HITS
Heidelberg Institute for Theoretical Studies

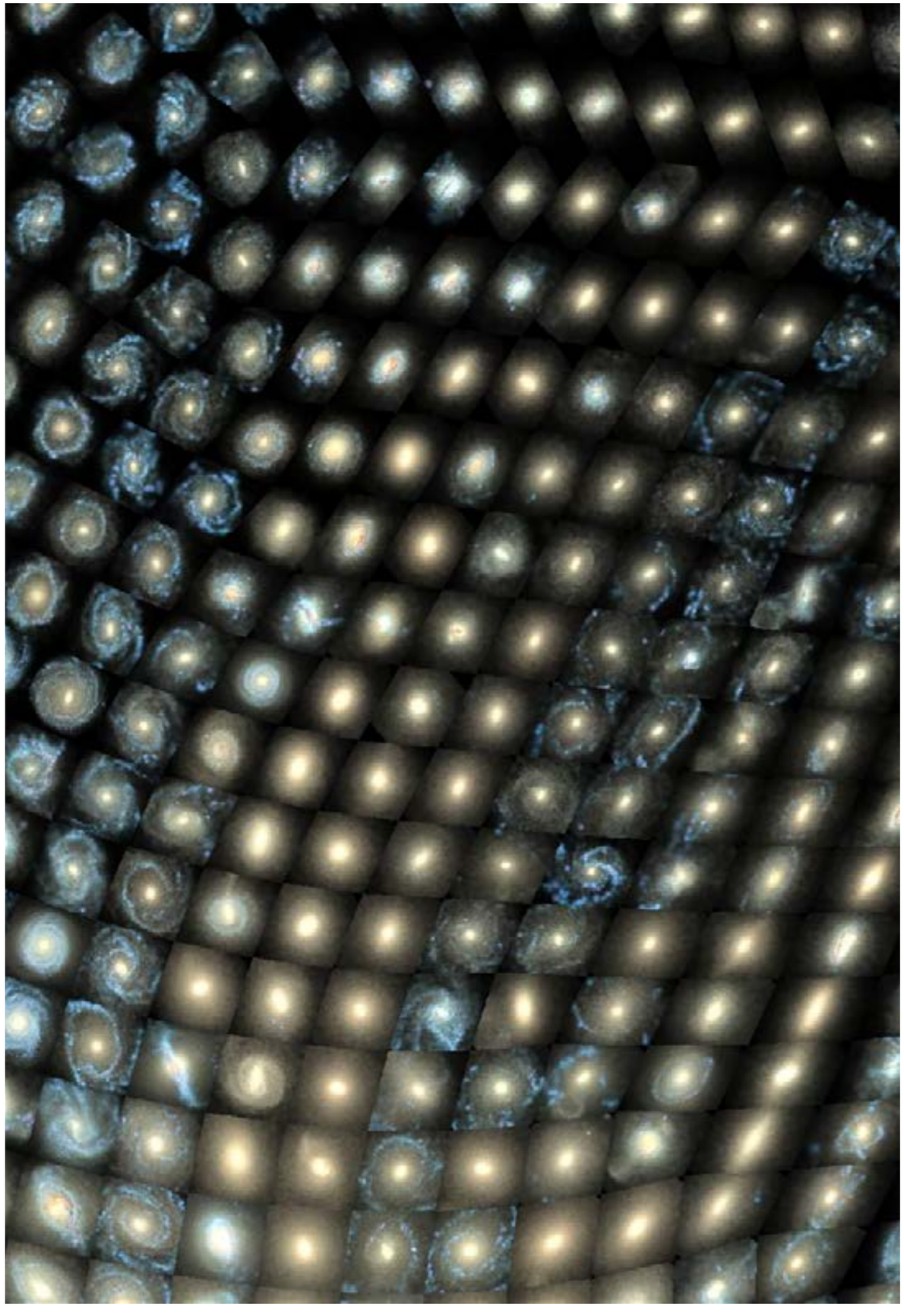
2023

Annual Report
Jahresbericht

# Think beyond the limits!

Simulations are the best approximations to laboratories in astrophysics and cosmology, but the large size and high dimensionality of their outputs severely limit their interpretability. The Astroinformatics group (AIN) is developing software tools to enable intuitive explorative access, interpretation, and analysis of large cosmological simulations in preparation for the next generation of supercomputers. These projects will produce simulation datasets one thousand times larger than the largest available today, creating enormous challenges to their scientific exploitation. The image shows a snapshot of the interactive visualization of the galaxies in the IllustrisTNG simulation (https://www.tng-project.org/) created with the Spherinator tool (https://github.com/HITS-AIN/Spherinator). The tool automatically projects arbitrarily large datasets of simulated galaxies onto a hierarchical 2-dimensional similarity space that allows intuitive exploration, catalog creation, and data analysis. An interactive demo of the prototype is available at https://space.h-its.org.

Simulationen stellen die bestmögliche Annäherung an Labore in der Astrophysik und Kosmologie dar. Allerdings schränken das Ausmaß und die Größe hohe Dimensionalität ihrer Ergebnisse die Interpretierbarkeit stark ein. Die HITS-Gruppe Astroinformatik (AIN) entwickelt Software-Tools, die einen intuitiven, explorativen Zugang zu großen kosmologischen Simulationen sowie deren Interpretation und Analyse ermöglichen, um auf die nächste Generation von Supercomputern vorbereitet zu sein. Dort werden künftig Simulationsdatensätze erzeugt, die tausendmal größer sind als die größten heute verfügbaren, was enorme Herausforderungen für ihre wissenschaftliche Nutzung mit sich bringt. Das Bild zeigt einen Schnappschuss der interaktiven Visualisierung der Galaxien in der IllustrisTNG-Simulation (https://www.tng-project.org/), die mit dem Tool Spherinator (https://github.com/HITS-AIN/Spherinator) erstellt wurde. Es projiziert automatisch beliebig große Datensätze von simulierten Galaxien auf einen hierarchischen zweidimensionalen Ähnlichkeitsraum, der eine intuitive Erkundung, Katalogerstellung und Datenanalyse ermöglicht. Eine interaktive Demo des Prototyps ist unter https://space.h-its.org verfügbar.



# 1 Think beyond the limits!

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

*Prof. Dr. Tilmann Gneiting*
*(Scientific Director / Wissenschaftlicher Direktor)*

Sometimes, an outside perspective helps to put things into focus: "You deliver the vision of your founder Klaus Tschira, making it a reality." This quote from Olexandr Isayev (Klaus Tschira Visiting Professor at HITS in spring 2023) serves as both an impulse and an affirmation for us.

We are proud of what our scientists and research groups achieved last year and are pleased to be able to give you a brief overview of the scientific highlights below:

At the beginning of the year, Friedrich Röpke – leader of the Physics of Stellar Objects (PSO) group – received an ERC Advanced Grant for his project "ExCEED: Explaining Common-Envelope Evolution and Dynamics in binary stellar systems." That means that six out of thirteen research groups at HITS were working with ERC grant funding in 2023.

Moreover, Alexandros Stamatakis – leader of the Computational Molecular Evolution (CME) group – founded the Biodiversity Computing research group on Crete in January 2023, which is funded by the EU's ERA Chair Program for a period of five years (see Chapter 2.3). During this time, Alexandros will continue to supervise his group at HITS from Crete and will participate in the HITS activities as an associated group leader.

In addition to the successful acquisition of third-party funding, the quality of the Institute's research is also reflected in the staff who join and leave HITS as well as in the guest researchers. Our second Independent Postdoc, Fabian Grünewald, took up his position at HITS in the fall of 2023 and works with the Molecular Biomechanics research group on developing computer-assisted high-throughput protocols that can enable the design of RNA nanoparticles as therapeutics (see Sections 2.14 and 6.2). Furthermore, our two Klaus Tschira Guest Professors – Olexandr Isayev from Carnegie Mellon University (Pittsburgh, USA) and Philipp Podsiadlowski (Oxford, UK) – used their time at HITS last year to intensify their collaboration with the research groups of Ganna (Anya) Gryn'ova (Computational Carbon Chemistry) and Fabian Schneider (Stellar Evolution Theory) (see Chapter 6.1). Another highlight in 2023 was Indian-American science journalist Anil Ananthaswamy's stay at HITS as Journalist in Residence, starting in April. During his time at the Institute, he gave a very well-attended public lecture on large language models such as Chat GPT ("ChatGPT and its ilk") at the Heidelberg Mathematics Informatics Station MAINS (see Chapter 4).

We also had to say a fond farewell last year to Anna Wienhard, head of our associated Groups and Geometry (GRG) group, who accepted an appointment as Director at the Max Planck Institute for Mathematics in the Sciences in Leipzig in fall 2022. We officially bid her farewell and honored her as a HITS Fellow at a special mathematics colloquium on 8 July 2023 (see Chapter 5.3). We are grateful to Anna and proud of her time at HITS, where she worked as a group leader for seven years. We wish her all the best in the next chapter of her career.

At HITS, we are highly dedicated to sustainability. We are aware that our research – and especially high-performance computing – consumes resources and energy. In fact, well over 50% of the electricity consumption at HITS is currently used for high-performance computing. By analyzing our energy consumption and raising awareness among our employees, we were able to take first steps toward future-oriented and more sustainable research. Since March 2023, we have been supported by a sustainability advisor with a focus on high-performance computing. Test runs for power limiting the HITS compute cluster were designed, carried out, and evaluated together with our IT group.

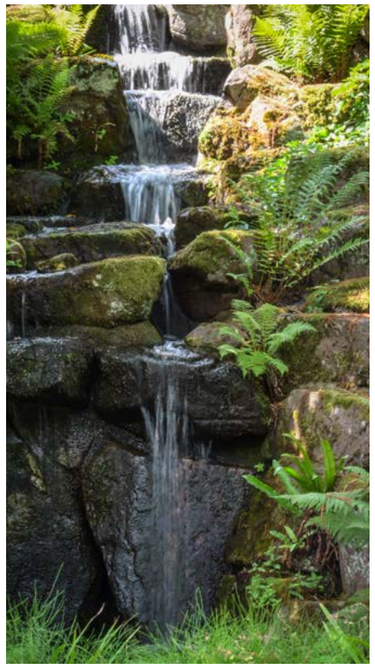
Reducing energy consumption is one thing, but actively investing in climate protection is another. That's why HITS has been offsetting air travel since 2021 and car travel for business purposes since 2023.

Last year, a team of astrophysicists at HITS playfully approached the topic of sustainability by developing a board game called "Habitable", in which the aim is to create a planet that remains habitable despite changing environmental conditions. With this concept, the team was among the winners of the "Our Universe" competition – which is organized by Wissenschaft im Dialog – and won a prize for the further development and implementation of the game (see Chapters 2.12 and 4).

We are delighted that sustainability is attracting a great deal of interest from our researchers. This can also be seen in the lively participation in colloquium lectures and scientific seminars on topics such as the carbon footprint of astrophysical research (Jürgen Knödlseder, IRAP, GAHEC, Toulouse, France), probabilistic forecasts for climate change (Adrian Raftery, University of Washington, Seattle, USA, see Chapter 5.2), and energy-aware coding (Vincent Heuveline, HITS DMQ group).

There is no doubt that all this is just the beginning. Furthermore, we will continue to do everything we can in the coming decades to enable, facilitate, and support sustainable and excellent research in order to realize the vision of our founder, Klaus Tschira: Think beyond the limits!

We hope you enjoy reading our Annual Report 2023!



Manchmal hilft der Blick von außen, um die eigene Wahrnehmung zu schärfen: „Ihr setzt am HITS die Vision eures Gründers Klaus Tschira in die Tat um." Dieses Zitat von Olexandr Isayev, Klaus Tschira Gastprofessor im Frühjahr 2023, ist für uns Ansporn und Bestätigung zugleich.
Wir sind stolz auf das, was unsere Wissenschaftler*innen und Forschungsgruppen im vergangenen Jahr geleistet haben, und freuen uns, Ihnen an dieser Stelle einen kurzen Überblick über die wissenschaftlichen Highlights geben zu dürfen:
Anfang des Jahres erhielt Friedrich Röpke, Leiter der Gruppe „Physics of Stellar Objects" (PSO), einen ERC Advanced Grant für sein Projekt „ExCEED: Explaining Common-Envelope Evolution and Dynamics in binary stellar systems". Damit arbeiteten 2023 sechs von dreizehn Forschungsgruppen am HITS mit Mitteln eines ERC Grants.
Der Leiter der Gruppe „Computational Molecular Evolution" (CME), Alexandros Stamatakis, gründete im Januar 2023 die Forschungsgruppe „Biodiversity Computing" auf Kreta, die durch das ERA Chair Programm der EU über einen Zeitraum von fünf Jahren finanziert wird (siehe Kapitel 2.3). Während dieser Zeit betreut Alexandros Stamatakis seine Gruppe am HITS von dort aus und beteiligt sich als assoziierter Gruppenleiter an den Aktivitäten des Instituts.
Neben erfolgreich eingeworbenen Drittmitteln spiegeln auch die personellen Zu- und Abgänge sowie die Gastwissenschaftler*innen am HITS die Forschungsqualität des Instituts wider. Unser zweiter Independent Postdoc, Fabian Grünewald, nahm im Herbst seine Tätigkeit am HITS auf, wo er in Zusammenarbeit mit der Forschungsgruppe „Molecular Biomechanics" computergestützte Hochdurchsatzprotokolle entwickelt, die das Design von RNA-Nanopartikeln als Therapeutika ermöglichen (siehe Kapitel 2.14 und 6.2). Die beiden Klaus Tschira Gastprofessoren Olexandr Isayev von der Carnegie Mellon University in Pittsburgh und Philipp Podsiadlowski aus Oxford nutzten ihre Zeit am HITS, um ihre Zusammenarbeit mit den Forschungsgruppen von Anya Gryn'ova (Computational Carbon Chemistry) bzw. Fabian Schneider (Stellar Evolution Theory) zu intensivieren (siehe Kapitel 6.1).
Ein weiteres Highlight 2023 war der Aufenthalt des indisch-amerikanischen Wissenschaftsjournalisten Anil Ananthaswamy, der als Journalist in Residence im April ans HITS kam. Während seiner Zeit am Institut hielt er einen sehr gut besuchten öffentlichen Vortrag über große Sprachmodelle wie Chat GPT ("ChatGPT and its ilk") in der Heidelberger Mathematik-Informatik-Station MAINS (siehe Kapitel 4).

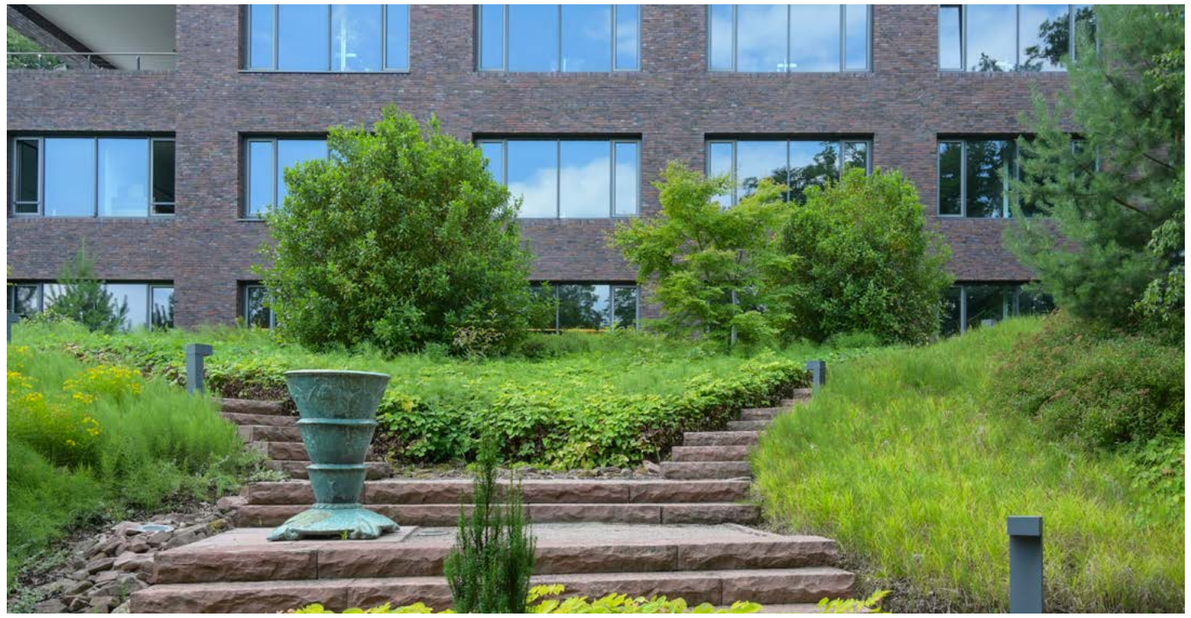
Doch es hieß auch Abschied nehmen im vergangenen Jahr. Nachdem Anna Wienhard, Leiterin unserer assoziierten Gruppe „Groups and Geometry" (GRG), im Herbst 2022 einen Ruf als Direktorin an das Max-Planck-Institut für Mathematik in den Naturwissenschaften in Leipzig angenommen hatte, wurde sie am 8. Juli im Rahmen eines speziellen Mathematik-Kolloquiums offiziell verabschiedet und als HITS Fellow ausgezeichnet.



Wir sind zugleich dankbar und stolz auf Anna, die sieben Jahre am HITS als Gruppenleiterin wirkte, und wünschen ihr für ihren nächsten Karriereschritt alles Gute.
Ein weiteres Thema, das uns am HITS bewegt, ist Nachhaltigkeit, und wir sind uns bewusst, dass unsere Forschung und insbesondere die Nutzung von Hochleistungsrechnern Ressourcen und Energie verbrauchen. Weit über 50% des Stromverbrauchs am HITS wird derzeit für Hochleistungsrechner aufgewendet. Mit der Analyse des Energieverbrauchs und der Sensibilisierung unserer Mitarbeiter*innen haben wir erste Schritte in Richtung einer zukunftsorientierten und nachhaltigeren Forschung unternommen. Seit März 2023 unterstützt uns dabei ein Nachhaltigkeitsbeauftragter mit Schwerpunkt Hochleistungsrechnen. Gemeinsam mit unserer IT-Gruppe wurden Testläufe zur Leistungsbegrenzung des HITS-Rechenclusters konzipiert, durchgeführt und ausgewertet.
Die Senkung des Energieverbrauchs ist ein Aspekt, aktiv in den Klimaschutz zu investieren ein anderer. Deshalb kompensiert HITS seit 2021 Flugreisen und seit 2023 die bei der dienstlichen Nutzung von PKWs angefallenen Kilometer.
Dem Thema spielerisch angenähert hat sich ein Team von Astrophysiker*innen am HITS:

Sie entwickelten ein Brettspiel namens „Habitable", bei dem es darum geht, einen Planeten zu erschaffen, der trotz wechselnder Umweltbedingungen bewohnbar bleibt. Mit diesem Konzept gehörten sie zu den Preisträger*innen des Wettbewerbs „Unser Universum" von Wissenschaft im Dialog und gewannen einen Preis für die Weiterentwicklung und Umsetzung des Spiels (siehe Kapitel 2.12 und 4).
Wir freuen uns, dass Nachhaltigkeit bei unseren Forschenden auf großes Interesse stößt. Dies zeigt auch die rege Teilnahme an Kolloquiumsvorträgen und wissenschaftlichen Seminaren, die sich u.a. mit dem Kohlenstoff-Fußabdruck der astrophysikalischen Forschung (Jürgen Knödlseder), Wahrscheinlichkeitsprognosen für den Klimawandel (Adrian Raftery, siehe Kapitel 5.2) und energiebewusstem Programmieren (Vincent Heuveline, DMQ Gruppe am HITS), befassten.

Zweifellos ist all dies erst der Anfang, und wir werden auch in den kommenden Jahrzehnten alles daran setzen, nachhaltige und exzellente Forschung zu ermöglichen, zu erleichtern und zu unterstützen, um die Vision unseres Gründers Klaus Tschira umzusetzen: Think beyond the limits!
Wir wünschen Ihnen beim Lesen des Jahresberichts viel Vergnügen!



# 2 Research
# 2.1 Astroinformatics (AIN)

## Group leader
Dr. Kai Polsterer

## Team
Iliana Isabel Cortés Pérez (PhD student; since September 2023)
Mariia Demianenko (visiting scientist; since September 2023)
Dr. Nikos Gianniotis (staff scientist)
Fenja Kollasch (master's student; until March 2023)
Sebastian Müller (since October 2023)
Dr. Franciso Pozo Nuñez
Johanna Riedel (bachelor's student)
Solomiya Serkiz (bachelor's student)
Dr. Sebastian Trujillo Gomez
Renuka Velu (master's student; since February 2023)

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 at 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 analysis, 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 both to the analysis of data in large upcoming survey projects, such as SKA, Gaia, LSST, and Euclid, as well as to the next generation of exa-scale simulations.

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



## Reconstructing Spectra From Photometry

Two common ways that spectral properties of objects can be observationally accessed in astronomy are via spectroscopy and photometry. In spectroscopy, the incoming photons are dispersed, and the recorded fluxes are assigned their corresponding wavelengths. Spectroscopic data therefore contain a large amount of information, but their acquisition requires long integration times. On the other hand, in photometry, fluxes are measured as they pass through dedicated filters. This process results in data of considerably lower spectral resolution, but it has the advantage of obtaining measurements much more quickly. We are currently developing a probabilistic model that reconstructs spectroscopic representations from photometric data. The method can cope with any filter system, which means that photometric data that are observed with any set of filters can be used to partially reconstruct the latent spectral information. As a result, the method makes it possible to compare photometric data observed in different filters by first projecting them in a common latent space and then transferring them to any other desired set of filters.

Probabilistic principal component analysis (PPCA) is a widely used approach that simplifies high-dimensional datasets by compressing them into fewer dimensions. PPCA is linear, and unlike example neural networks, it is easy to interpret. Unfortunately, the standard PCA is not applicable in our case as it assumes that all observed spectra are subjected to the same measurement uncertainty. Instead, we are developing a so-called heteroscedastic PPCA, which can deal with the fact that the uncertainties differ between different spectra as well as across the individual measurements along the wavelengths. Our model allows us to learn a general representation of spectra that can be used to map a spectrum in a latent space with fewer dimensions. Conversely, we can also map a point from this latent space back to a spectrum and obtain a reconstruction.

We have taken advantage of the fact that (a) the PPCA mappings are linear and (b) spectra can also be linearly mapped to their photometric representation. Using these properties, we can linearly map between the photometry and the latent low-dimensional space of the spectra, as shown in Figure 1. By following the arrows and first transferring photometric data into the latent space, these data can

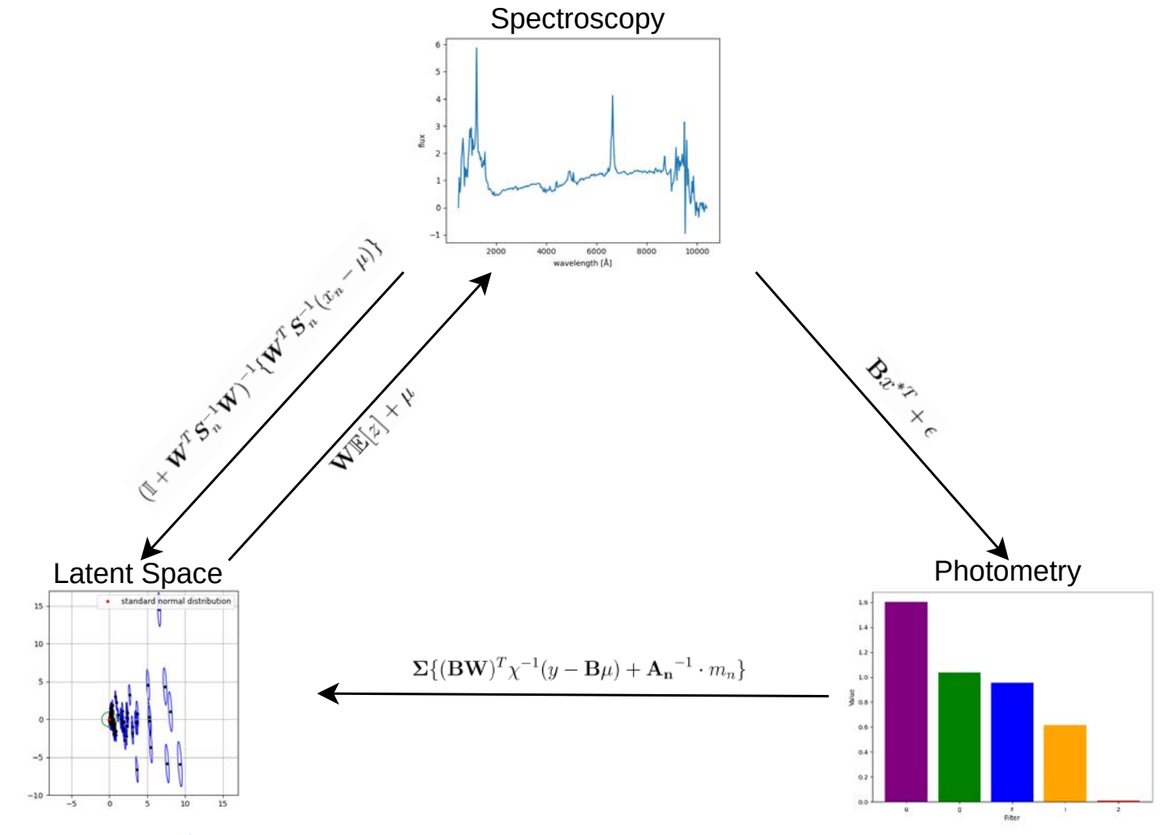



*Figure 1: Illustration of possible transitions between photometry, spectroscopy, and the latent space.*

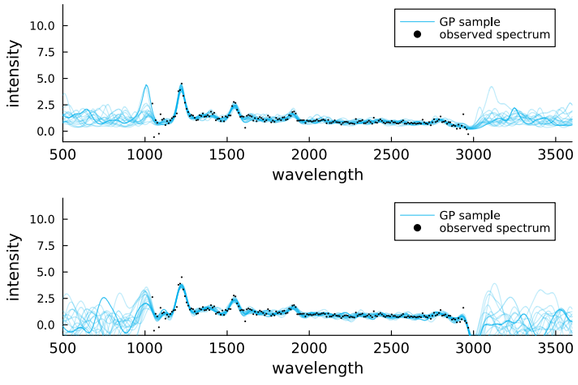



*Figure 2: Black dots represent measured spectra, and blue curves represent predictions that have been sampled from the Gaussian process. The plot above displays the positively constrained predictions given by our model, while the plot below depicts the unconstrained predictions of a standard Gaussian process.*

then be converted into a full spectrum. This process works for any filter system whose filter curves are available (and it also works between filter systems).

Gaussian processes (GPs) excel in regression problems and provide practitioners with the possibility of modeling particular aspects of their data (e.g., periodicity, trends) by selecting or even designing an appropriate kernel. Roughly speaking, a kernel quantifies the way in which observed data are similar to one another. GPs are particularly useful in small dataset settings because they are capable of producing predictive distributions that inform us not only of the estimate of a value in question, but also of how uncertain this estimate might be. This information is important when the estimate needs to be propagated into further computations or decisions.

Even though GPs are flexible in accommodating complex non-linear relationships, it is often difficult to employ them under certain constraints. We are interested in modeling astronomical spectra with GPs under the constraint that spectra be positive-valued. We consider negative-valued observations to be the result of Gaussian measurement noise corrupting positive-valued observations that we cannot directly observe. There are techniques that attempt to constrain GPs, such as warped Gaussian processes which apply an invertible non-linear transformation to the density of GPs but fail to properly accommodate the noise on the data. Therefore, we followed a Bayesian approach based on variational learning that analytically propagates the Gaussian process prior a non-linearity that results in a regression model that generates only positive values. Our solution is simple and flexible and can be easily extended for handling larger datasets (via inducing points).

Figure 2 illustrates the application of our approach when modeling a spectrum. Black dots represent the available spectral measurements, while blue curves represent predictions of the Gaussian process. The upper subplot reveals that predictions are indeed all constrained to be positive, even when observations are negative or are very close to zero. On the other hand, the standard Gaussian process produces negative values both at wavelengths where data have been observed and at wavelengths where we lack observations. Propagating such predictions that do not support negative values in order to further our calculations is paramount to the soundness of the statistical methods.

## Insights Into Active Galactic Nuclei From Light Curve Modeling and Time Delay Estimation

The interplay between the inner regions of active galactic nuclei (AGN) – including the broad-line region (BLR), the accretion disc (AD), and the connection of both regions with the supermassive black hole (SMBH) – is pivotal in the application of AGN in cosmology. The complexities of these interactions are not yet fully understood, which makes the modeling of such regions a subject of active research.

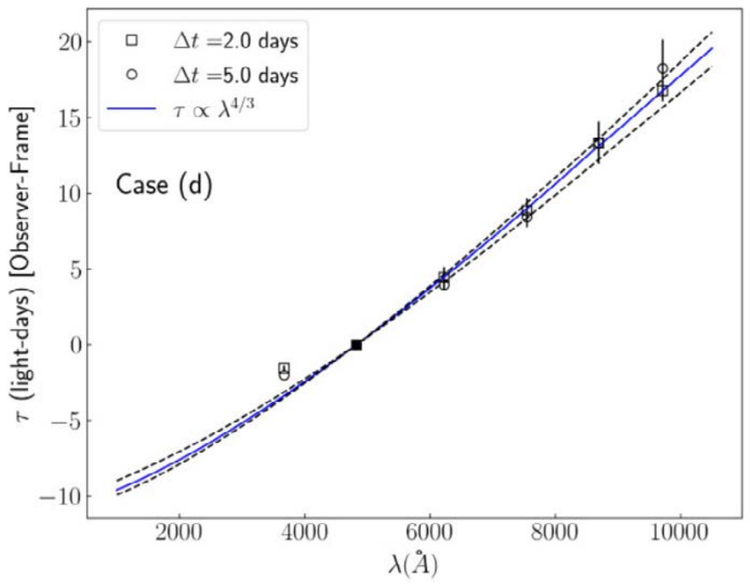



*Figure 3: AD time delay spectrum as predicted from the transfer functions for a quasar at a redshift of 1.5. Filled squares and open circles denote the recovered delays at a sampling of 2 and 5 days, respectively. The dotted lines show the delay spectrum obtained for a black hole mass with 30% uncertainty.*

Our study concentrates on modeling AGN light curves under various observational conditions in order to more precisely determine the contribution of each component. Specifically, we aim to assess the feasibility of accurately recovering time delays between light curves across different continuum and BLR regions. These time delays – which are observed between different wavelengths – provide a preliminary approximation of the sizes of the AD and the BLR. Our results were discussed in the context of the next Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory, which serves as a preparatory basis for the upcoming AGN survey at the Cerro Armazones Observatory.

Our preliminary findings indicate that a minimum signal-to-noise ratio (S/N) of 100 with a BLR emission line contribution of less than 10% allows for recovering time delays with accuracies of 5% and 10% for time sampling intervals of 2 and 5 days, respectively. This finding is particularly relevant for quasars in the 1.5 < z < 2.0 range. For quasars at z < 1.5, achieving an accuracy of 10–20% is possible only if the BLR emission line contribution is below 5% (see Figure 3). Increasing the S/N does not significantly improve the results. Improving the time sampling and reducing the contamination by BLR emission lines is the only way to improve the accuracy of the time delay.

Beyond the traditional time delay estimation methods, such as the interpolated cross-correlation function (ICCF) and Chi-squared minimization (Chi2), we created an optimized ICCF approach. This novel method integrates Gaussian processes with a cross-validation technique in order to estimate delays. The proposed method delivers a posterior distribution for the delay that accounts for observational noise and the non-uniform sampling of the light curves. This feature enables us to fully quantify the uncertainty on the delay and to propagate this uncertainty to subsequent calculations of dependent physical quantities, such as black hole masses. The method delivers out-of-sample predictions, which enables us to subject it to model selection. Moreover, it can also calculate the joint posterior delay for more than two light curves. Figure 4 displays an application of our approach to the light curves of the AGN MCG+08-11-011 as part of a reverberation mapping campaign for the AD. We provide the algorithms and examples of their application as part of a Julia GPCC package.

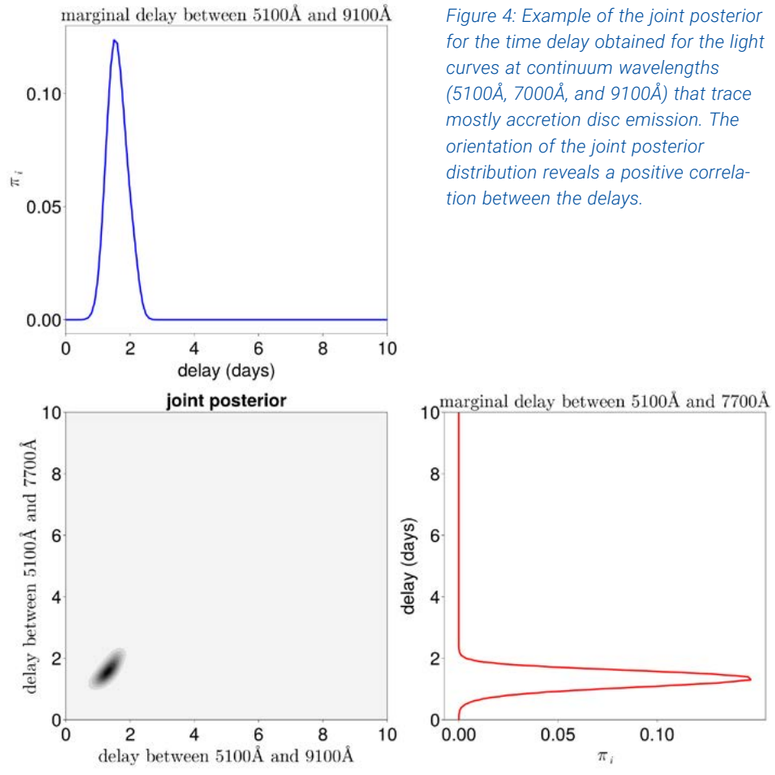



*Figure 4: Example of the joint posterior for the time delay obtained for the light curves at continuum wavelengths (5100Å, 7000Å, and 9100Å) that trace mostly accretion disc emission. The orientation of the joint posterior distribution reveals a positive correlation between the delays.*

## Representation Learning for Explorative Knowledge Discovery in Cosmological Simulations

Cosmological hydrodynamical simulations are excellent numerical laboratories for understanding the formation of galaxies and large-scale structures. They provide a highly detailed realization of structures in the Universe across a vast range of spatial and temporal scales that begin shortly after the Big Bang and extend up to the present day. The simulation output consists of a series of more than one hundred snapshots of the evolution of a large representative volume of the Universe that contains all its main components: namely dark matter, gas, stars, and black holes. Each simulation snapshot is information-rich and contains 3D positions and velocities as well as dozens of physical properties for each component (e.g., gas density, temperature, metallicity, elemental abundances). Figure 5 (next page) illustrates the complexity of these datasets.

State-of-the-art large-volume cosmological simulations model the evolution of the Universe using more than 1011 resolution elements and follow the formation of more than one million structures in detail across a dynamic range of ~7 orders of magnitude in space and time. The computational cost of these simulations is extremely high and lies somewhere in the order of 100 million CPU hours, with

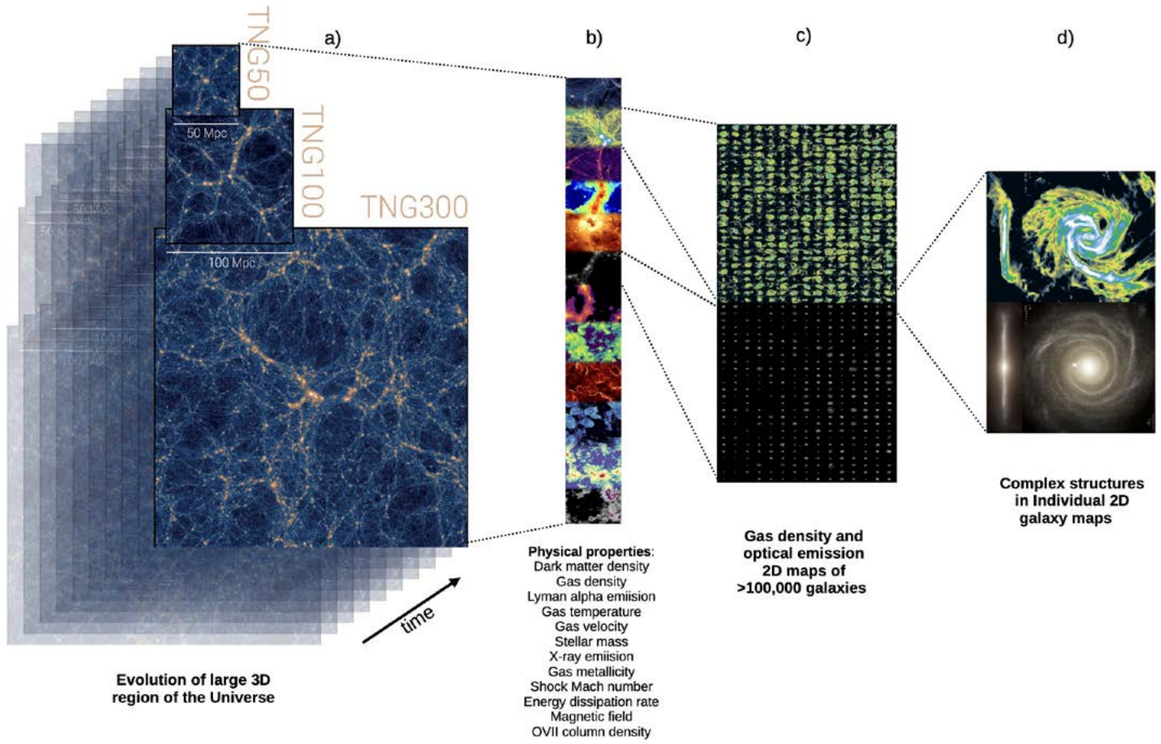



*Figure 5: Cosmological hydrodynamical simulations like "Illustris TNG" provide a detailed realization of structures in the Universe from the Big Bang until today, producing a high complexity of datasets.*

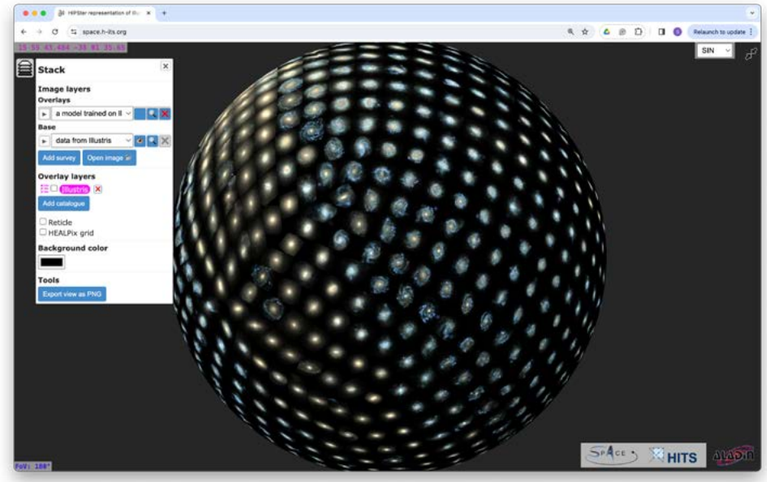
the output data being measured in petabytes. The volume and complexity of these datasets already far exceed the exploration, synthesis, and interpretation capacity of humans. In addition, architectures and data formats vary greatly across simulation codes, thereby imposing a barrier to the application of code-specific tools more generally to all cosmological datasets. In the exascale computing era, simulations are expected to increase in size by a factor of one thousand, thereby rendering traditional human- and code-centered simulation exploration and analysis techniques obsolete.

The approach to the problem of increasing simulation data volume and complexity has changed very little over the past three decades ever since cosmological simulations became widespread tools. Simulation output is typically represented in catalogs by first collapsing the rich multidimensional data onto simplified representations of galaxy properties (e.g., single values or spherically averaged profiles). These simplifications allow the large datasets to be synthesized for analysis by researchers. However, describing these high-dimensional structures using single values or spherically averaged profiles ignores most of the detailed morphological data and wastes valuable information on the physics behind the formation of galaxies.

Representation learning (RL) methods offer an ideal solution to simulation interpretability. These methods learn compact representations of complex high-dimensional data that efficiently compress the information in order to allow for easier visualization, exploration, and interpretation. RL also offers a powerful way to understand the intrinsic distribution of large datasets and to sample from these datasets in order to generate new data points. The models can be trained using raw data, thereby avoiding the costly need to produce labels.

"Spherinator" is a software tool that we developed in order to address the challenge of simulation analysis in the exascale era for the SPACE (Scalable Parallel and distributed Astrophysical Codes for Exascale) Center of Excellence. The tool takes the raw simulation data as input, performs all the preprocessing under the hood, and trains a model to learn a compact interpretable representation of galaxies in arbitrary user-selected physical components and fields. Instead of collapsing the galaxy properties to single values that are guided only by human intuition, "Spherinator" uses a generative deep learning algorithm to learn the most efficient representation of the simulated galaxies in a low-dimensional latent space. The latent space can be visualized by projecting the data onto a tiling of a spherical surface that can be easily explored by humans and that has the capacity to be extended to arbitrarily large datasets using multiple hierarchical levels. This ability guarantees that the data can be inspected interactively, even for the largest cosmological simulations of the exascale era. The visualization is lightweight, which allows it run on any laptop via a web server, and it provides the functionality to interactively inspect and select subsamples of data for local analysis. Figures 6 and Figure 7 present examples of the interface.

In addition to their challenging interpretability, cosmological simulation data are difficult to find and access. In order to adopt the “FAIR” data principles, these large projects should make their data easily findable, accessible, interoperable, and reusable. Our approach addresses this need by enabling users to find existing datasets, to interactively explore them, to compare them, and to directly access their raw and post-processed data products, all of which can be done using tools that are agnostic to the simulation and the code used to run it. Beyond our tool’s broad applicability to all existing cosmological codes, we expect it to prove useful across many disciplines when it comes to exploring simulations that model the evolution of a large number of structures.



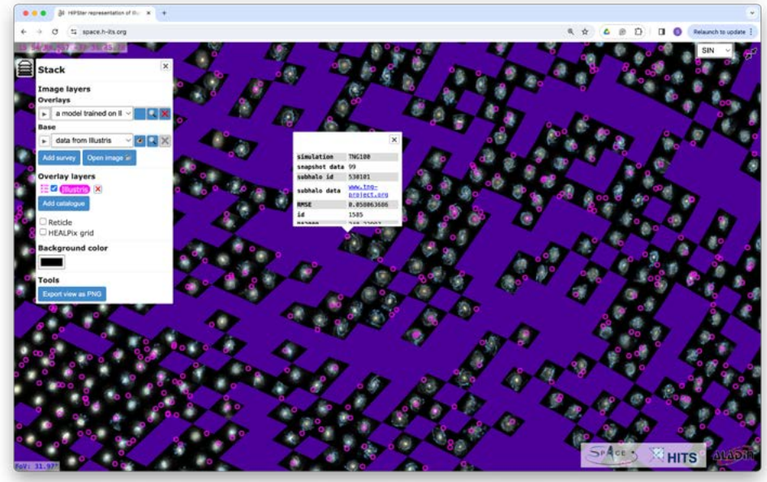




*Figure 6 and Figure 7: Two interface examples of the software tool "Spherinator."*

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

Die **Astroinformatik** Gruppe beschäftigt sich mit den Herausforderungen, die durch die Analyse und Verarbeitung dieser komplexen, heterogenen und großen Daten entstehen. In der Astronomie beschäftigen uns die Fragestellungen im Bereich der Galaxienentwicklung sowie die extremen physikalischen Vorgänge, wie man sie z. B. in der Umgebung von aktiven supermassereichen schwarzen Löchern in den Zentren von Galaxien findet. Auf diesen Fragestellungen basierend, entwickeln wir neue Methoden und Werkzeuge, die wir frei zur Verfügung stellen. In der Informatik liegt unser Interesse 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 befindende Projekte, wie SKA, Gaia, LSST und Euclid sowie Daten aus exa-scale Simulationen.

Unser Ziel ist es, einen möglichst unvoreingenommenen Zugang zu dieser enormen Menge an Information zu gewährleisten.

# 2 Research

# 2.2 Computational Carbon Chemistry (CCC)

## Group leader

Dr. Ganna Gryn'ova

## Team

Alessandro Calzolari (PhD student; since March 2023)
Sebastian Cremer (student intern, Heidelberg University; November–December 2023)
Dr. Christopher Ehlert
Dr. Michelle Ernst (until March 2023)
Rostislav Fedorov (PhD student)
Fabian Grieser (student intern, Heidelberg University; November–December 2023)
Anastasiia Grynyova (student intern, Heidelberg University; September–October 2023)
Carina Herrle (student intern, Heidelberg University, September-October 2023; student assistant; since November 2023)
Prof. Olexandr Isayev (Klaus Tschira Guest Professor, Carnegie Mellon University, USA; May–June 2023)
Prof. Heather J. Kulik (Romberg Visiting Scholar, Massachusetts Institute of Technology, USA; April–May 2023)
Gregor Lauter (PhD student; since October 2023)
Lukas Lehr (student intern, Heidelberg University; November–December 2023)
Stiv Llenga (PhD student)
Anastasiia Nihei (visiting scientist; MSc student, V. N. Karazin Kharkiv National University, Ukraine; until May 2023)
Dominique Ostermayer (PhD student; since March 2023)
Owen Paine (student assistant; until September 2023)
Anna Piras (PhD student)
Oliver Riccio (student intern, Heidelberg University; September–October 2023)
Florian Schmeller (student intern, Heidelberg University; March–April 2023)
Dr. Abderrezak Torche (until December 2023)

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, physics, and materials science in combination with chemical machine learning to explore and exploit diverse functional organic and hybrid materials. We are particularly interested in several classes of materials – including graphene-based materials, covalent organic frameworks, and hyperbranched polymers – in the context of their applications in capturing, storing, transporting, and catalytically transforming therapeutic molecules and environmental pollutants. Functional organic materials are central to our research efforts (1) to establish the role of topology in materials chemistry, (2) to predict emergent properties in molecule–material complexes from their individual components, and (3) to build reliable yet interpretable machine learning models of the chemical properties of materials.

The group is part of the SFB1249 "N-Heteropolycycles as Functional Materials" and the SIMPLAIX strategic initiative on bridging scales from molecules to molecular materials via multiscale simulation and machine learning. We also collaborate fruitfully with several theoretical and experimental groups, including those of Frauke Gräter (HITS), Lutz Greb (Heidelberg University), Ulrich Paetzold (Karlsruhe Institute of Technology), Bernd Schmidt (Heinrich Heine University Düsseldorf), and Amir Karton (University of Western Australia).



## What happened in the group in 2023?

In its final year at HITS, the CCC group continued developing computational workflows in order to simulate complex materials. The group 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 sophisticated multiscale high-throughput workflow was implemented and applied in order to generate a dataset of physisorption complexes between graphene-based materials (GBMs) and small molecular targets. Within the SFB1249 "N-Heteropolycycles as Functional Materials," several new datasets of N-heteropolycycles were constructed and used to elucidate structure–property relationships.

Machine learning (ML) continued to play a prominent role in the group's research. For example, a new dimensionality reduction and molecular representation technique called the Matrix of Reference Similarity was developed. Within the SIMPLAIX strategic initiative, the previously constructed message passing graph neural network was expanded in order to enable the prediction of redox potentials in any solvent. The group hosted two visiting US scientists: Olexandr Isayev from Carnegie Mellon University and Heather Kulik from the Massachusetts Institute of Technology. Olexandr is a leading expert in chemical machine learning, materials informatics, and de novo design and is also a developer of one of the most popular machine learning potentials, ANI-1. Heather's research lies at the interface of chemical engineering and computational materials science and has a strong focus on both the data-driven design of metal–organic frameworks and the highly accurate modeling of transition metal chemistry. The many stimulating scientific discussions between Olexandr, Heather, and the group members led to exciting new ideas and the launch of promising collaborative projects.

In 2023, Anya Gryn'ova taught the "Applied Computational Chemistry" course at the Faculty of Chemistry and Earth Sciences, Heidelberg University, for the first time in person. In addition to the usual lectures, this year, the course also included practical sessions on computational chemistry and chemical machine learning, which were led by Christopher Ehlert and Rostislav Fedorov. The course attracted an unprecedented number of bachelor's and master's students, many of whom later joined the group as interns and student assistants.

Finally, in 2023, Anya received several offers of tenured professorships and accepted the position of Associate Professor of Computational Chemistry at the School of Chemistry, University of Birmingham, UK. Thus, in the spring of 2024, after five fruitful and exciting years, the group will be leaving HITS.

## Introduction

Graphene – a flat two-dimensional monolayer of sp2-hybridized carbon atoms – is characterized by its extraordinary physical properties, including its large surface–volume ratio, its mechanical strength and flexibility, its high carrier mobility, and its thermal conductivity. The chemistry of graphene is equally rich and comprises physi- and chemisorption on its surface as well as – inter alia – reversible oxidation and reduction, which makes the material an attractive, environmentally beneficial, economically viable metal-free agent in a plethora of chemical transformations. For example, carbocatalysis – that is, catalysis with carbonaceous materials (i.e., pristine graphene, graphene oxide, reduced graphene oxide, doped and/or functionalized graphene oxide) – has been adapted in a broad scope of organic reactions, including substitution, addition, hydration and dehydration, hydrogenation and dehydrogenation, reduction and oxidation, hydrolysis, and polymerization. In fuel cells, graphene-based materials represent ideal electrocatalyst supports for oxygen reduction, oxygen evolution, and hydrogen evolution reactions. Due to their ability to reversibly capture, transport, and release small molecules, graphenes are well-suited for molecular delivery as well as for energy storage and conversion. Their rich redox chemistry – combined with their inherent electric conductivity – enables the highly selective and sensitive electrochemical sensing of various environmental pollutants, such as heavy metal ions, volatile organic compounds, and nitroaromatic and phenolic contaminants. Upon introducing vacancies and generating micro- and nanopores within the graphene lattice, GBMs can act as membranes, filters, or sieves for separating gases, liquids, and mixtures, which has applications in water remediation and pollution mitigation.

Future advances in these practical applications of GBMs require an in-depth understanding of the underlying chemistry and of the relevant structure–property relationships. Research efforts in the CCC group that are dedicated to gaining this knowledge are discussed below.

## High-throughput Multiscale Modeling of Molecular Adsorption on Graphenebased Materials

**Abderrezak Torche**

In the age of artificial intelligence, the efficient and application-driven design of new materials – as well as the selection from among existing materials – is greatly facilitated by chemical machine learning. Consequently, there is a pressing need for clean and reliable data that can be used to train and test the models. In materials science, this need is reflected in the unprecedented increase over the past 5 years in the number of both experimental and theoretical databases, such as CSD, Nomad, and Materials Project. However, none of these databases contains sufficient consistent and application-oriented data for GBMs.

Building a computational database for the adsorption of molecules on GBMs entails many challenges. The theoretical model of the adsorption phenomena requires identifying an appropriate structural representation (cluster or periodic) and choosing an in silico methodology that is compatible with the limitations that the high-throughput screening imposes on computing memory and time. In order to address these bottlenecks, we constructed a multiscale workflow that is designed to identify the adsorption geometry and

evaluate the binding energy of a given molecule on a given GBM. This automated workflow is implemented in the AiiDA (Automated Interactive Infrastructure and DAtabase for computational science) manager in order to ensure transparency, reproducibility, and sharing. It combines simulated annealing molecular dynamics (SAMD) and dispersion-corrected density functional theory (DFT) and involves several data refinement stages.

