20-21 October 2022. *"ISO/TC 276/WG 5 Data Processing and Integration - Report of the Convenor"*, Committee Meetings of ISO/TC 276/WG 5, online, 14 March and 3 November 2022 and ISO/TC 276 Biotechnology - Plenary Meeting, online, 29 November 2022. *"Analysing current landscape of standards, identifying needs and gaps"*, Meetings of the Ecosystem Digital Twins in Healthcare (EDITH) Consortium, online, 16 November 2022 and Leuven (Belgium), 29 - 30 November 2022. *"State of the Art: Standards for Virtual Human Twin Data Integration and Modelling"*, High-level partners meeting of the Ecosystem Digital Twins in Healthcare (EDITH), Leuven (Belgium), 29 - 30 November 2022. *"NFDI4Health Success Story Metadata"*, NFDI4Health Friday Meeting, online, 2nd December 2022.

Anton Hanke:
"Engineering Antibody Evolution", Workshop on Synthetic Immunology, Engineering Molecular Systems Flagship Initiative, Heidelberg University, Heidelberg, 26 April 2022.

Eric Hartmann:
"Reactive Molecular Mechanics", BioExcel Summer School on Biomolecular Simulations, Pula (Italy), 12-17 June 2022.

Saskia Haupt:
"Mathematically modeling multiple pathways of carcinogenesis using the Kronecker structure", talk at USACM Thematic Conference: The role of mathematical and computational modeling in cancer research, online, 10–11 January 2022.
"Why the combination of modeling and machine learning could be the future direction in mathematical oncology", talk at 6th EHTG Meeting 2022, Palma de Mallorca, Spain, 29 September 2022 – 1 October 2022. *"Project proposal: A proportional Cox model for subgroup differences in cancer risk based on the PLSD data"*, talk at PLSD Meeting at the 6th EHTG Meeting 2022, Palma de Mallorca, Spain, 29 September 2022 – 1 October 2022.

Jan Henneco:
"Contact tracing of binary Stars - from sub-solar to high-mass components", Stars 2020 Conference, IoA, University of Cambridge, United Kingdom, 16 August 2022. *"Contact tracing of binary Stars - from sub-solar to high-mass components",* MIAPbP Workshop: The Fundamental Role of Stellar Multiplicity in Stellar Dynamics and Evolution, MIAPbP, Garching, Germany, 21 November 2022.

Alexander Holas:
"Improvements to modelling tools for Type Ia supernovae – A focus on radiation transport", talk at Partially Ionised Plasmas in Astrophysics 2022, Hungary, 6 June 2022. *"A hierarchical Bayesian framework for standardising Type II-P supernovae"*, talk at the XVI. Würzburg Workshop on Stellar Astrophysics in Heidelberg, Germany, 13 December 2022.

Olga Krebs:
"FAIRDOM: Promoting and Supporting FAIR Data and Model Management in Systems Biology", Computational Modeling in Biology Network (COMBINE) Conference, 6-7 October 2022, Berlin, Germany.

Markus Kurth:
"Dihydroxyphenylalanine (DOPA) in collagen acts as a very efficient radical scavenger", 3DMM2O Winter Cluster Meeting (Talk, Neustadt a.d. Weinstraße), 19 -20 April 2022; *"A role for dihydroxyphenylalanine (DOPA) as a marker of mechanical stress and radical scavenger in tendon collagen"*, 9th World Congress of biomechanics (Talk, online), 10 -14 July 2022; 6th Collagen Café (Talk, online), 26 October 2022. *"A molecular marker for mechanical stress and aging in connective tissues"*, BioRN Annual Conference 2022 (Talk, Heidelberg), 7 October 2022.

Giovanni Leidi:
"A finite-volume scheme for modelling compressible MHD flows at low Mach numbers with gravity", talk at the Partially Ionised Plasmas in Astrophysics conference, Budapest, Hungary, 6 June 2022. *"A finite-volume scheme for simulating fully compressible MHD flows at low Mach numbers in stellar interiors"*, talk at the Challenges and Innovations in Computational Astrophysics IV conference, online, 23 November 2022.

Stiv Llenga:
"Theoretical Investigations of Anthracene-Based Open-Shell N-Heteropolycycles", SFB 1249 Werkstattgespräch, Heidelberg University, Heidelberg, Germany, 4 February 2022. *"New Molecular Representations for Radical Ions"* (online), Theoretical Physical Organic Chemistry (TPOC) seminar series, 8 December 2022.

