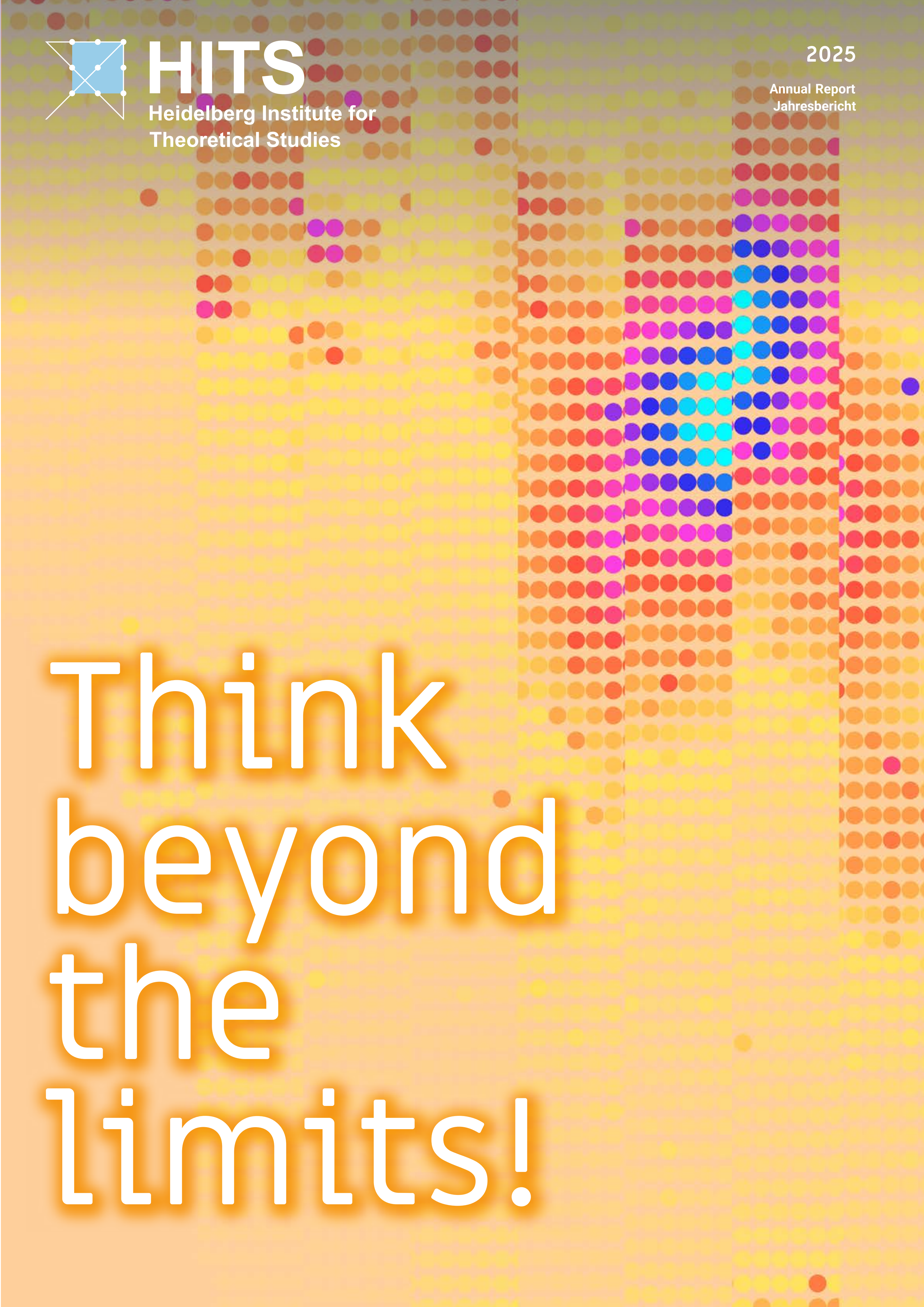


HITS
Heidelberg Institute for Theoretical Studies

2025

Annual Report
Jahresbericht

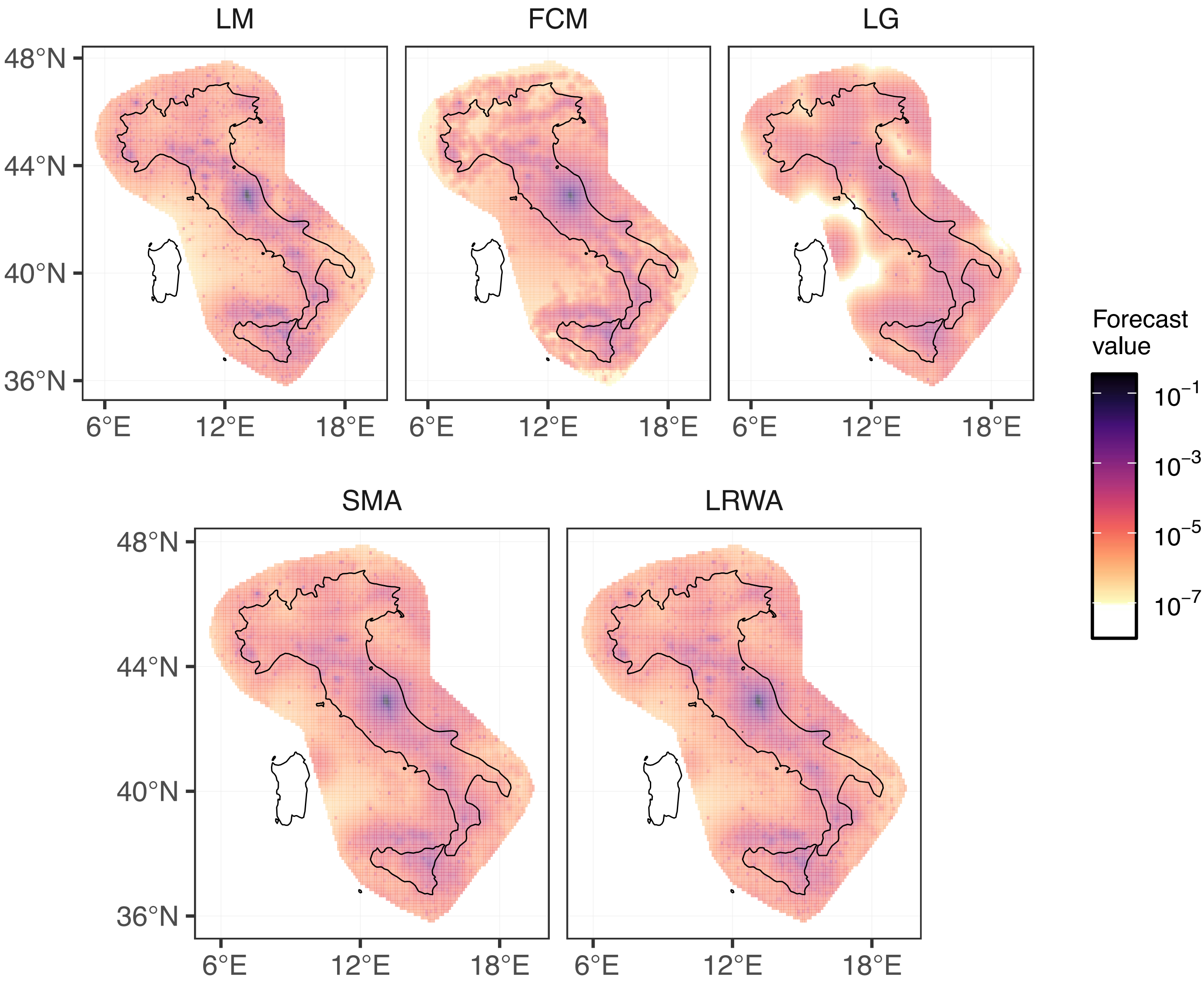
Think beyond the limits!

The Italian region is strongly affected by frequent earthquakes. The Computational Statistics (CST) group contributes to the scientific evaluation of earthquake predictions in Italy in collaboration with partners from the University of Naples.