Using this workflow, we created a database that contains over 5,500 adsorption complexes between 40 representative small molecules (guests) and various defected, doped, and functionalized GBMs (hosts) (Figure 8). This database – which we named GRADS (GRAphene ADSorption) – includes computed binding energies, equilibrium adsorption geometries, charge transfer rates, and other characteristics for each host–guest system.

In the future, we plan to use GRADS (1) to elucidate structure–property relationships that are pertinent to molecular adsorption on graphene and (2) to assess the performance of machine learning potentials and density functional tight binding in predicting the interaction energies of two-dimensional organic materials.

## Oxygen Reduction Reaction Catalysts on a Graphite Electrode

**Carina Herrle and Christopher Ehlert**

The oxygen reduction reaction (ORR) is an important electrochemical process that lies at the heart of sustainable energy conversion technologies, such as fuel cells and metal–air batteries. The reaction takes place at the negatively charged cathode and exhibits rather sluggish kinetics, which necessitates the use of catalysts. In the past, platinum-based alloys were among the best-performing catalysts; however, the scarcity of platinum, its high price, and its poor long-term durability inhibit its large-scale application. Recently, metal-free heteroatom-doped carbon catalysts have emerged as more efficient, stable, sustainable, inexpensive, and earth-abundant alternatives to conventional Pt-based systems. Among these catalysts, graphene-based materials that are doped with nitrogen, boron, sulfur, and phosphorus show remarkable performance. However, these systems suffer from ambiguity in doping patterns that contain multiple heteroatoms, thereby obfuscating the identification of active sites. Alternative, bottom-up approaches that begin with well-defined heteroatom-containing polyaromatic hydrocarbons (PAHs, sometimes termed graphene nanoflakes) address this drawback while potentially preserving the catalytic effect.

Recently, ORR catalyst activities of six boron- and nitrogen-doped graphene nanoflakes were experimentally investigated and showed diverse performance despite relatively similar structures [Kahan et al. (2019). Well-defined boron/nitrogen-doped polycyclic aromatic hydrocarbons are active electrocatalysts for the oxygen reduction reaction. Chem. Mater. 31: 1891–1898]. In order to solve this puzzle, we investigated the first step of the ORR pathway – that is, chemisorption – using density functional theory [Ehlert et al., 2023a] and discovered that the catalyst must be negatively charged in order to effectively adsorb oxygen. Correspondingly, electron affinity was put forward as a simple activity descriptor of charged (activated) catalysts on an electrode.

We next expanded this investigation by employing a more realistic theoretical model of the experimental setup – that is, a single layer of graphene that represents the surface of the graphite electrode and on which the molecular catalyst is adsorbed. Correspondingly, a broad range of computational techniques – including periodic DFT and semiempirical GFN2-xTB methods – were employed to evaluate electron affinities as well as interaction energies, charge transfer characteristics, and the energies of the lowest unoccupied molecular orbital (LUMO) for the previously investigated molecular catalysts, which were adsorbed on graphene. We found that the GFN2-xTB LUMO energies are well-correlated with the DFT electron affinities and can serve as reliable indicators of catalytic activity in the ORR (Figure 9). These findings pave the way toward the high-throughput screening of thousands of possible molecular catalysts using highly efficient and computationally inexpensive semiempirical tight binding.

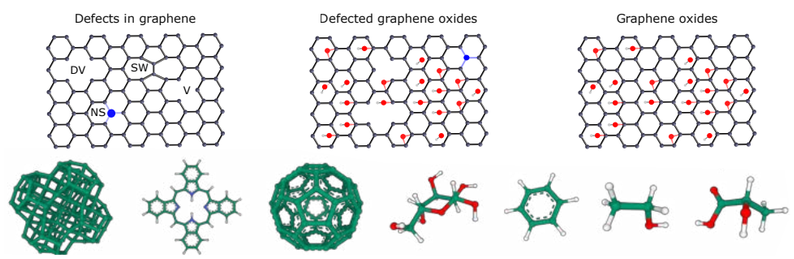



*Figure 8: Examples of GBM hosts and molecular guests in the GRADS database. Hosts (top row) include pristine and oxidized graphene both with and without defects (V = vacancy, DV = double vacancy, SW = Stone-Wales defect, NS = nitrogen substitution). Guests (bottom row) range from small molecules (e.g., CO2, methane, benzene) to large and topologically complex structures (e.g., fullerene, carbon quantum dots).*

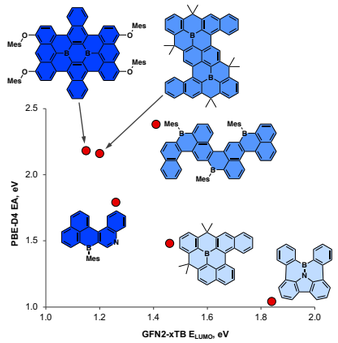



*Figure 9: Computed LUMO energies vs. electron affinities for six investigated molecular catalysts. Catalyst structures are colored according to experimentally measured ORR catalytic activities: dark blue = active, blue = mid-active, light blue = inactive. "Mes" stands for mesityl.*



## Size Dependence of the Redox Properties of GBMs

**Anna Piras**

The electronic structure of graphene-based materials is crucial to the development of more efficient and sustainable batteries and catalysts. In both electrochemical sensing and catalysis, the first electron transfer is considered the rate-determining step. Therefore, when it comes to enhancing the performance of electro GBMs in chemical applications, it is crucial to understand how the electronic (i.e., ionization energy (IE) and electron affinity (EA)) and redox (e.g., reduction potential) properties of graphene nanoflakes depend on their size, functionalization, and doping.

In order to establish relevant structure–property relationships, we studied the redox properties of diverse graphene nanoflakes across a range of sizes and chemistries (Figure 10A,B). Such large polycyclic π-conjugated systems tend toward diradicaloid ground states due to strong static electron correlation. In order to obtain reliable results, we examined the quality of the electronic structure of the investigated molecules using various DFT methods ranging from generalized gradient approximation (PBE) to range-separated functionals (ωB97X-D) in conjunction with several spin contamination diagnostics, including the fractional occupation number weighted density. Despite the multireference nature of the underlying wavefunctions, all methods resulted in very similar estimates for IEs and EAs.

We discovered that while doping and functionalization predictably improve both the electron-accepting (higher EAs) and electron-donating (lower IEs) properties of GBMs compared with pristine graphene, the precise chemical composition of the material does not have a strong effect on these properties, especially at larger nanoflake sizes (Figure 10C,D). Moreover, this dependence appears to be polynomial rather than linear with respect to the model size. This is particularly interesting considering that according to our previous studies, adsorption energies instead follow a linear relationship with the nanoflake size and – most importantly – show a much greater variation depending on the chemistry of the material.

These findings suggest that in order to boost the efficiency of graphene-based materials for electrochemical applications, a greater weight should be placed on the adsorbent properties, whereas reasonable electron shuttling can be expected with almost any doped graphene. This finding is reminiscent of the results reported in the fittingly titled paper "Will any crap we put into graphene increase its electrocatalytic effect?" [Wang et al. (2020), in ACS Nano, 14, 21–25, doi.org/10.1021/acsnano.9b00184], in which doped GBM electrocatalysts performed much better than its pristine counterpart in oxygen reduction and hydrogen evolution reactions, independent of the dopant's chemical nature.

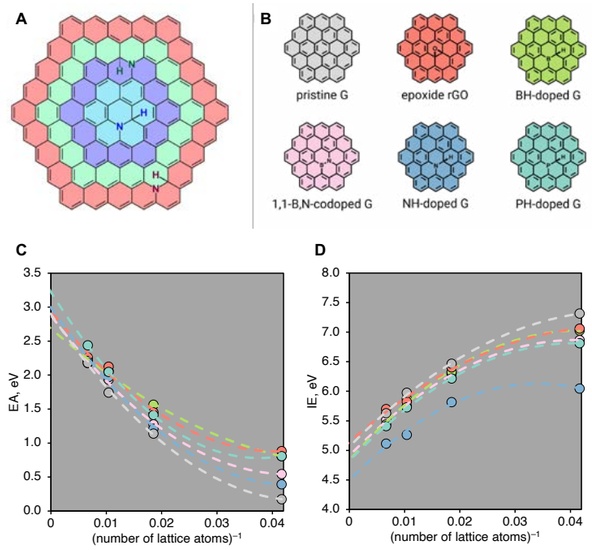



*Figure 10: A. Investigated model sizes illustrated for NH-doped graphene. B. Models of investigated GBMs shown for size 1 and color-coded according to their chemical nature. Vertical electron affinities (C) and ionization energies (D) computed at the ωB97X-D/def2-TZVP level of theory for various GBMs as a function of the size of the nanoflake and expressed as the reciprocal of the number of non-hydrogen atoms that the nanoflake contains. Dashed lines are 2nd-order polynomial fits.*

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.

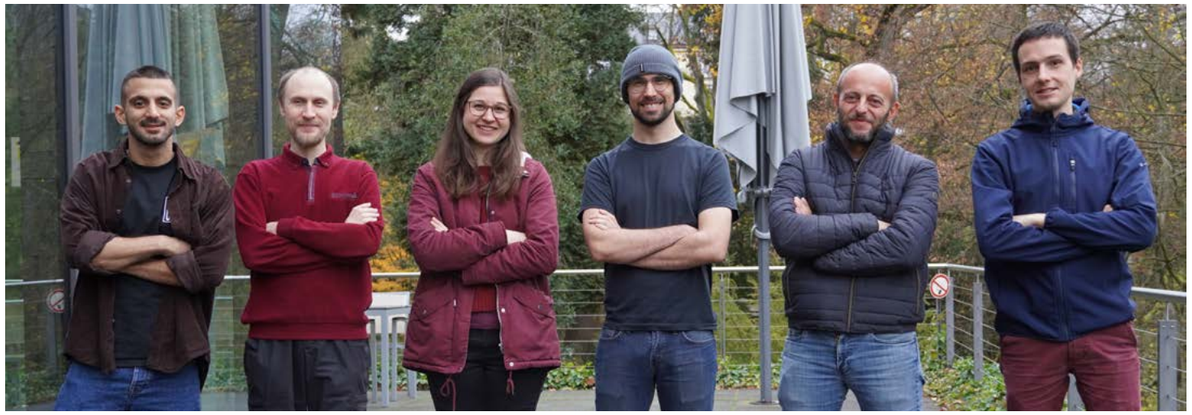
Die Forschungsgruppe **Computational Carbon Chemistry** (CCC) nutzt theoretische und computergestützte Chemie, Physik und Materialwissenschaft in Kombination mit maschinellem Lernen in der Chemie, um verschiedene funktionale organische und Hybrid-Materialien zu untersuchen und auszuwerten.

Wir sind insbesondere an bestimmten Materialklassen interessiert – einschließlich Graphenbasierten Materialien, kovalenten organischen Strukturen und hyperverzweigten Polymeren – im Zusammenhang mit ihren Anwendungen beim Einfangen, Speichern, Transportieren und katalytischen Umwandeln von therapeutisch genutzten Molekülen und Umweltschadstoffen. Funktionelle organische Materialien stehen im Mittelpunkt unserer Forschung, um (1) die Rolle der Topologie in der Materialchemie zu ergründen, (2) neu entstehende Eigenschaften in Molekül-Material-Komplexen aus ihren einzelnen Komponenten vorherzusagen und (3) zuverlässige und dennoch interpretierbare maschinelle Lernmodelle für die chemischen Eigenschaften von Materialien zu entwickeln.

Die Gruppe ist Teil des SFB1249 „N-Heteropolyzyklen als funktionale Materialien" und der SIMPLAIX-Initiative zur Überbrückung der Skalen von Molekülen zu molekularen Materialien mithilfe von Multiskalensimulation und maschinellem Lernen (siehe Kapitel 7). Wir kooperieren außerdem erfolgreich mit mehreren theoretisch und experimentell arbeitenden Gruppen, neben der Gruppe von Frauke Gräter (siehe Kapitel 2.7) auch mit Lutz Greb (Universität Heidelberg), Ulrich Paetzold (Karlsruher Institut für Technologie KIT), Bernd Schmidt (Heinrich-Heine-Universität Düsseldorf) und Amir Karton (University of Western Australia, Perth).

# 2 Research

# 2.3 Computational Molecular Evolution (CME)



### Group leader

Prof. Dr. Alexandros Stamatakis

### Team

Benjamin Bettisworth (PhD student)
Erik Borker (student; since December 2023)
Julia Haag (PhD student; HITS Scholarship)
Luise Häuser (PhD student; HITS Scholarship; since October 2023)
Johannes Hengstler (student)
Dimitri Höhler (PhD student)
Lukas Hübner (visiting scientist from KIT)
Lukas Knirsch (bachelor's student; since July 2023)
Dr. Alexey Kozlov (staff scientist)
Luc Mercatoris (master's student; since July 2023)
Dr. Benoit Morel (postdoc)
Konstantin Ntounas (bachelor's student; January–March 2023)
Angeliki Papadopoulou (visiting scientist, Hellenic Republic University of Crete; February–April 2023)
Ioannis Reppas (master's student; until February 2023)
Dominik Siebelt (student)
Christoph Stelz (master's student)
Jan Strehmel (student assistant; March–May 2023)
Anastasis Togkousidis (PhD student; HITS Scholarship)
Noah Wahl (master's student; July–December 2023)
Julius Wiegert (master's student; since July 2023)

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 2023?

On 1 January 2023, Alexis began his five-year EU-funded ERA chair project at the Institute of Computer Science within the Foundation for Research and Technology Hellas (ICS-FORTH). The goal is to set up a second research group – the Biodiversity Computing Group (BCG) at ICS-FORTH – that is closely connected to the CME group at HITS. The two research groups can also be regarded as one larger group that has offices in two countries and that uses shared calendars, Slack instances, etc. Within this context, many exchanges have already taken place, with CME PhD students Julia Haag and Lukas Hübner as well as KIT master's students Luise Häuser and Noah Wahl having visited Crete. Moreover, Angeliki Papadopoulou – a PhD student of former CME postdoc Pavlos Pavlidis – visited CME in Heidelberg. Finally, newly hired BCG postdocs Ben Bettisworth, Panos Ioannidis, and Georgios Koutsovoulos visited HITS with Alexis in order to tighten the links between the two groups. Three PhD students from three different European countries will join BCG in early 2024, thereby leaving BCG and CME with approximately the same number of team members. This new venture has shaped the activities of the past year.

With respect to teaching, in the winter of 2022/2023, 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 2023, we again taught our main seminar, Hot Topics in Bioinformatics.

In winter 2023/24, we launched a new teaching endeavor with the joint University of Crete (UoC)–KIT course on Introduction to Bioinformatics for Computer Scientists, which is taught simultaneously at the computer science departments of KIT and UoC. Moreover, four and five live lectures took place at KIT and UoC, respectively, which were streamed via Zoom to the other university.

Luise Häuser – our master's student from the Department of Computer Science at KIT – joined the lab as a PhD student in fall 2023 and will initially focus on computational aspects of inferring phylogenetic trees of natural languages.

Furthermore, Ben Bettisworth successfully defended his PhD at KIT in summer 2023 and joined BCG on Crete as a postdoc in fall 2023, which will help to even better connect the two groups. Our KIT master's student Noah – who completed his master's thesis in late 2023 – will also join BCG as a PhD student. For reasons we do not yet understand, there has been a substantial increase in the number of bachelor's and master's theses we supervise at KIT. This year, 2 bachelor's and 2 master's theses were completed, and a total of 6 student projects are currently ongoing. However, when CME was established in 2010, there used to be only 2 or 3 theses per year. Finally, our staff scientist Alexey Kozlov assumed his new part-time role as Sustainability Advisor of HITS with a key focus on technical and computer-science-related energy efficiency issues. After a successful first year in his new role, Alexey will continue in 2024 and will thereby contribute to the general institutional development of HITS.

Our recurring highlight – namely the summer school on Computational

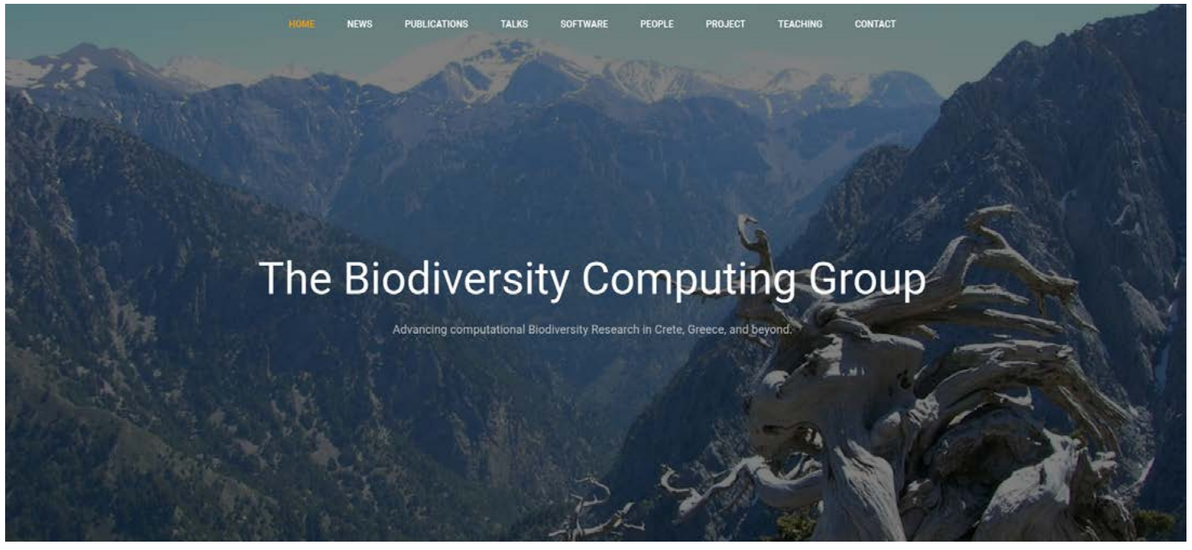



*Figure 11: The new webpage of the Biodiversity Computing Group, the sister lab of CME.*

Molecular Evolution – finally took place again in 2023 for the first time since 2018 at the Hellenic Center for Marine Research with a generous co-sponsorship by HITS (see Chapter 5.1.3). This iteration of the summer school on Crete had been postponed several times in 2020 and 2021 and was then finally canceled due to the pandemic. Lukas supported the summer school as a teaching assistant, and Alexis served as a lecturer and co-organizer. Both scientists thoroughly enjoyed the school and the subsequent satellite workshop on biodiversity informatics. The next iteration of our summer school is planned for May 2024, this time in the UK again.

Also in 2023, Alexis was listed on the Clarivate Analytics list of highly cited researchers for the eighth year in a row (see Chapters 4 and 10.5).

The year additionally consisted of a plethora of public outreach activities. Alexis gave three outreach presentations at three public primary schools on Crete using the now-established material on "The Aegean Archipelago: a living laboratory of evolutionary biology." Alexis also contributed to the setup of a permanent touchscreen in the bird section of the Natural History Museum of Crete that includes a publicly displayed web game as part of the aforementioned material. Furthermore, Alexis also developed publicly available material for teaching Air Traffic Control to elementary school children in collaboration with a Greek air traffic controller and a primary school director. Finally, Alexis also gave a public outreach talk within the framework of the Darwinian Monday series at the Natural History Museum of Crete.

In sum, 2023 was dominated by the establishment of the CME sister group – BCG – on Crete and by the tightening of the connections between the two groups, while all activities at HITS and KIT in Germany continued as in the years before.

## Introduction

The term "computational molecular evolution" refers to computer-based methods of reconstructing evolutionary trees from DNA or – for example – from protein data or morphological data. The term also refers to the design of programs that estimate statistical properties of populations – that is, programs that disentangle evolutionary events within a single species.
The very first evolutionary trees were inferred manually by comparing the morphological characteristics (i.e., 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.

Ever 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), scientists have been able to generate over 10,000,000 short DNA fragments (each containing between 30 and 400 DNA characters) as well as continuous improvements in sequencing technology that also allow substantially longer reads to be generated. As a result, the community as a whole now faces novel challenges. One key problem that needs to be addressed is the fact that the volume of molecular data available in public databases is growing at a significantly faster rate than the computers that are capable of analyzing the data can keep up with.

In addition, the costs of sequencing a genome are decreasing at a faster rate than are the costs of computation, although the curve seems to have been flattening out in the last 3–4 years (see https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost).

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

This scalability challenge additionally entails reproducibility challenges, especially when using parallel programs. Together with our KIT master's student Christoph Stelz, we are currently investigating the trade-offs between the increased reproducibility of parallel software on the one hand and execution times on the other hand. We are also re-investigating whether GPUs can be deployed to accelerate phylogenetic likelihood calculations. Related work that we conducted around one decade ago showed that porting this specific computational kernel to GPUs does not yield substantial efficiency gains.

Another emerging line of research is the deployment of machine learning techniques and AI methods to address problems in phylogenetics and bioinformatics, partially also because most current KIT master's students are looking for thesis topics in the area of data science. For instance, we currently supervise projects that attempt to predict the dataset-specific difficulty of phylogenetic placement – that is, we place anonymous DNA sequences onto a phylogenetic reference tree with known/named sequences/species in order to identify them, to predict the difficulty of aligning multiple sequences with one another, or to try to predict statistical support values that reflect the degree of certainty we have in specific branches of an evolutionary/phylogenetic tree.

Overall, phylogenetic trees (i.e., the evolutionary histories of species) and the application of evolutionary concepts in general are important in numerous domains of biological and medical research. Programs for tree reconstruction that have been developed in our lab can be deployed to aid in inferring evolutionary relationships between viruses, bacteria, green plants, fungi, mammals, etc. – in other words, these phylogenetic tree inference methods 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, or to shed light on population movements during the Greek Dark Ages (i.e., ca. 1100–750 BCE). Phylogenies can also be used to disentangle the evolution of natural languages in linguistics. With our new PhD student Luise Häuser, we have already begun exploring and understanding this application area of phylogenetic inference. Our collaborator Elena Anagnostopoulou – a theoretical linguist at the University of Crete – obtained an ERC grant this year to study the evolution of syntax structures. Both our group and the group of former CME postdoc Pavlos Pavlidis (who is now Associate Professor of Biology at the University of Crete) will attempt to help Elena address the inherent computational challenges of this project.

Finally, phylogenies play an important role in analyzing the dynamics and evolution of viruses, as we have all witnessed during the ongoing SARS-CoV-2 pandemic.

Accelerating Phylogenetic Tree Inference via Machine Learning

In 2022, 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. In 2023, we introduced difficulty values that range between 0 (easy) and 1 (hopeless). These values can be used not only to adjust the prior expectations of the end user with respect to the stability of the phylogeny to be inferred, but also to inform the behavior of the tree search algorithm. Since the problem of finding the optimal maximum likelihood tree is NP-hard, tools such as our RAxML-NG software typically deploy heuristic search strategies to find a "good tree" in the vast tree search space. Thus far, these search strategies have been dataset-agnostic, but difficulty prediction – which is substantially faster to compute than an ML tree search – now allows the search strategy to be adapted to the degree of difficulty (or signal strength) in the data to be analyzed. In other words, for easy datasets with a strong signal and for difficult datasets (which we also label "hopeless") with a weak signal, we do not need – or it does not make sense – to invest much computational effort into the tree search. To that end, we developed adaptive RAxML-NG, which uses such adaptive tree search strategies as a function of the predicted difficulty. Averaged over all difficulty value levels, adaptive RAxML-NG yields results that are statistically equally as good as those of the naïve dataset-agnostic standard search algorithm implemented in RAxML-NG, but it runs 3.4 times faster. Our experimental results for the adaptive heuristic on 9,515 empirical datasets and 5,000 simulated datasets with varying difficulty levels also revealed substantial speedups on easy and difficult datasets (53% of total MSAs), for which we observed average speedups exceeding an order of magnitude (Figure 12).

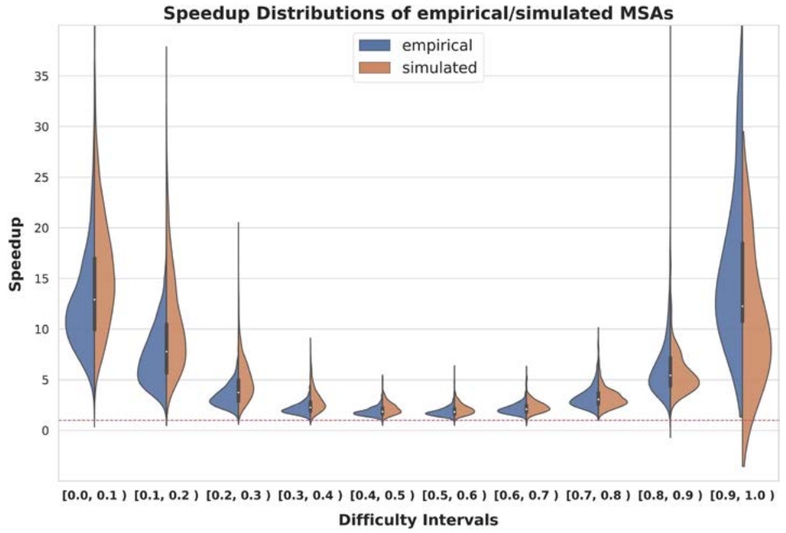



*Figure 12: Speedups of the adaptive RAxML-NG over the standard RAxML-NG search algorithm for difficulty intervals between 0 and 1. The blue curves represent speedup distributions on empirical datasets, and the orange curves represent speedup distributions on simulated datasets.*

## More Machine Learning Stuff

Speaking of simulated versus empirical datasets, the computational phylogenetics community has long agreed that search algorithms behave differently on simulated data than on empirical data. In general, simulated datasets tend to be smoother, and search algorithms require fewer search iterations in order to converge on simulated datasets than on empirical datasets. Simulated datasets also tend to be easier to analyze and might not reveal differences between search algorithms that appear with empirical data. In order to investigate and formalize this observation, we created a type of Turing test for datasets. We posed the question as to whether simple and more complicated machine learning tools are able to tell apart empirical datasets from simulated datasets. In other words, simulated data were said to be realistic if they could not be reliably distinguished from empirical data in the context of a simple machine learning binary classification task. In our study, we simulated DNA and protein MSAs under increasingly complex statistical models of evolution using a state-of-the-art sequence simulator. We assessed the realism of the MSAs by quantifying how accurately supervised learning methods were able to predict whether a given MSA was simulated or empirical. We found that we could easily distinguish between empirical and simulated MSAs with high accuracy across all tested models of sequence evolution using two distinct and independently developed classification approaches. We concluded that current state-of-the-art models fail to accurately replicate several aspects of empirical datasets, and research effort thus needs to be directed toward developing more realistic sequence simulation tools (Figure 13).

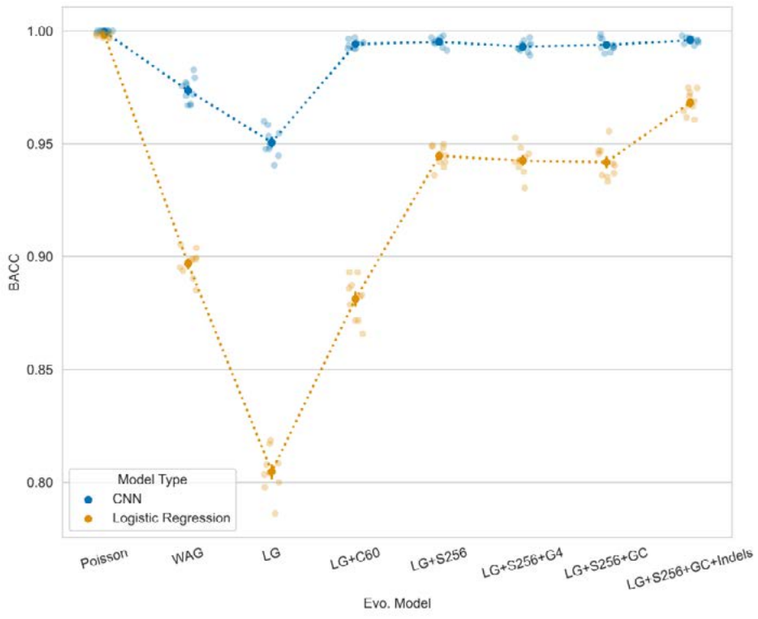



*Figure 13: Balanced ACCuracy (BACC) for telling apart simulated datasets from empirical datasets in a logistic regression (simple machine learning method) and a convolutional neural network (CNN, complex machine learning method) over increasingly complex statistical models of evolution (complexity increases from left to right).*

## Re-Engineering a Biogeography Tool

In the context of handling the aforementioned scalability challenge in bioinformatics, we turned our attention to the field of biogeography, and specifically to the inference of ancestral species ranges – that is, where a species once lived geographically. The statistical computational methods that are deployed for this task are characterized by an exponentially increasing number of model states as a function of the number of regions that are considered. In other words, the finer the geographical granularity, the more compute-intensive the models become. This computational complexity stems from a large matrix exponential, which is a standard numerical operation that typically accounts for up to 80% of overall run-time in these biogeography models. Therefore, the types biogeographical analyses that can be conducted under this model are extremely limited by the number of regions under consideration. To that end, we deployed algorithmic and numerical engineering techniques in order to develop a completely re-designed and substantially more efficient version of the popular Lagrange tool. Our new code – which we named Lagrange-NG (next generation) – is up to 49 times faster with multithreading enabled and is also 26 times faster when using only a single CPU core. This increased computational efficiency allows Lagrange-NG to analyze datasets with a larger number of geographical regions in a reasonable amount of time – for instance, up to 12 regions in approximately only 18 minutes. We achieved these speedups by using a comparatively new method of computing the matrix exponential based on so-called Krylov subspaces. In order to verify the correctness of Lagrange-NG, we also introduced a novel metric for comparing inferred range distributions for evolutionary trees that is able to quantify the difference between any two distinct ancestral geographical range inferences. Finally, Lagrange-NG exhibits substantially higher adherence to coding quality standards, which is also a key focus of our lab. Lagrange-NG improves a software quality indicator as implemented in our lab's SoftWipe tool from average (5.5; Lagrange) to high (7.8; Lagrange-NG) (Figure 14).

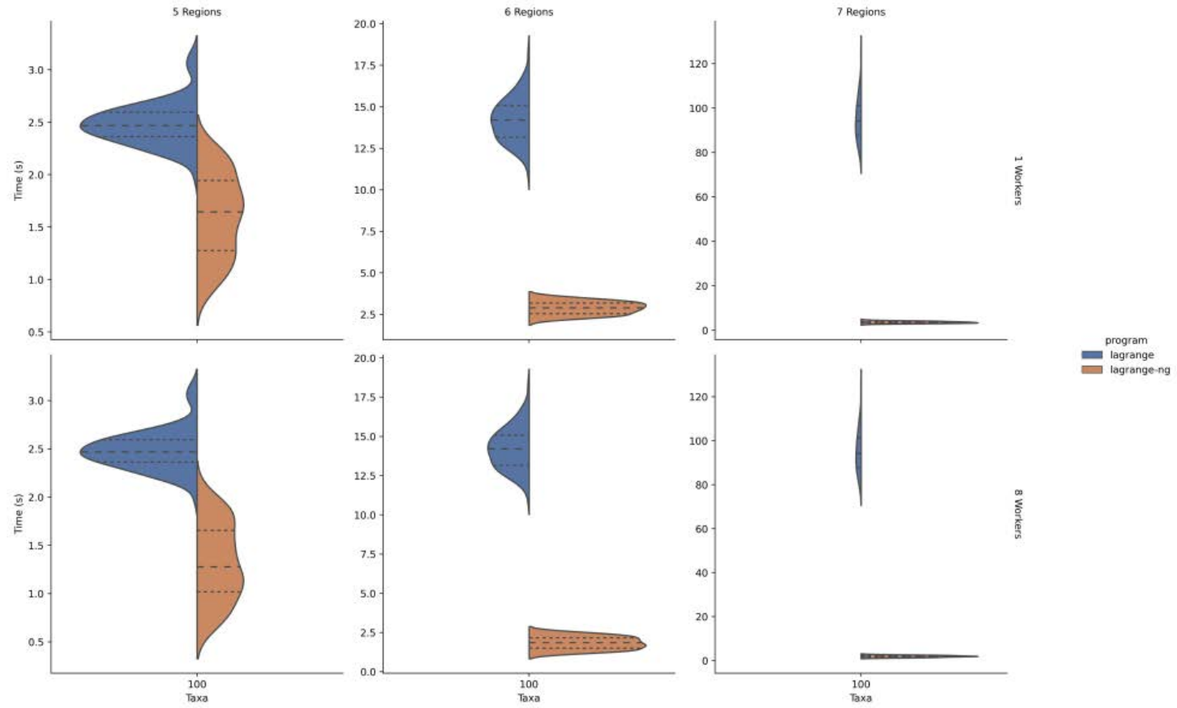



*Figure 14: Comparison of run times (in seconds on the x-axis) between Lagrange (left; blue distribution) and Lagrange-NG (right; orange distribution) with sequential Lagrange-NG (top row) and parallel Lagrange-NG (bottom row) using 8 cores. The reference trees contain 100 taxa (species) for all experiments, and the number of regions considered increases from 5 on the left to 6 in the middle to 7 on the right. Note the increasing run-time scale on the y-axis from left to right, which nicely displays the exponential run-time increase as a function of the number of regions.*

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 2023)
Dr. Johannes Bracher (visiting scientist, Karlsruhe Institute of Technology, Germany)
Jun.-Prof. Dr. Timo Dimitriadis (visiting scientist, Heidelberg University, Germany)
Sebastian Gottheil (student; until June 2023)
Davide Hailer (intern; from May until July 2023)
Dr. Alexander I. Jordan (staff scientist)
Kristof Kraus (student; until August 2023)
Dr. Sebastian Lerch (visiting scientist, Karlsruhe Institute of Technology, Germany)
Dr. Marc-Oliver Pohle (since April 2023)
Dr. Ghulam Abdul Qadir (until November 2023)
Dr. Johannes Resin (visiting scientist, Heidelberg University, Germany)
Prof. Dr. Melanie Schienle (senior researcher)
Evgeni Ulanov (student; until May 2023)
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 transdisciplinary 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.



## General News

In July 2023, Sándor Baran (University of Debrecen, Hungary) visited us, and on 28 July, we held a mini-symposium on an array of topics ranging from statistical post-processing and the machine learning of spatial covariance functions to recent advances in the theory of forecast evaluation.

We enjoyed another CST group excursion as a summer hike across the Weinbiet mountain starting and ending in the town of Neustadt. This time, we were joined by colleagues from the MLI group, and everyone had a great time spending the hot day together under the protective canopy of the Palatinate Forest.

In this year's scientific report, we highlight three projects that together provide an overview of the group's diverse activities. First, we describe how our work facilitates decision-making regarding the response to the COVID-19 outbreak. Accurate assessments of the current state of events via nowcasting are vital when it comes to putting appropriate control measures in place in a timely fashion. Second, we illustrate that much of the methodology in weather forecasting – which continues to adopt modern machine learning techniques – can also be greatly beneficial in solar forecasting. Third, we discuss our continued development of evaluation methods that are rooted in sound mathematical theory, and we describe the concept of top-list predictions, which should prove useful in classification problems.

## Collaborative Nowcasting of COVID-19 Hospitalization Incidences in Germany

*Note: Figs. 15, 16, and 17 as well as the original text by [Wolffram et al., 2023] are licensed under CC BY 4.0. The original text has been adapted into the shorter form presented below.*

During infectious disease outbreaks, such as the COVID-19 pandemic, real-time surveillance data contribute to situational awareness and risk management. These data can play an important role in the allocation of resources and in making decisions on control measures that limit exposure. However, real-time data are often preliminary, which leads to difficulties in the timely interpretation of epidemiological indicators.

In Germany, decisions have been based on COVID-19 hospitalization incidences. These indices are particularly affected by delays that can take several weeks or months to fully resolve for two main reasons: (1) hospital admission necessarily precedes the report of the hospitalization, and (2) hospitalization counts are aggregated by the date when the positive test of an ultimately hospitalized person is digitally registered rather than by the date of hospital admission. Figure 15

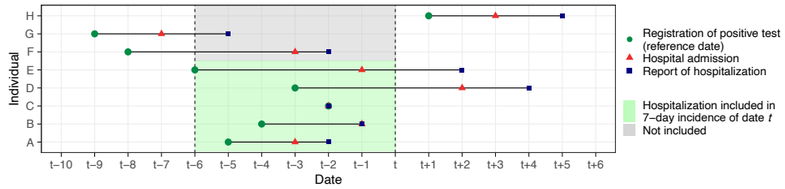



*Figure 15: Illustration of 7-day hospitalization incidences via individual-level timelines. The reference date that is used to count a hospitalization is the date when the positive test of an ultimately hospitalized person is reported (green dots). Actual hospitalization often takes place much later (red triangles) and is reported afterward (blue squares). Only individuals A–E are included in the 7-day hospitalization incidence of date t because their reference date falls within a 7-day window from t − 6 to t. Some hospitalizations (i.e., individuals D and E) appear only after a delay in updated versions of the data due to their late reporting dates.*

illustrates these delays and shows how individuals are counted in the official definition of the German COVID-19 hospitalization incidence via a 7-day accumulation of registered cases per 100,000 inhabitants. Since new hospitalizations retrospectively increase the count of registered cases, the most recent hospitalization incidence tends to be much lower than the actual value.

This bias can lead to incorrect conclusions about current trends. Statistical nowcasting methods aim to remedy this problem by predicting how strongly preliminary data points will be corrected upward when considering the associated uncertainty. Nowcasts thus help to uncover current trends that are not yet visible in reported numbers. Figure 16 illustrates the nowcasting task in the context of the 7-day hospitalization incidence. It shows real-time nowcasts from 1 December 2021, 1 February 2022, and 1 April 2022. Comparison with a more stable data version from 8 August 2022 shows that in these instances, the nowcasts were able to correctly reveal the actual trends, which differed sharply from the apparent declines found in the data at the time of nowcasting.

While numerous methods exist for nowcasting, little is known about the behavior of these methods in real-time settings or about their relative performance. In [Wolffram et al., 2023], we compared nowcasting methods used for COVID-19 hospitalization incidences in Germany from November 2021 to April 2022. Additionally, we combined the predictions of these methods to create so-called ensemble nowcasts. Figure 17 shows same-day nowcasts at the national level. The median predictions are displayed along with the central 50% and 95% prediction intervals. The light gray line shows the preliminary data as they were available on that day, and the black line shows the respective final value as they were available on 8 August 2022.

Nowcasts from eight independently run models were collected for the duration of our study. Six models were contributed by groups of academics (i.e., Epifore-

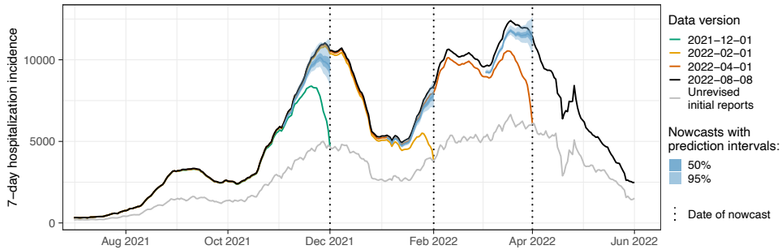



*Figure 16: Illustration of the nowcasting task. Data that are available in real time (colored lines) are incomplete, and the values are considerably lower than the final corrected values (black line), especially for recent dates. Nowcasts with prediction intervals (blue-shaded areas) aim to predict in real time what the final data points will be. The light gray line shows the initially reported value as it was available on the respective date.*

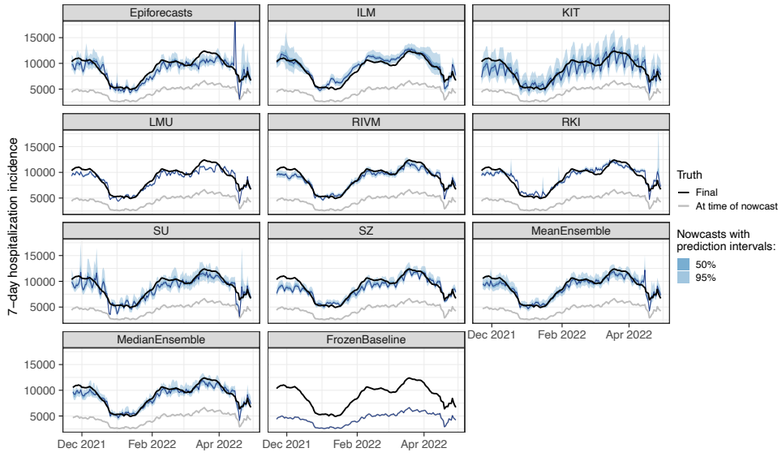



*Figure 17: Nowcasts with a horizon of 0 days back. Same-day nowcasts of the 7-day hospitalization incidence as it was issued on each day of the study period. Nowcasts are shown for the German national level.*

casts, ILM, KIT, LMU, RIVM, SU), one was contributed by the Robert Koch Institute (RKI), and one was contributed by the data science team at the newspaper Süddeutsche Zeitung (SZ). Most approaches took preliminary hospitalization numbers as their only input and applied various techniques to model delay distributions and the underlying time series of hospitalizations. The nowcasts from all models are generally close to the final predicted values; however, considerable variability in interval widths is apparent and ranges from rather wide (KIT) to very narrow intervals (LMU, RKI). Some models – especially those from the KIT and SZ – display pronounced weekday patterns in their same-day nowcasts that also carry through to the ensemble nowcasts to a lesser degree.

For the MeanEnsemble, each predictive quantile was obtained as the arithmetic mean of the respective quantiles of the member nowcasts. The ensemble mean was obtained as the mean of the submitted predictive means. For the MedianEnsemble, the same procedure was applied using the median rather than the arithmetic mean for aggregation. The ensemble nowcasts showed strong relative performance. Additionally, our collaborative approach fostered frequent exchange and interaction among modelers via bi-weekly coordination calls, thereby creating a valuable platform for knowledge sharing, feedback, and collaboration on methodological advancements. We thus concluded that ensemble approaches are a promising avenue for improving disease nowcasting. The nowcasts produced for this project were routinely displayed by numerous German media outlets, including Die Zeit, Neue Zürcher Zeitung, and Norddeutscher Rundfunk.

## Probabilistic Solar Forecasting

*Note: Figs. 18 and 19, Table 1, and the original text by [Gneiting et al., 2023a] are licensed under CC BY 4.0. The original text has been adapted into the shorter form presented below.*

The most critical recent development in the science of forecasting solar resources and solar energy is arguably the transition from single-valued deterministic forecasts to probabilistic forecasts. A broad consensus has developed around the idea that solar forecasts should rely on the output of physics-based numerical weather prediction (NWP) models in concert with post-processing by using techniques of statistics and machine learning. Forecasts can be issued for solar irradiance or solar power, with the former referring to the power delivered to a unit area in the form of radiation by the Sun and the latter referring to the generated electricity under conversion.

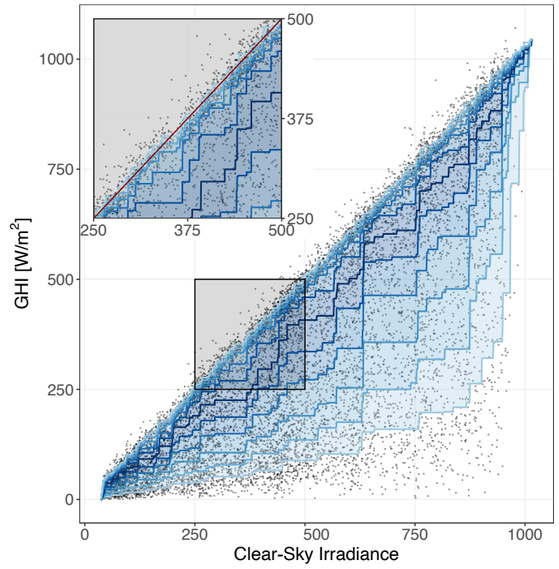



*Figure 18: Illustration of the CSD-IDR approach to training data from 2017–2019 at the Bondville SURFRAD station. The predictive distributions of global horizontal irradiance (GHI) as a function of clear-sky irradiance are represented by predictive quantiles at levels 1/12, ..., 11/12. While the quantile functions are non-crossing, they cluster at high levels, and we thus show a close-up view in the upper-left corner of the figure. The black dots represent clear-sky irradiance and GHI observations from the training archive. The close-up view also shows the diagonal.*

State-of-the-art solar forecasting relies on output either from a single NWP model or – often, but not necessarily – from ensembles consisting of multiple NWP model runs.

The generation of the NWP ensemble can be tailored to solar applications in various ways, including but not limited to perturbations of initial conditions, stochastic perturbations, as well as multi-physics and multi-model approaches. In this context, the term post-processing refers to the conversion and improvement of raw output from NWP models to skillful solar forecasts by using statistical and machine learning techniques. While developments of this type were pioneered in weather prediction, they are now commonly applied in solar forecasting.

Regardless of the field of application, progress in forecasting techniques needs to be demonstrated relative to reference methods. In [Gneiting et al., 2023], we proposed the use of isotonic distributional regression (IDR) as a probabilistic benchmark based on a deterministic forecast of clear-sky irradiance (CSD-IDR). This nonparametric technique yields simple and flexible probabilistic forecasts and can be used whenever training data on a deterministic predictor variable (e.g., clear-sky irradiance and/or smart persistence) and associated outcomes (e.g., irradiance or power) are available. Figure 18 illustrates the IDR approach in the context of day-ahead forecasts of hourly irradiance at the Bondville station of the Surface Radiation Budget (SURFRAD) network in the contiguous United States.

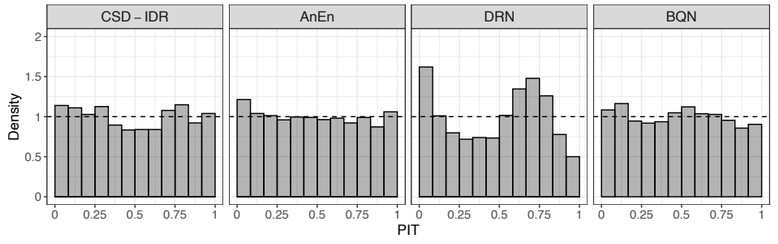



*Figure 19: PIT histograms for day-ahead probabilistic forecasts of hourly solar irradiance with the CSD-IDR, AnEn, DRN, and BQN methods at the Bondville SURFRAD station.*

As new methods of forecasting solar irradiance based on the operational deterministic high-resolution NWP model run by the European Centre for Medium-Range Weather Forecasts (ECMWF), we adapted the distributional regression network (DRN) and Bernstein quantile network (BQN) techniques, which were developed for weather prediction. The DRN approach builds on and extends the framework of non-homogeneous Gaussian regression, in which a typically linear, fixed-form link function expresses the Gaussian parameter vector in terms of input predictions. The DRN technique uses modern neural networks to learn flexible, nonlinear relations between predictor variables and parameter vectors. Owing to the increased flexibility, the vector of predictor variables may now include NWP forecasts for a range of weather quantities well beyond irradiance. The output of the neural network thus consists of the parameters of the forecast distribution based on the predictors at hand. The BQN technique is a fully nonparametric approach in which quantile functions are expressed in terms of basis polynomials. The BQN forecast distribution is defined by the basis coefficients, and the neural network is trained to link the predictors to these coefficients using training data. For both DRN and BQN, we adopted and adapted recent implementations, with the DRN forecast taking the form of a truncated normal distribution. Another competing post-processing technique is the previously proposed analogue ensemble (AnEn) method.

In Figure 19, we present histograms of the probability integral transform (PIT) values, which are derived by transforming the outcome using the predictive cumulative distribution function. If the outcomes are statistically indistinguishable from a sample drawn from the predictive distributions, we expect a flat histogram.

We then say that the forecast is probabilistically calibrated. Deviations from a uniform histogram – as seen for the DRN forecast – can be used to diagnose systematic errors that are present in a forecast.

When comparing forecasts, it is critical to assess the costs or benefits incurred by each competing probabilistic forecast method. From a theoretical perspective, an economic assessment of this type is equivalent to the use of proper scoring rules. The following approach is equivalent to using the mean score under a proper scoring rule: (1) Consider the actual cost–loss structure in an applied problem, (2) find an action that minimizes the expected costs under the probabilistic forecast distribution at hand, (3) compute the actual loss based on this action and the outcome, and (4) average monetary results over a test set. In Table 1, we present the results for the continuous ranked probability score (CRPS), which is reported in the same physical unit as the outcome. Lower values are better, and we see a stable ranking across the SURFRAD stations in the United States, where the BQN method is preferable to the DRN method, which beats the AnEn method, with the clear-sky-irradiance-IDR reference method trailing behind.

Table 1: Mean CRPS (in watts per square meter) (lower is better) for day-ahead probabilistic forecasts of hourly solar irradiance with the CSD-IDR, AnEn, DRN, and BQN methods at the Bondville (BON), Desert Rock (DRA), Fort Peck (FPK), Goodwin Creek (GWN), Penn State (PSU), Sioux Falls (SXF), and Table Mountain (TBL) SURFRAD stations in the United States.

| **Method** | **BON** | **DRA** | **FPK** | **GWN** | **PSU** | **SXF** | **TBL** |
|---|---|---|---|---|---|---|---|
| CSD-IDR | 82.9 | 37.9 | 62.4 | 85.6 | 84.6 | 74.8 | 74.2 |
| AnEn | 56.2 | 31.5 | 49.2 | 59.9 | 59.5 | 54.4 | 61.3 |
| DRN | 52.5 | 28.9 | 45.5 | 55.2 | 55.9 | 52.5 | 57.5 |
| BQN | 50.8 | 28.3 | 44.2 | 53.9 | 55.1 | 50.4 | 55.6 |

Five major aspects of solar forecasting can be distinguished: forecasting methodology, post-processing, irradiance-to-power conversion, verification, and the materialization of values. In [Gneiting et al., 2023a], we touched on essentially all these facets and emphasized areas in which the solar forecasting community might benefit from recent advances in statistical theory and methodology.

## Elicitability of Probabilistic Top-List Predictions

The original text by [Resin, 2023b] is licensed under CC BY 4.0 and has been adapted into the shorter form presented below.

Ideally, a prediction should specify a probability distribution over all potential outcomes. Such predictions are evaluated and compared by means of proper scoring rules, which quantify their value in a way that rewards truthful prediction. In statistical classification and machine learning, the need for reliable uncertainty quantification has not gone unnoticed. However, classifier evaluation often focuses on the most likely class by using classification accuracy – that is, the focus is on whether the most likely class coincides with the outcome.

In [Resin, 2023b], we proposed probabilistic top lists as a way of producing probabilistic classifications in settings in which specifying entire predictive distributions may be undesirable, impractical, or even impossible. A probabilistic top list is a subset of all potential outcomes in which each outcome in the top list is associated with a predictive probability, or confidence score. However, the subset may be chosen by the individual forecaster. For example, a top-1 list may contain only the most likely outcome and its assigned probability, whereas in the other extreme, a top list may be a full predictive distribution over all outcomes. An example of a top-2 list is given in Figure 20.

Consistent evaluation is achieved by using a padded score, which is derived from a symmetric proper scoring rule that is applied to a padded distribution – that is, a distribution that is the result of splitting the unassigned remaining probability mass of a top-list prediction evenly between the remaining classes. In order to prove the validity of the resulting scores, we show that the top list – which comprises the most likely classes and the correct corresponding probabilities – is the unique best prediction for optimizing the score. To that end, results on proper scoring rules are leveraged using the mathematical concepts of majorization and Schur-concavity. An interesting result that is vital when it comes to showing validity is that probabilistic top lists of different sizes remain comparable under padded scores.

Probabilistic top-list predictions offer a flexible framework for the general classification problem. The padded Brier score (which corresponds to squared error) appears particularly suitable for evaluation because top lists that distribute most of the probability mass obtain an expected padded Brier score that is close to optimal. The prediction of probabilistic top lists appears particularly useful in problems in which classification accuracy is not particularly high, as is frequently the case in multi-label classification. Shifting the focus toward probabilistic predictions may well increase prediction quality and usefulness in decision problems in which misclassification losses are not uniform.