Kiril Maltsev:
"Deep learning emulation of the MESA/MIST stellar evolution models", talk at VFTS consortium meeting, Heidelberg, Germany, 22 June 2022. *"Cost-efficient stellar evolution and final fate forecasting continuous parameter spaces"*, talk at XVI Würzburg winter workshop on Stellar Astrophysics, Heidelberg, Germany, 14 December 2022.

Nicholas Michelarakis:
"Avidity of Malaria Adhesin VAR2CSA is Mechano-Controlled by Exposure of a Second Cryptic CSA Binding Site": Talk at Workshop: Theory and Simulation of Macromolecules, Hünfeld (Germany), 8-9 April 2022.

Wolfgang Müller:
"Data Management: General Introduction on Data and Repositories", LiSyM-Cancer Young Scientists Retreat, Hofgeismar, Germany, 7-9 September 2022. *"ReStoRunT: Recording, Storing, (Re-)Running, Tracing changes in Excel tables"*, International Study Group for Systems Biology (ISGSB) conference, Innsbruck, Austria, 19-23 September, 2022.

Abraham Muniz Chicharro:
"A multiscale approach to compute conformation-gated ligand-receptor binding kinetics", American Chemical Society Spring 2022 Meeting, San Diego, USA, 24 March 2022.

Anthony Noll:
"Impact of central mixing and nuclear reactions network on the size of convective cores", Workshop on codes for stellar physics, Paris, France, July 2022.

Giulia Paiardi:
"Molecular dynamics simulations to investigate the antiviral effect of heparin in SARS-CoV-2 spike infection", MINDS retreat, H+ Hotel Wiesbaden-Niedernhausen, Germany, 29-30 March 2022; Annual meeting, Innogly COST action, Lugano, Switzerland, 4-6 May 2022.

Anna Piras:
"Nitroaromatic Pollutants on Graphene: from Physisorption to Redox Properties", 9th Young Scientists Workshop on Theoretical Chemistry and Computational Modelling, Vigo, Spain, 20 June 2022.

Benedikt Rennekamp:
"Hybrid Simulations of Collagen Breakages Reveal Mechanical and Oxidative Buffering", Workshop: Theory and Simulation of Macromolecules, Hünfeld (Germany), 8-9 April 2022; Matter-to-Life Spring Days, Online,13-14 April 2022. *"HYBRID SIMULATIONS OF COLLAGEN FAILURE"*, Biophysics at the Dawn of Exascale Computers by Biophysical Society, Hamburg (Germany), 16–20 May 2022; Matter-to-Life Ringberg Symposium, Schloss Ringberg (Germany), 10–16 July 2022. *"Where does collagen break?"*, Virtual department seminar of Max-Planck-Institute for Biophysical Chemistry in Göttingen, 7 July 2022. *"Where collagen breaks: Weak sacrificial bonds tame its mechanoradicals"*, Matter-to-Life Fall Days, Aachen (Germany), 28-30 September 2022.

Fabian Schneider:
"Stellar mergers as the origin of magnetic massive stars", IAU361: Massive Stars Near and Far, Ireland, 10 May 2022. *"Magnetogenesis in stellar mergers and common-envelope events"*, Stars 2020 Conference, IoA, University of Cambridge, United Kingdom, 16 August 2022.

Jonathan Teuffel:
"Learning the determinants of interprotein electron transfer from combined quantum mechanics and molecular mechanics simulations", HGS MathComp Graduate School Annual Retreat on *"The future of simulations"*, Speyer, Germany, 5 September 2022.
"Computational modeling of cytochrome P450 redox protein complexes" 5th SIMPLAIX scientific meeting, Heidelberg, Germany, 15 September 2022.

Alexandros Tsengenes:
"Computational Studies of the p75 Neurotrophin Receptor", Heidelberg University Center for Molecular Biology (ZMBH) ZAPP seminar, online, 2 June 2022.

Aysecan Ünal:
"Structural analysis of collagen fibrils with cryoEM", Matter to Life Ringberg Symposium, Schloss Ringberg (Germany), 10-16 July 2022 *"All-atom 3D model of crosslinked collagen fibrils"*, 9th Electron Tomography Congress, Egmond aan Zee (Netherlands), 11-14 September 2022. *"Structure of collagen fibrils with cryoEM"*, Matter to Life Fall Days, Aachen (Germany), 28-30 September 2022; Cryo-electron Microscopy and Beyond Workshop, Los Angeles (USA), 14-18 November 2022.