The image shows predictions of the expected number of earthquakes over the seven-day period starting on 30 October 2016 in each of the 8,993 grid cells of the Italian testing region. We use the predictions from five forecasting models to illustrate the evaluation methodology that was tailored to the seismological problem (see Chapter 2.3 Computational Statistics, pp. 22ff). One of the main challenges in predicting earthquakes and evaluating the corresponding forecasts lies in the low probabilities of occurrence.

Italien ist eine Region, die stark von Erdbeben betroffen ist. In Zusammenarbeit mit Partnern der Universität Neapel trägt die Arbeitsgruppe Computational Statistics (CST) zu einer wissenschaftlich fundierten Bewertung von Erdbebenvorhersagen in Italien bei.

Die Abbildung zeigt Vorhersagen der erwarteten Anzahl von Erdbeben während des Sieben-Tage-Zeitraums ab dem 30. Oktober 2016 in jeder der 8.993 Gitterzellen der Testregion. Anhand von fünf Vorhersagenmodellen veranschaulichen wir Methoden zur Bewertung von Vorhersagen, die an das seismologische Problem angepasst sind (siehe Kapitel 2.3 Computational Statistics, S. 22ff). Eine große Herausforderung bei der Vorhersage von Erdbeben und der entsprechenden Bewertung liegt in den geringen Auftrittswahrscheinlichkeiten.


LM
FCM
LG
SMA
LRWA
48°N
44°N
40°N
36°N
6°E
12°E
18°E
Forecast value
10^{-1}
10^{-3}
10^{-5}
10^{-7}

Gesa Schönberger

Dr. Gesa Schönberger
(Managing Director / Geschäftsführerin)

Dr. Kai Polsterer
(Scientific Director / Wissenschaftlicher Direktor)

1 Think beyond the limits!

The world around us is constantly changing. Some of these changes are perceived as turbulent and uncertain – not only in a scientific context, but also in a personal one. Rapid change creates a palpable sense of crisis that lingers. Science as an institution is not immune to this. As society has developed and grown, so has scientific research and practice. On that account, it is only natural that an ever-shifting societal landscape would also affect scientific context.

Luckily, change is an integral part of the scientific process. Theories are revised when new evidence becomes available, when advances in technology are made, or simply when research is approached differently. In 2025, a few things also changed at HITS. The year began with **Kai Polsterer** taking on the role of Scientific Director, replacing **Tilmann Gneiting**, who had filled the post during 2023 and 2024. Moreover, **Rebecca Wade** – supporting Polsterer – was appointed Deputy Scientific Director.

In May 2025, the contributions of both Polsterer – with "Dynaverse" – and Wade – with "SynthImmune" – were approved with two **Clusters of Excellence** in the Excellence Strategy of the Federal and State Governments of Germany (see Chapter 4). Participation in these Clusters is of great significance to HITS: Not only does this validate the work of our researchers, but it also promotes interdisciplinarity and facilitates cross-consortial collaboration.

Furthermore, Klaus Tschira Guest Professor **Guillermo Cabrera-Vives**' (Chile) seven-month stay at HITS came to an end in March 2025. Cabrera-Vives was succeeded by **Raphael Hirschi** (United Kingdom), who worked together with scientists at HITS from April until the end of the year. This program highlights the increasing value of collaboration and communication within the scientific community when it comes to offering solutions to problems encountered by society at large, particularly during periods of social instability (see Chapter 6.1).

For **Jackson Ryan**, the 13th HITS Journalist in Residence, transparent and trustworthy scientific communication is the pillar of keeping the public informed, engaged, and educated. Ryan joined HITS in April 2025 and stayed until October, delivering an internal seminar on scientific integrity in the face of uncertainty in addition to a public talk on quality journalism at the German-American Institute (DAI) (see Chapter 4).

Language is part and parcel of science communication. Its usage and fairness are the primary focus of **Vagrant Gautam**, who joined HITS in September 2025 as the 3rd Independent Postdoc and has since been working closely with the Natural Language Processing group (see Chapters 2.12 and 6.2).

Natural Language Processing was also the theme of the first symposium in our new **"The Future of…" series**, which was held in November 2025 (see Chapter 5.4). The aim of this future-oriented event format is to establish HITS as one of the pioneers in its research areas and to continue increasing the visibility of the institute's collective work in alignment with its motto, "Think beyond the limits!"

Without doubt, part of HITS' mission is for its scientists to become trusted international leaders in its research areas. The year 2025 marked the 10th year in a row during which group leader **Alexandros Stamatakis** – working in Crete – was named one of the most-cited researchers worldwide in the biology and biochemistry field, earning him a place on Clarivate Analytics' "Highly Cited Researchers" list (see Chapters 2.2, 4 and 10.4).

Society may change science, but science also grows alongside society and brings about changes within it. Science does not offer one simple solution to a problem; rather, it assesses situations, poses questions, and reconsiders challenges. Year after year, this is what we aim to do at HITS, and 2025 was a testament to this endeavor.

Take a look at this year's Annual Report and find that out for yourself. Happy reading!

Die Welt um uns herum verändert sich kontinuierlich. Manche dieser Veränderungen werden als turbulent und unsicher empfunden – nicht nur im wissenschaftlichen, sondern auch im persönlichen Kontext. Der rasante Wandel erzeugt spürbar das Gefühl einer andauernden Krise. Die Wissenschaft als Institution ist dagegen nicht immun. So wie sich die Gesellschaft entwickelt und gewachsen ist, haben sich auch wissenschaftliche Forschung und Praxis fortentwickelt. Deshalb wirkt sich natürlich eine Gesellschaft, die sich permanent verändert, auch auf die Wissenschaft aus.

Doch zum Glück ist die Veränderung ein integraler Bestandteil des wissenschaftlichen Prozesses. Theorien werden überarbeitet, wenn es neue Erkenntnisse oder technologische Fortschritte gibt, oder einfach, wenn ein neuer Forschungsansatz entsteht. Das Jahr 2025 brachte auch am HITS einige Veränderungen mit sich: Zu Beginn des Jahres wurde **Kai Polsterer** Wissenschaftlicher Direktor und löste damit **Tilmann Gneiting** ab, der diese Position in den Jahren 2023 und 2024 innehatte. Stellvertretende Wissenschaftliche Direktorin wurde **Rebecca Wade**, die Polsterer in dieser Funktion unterstützt.

Mit ihrer wissenschaftlichen Expertise waren sowohl Polsterer als auch Wade erfolgreich: Im Mai 2025 wurden ihre Beiträge im Rahmen der Exzellenzstrategie des Bundes und der Länder mit zwei **Exzellenzclustern** gefördert (Polsterer mit „Dynaverse“ und Wade mit „SynthImmune“, siehe Kapitel 4). Die Beteiligung an diesen Clustern ist für HITS von großer Bedeutung: Sie bestätigt nicht nur die Arbeit unserer Forschenden, sondern fördert auch die Interdisziplinarität und erleichtert die Zusammenarbeit in Konsortien.

Im März ging der siebenmonatige Aufenthalt von Klaus-Tschira-Gastprofessor **Guillermo Cabrera-Vives** (Chile) am HITS zu Ende. Sein Nachfolger **Raphael Hirschi** (Großbritannien) kam im April ans Institut und arbeitete bis zum Ende des Jahres mit HITS-Forschenden zusammen. Das Programm unterstreicht den zunehmenden Wert von Kooperation und Kommunikation innerhalb der wissenschaftlichen Gemeinschaft, wenn es darum geht, Lösungen für Probleme anzubieten, mit denen die Gesellschaft insgesamt konfrontiert ist, insbesondere in Zeiten sozialer Instabilität (siehe Kapitel 6.1).

Für **Jackson Ryan**, den 13. HITS „Journalist in Residence“, ist eine transparente und vertrauenswürdige Wissenschaftskommunikation die Grundlage dafür, dass die Öffentlichkeit informiert und aufgeklärt wird und sich engagiert. Ryan kam im April 2025 ans HITS und blieb bis Ende Oktober. Unter anderem hielt er ein internes Seminar zum Thema wissenschaftliche Integrität in Zeiten der Unsicherheit sowie einen öffentlichen Vortrag über Qualitätsjournalismus am Deutsch-Amerikanischen Institut (DAI) in Heidelberg (siehe Kapitel 4).