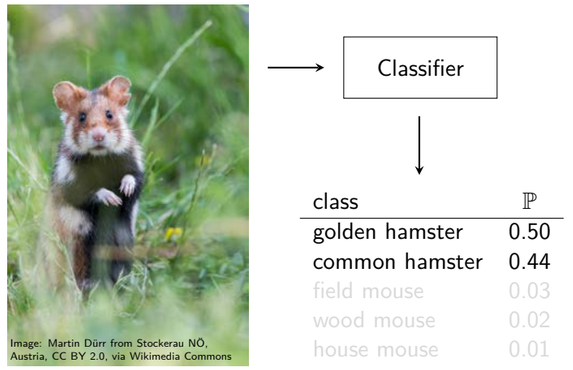



*Figure 20: Consider a (fictious) image classifier that has been trained to identify types (classes) of rodents but that fails to reliably distinguish different kinds of hamsters. A probabilistic top-2 list prediction accounts for most of the classification uncertainty.*

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.

# 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)
Isabel Gernand (PhD student; since November 2023)
Dr. Saskia Haupt (visiting scientist; Heidelberg University)
Alejandra Jayme (visiting scientist; Heidelberg University)
David Lehmann (visiting scientist; Heidelberg University)
Dr. Philipp Lösel (visiting scientist; Heidelberg University, until March 2023)
Stefan Machmeier (visiting scientist; Heidelberg University)
Jacob Jonas Relle (student, until July 2023)
Jonas Roller (visiting scientist; Heidelberg University)
Valentin Schmid (PhD student)
Elaine Zaunseder (visiting scientist; Heidelberg University)
Alexander Zeilmann (visiting scientist; Heidelberg University)
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 Heidelberg University's Engineering Mathematics and Computing Lab (EMCL) at the Interdisciplinary Center for Scientific Computing (IWR), which is also headed by Vincent Heuveline. The DMQ group's research focus lies in gaining knowledge from extremely large and complex datasets, with a particular concentration on computational fluid dynamics (CFD) and biomedical research. 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 2023, the group's research mainly involved developing robust and efficient machine learning methods for uncertainty quantification as well as for the analysis of the acquired data.



## Finding Knowledge in Petabytes of Data

"We are drowning in data, but starved for knowledge" is a famous quote by the author and researcher John Naisbitt, referring to the fact that while we have an incredible volume of data for many research questions, the information hidden inside these data is not given to us for free; instead, we have to "mine" the data in order to uncover the hidden knowledge. This data mining is exactly one of the main research interests in our group, the DMQ (Data Mining and Uncertainty Quantification). Another main research interest lies in uncertainty quantification, which involves analyzing the uncertainty in the data, in the developed mathematical models, and even in the varying user input. While there was already a large volume of research data in 1982, when John Naisbitt uttered his famous statement, now, more than 40 years later, this data abundance has become vastly larger.

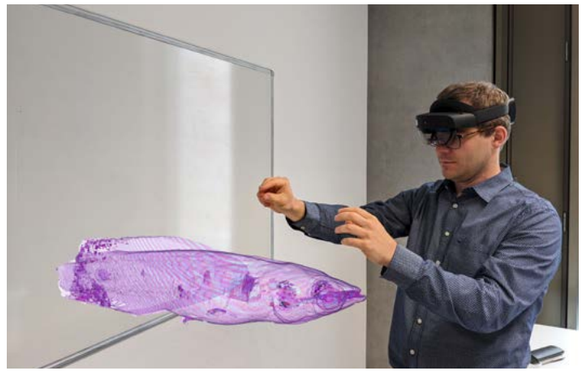


*Figure 21: We use various augmented and virtual reality devices – such as the HoloLens – to interactively visualize our research results.*

In our research project KI-Morph, which is a collaborative effort that we undertake together with the Karlsruhe Institute of Technology, the Museum of Natural History Stuttgart, the Center for Organismal Studies (COS), and the University Computing Center here in Heidelberg, we tackle the challenge of mining petabytes of research data in order to discover the information it contains. We work closely with several biologists who research various biological structures ranging from single cells to insects, fish, and even fossils. These research subjects are scanned using a synchrotron (a type of particle accelerator) in order to produce a digital three-dimensional image volume. While this scanning only takes a few minutes, the subsequent analysis of the volumes by the biologists takes significantly more time that can range from hours to days. Therefore, we are currently developing a platform for petabyte-scale image analysis that automates the repetitive manual tasks that once had to be done by the biologists themselves but leaving them all the research freedom required.

Following the mantra that "science isn't finished until it's communicated," we engage in a variety of outreach activities. One particular focus is on exploring different techniques for augmented and virtual reality in order to interactively display the generated 3D models in the context of workshops and exhibitions.

## Transfer Learning for Extracting Radiomic Features

Radiomics (Figure 22, next page) involves the extraction and analysis of quantitative features from medical images. These features capture a wide range of information, such as texture, shape, intensity, and the spatial relationships of structures within the images. Statistical methods and machine learning techniques have been applied to these extracted features in order to identify patterns, correlations, and associations that might be indicative of the presence, progression, or response to treatment of diseases.

In this work, we study whether deep learning can extract features that are similar or complementary to those obtained through traditional radiomics methods. We focus on the concept of transfer learning (Figure 23, next page), which is a machine learning technique in which a model that is developed for a particular task is repurposed as the starting point for a model for a second task. This concept is especially useful in cases in which a limited volume of data is available for the new task, thereby allowing the model to leverage knowledge from the previously learned task in order to improve performance. In this context, we take general image recognition models that have been trained on large datasets and fine-tune them to specialize in identifying radiomic features. Our study examines the use of image classification models such as Resnet50, VGG19, and InceptionV3 as well as image segmentation models such as FCN and DeeplabV3 as deep learning feature extractors.

Findings indicate that features extracted from pre-trained image classification and segmentation models are highly effective at both classification and regression tasks. The performance of these deep learning-extracted features is comparable to the performance of features extracted via quantitative and data-characterization algorithms. It is noteworthy that similar performance levels can be achieved using only a single slice of an MRI image, which stands in stark contrast to traditional radiomic methods, which typically require a 3D image and segmentation labels. The test accuracies for the pre-trained models were also found to be superior, thereby indicating better generalization. Moreover, the combination of traditional radiomics with

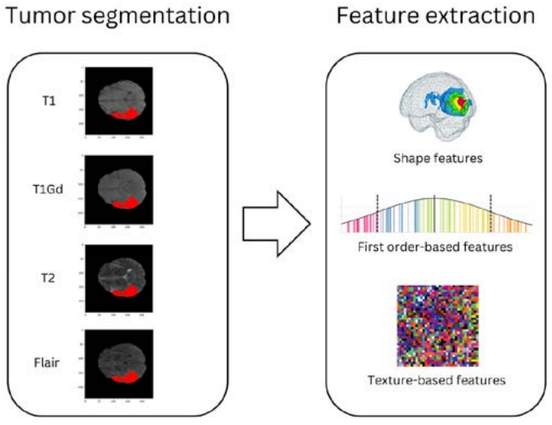



*Figure 22: Radiomics*

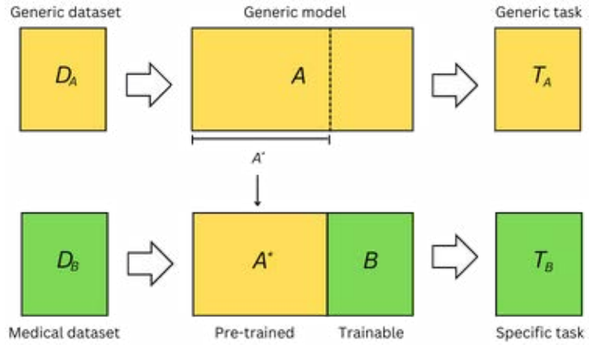



*Figure 23: Transfer learning.*

deep learning-extracted features – which leverages the strengths of both methodologies – leads to better outcomes.

This finding indicates that leveraging pre-trained deep learning models within radiomics can improve the precision and depth of medical imaging analyses. Further work involves investigating whether these extracted features can be linked to or explained by specific physical characteristics, with the goal of understanding the correlation between the features and their underlying biological or pathological attributes.

## Mathematical Modeling of Infant Metabolism for Newborn Screening

The complex system of human metabolism is at the core of pediatric metabolic research, in which metabolic processes are investigated in order to understand child metabolism. This is especially important in the context of newborn screening programs that aim at the early ideally, pre-symptomatic, identification of treatable inherited metabolic diseases in order to reduce morbidity and mortality. Mathematically modeling infant metabolism enables a detailed causal analysis of different metabolic processes that are related to these severe metabolic diseases. Whole-body metabolism can be represented by a genome-scale metabolic model using the constraint-based reconstruction and analysis approach.

In a collaborative initiative with both the Center for Child and Adolescent Medicine at the Heidelberg University, Medical Faculty, and the Molecular Systems Physiology group at the University of Galway, we are currently working on integrating mathematical modeling into newborn screening research. Therefore, we developed organ-resolved, sex-specific, anatomically accurate mathematical models of newborn and infant metabolism (infant-WBMs). These infant-WBMs are parameterized in order to accurately represent the distinct metabolic characteristics of newborns and infants and include over 80.000 metabolic reactions and 26 organs. In particular, the infant-WBMs account for infant-specific organ weights, energy requirements for brain development, heart function, and thermoregulation; both dietary and energy requirements for physical activity. By applying a flux balance analysis, the infant-WBMs can be optimized for growth and can predict neonatal and infant growth that is consistent with the growth recommended by the World Health Organization. By this, the infant-WBM resource can provide valuable insights into infant metabolism on an organ-resolved level and can additionally enable a holistic view of the metabolic processes that occur in infants (Figure 24).

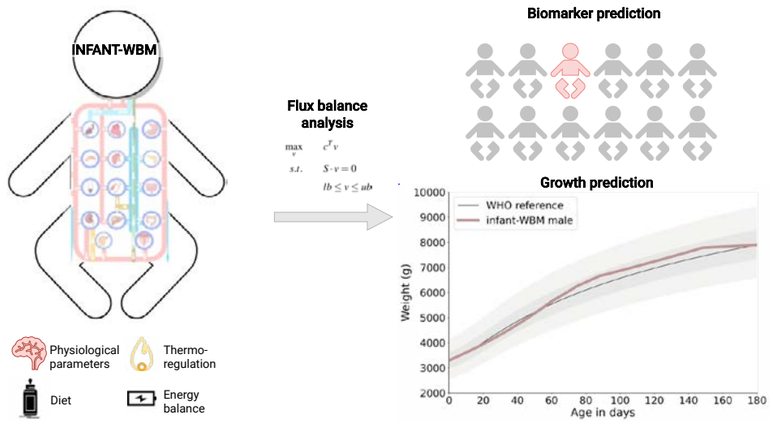



*Figure 24: Mathematical model of infant metabolism (infant-WBM) that uses flux balance analysis to predict infant growth and known biomarkers of inherited metabolic diseases*

These models can be applied to in silico simulations in newborn screening. Using sex, birth weight, and metabolite concentration measurements, the models can be personalized in order to investigate an individual's metabolism. As such, the infant-WBM resource holds promise for personalized medicine, and could represent a first step to personalized systematic simulations of newborn screening conditions and corresponding treatment planning.

## Model Order Reduction in Convective Heat Transfer Simulation

Heat exchanger systems in engineering and home appliances commonly consist of fluids that are contained between layers of solid materials, such as double-pane glass windows in which air is trapped between the two layers, which greatly improves thermal insulation. Enhancing the heat transfer in such systems via efficiently induced convection is of fundamental interest in order to develop sustainable and low-cost heating and cooling devices. In recent years, such enhancement that utilizes electric fields has gained a wider scientific audience, thanks to numerous international research projects. This approach is referred to as thermal electro-hydrodynamic (TEHD)-driven heat transfer augmentation. As part of our project in co-operation with partners at BTU Cottbus, we aim to gather insights into the dynamics of a dielectric fluid between two differentially heated cylinders, with an applied voltage at the inner cylinder and a grounded outer cylinder. Our insights stem both from the numerical simulation of the physical TEHD model equations and from comparisons with experimental data (Figure 25).

In order to obtain the numerical solution for this set of partial differential equations (PDE), we use the finite element method (FEM). This type of discretization has been shown in the previous project phase to converge to the exact solution of the TEHD equations by way of a priori error estimates. However, since the TEHD equations are a highly coupled and nonlinear PDE system in three spatial dimensions, computation of the numerical solution by FEM requires a considerable amount of time and energy. We therefore investigate model order reduction techniques that could speed up the computations and to enable further research in areas that rely on cheap model evaluations, such as optimal control of the TEHD system.

Our approach to model order reduction centers on Galerkin reduced order models (G-ROM) that are based on proper orthogonal decomposition (POD). The POD enables us to compute a lower-order approximation of the finite element space, while the G-ROM enables us to integrate the POD basis seamlessly into the TEHD model equations, allowing us to compute a numerical solution several orders of magnitude faster than the FEM solution. For an unsteady TEHD flow in a two-dimensional circular geometry, thus ignoring the axial extent of the cylinders, we observed a clear convergence of the ROM solution to the FEM solution with regard to the dimension of the POD space. This convergence can be seen in Figure 26 for the temperature variable. Computational investigations into the full three-dimensional annulus are currently ongoing.

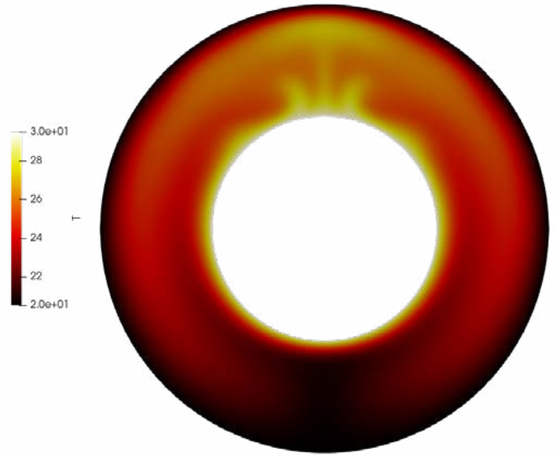



*Figure 25:*
*Snapshot of the reduced-order numerical temperature solution in the circular annulus with the inner boundary being 9K warmer than outer boundary. A voltage of 10kV is applied.*

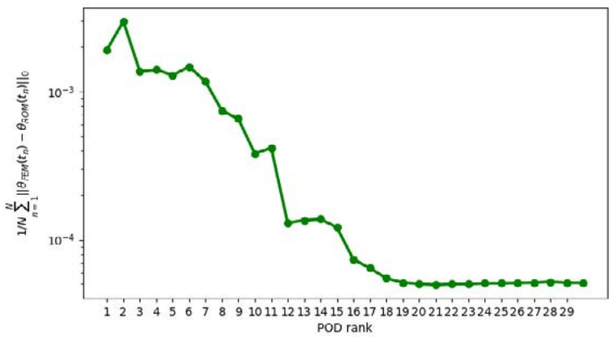



*Figure 26:*
*Convergence of the difference between the FEM and G-ROM solutions for two-dimensional TEHD flow with regard to the temperature variable. The other solution variables exhibit identical qualitative behavior.*

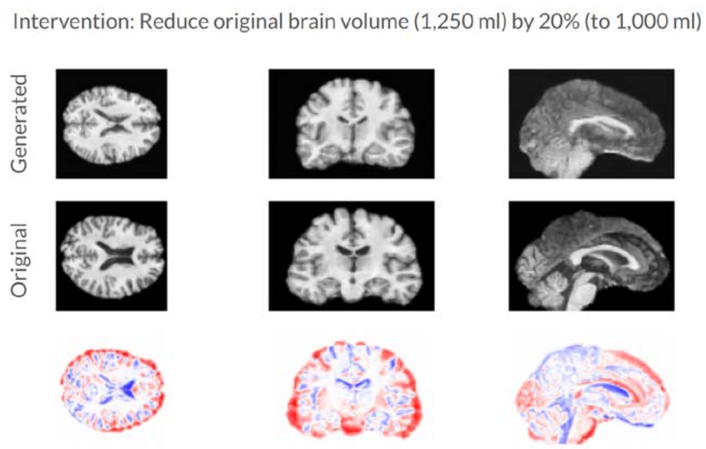



*Figure 27: Counterfactual image generation based on an intervention on the image causal parents (here, brain volume).*

## Counterfactual Image Generation Using Latent Diffusion Models

Investigating causal dependencies between images that are acquired in the medical context one the one hand and data of the image generation process on the other hand is fundamental to medical research. Whereas treatment planning, annotation, and referral rely on medical images and the analysis of these images, the images themselves are subject to acquisition parameters, patient characteristics and potential medical intervention. The potential to infer relationships between patient characteristics and structural changes in the anatomy or phenotypes of diseases that are depicted in an image is particularly interesting. However, modeling counterfactual queries such as “How would an MRI image of a patient change if the patient had a different genetic variation?” is a problem that is difficult to solve due to potential confounding factors outside the considered data.

In 2023, we began working with the Mannheim Institute for Intelligent Systems in Medicine (MIISM) on developing a novel deep learning approach for causal image generation that enables counterfactual reasoning. The proposed model consists of a conditional latent diffusion model (LDM) and a variational graph auto-encoder (VGAE). The image generation is carried out by the LDM and is guided by the direct causal parents of the image. These are generated using a VGAE that learns the structural assignments of all the variables of the data generation process, such as patient variables. While data from Alzheimer's patients is investigated as an example use case, this approach can be applied to any kind of causal image generation for which the casual graph of the image generation process is given. For instance, if provided a structured graph of the additional information, the proposed model allows for counterfactual dose distributions when intervening on radiation treatment parameters or when generating images of the patient after the treatment, i.e., predicting the radiation therapy effect.

By learning the causal relationships of the image generation process, the framework provides a principled approach to defining and assessing hypotheses of the underlying effect of a variable's change in the image. Developing such models allows for assessing the effects of intervening on variables outside a randomized controlled trial setting and could ultimately lead to a better understanding of the disease generation process. Furthermore, these models could enable treatment effect estimates, disease progression modeling, and precision medicine for patients.

## Modeling and Simulating Early Colorectal Cancer Development

Colorectal cancer is one of the most common types of cancer worldwide. We are currently in need of a profound understanding of its development in order to create effective intervention. Using mathematical modeling, we explore the complex process that governs colorectal cancer development at the crypt level, a microscopic structure in the human colonic epithelial layer, where cell proliferation and differentiation occur.

As part of the research project "Mathematics in Oncology – Towards optimal prevention and treatment in patients with inherited cancer syndrome", we have been working with three-dimensional mathematical models that use partial differential equations to simulate the cancer development in the human colon.

First, the special geometry of the colon is treated. We developed a model that describes the finger-like invaginations in the colon (i.e., so-called crypts) in a natural way using elastic growth equations. We apply the finite element method in order to solve these equations using the in-house open-source library HiFlow3.

This process makes it possible to generate geometries that very closely approximate the actual epithelial layer in the intestine, as shown in Figure 28.

Parameters in this model are used to describe important properties, such as cell growth and cell death. This process enables us to investigate the effects of uncontrolled cell growth or cell death resistance on the tissue structure in a virtual environment. In addition, it is also possible to virtually investigate how signaling pathways may change as a result of changes in the tissue.

Furthermore, by using this accurate spatial configuration, other known mathematical models can now be applied to the special structure of the colonic crypts. The aim is to make predictions about cancer growth that are as accurate as possible depending on the structure of the epithelium. An example of a cancer development simulation can be seen in Figure 29.

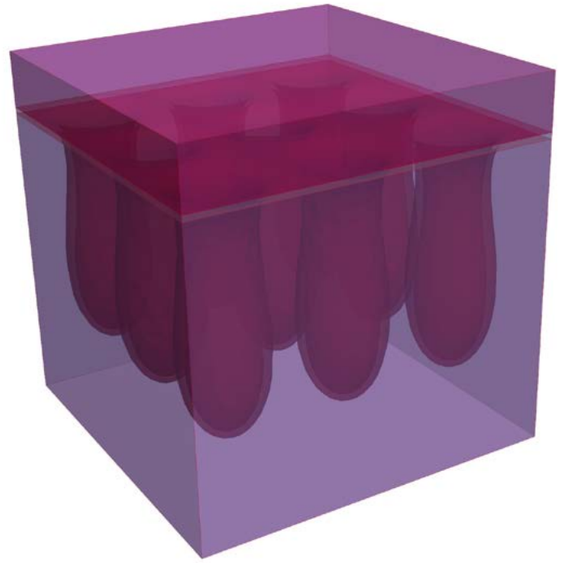


*Figure 28: Human colonic crypt geometry generated with elastic growth equations with the goal of being as close as possible to the actual structure of the colon. The colonic crypts can be seen as tubular invaginations lining the human colon.*

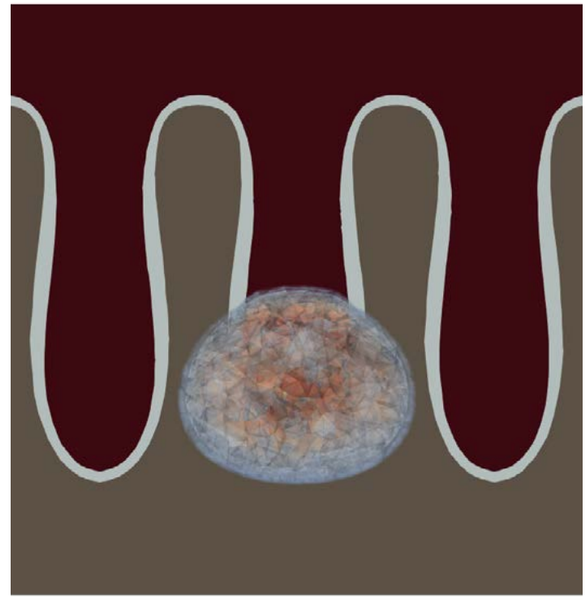


*Figure 29: Simulating cancer development with a standard growth equation from the geometry shown in Fig. 28. Cancer cell density above a certain level can be seen in the middle crypt, where the red color indicates a higher density*

Die Gruppe **Data Mining and Uncertainty Quantification (DMQ)** unter der Leitung von Vincent Heuveline begann ihre Forschung im Mai 2013. Die Gruppe arbeitet eng mit dem Engineering Mathematics and Computing Lab (EMCL) am Interdisziplinären Zentrum für Wissenschaftliches Rechnen (IWR) an der Universität Heidelberg zusammen, das ebenfalls von Vincent Heuveline geleitet wird. Im Fokus der Forschungsarbeit steht ein zuverlässiger und strukturierter Erkenntnisgewinn aus großen und komplexen Datensätzen mit Schwerpunkt auf Computational Fluid Dynamics (CFD) und biomedizinischer Forschung. Beide Bereiche – Data Mining und Uncertainty Quantification – erfordern einen interdisziplinären Ansatz für mathematische Modellierung, numerische Simulation, hardwarenahe Programmierung, Hochleistungsrechnen und wissenschaftliche Visualisierung. Im Jahr 2023 hat die DMQ Gruppe besonders die Entwicklung robuster und effizienter maschineller Lernmethoden für UQ sowie die Analyse der erfassten Daten weiterentwickelt.

# 2.6 Machine Learning and Artificial Intelligence (MLI)

### Group leader

Jun.-Prof. Dr. Jan Stühmer

### Team

Benjamin Barth (master's student; since September 2023)
Romain Chazotte (master's student; since June 2023)
Mila Coetzee (PhD student; since July 2023)
Moritz Feik (bachelor's student; since December 2023)
Xiaoxuan Fu (bachelor's student; August–November 2023)
Sanjana Satish (intern; May–July 2023)
Steven Schürstedt (master's student; since September 2023)
Leif Seute (PhD student with the MBM group; since October 2023)
Simon Wagner (master's student; since November 2023)
Zixiang Zhou (student; May–August 2023)

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 we explore enable us to take an existing model (e.g., a model that was trained on a large 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. This process enables the use of machine learning in application areas that have thus far not been able to benefit from such technology.

Another research focus in the group is on learning interpretable representations with the goal of leading to a better understanding of the underlying principles of a machine learning model. 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, several collaborations with other research groups have been initiated, including with the Molecular Biomechanics (MBM) and the Computational Carbon Chemistry (CCC) groups. Within these collaborations, the MLI group contributes its expertise in geometric deep learning, which is 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.



## Neuro-Symbolic Reasoning for Factually Grounded AI

If you have interacted with one of the recently emerging language models or chat bots, then you have likely noticed a problem: The models and bots can make up facts, which is referred to as "hallucination." These made-up facts are often consistently integrated into the rest of the presented knowledge, thereby making them difficult to spot as incorrect, even for users familiar with the general subject. This situation has already led to negative consequences for individuals who use such language models without checking that the answers they receive are trustworthy. For example, there have been cases in which a lawyer has submitted documents to a court with references to presumably related but non-existent cases. Moreover, as recently demonstrated, these hallucinations also pose security risks for software developers when references to non-existent software libraries are inserted into generated code (The Register, AI hallucinates software packages and devs download them – even if potentially poisoned with malware, 2024). This problem is clearly too important to ignore.

Currently, the most prominent solution that is being developed in industry is based on "retrieval-augmented generation" (Lewis P et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, NeurIPS, 2020). In addition to the user's query, we append all sorts of results to related search queries from databases in addition to appending on-the-fly internet search results to the input. Perhaps it is only a slight exaggeration to say that through this process, we expect some of the additionally presented data to include the true answer to the user's query. This approach raises the question as to why we then spend so much time and utilize so many resources for training a large language model with billions of parameters on enormously large volumes of data in the first place if we do not trust the data stored in the language model at inference time.

Why do language models make up facts? While we do not yet have a conclusive answer to this question, some recent findings about where language models store factual information have proven interesting (Meng K et al. Locating and Editing Factual Associations in GPT, NeurIPS 2022 and Vig J et al. Investigating gender bias in language models using causal mediation analysis. NeurIPS, 2020). Moreover, this research has shown that when we "overwrite" some of this knowledge with incorrect information, this incorrect information is then consistently integrated into the rest of the output in the same way that "hallucinated" (i.e., made-up) facts are consistently integrated into the model's response. It would thus appear that the problem of hallucinations in large language models is nothing more than a problem of incorrectly stored or retrieved factual knowledge. Knowledge can be incorrectly stored because although modern language models have a vast number of trainable parameters, there is a limit to the amount of information that can be stored consistently. Furthermore, knowledge can also be incorrectly retrieved since there is a limitation to the extent that language models can perform logical inference and thus retrieve consistent information at inference time.

In our research, which received a "top 5% honorable mention" award at the 2023 International Conference on Learning Representations (ICLR), we developed not only an inference process that incorporates the uncertain nature of neural network inference, but also an architecture that integrates two thus-far mostly disjoint worlds: namely the uncertain inference process of a neural network on the one hand and exact symbolic inference on the other hand [Cornelio et al., 2023]. Neural network inference and probabilistic machine learning have long been mostly disjunct from exact symbolic inference in the form of symbolic manipulation based on logical rules. In our proposed architecture (depicted in Figure 30, next page), both worlds are combined while utilizing their respective strengths.

In the first step, we use the much faster neural network inference to gain a partial solution to a problem, for example, when solving a Sudoku board. To achieve this partial solution, two separate neural network models are used: First, a Neuro-Solver tries to predict a number for every field in the Sudoku board. Second, another neural network component then classifies which of the predicted numbers is likely correct and which is likely incorrect. Thus, the uncertainty involved in neural network prediction is taken into account during the inference process. The second stage of the method involves an exact symbolic inference process – that is, a Logic Solver – that solves the Sudoku board via logical inference. The benefit is that the symbolic Logic Solver is provided with a partial solution, and the logical inference process is thus much faster than it would be without the neural component. This way, even Sudoku problems that appear to be too difficult to solve via logical inference alone – due to the large size of the space of possible solutions – are solvable with our combined approach. It is crucial that the false positive rate of the Mask Predictor not be too high; otherwise, it would be possible to not have an admissible solution to the board in too many cases. Therefore, after pre-training each component individually, the entire system is carefully tuned via reinforcement learning.

Of course, the question then arises as to how solving Sudoku boards is related to language models and their hallucinations. In the end, in both cases, we use machine learning models to predict an answer to a given question. Instead of merely relying on possibly incorrect predictions, as state-of-the-art large language models do, our proposed neuro-symbolic approach allows the uncertain nature of neural network

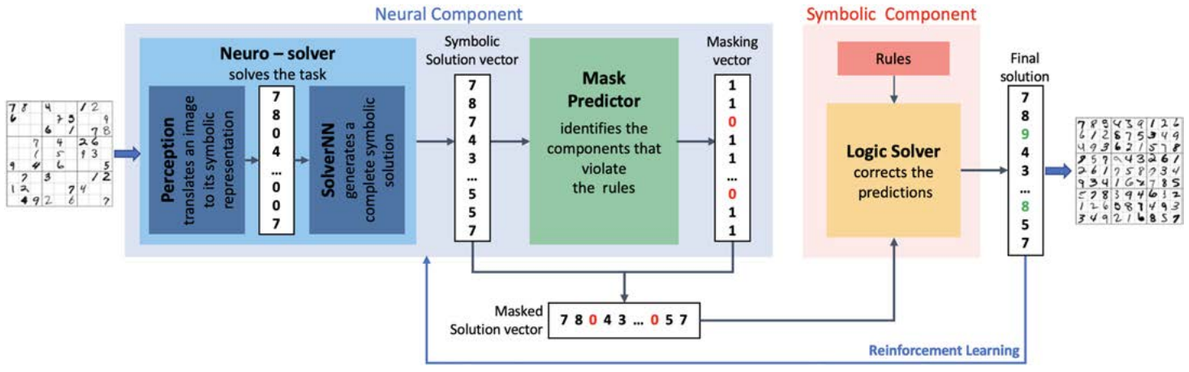



*Figure 30: The architecture of our proposed neuro-symbolic approach. First, a neural network model – the Neuro-Solver – is used to predict a solution. Then, another neural network model – the Mask Predictor – is used to classify which of the predicted values is correct, thereby leading to a partial solution. This partial solution is then used as initialization for the Logic Solver, which solves the board to completion [Cornelio et al., 2023].*

prediction to be taken into account during the inference process. By combining the neural network inference process with symbolic inference, the approach can generate provably correct answers. As an additional benefit, the approach is also intrinsically interpretable – that is, the symbolic part is interpretable because the symbolic variables usually carry interpretable semantics. Moreover, even the neural network model – which is a transformer – becomes interpretable, as we demonstrated when visualizing the attention mechanism of the transformer, which revealed that the model learned the three rules of Sudoku: namely that there can be only one distinct digit per row, column, and 3x3 block (Figure 31).

## Invariance and Equivariance as Design Principles in Machine Learning

A recent success story of machine learning is the development of so-called *foundation models*, which are large models for extracting robust features from data and that are suitable for solving a large variety of downstream tasks. A prominent example can be found in large language models, which learn a rich feature representation from the surrogate task of predicting the next word of a given text. After model training, not only can the feature representation be used to solve the surrogate task of predicting the next word, but the learned feature extractor can also be used for various other tasks, such as dialog systems and text summarization.

Because no additional information beyond the text itself is needed to formulate this training task, these methods are also referred to as *self-supervised learning*. For text data, the task of predicting the next word has been shown to be highly efficient at learning a rich feature representation. When performing self-supervised learning for image data, the surrogate task is adapted to images: Instead of predicting the next word, the surrogate task is to predict the content of a cropped-out region of the image. Additionally, a large number of transformations are applied to the image data in order to achieve invariance to unnecessary details and robustness to transformations, such as changes in image brightness, changes in color, and other characteristics that do not change the semantic information contained in the image. This transformation of the data is called *data augmentation* and has the goal of achieving invariance of the feature representation to the transformations used during augmentation. In the end, we would like to achieve feature representations that are invariant to unnecessary details and noise in the data while preserving as much information as possible.

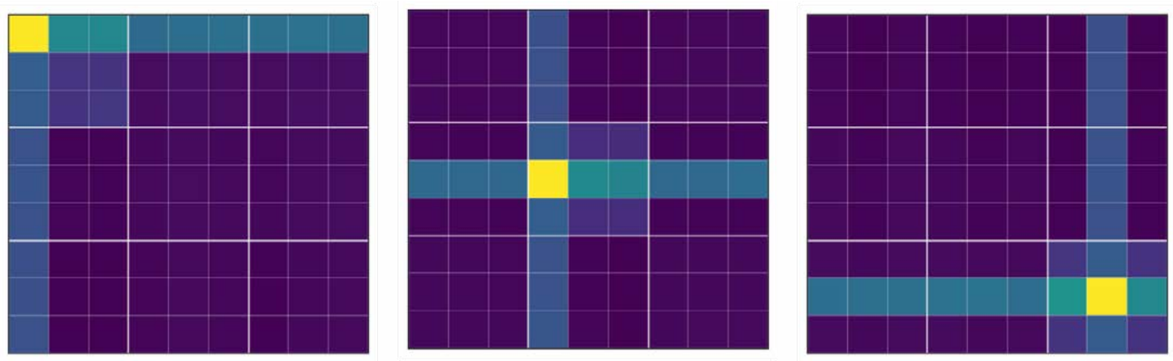


*Figure 31: The attention mechanism of the Neuro-Solver is intrinsically interpretable and reveals that the model learned the three rules of Sudoku: namely that there can be only one distinct digit per row, column, and 3x3 block [Cornelio et al., 2023].*

A challenge that remains is the choice of the correct data augmentations: The desired invariances are often task dependent. In object recognition, for example, we would like to have a classifier that is rotation invariant because the identified class of an object should not depend on the orientation or position of the object. However, when the task is to determine the orientation and position of an object, both rotation and position become the desired outcome and belong to the crucial information that should be preserved. This process usually requires manually choosing the correct augmentations during training. In order to make the process fully self-supervised, we developed a methodology that first learns a representation that supports multiple invariances, and include a differentiable hyper-parameter that encodes the desired invariances of each training task [Chavhan et al. 2023]. Then during fine-tuning on a later downstream task, the identification of the correct task-specific invariances simply amounts to finding the correct hyper-parameter for encoding these invariances, which is a task that can be solved fully automatically during fine-tuning.

Another field in machine learning in which the explicit modeling of invariances to transformations of the input data has significantly improved predictive accuracy is machine learning for computational chemistry. For example, when predicting the chemical properties of a molecule, we know that these properties are usually not affected by rotating the molecule in space. Thus, when our input features depend on the atom coordinates, we want the predicted chemical properties to be invariant to rotations of these coordinates. In order to achieve this invariance, one of two methods is usually used: Either we choose intrinsic input features that are invariant to rotation, such as bond lengths between atoms, or we use equivariant feature representations that transform *equivariantly* when the molecule is rotated. These equivariant representations can be understood as representations that change together with the transformation while leaving the rest of the information about the molecule intact.

Whereas intrinsic input features are usually restricted to simple features (e.g., bond lengths), equivariant representations – in turn – can quickly lead to complicated model architectures as well as to increased computational requirements during training. In collaboration with Pascal Friederich from the Karlsruhe Institute of Technology, we developed an efficient model for predicting the property of materials. The model utilizes the symmetry of the material's crystal lattice in combination with higher-order intrinsic features [Ruff et al. 2023]: First, we use an efficient representation that captures all the symmetries of the material's crystal lattice in the form of a single unit cell. Second, we compute higher-order features, such as angles and dihedral angles, which are intrinsic features of higher-order representations of the topology of the molecular graph (i.e., the first- and second-order line graph). In the future, we would like to extend these concepts and develop more efficient and more accurate equivariant representations that utilize this higher-order topological information.

Die **MLI Gruppe**, die im September 2022 am HITS gegründet wurde, 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. Dadurch können maschinelle Lernverfahren 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 Molecular Biomechanics (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.

# 2.7 Molecular Biomechanics (MBM)

## Group leader
Prof. Dr. Frauke Gräter

## Staff Scientist
Dr. Camilo Aponte-Santamaria

## Team
Saber Boushehri
Matthias Brosz
Johanna Buck (student, since April 2023)
Jannik Buhr
Svenja de Buhr
Lucas Diedrich (student, January–March 2023)
Guido Giannetti (visiting scientist, University of Vienna, Austria)
Christina Goß (student, April–September 2023)
Eric Hartmann
Marie Koerner (student, April–June 2023)
Dr. Markus Kurth
Dr. Debora Monego
Dr. Erik Poppleton (visiting scientist, Heidelberg University)
Benedikt Rennekamp
Kai Riedmiller
Andrea Sassoli
Leif Seute (since October 2023)
Moritz von Stetten (student, April–June 2023)
Daniel Sucerquia
Nuriza Suleimenova (since October 2023)
Aysecan Ünal
Evgeni Ulanov (student, until May 2023)

Machine learning enhances molecular simulations in many ways and can even partly or fully replace them. It also boosts the accuracy and widens the scope of physics-based modeling and can lengthen the accessible time scales. SIMPLAIX is a research initiative that aims at leveraging the capacity of machine learning for use in molecular problems in interdisciplinary projects. It was initiated by HITS, KIT, and Heidelberg University and is run in collaboration across the three participating institutions. Our group has greatly profited from this initiative, which has now reached its full momentum.

With our interest in the mechanochemistry of materials such as proteins and polymers, we aim to incorporate chemical reactivity into classical molecular dynamics simulations. To that end, we introduced a hybrid simulation scheme that is highly efficient but that relies on accurate predictions of the involved barriers. In collaboration with Pascal Friederich, KIT, and SIMPLAIX PI as well as with Anya Grynova, HITS, and SIMPLAIX PI, we were able to use graph neural networks to investigate chemical reactivity with the help of molecular dynamics simulations. It was thus possible to obtain satisfying accuracy in light of the challenge of measuring hydrogen transfer reactions amidst the jiggling and wiggling of the dynamic and dense protein environment. Our results can be used to build the next generation of machine learning–enhanced reactive simulations, especially by using the expertise available within SIMPLAIX.

The incorporation of machine learning into simulations has not changed the primary goals of our computational research, which are to provide an understanding of mechanisms behind biological functions and to generate hypotheses that can be experimentally tested. The MBM group proudly published its first primarily experimental study in 2023, with most of its results having been generated in the group's own wet lab. A highly related and largely computational second study published in 2023 serves as a computational microscope for zooming into some of our experimental findings. Our wet lab continues to be generously supported by the European Research Council and the Cluster of Excellence 3DMM2O.

## Research Highlights

**Sacrificial Bonds in Collagen**
One of the more unusual ways that objects can increase their longevity is by sacrificing a part of themselves. One example is when a lizard's tail breaks off in order to enable the animal to escape danger. By using extensive simulations of collagen atomistic fiber models, we revealed that sacrificial parts can also be found within collagen, which is the most abundant protein in our bodies. We were able to show how the rupture of weak sacrificial bonds within collagen tissue helps (1) to localize damage caused by excessive force, (2) to minimize negative impacts on the wider tissue, and (3) to promote recovery.

By using complementary computational and experimental techniques to study collagen in rat tissue, our findings indicate that weak bonds within the crosslinks of collagen have a strong propensity to rupture before other bonds, such as those in the collagen's backbone. This propensity serves as a protective mechanism, localizes the detrimental chemical and physical effects of radicals caused by ruptures, and likely supports molecular recovery processes. We were able to validate our findings via gel electrophoresis and mass spectrometry experiments conducted on rat tails, flexors, and Achilles tendons.

Collagen comprises roughly 30 percent of all proteins in the human body. It provides strength to bones, elasticity to skin, protection to organs, and flexibility to tendons in addition to aiding in blood clotting and supporting the growth of new cells. Structurally, collagen resembles a triple-braided helix, with three chains of amino acids intertwining to form a strong and rigid backbone. Each collagen fiber contains thousands of individual molecules that are staggered and bound to one another by crosslinks, thereby contributing to the collagen's mechanical stability. Collagen crosslinks were once thought to be susceptible to rupture; however, at the time, little was known about the specific sites of bond ruptures or about why these ruptures occur where they do. For the first time, we were able to show that sacrificial bonds in collagen serve a vital role in maintaining the overall integrity of the protein and in helping to localize the impacts of mechanical stress, which could otherwise have catastrophic consequences for the tissue.

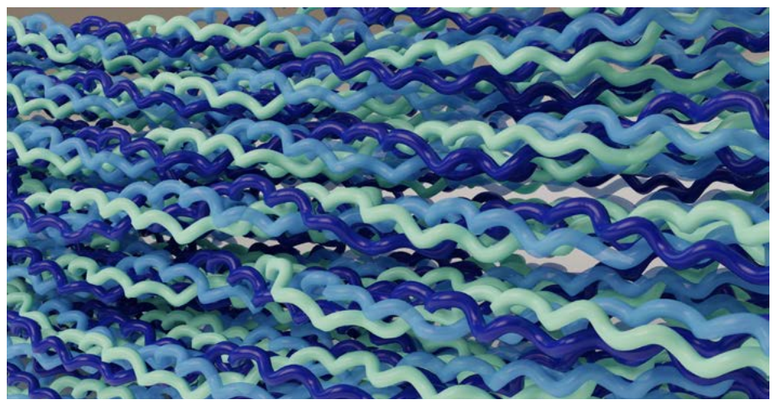
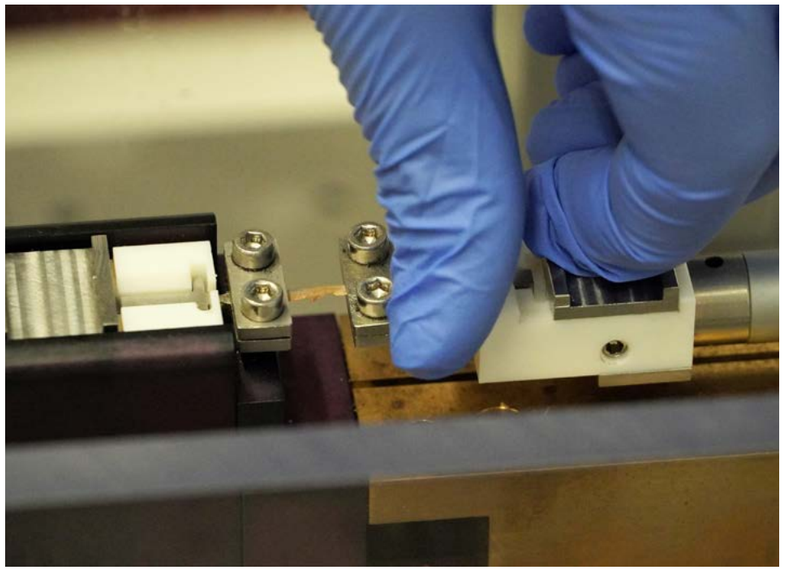


*Figure 32: The collagen triple helix.*

Our work sheds light on the rupture mechanisms of collagen, which is crucial when it comes to understanding tissue degradation and material ageing and can also potentially help to advance tissue engineering techniques.

**Challenging Predictions from Computer Simulations: The MBM Group's Wet Lab**
We first began to build the experimental laboratory of the MBM group at Heidelberg University's Neuenheimer Feld campus three years ago. Since then, we have not only recruited new members, but also expanded our equipment and protocols. Our endeavors culminated in the group's inaugural experimental paper, which was published in 2024 in the journal Angewandte Chemie.



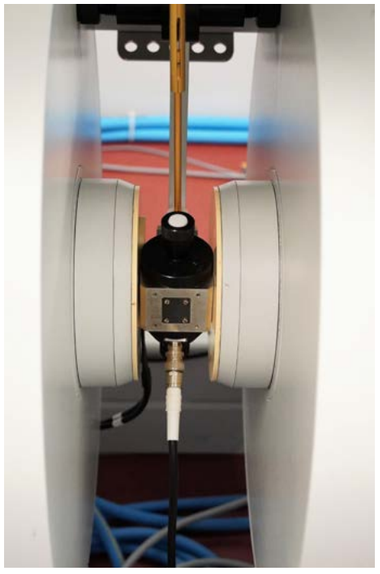


*Figure 33: Stretching collagen to its limits, then detecting radicals with very high sensitivity in the magnetic field of an electron paramagnetic resonance spectrometer.*

Our laboratory at the Institute for Molecular Systems Engineering and Advanced Materials (IMSEAM) at Heidelberg University was initially designed as a facility for an electron paramagnetic resonance (EPR) spectrometer as part of an infrastructure program that was made available through the Cluster of Excellence 3DMM2O, which we are part of. However, once the first students became involved, we began to extend our capabilities. Currently, we have facilities for sample storage, tissue preparation, bioassays, and material testing. Our most recent expansion added a climate chamber to our extensometer, which is frequently utilized in experiments on Achilles tendons in rats and on other tissues. Additionally, through IMSEAM, we have access to a state-of-the-art device and spectrometer pool as well as to the newly equipped biomechanics core facility.

All of our projects follow a highly collaborative approach that integrates a set of experiments as well as the simulations performed in our group at HITS. This approach enables us to test and challenge computational findings by experiments and vice versa. Importantly, we were able to identify dihydroxyphenylalanine (DOPA) modifications as radical scavengers and post-translational modification across various collagens ranging from the tail and the Achilles tendon to the meniscus. In collaboration with groups in Stockholm and at the German Cancer Research Center in Heidelberg, we were able to show that these special modifications are formed from the standard building blocks in proteins. However, one puzzle that remains unsolved involves determining how this modification comes about, which is a question that we plan to address next via simple model systems of collagen. We are convinced that our insights on collagen as a radical scavenging material that is used to cope with large mechanical stresses can aid in tissue engineering and can also be harnessed to combat oxidative stress and tissue ageing.

**Substituting Density Functional Theory in Reaction Barrier Calculations for Hydrogen Atom Transfer in Proteins**

Free radicals can have a critical and deleterious impact on biological systems. Moreover, they are also highly reactive and lead to non-specific damage in proteins, DNA, and lipids, thereby causing various diseases and aging. Radical formation is followed by a plethora of subsequent reactions, with the most important being radical propagation through hydrogen atom transfer (HAT). Due to the high reactivity of radicals, intermediate products of radical reactions can be very short-lived and therefore difficult to capture experimentally. Computationally, these reactions can be described via highly costly ab initio calculations. Molecular dynamics (MD) simulations can simulate proteins fast and efficiently, but they do not allow reactions to occur. Therefore, in order to search through the MD trajectory for likely reactions, it is necessary to choose a method that is capable of quickly assessing a vast number of reactions.

In order to assess hundreds of reactions in mere seconds instead of in dozens of days (as is the case with ab initio calculations), we built a machine learning model that predicts energy barriers of HAT reactions in proteins. Our model is based on graph neural networks, which operate directly on the molecular graph.

The developed method begins with structures from an MD simulation of the radical system of interest. Based on these structures, the model predicts HAT barriers for every possible reaction. By examining the barrier, it is possible to infer the likelihood that a particular reaction will occur. The model will be used in the future in the KIMMDY software to automatically change the topology of the system based on the prediction of the most likely reaction, and the MD simulation will continue from there.

As the method is based on machine learning, the model must first be trained on structures with known energy barriers. Therefore, a large data set of ~19,000 energy barriers was created using the hybrid DFT functional BMK/6-31+G(2df,p). In order to ensure an efficient training process, the model learned to predict accurate barriers in two phases. In the first phase, the model was trained with barriers from structures that had been taken directly from MD simulations. These barriers are cheaper to calculate since they require only seven single-point DFT calculations. In the second phase, the model was refined by training a transfer learning process on higher-quality barriers. These barriers were calculated by optimizing the start state and transition state of the reaction on the DFT level, thereby making the calculations far more expensive than in the first phase, but also more accurate. The resulting model achieves a mean absolute error of 3.6 kcal/mol on structures from MD and predicts optimized reaction barriers. (See publication: Riedmiller K, Reiser P, Bobkova E, Maltsev K, Gryn'ova G, Freiderich P, Gräter F: Substituting density functional theory in reaction barrier calculations for hydrogen atom transfer in proteins. Chemical Science, Issue 7. 2024. DOI: https://doi.org/10.1039/d3sc03922f)

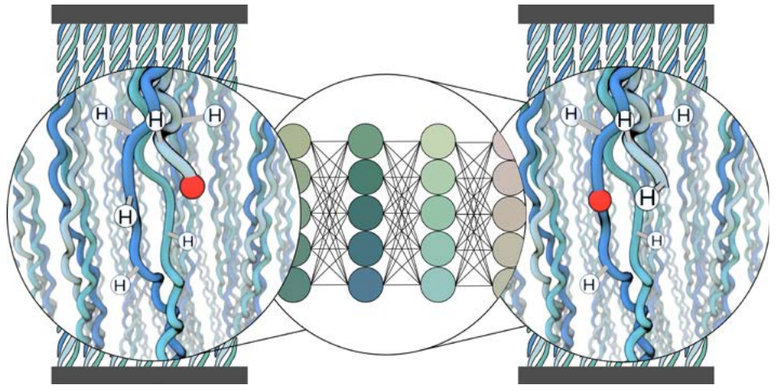



*Figure 34: Predicting reactivity with machine learning.*

Maschinelles Lernen verbessert molekulare Simulationen in vielerlei Hinsicht oder ersetzt sie teilweise oder vollständig. Es erhöht die Genauigkeit, die zugänglichen Zeitskalen und den Umfang der physikbasierten Modellierung. SIMPLAIX wurde vom HITS, dem KIT und der Universität Heidelberg initiiert, um das Potenzial des maschinellen Lernens für molekulare Probleme in interdisziplinären Projekten zu nutzen, in Zusammenarbeit der drei beteiligten Institutionen. Unsere Gruppe hat stark von dieser Initiative profitiert, die nun ihre volle Dynamik erreicht hat.

Mit unserem Interesse an der Mechanochemie von Materialien wie Proteinen und Polymeren zielen wir darauf ab, chemische Reaktivität in klassische Molekulardynamiksimulationen zu integrieren. Zu diesem Zweck haben wir ein hybrides Simulationsschema eingeführt, das hocheffizient ist, aber auf einer genauen Vorhersage der beteiligten Barrieren beruht. In Zusammenarbeit mit Pascal Friederich (KIT und SIMPLAIX PI), und Anya Gryn'ova (CCC Gruppe am HITS und SIMPLAIX PI), konnten wir hierfür graphische neuronale Netze nutzen. Angesichts des anspruchsvollen Problems erreichten wir eine zufriedenstellende Genauigkeit: Wasserstofftransferreaktionen inmitten des Wackelns und Zuckens einer dynamischen und dichten Proteinumgebung. Dies ist ein Ergebnis, auf dem wir eine nächste Generation von reaktiven Simulationen, die durch Maschinelles Lernen ermöglicht wird, aufbauen können, insbesondere mit Hilfe der Expertise von SIMPLAIX.

Die Einbeziehung von maschinellem Lernen in Simulationen hat die primäre Funktion unserer Computer-gestützten Forschung nicht verändert: ein mechanistisches Verständnis und die Generierung von Hypothesen, die experimentell überprüft werden können. Die **Molecular Biomechanics (MBM)** Gruppe ist stolz darauf, im Jahr 2023 ihre erste hauptsächlich experimentelle Studie zu veröffentlichen. Die meisten Ergebnisse wurden im eigenen Nasslabor unserer Gruppe generiert, gehen aber Hand in Hand mit einer zweiten größtenteils simulationsbasierten Studie. Unsere experimentelle Arbeit wurde auch im Jahr 2023 großzügig vom European Research Council (ERC) und dem Exzellenzcluster 3DMM2O unterstützt. Lesen Sie in diesem Kapitel mehr über diese und andere Aktivitäten des vergangenen Jahres.