Rebecca C. Wade, Abraham Muniz Chicharro, Stefan Richter:
Lecture and Practical on *"Brownian dynamics simulations with SDA to compute association rates"*, HBPMolSIM Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades", online, 7-10 March 2022.

Rebecca C. Wade:
"PIPSA (Protein Interaction Property Similarity Analysis), τRAMD, SDA (Simulation of Diffusional Association)", HBP Solution-Workshop for Start-ups in Neuropharma, on-line, 12 April 2022. *"The Molecular and Cellular Modeling (MCM) Group: Zooming in on the dynamic interactions of proteins and drugs by computer simulation"*, Heidelberg Laureate Form Young Researchers@HITS, Heidelberg, 21 September 2022. *"Molecular Simulations"*, Informatics for Life annual meeting, Heidelberg, Germany, 19 November 2022.

Ulrike Wittig:
"ELIXIR's infrastructure - Updates from the Data Platform", ELIXIR All Hands Meeting, Amsterdam, Netherlands, 6-9 June 2022.

Alexander Zeilmann:
"Uncertainty Quantification in Oncology - How do uncertain data influence model results and clinical decision making?", talk at 6th EHTG Meeting 2022, Palma de Mallorca, Spain, 29 September 2022 – 1 October 2022.

Yaroslav Zharov:
"Artificial Intelligence for Large-scale Facilities" talk at *"5. Konferenz zur Forschung mit Synchrotronstrahlung, Neutronen und Ionenstrahlen an Großgeräten,"* Berlin, 5-7 September 2022.

Posters

Camilo Aponte-Santamaría:
"Coarse-grained models for intrinsically disordered domains ported into GROMACS: how phosphorylation shifts ensembles" Poster at Workshop: Theory and Simulation of Macromolecules, Hünfeld (Germany), 8-9 April 2022, and Biophysics at the Dawn of Exascale Computers by Biophysical Society, Hamburg (Germany), 16 -20 May 2022.

Christina Athanasiou:
"Modeling and Simulation of the TrkA and TrkB Neurotrophin Receptors", 23rd European Symposium on Quantitative Structure-Activity Relationship, Heidelberg, Germany, 26-30 September 2022.

Giulia D'Arrigo:
"Computation of Protein-Protein Dissociation Rates and Mechanisms", 23rd EuroQSAR. Heidelberg, Germany, 26-30 September 2022.

Matthias Brosz:
"Bridging the gap between simulations and native tissue -molecular biomechanics of collagen", Summer Cluster Meeting Cluster 3DMM2O (Poster, Heidelberg), 10 – 11 October 2022.

Lynn Buchele:
"Structure Inversions of Main Sequence Stars", TASC6/KASC13 Workshop, KU Leuven, Belgium, 11-15 July 2022.

Quentin Coppee:
"Slender radial modes in red giant stars with suppressed non-radial modes", TASC6/KASC13 Workshop, KU Leuven, Belgium, 11-15 July 2022.

Lucas Diedrich:
"Energetics and permeation of photo-resists used for 3D-laser printing across biological lipid bilayers" Winter Cluster Meeting 3DMM2O (Poster, Neustadt a.d. Weinstraße), 19-20 April 2022; *"Energetics and permeation of photo-resists used for 3D-laser printing across biological lipid bilayers"*; German Biophysical Society Meeting 2022 (Poster, Constance), 25-28 September 2022; *"Energetics and permeation of photo-resists used for 3D-laser printing across biological lipid bilayers"* Summer Cluster Meeting 3DMM2O (Poster, Heidelberg) 10-11 October 2022.

Dorotea Dudaš, Maja Rey, Ulrike Wittig, Andreas Weidemann, Wolfgang Müller:
"Deep insight into SABIO-RK data via visualization", ICSB 2022, Berlin, Germany, 8-12 October 2022.

Christopher Ehlert:
"Stereoinversion of Tetracoordinated p-Block Element Hydrides", 12th Triennial Congress of the World Association of Theoretical and Computational Chemists (WATOC 2020), Vancouver, Canada, 3-8 July 2022.; 58th Symposium on Theoretical Chemistry, Heidelberg, Germany, 18–22 September 2022.