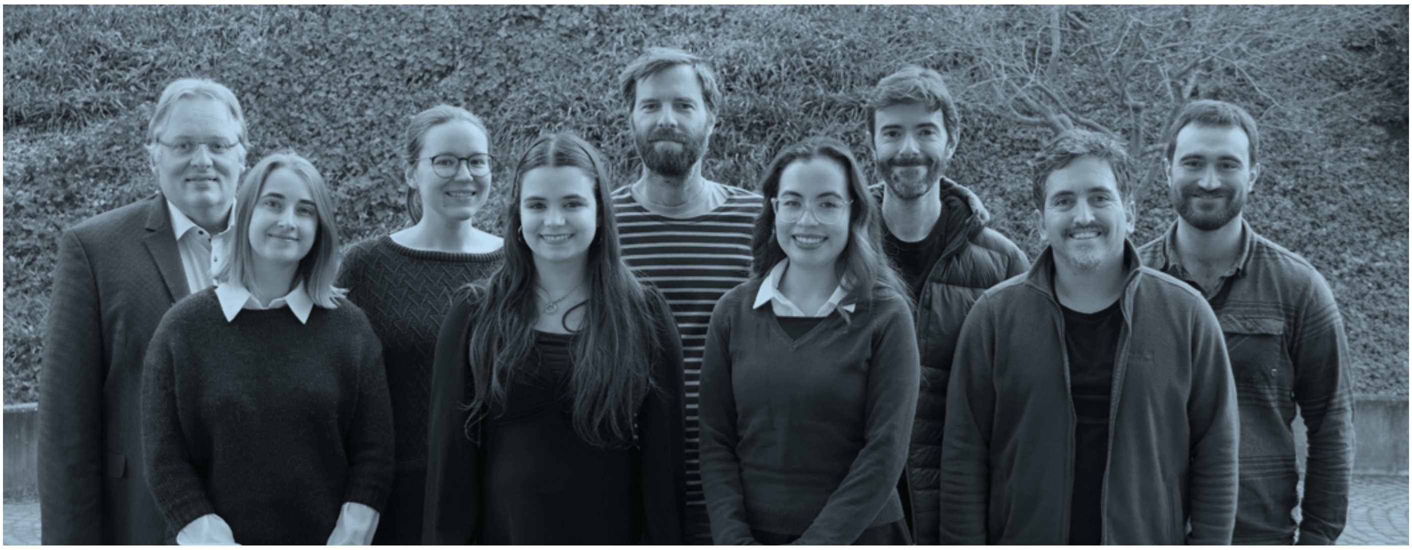
Sprache ist ein wesentlicher Bestandteil der Wissenschaftskommunikation. Wie sie fair und gerecht verwendet wird, steht im Mittelpunkt der Forschung von **Vagrant Gautam**. Vagrant kam im September 2025 als 3. Independent Postdoc ans HITS und arbeitet seitdem eng mit der Gruppe Natural Language Processing zusammen (siehe Kapitel 2.12 und 6.2).

Natural Language Processing war auch das Thema des ersten Symposiums unserer **neuen Reihe „The Future of…“**, das im November 2025 stattfand (siehe Kapitel 5.4). Ziel dieses zukunftsorientierten Veranstaltungsformats ist es, das HITS in seinen Forschungsbereichen als Vorreiter zu etablieren und die Sichtbarkeit der kollektiven Arbeit des Instituts im Einklang mit unserem Motto „Think beyond the limits!“ weiter zu steigern.

Ganz ohne Zweifel gehört es zum Auftrag des HITS, dass seine Forschenden in ihren Forschungsfeldern international führend sind. Das Jahr 2025 war das zehnte Jahr in Folge, in dem Gruppenleiter **Alexandros Stamatakis** – zurzeit auf Kreta als ERA Chair wirkend - als einer der weltweit meistzitierten Forscher im Bereich Biologie und Biochemie auf der Liste der „Highly Cited Researchers“ von Clarivate Analytics stand (siehe Kapitel 2.2, 4 und 10.4).

Gesellschaftliche Einflüsse können Wissenschaft verändern, doch auch die Wissenschaft entwickelt sich gemeinsam mit der Gesellschaft weiter und bewirkt dort Veränderungen. Die Wissenschaft bietet keine einfache Lösung für ein Problem, sondern bewertet Situationen, stellt Fragen und bewertet Herausforderungen neu. Genau das ist es, was wir beim HITS Jahr für Jahr anstreben, und das Jahr 2025 war ein Beleg dafür.

Überzeugen Sie sich selbst - wir wünschen Ihnen viel Vergnügen bei der Lektüre des Jahresberichts!



Research

2.1 AIN

Astroinformatics

Group Leader

Dr. Kai Polsterer

Team

Guillermo Cabrera-Vives
(Klaus Tschira Guest Professor; until August 2025)

Romain Chazotte
(PhD student)

Catarina Corte-Real
(PhD student)

Iliana Isabel Cortés Pérez
(PhD student; HITS Scholarship)

Mariia Demianenko
(visiting scientist)

Dr. Nikos Gianniotis
(staff scientist)

Fenja Schweder
(visiting scientist)

Dr. Franciso Pozo Nuñez

Johanna Riedel
(master's student)

Solomiya Serkiz
(bachelor's student)

Dr. Sebastian Trujillo Gomez

Renuka Velu
(master's student; until January 2025)

In recent decades, computers have revolutionized astronomy. Indeed, 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.

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 supermassive black holes in the centers of galaxies. Driven by these scientific interests, we develop new methods and tools that we share with the community. From a computer science perspective, we focus on time series analyses, sparse-data problems, morphological classification, the proper evaluation and training of models, and the development of exploratory research environments. These methods and tools will prove critical to the analysis of data in large upcoming survey projects, such as SKA, Gaia, LSST, and Euclid as well as next generation of exa-scale simulations.

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.

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.

High-Cadence Optical Reverberation Mapping of AGN in the Local Universe

As part of the ERC Universcale project, we initiated a long-term, high-cadence photometric monitoring program of active galactic nuclei (AGN) entitled "Hubble constant constraints through AGN Light curve Observations" (HALO), which aims to characterize accretion disk reverberation signals with minimal systematic contamination. In this first study, we presented optical continuum reverberation mapping results for the nearby Seyfert galaxy Fairall 9 based on observations obtained between September 2024 and February 2025 with the Cerro Murphy Observatory in Chile. The monitoring was carried out using four Strömgren intermediate-band filters and one Johnson–Cousins filter that had been specifically chosen in order to minimize emission-line contamination and isolate the intrinsic accretion disk continuum. The resulting multi-band light curves showed coherent variability across all bands with sub-day cadence, thereby providing an optimal dataset for measuring inter-band time delays (see Figure 1).

Using these light curves, we measured wavelength-dependent continuum time delays relative to the shortest-wavelength band and built a detailed optical lag spectrum for Fairall 9. The lag measurements revealed clear deviations from a simple smooth wavelength dependence: Excess delays were detected at wavelengths associated with the Balmer and Paschen continua, thereby indicating a non-negligible contribution from diffuse reprocessing in the broad-line region and from dust emission at longer wavelengths. These features introduced “wiggles” into the lag spectrum that would not be able to be captured by standard power-law descriptions and demonstrated that accretion disk sizes that are inferred from such simplified models are highly sensitive to wavelength coverage and filter choice.

We combined our new ground-based data with a homogeneous re-analysis of archival space-based data in order to extend the lag spectrum over a wider wavelength range. The result (see Figure 2, next page) confirms that the observed continuum delays are systematically larger and more structured than standard thin-disk models predict. However, due to the dense cadence and reduced line contamination of the Universcale observations, the inferred accretion disk size for Fairall 9 is smaller than that of previous estimates based on broad-band monitoring, thereby highlighting the value of optimized observing strategies. This result directly impacts ongoing efforts to use continuum reverberation mapping for precision accretion disk studies and cosmological applications.