# 2.8 Molecular and Cellular Modeling (MCM)

## Group leader

Prof. Dr. Rebecca Wade

## Team

Christina Athanasiou (until January 2023)
Riccardo Beccaria (since November 2023)
Sophia Ber (since July 2023)
Tommaso Bartoloni (since March 2023)
Mislav Brajkovic (since February 2023)
Elizaveta Chernova (July–September 2023)
Dr. Giulia D'Arrigo
Matheus Ferraz (visiting scientist; Federal University of Pernambuco, Recife, Brazil; DAAD scholarship) (until April 2023)
Manuel Glaser
Nico Herb (October–December 2023)
Melanie Käser (until January 2023)
Marcel Meyer (visiting scientist; Heidelberg University; until June 2023)
Emanuele Monaci (until January 2023)
Abraham Muñiz Chicharro
Jakob Niessner
Dr. Giulia Paiardi
Dr. Stefan Richter
Marco Rizzi (visiting scientist; University of Genoa, Italy; until July 2023)
Michel Tarnow (April–July 2023)
Jonathan Teuffel
Alexandros Tsengenes (until January 2023)
Congcong Xu

Molecular recognition, binding, and catalysis are fundamental processes of 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 during 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 2023. Following a general overview of what was new in the group last year, we describe our results in three research areas:
(i) the optimization of drug–polymer interactions for biopharmaceutical performance,
(ii) the exploration of protein-binding cavities, and
(iii) the prediction of protein–protein and peptide–protein interactions.

## What happened in the MCM Group in 2023?

One of our major achievements was the completion of the EU-supported Human Brain Project (HBP), which came to an end after ten years in 2023. The approach taken to the multiscale modeling of neuronal signaling cascades is described in a review [van Keulen, 2023] and our tools are available on the EBRAINS infrastructure (ebrains.eu). Our work on the development of molecular-level tools for brain simulation will continue in EBRAINS 2.0, which is an international project designed to support the further development of services for the EBRAINS (European Brain Research Infrastructures) collaborative research platform, whose aim is to advance neuroscience and brain health. During the final year of the HBP, Giulia D'Arrigo, Stefan Richter, and Rebecca Wade organized the second Human Brain Project Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades (HBPMolSim) as a hybrid event in June together with an open workshop on the results of the BRAVE HBP partner project (for details, see Section 5.1.4).

The MCM group participates in two of the SIMPLAIX projects (www.simplaix.org). The first SIMPLAIX international scientific symposium – chaired by Rebecca Wade – was held in May in Studio Villa Bosch (see Chapter 5.1.2). For details on this and other SIMPLAIX activities, see Chapter 7.

Amongst our visiting scientists Matheus Ferraz completed his stay as a DAAD sandwich PhD student and defended his doctoral thesis in the laboratory of Roberto Lins (Instituto Aggeu Magalhães, Oswaldo Cruz Foundation Recife, Brazil) in June. Rebecca Wade traveled to Brazil for the PhD defense and project discussions and also gave a lecture at the institute in Recife. Matheus worked on the prediction and design of protein–protein interactions using enhanced sampling molecular simulations and machine learning with a focus on Sars-CoV-2 spike-binding proteins (see, e.g., [Ferraz, 2023]). Marco Rizzi – a visiting medicinal chemist from Michelle Tonelli's group (University of Genoa, Italy) – completed his stay and returned to Genoa to synthesize some of the compounds he had designed as anti-parasitic agents for use against sleeping sickness during his time in Heidelberg. Marco's research built on our previous structure-based drug design studies in the EU-supported NMTrypI consortium (see, e.g., [Roussaki, 2023]).

Several group members completed their studies at Heidelberg University. In October, Christina Athanasiou successfully defended her doctoral thesis at Heidelberg University on the modeling of the mechanistic effects of neurotrophin modulators. Furthermore, Melanie Käser and Jakob Niessner completed their master's theses in molecular biotechnology and physics, respectively. Michel Tarnow carried out his bachelor's thesis research in molecular biotechnology. Moreover, Elizaveta Chernova (biochemistry) and Nico Herb (molecular biosciences) completed internships in the group as part of their master's studies at Heidelberg University.
Mislav Brajkovic and Tommaso Bartoloni joined the group as new PhD students to work on the HBP and Informatics4Life (see Chapter 7) projects, respectively.

## Optimizing Drug–Polymer Interactions for Biopharmaceutical Performance

The active ingredients in many types of medicine are poorly water-soluble and require suitable formulation in order to enable them to be orally administered and absorbed from the gastrointestinal tract. One approach to this problem is to add polymers to the active drug ingredient. The selection of the types of polymers to use in drug formulation is classically done via trial-and-error or via the application of classical empirical models. However, advances in molecular simulation methods and computer hardware mean that atomistic molecular dynamics simulations of systems of drug molecules and polymers in aqueous solution are becoming feasible and could provide a means for rationally choosing drug–polymer combinations based on their physicochemical behavior. To pilot this approach, we collaborated with Abhay Sangamwar and his colleagues at the Indian National Institute of Pharmaceutical Education and Research (NIPER SAS Nagar) – the foremost pharmaceutical research institute in India – in order to employ molecular simulations and experiments to investigate amorphous drug–polymer–salt solid dispersions of the widely used anti-inflammatory

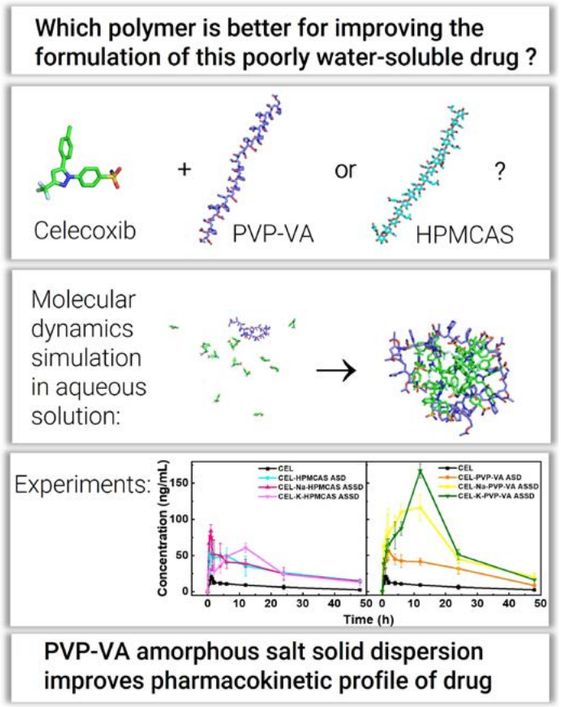



*Figure 35: Illustration of the concerted use of both molecular dynamics simulations of drug–excipient–polymer mixtures in aqueous solution and experiments to identify optimal drug–polymer combinations for formulating the drug celecoxib as an amorphous salt solid dispersion. Figure derived from [Mukesh, 2023].*

celecoxib. Classical empirical models did not correctly rank the two polymers polyvinylpyrrolidone vinyl acetate (PVP-VA) and hydroxypropyl methylcellulose acetate succinate (HPMCAS) for use as excipients with celecoxib. We therefore carried out molecular dynamics simulations and found that the computed strength of drug–polymer interactions correlates with the dissolution and pharmacokinetic profiles of amorphous drug–polymer–salt systems (see Figure 35). Moreover, the celecoxib–salt–PVP-VA formulation was stable and highly bioavailable and in addition demonstrated excellent biopharmaceutical performance [Mukesh, 2023]. This case study points the way to the molecular dynamics simulation-driven selection of polymer excipients for formulating a wide range of poorly water-soluble drugs that have the potential both to require lower and less frequent doses and to have fewer side-effects. In ongoing work, we are currently applying this approach in order to investigate the interaction of a drug with a mixture of polymers.

The study was carried out with bilateral German–Indian funding at Heidelberg University and HITS (German Federal Ministry for Education and Research; BMBF) as well as in India (Department of Biotechnology, New Delhi).

## Exploring Protein Binding Cavities

*For drug design:* A common strategy in structure-based drug design is to identify small molecules that bind in cavities in proteins (see, e.g., the multi-team large-scale virtual screening campaign to find low-molecular-weight inhibitors of SARS-CoV-2 virus proteins in Schimunek, 2023). Therefore, an important step is to identify and characterize the structure, dynamics, and binding properties of such protein cavities. We developed the TRAPP toolbox (trapp.h-its.org) for this purpose. This year, we published an application of TRAPP to human ecto-5'-nucleotidase, which is a key enzyme in purinergic signaling pathways that is a potential target for several diseases, including cancer and inflammatory, infectious, and autoimmune diseases [Viviani, 2023]. Lucas Viviani – who had previously visited HITS from the group of Antonia T.-do Amaral (Institute of Chemistry, University of Sao Paulo, Brazil) – ran molecular dynamics simulations of this highly flexible protein in active and inactive states and used TRAPP to identify both transient subpockets in the substrate-binding site and conformations with higher druggability scores than the experimentally determined crystal structure (see Figure 36).

*For protein engineering:* Another application of in silico modeling and the simulation of protein cavities is protein engineering. We applied this approach in collaboration with Kaija Autio and Kalervo Hiltunen and colleagues (Faculty of Biochemistry and Molecular Medicine, University of Oulu, Finland) in a study that demonstrated how long-chain fatty acids are required to regulate cellular respiration [Rahman, 2023]. This is a complex process by which mitochondria energy factories inside cells) extract cellular energy from nutrition. The discovery of the importance of long-chain fatty acids that are produced by mitochondria in cellular respiration opens up new approaches to tackling diseases that involve impaired mitochondrial function and cellular respiration. In this study, mutants of the human 2E-enoyl-ACP reductase (MECR) enzyme – which is a key enzyme in the mitochondrial fatty acid synthesis pathway – were designed using computational molecular modeling along with structure determination via crystallography and other experiments in order to validate the predictions. Specifically, the substrate-binding cavity was shortened in order to prevent the synthesis of long-chain fatty acids (see Figure 37). Experiments confirmed the validity of the computational approach to engineer substrate length specificity, and the mutant enzymes provided the basis for fundamental mechanistic studies of mitochondrial function. The modeling was carried out by Joanna Panecka-Hofman both when she was a postdoc in the MCM group and subsequently at the University of Warsaw, Poland.

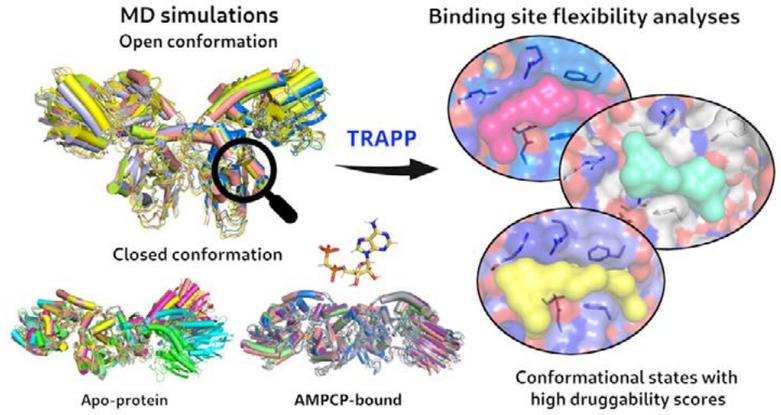



*Figure 36: Illustration of a TRAPP analysis of molecular dynamics (MD) simulations of human ecto-5'-nucleotidase, which revealed conformations that could be targeted in drug design. From [Viviani, 2023], reproduced with permission.*

*For tackling drug resistance:* One of the mechanisms by which drug resistance can arise is through proteins that transport drug molecules. The chloroquine resistance transporter (PfCRT) in the malaria-causing parasite Plasmodium falciparum expels quinoline drugs from the parasite's digestive compartment (vacuole), which is the site at which the drugs exert their antimalarial activity. In the 1950s, chloroquine (CQ) was the first drug that the malaria parasite became resistant to. Piperaquine (PPQ) was introduced as an alternative drug in areas where CQ resistance was widespread, but resistance emerged in the 1980s. Interestingly, this resistance occurred concomitantly with an increase in sensitivity to CQ. Changes in the sensitivity and resistance to these and other malaria drugs arise from mutations in PfCRT. In a collaboration with Michael Lanzer and colleagues (Center of Infectious Diseases, Parasitology, Heidelberg University Hospital), who performed detailed kinetic experiments, we carried out docking and molecular dynamics simulations in order to investigate how mutations in the binding cavity of PfCRT affect the transport of CQ and PPQ [Gomez, 2023]. PfCRT is a transmembrane protein with a cavity that is large enough to accommodate both CQ and PPQ. Based on our experiments and computations, we found that the CQ-resistant Southeast Asian Dd2 strain of Plasmodium falciparum could simultaneously accept both CQ and PPQ at distinct but allosterically interacting sites in the binding cavity (see Figure 38). Furthermore, PfCRT could be engineered to alter the relative resistance to different drugs or to transport both CQ and PPQ equally well. The study thus provides insights into how PfCRT has evolved under drug pressure and how it may evolve in the future.

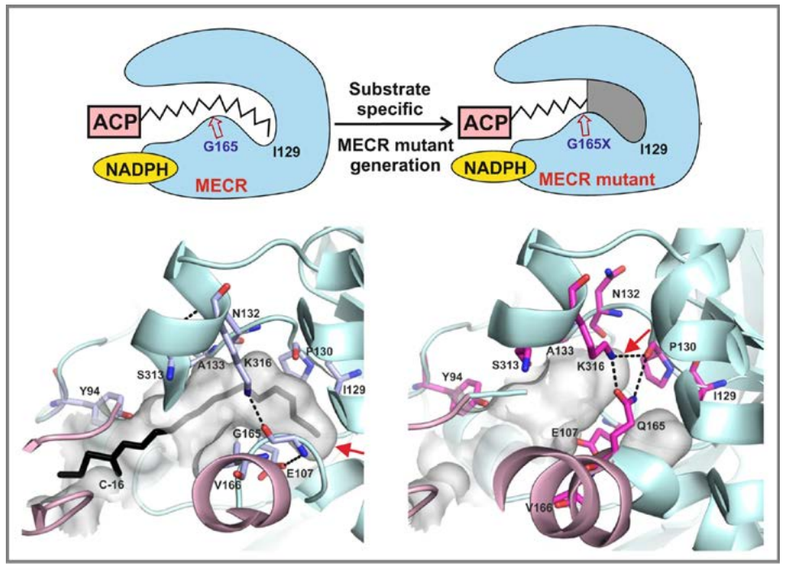



*Figure 37: Upper panel: Schematic diagram of the wild-type human 2E-enoyl-ACP reductase (MECR) enzyme and a designed mutant MECR enzyme. The wild-type enzyme can accommodate a long hydrocarbon chain with 16 carbons attached to acyl carrier protein (ACP). In the mutant, glycine 165 is mutated in order to introduce an amino acid sidechain that blocks the binding pocket so that it can only accommodate hydrocarbon chains up to 8 carbons in length. Lower panel: View of the fatty-acyl-binding cavity in crystal structures of (left) wild-type and (right) G165Q mutant MECR, which confirm the shortened cavity in the mutant (light gray surface, with the end of the cavity indicated by the red arrow). A substrate with 16 carbons is modeled into the wild-type structure. The mutation at residue 165 leads to readjustments of the positions of other amino acid residues that line the substrate cavity. Adapted from [Rahman, 2023], with permission.*

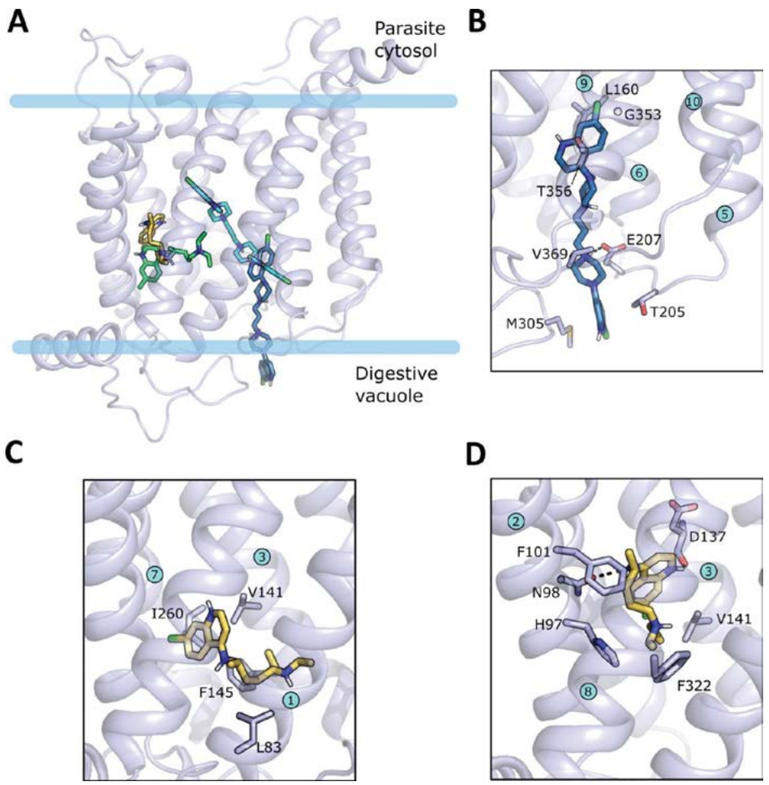



*Figure 38: Docking and molecular dynamics (MD) simulation of the chloroquine resistance transporter (PfCRTDd2) in a complex with two anti-malaria drugs, chloroquine (CQ) and piperaquine (PPQ). (A) The last frame from an MD simulation is shown superimposed on the original docking poses of CQ and PPQ in the protein cavity. CQ and PPQ occupy different regions of the binding cavity consistent with their non-competitive transport kinetics in PfCRTDd2. (B) Interactions established by PPQ during the last hundreds of ns. During the simulation, PPQ moves toward the digestive vacuole, where it interacts with the M305 residue of a highly flexible vacuolar loop on the transporter. In MD simulations, we found that the conformation of this loop is dependent on which drugs are present in the transporter. (C) Pose and interactions of CQ in the central part of the simulation (~400–600 ns). (D) Final pose and interactions established by CQ in the last hundreds of ns. CQ moves between favored positions on the left side of the channel. The docking and the MD final poses are shown in green and yellow, respectively, for CQ and in cyan and blue, respectively, for PPQ. The protein backbone and the amino acid residues that interact with ligands are shown in gray. From [Gomez, 2023], with permission.*

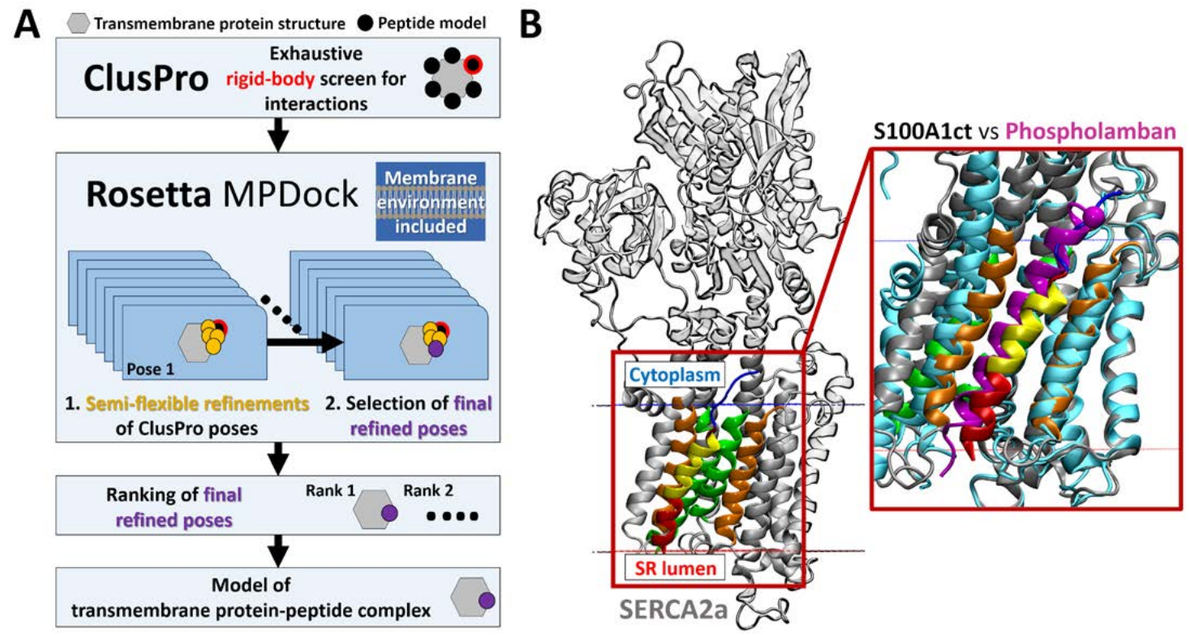



*Figure 39: Computational docking of the S100A1ct peptide to the transmembrane protein SERCA2a. Left: The computational docking pipeline that was developed in order to predict the structures of transmembrane protein–peptide complexes makes use of a combination of docking and structural refinement methods. Right: A predicted docking pose of the S100A1ct peptide to the SERCA2a protein. The S100A1ct peptide is colored blue, yellow, and red. SERCA2a is colored grey, and helices that interact with S100A1ct colored green and orange. The sarcoplasmic reticulum (SR) membrane is indicated by dashed lines. The inset compares the docking pose of S100A1ct with the experimentally determined binding pose of the transmembrane part of the natural inhibitor phospholamban (magenta) to SERCA1a (cyan). The overlap in the poses indicates that S100A1ct could interfere with the interactions of SERCA2a with natural peptide inhibitors and could thereby enhance SERCA2a activity.*

## Predicting Protein–Protein and Peptide–Protein Interactions

In contrast to small molecules that tend to bind in concave cavities on proteins, proteins and peptides tend to bind to much more extended, often convex surfaces on proteins. In addition, peptides can be highly flexible, even when bound to a target protein, which may be decorated with highly flexible carbohydrate chains. Their interactions thus pose different challenges to computational approaches. In several projects, we are currently developing and applying approaches to predicting structures of complexes, to computing their binding kinetics, to exploring their association and dissociation mechanisms (for a review, see [Muñiz-Chicharro, 2023]), and to designing therapeutic peptides. One of these projects is part of the Informatics4Life consortium, which is supported by the Klaus Tschira Foundation and aims to find a new therapeutic approach to cardiac disease (see [Glaser, 2023]). We are collaborating with Patrick Most and Julia Ritter (Heidelberg University Hospital) in order to investigate how the synthetic peptide S100A1ct – which is derived from the terminal region of the protein S100A1 – exerts its promising activity for the treatment of heart failure. These studies provide a basis for designing peptide variants and peptidomimetics with improved performance compared with S100A1ct. Experiments indicate that S100A1ct enhances the activity of the SERCA2a Ca2+-ATPase, which is a transmembrane protein in muscle cells that transports calcium, but the mechanism remains unknown. Our previous molecular dynamics simulations revealed that S100A1ct could adopt a predominantly helical structure, and bioinformatics tools suggested that it could form a transmembrane helix. As standard computational docking procedures do not account for the presence of a membrane, we developed and validated a computational pipeline for docking peptides to the membrane-exposed surfaces of transmembrane proteins (see Figure 39). By applying the pipeline, we identified five possible binding sites for the S100A1ct peptide on SERCA2a (see Figure 39). The docking results suggest that the SERCA2a activation effect of S100A1ct may arise from perturbation of the binding of natural inhibitory helical peptides to SERCA2a.

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.

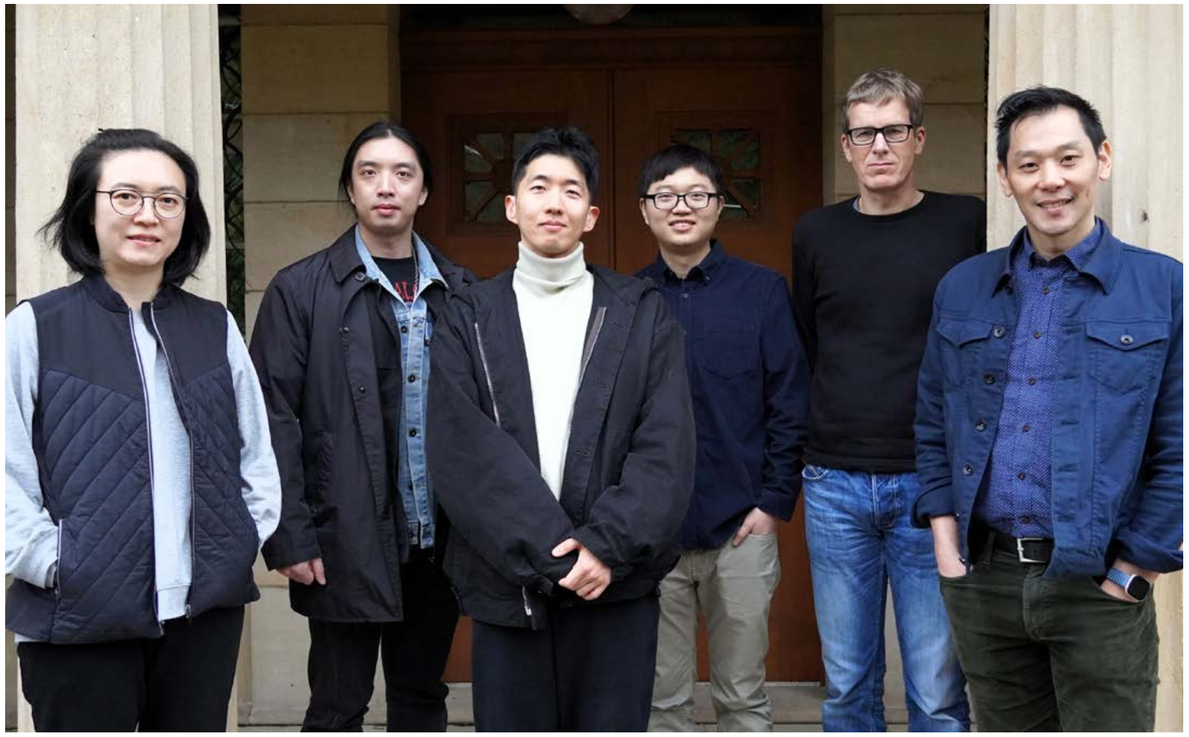
In der **Molecular and Cellular Modeling Gruppe (MCM)** sind wir in erster Linie daran interessiert zu verstehen, wie Moleküle interagieren. Was bestimmt die spezifische und selektive Wirkung beim Zusammenspiel von Wirkstoff und Rezeptor? Wie werden Proteinkomplexe gebildet und welche Formen können sie 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 Forscher*innen 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 2023. Nach einem allgemeinen Überblick über Neuigkeiten in der Gruppe konzentriert sich der Bericht auf Projekte zu (i) Optimierung der Wirkstoff-Polymer Wechselwirkungen für die biopharmazeutische Performance, (ii) Erkundung von Protein-Bindungstaschen, und (iii) Vorhersage von Protein-Protein und Peptid-Protein Wechselwirkungen.

# 2 Research

# 2.9 Natural Language Processing (NLP)



## Group leader

Prof. Dr. Michael Strube

## Team

Haixia Chai (HITS Scholarship holder)
Yi Fan
Mehwish Fatima (visiting scientist; HEC-DAAD Scholarship; until March 2023)
Sungho Jeon (HITS Scholarship holder)
Wei Liu (HITS Scholarship holder)
Xianghe Ma (student)
Prof. Dr. Shimei Pan (visiting scientist; Fulbright Award; until July 2023)
Dr. Stephen Wan (visiting scientist; CSIRO Award; since September 2023)
Dr. 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 proudly hosted Shimei Pan as a guest scientist until July 2023. Shimei is a professor in the Information Systems program at the University of Maryland in Baltimore, USA. In recent years, her research has focused on biases in both 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 also co-taught a seminar at the Institute for Computational Linguistics at Heidelberg University together with NLP group leader Michael Strube, and she regularly interacted with researchers at the university.



Soon after Shimei left, the NLP group welcomed Stephen Wan, a team leader and research scientist at CSIRO in Sydney, Australia. His research centers around information extraction from scientific documents, which aligns nicely with the NLP group's research focus on discourse processing. Stephen worked together with members of the NLP group – in particular, Wei Liu – on discourse relation classification.

Scientifically, 2023 turned out to be a very successful year for the NLP group, with five publications at *ACL conferences (4 main track, 1 findings) and several other publications at smaller conferences and workshops. Wei Liu and Yi Fan participated in the "DISRPT 2023 Shared Task on Elementary Discourse Unit Segmentation, Connective Detection, and Relation Classification." They did an excellent job and won two out of three subtasks. Congratulations!

In March 2023, Mehwish Fatima submitted her PhD thesis and left HITS to take up a job as a content architect in a company that specializes in online education.

Wei Zhao continued the highly successful collaboration between the NLP and GRG groups within the framework of the HITS Lab project on geometric deep learning. The project concluded as planned at the end of 2023. Wei Zhao then left HITS and took up a faculty position in the Computing Science department at the University of Aberdeen, Scotland.

Michael Strube was co-chair of the "Fourth Workshop on Computational Approaches to Discourse," which took place at ACL 2023 in Toronto, Canada. For the first time, the workshop felt like a real one after all three previous iterations had been hampered by the COVID-19 pandemic. The next iteration of the workshop is planned to be held at EACL 2024. Michael was also busy the rest of the year after being appointed General Chair of EACL 2024, which is the European flagship conference on computational linguistics and natural language processing and will be held in Malta in March 2024.

## Modeling Structural Similarities Between Documents for Coherence Assessment with Graph Convolutional Networks

**Wei Liu**

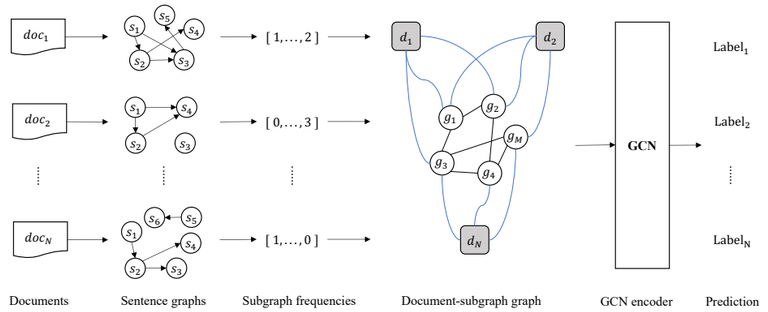
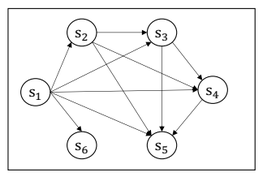



*Figure 41: Overview of the proposed approach. Our method identified the document's graph structure, converted the graph into a subgraph set, constructed a corpus-level graph based on the shared subgraphs between structurally similar documents, and finally, encoded these connections with a GCN.*

Coherence is an important aspect of text quality, and its modeling is beneficial for many downstream tasks, such as summarization and document-level text generation. Given the importance of the task, many methods have been proposed for performing it.
However, existing methods focus solely on extracting features within a single document while ignoring the underlying correlations between documents.
Coherence describes how sentences of a text connect to one another. Theoretically, documents with similar connection structures should tend to have similar degrees of coherence, which can be useful prior knowledge for coherence modeling. Figure 40 shows two examples with similar connection structures.
In order to fill this gap, we investigated a graph-based approach to modeling the correlation between documents from the perspective of structural similarity. Specifically, we first represented each document as a sentence graph and then further converted it into a subgraph set. Next, we constructed a heterogeneous graph over the whole corpus that connected documents through subgraphs that were shared between documents. Finally, a graph convolutional network was applied to the graph in order for it to be able to learn representations of documents while also considering the connections between them.
We evaluated our method on two corpora: GCDC and TOEFL. Experimental results revealed that our model outperformed various baselines and achieved a new state of the art on two benchmarks [Liu et al., 2023a].

1. The **Internet** is changing **Africa**.
2. In **South Africa**, people can look for jobs online without leaving home.
3. Movies from **Nigeria** can easily spread around the world.
4. Playing music on mobile phones is becoming popular in **Senegal**.
5. Farmers in **Tanzania** can learn how to grow vegetables from videos.
6. These results show the power of the **Internet**.




1. Different **exercise** have different benefits for the **body**.
2. **Jogging** can increase your breathing and heart rate.
3. **Table tennis** keeps you away from shortsightedness.
4. Playing **basketball** can strengthen your muscles.
5. **Yoga** helps to relieve your back pain.
6. So, pick the one your **body** needs the most.

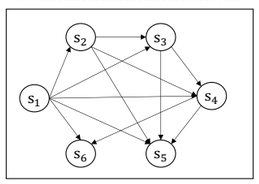



*Figure 40: Two example texts with similar connection structures, both of which are coherent.*

## Geometric Deep Learning

**Wei Zhao**

Real-world graph data – such as social networks – often feature complicated structures with diverse topologies, including grids, trees, and combinations of thereof. According to the manifold hypothesis, the quality of graph embeddings depends primarily on how the inductive biases (e.g., the symmetry and curvature) of a geometric space match the underlying graph structures. It has been shown that neither Euclidean nor hyperbolic geometry has the ability to embed complex graph structures with low distortion due to the inherent geometric limitations of these structures. Since 2021, a collaboration between the GRG and the NLP groups within the HITS Lab Project Geometry and Representation Learning has marked a breakthrough in this regard. The collaboration introduced a class of advanced mathematical spaces known as high-rank, non-compact symmetric spaces. Mathematically speaking, these spaces offer the ability to accommodate compound geometry with flat and negatively curved subspaces, thereby making the spaces particularly suitable for modeling complex graphs. Based on this momentum, the HITS Lab project continued to explore this research area with the following two subprojects:
For the first subproject [Zhao et al, 2023b], we extended the ideas conceived earlier in the HITS Lab project in order to develop graph neural networks in the space of symmetric positive definite (SPD) matrices, which are computationally tractable spaces within the broader family of symmetric spaces (see Figure 42). However, the challenge in doing so arises from the fact that the building blocks of graph neural networks are exclusively designed for Euclidean spaces, thereby rendering them impossible to operate in SPD. We addressed this issue by using gyrocalculus to reconfigure these building blocks via feature transformation and graph propagation in SPD. We released a library that features five popular graph neural networks in SPD. Through extensive experiments, graph neural networks in SPD were found to demonstrate substantial improvements over Euclidean and hyperbolic spaces on complex graphs; however, these improvements came at the cost of doubling (or quadrupling) the training time of graph neural networks compared with hyperbolic space (or Euclidean space). As such, the computational challenge limits the practical applicability of our approach to large-scale graphs, thereby leading to the initialization of a second subproject that focuses on improving efficiency.

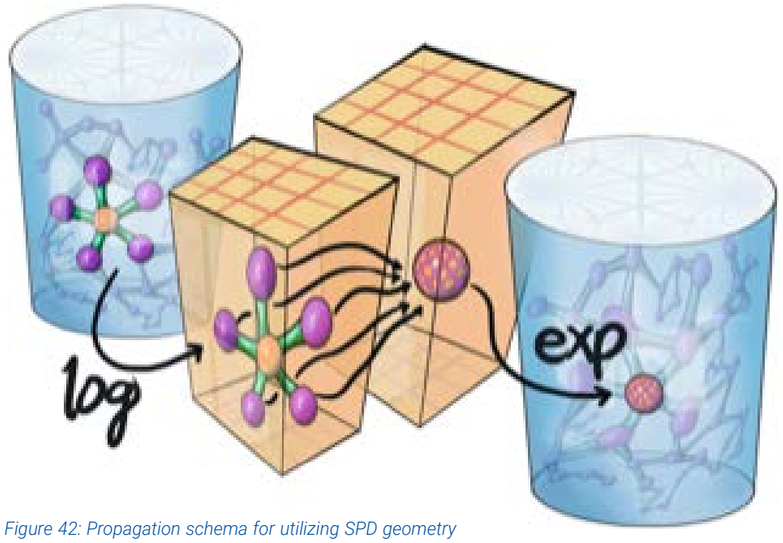



*Figure 42: Propagation schema for utilizing SPD geometry while performing calculations in the tangent (Euclidean) space: Beginning with an SPD embedding, map a node and its neighbors to the tangent space via the logarithm, and perform a modified Euclidean aggregation before returning to SPD via the Riemannian exponential map.*

For the second subproject, we searched for geometric spaces that not only are efficient to compute, but also provide suitable inductive biases that are needed to produce high-quality graph embeddings. In discrete geometry, theoretical results suggest that normed spaces can abstractly embed finite metric spaces with surprisingly low theoretical bounds on distortion in low dimensions. Inspired by this suggestion, we chose normed space as a flexible and computationally efficient alternative to SPD and other popular Riemannian spaces for graph embeddings (see Figure 43). Our empirical results demonstrate that normed space embeddings outperform other spaces on many synthetic and real-world graph data, with computational resource requirements growing much slower as graph size increases. Beyond graph embeddings, this work also provides valuable insights into experimental mathematics in the field of finite metric space embeddings.

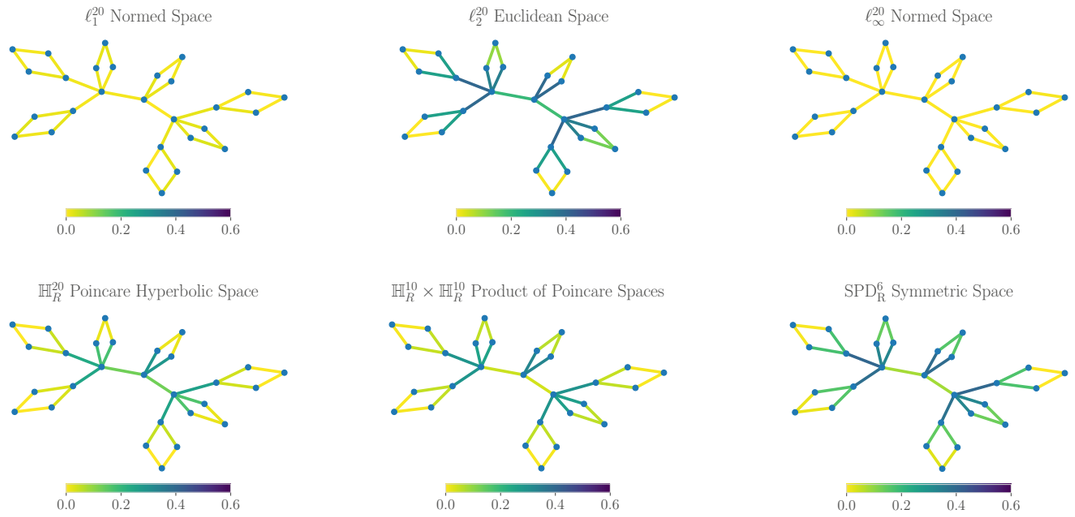



*Figure 43: Embedding distortion across spaces on a small synthetic graph, with color indicating distortion levels (i.e., the absolute difference between graph edge and norm distances). The graph embeds well in the ℓ1 and ℓ∞ normed spaces but becomes distorted in other spaces.*

**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 beherbergte bis Juli 2023 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 bei HITS wurde durch einen US Fulbright Award ermöglicht. Sie arbeitete aber nicht nur mit Wissenschaftler*innen am HITS, sondern auch am Institut für Computerlinguistik der Universität Heidelberg. Dort unterrichtete sie auch zusammen mit Michael Strube ein Seminar über Ethik in NLP.

Schon bald, nachdem uns Shimei verließ, begrüßte die NLP Gruppe einen weiteren Gastwissenschaftler, Stephen Wan, Team Leader und Research Scientist bei CSIRO, Sydney, Australien. Sein Forschungsinteresse ist Informationsextraktion aus wissenschaftlichen Dokumenten, was sich gut mit den Forschungsinteressen der NLP Gruppe im Bereich Diskursverarbeitung ergänzt. Stephen arbeitet eng mit der NLP Gruppe zusammen, insbesondere mit Wei Liu über die Klassifikation von Diskursrelationen.

Das Jahr 2023 war ein wissenschaftlich sehr erfolgreiches Jahr für die NLP Gruppe mit fünf Veröffentlichungen bei *ACL-Konferenzen (4 Main Track, 1 Findings) und weiteren Publikationen bei kleineren Konferenzen und Workshops. Wei Liu und Yi Fan nahmen teil an der „DISRPT 2023 Shared Task on Elementary Discourse Unit Segmentation, Connective Detection, and Relation Classification“. Sie leisteten exzellente Arbeit und gewannen zwei von drei Teilaufgaben. Herzlichen Glückwunsch!

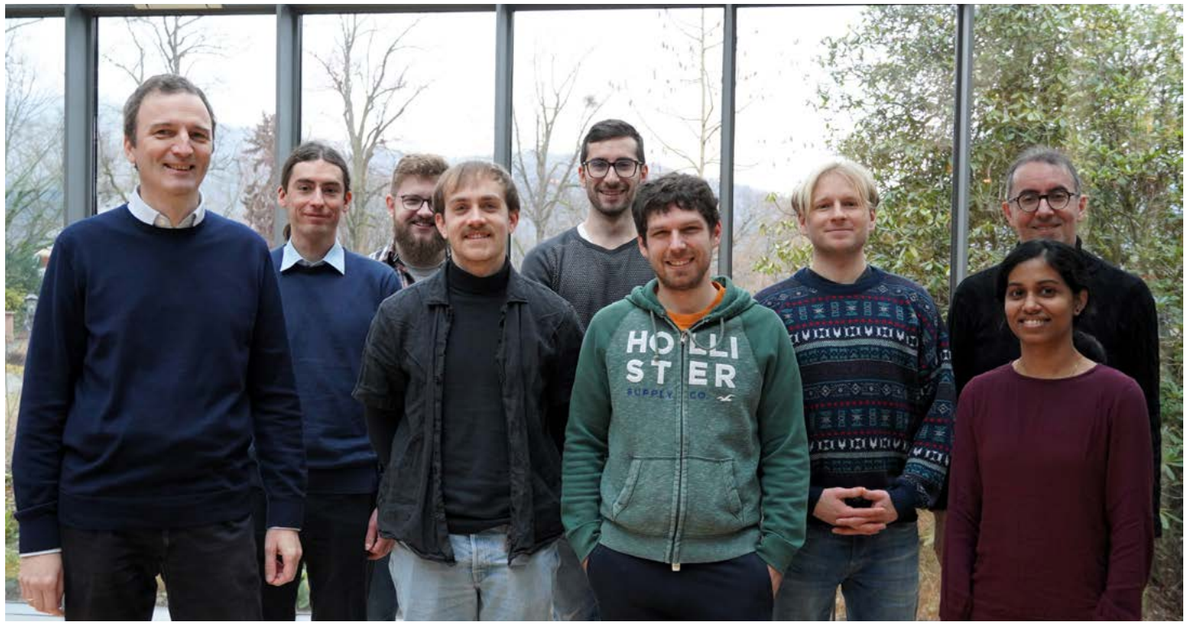
Mehwish Fatima reichte im März 2023 ihre Dissertation ein. Sie verließ HITS und nahm in der Industrie eine Stelle als Content Architect im Bereich Online Education an.

Wei Zhao brachte Ende 2023 das höchst erfolgreiche HITS Lab Projekt „Geometry and Representation Learning” (zwischen der GRG und der NLP Gruppe) zum Abschluss. Wei verließ HITS und trat eine Position als Hochschullehrer im Computer Science Department an der University of Aberdeen an.

Michael Strube war Programm Co-Chair des „Fourth Workshop on Computational Approaches to Discourse", der im Rahmen der ACL 2023 in Toronto, stattfand. Dies war das erste Mal, dass er sich wie ein richtiger Workshop anfühlte, da die letzten Ausgaben durch die Covid Pandemie behindert wurden. Der fünfte Workshop der Reihe ist schon in der Vorbereitung und wird im Rahmen der EACL 2024 stattfinden. Michael wurde für den Rest des Jahres auf Trab gehalten, nachdem er zum General Chair der EACL 2024 berufen wurde, der wichtigsten europäischen Konferenz im Bereich Computerlinguistik und automatische Sprachverarbeitung, die im März 2024 in Malta stattfinden wird.

# 2 Research

# 2.10 Physics of Stellar Objects (PSO)



## Group leader

Prof. Dr. Friedrich Röpke

## Team

Dr. Róbert Andrássy
Ferdinand Berwig (until January 2023)
Paul Christians (visiting scientist; TU Darmstadt, since November 2023)
Prof. Dr. Robert Fisher (visiting scientist; University of Massachusetts, Dartmouth, since October 2023)
Javier Morán Fraile
Alexander Holas (visiting scientist; Heidelberg University)
Dr. Alexandra Kozyreva (March-June 2023)
Dr. Mike Lau (Croucher fellow, since September 2023)
Giovanni Leidi
Kiril Maltsev
Vijayalakshmi Vijayakumaran Nair (since October 2023)
Marco Vetter

"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.

## Simulating Mergers of White Dwarfs with Neutron Stars

Compact objects such as black holes, neutron stars, and white dwarfs represent the fascinating end stages of stellar evolution. These objects have masses comparable with that of our Sun but are compacted into sizes in the order of kilometers, thereby making them the perfect laboratories for extreme physics. When two of these stellar remnants interact, the immense amount of energy that they contain can give rise to some of the brightest observable events in the Universe, such as Type Ia supernovae (which are most likely caused by the interaction of two white dwarf stars) and kilonovae (which are produced as a result of mergers between neutron stars).

However, these events are highly challenging to model. In order to accurately represent them, a wide range of physical phenomena need to be taken into account, including nuclear reactions, magnetic fields, and fluid dynamics. This situation makes three-dimensional (3D) magnetohydrodynamics (MHD) simulations the best tool for studying such events. Mergers between white dwarfs have been studied in depth because these stars are promising candidates for explaining Type Ia supernovae. In fact, a special case was recently proposed by the PSO group and is discussed further below. Black holes and neutron stars have also received a great deal of attention from the scientific community because their mergers result in gravitational wave signals, the detection of which was awarded with the 2017 Nobel Prize in physics.

However, one type of merger between compact objects has received less attention thus far: namely the merger between neutron stars and white dwarfs. The reason for this neglect is the extreme challenge that such mergers present due to the exceptionally large disparity in the scales of the objects: Black holes and neutron stars have relatively similar sizes, with neutron stars having radii of a couple tens of kilometers and stellar-scale black holes having similarly sized Schwarzschild radii, whereas white dwarfs can have radii larger than that of the Earth.

Simulating the merger of stars with such substantial differences in radii has historically been out of reach for 3D MHD simulations because the poor spatial resolution reached with conventional numerical schemes and the insufficient computational resources available have long prohibited us from capturing the relevant physical processes.

However, thanks to advances in numerical methods and computational power, the PSO group was able to perform the first 3D MHD simulation of the merger between a neutron star and a white dwarf.

This simulation (cf. [Moran-Fraile et al. 2023a]) revealed how such events lead to the tidal disruption of the white dwarf. The white dwarf's debris forms a disk around the neutron star that is expected to be slowly accreted. Magnetic fields that are initially present in the white dwarf are amplified during this process and reach values of up to $10^{13}$ G. Material from the disk that gets too close to the neutron star is launched perpendicular to the orbital

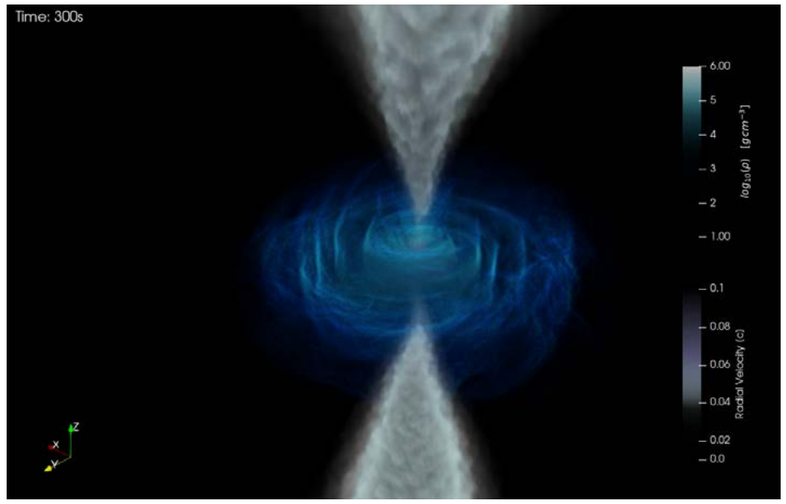



*Figure 44: Rendering of the disk of material from a disrupted white dwarf star (blue) around a neutron star. The magnetically driven bipolar jet is shown in white.*

plane by these immensely strong magnetic fields and forms spectacular bipolar jets in which gas is accelerated to velocities of up 10% the speed of light (see Figure 44, previous page).

It is not yet clear what such an event would look like if observed from Earth, but it is estimated to resemble so-called "fast blue optical transients" (FBOTs) with luminosities greater than those of kilonovae. The neutron star could also potentially collapse into a black hole as a result of accreting enough mass from the disk, which would make the jets reach speeds much closer to the speed of light and cause them to resemble long or ultra-long gamma-ray bursts.

## Thermonuclear Type II Supernovae

The classical distinction between Type I supernovae – which do not show hydrogen features in their spectra – and Type II supernovae – which show clear signs of hydrogen features in their spectra – was drawn as early as in the 1940s. This purely observational classification was interpreted with two distinct theoretical models: The collapse of the core of a massive star inside its hydrogen-rich envelope served as an explanation for Type II supernovae, and the thermonuclear explosion of a white dwarf star consisting of carbon and oxygen served as an explanation for Type I supernovae. Later, sub-classifications were introduced, but the general picture of these two types of explosions was largely confirmed by more detailed observations and theoretical models.

Any proposed model of Type I supernovae has two challenges: First, it has to explain the absence of the most common element in the Universe – that is, hydrogen – in the spectra of the supernova explosion, and second, it has to explain why an inert object such as a white dwarf star would initiate an explosion. In principle, the white dwarf star is eternally stabilized by the Fermi pressure of a degenerate electron gas. One possibility that is discussed in the literature is that of two white dwarfs merging violently. Because these white dwarfs are the remnants of the cores of low-mass stars at the end of their evolution (i.e., after the envelope has been lost), little or no hydrogen is expected to be present in this scenario. Several possibilities have been considered as causes of the detonation that can occur as a result of the violent interaction of the two stars. One of these possibilities is known as the "double detonation mechanism," in which a thin helium shell around the carbon–oxygen core is ignited due to the interaction and burns in a detonation. This detonation then drives a shock wave into the core and sets off a secondary detonation in the carbon–oxygen material.

The PSO group recently explored a complementary scenario [Moran-Fraile et al. 2023c]: namely the merger of a carbon–oxygen white dwarf with the core of a red giant star that resembles a helium white dwarf. In the new scenario, the merger is initiated by a common-envelope interaction, in which two stellar cores orbit each other inside a shared stellar envelope and transfer their angular momentum and orbital energy onto the envelope due to gravitational drag. In many cases, the energy transfer is thought to expel the envelope while the orbit of the cores shrinks. This process would leave behind a tight binary system consisting of two stellar cores. However, it is also possible for the envelope to not be successfully ejected, in which case, the two cores would merge inside it. This is the scenario that the PSO group simulated. Compared with the parameters usually studied in the context of thermonuclear supernova explosions, the considered carbon–oxygen white dwarf had a low mass (i.e., 0.6 solar masses). The helium white dwarf was disrupted in the last stages of its evolution and engulfed the carbon–oxygen white dwarf in an unusually massive helium shell. This shell was subjected to dynamic interaction and was highly non-homogeneous. Surprisingly, the conditions reached in the simulation revealed that even with these extreme system parameters, a double detonation leads to a thermonuclear explosion. Snapshots from the corresponding simulation of the PSO group are shown in Figure 45.

The low mass of the exploding carbon–oxygen star implies that the explosion is far less bright that that of a normal Type Ia supernova. The detonation of the thick helium shell produces relatively large amounts of calcium; therefore, the event may resemble a class of astronomical events known as "calcium-rich transients."