Michelle Ernst:
"Analysing and Utilising Host-Guest Interactions in MOFs and COFs", IWR Evaluation, Heidelberg, Germany, 16-17 May 2022. *"Computational Chemistry to Analyse Host-Guest Interactions in MOFs and COFs"*, 12th Triennial Congress of the World Association of Theoretical and Computational Chemists (WATOC 2020), Vancouver, Canada, 3-8 July 2022; Leverhulme Research Centre for Functional Materials Design 2nd Biennial Symposium, Liverpool, England, 21-23 September 2022.

Matheus Ferraz:
"Structure-based computational design of VHHs targeting the SARS-CoV-2 spike protein", Engineering Molecular Science Flagship Initiative Young Scientist Retreat, Heidelberg University, Dannenfels, 1-3 August 2022. *"Artificial neural network to predict the free energy of binding for protein-protein complexes from Rosetta-calculated parameters"*, 2022 Virtual Workshop on Free Energy Methods in Drug Design, online, 21-22 September 2022. *"Enhanced Sampling the Dissociation Pathway of VHH-72 against SARS-CoV-1 and -2 Receptor Binding Domain"*, 23rd European Symposium on Quantitative Structure-activity relationship, Heidelberg, 26-30 September 2022.

Manuel Glaser:
"Towards molecular design of peptide-based therapeutics against striated muscle disorders: unraveling the biophysics of the inotropic peptide S100A1ct by molecular modeling and simulation", ISQBP President's meeting 2022. Innsbruck, Austria. 10-14 July 2022; 23rd EuroQSAR Symposium, Heidelberg, Germany, 26-30 September 2022. *"Towards molecular design of peptide-based therapeutics against striated muscle disorders: unraveling the mechanism of action of the inotropic peptide S100A1ct"*, Informatics for Life annual meeting 2022, Heidelberg, Germany, 19 November 2022.

Saskia Haupt:
"Mathematical modeling of Lynch syndrome carcinogenesis", poster at Joint ECMTB & SMB 2022, Heidelberg, Germany, 19–23 September 2022.

Saskia Hekker:
"New oscillating red-giant stars in NGC 6791 & NGC 6819: the mass and radius for H-shell burning & core -He burning giants scale differently to the Sun", Fundamental stellar parameters from asteroseismology in the era of Gaia, Aarhus, Denmark, 13-15 June 2022.

Jan Henneco:
"Contact Tracing of Massive Binary Stars", IAU361: Massive Stars Near and Far, Ireland, 9-13 May 2022. *"Contact Tracing of binary Stars"*, Topical Workshop: The Impact of Binaries on Stellar Evolution, Max Planck Institute for Astrophysics of Garching, Germany, 14-18 November 2022.

Alexander Holas:
"Thermonuclear electron-capture supernovae – New production sites completing the solar inventory of isotopes?", poster at Nuclear Physics in Astrophysics X, Switzerland, 5 September 2022.

Olga Krebs, Martin Golebiewski, Xiaoming Hu, Ulrike Wittig, Alain Becam, Stuart Owen, Wolfgang Müller, Carole Goble:
"FAIRDOM-SEEK: FAIR data and model management platform for systems biology and systems medicine", International Study Group for Systems Biology (ISGSB) Conference, Innsbruck, Austria, 19-23 September, 2022.

Markus Kurth:
"Bridging the gap between simulations and native tissue -molecular biomechanics of collagen", Summer Cluster Meeting Cluster 3DMM2O (Poster, Heidelberg), 10 -11 October 2022.

Stiv Llenga:
"Property Patterns in N-Heteropolycyclic Radicals", 2nd Symposium *"Materials for Organic Electronics: Synthesis, Spectroscopy and Theory"*, Heidelberg, Germany, 10-11 June 2022.

Philipp Lösel:
"Large-scale analysis of the bee brain using micro-CT imaging and deep learning", poster at Heidelberg Laureate Forum, Young Researchers' visit at HITS, Heidelberg, Germany, 21 September 2022. *„HEDI: 3D surface imaging of the unstable anterior abdominal wall before incisional hernia repair"* poster at Informatics for Life - Yearly Meeting, Medical Hospital, Heidelberg, Germany, November 19, 2022.