This work validates the observational design and analysis framework of Universcale and reveals that the project can separate intrinsic disk reverberation from reprocessing effects in surrounding AGN structures. These first results provide a benchmark for future targets in the program and lay the foundation for building a homogeneous AGN sample that is suitable for robust scaling relations and independent constraints on cosmological parameters.

The result confirms that the observed continuum delays are systematically larger and more structured than standard thin-disk models predict

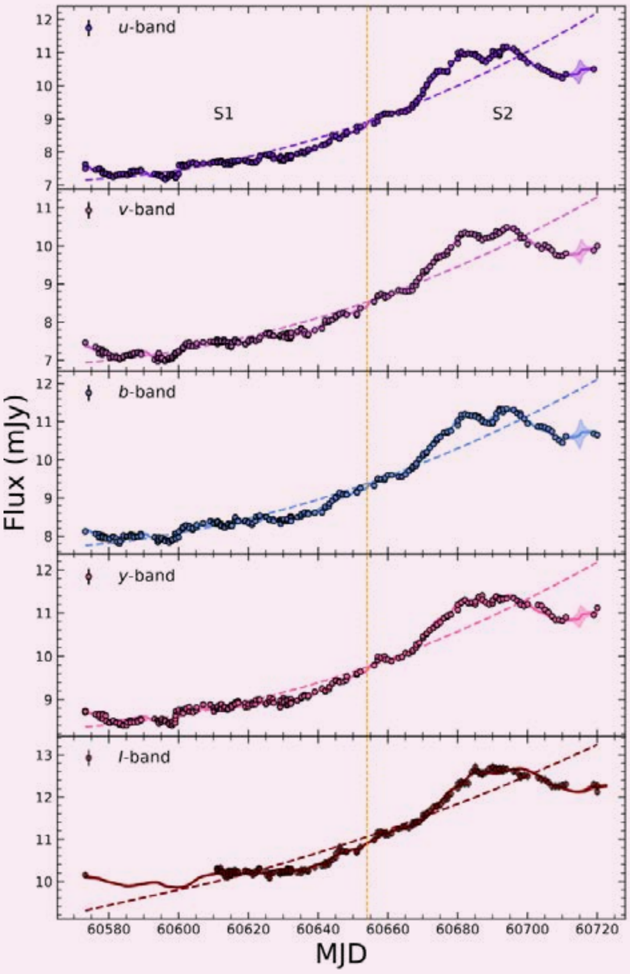


Figure 1
Multi-band optical light curves of the Seyfert galaxy Fairall 9 obtained within the Universcale monitoring program, together with the corresponding inter-band time-delay measurements. The data indicate coherent variability across all filters, thereby enabling robust measurements of wavelength-dependent continuum delays.

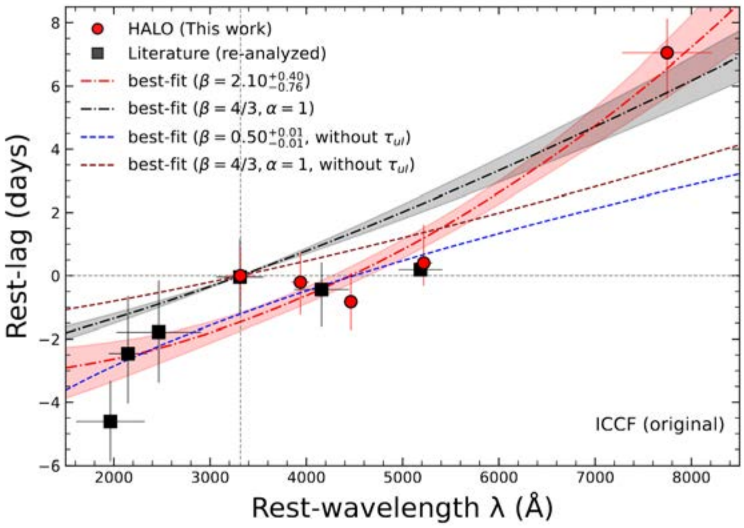


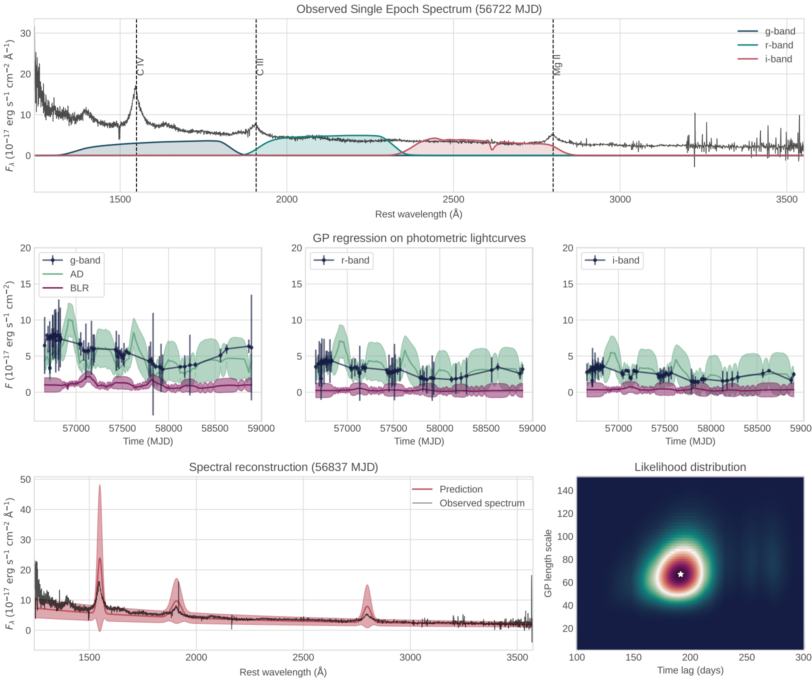
Figure 2
Combined optical–UV continuum lag spectrum of Fairall 9 based on new HALO observations and re-analyzed archival data. The lag spectrum exhibits significant deviations from a smooth wavelength dependence, thereby revealing contributions from reprocessing in the broad-line region and dust in addition to the accretion disk.

> We can estimate both the time delay and spectra from a single epoch spectrum and photometry alone

Photometric Decomposition of AGN Variability via Gaussian Process Modeling

Galaxies that host a central supermassive black hole (SMBH) with a high accretion rate are known as active galactic nuclei (AGNs). Their high luminosity makes them excellent probes of the high-redshift Universe. The optical spectrum of AGNs is dominated by the thermal radiation from the accretion disk (AD) surrounding the SMBH as well as by emission lines from the broad line region (BLR), which consists of dust-free high-density gas clouds. The emission from the BLR is delayed relative to that from the AD. Classical methods of determining this time lag require steady, densely sampled observations over long periods; consequently, only a few hundred sources have been monitored to meet these requirements (e.g., the Reverberation Mapping Project from the Sloan Digital Sky Survey (RM-SDSS)). However, procuring such densely sampled observations for the number of known AGNs placed prohibitively high demands on the monitoring resources.

Our approach exploits AGN variability – which is driven by SMBH activity – in order to infer the time lag between the BLR and AD. We model the latent SMBH with a Gaussian process. We test our method on the RM-SDSS dataset using a 75% training and 25% testing split with photometric light curves artificially constructed by folding the spectral time series with SDSS filters. Our results (see Figure 3) show that we can estimate both the time delay and spectra from a single epoch spectrum and photometry alone.


Observed Single Epoch Spectrum (56722 MJD)

GP regression on photometric lightcurves

Spectral reconstruction (56837 MJD)

Likelihood distribution

Figure 3
Analysis of the source RM-ID 31, located at RA=214.170 deg, DEC=53.115 deg, at redshift z=1.907. Top panel: observed optical spectrum (grey); the transmission curves of the SDSS photometric filters are overlaid, thereby illustrating that these bands are able to capture both radiation from the AD continuum and BLR emission lines. Middle panel: light curves of the same source observed over 57 epochs from the train dataset, which were created by folding the multi-epoch spectra with the three filters displayed in the top panel; the disentangled contributions AD and BLR are shown in green and magenta, respectively. Bottom-left panel: reconstructed spectrum from the test dataset at a randomly selected epoch. Bottom-right panel: likelihood distribution as a function of the time delay and GP length scale.