The special feature of the scenario is that the thermonuclear explosion occurs inside an extended, hydrogen-rich stellar envelope. Therefore, in the early stages, the emission that is observable from the event should result from processes that are similar to those found in Type II supernovae. This hypothesis was tested in radiative transfer simulations that followed up on the explosion simulations carried out by the PSO group [Kozyreva et al.: Thermonuclear explosions as Type II supernovae, Astronomy & Astrophysics, Vol. 684, id.A97, 12 pp., 2024]. Indeed, the interaction of the ejecta of the thermonuclear explosion with the stellar envelope was found to transfer part of the kinetic energy of these ejecta into emitted light, and the shape of the resulting light curve (i.e., the measured luminosity of the event as a function of time) was found to have a pronounced plateau similar to that observed in Type IIP supernovae. In fact, the event simulated by the PSO group would be a Type II supernova powered by a thermonuclear explosion rather than by the commonplace gravitational collapse of the core of a massive star into a neutron star.

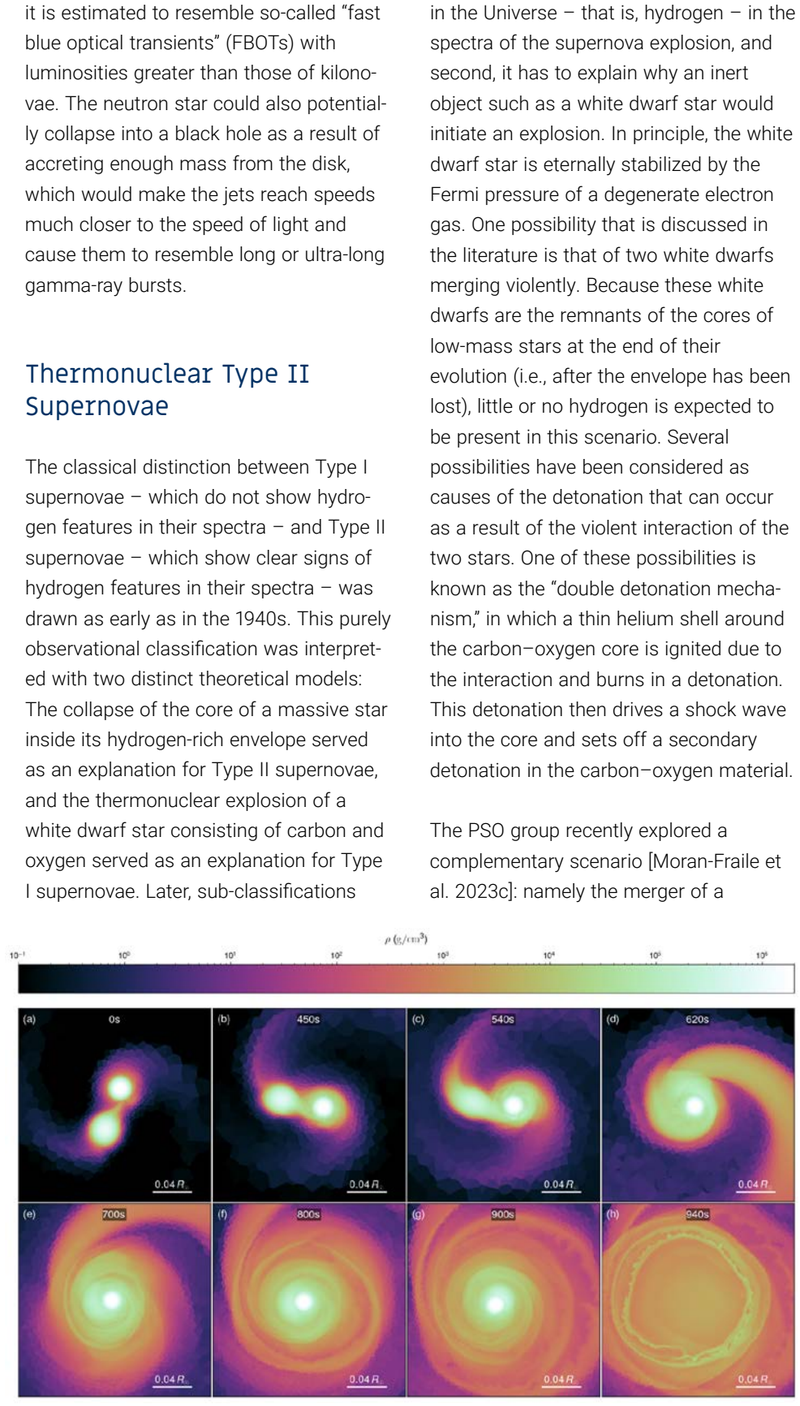



*Figure 45: Density slices through the orbital plane showing the inspiral of the two stellar cores (a–c), the disruption of the helium core (d–f), and the outcome of the double detonation (g).*

„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

# 2.11 Scientific Databases and Visualization (SDBV)

## Group leader

PD Dr. Wolfgang Müller

## Team

Dr. Haitham Abaza
Dr. Alain Becam (until October 2023)
Dr. Ina Biermayer
Dr. Susan Eckerle
René Geci (student; until June 2023)
Dr. Sucheta Ghosh
Martin Golebiewski
Xiaoming Hu
Jan Koß (student; until August 2023)
Dr. Olga Krebs
Gerhard Mayer
Lukrécia Mertová
Ghadeer Mobasher
Dr. Maja Rey
Maria Paula Schröder (student; until July 2023)
Fabian Springer (student; until August 2023)
Dr. Andreas Weidemann
Dr. Ulrike Wittig
Yueyang Xie (student; until July 2023)

The mission of the SDBV group is to provide FAIR (findable, accessible, interoperable, reusable) data to scientists. To that end, we develop, run, and maintain SABIO-RK as a professionally curated resource for FAIR reaction kinetics data. Moreover, we continue to contribute to the development of FAIRDOM SEEK, a system that enables projects to manage and "FAIRify" their data. These systems – as well as our standardization activities – are well integrated into national and European data infrastructure networks.
2023 was marked by a group evaluation in July. As during our last evaluation, we found the preparation for the evaluation and the subsequent presentation by the whole group stressful, fun, and stimulating. We thank the reviewer board for their vivid interest in – and thoughtful advice on – our work.

The year was also full of structural changes. The HITS de.NBI activities began to be funded through the MWK Baden Württemberg via Heidelberg University, as well as by the BMBF via the FZ Jülich. The funds are subject to evaluation but are potentially unlimited in terms of time. We wish to thank our colleagues from Baden-Württemberg (in particular Ursula Kummer, Heidelberg University) and Germany as well as the funders for going new ways in providing funding for science infrastructure.

Moreover, LiSyM-Cancer had its mid-term evaluation, and we were asked to submit a proposal in 2024 for the next phase for our project that comprises the program directorate and the data management of the network.

NFDI4Health – the National Research Data Infrastructure for Personal Health Data – also submitted its mid-term report. The SDBV group plays a key role both in leading standardization efforts and in building a metadata schema for studies in health research.

In the following we will describe our work on metadata, and we describe NFDI4Health Local Data Hubs as metadata hubs and connectors within the NFDI4Health infrastructure.

## The European Roadmap of Virtual Human Twins

The SDBV group takes part in the EU EDITH project (https://www.edith-csa.eu/), which defines a roadmap of a European Ecosystem for Digital Twins in Health Applications. Virtual human twins (VHTs) are digital representations of human physiology and pathology that can be used in medical research to accelerate the development of new medicines and medical devices and to enable personalized, patient-centered medicine.
VHTs include artificial intelligence and in silico (i.e., computing-based) technologies for diagnoses, medical decision support, treatment development, and intervention planning. EDITH aims to prototype computing, data, and research infrastructure and comprises technical as well as ethical, legal, and cultural elements that are designed to safeguard patient rights and privacy.
Within EDITH, the SDBV group is responsible for defining a standardization framework for VHTs as well as for collecting standards, terminologies, and metadata guidelines that are relevant to the development of the VHTs. We are involved in defining and further developing standards in standard defining organizations (SDOs) – such as the ISO and the IEC – as well as in relevant initiatives of communities such as the Computational Modeling in Biology Network (COMBINE) and the Global Alliance for Genomics and Health (GA4GH). These standards define and regulate the creation, semantic annotation, execution, and validation of models of the VHT as well as the data and metadata that are used to create or validate the models and the results of the simulations.

In 2023, we additionally defined and published the first draft of the VHT roadmap (https://zenodo.org/records/8200955), which has already been adopted by the European Commission as the masterplan for building a VHT infrastructure in Europe. This accomplishment was complemented by a more comprehensive description of the standards for formats, terminologies, and metadata guidelines as well as for modeling processes, model quality and validation, data provenance, and interoperability. The roadmap was published in January 2024 (https://zenodo.org/records/10492796, https://fairsharing.org/4787). To provide hands-on guidance on standardization, we also defined an implementation guide (https://zenodo.org/records/10524795). These activities are supported by our ongoing standardization activities in SDOs, e.g. in ISO 23494 (see [Wittner et al., 2023]) and scientific communities (see [König et al., 2023]).
Last year, we also coordinated the drafting of the international ISO standard ISO/TS 9491-1 (see [Kirschner, Golebiewski et al., 2023]), which is a guideline for modeling in the health research domain, including in the field of VHTs. The European Commission has already adopted ISO/TS 9491-1 as a prerequisite for compliance with proposals that address European calls for research projects on virtual human twins.

## The National Research Data Infrastructure for Personal Health Data (NFDI4Health)

The National Research Data Infrastructure (NFDI) for Personal Health Data (NFDI4Health) is a long-term project that is jointly funded by federal and state governments via the German Research Foundation (DFG). The initial phase of the project is set to run for 5 years, but the whole construction phase of NFDI4Health is planned as a long-term effort that will last at least 10 years. The goal of NFDI4Health is to support clinical and epidemiological researchers in sharing their data with the user community in a FAIR, privacy-preserving, and ethical manner. NFDI4Health creates new opportunities for data analyses that should improve population health. The main objectives are

1. to make structured health data from clinical and epidemiological studies, disease registries, administrative health databases, and health reporting in Germany FAIR;
2. to implement a framework for centralized search and access to existing decentralized infrastructures for health data;
3. to facilitate data sharing, record linkage, and harmonized data quality assessment;
4. to enable collaborative (federated) analyses of personal health data;
5. to enable the development and use of new machine-readable consent mechanisms and innovative data access services;
6. to support collaboration between clinical research, epidemiology, and public health; and
7. to promote the interoperability of currently fragmented IT solutions for metadata repositories, cohort searching, data quality, and harmonization.

NFDI4Health consists of 18 institutions sending more than 60 researchers in total and unites a broad range of expertise in

the field of research data management among both data-holding organizations and infrastructure specialists.

NFDI4Health established a data publication workflow in order to make health-related studies and their sensitive data centrally searchable (see [Fluck, Golebiewski, and Darms, 2023]). This workflow includes the publication of metadata that describe the studies as well as health datasets and their resources (study documents, instruments, data collections, etc.). To that end, the NFDI4Health metadata schema (MDS) for health studies was developed (see corresponding section below) in addition to data publication guidelines for health data and a suite of tools and services, including the German Central Health Study Hub (Health Study Hub) and the Local Data Hub, the latter of which builds on our SEEK software system (see corresponding section below).

We also engage directly with our user communities by raising awareness of and supporting capacity-building for research data management and NFDI4Health-specific services by organizing workshops and hands-on virtual or on-site training sessions and by conducting user research on the usability and functionality of our services. As part of a network of health data infrastructures in Europe and worldwide, NFDI4Health is well integrated into national and international activities, such as the emerging European Health Data Space (EHDS).

## The NFDI4Health Metadata Schema

The major contributions of the SDBV to NFDI4Health metadata standards are the program's tailored metadata schema for publishing personal health data (see [Fluck, Golebiewski, and Darms, 2023]) and the NFDI4Health FAIRsharing collection of interoperability standards (https://fairsharing.org/4940). Both tasks were accomplished in close cooperation with our user communities, international standardization initiatives, and committees of standard-defining organizations (SDOs), in which we are active members. Our developments are guided by international standards in the field in order to ensure international interoperability.
In an effort to make clinical, epidemiological, and public health research data FAIR, the NFDI4Health metadata schema was developed in order to support the standardized publication of high-level metadata and their content for health studies through two main services of NFDI4Health: the German Central Health Study Hub (GCHSH) and Local Data Hubs (LDHs; see below). This study metadata schema (MDS) is tailored to the publication of metadata from health studies and their respective study resources (e.g., study protocols, instruments, documents). The generic nature of the schema enables further types of resources to be registered, such as registries, secondary data sources, and study documents. The schema also extends to other health domains by adopting a modular structure that consists of core and domain-specific metadata items in generic and use-case-dedicated modules (see Figure 46). Most data items were primarily adapted from established standards and models, including DataCite, ClinicalTrials.gov, DRKS, Maelstrom, and MIABIS.
The schema's core module captures information that is commonly collected by any type of health resource, while further resource-type- and/or use-case-specific modules gather descriptions of certain types of resources or of health domains. Bibliographic information – such as the resource's title, description, and acronyms – is included in the core module along with information about contributors, identifiers, and relevant resources.

For design and data access information, the schema provides a design module that consists of characteristics that are specific to certain resource types. The module distinguishes between interventional and non-interventional study designs and provides descriptive information about studies and their populations. Information about data sharing is also included, which triggers the record linkage module when applicable.
The nutritional epidemiology module provides domain-specific information that is mainly related to the dietary assessment instruments that are applied in relevant studies. The chronic diseases module specifies whether prevalent or incident disease data were collected and indicates the sources from which the data were generated. The third dedicated module provides legal, consent, and budget information that is required for conducting record linkage. Modules that provide clinical trials' and imaging / radiomics' metadata are yet to be implemented.
The NFDI4Health metadata schema is currently available in human-readable spreadsheet format (https://doi.org/10.4126/FRL01-006472531). With the aim of a machine-readable and more relational version, it has also been represented on in the "art decor" repository (https://art-decor.org/ad/#/nfdhtfcov19-/project/overview). The interoperability standard "Fast Healthcare Interoperability Resources (FHIR)" developed by Health Level 7 (HL7) will play a key role in connecting the NFDI-4Health services with one another, including connecting the Health Study Hub with the Local Data Hubs and with external (inter)national infrastructures. Our collaboration partners at the Berlin Institute of Health (BIH) thus mapped the NFDI4Health MDS to the HL7 FHIR standard, translated the MDS into a FHIR logical model, and defined FHIR profiles – including extensions – to represent the components of the MDS (https://simplifier.net/nfdi4health-metadata-schema).

In order to ensure further interoperability with other systems, the schema was mapped to data structures of clinical trial registries, such as ClinicalTrials.gov, DRKS, and the World Health Organization's International Clinical Trials Registry Portal (ICTRP). Based on these mappings, studies are automatically uploaded from the portals to the GCHSH. Mappings to the schemas of the European Clinical Research Infrastructure Network (ECRIN) and the German Human Genome Phenome Archive (GHGA) are also in progress, as are discussions with the European Commission's Joint Research Center on interoperability with their European Platform on Rare Disease Registration (EU RD Platform). In order to facilitate the sharing of metadata by data-holding organizations, the schema is also being implemented by Local Data Hubs.

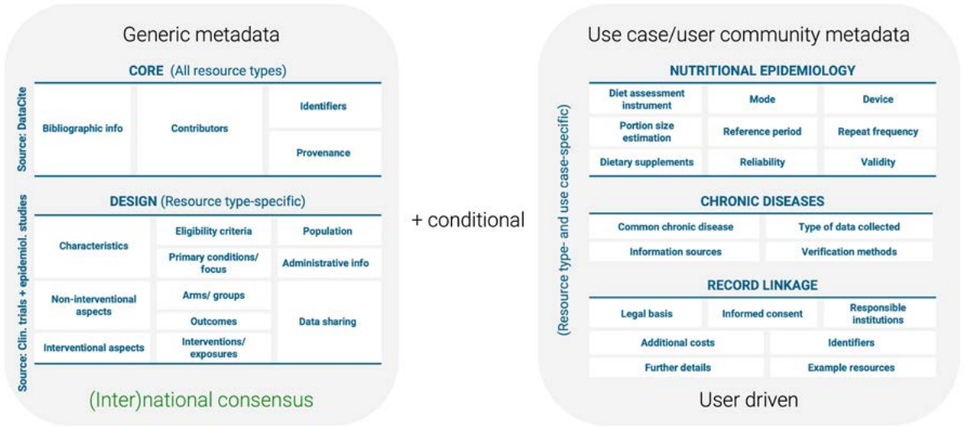



*Figure 46: The NFDI4Health metadata schema.*

## The NFDI4Health Local Data Hub (LDH)

The Local Data Hub (LDH) is the main local component and connector of the federated distributed services of NFDI-4Health (see [Meineke et al., 2023]). It enables local (meta-)data structuring, bundling, and sharing within data-holding organizations (DHOs) and beyond. The LDH's software operationalizes NFDI-4Health standards, including a tailored metadata schema (MDS) for health studies (see above), thereby facilitating communication with the German Central Health Study Hub. Integration into DHO processes and workflows is imperative. These LDHs target (bio)medicine, epidemiology, biostatistics, modeling, and bio- and medical informatics researchers, thereby providing standardized platforms for collaboratively sharing various data elements, such as projects, studies, publications, and software tools (see Figure 47).
By adhering to the FAIR principles (findable, accessible, interoperable, reusable), the LDH ensures that stored information is easily discoverable, accessible to authorized users, interoperable across various systems, and available for reuse in subsequent research endeavors.


*Figure 47: The Local Data Hub.*

The Local Data Hub within the NFDI-4Health ecosystem offers a comprehensive suite of features for enhancing research collaboration and data management. The hub systematically organizes information related to local research, thereby ensuring easy navigation and accessibility. It additionally creates meaningful links between domain-specific metadata and data from local studies, thereby facilitating a cohesive research environment. With fine-grained access rights management, the hub enables precise control over data access. The Local Data Hub supports programmatic access (API), thereby promoting interoperability and data exchange with analytical tools such as DataSHIELD and Personal Health Train. In addition, it also facilitates the sharing of public metadata by submitting data to the German Health Study Hub, which is another core service of NFDI4Health.

Due to its inherent FAIR principles, FAIRDOM SEEK emerged as the chosen platform for the Local Data Hub (LDH) in the NFDI4Health project for various compelling reasons. The platform's well-established user community spans diverse research areas in the life sciences. SEEK's compatibility with NFDI-4Health's requirements simplifies implementation, thereby ensuring a smooth and efficient integration process. The decision is reinforced by a proven track record in data management, as evidenced by the success of projects such as the Leipzig Health Atlas, FAIRDOMHub, and LiSyM SEEK. Furthermore, the development expertise at HITS – including the Institute's role in co-founding the FAIRDOM initiative –shows the deep understanding of SEEK's capabilities, thereby enhancing SEEK's effective use as the basis for the LDH. The LDH software platform has already been implemented at 3 different DHOs in Germany, including at one clinical study center, which demonstrates the usefulness of our development for users.

## Extended Metadata Feature

In response to the richness of the NFDI4Health metadata schema (MDS), SEEK incorporates an Extended Metadata feature that enables flexible user-defined metadata extensions. The SEEK-based Local Data Hub fully represents the comprehensive NFDI4Health metadata schema through this feature. The hub is also rolled out across the entire FAIRDOM community, thereby facilitating the description of complex metadata structures.

Extended Metadata supports the definition of additional metadata attributes for a particular type of data, thereby enabling adherence to specific standards (e.g., NFDI4Health MDS). While not directly visible to users, these attributes are associated with a particular attribute type, which is marked as optional or mandatory. The attribute's content is validated with respect to the type. [

Extended Metadata types are linked to specific resource types in SEEK, including Collection, DataFile, Document, Event, Model, Presentation, SOP, and Project.

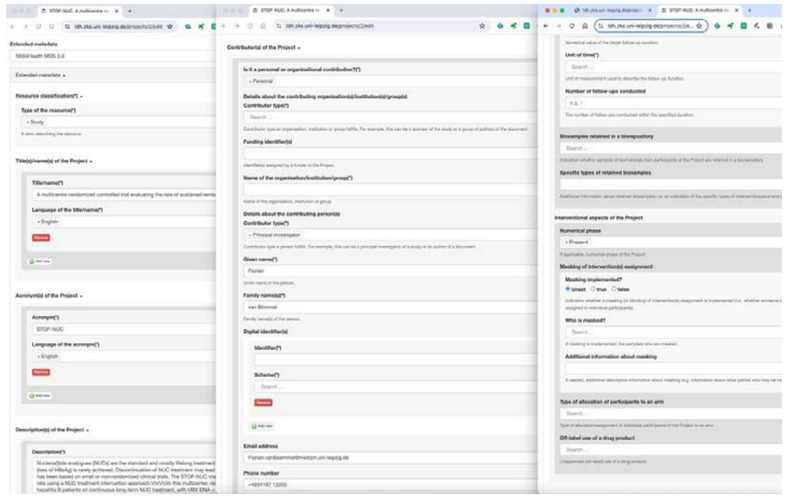


*Figure 49: Screenshot of MDS implementation using the Extended Metadata function.*

Extended Metadata can currently only be defined directly in the database, typically through a seed file, such as the NFDI-4Health MDS Extended Metadata Seed (https://github.com/nfdi4health/ldh/blob/main/db/seeds/018_MDS_3_3_Project_gen.seeds.rb).

Future plans include making it easier for administrators to define Extended Metadata through JSON, an Excel template, or a user interface.

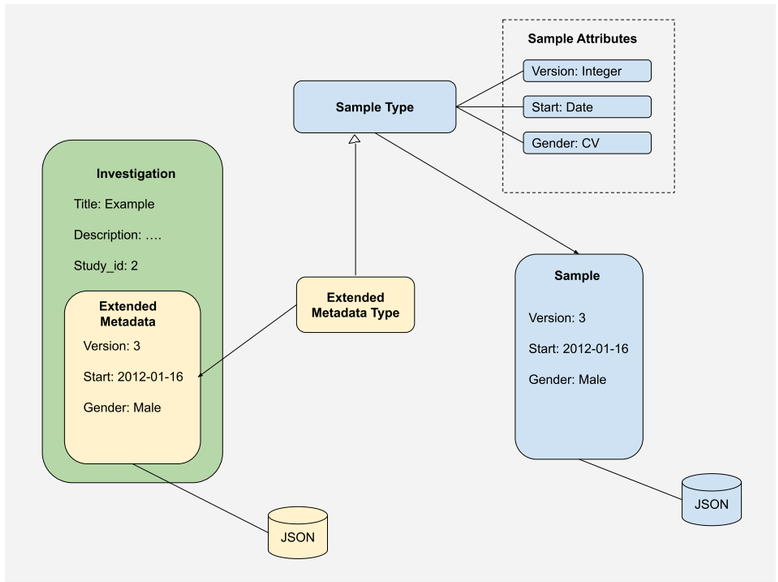



*Figure 48: Screenshot of the architecture of Extended Metadata.*

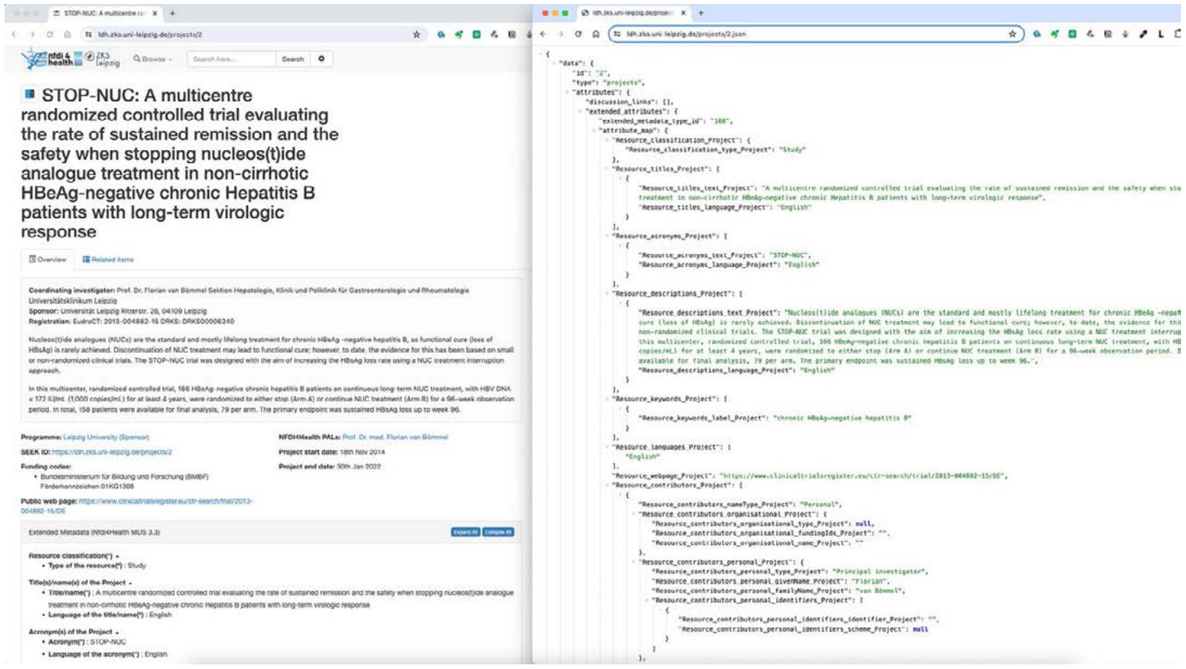



*Figure 50: Screenshot of MDS access in SEEK through the user interface or the JSON API.*

Die Mission der **Scientific Databases and Visualization Gruppe (SDBV)** ist die Bereitstellung „FAIRer" Daten für Wissenschaftler. Mit diesem Ziel entwickeln, betreiben und warten wir SABIO-RK, eine professionell kuratierte Datenressource für reaktionskinetische Daten. Weiterhin tragen wir weiter zu FAIRDOM SEEK bei, einem System, das Nutzern erlaubt, ihre Daten zu „FAIRifizieren". Diese Systeme - wie auch unsere Aktivitäten zur Standardisierung - sind in nationale und europäische Infrastrukturnetzwerke gut integriert. Forschung über Themen wie Datenextraktion und „Research Software Engineering" komplettieren das Bild.

Das Jahr 2023 stand im Zeichen der Evaluation unserer Gruppe im Juli. Wie bei der letzten Evaluation empfanden wir zwar die Vorbereitungen und die Präsentation als stressig, sie machten aber auch Spaß und waren für unsere Arbeit sehr stimulierend. Wir danken allen Reviewern für ihr Interesse an unserer Arbeit und ihren wohldurchdachten Rat.

Strukturelle Veränderungen waren ebenfalls maßgebend: Die Aktivitäten des HITS innerhalb von de.NBI werden seit 2023 durch das Ministerium für Wissenschaft und Kunst (MWK) Baden-Württemberg sowie durch das Bundesministerium für Bildung und Forschung (BMBF, via Forschungszentrum Jülich) finanziert. Diese Mittel werden jeweils abhängig von Evaluationen vergeben, aber sie sind potentiell zeitlich unbegrenzt. Wir möchten uns bei unseren Kolleg*innen in Baden-Württemberg (insbesondere Ursula Kummer, Universität Heidelberg), in Deutschland, sowie bei den Forschungsförderungsorganisationen bedanken, die hier neue Wege zur Förderung von Wissenschaftsinfrastruktur beschreiten.

Das Projekt „LiSyM-Cancer" durchlief mit Erfolg eine Zwischenevaluation. Wir wurden inzwischen darum gebeten, einen Antrag für die nächste Phase einzureichen. In LISyM Cancer hat das Programmdirektorat um Beat Müllhaupt (Universitätsspital Zürich, Schweiz) sowie das Datenmanagement seinen Sitz am HITS.

„NFDI4Health", die Nationale ForschungsDatenInfrastruktur für Gesundheitsdaten, hat ihren Zwischenbericht vorgelegt. Darin spielt die SDBV-Gruppe eine Schlüsselrolle bei der Standardisierung und der Erstellung des Metadatenschemas für Studien in der Gesundheitsforschung.

In diesem Kapitel beschreiben wir unsere Arbeit an Metadaten, und wir erläutern die Funktion von „NFDI4Health Local Data Hubs" als Verteiler von Metadaten und Verbinder innerhalb von NFDI4Health.

# 2 Research

# 2.12 Stellar Evolution Theory (SET)

### Group leader

Dr. Fabian Schneider (since January 2021)

### Team

Vincent Bronner
Max Marius Heller (master's student; since August 2023)
Jan Henneco
Dr. Rajika Kuruwita (HITS Independent Postdoc)
Dr. Eva Laplace
Aina Sophie Nambena (bachelor's student; March–July 2023)
Prof. Dr. Philipp Podsiadlowski (Klaus Tschira Guest Professor; May, November, & October 2023)
Julian Saling (student assistant; until May 2023)
Simon Speith (student assistant; May–September 2023)
Duresa Temaj (until March 2023)
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 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 lifespans, 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 being 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, which 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 the 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 can 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.



## Group News

One of the most notable events in 2023 was our group's successful participation in the Hochschulwettbewerb ("University Competition"), at which we were awarded a grant for developing the board game "Habitable" (see below). One of the two 2023 Klaus Tschira Guest Professors – Philipp Podsiadlowki from Oxford University – collaborated extensively with our group during his three-month stay at HITS (see also Chapter 6.1). His visit was extremely fruitful, and several projects were initiated and finished with his involvement. In fact, Philipp enjoyed his stay at HITS so much that he plans to come back for further extended periods. Our bachelor's student Aina Nambena finished her project on constraining the birth distribution of black holes in X-ray binaries using a Bayesian hierarchical model, and new master's student Max Heller joined our team in the summer. Max is currently investigating the outcome and further evolution of remnants of stellar mergers. Our postdoc Dandan Wei left HITS at the end of 2023. We wish her all the best and are very much looking forward to our continued collaboration. Furthermore, our former master's student Duresa Temaj began a PhD project at the Max Planck Institute for Solar System Research.

*Figure 51: The Habitable team (from left to right): Dandan Wei (with Jan Henneco joining remotely), Julian Saling, Duresa Temaj, Eva Laplace, Vincent Bronner, and Rajika Kuruwita (Simon Speith is missing).*

## Habitable: More Than Just a Game

Our team was involved in a special public outreach project this year called Habitable, which is a fun board game that links astronomy to the climate crisis. For the project, our team (i.e., Vincent Bronner, Jan Henneco, Rajika Kuruwita, Eva Laplace, Julian Saling, Simon Speith, Duresa Temaj, and Dandan Wei; Figure 51) was one of the winners of the Hochschulwettbewerb ("University Competition") and earned a 10,000-euro prize. The annual competition is organized by "Wissenschaft im Dialog" ("Science in Dialogue") and was funded in 2023 by the German Ministry of Education and Research as part of the Year of Science 2023 theme, "Our Universe." At the beginning of the 9-month-long project, we learned about board game design by attending a workshop led by a professional game designer. By testing our first prototypes with our HITS colleagues, a school class, board game enthusiasts, and the general public, we collected useful feedback. Over time, we developed a version of the game that was consistently fun to play. We hired a student worker, who – inter alia – created an online version of the game, which is now freely available to play on the platform Tabletopia. By working closely with the HITS communications team, we were able to help our project reach a wide audience on social media and in the press. We additionally organized and joined multiple events where we presented and tested the game. We had the chance to show Habitable at several board game fairs (i.e., in Essen, Mannheim, and Nuremberg) and received encouraging feedback from game authors and publishers alike. At the end of the project, we hired a professional graphics designer to create a beautiful design. Ninety copies of the final version are now being professionally printed as an actual board game that we will donate to astronomy institutes and public outreach centers (cf. more information on Habitable in Chapter 4).

## The Universal Sound of Black Holes

Scientifically, we devote a lot of our resources to studying some of the most exotic objects in the Universe: namely mysterious, exciting, and inescapable black holes. Using gravitational-wave detectors, it is now possible to detect the chirp sound that two black holes emit when they merge. Approximately 70 such chirps have been found, and the fourth observing run of the Advanced LIGO, VIRGO, and KAGRA detectors remains ongoing. In [Schneider et al., 2023], we predicted that – in this "ocean of voices" – chirps preferentially occur in two universal frequency ranges – that is, in ranges that are associated with universal black hole masses.
The discovery of gravitational waves in 2015 – which Einstein postulated 100 years ago – led to the 2017 Nobel Prize in Physics and initiated the dawn of gravitational-wave astronomy (Figure 52, next page). When two stellar-mass black holes merge, they emit gravitational waves of increasing frequency – that is, they produce the so-called "chirp signal," which can be "heard" on Earth. By observing this frequency evolution (i.e., the chirp), it is possible to infer the so-called "chirp mass," which is a mathematical combination of the two individual black hole masses.
Merging black holes have thus far been assumed to be capable of having any mass. However, our models suggest that some black holes come in standard masses that then result in universal chirps. The existence of universal chirp masses not only tells us how black holes form, but can also be used to infer which stars explode as supernovae. Moreover, these universal chirp masses provide insights into the supernova mechanism as well as into uncertain nuclear and stellar physics, and they also offer a new way for scientists to measure the

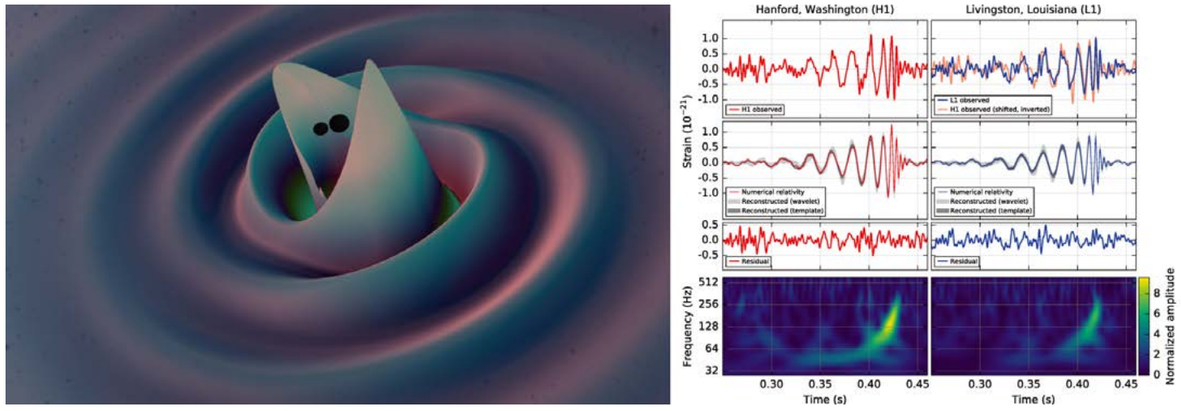



*Figure 52: Left: Ripples in the spacetime around a merging binary black hole system from a numerical-relativity simulation (image credit: Deborah Ferguson, Karan Jani, Deirdre Shoemaker, Pablo Laguna, Georgia Tech, MAYA Collaboration). Right: The gravitational-wave event GW150914 observed by the LIGO Hanford (H1) and Livingston (L1) detectors on 14 September 2015. The strain of the event and the frequency evolution (i.e., the chirp signal) are shown (image credit: Abbott et al. 2016, Phys. Rev. Lett. 116, 061102).*

accelerated cosmological expansion of the Universe.

Stellar-mass black holes with masses of approximately 3–100 times that of our Sun are the endpoints of massive stars that do not explode as supernovae but collapse into black holes. The progenitors of black holes that lead to mergers are originally born in binary star systems and experience several episodes of mass exchange between their components: In particular, both black holes form from stars that have been stripped of their envelopes. This envelope stripping has severe consequences for the ultimate fate of stars, for example, by making it easier for stars to explode as supernovae or leading to universal black hole masses, as is now predicted by our simulations.

The "stellar graveyard" – which is a collection of all known masses of the neutron star and the black hole remains of massive stars – is quickly growing thanks to both the ever-increasing sensitivity of gravitational-wave detectors and the ongoing search for such objects. In particular, there seems to be a gap in the distribution of the chirp masses of merging binary black holes, and evidence has emerged on the existence of peaks at roughly 8 and 14 solar masses (Figure 53). These features correspond to the universal chirps predicted by our models. In general, any features in the distribution of black hole masses and chirp masses can tell us a great deal about how these objects formed.

With the initial discovery of merging black holes, it became evident that some black holes have much larger masses than those found in our Milky Way. This feature is a direct result of the fact these large black holes originated from stars with a chemical composition different from the composition of our Milky Way Galaxy. We have been able to show that regardless of their chemical composition, stars whose envelopes are stripped in close binaries form black holes of <9 and >16 solar masses, but almost none of sizes in between.

In merging black holes, the universal black hole masses of approximately 9 and 16 solar masses logically imply universal chirp masses – that is, universal sounds. Because the number of observed black hole mergers is still rather low, it is not yet clear whether this signal in the data is merely a statistical fluke. The ongoing fourth observing run may already hold the answer to this question. Whatever the outcome of future gravitational-wave observations, the results are sure to be exciting and should help us better understand where the singing black holes in this ocean of voices come from.

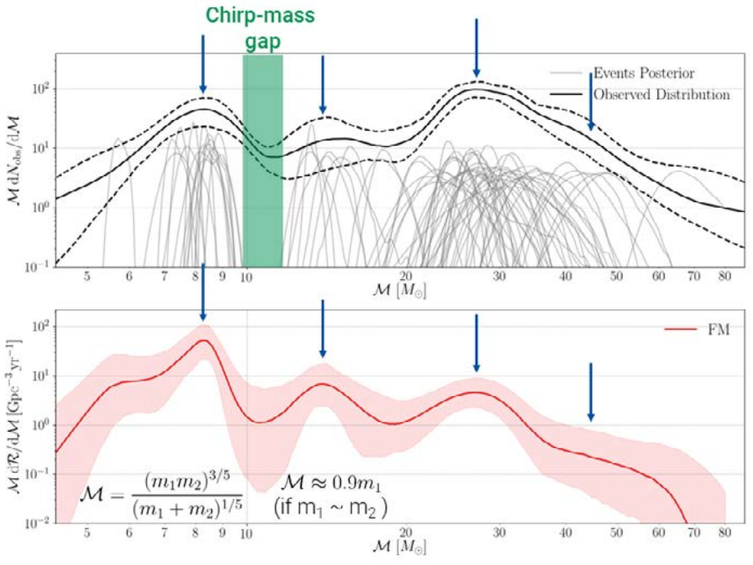



*Figure 53: Distribution of the chirp masses of all binary black hole mergers observed today. The top panel shows the raw data and probability distributions of the chirp masses of each individual event, while the bottom panel shows a model inferred from the combined observations. The gap in chirp masses at 10–12 solar masses and the thus-far identified features at about 8, 14, 27, and 45 solar masses are indicated. Figure reproduced from Abbott et al. 2021.*

## Contact Tracing of Binary Stars

It was only around a decade ago that stars – especially those more massive than our Sun – were firmly established to not live alone. Instead, stars mostly live in pairs, triples, or even quadruples. Those living in pairs that revolve around a common center of mass are called binary stars. Living close together can complicate things, and this is no different for stars.

When the stars in a binary system (or in a triple or quadruple system) grow in size or their orbit shrinks, mass transfer can occur, which severely alters the further evolution of both the donor and the accretor (i.e., the star that receives mass). Hence, in order to understand their evolution, we cannot rely solely on the theory of single stars. As a response to mass being added to their envelopes, accretor stars can grow in size, which can potentially lead to the formation of a so-called contact binary (Figure 54). These peanut-shaped contact binaries can be stable and hence observable as such, or they can be unstable and lead to a stellar merger. Alternatively, stellar mergers can also occur when the process of mass transfer becomes unstable, which leads to increasing amounts of mass being dumped onto the accretor star.

In both cases, stellar mergers leave behind a single star, which is called a merger product. We are interested in these merger products because they can explain a multitude of objects, such as stars that appear younger than they are (i.e., blue stragglers), highly magnetic stars, and stars with peculiar chemical and rotational properties.

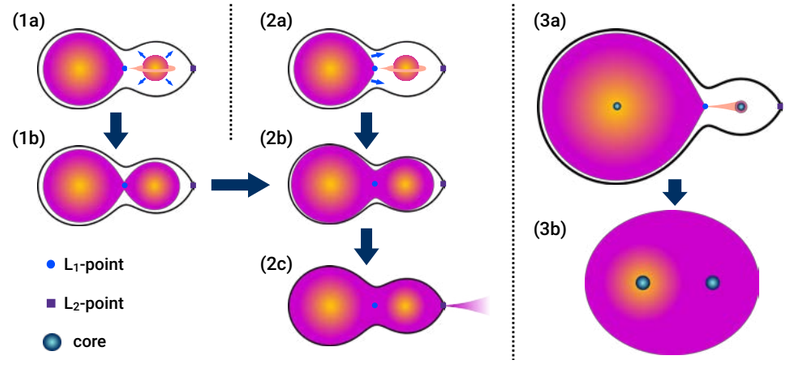



*Figure 55: Schematic representation of the pathways that lead to the formation of contact binaries (1 and 2) and classical common-envelope phases (3).*

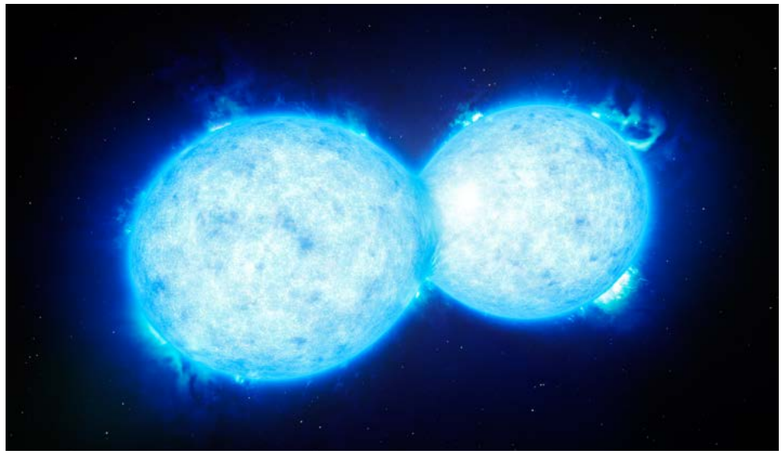


*Figure 54: Artist's impression of a massive contact binary. Credit: ESO/L. Calçada*

In order to find out which binary systems end up as contact binaries and/or lead to stellar mergers, we computed a grid of one-dimensional binary star evolution models and found that at least 12% of binary systems with component masses of between 5 and 20 solar masses merge. Additionally, in more than 19% of binary systems, one of the components is engulfed by the stellar envelope of its supergiant companion. This process leads to a classical common envelope (Figure 55), which is the focus of numerous three-dimensional hydrodynamics computations that have been performed by members of the PSO group.

Moreover, we found that most contact binaries form with relatively unequal component masses. However, the components of observed contact binaries have more equal masses. We were able to resolve this apparent discrepancy by looking at the stability of the contact binaries in our models. The bulk of the unequal mass systems were found to either quickly merge (i.e., they were unstable) or to have one of the components shrink again. In both cases, these systems were not observable as contact binaries for long.

## Common-Envelope Evolution on a Laptop

Planetary nebulae are some of the most colorful objects visible in the night sky and come in many different shapes and appearances. Unfortunately, planetary nebulae are invisible to the naked eye because they are too faint, and either binoculars or a telescope is required to fully appreciate them. Researchers have used the recently launched James Webb Space Telescope to take high-resolution images of the Southern Ring Nebula NGC 3132 (Figure 56). At the center of this planetary nebula, at least two stars orbit each other. The Southern Ring Nebula is not a special planetary nebula since many planetary nebulae have tight binary stars at their center.

*Figure 56: The Southern Ring Nebula NGC 3132 imaged by the NIRCam instrument aboard the James Webb Space Telescope. In the center, there are at least two stars that cannot be distinguished in this image. The elongated, oval shape of the planetary nebula is thought to have originated from the interaction between the central stars. Credit: NASA, ESA, CSA, and STScI.*

These planetary nebulae are thought to have originated in a process called common-envelope evolution (cf. Figure 55). As stars deplete their nuclear fuel, they tend to expand in size. If two stars orbit each other and one star grows in size, it might end up engulfing its companion star inside its extended envelope. During this evolutionary phase, the core of the giant star and the companion orbit each other within the shared envelope, hence the name "common-envelope evolution." Frictional forces dissipate orbital energy, thereby causing the orbit to contract while the envelope expands, which can lead to its ejection and to the formation of a planetary nebula with a tight binary star at its center.

Given its inherently 3D nature, the common-envelope phase is best understood through comprehensive 3D simulations, such as the simulation created by researchers in the PSO group. However, these 3D simulations are computationally highly demanding, with a single simulation easily requiring 100,000 core hours. Therefore, researchers from the SET group took on the challenge of using a simplified model for the common-envelope phase that could be implemented in standard 1D stellar evolution codes. This simplified model was then calibrated to existing 3D simulations of the common-envelope phase such that the 1D model was able to reproduce the orbit of the two stars as well as the mass of the ejected envelope. By using the 1D model, the computational cost was reduced from 100,000 core hours to less than 10 core hours. These simulations can now even be run on a laptop as compared with the high-performance compute clusters needed for 3D simulations. The 1D simulations also have the advantage that the energy transport via radiation (i.e., via photons) can be better treated than in most 3D simulations of common envelopes (see Figure 57). This situation enables us to study how such objects might appear in the sky and how their brightness changes over time. Another advantage of the 1D simulations is that the entire star – from its core to its surface – can be modeled. Such modeling is a huge challenge in 3D simulations because the dense core requires small simulation time steps, which increases the overall cost. Being able to also resolve the core in 1D simulations makes it possible to study the response of the core to envelope ejection. The core is expected to expand upon envelope ejection due to the weight removed from the envelope.

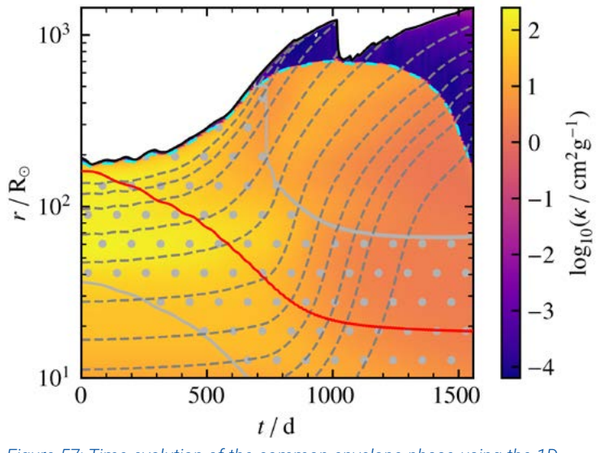



*Figure 57: Time evolution of the common-envelope phase using the 1D model. The red line indicates the companion moving through the envelope. As the binary orbit tightens, the entire envelope begins to expand. The coloring shows the opacity of the envelope (i.e., how transparent the envelope is to light). The darker the color, the more transparent the envelope is. At the cyan dashed line, which marks the photosphere, photons are expected to no longer interact with the envelope and thus to freely escape (Bronner et al., A&A, Volume 683, March 2024, in print).*

Once the 1D model has been fully calibrated on the 3D simulations, it can be used to efficiently and accurately predict the outcome of a common-envelope event. This feature represents a vast improvement over current methods, which rely mostly on simple arguments that revolve around energy conservation and are known to have large uncertainties.

In the future, the combination of 1D and 3D simulations holds promise for advancing our understanding of the common-envelope phase. While 1D simulations help to pinpoint areas where 3D models lack critical physics, such as radiation transport and core expansion, 3D simulations refine 1D models and enhance predictive capabilities regarding the post-common-envelope orbital separation and the mass of the ejected envelope.

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

# 2.13 Theory and Observations of Stars (TOS)

### Group leader

Prof. Dr. Ir. Saskia Hekker

### Team

Dr. Felix Ahlborn
Dr. Michaël Bazot
Beatriz Bordadagua
Lynn Buchele
Jeong Yun Choi
Quentin Coppèe
Francisca Espinoza
Jan Henneco (joint member of TOS and SET)
Anastasiya Kapinskaya (student; since October 2023)
Jonas Müller
Dr. Anthony Noll

Stars are an important source of electromagnetic radiation in the Universe that enable us to study many phenomena ranging from distant galaxies to the interstellar medium and extra-solar planets. However, due to their opacity, Sir Arthur Eddington once noted 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, thanks to modern mathematical techniques and high-quality data, it has become possible to probe and study the internal stellar structure directly through global stellar oscillations via a method known as asteroseismology.

Asteroseismology uses similar techniques to helioseismology – which is carried out on our closest star, the Sun – to study the structure of other stars. The properties of waves are used to trace internal stellar conditions. Oscillations that impact the whole star reveal information that is hidden by the star's opaque surface. This asteroseismic information from the CoRoT, Kepler, K2, TESS, SONG, and Plato space observatories – combined with astrometric observations from Gaia, spectroscopic data from SDSS-V APOGEE, interferometry, photometry, and state-of-the-art stellar models, such as MESA – provides insights into the stellar structure and the physical processes that take place in stars.

Understanding these physical processes and how these change as a function of stellar evolution is the ultimate goal of the Theory and Observations of Stars (TOS) group at HITS. Our research focusses on low-mass main-sequence stars, subgiants, and red giants. These stars are interesting as they go through a series of internal structure changes. Furthermore, these stars are also potential hosts of planets and serve as standard candles for Galactic studies (e.g., studies on core-helium-burning red giant stars). Exoplanet studies as well as Galactic archaeology can therefore also benefit from a better understanding of these stars.



## Background

In the TOS group, we focus on stars with oscillations similar to those present in the Sun. These so-called solar-like oscillations are low-amplitude oscillations that are stochastically excited by the turbulence in the near-surface convection layer of a star. The oscillations are sound waves and are expected to be present in all stars with convective outer layers. A convective envelope is typically present in low-mass main-sequence stars, 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), spherical degree (i.e., the number of nodal lines on the surface), and azimuthal order (i.e., the number of nodal lines that cross the spin axis). The typical frequencies and frequency differences between modes of the same degree provide the mass and radius of the stars, whereas the individual oscillation modes provide information on the internal stellar structure.

Contrary to main-sequence stars, in more evolved, so-called red-giant stars, the dipole modes (i.e., those with a spherical degree of 1) have sensitivity to both the deep interior and the outer layers. In other words, the oscillations resonate in an inner (gravity) and an outer (acoustic) cavity that are separated by an evanescent zone (i.e., the area between the cavities where oscillations cannot propagate and decay exponentially). The coupling between the two oscillating cavities and the phases of the waves in each cavity can be derived from the resulting mixed pressure–gravity oscillations and can provide information on the physical conditions in the evanescent region. Furthermore, the difference in period between pure gravity dipole modes with consecutive radial orders (i.e., 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 aims of the TOS group.

## Structure Inversions of Main-Sequence Stars

The asteroseismic data from space-based telescopes such as Kepler provide many ways to test our understanding of stellar evolution theory. One technique that directly highlights the insights into the interior of stars that asteroseismology can yield is the technique of stellar structure inversions. This technique was used to study 12 solar-type stars in research that is described in an article that has been accepted for publication in The Astrophysical Journal (Buchele, L. et al.: "Asteroseismic Inversions for Internal Sound Speed Profiles of Main-sequence Stars with Radiative Cores", ApJ 961 198 DOI 10.3847/1538-4357/ad1680). The article focuses on stars observed by Kepler that are similar to our Sun in terms of both mass and evolutionary stage.

We found best-fit models for each of our target stars using the MESA stellar evolution code. Through this process, we sought to reproduce both the observed frequencies and surface properties (i.e., effective temperature and metallicity) of the star. Whereas our models were able to get close to reproducing the observations, the high precision of asteroseismic measurements meant that there were still statistically significant differences between the best-fit model and the observations. As a result, our models could not fully reproduce the interior conditions of the stars.