Nicholas Michelarakis:
"Avidity of Malaria Adhesin VAR2CSA is Mechano-Controlled by Exposure of a Second Cryptic CSA Binding Site": Poster at Biophysics at the Dawn of Exascale Computers by Biophysical Society, Hamburg (Germany), 16 -20 May 2022.

Wolfgang Müller:
"Excel, das ungeliebte, wichtige Datenvehikel", NFDI4Biodiversity Annual Conference, Berlin, Germany, 12-14 October 2022. *"INCOME post project progress"*, e:Med Meeting 2022 on Systems Medicine, Heidelberg, Germany, 28-30 November 2022.

Abraham Muniz-Chicharro:
"Multiscale approach to predict protein-ligand binding kinetics", 23rd EuroQSAR, 26-30 September 2022.

Anthony Noll:
"Impact of central mixing and nuclear reactions network on the size of convective cores", TASC6/KASC13 Workshop, KU Leuven, Belgium, July 2022.

Giulia Paiardi:
"Molecular dynamics simulations to investigate the dynamics and mechanism of action of carbohydrate-based polymers", Gordon research conference on *"Multiscale Modelling of Complex Systems: Methods and Applications"*, Barcelona, Spain, 17-22 July 2022; 23rd EuroQSAR European Symposium on Quantitative Structure-Activity Relationship, Heidelberg, 26-30 September 2022; NIPER-PHARMACON 2022 International Symposium on Recent Trends and Future Opportunities in Pharmaceuticals. Chandigarh, India, 9-11 November 2022.

Alba Covelo Paz:
"Difference imaging of red giants in NGC 6791 and NGC 6819", TASC6/ KASC13 Workshop, KU Leuven, Belgium, 11-15 July 2022.

Anna Piras:
"Nitroaromatic Pollutants on Graphene: from Physisorption to Redox Properties", 12th Congress on Electronic Structure Principles and Applications, Vigo, Spain, 21-24 June 2022.

Kai Riedmiller:
"Predicting Radical Migration in Collagen", Poster at the Workshop: Theory and Simulation of Macromolecules, Hünfeld (Germany), 8-9 April 2022. METHODS IN MOLECULAR SIMULATIONS AND MACHINE LEARNING, Barcelona, 14-16 July 2022.

Jonas Roller:
"Numerical Simulation of Thermal Electrohydrodynamical (TEHD) Convection in Annular Geometry", 14th European Fluid Mechanics Conference, Athens, Greece, 15 September 2022.

Valentin Schmid:
"Efficient 3D modeling of colonic crypts for large microscale simulations of cancer development", poster at Joint ECMTB & SMB 2022, Heidelberg, Germany, 19–23 September 2022.

Andreas Weidemann, Dorotea Dudaš, Maja Rey, Ulrike Wittig, Wolfgang Müller:
"More findability, more interoperability for SABIO-RK, the curated database for biochemical reaction kinetics", ELIXIR All Hands Meeting 2022, Amsterdam, Netherlands, 7-10 June 2022.

Alexander Zeilmann:
"Mathematical modeling of Lynch syndrome carcinogenesis", poster at the Heidelberg Laureate Forum, Young Researchers' visit at HITS, Heidelberg, Germany, 21 September 2022.

Yaroslav Zharov:
"Deep Learning for Synchrotron-based Tomography" poster at 14th International Conference on Synchrotron Radiation Instrumentation (SRI 2021), virtual, 28 March to 1 April 2022.

10.3 Memberships

Michaël Bazot:
Member Plato Science Management (PSM).

Martin Golebiewski:
Convenor (chair) of the ISO/TC 276 Biotechnology working group 5 *"Data Processing and Integration"*, International Standardization Organization (ISO), reelected 2021. Chair of the working 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). German delegate at the ISO technical committee 276 Biotechnology (ISO/TC 276), International Organization for Standardization (ISO). Co-Chair of the national German standardization committee *("Nationaler Arbeitsausschuss")* NA 063-09-03 AA Organ-on-Chip, German Institute for Standardization (DIN). Member of the national German standardization committee *("Nationaler Arbeitsausschuss")* NA 057–06–02 AA Biotechnology, German Institute for Standardization (DIN). Member of the IEC SEG 12 Bio-Digital Convergence, International Electrotechnical Commission (IEC). Member of the steering committee of the German National Research Data Infrastructure for Personal Health Data (NFDI4Health). Member of the steering committee of the AIMe registry for artificial intelligence in biomedical research. Member of the Virtual Physiological Human Institute for Integrative Biomedical Research (VPH Institute).