Physical Analysis of Galaxies via the Spectral Reconstruction of Deep Imaging

Integral field spectroscopy (IFS) provides highly complete information about extended astronomical sources – such as galaxies – by measuring a spectrum at each position across the object. Reconstructing IFS data from photometric imaging is a mathematically ill-posed problem because images lack the information required to uniquely recover highly resolved spectra. We have been developing a machine-learning model that reconstructs the latent spectra that underlie a given set of galaxy images in different wavelength bands while taking into account the spatial information inherent in IFS observations. This approach combines (i) a generative spectral model that captures the main features of spectra with few degrees of freedom and (ii) probabilistic inference that calculates the posterior distribution of spectral reconstructions. The end goal is not only spectral recovery, but also the estimation of astrophysical properties, such as stellar age and the metallicity of the galaxies.

To date, the project has used data from the open-access CALIFA (Calar Alto Legacy Integral Field Area Survey). The developed methods were designed to be generalizable and applicable to other observational or simulated datasets, which will be explored in the future. A working prototype (see Figure 4) can already reconstruct IFS data cubes, although spatial dependencies are not yet taken into account. At present, the prototype uses a basis-function model to parameterize galaxy spectra. Performance has been evaluated by comparing reconstructed spectra with those from open astronomical surveys, and it is expected to improve when more dependencies are accounted.

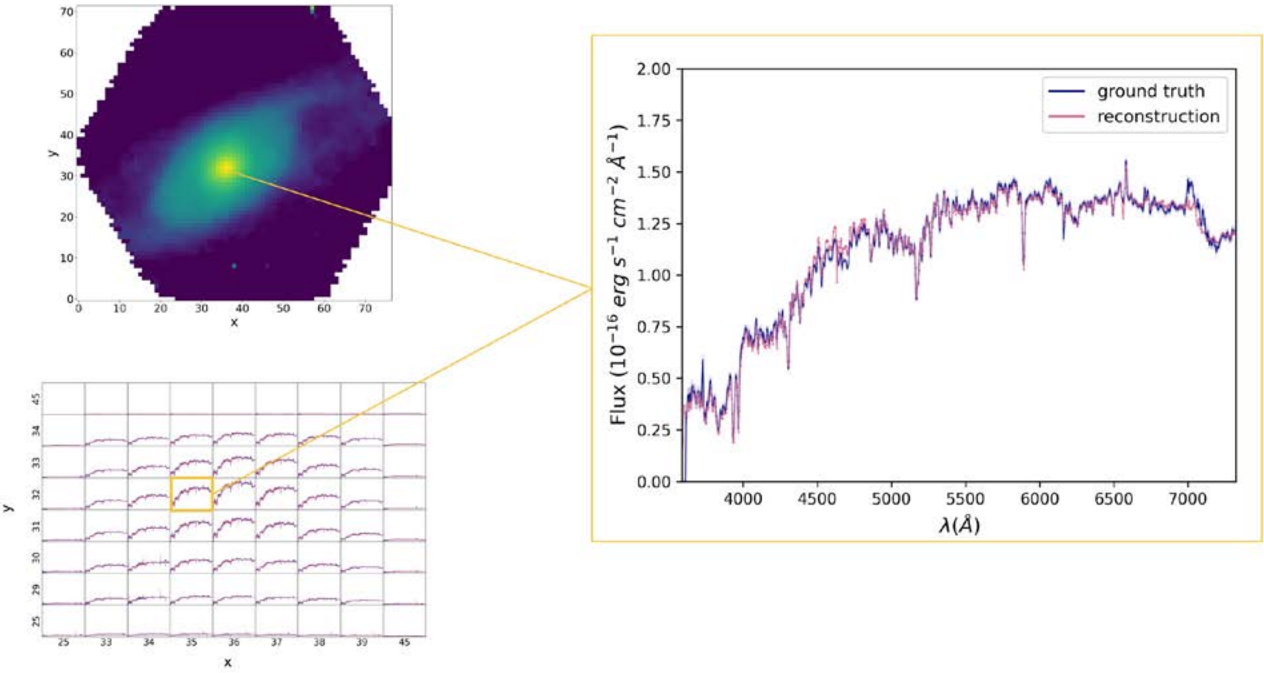


Figure 4
Illustration of the reconstruction of integral field spectroscopy (IFS) from photometric imaging. The top-left figure corresponds to an image of a galaxy, and the bottom-left figure represents the reconstruction of its spectra based on several images in different wavelength bands. The figure on the right is zoomed in on the bottom-left figure in a specific spatial location.

Rotation-Invariant Compression of Data via Zernike Polynomial Representations

With emerging large-scale simulations, the question of how to handle large amounts of data arises. Specifically, exploring the immense data space is no longer possible via simple visual inspection by humans. Neural networks can automate such tasks because they are designed to process immense loads of data at unprecedented speeds. However, this process introduces many new challenges. While humans are able to recognize objects independently of the position of these objects in space, machines struggle significantly with this task. For tasks such as the automatic classification of large sets of data, this introduces a significant problem. A naïve model that groups together data is likely to focus on the orientation of the data rather than on their truly distinctive features. To that end, we constructed a transformation of the data into the space of Zernike polynomials that allows us to treat data invariantly in terms of orientation (see Figure 5). This approach is computationally highly stable, thereby making it applicable beyond its theoretical groundwork. We use it to construct a compressed representation of galaxy data that contains information about the input orientation without taking this information into account for grouping.

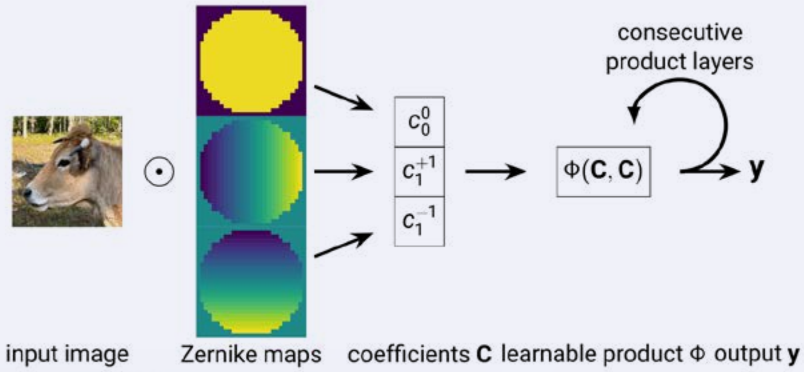


Figure 5
Overview of the network construction: an input image is decomposed into Zernike coefficients that are propagated through multiple layers that preserve O(2) equivariance.

A New Method of Inferring the Structure of Dark Matter Halos Reveals Tension With the Standard Cosmological Model

A key mystery in our understanding of the Universe is the nature of dark matter, which comprises nearly 84% of all matter. The only evidence of its existence comes from its gravitational effects on baryonic matter in galaxies, galaxy clusters, and the cosmic web. The study of this mysterious substance is further complicated by its own gravitational dominance: Dark matter is the scaffolding that sustains the formation of galaxies, and every large clump (a.k.a. halo) of dark matter should gather enough baryonic matter at its center to form a galaxy. Densely packed atoms easily dissipate energy and condense at the centers of galaxies, with the potential to have a profound influence (again, via gravity) on the mass distribution of dark matter halos. This would not be a problem if we could predict precisely how galaxies form and evolve.