Asteroseismic structure inversions allow us to target specific radii in a star and to infer differences in stellar structure variables, which in our case is the squared isothermal sound speed. Knowledge about the location and magnitude of these differences can then provide clues as to what can be improved in our stellar models. Given the current quality of asteroseismic data, we can infer these sound speed differences at distances from the center of a star that are between 5 and 30% of the total stellar radius. The process of a structure inversion requires as many observed frequencies as possible, and we are therefore restricted to using only the best of the best of the observed solar-like stars. Currently, there are 12 stars for which we are able to perform structure inversions.

Of these 12 stars, our inversions revealed that the internal sound speed profile is consistent with the observations in six cases. In one case, we found that the sound speed of our model is too high, whereas in the remaining five cases, our sound speed is too low. Correlations between these inversion results and several parameters that correspond to the age of our models suggest that our models work better when a star is younger.

To explore potential changes to the models, we focused on the two stars for which our inversions revealed the largest differences, which lay at around 15%. We tested three changes to the microphysics: namely the opacity in the core of the model, and two nuclear reaction rates. One reaction rate is important in the ppII and ppIII reaction chains, and one is the limiting reaction of the CNO cycle. In all three cases, the changes are able to increase the sound speed in the inner part of our models (see left panels of Figure 58, next page), resulting in the lower difference inferred by the inversions (see middle panels of Figure 58).

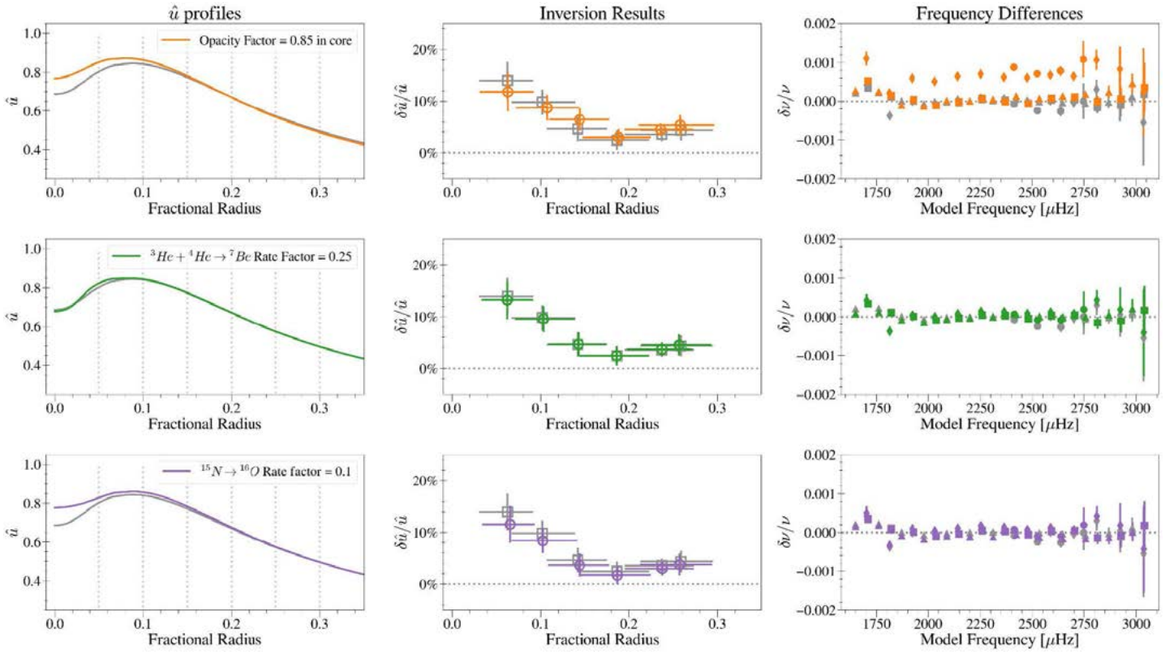



*Figure 58: Results of testing the microphysics used to evolve a best-fit model for the star KIC 6623624. The squared isothermal sound speed ($u\hat{}$) profile of each model is shown in the left plot, with light gray dashed vertical lines indicating the target radii of the inversions. The center plot shows the result of structure inversions using each model as the reference model. The right plot shows the frequency differences between each model and the observed modes of KIC 6603624. The shape of the marker denotes the spherical degree of the mode, with l = 0,1,2,3 denoted by squares, triangles, diamonds, and circles, respectively. In each plot, gray lines and points represent the values of the reference model of KIC 6603624 that are obtained using unmodified microphysics.*

However, the overall agreement between the frequencies of these models and the observations is reduced (right panels of Figure 58), which suggests that these fixes bring the center of the star into closer alignment but take the overall model farther away from the star.

## Seismic Properties of Red-Clump Stars

Within the framework of the ERC consolidator grant "Dipolar Sound," we investigated the seismic properties of red-clump (RC) stars. RC stars are considerably more evolved than our Sun and burn helium into carbon and oxygen in their cores. However, many uncertainties remain regarding their internal structures and especially regarding their cores.

The amount of fuel (i.e., helium and hydrogen) and the reaction rates of the nuclear fusion that these stars undergo determine the energy production and thus also their evolution. Properly understanding this information is therefore necessary to understand the evolution of the star. However, our knowledge of the conditions in the cores of RC stars is subject to several uncertainties. In particular, the size of the convective core is ill-defined by theory in our models due to the occurrence of complex hydrodynamical processes such as overshoot and semi-convection. Additionally, the rates of the helium-burning nuclear reaction are prone to experimental uncertainties and to extrapolation to the extreme conditions in stellar cores, which cannot be reproduced in experiments.

We investigated these issues by performing an asteroseismic study of RC stars which exhibit mixed pressure–gravity modes that probe the full structure of the star and are especially sensitive to the features in the region around the stellar core. We particularly focused on the "period spacing" – that is the difference in period between consecutive mixed modes that has been measured in several thousands of stars thanks to the data obtained by the Kepler satellite.

Earlier research indicated that current stellar evolution models fail to reproduce the distribution of the observed period spacing as the models yield systematically lower values.

From the theory of stellar oscillations, we know that the value of the period spacing is directly related to the size of the convective core. Using the MESA stellar evolution code, we computed RC stellar models assuming different core boundary mixing prescriptions and different values for the nuclear reaction rates. We found that the properties of the mixing around the core had the greatest impact on the value of the period spacing. The so-called "maximal overshoot" scheme – which was designed to yield large period spacings – was found to be the mixing prescription that is the most compatible with the observations. However, the construction of this overshoot is intrinsically unphysical, and the question of the core boundary mixing thus has yet to be solved.

We additionally tested the effect of changing the nuclear rate of helium-burning reactions and found that increasing the rate of the carbon-alpha reaction (in which carbon and helium react to form oxygen) lengthens the helium-burning phase, which – in turn – increases the maximum period spacing value reached by the models, as illustrated in Figure 59. The effect is greater than the observational uncertainties, which means that the seismology of RC stars could potentially be used to retrieve astrophysical constraints on the rate of the carbon-alpha reaction. However, this discovery would require more constraints on the core boundary mixing – potentially through hydrodynamical simulations – to isolate the effects of the nuclear rates from the effects of the mixing.

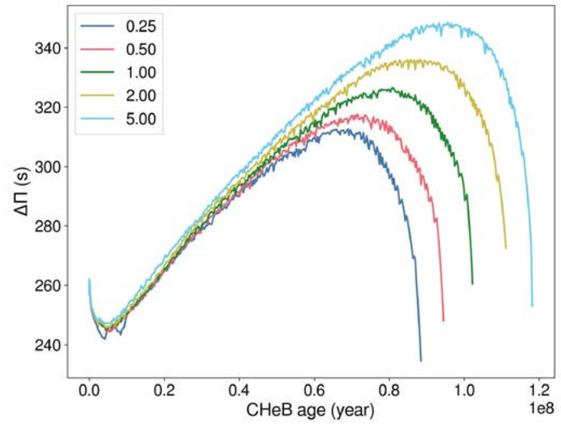



*Figure 59: Evolution of the period spacing (ΔP) against the time since the start of the helium-burning phase for models computed with different factors of the carbon-alpha nuclear reaction rate.*

Our article on this work has been accepted in Astronomy & Astrophysics (Noll. A., Basu, S., & Hekker, S. (2024). "Effect of nuclear reactions rates and core boundary mixing on the seismology of red clump stars." arXiv preprint arXiv:2401.05519).

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.

# 2.14 HITS Independent Postdoc Research

**HITS Independent Postdocs**
Dr. Rajika Kuruwita
Dr. Fabian Grünewald

The HITS Independent Postdoc Program offers a great opportunity for highly talented young scientists who wish to transition from PhD students to junior group leaders. The program supports 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 (see Chapter 6.1).

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 as well as 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 the various courses offered at HITS.

Astrophysicist Rajika Kuruwita is the first HITS Independent Postdoc. 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. She collaborates closely with the SET group (see Chapter 2.12).

The second HITS Independent Postdoc is Fabian Grünewald. Born in Germany, Fabian studied chemistry at the University of Groningen, the Netherlands, and received his PhD with honors in physical/computational chemistry from the University of Groningen. He joined HITS in October 2023 and collaborates closely with the MBM group (see Chapter 2.7) and the MCM group (see Chapter 2.8).

## Rajika Kuruwita: The Formation of Multiple Star Systems

Since beginning her position in October 2022, Rajika has continued investigating various aspects of binary and multiple star formation.

In May, her paper on “The contribution of binary star formation via core fragmentation on protostellar multiplicity” was published (Kuruwita & Haugbølle, 2023). There are two main pathways for binary/multiple star formation: (1) core fragmentation, in which a turbulent cloud of gas fragments into multiple stars, and (2) disk fragmentation, in which a star hosts a massive gas disk that is unstable and fragments. Rajika conducted a comprehensive study on the formation of multiple star systems through core fragmentation in different star-forming environments and on how these systems later evolved. The work was then compared with observations of young multiple stars and was able to reproduce a peculiar bimodal distribution, thereby calling into question how common disk fragmentation is in the formation of close binaries.

The work of Kuruwita & Haugbølle produced a vast treasure trove of data that continue to be explored and have helped observers. Recently, Rajika tested these models of fragmentation against observations of a massive star-forming complex. The paper was published in Nature Astronomy (Li S et al (2024), “Observations of high-order multiplicity in a high-mass stellar protocluster”, Nat Astron 1, https://doi.org/10.1038/s41550-023-02181-9).

Rajika was additionally awarded an Isabel Rojas Travel Award, which enabled her to collaborate with Professor Christoph Federrath at the Australian National University and to study how stars get their spin. There exists a class of low-mass stars that have rotation periods of less than 2 days (for context, our Sun has a rotation period of 25 days). Many of these stars are found in binaries, and Rajika has thus been investigating whether the binary star formation pathway inherently produces fast-rotating stars, which would have implications for the lifetime and evolution of stars in these systems.

## Fabian Grünewald: Accelerating the Design of Therapeutics Through Simulations

Fabian Grünewald develops and applies simulation methods in order to decipher structure–function relationships in biomaterials that push the boundaries of (bio)-polymer design at the interface of material science and nanomedicine. Fabian’s main project focuses on simulating nanoparticles (NP) for delivering RNA therapeutics, such as the COVID-19 vaccines. Broadly speaking, these NPs come in three varieties: synthetic NPs, which are made of polymers; lipid-bilayer-coated NPs (LNP); and hybrid NPs, which consist of a combination of a polymer matrix and a lipid bilayer coating. Additionally, all classes of NPs usually employ “camouflaging” techniques, in which synthetic polymers, carbohydrates, or protein coatings hinder detection from the immune system. The coating can also be used to selectively target certain cells. Tuning the polymer properties, the lipid composition, and even the RNA sequence of the NPs gives researchers the ability to adjust their specificity, stability, and release dynamics. At the same time, this tuning presents an enormous challenge given the large number of tunable components.
In silico screening methods – such as those based on molecular dynamics (MD) simulations – have the potential to accelerate the rational design of RNA-based therapeutics because these methods are typically faster and more cost-efficient than is experimental exploration. However, in order to utilize in silico screening methods, it is necessary to first develop simulation parameters, automated simulation protocols, and benchmarks. At HITS, Fabian is addressing these key challenges by extending the Martini coarse-grained force field with parameters that describe polymers and RNA, both of which are used in NP therapeutics. Furthermore, he also leads an international team that is working on extending the functionality of the open-source Python suite Polyply in order for it to be able to routinely set up and run simulations of these NPs. The simulation protocols are being tested against available data from the literature.

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 (siehe auch Kapitel 6.1).

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 2.12) zusammen.
Der zweite „HITS Independent Postdoc“ ist der gebürtige Deutsche **Fabian Grünewald**, der seine Promotion in physikalischer und computergestützter Chemie an der Universität Groningen / Niederlande Anfang 2023 mit Auszeichnung abschloss und im Oktober 2023 ans HITS kam. Er arbeitet eng mit der MBM-Gruppe (siehe Kapitel 2.7) und der MCM-Gruppe (siehe Kapitel 2.8) zusammen.

# 3 Centralized Services

## 3.1 Administrative Services

### Group leader
Dr. Gesa Schönberger

### Team
Yashasvini Balachandra (controlling)
Christina Blach (office)
Frauke Bley (human resources)
Christina Bölk-Krosta (controlling)
Benedicta Frech (office)
Silvia Galbusera (human resources)
Harald Haas (Grant Office team leader; since March 2023)
Jessica Herbert (accounting; since May 2023)
Ingrid Kräling (controlling)
Dr. Barbara Port (Scientific Manager)
Thomas Rasem (controlling; until January 2023)
Rebekka Riehl (human resources and Assistant to the Managing Director)
Jason Vay-Disterhöft (human resources; since July 2023)
Irina Zaichenko (accounting)

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

Due to the constantly recurring tasks and the regular handling of the same (or very similar) processes, the work of the HITS administration may not seem to differ significantly from year to year when viewed from the outside. But looks can be deceiving. In fact, every year, there are many unique details to be worked out, individual cases to review, and new decisions to be made. This type of work is usually unavoidable, often takes place in the background, and only becomes apparent when something doesn't go as it should. A good administration takes on its tasks effectively and in a solution-orientated way.

Moreover, it requires specialists who are reliable in their work and helpful with complicated issues.
With these goals in mind, the HITS administration is constantly evolving – including in 2023. Last year, in addition to its extensive day-to-day business in the areas of HR, accounting, and controlling, the administration trained four new colleagues and bid farewell to two long-standing controllers. In addition, a Grant Office team leader was set up for accounting and controlling. The team was additionally intensively involved in structuring and digitalizing many processes, including guidelines for good scientific practice.

In the end, no two years are the same for the HITS administration. Instead, HITS strives to be a modern administration: The goal is to create a group of service-oriented professionals who guarantee researchers a high degree of independence, keep processes efficient, act transparently, and concentrate on making things possible within the given framework.



## 3.2 IT Infrastructure and Network

### Group leader
Dr. Ion Bogdan Costescu

### Team
Dr. Bernd Doser (Senior Software Developer)
Dr. Simon Kreuzer (System Administrator)
Christiane Luttermann (Team Assistant; since April 2023)
Norbert Rabes (System Administrator)
Andreas Ulrich (System Administrator)
Taufan Zimmer (System Administrator)

2023 started quite abruptly: Much to our surprise, at the end of 2022, Microsoft changed their licensing policy. Consequently, as of February 2023, HITS is no longer eligible for non-profit Office365 licenses, which has led to multiple changes in the way we manage these licenses.

Due to a change in the housing conditions for the computer room of the University Computing Centre (URZ) at Heidelberg University, the HITS cluster located there is scheduled to remain in operation until around mid-2024. In order to provide enough disk space for the scientific data generated during this time and beyond, at the beginning of February 2023, we put a new storage system of around 4.5PiB into production and transferred around 2.5PiB of data from the old system to it. At the same time, we extended the storage system located at HITS by adding 2 servers. The overall storage capacity available in our HPC environment has thus grown to slightly over 7PiB. Despite such impressive numbers, the storage capacity will very likely need a further increase over the course of the next year due to the rate at which our scientific data is growing.

In order to partly compensate for the delayed cluster replacement at URZ, HITS invested in an extension to the Helix cluster of Heidelberg University. This extension comprises 84 nodes with AMD Milan CPUs and has been available since November to all HITSters – in particular, enabling scientific groups that are associated with neither Heidelberg University nor KIT to gain access to Helix. In order to implement this access, we joined the DFN-AAI – that is, the Authentication and Authorization Infrastructure of the German Research Network – and set up our own Identity Provider. This service can be easily extended to enable access to resources offered by other educational and research institutions, such as computing and storage.

Toward the end of the year, together with our KTA IT colleagues, we made some significant changes in the domain name services (DNS) setup after an aging DNS server had been replaced. Such changes are very delicate because they have the potential to seriously disrupt most network services, including for our main domain, h-its.org. We believe that the changes have gone unnoticed, which means that we have done a good job!

Throughout the year, we had to handle multiple planned and unplanned interruptions in the delivery of electricity and cooling. Luckily, no serious problems with our computing equipment occurred. On a few occasions, we could even taken advantage of the interruptions to improve the redundancy and monitoring of our systems. Nonetheless, we hope to have fewer such interruptions in 2024 and beyond.

# 4 Communication and Outreach

### Head of Communications
Dr. Peter Saueressig

### Team
Marisa de Sa Almeida (since September 2023)
Anna Cap (student)
Angela Michel

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 2023, we were involved in a plethora of events and contributed to special projects focused on stars and planets. Furthermore, we finally managed to fill a long-vacant position by welcoming Marisa de Sa Almeida as a new member of our team.



## Successful Research Sparks 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 2023, the Communications team was pleased to announce several success stories to the public.

Once again, a HITS researcher received a grant from the European Research Council (ERC) when astrophysicist and group leader Friedrich Röpke (Physics of Stellar Objects) was awarded an ERC Advanced Grant of €2.5 million for his project ExCEED. Friedrich's proposal is one of only 13.2% of project applications that were selected for funding. Moreover, according to this year's Highly Cited Researchers list from Clarivate, group leader Alexandros Stamatakis (Computational Molecular Evolution) was named one of the most-cited researchers worldwide for the eighth year in a row. The ranking is an important indicator of the impact of a researcher's scientific publications.

HITS researchers were also busy publishing papers, for instance, on collagen, which is the most abundant protein in our bodies: Scientists from the Molecular Biomechanics group revealed how the rupture of weak sacrificial bonds within collagen tissue helps to localize damage caused by excessive force, to minimize negative impact on other tissue, and to promote recovery. Their work was published in both Nature Communications and Angewandte Chemie ("Applied Chemistry") (international edition) as well as in the popular science magazine "Spektrum der Wissenschaft" ("Spectrum of Science") (see Chapter 2.7). Furthermore, members of the Stellar Evolution Theory group published a study on black hole mergers. Using gravitational-wave detectors, the group found that it is possible to detect the "chirp" sound that two black holes produce when they merge. The researchers predicted that in this "ocean of voices," chirps tend to occur in two universal frequency ranges. The study was published in The Astrophysical Journal Letters (see Chapter 2.12).

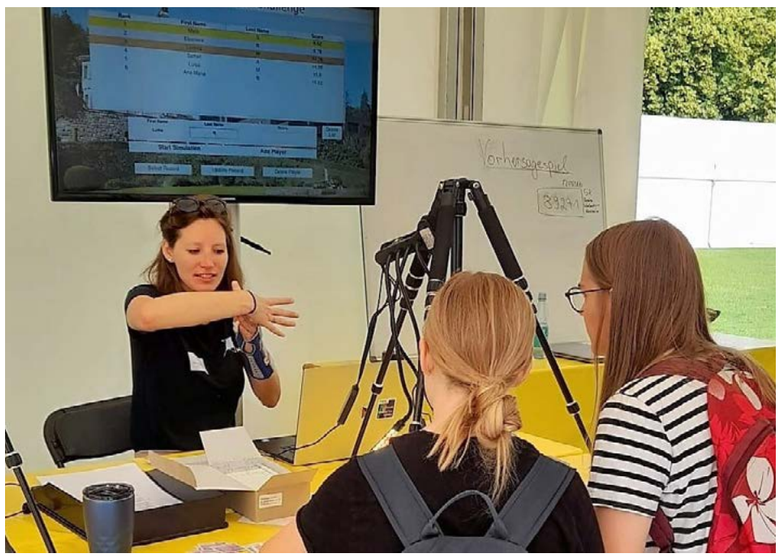
## Events, Events, Events!

The post-pandemic era in 2023 saw a comeback of on-site events that attracted many people. This particularly held true for the "Explore Science" show at Herzogenriedpark in Mannheim (21–25 June), which welcomed 29,000 visitors on the topic of "Mathematics." HITS participated in the event with two hands-on activities in which science enthusiasts could test their streamlining and weather forecasting skills. Members of the DMQ and CST groups designed the stations with the support of their colleagues from CME, MBM, Administration, and Controlling, which made the festival a truly interdisciplinary experience.

In January, HITS researchers from the SET and MCM groups welcomed young women who were participating in ZOrA, which is an initiative that supports female high- students in STEM subjects. Eva Laplace and Jan Henneco (both SET) told their keen audience about the lives of stellar objects. Moreover, Giulia Paiardi and Jonathan Teuffel (both MCM) focused on the basis of their research and on the increasing importance of molecular simulations.

*Giulia Paiardi (MCM) during her talk at the ZOrA meeting in the Studio Villa Bosch.*

In July, the biannual HITS Alumni Meeting was held at the Studio Villa Bosch and the HITS campus (see Chapter 5.4). The meeting was intertwined with a "Math colloquium," which was a special event in honor of former group leader and HITS alumna Anna Wienhard (Groups and Geometry group), who now serves as Director at the Max



*The HITS contribution to Explore Science: Elaine Zaunseder (DMQ) explaining the "Go-With-The-Flow" challenge hands-on station.*

*The HITS – ISH interns Shelly Shomron (Rehovot, left) and Clarisse Huchet-Marliacy (Montpellier, right) with their tutors Johanna Riedel and Kai Polsterer.*

Planck Institute for Mathematics in the Sciences in Leipzig (see Chapter 5.3).

Also in July, HITS participated in the "Universe on Tour" project – a mobile planetarium on astronomy that toured 15 cities in Germany, including Heidelberg. The project was initiated and sponsored by the Federal Ministry of Education and Research (BMBF) during the "Wissenschaftsjahr" ("Science Year") entitled "Unser Universum" ("Our Universe"). HITS contributed to the presentation at University Square in Heidelberg, and group leader Saskia Hekker (TOS) gave a talk on the "sound of stars."

During the summer, the Institute once again took part in the International Summer Science School Heidelberg (ISH) and welcomed one student each from Heidelberg's twin cities of Montpellier (France) and Rehovot (Israel). Members of the AIN group introduced the girls to AI in astronomy. At the end of the four-week summer school in August, Angela Michel organized the final scientific presentations of all participating students at Studio Villa Bosch, moderated by AIN group leader Kai Polsterer.

A special event took place in September when the Heidelberg Laureate Forum celebrated its 10th anniversary. Additionally, HITS again hosted about 20 young researchers from five continents at the Studio Villa Bosch (see Chapter 7).

By the end of 2023, the Events team (Communications plus Christina Blach and Benedicta Frech from HITS Administration) had managed to organize a considerable number of scientific events, including the SIMPLAIX workshop (see Chapters 5.1.2 and 7) and more than one dozen colloquium talks in hybrid format, which were streamed, recorded, and broadcast on the HITS YouTube channel (see Chapter 5.2).

## "Habitable": A Very Special Board Game

What makes a planet habitable? How can it be kept habitable, how can life be developed, and what strategic decisions put this life in danger? Members of the Stellar Evolution Theory group tackled these existential questions by playfully testing the habitability of planets. The result was a board game that combines astronomy and the climate crisis. In a competition that was part of the 2023 German Science Year "Our Universe," the international team was awarded a prize for their idea and received funding of EUR 10,000 to implement the game. In less than one year, they managed to develop "Habitable" one step at a time with game testing events at HITS and elsewhere.


The team finally published an online version and produced a prototype of the board game. The Communications team supported them in organizing the testing events and in spreading the news, which gained a remarkable amount of public awareness.



*Members of the SET group developed the "Habitable" board game and ran several tests with school classes and science fans.*

"Habitable" is a strategy game for three to five players in which the aim is to make planets sustainably habitable and to enable and develop life on them. Target groups include families, game fans, astronomy enthusiasts, and educators alike. What distinguishes "Habitable" from many other board games is that it is based entirely on scientific findings – namely from both astronomy and climate research.

## And on and on It Goes: The "Journalist in Residence" Program

We firmly believe that an important prerequisite for successful science communication is the development of reliable and sustainable journalistic contacts. Since the establishment of the "Journalist in Residence" program at HITS in 2012, the Institute has continued to refine the program, which is geared toward experienced science journalists and offers them a paid sojourn at HITS. During their stay, these journalists can learn more about data-driven science and get to know researchers and new research topics without the pressure of the "daily grind."

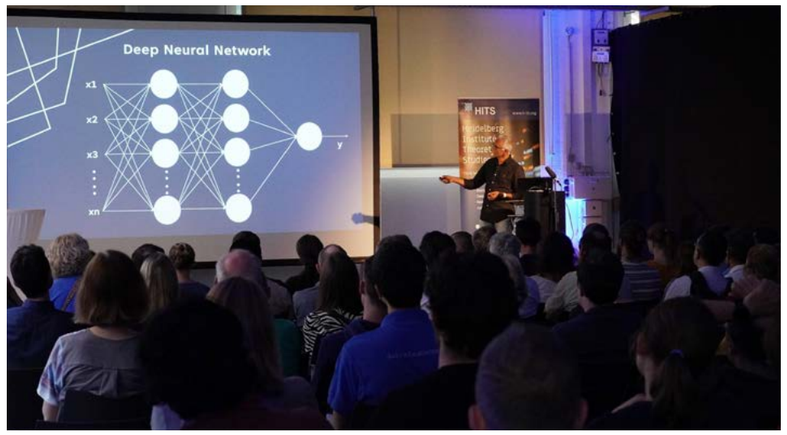
Our 11th "Journalist in Residence" was award-winning science journalist and author Anil Ananthaswamy (India/USA). Anil stayed for six months, gave an internal seminar on "the seeds of stories," and delivered a public talk on a hot topic of the year entitled "Chat GPT and its ilk." The talk was held in the MAINS (Mathematics and Informatics Station) in downtown Heidelberg and attracted many people from the region and beyond. Moreover, Anil used his stay to visit the editorial offices of Springer Nature, moderated a session at the Heidelberg Laureate Forum, and gave talks at universities and science festivals in Germany.

In the summer, HITS announced the next call for applications, with candidates from six continents applying. A committee of science journalists and scientists selected science journalist and trained astrophysicist Felicitas Mokler (Germany) as the next "HITS Journalist in Residence." Felicitas will come to the Institute in April 2024 for a six-month stay.

In the more than one dozen years since its conception in 2011, the Journalist in Residence program has proven to be a best practice for good science communication at large as well as for science journalism, in particular.

Indeed, the European Research Council (ERC) recently launched the Science Journalism Initiative, which is modeled after the HITS residency. Peter Saueressig participated in the international committee that selected the consortium that set the project on track in fall 2023.


*Three journalists in residence on one picture (f.l.t.r.): Anil Ananthaswamy and his predecessors Michele Catanzaro (2014) and Pia Grzesiak (2013).*




*Anil Ananthaswamy, HITS Journalist in Residence 2023, during his talk in the MAINS venue in Heidelberg.*

# 5 Events

## 5.1 Conferences, Workshops & Courses

### 5.1.1 Matter to Life Spring Days

**Studio Villa Bosch, Heidelberg, 30–31 March 2023**


The Max Planck School "Matter to Life" (MtL) consists of an innovative network of German organizations and provides excellent graduate education. By joining leading research institutions and universities, students receive top-notch education and training from world-renowned scientists. HITS is partner of the MtL network, with MBM group leader Frauke Gräter serving as both MtL fellow and Supervisor. Among other activities, MtL organizes the so-called "Spring Days" and "Fall Days," which are annual internal symposia that are attended by MtL students, PhD candidates, and HITS Fellows alike.

The "Spring Days" 2023 was organized as a two-day multi-hub event in Göttingen and Heidelberg with the topics of "Science in Matter to Life" and "Sustainability in Matter to Life." HITS hosted the Heidelberg part of the event in the Studio Villa Bosch. Frauke Gräter welcomed the participants both on-site and online. The program comprised lectures as well as sessions entitled "get to know my lab" – at which students shared their experiences – in addition to group work and presentations throughout the second day.

https://mattertolife.maxplanckschools.org/spring-days-2023

### 5.1.2 2nd SIMPLAIX Workshop "Machine Learning for Multiscale Molecular Modeling"

**Studio Villa Bosch, Heidelberg, 2–4 May 2023**

SIMPLAIX is a collaboration between the Heidelberg Institute for Theoretical Studies (HITS), the Karlsruhe Institute of Technology (KIT), and Heidelberg University that focusses on bridging scales from molecules to molecular materials via multiscale simulations and machine learning (see www.simplaix.org and Section 7 of the Annual Report). The aim of the workshop was to bring together scientists from the field to share their research and discuss current challenges. The meeting was a great success and included excellent presentations as well as lively discussions that featured active participation by young and established scientists alike. The covered topics included machine-learned force fields at various scales, large language models for chemistry, simulations of transport processes in biomolecular systems and materials, machine learning–enhanced multiscale simulations of molecular interactions and reactions, and machine learning analyses of molecular simulation results. The broad range of discussed applications highlighted the impact of recent advances in machine learning approaches to the modeling and simulation of (bio) molecules and materials in terms of computational efficiency, accuracy, and scope.

The workshop included lectures by 13 invited speakers, 8 contributed talks and poster presentations. About 80 participants attended the workshop. SIMPLAIX is supported by the Klaus Tschira Foundation and the participating institutions. For more details, see https://simplaix-workshop2023.h-its.org/ and #simplaix_ml3m.

Scientific Organizing Committee: Rebecca Wade, Rostislav Fedorov, Frauke Gräter, and Ganna (Anya) Gryn'ova (all from HITS); Fred Hamprecht, Andreas Dreuw, and Ullrich Köthe (all from Heidelberg University); and Pascal Friederich and Marcus Elstner (both from KIT)



### 5.1.3 EMBO Practical Course on Computational Molecular Evolution

**Heraklion, Greece, 7–18 May 2023**

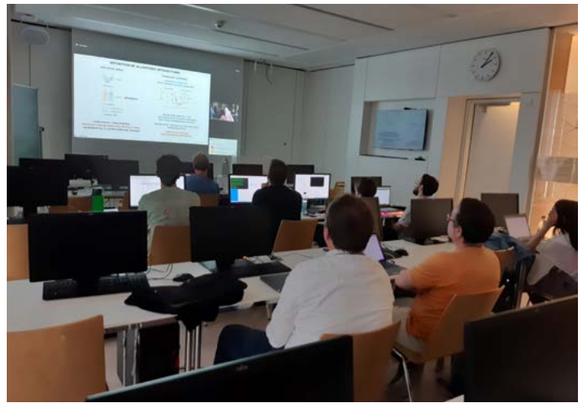
After a long pandemic break, the EMBO Practical Course on "Computational Molecular Evolution" finally took place again at the Hellenic Institute of Marine Research near Heraklion, Crete. It provided graduate and postgraduate researchers with the theoretical knowledge and practical skills necessary to carry out molecular evolutionary analyses on sequence data. The course offered data retrieval, alignment techniques, phylogeny reconstruction, hypothesis testing, and population genetic approaches. Moreover, participants had the opportunity to interact with leading scientists in evolutionary bioinformatics and with authors of famous analysis tools (e.g., Maria Anisimova, Bastien Bousseau, Bruce Rannala, Alexandros Stamatakis, Benjamin Redelings, Ziheng Yang, Rachel Warnock) in evolutionary bioinformatics.

CME group leader Alexandros Stamatakis was the main organizer of the event, and HITS was again the main co-sponsor of the course. Among the speakers was CME group member Lukas Hübner.

### 5.1.4 2nd HBPMolSim Training Workshop

**Mathematikon (hybrid), Heidelberg University, 21–23 June 2023**

One of the goals of the EU-supported Human Brain Project (HBP) was to develop computational approaches to the multiscale simulation of brain processes and to make the corresponding data, software, and workflows available in the EBRAINS infrastructure (https://ebrains.eu). The HBP training workshops aim to provide practical training in the use of these computational tools to interested scientists ranging from master's students to postdoctoral researchers. In 2023, we organized the second HBPMolSim training workshop on computational tools that were developed in HBP and that enable brain simulation and modeling at the molecular and subcellular levels. The three-day event included sessions on brain proteomics and subcellular network simulation, databases for in silico brain drug discovery, as well as tools for computing binding kinetics via molecular dynamics, quantum mechanics, and Brownian dynamics. The workshop included training on the SDA7 and tauRAMD tools, which were developed in the MCM group at HITS.



*Nobel Prize winner Jean-Pierre Changeaux (Institut Pasteur, France) speaking at the training workshop.*


*Giorgio Colombo (Univ. Pavia, Italy) describing his research at the BRAVE workshop.*

After the training workshop, the results of the "BRAVE" HBP partner project were presented in an open workshop. The BRAVE project addressed the problem of COVID-19's impact on the brain by tackling COVID-19 brain inflammation via computer-aided molecular design. The aim of the project was to block SARS-CoV-2-mediated neurodegeneration and brain damage by inhibiting the activation of the NLRP3 inflammasome, which is responsible for neuroinflamma-

tion in several neurodegenerative conditions, such as Alzheimer's disease and Long COVID-19. Promising new inhibitors were identified and are currently undergoing further experimental characterization. The project was a collaboration between scientists at the Universities of Turin and Pavia (both in Italy) as well as the Forschungszentrum Jülich and HITS (both in Germany).

For details (including slides and recordings), see https://flagship.kip.uni-heidelberg.de/jss/HBPm?m=SgD&mI=253.

Organizers: Giulia D'Arrigo, Stefan Richter, Rebecca Wade (HITS)

### 5.1.5 Tools for Systems Biology and Data Exchange: de.NBI Course

**Magdeburg, Germany, 28–30 November 2023**

SPONSORED BY THE
Federal Ministry of Education and Research

The goal of this three-day course was to familiarize participants with software tools as well as with analyzing, creating, editing, importing, simulating, and storing biochemical networks. The focus was on basic techniques for modeling biochemical networks, including data access and storage in line with the FAIR principles. The course comprised hands-on exercises using different tools, including CellNetAnalyzer, CNApy, COPASI, SABIO-RK, and FAIRDOM Hub/SEEK, which is a data and model management platform fitted to the needs of systems biologists. The hands-on exercises throughout the three days ensured that attendees became familiar with the software tools as well as with analyzing, creating, editing, importing, simulating, and storing biochemical networks. Among the educators were Wolfgang Müller, Maja Rey, Andreas Weidemann, and Ulrike Wittig, all from the Scientific Databases and Visualization group (SDBV) at HITS, as well as HITS alumna Ursula Kummer (Heidelberg University).

### 5.1.6 "Würzburg" Winter Workshop on Stellar Astrophysics

**Studio Villa Bosch, Heidelberg, 18–19 December 2023**

From 18–19 December 2023, the 17th traditional "Würzburg Workshop on Stellar Astrophysics" took place on-site. The group of more than 40 participants included scientists from several institutions from both within Germany and abroad, who discussed topics on supernova research, binary stellar evolution, neutron stars, stellar hydrodynamics, and machine learning.

The plenary session featured invited talks by Wolfgang Kerzendorf (Michigan State University, USA), Junming Duan (University of Würzburg), Jan Henneco (HITS), and Georgios Lioutas (GSI, Darmstadt).

The plenary session was followed by topical sessions in which the participants presented their latest results and discussed new projects for collaborations.



## 5.2 HITS Colloquia

### Peter Smillie

**Mathematical Institute, Heidelberg University, Germany**

23 January 2023: Counting fullerenes with modular forms – an application of number theory to carbon chemistry (Hybrid)

### Shimei Pan

**Information Systems Department, University of Maryland, Baltimore County, USA**

27 February 2023: Bias and Fairness in AI (Hybrid)

### Victor M. Panaretos

**Institute of Mathematics, EPFL, Switzerland**

27 March 2023: Sums of Squares from Pythagoras to Hilbert via Fisher (Hybrid)

### Jürgen Knödlseder

**IRAP, Toulouse, France**

24 April 2023: The carbon footprint of astronomical research infrastructures (Hybrid)

### Adrian Raftery

**Department of Statistics, University of Washington, Seattle, USA**

28 April 2023: HITS-IMK Joint Colloquium Downscaled Probabilistic Climate Change Projections, with Application to Hot Days (Hybrid)

### Philipp Podsiadlowski


**Klaus Tschira Guest Professor, Stellar Astrophysics, University of Oxford, UK**

11 May 2023: How Stars End Their Lives (Hybrid)

### Norman Schumann

**KLIMA vor Acht e.V., Hamburg, Germany**

22 May 2023: How German Media Manage Humanity´s Greatest Communication Challenge (Hybrid)

### Olexandr Isayev

**Klaus Tschira Guest Professor, Carnegie Mellon University, Department of Chemistry, Pittsburgh, Pennsylvania, USA**

13 June 2023: Accelerating Design of Organic Materials with Machine Learning and AI (Hybrid)

### Anil Ananthaswamy

**Journalist in Residence 2023, India/USA**

5 July 2023: Chat GPT and its ilk – The rise of large language models and why we should care (at MAINS, Heidelberg)

### Steve Trettel

**Department of Math & Statistics University of San Francisco, USA**

8 July 2023: Geometry from the Inside (Math Colloquium, hybrid)

### Richard Schwartz

**Brown University, Rhode Island, USA**

8 July 2023: Divide and Conquer: 5 Point Energy Minimization (Math Colloquium, hybrid)

### Rafal Weron

**Wroclaw University of Science and Technology, Poland**

24 July 2023: Electricity price forecasting in the 2020s (Hybrid)

### Fernanda Duarte

**Department of Chemistry, University of Oxford, UK**

28 September 2023: HITS-SIMPLAIX Joint Colloquium – Exploring chemical reactivity in the age of automation and ML (Hybrid)

### Volker Deringer

**Department of Chemistry, University of Oxford, UK**

23 October 2023: HITS-SIMPLAIX Joint Colloquium – Machine-learning-driven advances in modelling and understanding disordered materials (Hybrid)

### Laura Maria Sangalli

**MOX and Department of Mathematics, Politecnico di Milano, Italy**

20 November 2023: Physics-Informed Spatial and Functional Data Analysis over non-Euclidean Domains (Hybrid)

## 5.3 Math Colloquium and HITS Fellow ceremony

**Studio Villa Bosch, Heidelberg, 8 July 2023**
In a special ceremony, Anna Wienhard – former group leader at HITS and current Director at the Max Planck Institute for Mathematics in the Sciences in Leipzig – was bestowed with the title of "HITS Fellow." The event at the Studio Villa Bosch was part of a "Math Colloquium" that included two scientific talks in honor of Anna's achievements at HITS, where she had been active for seven years.

**Five points and geometry from the inside:**
**The Math Colloquium**
Steve Trettel from the University of San Francisco (California, USA) gave a talk on "Geometry from the Inside" in which he presented a menagerie of curved and twisted worlds. These new geometries have greatly expanded our mathematical universe and have found applications that range from describing the curvature of space–time to representing data in machine learning. Richard Schwartz from Brown University (Rhode Island, USA) spoke about "5 Point Energy Minimization" and presented a proof that used computer algebra, interval arithmetic, and symmetrization with a computer demo that showed the proof in action.


*Steve Trettel.*


*Richard Schwartz.*

After the talks, HITS Scientific Director Tilmann Gneiting bestowed Anna Wienhard with the title of "HITS Fellow." In her acceptance speech, Anna emphasized the interdisciplinary atmosphere at the Institute. "Over the years, this atmosphere had a huge impact on my further development," she stated. "HITS is a special place, and it has been very special for me."

**Anna Wienhard – A journey in geometry**
Anna Wienhard received her doctorate in 2004 from Bonn University (Germany). Between 2007 and 2012, she was Assistant Professor at Princeton University (USA). In 2012, Anna accepted the position as Chair of Differential Geometry at Heidelberg University. In the same year, she received an ERC Consolidator Grant. In 2015, Anna joined HITS as leader of the associated group "Groups and Geometry," which she led in addition to performing the duties of her professorship. Since October 2020, Anna has also been Scientific Chairperson of the Heidelberg Laureate Forum Foundation, which organizes the annual Heidelberg Laureate Forum (HLF). During her time at HITS, Anna contributed to the HITS Lab, and she additionally established the Research Station Geometry & Dynamics at Heidelberg University. In 2021, Anna was awarded an ERC Advanced Grant. She left HITS to become Director at the Max Planck Institute for Mathematics in the Sciences in November 2022. Anna has received numerous


*Anna Wienhard and Tilmann Gneiting.*

honors and awards throughout her career. She is a Fellow of the American Mathematical Society and an elected member of the Heidelberg, Berlin-Brandenburg, and European Academies of Science.

**HITS Fellows**
"HITS Fellow" is the highest distinction that the Institute awards. It is given to scientists in recognition of outstanding contributions to defining and implementing the scientific agenda at HITS and in gratitude for exceptional service to the Institute. Thus far, the only two other HITS Fellows are bioinformatician Isabel Rojas (1968–2013) and astrophysicist Volker Springel, who was appointed Director at the Max Planck Institute for Astrophysics in Garching in 2018.



## 5.4 HITS Alumni Meeting

**Studio Villa Bosch and HITS Campus, Heidelberg, 8 July 2023**
Since its inception in 2010, HITS has organized meetings with its alumni as well as with those of its predecessors, EML and EML Research. Due to the pandemic, however, the meeting in 2021 had to be held "virtually" – that is, online. Fortunately, HITS was able to hold an on-site meeting again this year. As with the previous events, the emphasis was on exchange and networking between alumni and current HITSters, who were also cordially invited. More than 70 participants registered and met on a summer's day in the Studio Villa Bosch and on the HITS premises. The meeting was intertwined with a "Math Colloquium," which was a special event in honor of former HITS group leader Anna Wienhard (see Chapter 5.3). After a lunch break, the meeting commenced with a welcome address by HITS Managing Director Gesa Schönberger and HITS Scientific Director Tilmann Gneiting.


*The career panel (f.l.t.r.): Antonio Disanto, Ariane Nunes Alves, Johannes Wagner, Camilo Aponte Santamaría.*

Afterward, three alumni who now work in academia and industry shared their experiences of the different career paths they have taken: Ariane Nunes Alves – a former Capes-Humboldt Research fellow in the MCM group and current group leader at the Institute of Chemistry at TU Berlin – discussed her career in academia. Johannes Wagner – who worked on his PhD thesis in the MBM group and is now CEO of his own company, Ginkgo Analytics, in Hamburg – spoke about the highs and lows of industry as well as about the different mindset in industry compared with in basic research. Finally, Antonio Disanto – Senior Data Scientist in a German software company and former AIN member – also discussed the differences between industry and basic research. The lively discussion that followed was moderated by MBM staff scientist Camilo Aponte Santamariá.
After the session, the participants went over to the HITS campus for a "sit and sizzle" barbecue. HITS chef Ralf Westermann and his team treated the participants to delicious food and beverages, and there was plenty of room for friendly discussions and laughs that lasted until sundown.


*Gesa Schönberger during her welcome address.*

*The "sit and sizzle" barbecue on the HITS campus.*

# 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 can last anywhere from three weeks to six months. Invited guest professors collaborate with scientists at the Institute and potentially develop joint research projects. Additionally, these guest professors 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 once again welcomed two Klaus Tschira Guest Professors.

**AI in Chemistry: Olexandr Isayev**

The first Klaus Tschira Guest Professor was chemist Olexandr Isayev from Carnegie Mellon University, Pennsylvania, USA. Olexandr arrived at the beginning of May and stayed until the end of June. During his time at HITS, he explored new collaborations with the groups led by Ganna "Anya" Gryn'ova (CCC; see Chapter 2.2) and Frauke Gräter (MBM; see chapter 2.7). "My lab has the machine learning methods," Olexandr stated, "[and] they have the applications. So, we look at the reactions of radicals in collagen and check the library of chemical compounds they have developed to find compounds with better properties." Olexandr also participated in the SIMPLAIX workshop as an invited speaker (see Chapter 5.1.2), held a HITS colloquium talk on machine learning in chemistry (see Chapter 5.2), and gave a lecture at Heidelberg University. Moreover, he took his time to discuss AI and ethics with HITS Journalist in Residence Anil Ananthaswamy (for more on that program, see Chapter 4).


*"How stars end their lives": Philipp Podsiadlowski at the HITS colloquium.*

**Dying stars and Black Holes: Philipp Podsiadlowski**

The second Klaus Tschira Guest Professor was astrophysicist Philipp Podsiadlowski (University of Oxford, UK), who split his stay into two parts: He spent one month at HITS in May and returned in October to spend the rest of the year at the Institute. Philipp's main focus is currently on how massive stars "die" and on the so-called common-envelope evolution of binary star systems, which is an important problem in astrophysics that remains poorly understood. Philipp refreshed long-standing collaborations with Fabian Schneider's SET group (see Chapter 2.12) and Friedrich Röpke's PSO group (see Chapter 2.10), which proved to be very fruitful. Indeed, during his stay, he co-authored a paper by members of the SET group on the universal sound of black holes, which was published in the Astrophysical Journal Letters. Philipp also gave a talk at the HITS colloquium on how stars end their lives (see Chapter 5.2). "This Institute allows scientists to grow and build up strong groups," he stated in an interview for the HITS newsletter, The Charts. "It's a great place to do research."


*Olexandr Isayev during his talk at the SIMPLAIX workshop.*

## 6.2 HITS Independent Postdoc Program

The HITS Independent Postdoc Program offers a great opportunity for highly talented young scientists who wish to transition from PhD students to junior group leaders. The program supports young scientists in exploring their own ideas and testing new hypothe-ses. 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 the HITS shareholders of Heidelberg University and the Karlsruhe Institute of Technology (KIT). In addition, successful candidates also benefit from outstanding computing resources and various courses that are offered at HITS.

Candidates for the program must hold a doctoral degree or an equivalent academ-ic qual-ification at the beginning of the fellowship. Application is open to candidates up to three years after the completion of their PhD at 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 may not have carried out research at HITS previously (except for brief visits), and the main thread of their research may 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 Annual Report 2022). In 2023, she published a paper in the journal Astronomy & Astrophysics (see Chapter 2.14). In addition to her scientific endeavors, Rajika was heavily involved in the "Habitable" game project, in which she contributed considerably to the development of the game and presented the idea at several game fairs and other events (see Chapters 2.12 and 4).

In October 2023, biologist Fabian Grünewald (University of Groningen, the Netherlands) joined the Institute as the second HITS Independent Postdoc (see also Chapter 2.14). Born in Germany, Fabian studied chemistry at the University of Groningen, the Netherlands. He received his PhD with honors in physical/computational chemistry from the University of Groningen, where he remained as a researcher in the Molecular Dynamics group before joining HITS. Fabian's research interests center around the use of computer simulations in the in silico design and understanding of polymeric materials, which lie at the interface of biology and traditional material science. In his project, he focuses on RNA and aims to develop computational methods and software that will help to design RNA nanoparticle therapeutics. Fabian collaborates closely with the Molecular Biomechanics group (MBM; see Chapter 2.7) and the Molecular and Cellular Modeling group (MCM; see Chapter 2.8).

The next opening for the program will be in 2024.

# 7 Collaborations

## SIMPLAIX

SIMPLAIX – which is a 3-way inter-institutional collaboration between HITS, Heidelberg University, and the Karlsruhe Institute of Technology (KIT) – aims to pool the expertise of these 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 in order to study a set of challenging problems in biomacromolecules 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.

From 2–4 May 2023, the first international SIMPLAIX Workshop on "Machine Learning for Multiscale Molecular Modeling" took place at the Studio Villa Bosch in Heidelberg. The aim of the event was to bring together scientists working in the field in order for them to share their research and discuss current challenges. In this on-site meeting, more than 80 participants attended the eight sessions, in which thirteen invited speakers gave talks on current developments in machine learning for chemistry and molecular biology in application areas ranging from metal chemistry to drug design (see Chapter 5.1.3).
In the course of the year, four internal project meetings were held at Studio Villa Bosch, at Heidelberg University and at KIT and the SIMPLAIX researchers engaged in a journal session discussion. Moreover, several joint HITS–SIMPLAIX colloquia (see Chapter 5.2) and scientific talks with external speakers were organized.
Plans for 2024 include the second international SIMPLAIX workshop in May, and several joint colloquia and project meetings are also in the offing.

*Rebecca Wade opening the workshop on 2 May 2023.*

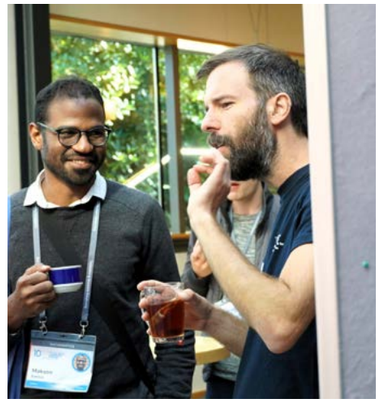
## Heidelberg Laureate Forum

The Heidelberg Laureate Forum (HLF) is a networking conference at which 200 carefully selected young researchers in mathematics and computer science spend a week interacting with laureates from these 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.

*Tilmann Gneiting welcomes young researchers in the Carl Bosch Auditorium.*

**A special anniversary: The 10th HLF**
The HLF celebrated its 10th anniversary from 24–29 September. The program included laureate lectures, master classes, lightning talks, discussions, and various interactive program elements. HITS group leader Rebecca Wade moderated a session on "synergies in research" between mathematics, computer science, and the sciences. Furthermore, HITS Journalist in Residence Anil Ananthaswamy (see Chapter 4) presented a panel discussion on Generative AI.

*Poster session in the Studio Villa Bosch.*



**HITSters meet young researchers**
HITS additionally hosted a group of 20 young researchers ranging from undergraduate students to postdoctoral scientists. Scientific Director Tilmann Gneiting welcomed the group and introduced the presentations by HITS group leaders Kai Polsterer (AIN), Ganna (Anya) Gryn'ova (CCC), and Jan Stühmer (MLI). Moreover, members of the AIN, CCC, and MLI groups presented their current research topics and publications in a poster session. As in the previous year, the event took place in the Studio Villa Bosch.

*The group during a break in the Villa Bosch garden.*

## Informatics 4 Life

Despite remarkable progress in the diagnosis and treatment of acute and chronic cardiovascular diseases in recent years, these illnesses remain the leading cause of death and hospitalization for all people worldwide. Informatics4life is a collaborative initiative founded by the Klaus Tschira Foundation that focuses on cardiovascular research and brings together experts in clinical research and computational methods. The initiative's patient-centric environment – a completely novel approach – is critical to translating research into clinical application. As Heidelberg is a "hot spot" for health research and medicine with internationally highly competitive institutes and researchers, it is the ideal location for this pioneering initiative.
The project represents an interdisciplinary alliance between cardiovascular physicians and computer scientists from Heidelberg University, University Hospital Heidelberg, and HITS. It consists of several subprojects in which HITS researchers are involved as principal investigators and contributors.

**HITS involvement: From structure-based design to machine learning**
Rebecca Wade (MCM) is a Principal Investigator of the subproject "Structure-based design of peptide-based pharmaceuticals against striated muscle disorders"; Wolfgang Müller (SDBV) is Collaborating Principal Investigator of the subproject "Research data warehouse"; and Vincent Heuveline (DMQ) is Principal Investigator of the subproject "Cognition and uncertainty quantification for numerical heart simulation." Moreover, Vincent is also Co-Coordinator of Informatics4life in cooperation with Hugo Katus and Benjamin Meder (both from Heidelberg University Hospital).