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 Graduate school HIDDS4HEALTH. 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.

Ganna Gryn'ova:
Affiliated junior research group leader: Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University. Member: working group on advancing women's careers in mathematics and computer science, Heidelberg Laureate Forum Foundation (HLFF). Member: Early Career International Advisory Board, Helvetica Chimica Acta. Associate Editor: Theoretical and Computational Chemistry – Frontiers.

Saskia Hekker:
Scientific Advisory board member of the TESS Asteroseismic Science Consortium, European Space Science Committee, Member IAU, Member EAS.

Wolfgang Müller:
Authorized representative for the LiSyM-Cancer network in the TMF - Technologie und Methodenplattform für die vernetzte medizinische Forschung e.V.; Leadership Team of LiSyM research network Liver Systems Medicine. Deputy Chairman of SIG 4 (Infrastructure & data management), German Network for Bioinformatics Infrastructure (de.NBI).

Giulia Paiardi:
Member of the Innogly COST action (CA18103).

Kai Polsterer:
Vice President of the International Astro Informatics Association. Member of the Standing Committee on Science Priorities of the International Virtual Observatory Alliance. Member of the Astronomische Gesellschaft. Member of the Deutsche Physikalische Gesellschaft, AKPIG. Member of the International Astrostatistics Association.

Fabian Schneider:
Board Member, International Max Planck Research School for Astronomy and Cosmic Physics at Heidelberg University.

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. Associate Editor, Joint Oxford & ISMB Journal: Advances in Bioinformatics.

Jonathan Teuffel:
Head of the local students' group of the German society for Biochemistry and Molecular Biology in Heidelberg.

Rebecca Wade:
Associate Editor, Journal of Molecular Recognition, PLOS Computational Biology. Editorial Board, Advances and Applications in Bioinformatics and Chemistry; BBA General Subjects; Journal of Chemical Information and Modeling; Journal of Computer-aided Molecular Design; Journal of Physical Chemistry B; Biopolymers; Protein Engineering, Design and Selection. Member, Scientific Advisory Council of the Computational Biology Unit (CBU), University of Bergen, Norway. Member, Scientific Advisory Board of the Max Planck Institute of Biophysics, Frankfurt. Member, Managing Board of Directors, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University. Member, Steering

Committee, Thematic Research Network *"Molecular Mechanisms in health and disease - from understanding to engineering (MINDS)"*, Heidelberg University. Coordinator, SIMPLAIX HITS-Heidelberg University-Karlsruhe Institute of Technology consortium on data-enhanced multi-scale molecular simulation of biomolecules and molecular materials. Member, Interdisciplinary Center for Scientific Computing (IWR), DK-FZ-ZMBH Alliance of the German Cancer Research Center and the Center for Molecular Biology at Heidelberg University. Faculty member, HBIGS (Heidelberg Biosciences International Graduate School) faculty, HGS MathComp Graduate School faculty, Helmholtz Information and Data Science School for Health (HIDSS4Health). Member, Faculty of Engineering Sciences and Faculty of Biosciences, Heidelberg University. President, QSAR, Chemoinformatics and Modeling Society (QCMS).

Ulrike Wittig:
Member of the STRENDA Commission (Standards for Reporting Enzymology Data). Editorial board member of ELIXIR Research Data Management Kit (RDMkit). Executive committee member (ExCo) of ELIXIR Data Platform. Steering committee member of ELIXIR Biocuration Focus Group. Member of ELIXIR Germany Central Coordination Unit (CCU). Working Group (WG1) leader of COST Action CA21111 - One Health drugs against parasitic vector borne diseases in Europe and beyond (OneHealthdrugs).

10.4 Contributions to the Scientific Community

Camilo Aponte-Santamaría:
Alumni panel at GGNB Science Day *"Experience Fair"* (Göttingen, Germany), 18 October 2022.

Giulia D'Arrigo, Rebecca Wade:
Co-organizers: HBPMolSim Human Brain Project Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades (with Giulia Rossetti, Forschungszentrum Jülich, and Jeanette Hellgren Kotaleski, Royal Institute of Technology, Stockholm). Virtual Event, 7-10 March, 2022.