Over the last two decades, N-body simulations have produced precise predictions for the distribution of dark matter predictions ranging from the largest scales of the cosmic web down to the structure of individual halos. These predictions require only the relatively simple physics of gravity. One such prediction is the tight correlation between the total mass of a halo and its concentration or central density. The concentration–mass relation of galactic halos is a powerful probe of cosmological and dark matter physics at scales smaller than most observational probes: This relation encodes information about the total matter content of the Universe, the amplitude of primordial fluctuations, and the masses and interactions of dark matter particles, among other things. However, the predictions are expected to break down in regions where baryonic matter has a strong gravitational influence, including within galaxies and clusters. In order to model these effects, we need to include – in addition to gravity – all the extra physics of ordinary matter. Despite our detailed knowledge of the physics of baryonic matter, the spatio-temporal dynamic range required to make such predictions is so large – at more than 6 orders of magnitude – that we are forced instead to use many simplifying assumptions in even the most detailed numerical simulations run on the largest supercomputers. The vast gap across the scales that can be accurately modeled cannot be bridged by current computational techniques in the foreseeable future; therefore, we need to develop more clever ways to approach the problem.

The physics of the unknown dark matter particle seems to be more complex than expected

As a first step, we developed a new inference method that combines data-driven data analysis techniques with a metamodel of the gravitational effects of gas, stars, and massive black holes on the pristine distribution of dark matter as predicted by cosmological models. The metamodel approximates simulation-based inference methods designed to learn the likelihood needed for probabilistic inference directly from synthetic data produced by complex simulations. In our case, we lack the suite of simulations necessary to learn the likelihood. As a rough approximation, we collected all published experiments and estimated the likely range of galactocentric radii where the gravitational effects of baryonic material in the galaxy can influence the structure of the dark matter halos predicted by cosmological models. Predicting the extent of these effects and their uncertainty then allows us to select the pristine outer regions of halos that have evolved unaffected by the galaxy that formed at their center.

One final piece is necessary to complete the puzzle: observational data. Nearly half a century ago, Vera Rubin and collaborators showed that the rotation velocity of gas or stars in a disk galaxy (a.k.a. its rotation curve) can be used to infer this galaxy's mass distribution and demonstrated that galaxies rotate much faster in their outskirts as compared with the expectation from the amount of visible matter they contain. In order to maximize the statistical significance of our results, we collected most of the available high-quality rotation curve data and baryonic mass models of gas-rich galaxies across a broad spectrum of masses and morphological types.

With our method, we can discern the pristine outer regions of the rotation curves. Hierarchical Bayesian inference helps us infer the total mass and concentration of each object as well as the shape of the concentration–mass relation, including properly calibrated uncertainties. The inferred relation is steeper, with denser low-mass halos and more diffuse massive halos than those predicted by the standard ΛCDM cosmological model (see Figure 6). The slope is more than 3σ steeper than the theoretical prediction, thereby suggesting that the physics of the unknown dark matter particle are more complex than expected. We tested the robustness of this result thoroughly and revealed that it is highly difficult to explain away via selection biases or systematics. This may mean – as suggested by some simulations – that rotation curves are poor tracers of underlying gravitational potential. This motivates our next step: developing more powerful probabilistic methods for reconstructing mass distribution directly from synthetic observations using deep learning.

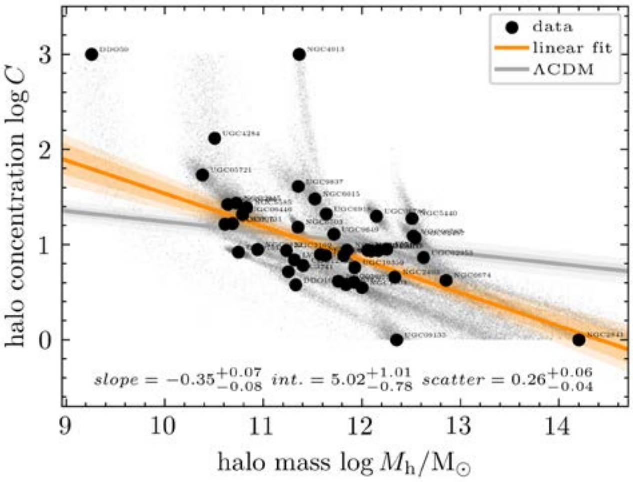
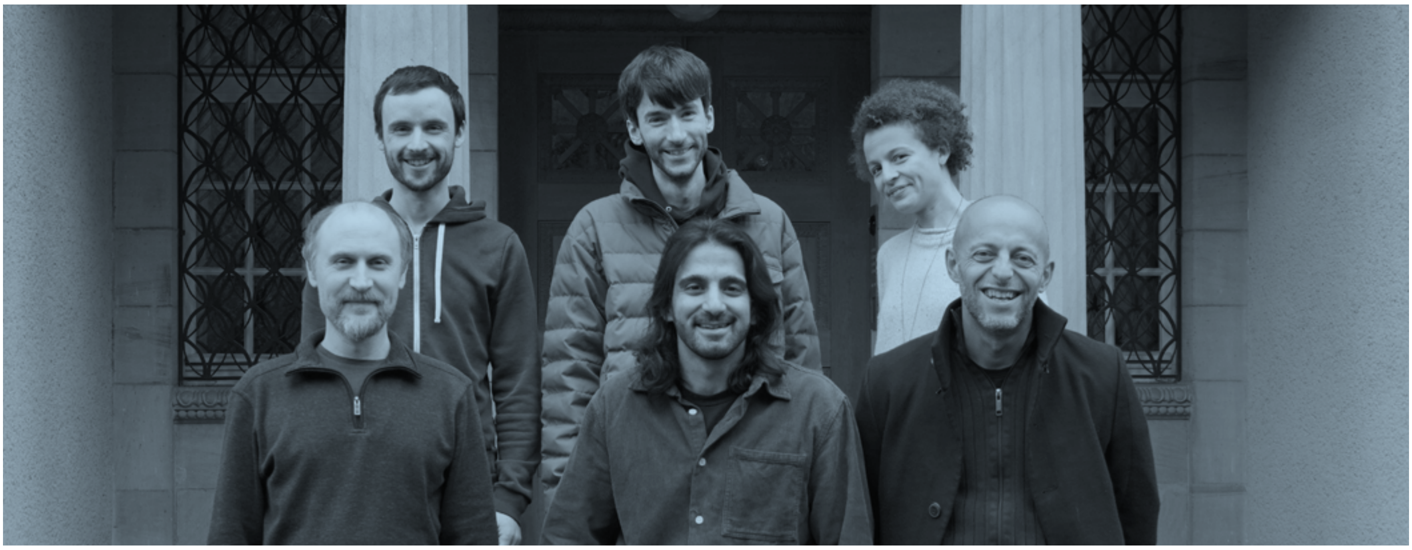


Figure 6
The tension between the concentration–mass relation of dark matter halos inferred from rotation curve data using our method on the one hand and the theoretical prediction from the standard cosmological model on the other hand. The orange line and shading displays a hierarchical Bayesian linear fit to the individual data points and its 1 and 2σ uncertainty bands, while the grey line displays the theoretical prediction.



Research

2.2 CME

Computational Molecular Evolution

Group Leader

Prof. Dr. Alexandros ("Alexis") Stamatakis

Team

Mattis Bodynek
(master's student; until September 2025)

Nam Do
(bachelor's student; since June 2025)

Julia Haag
(PhD student; HITS Scholarship; until June 2025)

Luise Häuser
(PhD student; HITS Scholarship)

Johannes Hengstler
(PhD student; HITS scholarship)

Dimitri Höhler
(PhD student)

Lukas Hübner
(visiting scientist from KIT; until October 2025)

Dr. Alexey Kozlov (staff scientist)

Victoria Kuch
(master's student, since September 2025)

Christoph Stelz
(master's student)

Anastasis Togkousidis
(PhD student)

Julian Wu
(master's student; until August 2025)

The Computational Molecular Evolution group focuses on developing algorithms, models, and high-performance computing solutions for bioinformatics. We focus mainly on

- computational molecular phylogenetics,
- large-scale evolutionary biological data analysis,
- supercomputing,
- quantifying biodiversity,
- next-generation sequence-data analysis, and
- scientific software quality and verification.

Below, we outline our current research activities, which are situated at the interface(s) between computer science, biology, and bioinformatics. The overall goal of the group is to provide evolutionary biologists with new methods, algorithms, computer architectures, and freely available/accessible tools for molecular data analysis.