Another subproject in which HITS is involved is a "multiparametric single-cell morphological analysis of cardiomyocytes for translational cardiovascular research." Kai Polsterer (AIN) works on this project and collaborates with both Mathias Konstandin (Heidelberg University Hospital) and Stefania Petra (Heidelberg University). Although the scales are extremely different, morphological analyses of cell tissues are highly similar to analyses of multi-band observations of galaxies – an observation that provides further support for Klaus Tschira's vision that interdisciplinary research is key to making progress in research.

# 8 Publications


**Abaza H, Shutsko A, Golebiewski M, Klopfenstein SAI, Schmidt CO, Vorisek CN, Brünings-Kuppe C, Clemens V, Darms J, Hanß S, Intemann T, Jannasch F, Kasbohm E, Lindstädt B, Löbe M, Orban E, Perrar I, Peters M, Sax U, Schulze M, Schupp C, Schwarz F, Schwedhelm C, Strathmann S, Waltemath D, Wünsche H, Zeleke A (2023a).** Metadata schema of the NFDI4Health and the NFDI4Health Task Force COVID-19 (V3_1). https://doi.org/10.4126/FRL01-006450625

**Abaza H, Shutsko A, Golebiewski M, Klopfenstein SAI, Schmidt CO, Vorisek CN, Brünings-Kuppe C, Clemens V, Darms J, Hanß S, Intemann T, Jannasch F, Kasbohm E, Lindstädt B, Löbe M, Orban E, Perrar I, Peters M, Sax U, Schulze M, Schupp C, Schwarz F, Schwedhelm C, Strathmann S, Waltemath D, Wünsche H, Zeleke A (2023b).** Metadata schema of the NFDI4Health and the NFDI4Health Task Force COVID-19 (V3_2). https://doi.org/10.4126/FRL01-006453422

**Abaza H, Shutsko A, Golebiewski M, Klopfenstein SAI, Schmidt CO, Vorisek CN, Brünings-Kuppe C, Clemens V, Darms J, Hanß S, Intemann T, Jannasch F, Kasbohm E, Lindstädt B, Löbe M, Orban E, Perrar I, Peters M, Sax U, Schulze M, Schupp C, Schwarz F, Schwedhelm C, Strathmann S, Waltemath D, Wünsche H, Zeleke A (2023c).** The NFDI4Health Metadata Schema (V3_3). https://doi.org/10.4126/FRL01-006472531

**Aguila AL, Jayme A, Montaña-Brown N, Heuveline V, Altmann A (2023).** Multi-view-AE: A Python package for multi-view autoencoder models. JOSS 8(85):5093

**Ahadova A, Stenzinger A, Seppälä T, Hüneburg R, Kloor M, Bläker H, Wittemann J-N, Endris V, Gerling L, Bertram V, Neumuth MT, Witt J, Graf S, Kristiansen G, Hommerding O, Haupt S, Zeilmann A, Heuveline V, Kazdal D, Gebert J, von Knebel Doeberitz M, Mecklin J-P, Nattermann J (2023).** A "two-in-one hit" model of shortcut carcinogenesis in MLH1 Lynch syndrome carriers. Gastroenterology, 16(1):267-270

**Alpay A, Heuveline V (2023).** One Pass to Bind Them: The First Single-Pass SYCL Compiler with Unified Code Representation Across Backends. International Workshop on OpenCL, pp.1-12, ACM

**Amaro-Seoane P, Andrews J, Arca Sedda M, Askar A, Baghi Q, 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, Cattorini 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, Duțan I, Dvorkin I, Fabj G, Foglizzo T, Ford S, Fouvry J-B, Franchini A, Fragos T, Fryer C, Gaspari M, Gerosa D, Graziani L, Groot P, 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, Ricarte A, Röpke FK, 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 NC, 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, Bonilla Rivera A, 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, Portegies Zwart S, Pound A, Schussler F, Sergijenko O, Spallicci A, Vernieri D, Vigna-Gómez A (2023).** Astrophysics with the Laser Interferometer Space Antenna. Living Rev Relativ 26(1), 2

**Bloot S, Miguel Y, Bazot M, Howard S (2023).** Exoplanet interior retrievals: core masses and metallicities from atmospheric abundances. Monthly Notices of the Royal Astronomical Society 523(4):6282-6292

**Bodensteiner J, Sana H, Dufton PL, Wang C, Langer N, Banyard G, Mahy L, de Koter A, de Mink SE, Evans CJ, Götberg Y, Hénault-Brunet V, Patrick LR, Schneider FRN (2023).** The young massive SMC cluster NGC 330 seen by MUSE. III. Stellar parameters and rotational velocities. A&A 680:A32

**Bosse NI, Abbott S, Cori A, van Leeuwen E, Bracher J, Funk S (2023).** Scoring epidemiological forecasts on transformed scales. PLOS Computational Biology, 19(8):e1011393

**Bracher J, Rüter L, Krüger F, Lerch S, Schienle M (2023).** Direction augmentation in the evaluation of armed conflict predictions. International Interactions, 49(6):989–1004

**Brockhaus EK, Wolffram D, Stadler T, Osthege M, Mitra T, Littek JM, Krymova E, Klesen AJ, Huisman JS, Heyder S, Helleckes LM, an der Heiden M, Funk S, Abbott S, Bracher J (2023).** Why are different estimates of the effective reproductive number so different? A case study on COVID-19 in Germany. PLOS Computational Biology, 19(11):e1011653

**Bruers B, Cruces M, Demleitner M, Duckeck G, Düren M, Eich N, Enßlin T, Erdmann J, Erdmann M, Fackeldey P, Felder C, Fischer B, Fröse S, Funk S, Gasthuber M, Grimshaw A, Hadasch D, Hannemann M, Kappes A, Kleinemühl R, Kozlov OM, Kuhr T, Lupberger M, Neuhaus S, Niknejadi P, Reindl J, Schindler D, Schneidewind A, Schreiber F, Schumacher M, Schwarz K, Streit A, von Cube RF, Walker R, Walther C, Wozniewski S, Zhou K (2023).** Resource-aware Research on Universe and Matter: Call-to-Action in Digital Transformation. arXiv:2311.01169 [physics.comp-ph]

**de Buhr S, Gräter F (2023).** Myristoyl's dual role in allosterically regulating and localizing Abl kinase. eLife 12:e85216

**Buhr J, Franz F, Gräter F (2023).** Intrinsically disordered region of talin's FERM domain functions as an initial PIP2 recognition site. Biophysical Journal 122(7):1277–1286

**Bétrisey J, Eggenberger P, Buldgen G, Benomar O, Bazot M (2023).** Testing angular momentum transport processes with asteroseismology of solar-type main-sequence stars. A&A 673:L11

**Claveras Cabezudo A, Athanasiou C, Tsengenes A, Wade RC (2023).** 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. 19(7):2109-2119

**Chai H, Strube M (2023).** Investigating multilingual coreference resolution by universal annotations. Proceedings of the Findings of the Association for Computational Linguistics: EMNLP , Singapore, 6-10 December, pp.10010-10024

**Chavhan R, Gouk H, Stuehmer J, Heggan C, Yaghoobi M, Hospedales T (2023).** Amortised Invariance Learning for Contrastive Self-Supervision. In The Eleventh International Conference on Learning Representations. arXiv:2302.12712 [cs.CV]

**Collins CE, Sim SA, Shingles LJ, Gronow S, Röpke FK, Pakmor R, Seitenzahl IR, Kromer M (2023).** Helium as a signature of the double detonation in Type Ia supernovae. Monthly Notices of the Royal Astronomical Society 524(3):4447-4454

**Cornelio C, Stuehmer J, Hu SX, Hospedales T (2023).** Learning where and when to reason in neuro-symbolic inference. In International Conference on Learning Representations

**Criswell AW, Miller J, Woldemariam N, Soultanis T, Bauswein A, Chatziioannou K, Coughlin MW, Jones G, Mandic V (2023).** Hierarchical Bayesian method for constraining the neutron star equation of state with an ensemble of binary neutron star postmerger remnants. Phys. Rev. D 107(4),043021

**Csörnyei G, Vogl C, Taubenberger S, Flörs A, Blondin S, Cudmani MG, Holas A, Kressierer S, Leibundgut B, Hillebrandt W (2023).** Family dispute: do Type IIP supernova siblings agree on their distance? arXiv:2302.03112 [astro-ph.SR]

**Czerny B, Cao S, Jaiswal VK, Karas V, Khadka N, Martínez-Aldama ML, Naddaf MH, Panda S, Pozo Nuñez F, Prince R, Ratra B, Sniegowska M, Yu Z, Zajaček M (2023).** Accretion disks, quasars and cosmology: meandering towards understanding. Astrophys Space Sci 368(2), 8

**Diedrich L, Brosz M, Abele T, Steinke S, Gräter F, Göpfrich K, Aponte-Santamaría C (2023).** Energetics and kinetics of membrane permeation of photoresists for bioprinting. Biophysics, doi:10.1101/2023.03.27.534360

**Dimitriadis G, Maguire K, Karambelkar VR, Lebron RJ, Liu (2023).** SN 2021zny: an early flux excess combined with late-time oxygen emission suggests a double white dwarf merger event. Monthly Notices of the Royal Astronomical Society 521(1):1162-1183

**Dimitriadis T, Liu X, Schnaitmann J (2023).** Encompassing tests for value at risk and expected shortfall multistep forecasts based on inference on the boundary. Journal of Financial Econometrics, 21(2):412–444

**Djorgovski SG, Mahabal AA, Graham MJ, Polsterer K, Krone-Martins A (2022).** Applications of AI in Astronomy. arXiv:2212.01493 [astro-ph.IM]

**Dominguez-Valentin M, Haupt S, Seppälä TT, Sampson JR, Sunde L, Bernstein I, Jenkins MA, Engel C, Aretz S, Nielsen M, Capella G, Balaguer F, 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, Della Valle A, Heinimann K, Dębniak T, Fruscio R, Lopez-Koestner F, Alvarez-Valenzuela K, Katz LH, Laish I, Vainer E, Vaccaro C, Carraro DM, Monahan K, Half E, Stakelum A, Winter D, Kennelly R, Gluck N, Sheth H, Abu-Freha N, Greenblatt M, Rossi BM, Bohorquez M, Cavestro GM, Lino-Silva LS, Horisberger K, Tibiletti MG, Nascimento I do, Thomas H, Rossi NT, Apolinário da Silva L, Zaránd A, Ruiz-Bañobre J, Heuveline V, Mecklin J-P, Pylvänäinen K, Renkonen-Sinisalo L, Lepistö A, Peltomäki P, Therkildsen C, Madsen MG, Burgdorf SK, Hopper JL, Win AK, Haile RW, Lindor N, Gallinger S, Le Marchand L, Newcomb PA, Figueiredo J, Buchanan DD, Thibodeau SN, von Knebel Doeberitz M, Loeffler M, Rahner N, Schröck E, Steinke-Lange V, Schmiegel W, Vangala D, Perne C, Hüneburg R, Redler S, Büttner R, Weitz J, Pineda M, Duenas N, Vidal JB, Moreira L, Sánchez A, Hovig E, Nakken S, Green K, Lalloo F, Hill J, Crosbie E, Mints M, Goldberg Y, Tjandra D, ten Broeke SW, Kariv R, Rosner G, Advani SH, Thomas L, Shah P, Shah M, Neffa F, Esperon P, Pavicic W, Torrezan GT, Bassaneze T, Martin CA, Moslein G, Moller P (2023).** Mortality by age, gene and gender in carriers of pathogenic mismatch repair gene variants receiving surveillance for early cancer diagnosis and treatment: a report from the prospective Lynch syndrome database. eClinicalMedicine:101909

**Dudaš D, Wittig U, Rey M, Weidemann A, Müller W (2023).** Improved insights into the SABIO-RK database via visualization. Database 2023:baad011

**Dupuy A, Aponte-Santamaría C, Yeheskel A, Hortle E, Oehlers SH, Gräter F, Hogg PJ, Passam FH, Chiu J (2023).** Mechano-Redox Control of Macrophage-1 Antigen De-Adhesion From ICAM-1 (Intercellular Adhesion Molecule 1) by Protein Disulfide Isomerase Promotes Directional Movement Under Flow. Circulation Research. 132(9):e151-e168

**EDITH consortium (2023).** Vision for the Virtual Human Twin and Roadmap Outline. https://doi.org/10.5281/zenodo.7791708

**EDITH consortium (2023).** EDITH CSA Deliverable 3.2: First draft of the VHT roadmap. https://doi.org/10.5281/zenodo.8200955

**Ehlert C, Piras A, Schleicher J, Gryn'ova G (2023a).** Metal-Free Molecular Catalysts for the Oxygen Reduction Reaction: Electron Affinity as an Activity Descriptor. J. Phys. Chem. Lett. 14(2):476-480

**Ehlert C, Piras A, Gryn'ova G (2023b).** CO 2 on Graphene: Benchmarking Computational Approaches to Noncovalent Interactions. ACS Omega 8(39):35768-35778

**Ehret T, Heißenberg T, de Buhr S, Aponte-Santamaría C, Steinem C, Gräter F (2023).** FERM domains recruit ample PI(4,5)P2s to form extensive protein-membrane attachments. Biophysical Journal 122(7):1325–1333

**Ernst M, Evans JD, Gryn'ova G (2023).** Host–guest interactions in framework materials: Insight from modeling. Chemical Physics Reviews 4(4), 041303

**Ernst M, Gryn'ova G (2023).** Engineering Host–Guest Interactions in Organic Framework Materials for Drug Delivery. Helvetica Chimica Acta, e00013

**Ernst M, Poręba T, Gnägi L, Gryn'ova G (2023).** Locating Guest Molecules inside Metal–Organic Framework Pores with a Multilevel Computational Approach. J. Phys. Chem. C 127(1):523-531

**Erozan A, Lösel P, Heuveline V, Weinhardt V (2023).** ACSeg: Automated 3D Cytoplasm Segmentation in Soft X-Ray Tomography. Preprint Series of the Engineering Mathematics and Computing Lab :Nr. 2

**Ertini K, Folatelli G, Martinez L, Bersten MC, Anderson JP, Ashall C, Baron E, Bose S, Brown PJ, Burns C, DerKacy JM, Ferrari L, Galbany L, Hsiao E, Kumar S, Lu J, Mazzali P, Morrell N, Orellana M, Pessi PJ, Phillips MM, Piro AL, Polin A, Shahbandeh M, Shappee BJ, Stritzinger M, Suntzeff NB, Tucker M, Elias-Rosa N, Kuncarayakti H, Gutiérrez CP, Kozyreva A, Müller-Bravo TE, Chen T-W, Hinkle JT, Payne AV, Székely P, Szalai T, Barna B, Könyves-Tóth R, Bánhidi D, Bíró IB, Csányi I, Kriskovits L, Pál A, Szabó Z, Szakáts R, Vida K, Vinkó J, Gromadzki M, Harvey L, Nicholl M, Paraskeva E, Young DR, Englert B (2023).** SN 2021gno: a Calcium-rich transient with double-peaked light curves. Monthly Notices of the Royal Astronomical Society, stad2705

**Ewers I, Rajter L, Czech L, Mahé F, Stamatakis A, Dunthorn M (2023).** Interpreting phylogenetic placements for taxonomic assignment of environmental DNA. J Eukaryotic Microbiology 70(5), e12990

**Farmer R, Laplace E, Ma J, de Mink SE, Justham S (2023).** Nucleosynthesis of Binary-stripped Stars. ApJ 948(2):111

**Fatima M, Kolber T, Markert K, Strube M (2023).** SimCSum: Joint learning of simplification and cross-lingual summarization for cross-lingual science journalism. Proceedings of the Forth Workshop on New Frontiers in Summarization. Singapore, 6 December, pp.24-40.

**Fatima M, Strube M (2023).** Cross-lingual Science Journalism: Select, Simplify and Rewrite Summaries for Non-expert Readers. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, Ontario, Canada, July, pp.1843-1861

**Fedorov R, Gryn'ova G (2023).** Unlocking the Potential: Predicting Redox Behavior of Organic Molecules, from Linear Fits to Neural Networks. J. Chem. Theory Comput. 19(15):4796-4814

**Ferraz MVF, Neto JCS, Lins RD, Teixeira ES (2023).** An artificial neural network model to predict structure-based protein–protein free energy of binding from Rosetta-calculated properties. Phys. Chem. Chem. Phys. 25(10):7257-7267

**Fluck J, Golebiewski M, Darms J (2023).** Data Publication for Personalised Health Data: A New Publication Standard Introduced by NFDI4Health. Proc Conf Res Data Infrastr 1.

**Franz F, Tapia-Rojo R, Winograd-Katz S, Boujemaa-Paterski R, Li W, Unger T, Albeck S, Aponte-Santamaria C, Garcia-Manyes S, Medalia O, Geiger B, Gräter F (2023).** Allosteric activation of vinculin by talin. Nat Commun 14, 4311

**Geen S, Agrawal P, Crowther PA, Keller BW, de Koter A, Keszthelyi Z, van de Voort F, Ali AA, Backs F, Bonne L, Brugaletta V, Derkink A, Ekström S, Fichtner YA, Grassitelli L, Götberg Y, Higgins ER, Laplace E, You Liow K, Lorenzo M, McLeod AF, Meynet G, Newsome M, André Oliva G, Ramachandran V, Rey MP, Rieder S, Romano-Díaz E, Sabhahit G, Sander AAC, Sarwar R, Stinshoff H, Stoop M, Szécsi D,**

**Trebitsch M, Vink JS, Winch E (2023).** Bringing Stellar Evolution and Feedback Together: Summary of Proposals from the Lorentz Center Workshop. PASP 135(1044):021001

**Gerstner P, Heuveline V (2023)**. Finite element approximation of dielectrophoretic force driven flow problems. ESAIM: M2AN 57(3):1691-1729

**Glaser M, Ritterhof J, Most P, Wade RC (2023).** Wird KI neue Medikamente gegen Herzkrankheiten hervorbringen? Aktuelle Kardiologie 12(06):450-458

**Gneiting T, Wolffram D, Resin J, Kraus K, Bracher J, Dimitriadis T, Hagenmeyer V, Jordan AI, Lerch S, Phipps K, Schienle M (2023).** Model diagnostics and forecast evaluation for quantiles. Annual Review of Statistics and Its Application, 10:597–621

**Gneiting T, Lerch S, Schulz B (2023a).** Probabilistic solar forecasting: Benchmarks, post-processing, verification. Solar Energy, 252:72–80

**Gneiting T, Resin J (2023).** Regression diagnostics meets forecast evaluation: Conditional calibration, reliability diagrams, and coefficient of determination. Electronic Journal of Statistics, 17(2):3226-3286

**Gomez GM, D'Arrigo G, Sanchez CP, Berger F, Wade RC, Lanzer M (2023).** PfCRT mutations conferring piperaquine resistance in falciparum malaria shape the kinetics of quinoline drug binding and transport. PLoS Pathog 19(6):e1011436

**Gryn'ova G (2023).** Crafting molecular architectures with guided diffusion. Nat Comput Sci. 3, pp.821–822

**Gryn'ova G (2023).** Author's profile, Helvetica Chimica Acta, Vol. 106, Issue 8, August 2023. https://doi.org/10.1002/hlca.202300093

**Haag J, Hübner L, Kozlov AM, Stamatakis A (2023).** The Free Lunch is not over yet—systematic exploration of numerical thresholds in maximum likelihood phylogenetic inference. Bioinformatics Advances 3(1)

**Herrera-Rodríguez AM, Dasanna AK, Daday C, Cruz-Chú ER, Aponte-Santamaría C, Schwarz US, Gräter F (2023).** The role of flow in the self-assembly of dragline spider silk proteins. Biophysical Journal 122(21):4241–4253

**Hoga Y, Dimitriadis T (2023).** On testing equal conditional predictive ability under measurement error. Journal of Business & Economic Statistics, 41(2):364–376

**Howard S, Guillot T, Bazot M, Miguel Y, Stevenson DJ, Galanti E, Kaspi Y, Hubbard WB, Militzer B, Helled R, Nettelmann N, Idini B, Bolton S (2023).** Jupiter's interior from Juno: Equation-of-state uncertainties and dilute core extent. A&A 672:A33

**van Keulen SC, Martin J, Colizzi F, Frezza E, Trpevski D, Diaz NC, Vidossich P, Rothlisberger U, Hellgren Kotaleski J, Wade RC, Carloni P (2023).** Multiscale molecular simulations to investigate adenylyl cyclase-based signaling in the brain. WIREs Comput Mol Sci 13(1), e1623

**Kirschner M, Golebiewski M, Moser H, EU-STANDS4PM consortium, ISO/TC 276/WG 5 (2023).** ISO/TS 9491-1:2023 Biotechnology — Predictive computational models in personalized medicine research — Part 1: Constructing, verifying and validating models. International Organization for Standardization (ISO), Geneva, Switzerland.

**Konig M, Gleeson P, Golebiewski M, Gorochowski TE, Hucka M, Keating SM, Myers CJ, Nickerson DP, Sommer B, Waltemath D, Schreiber F (2023).** Specifications of standards in systems and synthetic biology: status and developments in 2022 and the COMBINE meeting 2022. J Integr Bioinform. Mar 29;20(1):0004

**Koptekin D, Yüncü E, Rodríguez-Varela R, Altınışık NE, Psonis N, Kashuba N, Yorulmaz S, George R, Kazancı DD, Kaptan D, Gürün K, Vural KB, Gemici HC, Vassou D, Daskalaki E, Karamurat C, Lagerholm VK, Erdal ÖD, Kırdök E, Marangoni A, Schachner A, Üstündağ H, Shengelia R, Bitadze L, Elashvili M, Stravopodi E, Özbaşaran M, Duru G, Nafplioti A, Rose CB, Gencer T, Darbyshire G, Gavashelishvili A, Pitskhelauri K, Çevik Ö, Vuruşkan O, Kyparissi-Apostolika N, Büyükkarakaya AM, Oğuzhanoğlu U, Günel S, Tabakaki E, Aliev A, Ibrahimov A, Shadlinski V, Sampson A, Kılınç GM, Atakuman Ç, Stamatakis A, Poulakakis N, Erdal YS, Pavlidis P, Storå J, Özer F, Götherström A, Somel M (2023).** Spatial and temporal heterogeneity in human mobility patterns in Holocene Southwest Asia and the East Mediterranean. Current Biology 33(1):41-57.e15

**Kurth M, Barayeu U, Gharibi H, Kuzhelev A, Riedmiller K, Zilke J, Noack K, Denysenkov V, Kappl R, Prisner TF, Zubarev RA, Dick TP, Gräter F (2023).** DOPA Residues Endow Collagen with Radical Scavenging Capacity. Angew. Chem. Int. Ed., e202216610

**Kuruwita RL, Haugbølle T (2023).** The contribution of binary star formation via core fragmentation on protostellar multiplicity. A&A 674:A196

**Kutzki F, Butera D, Lay AJ, Maag D, Chiu J, Woon H-G, Kubař T, Elstner M, Aponte-Santamaría C, Hogg PJ, Gräter F (2023).** Disulfide Bond Reduction and Exchange in von-Willebrand-Factor's C4-Domain Undermines Platelet Binding. Journal of Thrombosis and Haemostasis 21(8):2089–2100

**Lahén N, Naab T, Kauffmann G, Szécsi D, Hislop JM, Rantala A, Kozyreva A, Walch S, Hu C-Y (2023).** Formation of star clusters and enrichment by massive stars in simulations of low-metallicity galaxies with a fully sampled initial stellar mass function. Monthly Notices of the Royal Astronomical Society 522(2):3092-3116

**Lauterbach S, Dienhart H, Range J, Malzacher S, Spöring J-D, Rother D, Pinto MF, Martins P, Lagerman CE, Bommarius AS, Høst AV, Woodley JM, Ngubane S, Kudanga T, Bergmann FT, Rohwer JM, Iglezakis D, Weidemann A, Wittig U, Kettner C, Swainston N, Schnell S, Pleiss J (2023).** EnzymeML: seamless data flow and modeling of enzymatic data. Nat Methods 20(3):400–402

**Leaman R, Islamaj R, Adams V, Alliheedi MA, Almeida JR, Antunes R, Bevan R, Chang Y-C, Erdengasileng A, Hodgskiss M, Ida R, Kim H, Li K, Mercer RE, Mertová L, Mobasher G, Shin H-C, Sung M, Tsujimura T, Yeh W-C, Lu Z (2023).** Chemical identification and indexing in full-text articles: an overview of the NLM-Chem track at BioCreative VII. Database, baad005

**Linciano P, Quotadamo A, Luciani R, Santucci M, Zorn KM, Foil DH, Lane TR, Cordeiro Da Silva A, Santarem N, B Moraes C, Freitas-Junior L, Wittig U, Mueller W, Tonelli M, Ferrari S, Venturelli A, Gul S, Kuzikov M, Ellinger B, Reinshagen J, Ekins S, Costi MP (2023).** High-Throughput Phenotypic Screening and Machine Learning Methods Enabled the Selection of Broad-Spectrum Low-Toxicity Antitrypanosomatidic Agents. J Med Chem 66(22):15230–15255.

**Liu Z-W, Röpke FK, Han Z (2023).** Type Ia Supernova Explosions in Binary Systems: A Review. Res. Astron. Astrophys. 23(8):082001

**Liu W, Fu X, Strube M (2023a).** Modeling Structural Similarities between Documents for Coherence Assessment with Graph Convolutional Networks. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, Ontario, Canada, July, pp.7792-7808

**Liu W, Fan Y, Strube M (2023b).** HITS at DISRPT 2023: Discourse segmentation, connective detection, and relation classification. Proceedings of the 3rd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT ), Toronto, Ontario, Canada, 14 July, pp.43–49

**Liu W, Strube M (2023c).** Annotation-Inspired Implicit Discourse Relation Classification with Auxiliary Discourse Connective Generation. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, Ontario, Canada, July, pp.15696-15712

**Llenga S, Gryn'ova G (2023).** Matrix of orthogonalized atomic orbital coefficients representation for radicals and ions. The Journal of Chemical Physics 158(21), 214116

**Machmeier S, Hoecker M, Heuveline V (2023).** Explainable Artificial Intelligence for Improving a Session-Based Malware Traffic Classification with Deep Learning. IEEE Symposium Series on Computational Intelligence (SSCI), pp.850-855, IEEE

**Machmeier S, Trageser M, Buchwald M, Heuveline V (2023).** A Generalizable Approach for Network Flow Image Representation for Deep Learning. 7th Cyber Security in Networking Conference (CSNet), pp.57-61, IEEE

**Maltsev K (2023).** Stellar gravitational collapse, singularity formation and theory breakdown. The Sixteenth Marcel Grossmann Meeting, WORLD SCIENTIFIC, Online, pp.3596-3625

**Mayer G, Golebiewski M (2023).** Standardization landscape, needs and gaps for the virtual human twin (VHT). https://doi.org/10.5281/zenodo.10492796

**Mayer G, Golebiewski M (2023).** EDITH standards implementation guide (IG). https://doi.org/10.5281/zenodo.10524795

**Meineke F, Golebiewski M, Hu X, Kirsten T, Löbe M, Klammt S, Sax U, Müller W (2023).** NFDI4Health Local Data Hubs for Finding and Accessing Health Data: Making Distributed Data Accessible Through a SEEK-Based Platform. Proc Conf Res Data Infrastr 1.

**Meyer J, Alpay A, Hack S, Fröning H, Heuveline V (2023).** Implementation Techniques for SPMD Kernels on CPUs. International Workshop on OpenCL, pp.1-12, ACM

**Mobasher G, Müller W, Krebs O, Gertz M (2023).** In The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks, pages 427–438, Toronto, Canada. Association for Computational Linguistics.

**Montisci F, Ernst M, Macchi P (2023).** Experimental and Computational Study on the Effects of High Pressure on the Crystal Structure of Boron Nitrilotriacetate. Crystal Growth & Design 23(4):2745-2754

**Morán-Fraile J, Roepke FK, Pakmor R, Aloy MA, Ohlmann ST, Schneider FRN, Leidi G (2023a).** Self-consistent MHD simulation of jet launching in a neutron star - white dwarf merger. arXiv:2310.08623 [astro-ph.HE]

**Morán-Fraile J, Schneider FRN, Röpke FK, Ohlmann ST, Pakmor R, Soultanis T, Bauswein A (2023b).** Gravitational wave emission from dynamical stellar interactions. A&A

**Morán-Fraile J, Holas A, Röpke FK, Pakmor R, Schneider FRN (2023c).** Faint calcium-rich transient from a double detonation of a 0.6 M ⊙ carbon-oxygen white dwarf star. A&A 683:A44

**Morel B, Williams TA, Stamatakis A, Szöllősi GJ (2023).** AleRax: A tool for gene and species tree co-estimation and reconciliation under a probabilistic model of gene duplication, transfer, and loss. Bioinformatics. bioRxiv. https://doi.org/10.1101/2023.10.06.561091

**Morel B, Williams TA, Stamatakis A (2023).** Asteroid: a new algorithm to infer species trees from gene trees under high proportions of missing data. Bioinformatics 39(1), btac832

**Mukesh S, Mukherjee G, Singh R, Steenbuck N, Demidova C, Joshi P, Sangamwar AT, Wade RC (2023).** Comparative analysis of drug-salt-polymer interactions by experiment and molecular simulation improves biopharmaceutical performance. Commun Chem 6(1), 201

**Muñiz-Chicharro A, Votapka LW, Amaro RE, Wade RC (2023).** Brownian dynamics simulations of biomolecular diffusional association processes. WIREs Comput Mol Sci 13(3), e1649

**Møller P, Seppälä TT, Ahadova A, Crosbie EJ, Holinski-Feder E, Scott R, Haupt S, Möslein G, Winship I, Broeke SWB, Kohut KE, Ryan N, Bauerfeind P, Thomas LE, Evans DG, Aretz S, Sijmons RH, Half E, Heinimann K, Horisberger K, Monahan K, Engel C, Cavestro GM, Fruscio R, Abu-Freha N, Zohar L, Laghi L, Bertario L, Bonanni B, Tibiletti MG, Lino-Silva LS, Vaccaro C, Valle AD, Rossi BM, da Silva LA, de Oliveira Nascimento IL, Rossi NT, Dębniak T, Mecklin J-P, Bernstein I, Lindblom A, Sunde L, Nakken S, Heuveline V, Burn J, Hovig E, Kloor M, Sampson JR, Dominguez-Valentin M (2023).** Dominantly inherited micro-satellite instable cancer – the four Lynch syndromes - an EHTG, PLSD position statement. Hered Cancer Clin Pract 21(1),19

**Müller W, Mertová L (2023).** ReStoRunT: Simple Recording, Storing, Running and Tracing changes in Spreadsheets. In: König-Ries, B., Scherzinger, S., Lehner, W. & Vossen, G. (Hrsg.), BTW. Gesellschaft für Informatik e.V. doi:10.18420/BTW-57

**Noll A, Deheuvels S (2023).** Impact of central mixing scheme and nuclear reaction network on the extent of convective cores. A&A 676:A70

**Pozo Nuñez F, Gianniotis N, Polsterer KL (2023).** A Gaussian process cross-correlation approach to time-delay estimation in active galactic nuclei. A&A 674:A83

**Pozo Nuñez F, Bruckmann C, Deesamutara S, Czerny B, Panda S, Lobban AP, Pietrzyński G, Polsterer KL (2023).** Modeling photometric reverberation mapping data for the next generation of big data surveys. Quasar accretion disks sizes with the LSST. Monthly Notices of the Royal Astronomical Society, stad286

**Ono M, Nozawa T, Nagataki S, Kozyreva A, Orlando S, Miceli M, Chen K-J (2023).** The impact of effective matter mixing based on three-dimensional hydrodynamical models on the molecule formation in the ejecta of SN 1987A. arXiv e-prints:arXiv:2305.02550

**Perez S, Makshakova O, Angulo J, Bedini E, Bisio A, de Paz JL, Fadda E, Guerrini M, Hricovini M, Hricovini M, Lisacek F, Nieto PM, Pagel K, Paiardi G, Richter R, Samsonov SA, Vivès RR, Nikitovic D, Ricard Blum S (2023).** Glycosaminoglycans: What Remains To Be Deciphered? JACS Au 3(3):628-656

**Psonis N, Vassou D, Nafplioti A, Tabakaki E, Pavlidis P, Stamatakis A, Poulakakis N (2023).** Identification of the 18 World War II Executed Citizens of Adele, Rethymnon, Crete Using an Ancient DNA Approach and Low Coverage Genomes. SSRN

**Rahman TM, Kristian Koski M, Panecka-Hofman J, Schmitz W, Kastaniotis AJ, Wade RC, Wierenga RK, Kalervo Hiltunen J, Autio KJ (2023).** An engineered variant of MECR reductase reveals indispensability of long-chain acyl-ACPs for mitochondrial respiration. Nat Commun 14(1), 619

**Rasheeda Satheesh A, Knippertz P, Fink AH, Walz E, Gneiting T (2023).** Sources of predictability of synoptic-scale rainfall during the West African summer monsoon. Quarterly Journal of the Royal Meteorological Society, 149(757):3721–3737

**Ray EL, Brooks LC, Bien J, Biggerstaff M, Bosse NI, Bracher J, Cramer EY, Funk S, Gerding A, Johansson MA, Rumack A, Wang Y, Zorn M, Tibshirani RJ, Reich NG (2023).** Comparing trained and untrained probabilistic ensemble forecasts of COVID-19 cases and deaths in the United States. International Journal of Forecasting, 39(3):1366–1383

**Remme R, Kaczun T, Scheurer M, Dreuw A, Hamprecht FA (2023).** KineticNet: Deep learning a transferable kinetic energy functional for orbital-free density functional theory. arXiv,physics.chem-ph,2305.13316

**Rennekamp B, Karfusehr C, Kurth M, Ünal A, Monego D, Riedmiller K, Gryn'ova G, Hudson DM, Gräter F (2023).** Collagen breaks at weak sacrificial bonds taming its mechanoradicals. Nat Commun 14, 2075

**Resin J (2023b).** From classification accuracy to proper scoring rules: Elicitability of probabilistic top list predictions. Journal of Machine Learning Research, 24(173):1–21

**Resin J (2023).** A simple algorithm for exact multinomial tests. Journal of Computational and Graphical Statistics, 32(2):539–550

**Roussaki M, Magoulas GE, Fotopoulou T, Santarem N, Barrias E, Pöhner I, Luelmo S, Afroudakis P, Georgikopoulou K, Nevado PT, Eick J, Bifeld E, Corral MJ, Jiménez-Antón MD, Ellinger B, Kuzikov M,**

**Fragiadaki I, Scoulica E, Gul S, Clos J, Prousis KC, Torrado JJ, Alunda JM, Wade RC, de Souza W, Cordeiro da Silva A, Calogeropoulou T (2023).** Design, synthesis and biological evaluation of antiparasitic dinitroaniline-ether phospholipid hybrids. Bioorganic Chemistry 138:106615

**Röpke FK, De Marco O (2023).** Simulations of common-envelope evolution in binary stellar systems: physical models and numerical techniques. Living Rev Comput Astrophys 9(1), 2

**Röpke FK, Lach F, Gronow S, Sim SA, Callan FP, Collins CE (2023).** Modeling Type Ia supernovae with explosions in white dwarfs near and below the Chandrasekhar mass. In: The Sixteenth Marcel Grossmann Meeting. WORLD SCIENTIFIC, Online, pp 4420–4426

**Schneider FRN, Podsiadlowski P, Laplace E (2023).** Bimodal Black Hole Mass Distribution and Chirp Masses of Binary Black Hole Mergers. ApJL 950(2):L9

**Schröder M, Machmeier S, Heuveline V (2023).** Vtable hijacking: Object Type Integrity for run- time type information. Preprint Series of the Engineering Mathematics and Computing Lab :Nr. 1 (2023): Vtable hijacking: Object Type Integrity for runtime type information. https://doi.org/10.11588/EMCLPP.2023.1.94354

**Schröter J, Dattner T, Hüllein J, Jayme A, Heuveline V, Hoffmann GF, Kölker S, Lenz D, Opladen T, Popp B, Schaaf CP, Staufner C, Syrbe S, Uhrig S, Hübschmann D, Brennenstuhl H (2023).** aRgus: Multilevel visualization of non-synonymous single nucleotide variants & advanced pathogenicity score modeling for genetic vulnerability assessment. Computational and Structural Biotechnology Journal 21:1077-1083

**Schulze S, Fransson C, Kozyreva A, Chen T-W, Yaron O, Jerkstrand A, Gal-Yam A, Sollerman J, Yan L, Kangas T, Leloudas G, Omand CMB, Smartt SJ, Yang Y, Nicholl M, Sarin N, Yao Y, Brink TG, Sharon A, Rossi A, Chen P, Chen Z, Cikota A, De K, Drake AJ, Filippenko AV, Fremling C, Freour L, Fynbo JPU, Ho AYQ, Inserra C, Irani I, Kuncarayakti H, Lunnan R, Mazzali P, Ofek EO, Palazzi E, Perley DA, Pursiainen M, Rothberg B, Shingles LJ, Smith K, Taggart K, Tartaglia L, Zheng W, Anderson JP, Cassara L, Christensen E, Djorgovski SG, Galbany L, Gkini A, Graham MJ, Gromadzki M, Groom SL, Hiramatsu D, Howell DA, Kasliwal MM, McCully C, Müller-Bravo TE, Paiano S, Paraskeva E, Pessi PJ, Polishook D, Rau A, Rigault M, Rusholme B (2023)**. 1100 Days in the Life of the Supernova 2018ibb – the Best Pair-Instability Supernova Candidate, to date. arXiv e-prints:arXiv:2305.05796

**Srivastav S, Moore T, Nicholl M, Magee MR, Smartt SJ, Fulton MD, Sim SA, Pollin JM, Galbany L, Inserra C, Kozyreva A, Moriya TJ, Callan FP, Sheng X, Smith KW, Sommer JS, Anderson JP, Deckers M, Gromadzki M, Müller-Bravo TE, Pignata G, Rest A, Young DR (2023).** Unprecedented early flux excess in the hybrid 02es-like type Ia supernova 2022ywc indicates interaction with circumstellar material. arXiv e-prints:arXiv:2308.06019

**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, Orinion 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, Schrader D, 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, Wingfield 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, Batchelder J, Berant J, Frohberg J, Rozen J, Hernandez-Orallo J, Boudeman J, Guerr 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, Walker M, Tiwari M, Bansal M, Aminnaseri M, Geva M, Gheini M, T MV, Peng N, Chi NA, Lee N, Krakover NG-A, Cameron N, Roberts N, Doiron N, Martinez 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, Risco R, 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, 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 (2023).** Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. arXiv:2206.04615 [cs.CL]

**Strube M, Braud C, Hardmeier C, Li JJ, Loaiciga S, Zeldes A (eds) (2023).** Proceedings of the 4th Workshop on Computational Approaches to Discourse. Proceedings of the 4th Workshop on Computational Approaches to Discourse, Toronto, Ontaria, Canada, 13-14 July

**Tempest KI, Craig GC, Brehmer JR (2023).** Convergence of forecast distributions in a 100,000-member idealised convective-scale ensemble. Quarterly Journal of the Royal Meteorological Society, 149(752):677–702

**Tiwari DD, Hoffmann N, Didi K, Deshpande S, Ghosh S, Nguyen TVN, Raman K, Hermjakob H, Sheriff R (2023).** BioModelsML: Building a FAIR and reproducible collection of machine learning models in life sciences and medicine for easy reuse. bioRxiv. https://doi.org/10.1101/2023.05.22.540599

**Togkousidis A, Kozlov OM, Haag J, Höhler D, Stamatakis A (2023).** Adaptive RAxML-NG: Accelerating Phylogenetic Inference under Maximum Likelihood using Dataset Difficulty. Molecular Biology and Evolution 40(10)

**Togkousidis A, Chernomor O, Stamatakis A (2023).** Parallel Inference of Phylogenetic Stands with Gentrius. IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp.139-148, IEEE

**Trost J, Haag J, Höhler D, Jacob L, Stamatakis A, Boussau B (2023).** Simulations of sequence evolution: how (un)realistic they are and why. Bioinformatics. bioRxiv. https://doi.org/10.1101/2023.07.11.548509

**Varma V, Müller B, Schneider FRN (2023).** 3D simulations of strongly magnetized non-rotating supernovae: explosion dynamics and remnant properties. Monthly Notices of the Royal Astronomical Society 518(3):3622-3636

**Vink JS, Mehner A, Crowther PA, Fullerton A, Garcia M, Martins F, Morrell N, Oskinova LM, St-Louis N, ud-Doula A, Sander AAC, Sana H, Bouret J-C, Kubátová B, Marchant P, Martins LP, Wofford A, van Loon JTh, Grace Telford O, Götberg Y, Bowman DM, Erba C, Kalari VM, Abdul-Masih M, Alkousa T, Backs F, Barbosa CL, Berlanas SR, Bernini-Peron M, Bestenlehner JM, Blomme R, Bodensteiner J, Brands SA, Evans CJ, David-Uraz A, Driessen FA, Dsilva K, Geen S, Gómez-González VMA, Grassitelli L, Hamann W-R, Hawcroft C, Herrero A, Higgins ER, John Hillier D, Ignace R, Istrate AG, Kaper L, Kee ND, Kehrig C, Keszthelyi Z, Klencki J, de Koter A, Kuiper R, Laplace E, Larkin CJK, Lefever RR, Leitherer C, Lennon DJ, Mahy L, Maíz Apellániz J, Maravelias G, Marcolino W, McLeod AF, de Mink SE, Najarro F, Oey MS, Parsons TN, Pauli D, Pedersen MG, Prinja RK, Ramachandran V, Ramírez-Tannus MC, Sabhahit GN, Schootemeijer A, Reyero Serantes S, Shenar T, Stringfellow GS, Sudnik N, Tramper F, Wang L (2023).** X-Shooting ULLYSES: Massive stars at low metallicity. I. Project description. A&A 675:A154

**Viviani LG, Kokh DB, Wade RC, T.-do Amaral A (2023).** Molecular Dynamics Simulations of the Human Ecto-5'-Nucleotidase (h-ecto-5'-NT, CD73): Insights into Protein Flexibility and Binding Site Dynamics. J. Chem. Inf. Model. 63(15):4691-4707

**Williams TA, Davin AA, Morel B, Szánthó LL, Spang A, Stamatakis A, Hugenholtz P, Szöllősi GJ (2023).** The power and limitations of species tree-aware phylogenetics. bioRxiv. https://doi.org/10.1101/2023.03.17.533068

**Wittner R, Holub P, Mascia C, Frexia F, Müller H, Plass M, Allocca C, Betsou F, Burdett T, Cancio I, Chapman A, Chapman M, Courtot M, Curcin V, Eder J, Elliot M, Exter K, Goble C, Golebiewski M, Kisler B, Kremer A, Leo S, Lin-Gibson S, Marsano A, Mattavelli M, Moore J, Nakae H, Perseil I, Salman A, Sluka J, Soiland-Reyes S, Strambio-De-Castillia C, Sussman M, Swedlow JR, Zatloukal K, Geiger J (2023).** Toward a common standard for data and specimen provenance in life sciences. Learning Health Systems, e10365

**Wolffram D, Abbott S, an der Heiden M, Funk S, Günther F, Hailer D, Heyder S, Hotz T, van de Kassteele J, Küchenhoff H, Müller-Hansen S, Syliqi D, Ullrich A, Weigert M, Schienle M, Bracher J (2023).** Collaborative nowcasting of COVID-19 hospitalization incidences in Germany. PLOS Computational Biology, 19(8):e1011394

**Yu J, Khanna S, Themessl N, Hekker S, Dréau G, Gizon L, Bi S (2023).** Revised Extinctions and Radii for 1.5 Million Stars Observed by APOGEE, GALAH, and RAVE. ApJS 264(2):41

**Zaunseder E, Mütze U, Garbade SF, Haupt S, Kölker S, Heuveline V (2023).** Deep Learning and Explainable Artificial Intelligence for Improving Specificity and Detecting Metabolic Patterns in Newborn Screening. IEEE Symposium Series on Computational Intelligence (SSCI), pp.1566-1571, IEEE

**Zaunseder E, Mütze U, Okun JG, Hoffmann GF, Kölker S, Heuveline V, Thiele I (2023).** Personalised metabolic whole-body models for newborns and infants predict growth and biomarkers of inherited metabolic diseases. bioRxiv 2023.10.20.563364

**Zaunseder E, Mütze U, Garbade SF, Haupt S, Feyh P, Hoffmann GF, Heuveline V, Kölker S (2023).** Machine Learning Methods Improve Specificity in Newborn Screening for Isovaleric Aciduria. Metabolites 13(2):304

**Zhao W, Strube M, Eger S (2023a).** DiscoScore: Evaluating text generation with BERT and discourse coherence. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, Dubrovnik, Croatia, May, pp.3865-3883

**Zhao W, López F, Riestenberg JM, Strube M, Taha D, Trettel S (2023b).** Modeling graphs beyond hyperbolic: Graph neural networks in symmetric positive definite matrices. Proceedings of Machine Learning and Knowledge Discovery in Databases: Research Track - European Conference. ECML PKDD, Turin, Italy, 18-22 September, pp.122-139

**Zharov Y, Ametova E, Spiecker R, Baumbach T, Burca G, Heuveline V (2023).** Shot noise reduction in radiographic and tomographic multi-channel imaging with self-supervised deep learning. Opt. Express 31(16):26226



# 9 Teaching

## Degrees

**Christina Athanasiou:**
*"Modeling of mechanistic effects of neurotrophin modulators"*. PhD thesis, Combined Faculty of Mathematics, Engineering and Natural Sciences, Heidelberg University and HITS: Rebecca C. Wade (2023).

**Ben Bettisworth:**
*"Uncommon Problems in Phylogenetic Inference"*, Ph.D. thesis, Karlsruhe Institute of Technology, Karlsruhe and HITS: Alexandros Stamatakis (2023).

**Matthias Brosz:**
*"Coarse-grained molecular dynamics of semi-flexible (bio-)polymers under force"*, PhD thesis, Faculty of Engineering Sciences, Heidelberg University and HITS: Frauke Gräter (2023).

**Johannes Cepicka:**
*"Uncertainty Quantification for an extended Lotka-Volterra model of tumor-immune interaction"*, Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**Felix Daas:**
*"Analyzing schemes for secure and memorable password generation"*, Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**Ben Dangelmayr:**
*"Identifying the heart region of Medaka using deep convolution neural networks for object detection"*, Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**Christina Goß:**
*"Simulating interactions between intrinsically disordered proteins and lipid bilayers: the case of Abl kinase"*, M.Sc. Thesis, Department of Physics, University of Göttingen, HITS: Frauke Gräter (2023).

**Saskia Haupt:**
*"Mathematical modeling of Lynch syndrome carcinogenesis"*, Ph.D. thesis, Heidelberg University and HITS: Vincent Heuveline (2023).

**Luise Häuser:**
*"Quantitative Analysis and Characterization of Natural Language Evolution Datasets"*, Master's thesis, Institute of Theoretical Computer Science, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2023).

**Melanie Kaeser:**
*"Computational Investigation of the Structure Kinetics Relationship of Small-Molecule Compounds Binding the Histamine-1-Receptor"*, Master's Thesis, Molecular Biotechnology, Faculty of Biosciences, Heidelberg University and HITS: Giulia D'Arrigo and Rebecca C. Wade (2023).

**Veton Kajtazi:**
*"Deep learning based tabular dataset synthetization for data privacy protection"*, Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heu-veline (2023).

**Lukas Knirsch:**
*"Distanzbasierte Inferenz von Genbäumen unter Berücksichtigung ihres Speziesbaumes"*, Bachelor's thesis, Institute of Theoretical Computer Science, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2023).

**Pascal Memmesheimer:**
*"Adding Interpretability to an anomaly-based method for deep packet inspection in intrusion de-tection systems"*, Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**Aina Nambena:**
*"Describing the mass distribution of black holes in X-ray binaries using Bayesian inference"*, Bachelor's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Fabian Schneider (2023).

**Jakob Niessner:**
*"Computational Models of Hen Egg White Lysozyme Physisorption to Silica- and Mica-like Surfaces"*, Master's thesis, Physics, Faculty of Physics, Heidelberg University and HITS: Abraham Muniz-Chicharro and Rebecca C. Wade (2023).

**Dennis Pfleger:**
*"Implementing an exploit as a Metasploit module and investigating the exploit ranking mechanism"*, Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**Lea Reisinger:**
*"Using DPI to analyse and reduce DDoS attacks on servers or applications"*, Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heu-veline (2023).

**Benedikt Rennekamp:**
*"Multi-Scale Simulations of Collagen Failure and Mechanoradicals"*, PhD thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Frauke Gräter (2023).

**Johannes Resin:**
*"Model Diagnostics meets Forecast Evaluation: Goodness-of-Fit, Calibration, and Related Topics"*, Ph.D. thesis, Department of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting (2023).

**Purusharth Saxena:**
*"Mathematical modeling of colorectal cancer: Exploring spatial heterogeneity of tumor microen-vironment with random fields and Fischer KPP"*, Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**Felix Schmäling:**
*"Analyzing the BrakTooth family of experimental attacks on specific Bluetooth chipsets"*, Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**Lea Schmierer:**
*"Design-Prinzipien für Softwarearchitekturen für Softwarearchitekturen mit erhöhtem Schutzbedarf"*, Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**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: Vincent Heuveline, Kai Polsterer (2023).

**Benedikt Schulz:**
*"Statistical Postprocessing of Numerical Weather Prediction Forecasts using Machine Learning"*, Ph.D. thesis, Department of Mathematics, Karlsruhe Institute of Technology and HITS: Sebastian Lerch and Tilmann Gneiting (2023).

**Leif Seute:**
*"Building Grappa, a Machine Learned Molecular Mechanics Force Field for Proteins"*, M. Sc. thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Frauke Gräter (2023).

**Olivia Stanossek:**
*"Vulnerabilities in medical data transfer based on blockchain technologies"*, Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heu-veline (2023).

**Jan Strehmel:**
*"Is there a Correlation between the Use of Swearwords and Code Quality in Open Source Code?"*, Bachelor's thesis, Institute of Theoretical Computer Science, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2023).

**Michel Tarnow:**
*"Optimization of a Drug:Polymer Formulation Exploiting a Computational Approach Based on Molecular Dynamics Simulations"*. Bachelor's Thesis. Molecular Biotechnology, Faculty of Engineering Science, Heidelberg University and HITS: Giulia Paiardi and Rebecca Wade (2023).

**Manuel Trageser:**
*"Flow-based traffic classification using deep vision"*, Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

**Evgeni Ulanov:**
*"Predicting Hydrogen Atom Transfer Activation Energies using Gaussian Process Regression"*, M. Sc. thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Frauke Gräter and Tilmann Gneiting (2023).

**Noah Wahl:**
*"Approximating Phylogenetic Tree Distributions with Distance-Based Methods"*, Master's thesis and HITS: Alexandros Stamatakis, Institute of Theoretical Computer Science, Karlsruhe Institute of Technology (2023).

**Holger Wünsche:**
*"Evaluation of OpenMP as device backend in hipSYCL"*, Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2023).

## Lectures, courses and seminars

**Vincent Bronner:**
Tutor for the course *"Computational Astrophysics"* (lecturer: Friedrich Röpke), Heidelberg University, summer semester 2023, Tutor for the course *"Einführung in die Astronomie und Astrophysik I"* (lecturer: Cornelis Dullemond), Heidelberg University, winter semester 2023/2024.

**Rostislav Fedorov:**
Neural networks training at the *"MOQS Heidelberg Workshop: from platforms to algorithms for quantum simulation"*, Internationales Wissenschaftsforum Heidelberg, 23-24 March 2023.

**Tilmann Gneiting:**
Lecture on *"Forecasting: Theory and Practice I"*, Karlsruhe Institute of Technology, winter semester 2022/2023.
Lecture on *"Forecasting: Theory and Practice II"*, Karlsruhe Institute of Technology, summer semester 2023.