Frauke Gräter:
Co-organizer, *"Workshop Multiscale modelling in materials science, chemistry, and biology"* 11/12 July 2022, Karlsruhe Institute of Technology. Session Chair and Co-organizer, 23rd EuroQSAR Symposium, 23-26 September 2022, Heidelberg University, Germany.

Giulia D'Arrigo, Giulia Paiardi:
Co-organizers: *"Orienting your career compass"* Workshop, 23rd Euro-QSAR, Heidelberg, Germany, 28 September 2022.

Christina Athanasiou, Giulia D'Arrigo, Matheus Ferraz, Manuel Glaser, Melanie Käser, Abraham Muniz Chicharro, Giulia Paiardi, Jonathan Teuffel, Alexandros Tsengenes:
Helpers, 23rd EuroQSAR, Heidelberg, Germany, 26-30 September 2022.

Susan Eckerle, Wolfgang Müller:
Organizer and chair of specific session of LiSyM-Cancer Status Seminar, Heidelberg, Germany, 18-19 May 2022. Organizer and chair of LiSyM-Cancer Young Scientists Retreat, Hofgeismar, Germany, 7-9 September 2022.

Sucheta Ghosh:
The 36th AAAI Conference on Artificial Intelligence (AAAI), Virtual Conference, 22 February- 1 March 2022 (Program Committee). Co-organizer *"Identification, collection and review of datasets for the assessment and development of supervised machine learning methods for biomedical applications"* at ELIXIR All Hands Meeting, Amsterdam, Netherlands, 6-9 June 2022. The 23rd Annual Conference of the International Speech Communication Association- Interspeech, Incheon Korea, 18-22 September 2022 (Technical Program Committee). The 44th Annual Meeting of the Cognitive Science Society, Toronto, Canada, 27–30 July 2022.

Nikos Gianniotis:
Astrohack week 2022, Heidelberg, Germany, 17-21 October 2022.

Martin Golebiewski:
Host and chair of Committee Meetings of ISO/TC 276 Biotechnology working group WG5 *"Data Processing and Integration"*, online, 24-28 January and 3-11 November 2022.Chair of the Workshop *"Wie mache ich Health-Daten FAIR: Benutzung der Forschungsdateninfrastrukturen von NFDI4Health und FAIR4Health"*, DMEA - Satellitenveranstaltung 2022 von GMDS und BVMI, Berlin (Germany), 25 April 2022. NFDI4Health Task Force COVID-19-Symposium, Fraunhofer-Institut für Digitale Medizin MEVIS, Bremen, Germany, 13 June 2022 (Organization Committee).

Martin Golebiewski, Wolfgang Müller:
Annual NFDI4Health Consortium Meeting, Studio Villa Bosch, Heidelberg, Germany, 20-21 October 2022.

Ganna Gryn'ova:
Scientific board member: European Young Chemists' Meeting 2022 (EYCheM2022).

Saskia Haupt:
Scoring team member and session chair: Best of visual posters. 6th EHTG Meeting 2022, Palma de Mallorca, Spain, 29 September 2022 – 1 October 2022. Organizing committee: Joint ECMTB & SMB 2022, Heidelberg, Germany, 19–23 September 2022. Assistant editor of the official Mathematical Oncology community website.

Saskia Haupt, Vincent Heuveline, Aysel Ahadova:
Mini symposium Organization *"How to get a big picture of cancer? Recent advances in mathematical oncology"*, Joint ECMTB & SMB 2022, Oncology subgroup, Heidelberg, Germany, 19–23 September 2022.

Saskia Hekker:
Scientific organizing committee *"Fundamental stellar parameters from asteroseismology in the era of Gaia"*, Aarhus, Denmark, 13-15 June 2022.

Kai Polsterer:
Co-organizer of IAU Symposium 368, *"Machine Learning in Astronomy: Possibilities and Pitfalls"*, Busan, Rep. of Korea, 2-11 August 2022. Co-organizer of E-Science and E-Infrastructure Splinter at AG-Tagung, Bremen, Germany, 12-16 September 2022.

Fabian Schneider:
Main organizer of VFTS Workshop, HITS, Heidelberg, Germany, 20-22 June 2022.

Alexandros Stamatakis:
Organizer of 2022 Computational Molecular Evolution Summer School, Welcome Trust Genome Campus, Hinxton, UK.