In other words, we strive to enable research. One aim of evolutionary biology is to infer evolutionary relationships between species as well as the properties of individuals within populations of the same species. As evolutionary biologist Theodosius Dobzhansky's famous and widely quoted dictum states, "Nothing in biology makes sense except in the light of evolution."

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*

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

> In the winter of 2023/24, we launched a new teaching endeavor with the University of Crete

What Happened in the Lab 2025?

Alexis completed the third year of his five-year EU-funded ERA chair project at the Institute of Computer Science within the Foundation for Research and Technology Hellas (ICS-FORTH). His Biodiversity Computing Group (BCG) at ICS-FORTH is fully staffed (with 3 postdocs and 3 PhD students) and is closely linked to the CME group at HITS. For instance, BCG members Georgios Koutsovoulos, Ben Bettisworth, Noah Wahl, Franziska Reden, and Lucia Martin Fernandez are actively involved in supervising several bachelor's and master's theses at the Karlsruhe Institute of Technology (KIT). In addition, KIT master's student Victoria Kuch moved to Crete to be closer to her supervisors.

In the winter of 2023/24, we launched a new teaching endeavor with the University of Crete (UoC)–KIT master's level course "Introduction to Bioinformatics for Computer Scientists," which is taught simultaneously at the computer science departments of KIT and UoC. This course was repeated in the winter of 2024/25, when it reached a record high in exam registrations (40 at KIT and 11 at UoC), and it is also being taught in winter 2025/26, with high enrollment numbers once again.

Because the summer semesters at UoC and KIT mostly do not overlap, Alexis additionally taught a seminar on "Reproducibility in Bioinformatics" during the 2025 UoC spring term for the second time. Finally, during the summer semester of 2025, we again taught our main seminar – "Hot Topics in Bioinformatics" – at KIT. This time, KIT students could choose between writing a standard (slightly boring) seminar report or a (more interesting) reproducibility analysis in analogy to the seminar taught at UoC.

Christoph Stelz – our long-term student member from the Department of Computer Science at KIT (bachelor's thesis and "Praxis der Forschung" ("research practice") project) is currently completing his master's thesis and will begin working on his PhD in January 2026. We also hope to be able to hire one or two new female PhD students in 2026. Finally, our staff scientist Alexey Kozlov continued his part-time role as Sustainability Advisor of HITS with a key focus on technical and computer science–related energy efficiency issues.

Our recurring highlight – namely the summer school on Computational Molecular Evolution – finally took place again in 2025 at the Hellenic Center for Marine Research on Crete for the 15th time. Local BCG lab members Franziska Reden and Ben Bettisworth (former PhD student at HITS/KIT) supported the summer school as teaching assistants. Alexis served as both a lecturer and the main organizer once again. The summer school also included a satellite seminar on Biodiversity Genomics. There will unfortunately be no summer school in 2026 due to restructuring at the Wellcome Trust, which is beyond our control. However, Alexis will organize the 2026 meeting on Mathematical and Computational Evolutionary Biology at the Hellenic Center for Marine Research on Crete (https://mceb2026.sciencesconf.org/).

Finally, Alexis also trained regional talent in a completely different part of the world: Brazil. From 14–17 July, upon invitation by Georgios Pappas, he taught a four-day intensive course on computational molecular evolution at the Biology Department of the University of Brasilia, Brazil, that included practical computer exercises. Figure 7 shows Alexis with some of the participating students in front of a phylogenetic tree at the University of Brasilia.

Figure 7
Alexis with master's and PhD students at the University of Brasilia in front of a phylogenetic tree. On the right-hand side, part of the famous quote by Theodosius Dobzhanskh can be seen in Portuguese: Translated, it states, "Nothing in biology makes sense except in the light of evolution."

Additionally in 2025, Alexis was listed on the Clarivate Analytics list of highly cited researchers for the tenth year in a row. He can also be found at rank #481 (out of 230,334 researchers) on the so-called Stanford list published by John Ioannidis (https://elsevier.digitalcommonsdata.com/datasets/btchxktzyw/8), which comprises the top 2% of the most highly cited researchers worldwide and relies upon different criteria than the Clarivate list.

A plethora of press articles also appeared in 2025, mainly covering the desolate state of current research policy in Greece, including the introduction of the term "brain re-drain" (see https://www.dnews.gr/eidhseis/ellada/542584/brain-gain-pos-i-ellada-xantakonei-tin-erevna-kai-dioxnei-ta-myala-tis-osa-apokalyptei-o-dr-al-stamatakis for an example from a national Greek news portal).

Another highlight of 2025 was that the second iteration of the "Machine Learning for Evolutionary Genomics Data" conference took place in the French Alps. The inaugural iteration of the conference had been organized by Alexis on Crete in 2024.

Another important highlight of the year was that Julia Haag and Lukas Hübner very successfully defended their PhD theses at KIT in the summer of 2025. After having attended Alexi's bioinformatics classes at KIT and carried out their master's theses in our lab, it was time to say goodbye to both long-term lab members, whom we will miss.

In sum, 2025 was characterized by the consolidation of the BCG research activities on Crete and by high productivity at CME.

> Alexis was listed on the Clarivate Analytics list of highly cited researchers for the tenth year in a row

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.

However, 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, the manual reconstruction of trees is no longer feasible. Evolutionary biologists thus have to rely on computers and algorithms for phylogenetic and population genetic analyses.

Ever since the introduction of so-called "next-generation sequencing machines" (i.e., machines that biologists use in the wet lab to extract DNA data from organisms), scientists have been able to generate billions of short DNA fragments (each containing between 30 and 400 DNA characters). Due to continuous improvements in sequencing technology, we can now also obtain substantially longer sequences spanning thousands of characters (also called long-read sequencing technology). As a result, the community as a whole currently 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.

> We are constantly trying to catch up with the data avalanche and to make molecular data analysis tools more scalable with respect to dataset size

This scalability challenge additionally entails reproducibility challenges, especially when using parallel programs. For instance, with our KIT master's student Christoph Stelz, we recently completed an investigation pertaining to the trade-offs between the increased reproducibility of parallel software on the one hand and execution times on the other hand (see preprint at https://www.biorxiv.org/content/10.1101/2025.06.02.656320v1.abstract). With Christoph, we are currently also implementing fast methods for selecting the best-fit statistical model of evolution for a given dataset in our flagship tool, RAxML-NG, which conducts maximum likelihood–based phylogenetic inferences.