**Tilmann Gneiting and Johannes Resin:**
Seminar on *"Statistical Forecasting and Classification"*, Karlsruhe Institute of Technology, winter semester 2023/2024.

**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 2023.

**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, 2022/2023.



**Ganna Gryn'ova, Rostislav Fedorov, and Christopher Ehlert:**
Special Lecture Course *"Applied Computational Chemistry"*, Heidelberg University, summer semester 2023.

**Saskia Hekker:**
Lecture on *"Asteroseismology"*, WS 2022/23. Seminar series on *"Applications of asteroseismology"*, Heidelberg University, summer semester 2023/2024. Lecture on *"Asteroseismology"*, Heidelberg University, winter semester 2023/2024.

**Jan Henneco:**
Tutor for course *"Fundamentals of Simulation Methods"*, Heidelberg University (lecturers: Ralf Klessen and Philipp Girichidis), winter semester 2023/2024.

**Vincent Heuveline:**
Seminar on *"Mathematical modeling in cancer research"*, Heidelberg University, winter term 2022/23. Lecture on *"IT-Sicherheit 1"*, Heidelberg University, winter term 2022/2023. Lecture on *"Finite Elements"*, Heidelberg University, summer term 2023. Lecture on *"IT-Sicherheit 2"*, Heidelberg University, summer term 2023. Lecture on *"IT-Sicherheit 1"*, Heidelberg University, winter term 2023/2024. Seminar on *"IT-Security"*, Heidelberg University, winter term 2023/2024. Seminar on *"Special Topics in FEM"*, Heidelberg University, winter term 2023/2024.

**Xiaoming Hu, Haitham Abaza:**
Tutorial on FAIR data management tools in biomedical research, 68. GMDS-Jahrestagung, Heilbronn (Germany), 17-21 September 2023.

**Giovanni Leidi:**
*"Fundamentals of simulation methods"*, tutorials accompanying the lecture course, Heidelberg University, winter semester 2023/2024.

**Kiril Maltsev:**
*"Stellar Astrophysics"* tutorial, Department of Physics and Astronomy, Heidelberg University, summer semester 2023.

**Giulia Paiardi:**
*"Meet-EU Heidelberg-team"*, European Bioinformatics Master Network (https://cu-bioinformatics.github.io/meet-eu-2023/), winter semester, 2023/2024.

**Kai Lars Polsterer:**
*"An interdisciplinary view on machine learning in Astronomy."* Workshop on *"Machine Learning in Astronomy"*, 1st Meeting of the Astronomical Society of India, IIT Indore, IIT, Indore, India, 1-5 March 2023.

**Maja Rey, Andreas Weidemann, Ulrike Wittig:**
de.NBI Course *"Tools for Systems biology modeling and data exchange: COPASI, CellNetAnalyzer, SABIO-RK, FAIRDOMHub/SEEK"*, Magdeburg, Germany, 28-30 November 2023.

**Friedrich Röpke:**
Lecture course *"Computational Astrophysics"*, Heidelberg University, summer semester 2022. Seminar on *"Physics of Stellar Objects"*, Heidelberg University, winter semester 2022/2023, summer semester 2023, and winter semester 2023/2024.

**Friedrich Röpke, Fabian Schneider:**
Lecture course *"The Stellar Cookbook: A practical guide to the theory of stars"*, Heidelberg University, winter semester 2022/2023, and winter semester 2023/2024.

**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, Lukas Hübner:**
Joint UoC-KIT Lecture *"Introduction to Bioinformatics for Computer Scientists"*, computer science Master's program at Karlsruhe Institute of Technology and computer science Master's program at Uiniversity of Crete, winter semester 2023/2024.

**Alexandros Stamatakis, Benoit Morel, Alexey Kozlov, Lukas Hübner, Anastasis Togkousidis, Julia Haag, Dimitri Höhler, Ben Bettisworth:**
Seminar *"Hot Topics in Bioinformatics"*, computer science Master's program at Karlsruhe Institute of Technology, summer semester 2023.

**Alexandros Stamatakis, Lukas Hübner:**
Summer School *"Computational Molecular Evolution"*, Hellenic Center for Marine Research, Crete, Greece, May 2023.

**Alexandros Stamatakis:**
*"Optimizing and Parallelizing Phylogenetic Likelihood Calculations"*, guest lecture, ECE 284 graduate level course on *"Parallel Computing in Bioinformatics"* (ECE 284), University of California at San Diego, March 2023.

**Michael Strube:**
PhD Colloquium, Department of Computational Linguistics, Heidelberg University (winter semester 2022/2023). Seminar *"Ethics in NLP: Bias and Dual Use"* (together with Prof. Shimei Pan, PhD.), Department of Computational Linguistics, Heidelberg University (winter semester 2022/2023). PhD Colloquium, Department of Computational Linguistics, Heidelberg University (summer semester 2023).

**Jan Stühmer:**
Geometric Deep Learning (compact course), Physics Master program, Heidelberg university, summer semester 2023. Geometric Deep Learning, Computer Science Master program, Karlsruhe Institute of Technology, winter semester 2023/2024.

**Marco Vetter:**
Tutorial on *„Theoretical Astrophysics"* within Master program of Physics, Heidelberg University, winter semester 2023/2024.

**Rebecca C. 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, Heidelberg University.

**Rebecca C. Wade, Stefan Richter, and Manuel Glaser:**
BSc. lecture and practical course on Bioinformatics, Heidelberg University, winter semester 2022/23.

**Ulrike Wittig:**
COST Action OneHealthDrugs Course *"FAIRDOM-SEEK Introduction"*, online, 8 March 2023. *"Research Data Management Guidelines - RDMkit"*, de.NBI Spring School *"Data Management"*, Gatersleben, Germany ,13-17 March 2023. *"FAIRDOM-SEEK Introduction"*, de.NBI Spring School *"Data Management"*, Gatersleben, Germany ,13-17 March 2023.

**Daniel Wolffram:**
Tutorials on *"Nicht- und Semiparametrik"*, Karlsruhe Institute of Technology, winter semester 2022/2023.

# 10 Miscellaneous

## 10.1 Guest Speaker Activities (invited talks):

**Felix Ahlborn:**
*"Turbulent convection in stellar structure and evolution models"*, Kavli summer-program in Astrophysics, MPA Garching, Germany, 28 June 2023. *"Turbulent convection models in stellar evolution"*, Workshop on the Modelling of geophysical and stellar flows, Wolfgang Pauli Institute, Vienna, Austria, 30 August 2023. *"Accurate internal rotation rates of red-giant stars"*, 11th Applied Inverse Problems Conference, University of Goettingen, Germany, 5 September 2023.

**Aksel Alpay:**
*"HipSYCL's Quest for Universal SYCL Binaries: One Binary from NVIDIA and AMD to Intel Data Center GPU Max Series"*, oneAPI Dev Summit 2023, 13 June 2023.

**Michaël Bazot:**
*"On the nature of the core of alpha Cen"*, The alpha Centauri system -- Towards new worlds, Nice, France, 26 June 2023.

**Lucas Czech and Alexandros Stamatakis:**
*"Leveraging Phylogenetic Placement to understand the Environmental Drivers of Microbial Community Composition"*, XL Annual Meeting of the Willi Hennig Society, Cornell University, USA, July 2023 (talk given by Lucas Czech).

**Christopher Ehlert:**
*"Metal-Free Molecular Catalysts for the Oxygen Reduction Reaction"*, Institute of Chemistry, University of Potsdam, Germany, 1 February 2023.

**Nikos Gianniotis:**
*"Probabilistic Cross Correlations for Density Estimation"*, Astroinformatics 2023, INAF National Auditorium Observatory of Capodimonte, Naples, Italy, 1-6 October 2023.

**Tilmann Gneiting:**
*"Isotonic distributional regression"*, Kolloquium Mathematische Statistik, Heidelberg University, Heidelberg, Germany, 26 January 2023.

**Martin Golebiewski:**
*"Data and metadata standards for personalised medicine research"*, EU Africa PerMed Summer School *"Standards in personalised medicine research"*, Cape Town (South Africa) and online, 22-23 February 2023. *"Standardization Activities of ISO/TC 276/WG 5 Data Processing and Integration"*, CEN Focus Group *"Organ-on-Chip"*, online, 29 March 2023.

**Ganna Gryn'ova:**
*"Engineering Organic Frameworks for Molecular Storage"*, Flagship Initiative *"Engineering Molecular Systems"* Colloquia Series, Heidelberg University, Germany, 24 April 2023; Mulliken Center for Theoretical Chemistry, University of Bonn, Bonn, Germany, 25 May 2023. *"New Representations for Chemical Machine-Learning"*, CECAM-Psi-k Research Conference *"Bridging Length Scales with Machine Learning: From Wavefunctions to Thermodynamics"*, Berlin, Germany, 19-23 June 2023; European Conference on Computational and Theoretical Chemistry (EuChemS CompChem 2023), Thessaloniki, Greece, 27-31 August 2023. *"Exploring Chemical Space with Machine Learning"*, Interdisciplinary Seminar Series, TU Braunschweig, Braunschweig, Germany, 8 November 2023.

**Julia Haag:**
*"Predicting the Difficulty of Phylogenetic Analyses"*, ERGA BioGenome Analysis and Applications on-line Seminar, November 2023.

**Saskia Haupt:**
*"The Importance of Female Role Models in STEM"*, Invited panelist in the round table, CRC1225 ISOQUANT Equal Opportunities Team, KAMERA Kino Heidelberg, 8 March 2023. *"How to use the Kronecker structure for modeling mutational rates and multiple pathways of cancer development."* CCO Seminar, online, 22 March 2023.

**Luise Häuser and Alexandros Stamatakis:**
*"General and Language-specific Aspects of Phylogenetic Inference"*, invited talk, Rethymno, Crete, May 2023.

**Saskia Hekker:**
*"The power of asteroseismology: stellar structure revealed by global oscillations"*, colloquium at TU Dresden, Germany, 9 May 2023. *"Red-giant stars and their many different flavours"*, MERAC-symposium, EAS week, ICE Krakow, Poland, 10-14 July 2023.

**Vincent Heuveline:**
*"Künstliche Intelligenz in der Medizin: Revolution mit Nebenwirkungen"*, invited talk at the Fraunhofer-Institut für Techno- und Wirtschaftsmathematik (ITWM), 28 June 2023. *"Uncertainty Quantification (UQ) for medical models"*, Data Science & Health Lecture Series, Helmholtz Information and Data Science School for Health, 23 November 2023 (online).

**Benoit Morel:**
*"Models and methods for disentangling the complexity of gene and species evolution"*, CMMS seminar series, Frankfurt, Germany, October 2023.

**Anthony Noll:**
*"Asteroseismology of Red Clump stars"*, Institut d'Astrophysique Spatiale (IAS) Seminar, Université Paris Saclay, Orsay, France, 6 December 2023.

**Kai Lars Polsterer:**
*"Accessing Complex Structures with Unsupervised and Deep-Learning Techniques"*. Keynote at the annual meeting of the Astronomical Society of India, Indore, India, 3 March 2023. *"Exploring Data with Unsupervised Machine Learning"*, Keynote at NextGen Data Analysis Workshop 2023, HZB, Berlin, Germany, 2 May 2023. *"From Supervised to Unsupervised Machine Learning: lessons learned from learning machines"*, Colloquium, Dep. Of Astronomy, University of Geneva, Geneva, Switzerland, 5 June 2023. *"From Photometric Redshifts to Improved Weather Forecasts"*, Astroinformatics 2023, INAF National Auditorium Observatory of Capodimonte, Naples, Italy, 1-6 October 2023.

**Alexandros Stamatakis:**
*"The role of Computer Science in Evolutionary Biology"*, outreach talk (in Greek), Darwinian Monday lecture series, Heraklion, Crete, May 2023. *"Why Greece needs a dedicated female researchers/professors program"*, Hybrid seminar for gender dimension in research @ FORTH in Horizon Europe, Foundation for Research and Technology Hellas, Crete, Greece, July 2023. *"Quantifying Uncertainty in Evolutionary Analyses"*, keynote talk, Ecology And Evolutionary Biology Symposium, Istanbul, Turkey, July 2023. *"Necessary Reforms in the Greek Academic System"*, 1st Greek ERA chairs meeting, Foundation for Research and Technology Hellas, Crete, Greece, September 2023. *"RAxML: 20 χρόνια ανάπτυξης λογισμικού ανοιχτού κώδικα για την εξελικτική βιολογία"* - *"RAxML: 20 years of open source software development for evolutionary biology"*, keynote talk in Greek at Free and Open source Software Communities Meeting, University of Crete, Crete, Greece, October 2023.

**Anastasis Togkousidis:**
*"Adaptive RAxML-NG: Accelerating Phylogenetic Inference under Maximum Likelihood using dataset difficulty"*, Rencontres Bioinformatiques du Muséum National d'Histoire Naturelle 2023, Paris, France, May 2023; ERGA BioGenome Analysis and Applications on-line Seminar, November 2023.

**Rebecca C. Wade:**
*"Research in the Molecular and Cellular Modeling Group at HITS"*, ML galore, IWR, Heidelberg University, 19 January 2023. *"Bridging timescales to predict protein-ligand binding kinetics: Applications to G protein-coupled receptors"*, Forschungszentrum Jülich, Germany, 6-8 February 2023; Interdisciplinary conference on Mathematical Life Science, Bonn, Germany 17-20 April 2023; Integrative Structural Modelling talk (http://mbu.iisc.ac.in/IMSS/index.html), online, 13 June 2023; MepAnti Drug Research Symposium, Basel, 9 November 2023. *"On the druggability of transient pockets and their binding kinetics"* Trabita retreat, Trifels, Germany, 1-3 March 2023. *"On the interplay between protein conformational switching and binding: Insights from molecular simulations"*, SFB meeting on protein dynamics and conformational switching, Venice, Italy, 4-7 May 2023. *"Computing drug binding kinetics"*, 55th Course of the International School of Crystallography entitled: *"Structural Drug Design 2020: Biology, Chemistry and Computers"*, Erice, Italy, 2-10 June 2023. *"Computing Protein Binding Kinetics"*, Instituto Aggeu Magalhães, Oswaldo Cruz Foundation, Recife, Brazil, 28 June 2023. *"In silico exploration of the*

*dynamics of biomolecular interactions"*, Heidelberg University Thematic Research Network 'Molecular Mechanisms in Health and Disease – from understanding to engineering' (MINDS), online,10 July 2023.

**Elaine Zaunseder:**
*"Artificial Intelligence in Medicine"*, SDW Workshop *"KI - Kann man das Essen?!"*, Paris, France, 17-19 March 2023. *"Personalized metabolic whole-body models for newborns and infants"*, Recon4IMD meeting, online, 21 September 2023; Euregio meeting, Freiburg, Germany, 11 November 2023.

**Alexander Zeilmann:**
*"Petabyte-scale Image Analysis"*, KIT IPS Meeting, Karlsruhe, 13 June 2023.

## 10.2 Presentations

**Talks (Contributed talks)**

**Haitham Abaza:**
The NFDI4Health Metadata Schema, Tutorial on FAIR data management tools in biomedical research, 68. GMDS-Jahrestagung, Heilbronn, 17-21 September 2023.

**Felix Ahlborn:**
*"Accurate internal rotation rates of red-giant stars"*, MIAPP program Stellar Astrophysics in the Era of Gaia, Spectroscopic, and Asteroseismic Surveys, MIAPbP, Garching, Germany, 25 August 2023.

**Aksel Alpay:**
*"One Pass to Bind Them: The First Single-Pass SYCL Compiler with Unified Code Representation Across Backends"*, talk at the 2023 International Workshop on OpenCL (IWOCL '23). Association for Computing Machinery, New York, NY, USA, 19 April 2023.

**Michaël Bazot:**
*"An asteroseismic butterfly diagram"*, MIAPP program Stellar Astrophysics in the Era of Gaia, Spectroscopic, and Asteroseismic Surveys, MIAPbP, Garching, Germany, 23 August 2023.

**Mislav Brajkovic:**
*"Calculating residence time of Histamine H1 receptor-binding compounds using τRAMD."*, Heidelberg University FI EMS Young Scientists' Retreat, Annweiler am Trifels, Germany, 27 November 2023.

**Lynn Buchele:**
*"Sound speed inversions of an ensemble of low-mass main-sequence stars"*, PLATO Stellar Science Conference 2023, Milazzo, Italy, 26 June 2023. *"Structure inversions for sound speed differences in solar-like stars"*, 11th Applied Inverse Problems Conference, Goettingen, Germany, 8 September 2023.

**Marcus Buchwald:**
*"3D Causal Image Generation using Deep Graph Conditioned LDMs"*, DKFZ Summer School on Medical Physics 2023 in Chile: *"The role of imaging in the radiotherapy process"*, 11-15 December 2023, Santiago de Chile.

**Giulia D'Arrigo:**
*"Computation of Unbinding Rates and Mechanisms in Protein-Protein Systems"*. Affinity 2023, Lisbon, Portugal, 5-7 June 2023. Lecture and Practical on *"τRAMD calculations of protein-ligand dissociation rates"*. 2nd HBPMolSIM Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades", Mathematikon, Heidelberg University and Hybrid, 21-23 June 2023. *"Computing residence time and investigating dissociation mechanisms of protein-protein interactions"*. EMBO Workshop: Computational structural biology, Heidelberg, Germany. 6-9 December 2023.

**Christopher Ehlert:**
*"Metal-Free Molecular Catalysts for the Oxygen Reduction Reaction"*, Computational and Mathematical Methods in Science and Engineering, Rota, Spain, 3-8 July 2023.

**Michelle Ernst:**
*"Insights into the Nature of Host-Guest Interactions in Emergent Framework Materials"*, 1st Swiss Symposium on Materials Chemistry, Dübendorf, Switzerland, 14 June 2023.

**Rostislav Fedorov:**
*"Deterministic Graph Retrosynthetic Algorithm for COFs Separation"*, RDKit UGM 2023, Mainz, Germany, 20-22 September 2023.

**Martin Golebiewski:**
*"Analysing current landscape of standards, identifying needs and gaps"*, Plenary meeting of the Ecosystem Digital Twins in Healthcare (EDITH), Leuven (Belgium), 30 - 31 January 2023. *"NFDI4Health Standardisation and Interoperability Roadmap"*, NFDI4Health Workshop *"Standards for Health Data"*, TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung, Berlin (Germany), 20-21 February 2023. *"ISO/TC 276/WG 5 Data Processing and Integration"*, Committee Meetings of the German standardization committe *"Arbeitsausschuss Biotechnologie"*, DIN, Berlin (Germany), 24 February and 7 September 2023. *"Metadatenschema und Standardisierung-Roadmap für NFDI4Health"*, NFDI Community Workshop Standardisierung, online, 21 March 2023. *"Development of community-based guidelines and normative documents"*, Final meeting of EU-STANDS4PM, online, 23 March 2023. *"ISO/TC 276/WG 5 Data Processing and Integration - Report of the Convenor"*, Committee Meetings of ISO/TC 276/WG 5, Atlanta, Georgia (USA), 12-16 June and online, 4-8 December 2023, and ISO/TC 276 Biotechnology - Plenary Meeting, Atlanta, Georgia (USA), 17 June 2023. *"NFDI4Health Metadata model and Data Quality Aspects"*, NFDI-4Health General Assembly, Cologne (Germany), 6 July 2023. *"NFDI4Health Local Data Hubs for Finding and Accessing Health Data"*, 1st Conference on Research Data Infrastructure (CoRDI 2023), KIT, Karlsruhe (Germany), 12-14 September 2023. *"NFDI4Health Strategische Ziele: TA2 - Standards for FAIR Data"*, NFDI4Health Steering Committee Meeting, Göttingen (Germany), 14-15 November 2023.

**Martin Golebiewski, Wolfgang Müller:**
*"SEEK Use Case"*, Plenary meeting of the Ecosystem Digital Twins in Healthcare (EDITH), Leuven (Belgium), 30 - 31 January 2023. *"NFDI4Health Success Story: Domain overarching health data harmonization and metadata bridging"*, NFDI4Health Steering Committee Meeting, ZBMed, Cologne (Germany), 1 March 2023.

**Julia Haag:**
*"Predicting the Difficulty of Phylogenetic Analyses"*, Peder Sather/Invertomics Symposium *"Progress and Development in Phylogenetic Methods"*, University of Oslo, Norway, March 2023.

**Saskia Haupt:**
Invited workshop participant: Banff International Research Station (BIRS): Computational Modelling of Cancer Biology and Treatments, hybrid (Banff, Canada and online), 22 – 27 January 2023. *"Math4INDICATE - Methodological challenges for investigating the influence of the HLA genotype on cancer risk in Lynch syndrome."* 7th EHTG Meeting 2023, Vilnius, Lithuania, 29 September 2023 – 1 October 2023. *"Mathematical modeling of Lynch syndrome cancer development - A multi-scale view from the DNA over the tissue level to clinical translation."* 7th EHTG Meeting 2023, Vilnius, Lithuania, 29 September 2023 – 1 October 2023.

**Saskia Hekker:**
*"Effects of rapid mass loss just before the onset of He burning in stars with masses of 2-3 Msol"*, SDSS-V/IReNA/CeNAM Science Festival, Leuven, Belgium, 6 April 2023.

**Xiaoming Hu:**
Tutorial on FAIR data management tools in biomedical research, 68. GMDS-Jahrestagung, Heilbronn (Germany), 17-21 September 2023.

**Lukas Hübner:**
*"The Computational Stumbling Blocks: Fault-Tolerance and Reproducibility"*, Biodiversity Bioinformatics Workshop, Heraklion, Greece, May 2023.

**Alexander Jordan:**
*"Model diagnostics and forecast evaluation for quantiles"*, International Conference on Computational and Financial Econometrics, Berlin, Germany, 16 December 2023.

**Olga Krebs:**
*"FAIRDOM-SEEK: FAIR and standardised management of models and related (meta)data"*, COMBINE 2023, Farmington, CT (USA), 6 October 2023. *"FAIR Data Management for Collaborative Systems Medicine Projects: From Instruments to Publication"*, Systems Medicine Session at the 22nd International Conference on Systems Biology (ICSB 2023), Hartford, CT (USA), 10 October 2023.

**Stiv Llenga:**
*"Machine Learning on Open-Shell N-Heteropolycycles"*, SFB1249 Flash-Talks, Heidelberg, Germany, 13 January 2023.

**Stefan Machmeier:**
*"A Generalizable Approach for Network Flow Image Representation for Deep Learning"*, talk at the 7th Cyber Security in Networking Conference (CSNet), Montreal, QC, Canada, 17 October 2023. *"Explainable Artificial Intelligence for Improving a Session-Based Malware Traffic Classification with Deep Learning"*, talk at the 2023 IEEE Symposium Series on Computational Intelligence (SSCI), Mexico City, Mexico, 7 December 2023.

**Gerhard Mayer:**
*"Current landscape of standards - Long term view on standardization, Part 2"*, Plenary meeting of the Ecosystem Digital Twins in Healthcare (EDITH), Leuven (Belgium), 30 - 31 January 2023.

**Benoit Morel:**
*"AleRax: massively parallel species tree inference and gene tree reconciliation from gene tree distributions"*, Peder Sather/Invertomics Symposium *"Progress and Development in Phylogenetic Methods"*, University of Oslo, Norway, March 2023.

**Anthony Noll:**
*"Impact of central mixing and nuclear reactions network on the size of convective cores"*, PLATO Stellar Science Conference 2023 Milazzo, Italy, 30 April 2023. *"Impact of mixing and nuclear reactions rates on the seismic properties of red clump stars"*, MIAPbP Workshop, Garching, Germany, 22 August 2023.

**Dominique Ostermayer:**
*"Curvature Effects in Reactivity"*, Paris International Summer School on Advanced Computational Materials Science 2023, Paris, France, 28 August - 1 September 2023.

**Giulia Paiardi:**
*"The role of heparin in SARS-CoV-2 infection: from a model for the heparan sulfates to a starting structure for antivirals"*. CECAM conference: Computational methods in biophysics for applications to drug discovery, online, 1 June 2023.

**Marc-Oliver Pohle:**
*"Testing quantile forecast optimality"*, 12th European Central Bank Conference on Forecasting Techniques, Frankfurt, Germany, 12 June 2023. *"Generalised covariances and correlations"*, HKMEtrics Workshop on Economic and Financial Forecasting, Karlsruhe, Germany, 12 July 2023. *"Regression diagnostics via generalized residuals"*, Statistical Week, TU Dortmund, Dortmund, Germany, 12 September 2023. *"Scores-based calibration testing for multivariate forecast distributions"*, HKMEtrics Workshop on Economic and Financial Forecasting, Karlsruhe, Germany, 5 December 2023. *"Uncertainty quantification in forecast comparisons"*, International Conference on Computational and Financial Econometrics, Berlin, Germany, 16 December 2023.

**Kai Lars Polsterer:**
*"Unsupervised learning for agnostic knowledge discovery from simulations"*, Astronomische Gesellschaft (AG) Meeting 2023, Berlin, Germany, 14 September 2023. *"From Photometric Redshifts to Improved Weather Forecasts"*, HLF Young Researchers Visit to HITS, Heidelberg, Germany, 27 September 2023. *"HiPSter: Using HiPS tilings to allow for explorative access to simulations"*, IVOA November 2023 Interoperability Meeting, Tucson, AZ, USA, 11 November 2023. *"Wie künstliche Intelligenz Astronomen bei der Arbeit hilft"*, Deutsches Museum, Bonn, Germany, 6 December 2023.

**Francisco Pozo Nuñez:**
*"Quasars accretion discs sizes with the LSST"*. Seminar of the *"Galaxies & Cosmology"* Department at the MPIA. MPIA, Heidelberg, Germany, 28 September 2023.

**Ghulam Qadir:**
*"Deep learning for spatial statistics via neural tangent kernels"*, MathSEE Symposium on Applications of Mathematical Sciences, Karlsruhe Institute of Technology, Karlsruhe, Germany, 27 September 2023.

**Alexandros Stamatakis:**
*"Quantifying Uncertainty in Evolutionary Analyses"*, talk, Biomedical Sciences Research Center (BSRC) Alexander Fleming, Athens, Greece, April 2023. *"Quantifying Uncertainty in Evolutionary Analyses"*, contributed talk, EMBO Satellite Workshop on Biodiversity Informatics, Hellenic Center for Marine Research, Heraklion, Crete, May 2023.

**Jonathan Teuffel:**
*"Multiscale simulation of cytochrome P450 electron transfer complexes: the reduction of CYP17A1 and its implications for the regulation of human sex hormone biosynthesis"*, ZAPP seminar series, Center for Molecular Biology (ZMBH), Heidelberg University, Heidelberg, 15 May 2023.

**Anastasis Togkousidis:**
*"Adaptive search heuristic for maximum likelihood phylogenetic tree inference based on the predicted difficulty of the dataset"*, Peder Sather/Invertomics Symposium *"Progress and Development in Phylogenetic Methods"*, University of Oslo, Norway, March 2023. *"Parallel Inference of Phylogenetic Stands with Gentrius"*, 22nd IEEE International Workshop on High Performance Computational Biology (HiCOMB), IPDPS 2023, St. Petersburg, Florida, USA, May 2023.

**Sebastian Trujillo Gomez:**
*"Unsupervised learning for agnostic knowledge discovery from simulations"*, Astroinformatics 2023, INAF National Auditorium Observatory of Capodimonte, Naples, Italy, 1-6 October 2023.

**Rebecca C. Wade, Abraham Muniz Chicharro, Stefan Richter:**
Lecture and Practical on *"Brownian dynamics simulations with SDA to compute association rates"*, 2nd HBPMolSIM Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades", Mathematikon, Heidelberg University and Hybrid, 21-23 June 2023.

**Ulrike Wittig:**
*"Improved insights into the SABIO-RK database via visualization"*, 16th Annual International Biocuration Conference, Padua, Italy, 24-26 April 2023. *"Data Resources and Biocuration at ELIXIR"*, ELIXIR All Hands Meeting, Dublin, Ireland, 5-8 June 2023.

**Daniel Wolffram:**
*"Collaborative nowcasting of COVID-19 hospitalization incidences in Germany"*, Talks at the HKMEtrics Workshop on Economic and Financial Forecasting, Karlsruhe, Germany, 12 July 2023, at the 5th Conference of the Central European Network, University of Basel, Basel, Switzerland, 7 September 2023, and at the MathSEE Symposium on Applications of Mathematical Sciences, Karlsruhe Institute of Technology, Karlsruhe, Germany, 28 September 2023.

**Elaine Zaunseder:**
*"Machine Learning Methods Improve Specificity in Newborn Screening for Isovaleric Aciduria"*, APS 2023, in Kassel, Germany, 8 – 10 March 2023. *"A digital-tier strategy based on machine learning methods improves specificity in newborn screening for isovaleric aciduria"*, SSIEM, Jerusalem, Israel, 28 August – 1 September 2023. *"Deep learning and explainable artificial intelligence for improving specificity and detecting metabolic patterns in newborn screening"*, 2023 IEEE Symposium Series on Computational Intelligence, Mexico City, Mexico, 5 – 8 December 2023.

**Alexander Zeilmann:**
*"Mathematical guidance for the interpretation of PCR-based MSI diagnostics."* 7th EHTG Meeting 2023, Vilnius, Lithuania, 29 September 2023 – 1 October 2023. *"Tumor3D – a platform for virtual 3D tumor models."* 7th EHTG Meeting 2023, Vilnius, Lithuania, 29 September 2023 – 1 October 2023.

**Posters**

**Felix Ahlborn:**
*"Accurate asteroseismic surface rotation rates for evolved red giants"*, PLATO Stellar Science Conference 2023 Milazzo, Italy, 26-30 June 2023.

**Tommaso Bartoloni:**
*"Peptide-based therapeutics against cardiac disease"*. BioExcel Summer School on Biomolecular Simulations 2023, Sardinia, Italy, 10-15 September 2023. *"Towards peptide-based therapeutics against cardiac disease: Prediction and simulation of the S100A1ct peptide with a membrane environment"*. EMBO workshop: Computational Structural Biology. Heidelberg, 6-9 December 2023.

**Beatriz Bordadagua:**
*"Mixed modes slow down red giants core rotation"*, PLATO Stellar Science Conference 2023 Milazzo, Italy, 26-30 June 2023.



**Mislav Brajkovic:**
*"Molecular Simulations of Histamine H1 Receptor-Binding Compounds."*. SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany. 2-4 May 2023. *"Calculating residence time of Histamine H1 receptor-binding compounds using τRAMD"*. BioExcel Summer School on Biomolecular Simulations 2023, Sardinia, Italy, 10-15 September 2023.

**Alessandro Calzolari:**
*"Exploring and Exploiting Interactions between Molecules and Framework Materials"*, Advanced Quantum ESPRESSO School: Hubbard and Koopmans Functionals from Linear Response, Pavia, Italy, 28 August - 1 September 2023.

**Quentin Coppée:**
*"Strong radial modes in red giant stars with suppressed non-radial modes"*, PLATO Stellar Science Conference 2023 Milazzo, Italy, 26-30 June 2023.

**Giulia D'Arrigo:**
*"Computation of Protein-Protein Interaction Kinetics using τ-RAMD"*. Human Brain Project Summit 2023, Marseille, France. 27-31 March 2023. *"Computation of Unbinding Rates and Mechanisms in Protein-Protein Systems"*. SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany. 2 – 4 May 2023.

**Christopher Ehlert:**
*"Metal-Free Molecular Catalysts for the Oxygen Reduction Reaction"*, 17th International Congress of Quantum Chemistry (ICQC), Bratislava, Slovakia, 26 June - 1 July 2023.

**Francisca Espinoza:**
*"Asteroseismic catalogue of Kepler Red Giants"*, Larim 2023, Montevideo, Uruguay, 29 November 2023.

**Rostislav Fedorov:**
*"Exploration of Redox Properties in Chemical Space"*, 1st SIMPLAIX Workshop on *"Machine Learning for Multiscale Molecular Modeling"*, Heidelberg, Germany, 2-4 May 2023. *"MPNN for Prediction of the Reduction Potential"*, Learning on Graphs Paris Meetup 2023, Paris, France, 14 June 2023, European Conference on Computational and Theoretical Chemistry (EuChemS CompChem 2023), Thessaloniki, Greece, 27-31 August 2023; *"Deterministic Graph Retrosynthetic Algorithm for COFs separation"*, RDKit UGM 2023, Mainz, Germany, September 20 -22 September 2023.

**Manuel Glaser:**
*"Structural prediction of transmembrane peptide-protein interaction and application to the potential peptide therapeutic S100A1ct"*, Informatics for Life Symposium *"HEALTHY AI – promises and limits"*, Heidelberg, Germany, 15 September 2023. *"Towards peptide-based therapeutics against cardiac disease – prediction of a mechanism of action for the S100A1ct peptide by docking"*, EMBO Workshop on Computational structural biology, Heidelberg, Germany, 6 – 9 December 2023.

**Manuel Glaser, Tommaso Bartoloni:**
*"Structural prediction of transmembrane peptide-protein interaction and application to a potential peptide therapeutic"*, SIMPLAIX Workshop on *"Machine Learning for Multiscale Molecular Modeling"*, Heidelberg, Germany, 2 – 4 May 2023.

**Olga Krebs, Maja Rey, Ina Biermayer, Andreas Weidemann, Susan Eckerle, Xiaoming Hu, Martin Golebiewski, Wolfgang Müller:**
*"FAIR data management for collaborative Systems Medicine projects : from Instruments to Publication"*, 22nd International Conference on Systems Biology (ICSB 2023), Hartford, CT (USA), 10 October 2023.

**Stiv Llenga:**
*"Fragmentation-Based Molecular Representation Suitable for ML/DL Applications"*, 1st SIMPLAIX Workshop on "Machine Learning for Multiscale Molecular Modeling", Heidelberg, Germany, 2-4 May 2023. *"Machine Learning on Open-Shell N-Heteropolycycles"*, SFB1249 Symposium on *"Materials for Organic Electronics 2023"*, Heidelberg, Germany, 16-17 June 2023. *"New Representations for Chemical Machine-Learning"*, CECAM-Psi-k Research Conference *"Bridging Length Scales with Machine Learning: From Wavefunctions to Thermodynamics"*, Berlin, Germany, 19-23 June 2023; European Conference on Computational and Theoretical Chemistry (EuChemS CompChem 2023), Thessaloniki, Greece, 27-31 August 2023.

**Gerhard Mayer, Martin Golebiewski, Wolfgang Müller:**
The role of standards in defining an ecosystem for Virtual Human Twins (VHTs), 1st Conference on Research data Infrastructure (CoRDI), Karlsruhe (Germany), Germany, 12-14 September 2023.

**Ghadeer Mobasher:**
*"WeLT: Improving Biomedical Fine-tuned Pre-trained Language Models with Cost-sensitive Learning"*, 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks, Association for Computational Linguistics, Toronto, Canada, 13 July 2023.

**Jonas Müller:**
*"The area enclosed by the H-burning shell dominates the magnetic frequency splittings"*, PLATO Stellar Science Conference 26-30 June 2023, Milazzo, Italy.

**Abraham Muniz-Chicharro:**
*"Multiscale approach to predict protein-ligand binding kinetics"*, Biophysical Society Annual Meeting, San Diego, USA, 18-22 February 2023; Annual Retreat, Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences, Heppenheim, 31 July – 2 August 2023.

**Dominique Ostermayer:**
*"Organic Materials by Geometric Design"*, Paris International Summer School on Advanced Computational Materials Science 2023, Paris, France, 28 August - 1 September 2023; 10th Heidelberg Laureate Forum 2023, Heidelberg, Germany, 24-29 September 2023.

**Giulia Paiardi:**
*"PepAISim: Combining AI and molecular simulation for anticancer peptide and peptidomimetic design."* Joachim Herz Stiftung Add-On Fellowship kick off meeting, Hamburg, 10-11 February 2023; SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, 2-4 May 2023; AI inScide Out Unconference, Heidelberg, 16-17 October 2023; EMBO workshop: Computational Structural Biology, Heidelberg, 6-9 December 2023.

**Anna Piras:**
*"Predicting Adsorbent and Electronic Properties of Nanographenes"*, 17th International Congress of Quantum Chemistry (ICQC), Bratislava, Slovakia, 26 June - 1 July 2023.

**Kai Lars Polsterer:**
*"Using unsupervised learning for explorative discovery in astrophysical simulations."*, ADASS XXXIII, Tucson, AZ, USA, 5-9 November 2023.

**Jonathan Teuffel:**
*"Effects of conformational transitions and redox protein binding on the catalytic properties of CYP17A1 and CYP2B4 revealed by ligand egress patterns"*, Hünfeld Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, 28-29 April 2023.
*"Multiscale simulation of cytochrome P450 electron transfer complexes: the reduction of CYP17A1 and its implications for the regulation of human sex hormone biosynthesis"*, SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, 2-4 May 2023; 2023 HGS MathComp retreat, Heppenheim, 2 August 2023.

**Sebastian Trujillo-Gomez, Kai Polsterer & Bernd Doser:**
*"Unsupervised learning for agnostic knowledge discovery from simulations"*. Poster at the Astronomische Gesellschaft (AG) Meeting 2023, Berlin, Germany, 11-15 September 2023.

**Ulrike Wittig:**
*"Improved insights into the SABIO-RK database via visualization"*, 16th Annual International Biocuration Conference, Padua, Italy, 24-26 April 2023.

**Daniel Wolffram:**
*"Collaborative nowcasting of COVID-19 hospitalization incidences in Germany"*, poster at a satellite meeting of the Royal Society, London, United Kingdom, 15 March 2023.

## 10.3 Memberships

**Michaël Bazot:**
Member Plato Science Management (PSM), Leader of the PLATO Working Group *"Error estimates and propagation through the SAS pipeline"*.

**Tilmann Gneiting:**
Affiliate Professor at the Department of Statistics, University of Washington, Seattle, Washington, United States. Deputy Co-Speaker for the Statistical Data Analysis and Forecasting area and Member of the Steering Committee at Center MathSEE (Mathematics in Sciences, Economics and Engineering), Karlsruhe Institute of Technology. Member of the Committee on Fellows at Institute for Mathematical Statistics.

**Martin Golebiewski:**
Convenor (chair) of the ISO/TC 276 Biotechnology working group 5 *"Data Processing and Integration"*, International Standardization Organization (ISO). 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). Member of the national German standardization committee ("Nationaler Arbeitsausschuss") NA 176-09-02 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).

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

**Vincent Heuveline:**
University Council of Universität Hamburg (since 2020) Scientific Advisory Board, Potsdam Institute for Climate Impact Research (since 2017).

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

**Anthony Noll:**
Plato Science Management, WP 121 (Stellar models).

**Giulia Paiardi:**
Member of the Innogly COST action (CA18103). Add-On Fellowship of the Joachim Herz Stiftung. Postdoctoral Fellow of the AI Health Innovation Cluster.

**Kai Polsterer:**
President of International Astro-Informatics Association (IAIA, since October 2023). 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 *„Verein für datenintensive Radioastronomie"* (VdR). Member of the International Astrostatistics Association.

**Alexandros Stamatakis:**
Member of scientific committee of the SMPGD (Statistical Methods for Post Genomic Data analysis) workshop series.

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

**Rebecca Wade:**
Associate Editor: Journal of Molecular Recognition. Academic Editor: PLOS Computational Biology. Editorial Board: Biophysical Journal, Advances and Applications in Bioinformatics and Chemistry; BBA General Subjects; Biopolymers; Journal of Chemical Information and Modeling; Journal of Computer-aided Molecular Design; Journal of Physical Chemistry B; Protein Engineering, Design and Selection. Member of Scientific Advisory Council of the Computational Biology Unit (CBU), University of Bergen, Norway. Member of Scientific Advisory Board of the Max Planck Institute of Biophysics, Frankfurt, Germany. Member of International Advisory Board, 'Complexity in Chemistry' C2 Programme, Department of Chemical Sciences, University of Padova, Italy. Scientific Chair and Secretary, QSAR, Chemoinformatics and Modeling Society (QCMS). Coordinator, SIMPLAIX HITS-Heidelberg University-Karlsruhe Institute of Technology consortium on data-enhanced multiscale molecular simulation of biomolecules and molecular materials. Member at Heidelberg University of: HBIGS (Heidelberg Biosciences International Graduate School) faculty, HGS MathComp Graduate School faculty, Helmholtz Information and Data Science School for Health (HIDDS4Health) Graduate School faculty, Interdisciplinary Center for Scientific Computing (IWR), DKFZ-ZMBH Alliance of the German Cancer Research Center and the Center for Molecular Biology at Heidelberg University, Faculties of Engineering Sciences and Biosciences, CellNetworks Core Facilities Steering Committee, Thematic Research Network "Molecular Mechanisms in health and disease - from understanding to engineering (MINDS)" Steering Committee.

**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

**Aksel Alpay:**
Reviewer for the journals and conferences Future Generation Computer Systems (FCGS), Concurrency and Computation: Practice and Experience (CCPE), IXPUG Annual Conference 2023, IXPUG Workshop at HPC Asia 2024. Panelist and Co-Organizer of the Birds-of-a-Feather session Heterogeneous Programming with C++ and Khronos SYCL, ISC '23, Hamburg. Instructor and Co-Organizer of the Tutorial session Heterogeneous Programming with C++ and Khronos SYCL, ISC'23, Hamburg. Instructor and Co-Organizer of the Tutorial session SYCL Techniques and Best Practices, IWOCL '23, Cambridge, UK. Panelist at the panel discussion OpenCL and SYCL, IWOCL' 23, Cambridge, UK.

**Giulia D'Arrigo, Stefan Richter, Rebecca Wade:**
Co-organizers: 2nd HBPMolSim Human Brain Project Training Workshop on Tools for Molecular Simulation of Neuronal Signaling Cascades and the HBP BRAVE project workshop. Mathematikon, Heidelberg University and Hybrid, 21-23 June 2023.

**Rostislav Fedorov:**
Co-organizer of the 1st SIMPLAIX Workshop on *"Machine Learning for Multiscale Molecular Modeling"*, Heidelberg, Germany, 2-4 May 2023.

**Ganna Gryn'ova:**
Scientific coordinator of the Lieseberg Colloquia series at the Faculty of Chemistry and Earth Sciences, Heidelberg University.

**Martin Golebiewski:**
NFDI4Health Workshop *"Standards for Health Data"*, TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung, Berlin (Germany), 20-21 February 2023. Chair of the 19th Committee Meeting of ISO/TC 276 Biotechnology working group WG5 *"Data Processing and Integration"*, Atlanta, Georgia (USA), 12-16 June 2023. Host and chair of Committee Meetings of ISO/TC 276 Biotechnology working group WG5 *"Data Processing and Integration"*, online, 4-8 December 2023.

**Martin Golebiewski, Gerhard Mayer:**
World Café Session: *"Standardization of data, models, metadata and workflows"*, EDITH Deep Thinkers Meeting, Rome (Italy), 17th May 2023.

**Saskia Haupt:**
Y-EHTG representative in the board of the European Hereditary Tumor Group (EHTG), since 2023: Assistant editor of the official MathOnco community website, since 2022. Reviewer for the Journals ActaBiotheoretica, BioSystems, International Journal of Cancer, npj precision oncology.

**Saskia Hekker:**
Scientific organizer of the MIAPP program Stellar Astrophysics in the Era of Gaia, Spectroscopic, and Asteroseismic Surveys.

**Giulia Paiardi:**
Co-organizer of *"Young Researchers on stage: International Meeting of Young Researchers – INNOGLY & GLYCONanoPROBES"* workshop. Lasi (Romania). 20-21 April 2023.

**Marc-Oliver Pohle:**
Session organizer *"Statistical Theory and Methods"*, Statistical Week, TU Dortmund, Dortmund, Germany, 12 September 2023. Member of the scientific program committee, MathSEE Symposium on Applications of Mathematical Sciences, Karlsruhe Institute of Technology, Karlsruhe, Germany, 27-29 September 2023. Session organizer *"Advances in forecasting and forecast evaluation"*, International Conference on Computational and Financial Econometrics, Berlin, Germany, 16 December 2023.

**Kai Polsterer:**
Co-organizer of E-Science and E-Infrastructure Splinter at AG-Tagung,Berlin, Germany, 11-15 September 2023. Scientific organizing committee of Astroinformatics 2023, Naples, Italy, 1-6 October 2023. Scientific organizing committee of Radio 2023, Bochum, Germany, 14-17 November 2023.

**Alexandros Stamatakis:**
Organizer of 2023 Computational Molecular Evolution Summer School, Hellenic Center for Marine Research, Crete, Greece. Organizer of 2023 Satellite workshop on Biodiversity Informatics, Hellenic Center for Marine Research, Crete, Greece.

**Rebecca Wade (Chair), Frauke Gräter, Ganna Gryn'ova:**
Scientific organizers (with Andreas Dreuw, Fred Hamprecht, Ullrich Köthe (Heidelberg University), Pascal Friederich, Marcus Elstner (KIT), SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Studio Villa Bosch, Heidelberg, 2-4 May 2023.

**Other contributions**

**Saskia Hekker:**
*"Der Puls der Sterne"*. Public Talk at the *"Universe on Tour"* roadshow, Heidelberg, 22 July 2023. *"Asteroseismologie: Schwingende Sterne und innere Geheimnisse."* Guest contribution in the "AstroGeo" podcast, 28 September 2023.

**Peter Saueressig:**
Member of the evaluation panel, ERC Science Journalism Initiative 2023.

## 10.5 Awards

**Marcus Buchwald:**
2nd Poster Price of the DKFZ Summer School on Medical Physics 2023 in Chile: *"The role of imaging in the radiotherapy process"*, 11-15 December 2023, Santiago de Chile.

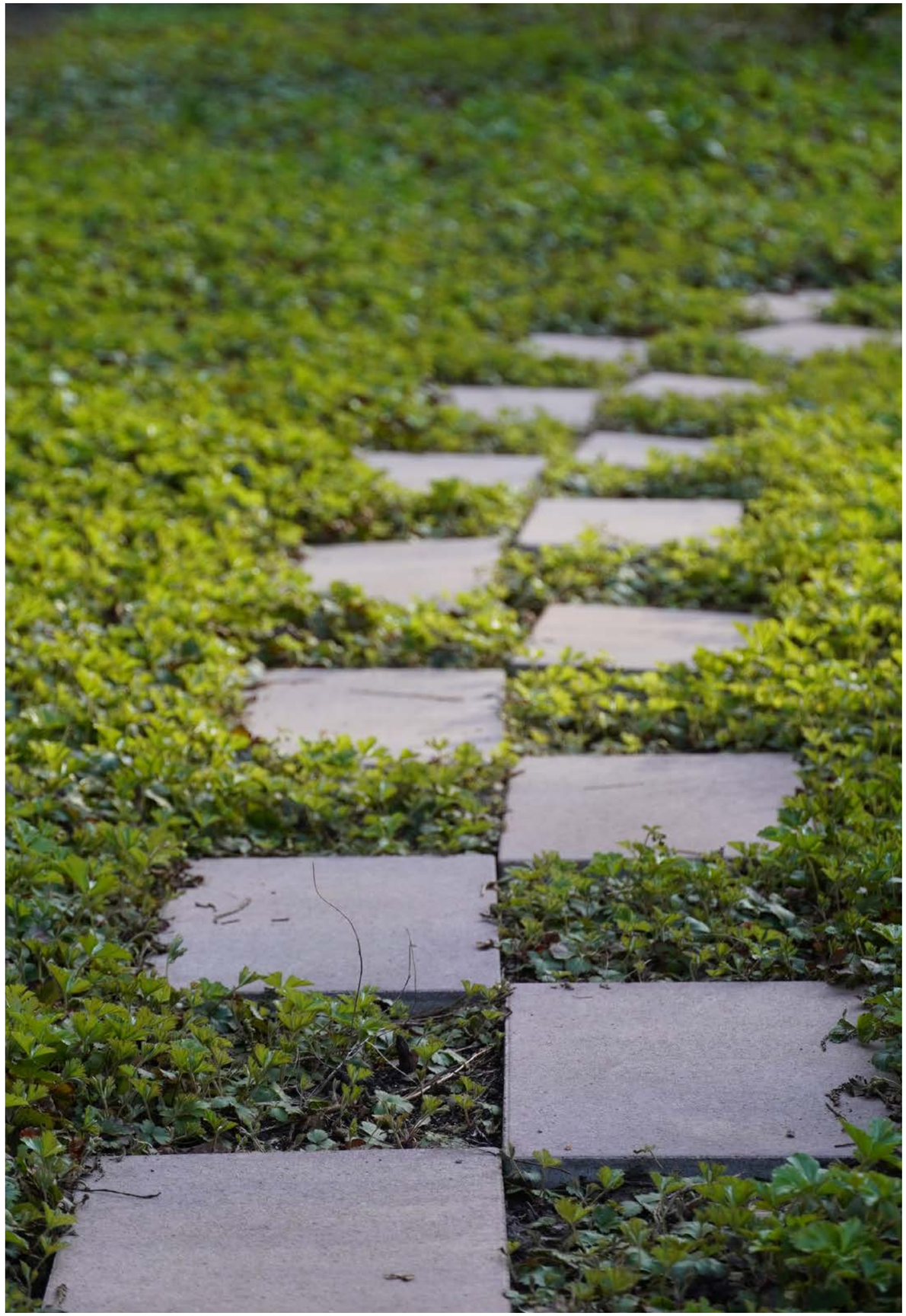
**Manuel Glaser:**
Poster Award 2023, Informatics for Life Symposium - *"HEALTHY AI – promises and limits"*, Heidelberg, Germany, 15 September 2023.

**Saskia Haupt:**
Young Marsilius Fellowship for Interdisciplinarity and Science Communication, awarded by the Marsilius Kolleg Heidelberg for one year.

**Friedrich Röpke:**
ERC Advanced Grant *"ExCEED: Explaining Common-Envelope Evolution and Dynamics in binary stellar systems"*, European Research Council, 2022.

**Alexandros Stamatakis:**
Highly Cited Researcher in the biology and biochemistry category, Clarivate Analytics, 2023. Appointed as honorary fellow of the Foundation for Research and Technology Hellas.

**Elaine Zaunseder:**
SSIEM Travel Award 2023, SSIEM, Jerusalem, Israel, 28 August – 1 September 2023.



# 11 Boards and Management

*The HITS Scientific Advisory Board and the HITS management (November 2023). From left to right: Tilmann Gneiting (HITS Scientific Director), Jeffrey Brock, Kai Polsterer (HITS Deputy Scientific Director), Thomas Lengauer, Gesa Schönberger (HITS Managing Director), Alex Szalay, Victoria Stodden.*

## 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 2023, 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
- **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 (until June 2023)
- **Prof. Dr. Thomas Lengauer**, Max Planck Institute for Computer Science, Saarbrücken, Germany (Chair, SAB)
- **Prof. Dr. Alex Szalay**, Johns Hopkins University, Baltimore, USA (Vice Chair, SAB)
- **Prof. Dr. Victoria Stodden**, University of Southern California, Los Angeles, (Vice Chair, SAB)
- **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
(until November 2023)

**Prof. Dr. Carsten Könneker**
Member of the Board of Directors
*© Mück/Klaus Tschira Stiftung*

**Prof. Dr. Dieter Kranzlmüller**
Member of the Board of Directors
(since July 2023)
*© Alessandro Podo*

**Heidelberg University**
**Prof. Dr. Jörg Pross**
Vice-President of Research and Structure (until September 2023)
*© Philip Benjamin*

**Prof. Dr. Andreas Dreuw**
Vice President of Research and Structure (since October 2023)
*© Tobias_Schwerdt*

**Karlsruhe Institute of Technology (KIT)**
**Prof. Dr. Michael Decker**
Head of Division II "Informatics, Economics, and Society"

## 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 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. Tilmann Gneiting**
(2023–2024)

**Deputy Scientific Director:**
**Dr. Kai Polsterer**
(2023–2024)

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 | © 2024 HITS gGmbH

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