Michael Strube:
Program Co-Chair of CODI 2022, The Third Workshop on Computational Approaches to Discourse at COLING 2022, Gyeongju, Republic of Korea, 16-17 October 2022. Program Co-Chair of CODI-CRAC 2022 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue at CODI 2022, Gyeongju, Republic of Korea, 17 October 2022.

Jan Stühmer:
Senior Program Committee member at the 36th AAAI Conference on Artificial Intelligence (AAAI), Virtual Event, 22 February – 1 March 2022. Scientific Q&A at the HLF film festival Mathematics – Informatics, 9 November 2022

Alexandros Tsengenes:
Contribution to the *"Neurotrophin mimetics"* Wikipedia page.

Rebecca Wade:
Chair, SIMPLAIX Inaugural Symposium, Heidelberg, 12 April 2022.Chair, 23rd European Symposium on Quantitative Structure-Activity Relationship (EuroQSAR), Heidelberg, 26-30 September 2022.

Ulrike Wittig:
Co-organizer of Mini Symposium *"Research Data Management"* at ELIXIR All Hands Meeting, Amsterdam, Netherlands, 6-9 June 2022.

Alexander Zeilmann:
Assistant editor of the official Mathematical Oncology community website.

Other contributions

Angela Michel:
"Universe inside – how computational astrophysics has changed traditional astronomy", poster presentation at ESOF 2022, Leiden, Netherlands, July 2022.

Peter Saueressig:
"HITS Journalist in Residence" (together with Carl Smith, Sydney, Australia), best practice presentation, ERC Host Institution Communication Partners Meeting (online), 31 March 2022. *"Wenn Journalisten im Haus sind – eine nachhaltige Geschichte?"*, Talk at the *„Europäische Akademie für Wirtschaft und Recht"* (online), 31 May 2022.

Elaine Zaunseder:
Podcast on *"Machine learning in newborn screening"*, JIMD Podcast, June 2022.

10.5 Awards

Camilo Aponte-Santamaria:
"Postdoc Take-Off Grant" on the topic of *"Energetics and permeation of photo-resists used for 3D-laser printing across biological lipid bilayers"*, awarded by 3DMM2O Cluster.

Christina Athanasiou:
Poster prize, 23rd European Symposium on Quantitative Structure-Activity Relationship, Heidelberg, Germany, 26-30 September 2022.

Michelle Ernst:
UZH postdoc grant 2022, University of Zurich, Switzerland.

Manuel Glaser:
Poster Prize, Informatics for Life - Yearly Meeting 2022, Heidelberg, Germany, 19 November 2022.

Ganna Gryn'ova:
ERC Starting Grant *"PATTERNCHEM: Shape and Topology as Descriptors of Chemical and Physical Properties in Functional Organic Materials"*, European Research Council, 2021.

Saskia Haupt:
Best visual poster: *"Why the combination of modeling and machine learning could be the future direction in mathematical oncology"* 6th EHTG Meeting 2022, Palma de Mallorca, Spain, 29 September 2022 – 1 October 2022.

Markus Kurth:
"Postdoc Take-Off Grant" on the topic of *"Artificial collagen peptides with radical scavenging functionalization and measurements of mechanoradicals in situ"*, awarded by 3DMM2O Cluster.

Giulia Paiardi:
"Add-on Fellowship for Interdisciplinary Life Science" 2022, Joachim Herz Foundation for project on *"Combining AI and molecular simulation for anticancer peptide and peptidomimetic design."*

Alexandros Stamatakis:
Highly Cited Researcher in the biology and biochemistry category, Clarivate Analytics, 2022.

11 Boards and Management

The HITS Scientific Advisory Board (December 2022). From left to right: Jeffrey Brock, Adele Goldberg, Gesa Schönberger (HITS Managing Director), Thomas Lengauer, Frauke Gräter (HITS Scientific Director), Dieter Kranzlmüller, Victoria Stodden, Tilmann Gneiting (HITS Deputy Scientific Director).

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 2022, 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. 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

HITS, the Heidelberg Institute for Theoretical Studies, was established in 2010 by 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, mathematical, and computer sciences. Major research directions include complex simulations across scales, making sense of data, and enabling science via computational research. Application areas range from molecular biology to astrophysics.
An essential characteristic of the Institute is interdisciplinarity, implemented in numerous cross-group and cross-disciplinary projects. The base funding of HITS is provided by the Klaus Tschira Foundation.

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
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 | © 2023 HITS gGmbH

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