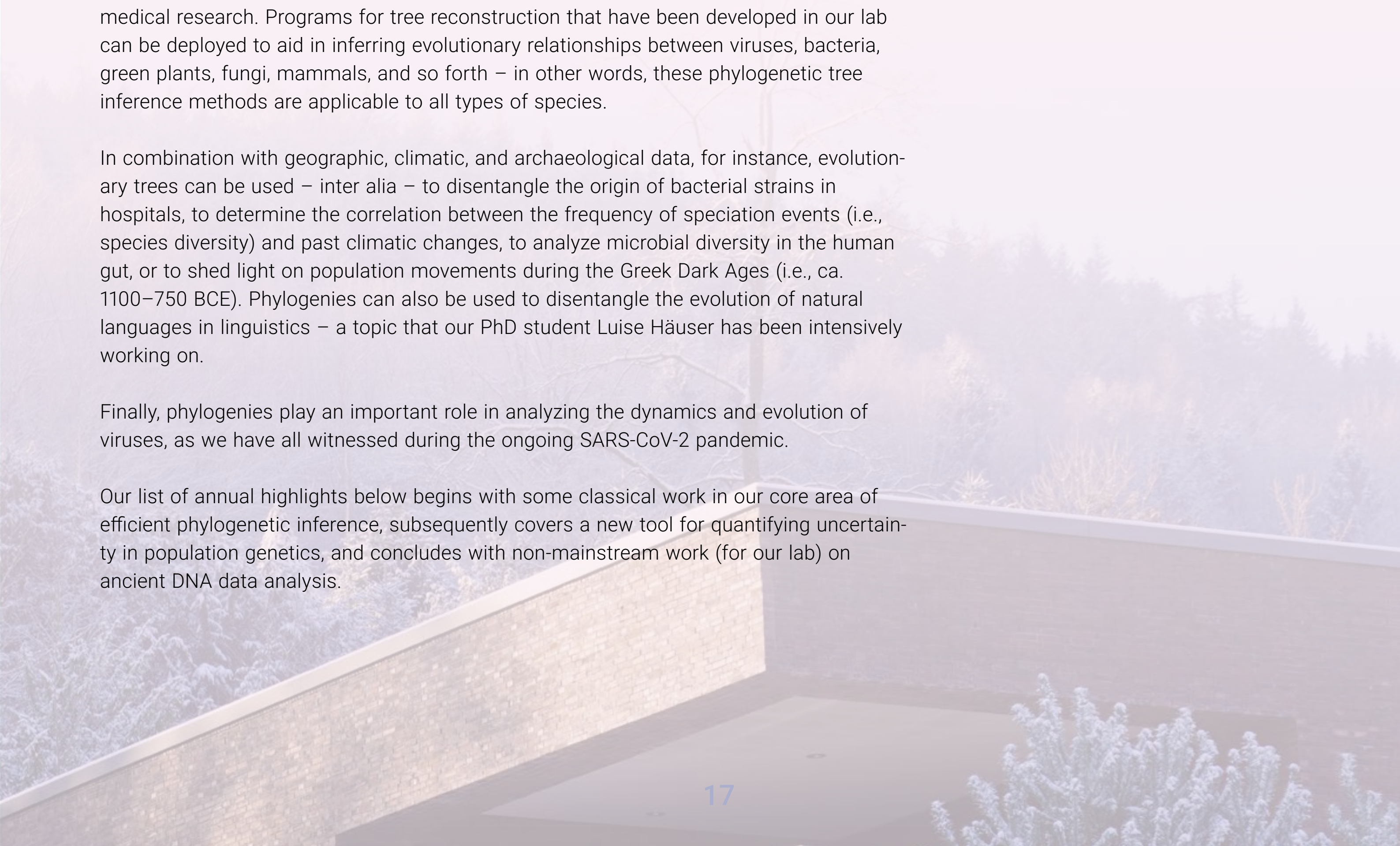
Another emerging line of research is the deployment of machine learning techniques and AI methods to address problems in phylogenetics and bioinformatics, partially also because most current KIT master's students are looking for thesis topics in the area of data science. For instance, with KIT master's student Mattis Bodynek, we managed to show that we can reasonably quantify as well as accurately predict the degree of difficulty of the so-called Multiple Sequence Alignment (MSA) problem (a fundamental bioinformatics analysis step) given a set of unaligned molecular sequences. This difficulty prediction can inform an adequate MSA analysis setup. Interestingly, we also found that MSA difficulty is anti-correlated with phylogenetic difficulty (as introduced by our former PhD student Julia Haag three years ago) – that is, it is generally difficult to infer a reliable tree on sequences that are easy to align, and vice versa.

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, and so forth – in other words, these phylogenetic tree inference methods are applicable to all types of species.

In combination with geographic, climatic, 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 – a topic that our PhD student Luise Häuser has been intensively working on.

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.

Our list of annual highlights below begins with some classical work in our core area of efficient phylogenetic inference, subsequently covers a new tool for quantifying uncertainty in population genetics, and concludes with non-mainstream work (for our lab) on ancient DNA data analysis.



Evading Over-Optimization Under Maximum Likelihood

Our PhD student Anastasis worked on a question we have been wondering about for a very long time: Given that Maximum Likelihood–based phylogenetic tree searches tend to exhibit an asymptotic convergence behavior over execution time with respect to likelihood improvements, can we come up with a method to terminate these tree searches earlier while still obtaining equally accurate trees?

The plain answer is: Yes. In a paper published in Systematic Biology (https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syaf043/8154032), Anastasis demonstrated that by modifying the search algorithm of our widely used RAxML-NG open-source tool and applying the so-called Kishino–Hasegawa Test, we are indeed able to terminate searches earlier during their asymptotic phase while obtaining equally good trees. More specifically, Anastasis showed that for 98% of the tested empirical DNA datasets, his early stopping method returned trees that were statistically equivalent to those inferred via the standard RAxML-NG baseline version. For protein datasets, the fraction of datasets in which Anastasis' version inferred statistically equivalent trees amounted to 92%. In conjunction with algorithmic tree search innovations, Anastasis' stopping criterion was on average 5 times faster on DNA datasets and 3.9 times faster on protein datasets compared with the baseline version. As always, the code is available as open-source code.

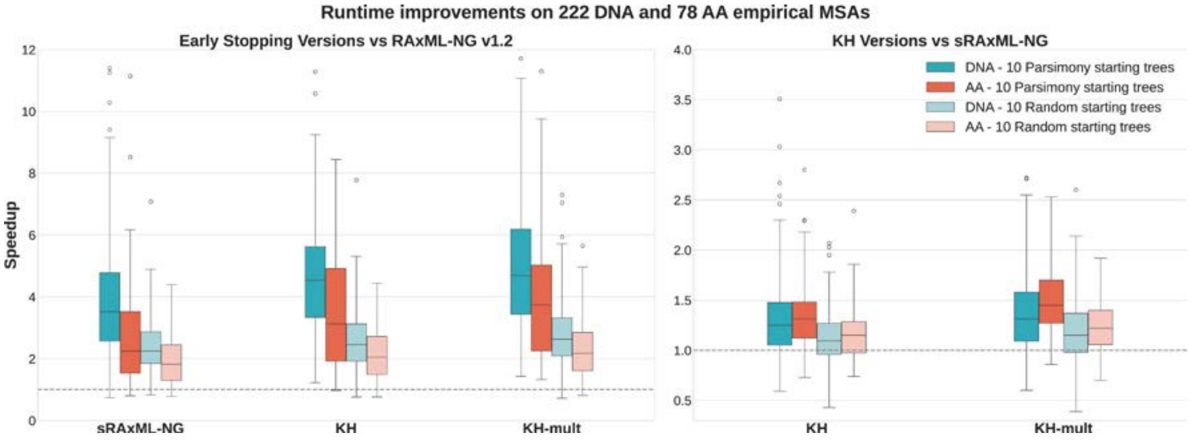


Figure 8
Run-time improvement factors (speedups) of the new RAxML-NG versions that Anastasis developed versus the baseline algorithm (dotted line with a speedup of 1 along the y-axis) on empirical DNA and protein datasets. The sRAxML-NG version is the version that only includes algorithmic innovations, while KH and KH-mult denote distinct applications of the Kishino–Hasegawa Test for stopping searches earlier during the asymptotic convergence phase."

Opening Pandora’s Box in Population Genetics

Ever since we began conducting some work related to ancient DNA (see below), we have been suspicious of the very low percentage of variation explained by the first two principal components of standard principal component analyses (PCAs). Conducting a PCA typically constitutes the first analysis step when beginning to explore population genetics data (i.e., sequences from distinct individuals belonging to the same species). Therefore, Julia attempted to devise a method for quantifying the variance of PCA analyses under slight alterations of the input data. This was achieved by bootstrapping the input data matrix, which comprises single nucleotide polymorphisms (so-called SNPs) that capture the genomic sites at which genomes from the same species differ from one another. This bootstrapping approach now allows for computing global stability indices (over all samples in the input dataset) as well as per-sample/per-individual stability indices, which essentially quantify the variance of the locations of the individuals in PCA plots – or in other words, whether these individuals remain at the same location or begin “hopping” around under bootstrapping. The approach was implemented in an open-source tool called Pandora (see related paper at https://academic.oup.com/bioinformaticsadvances/article/5/1/vbaf040/8051114), and we hope that it will be routinely used for assessing the PCA analysis stability of population genetics data, particularly when the dataset also comprises ancient DNA samples that are more fragmented and hence exhibit decreased signal strength. In fact, we were able to identify some published empirical datasets with unstable support values that would have gone unnoticed without Pandora, as shown in Figure 9.

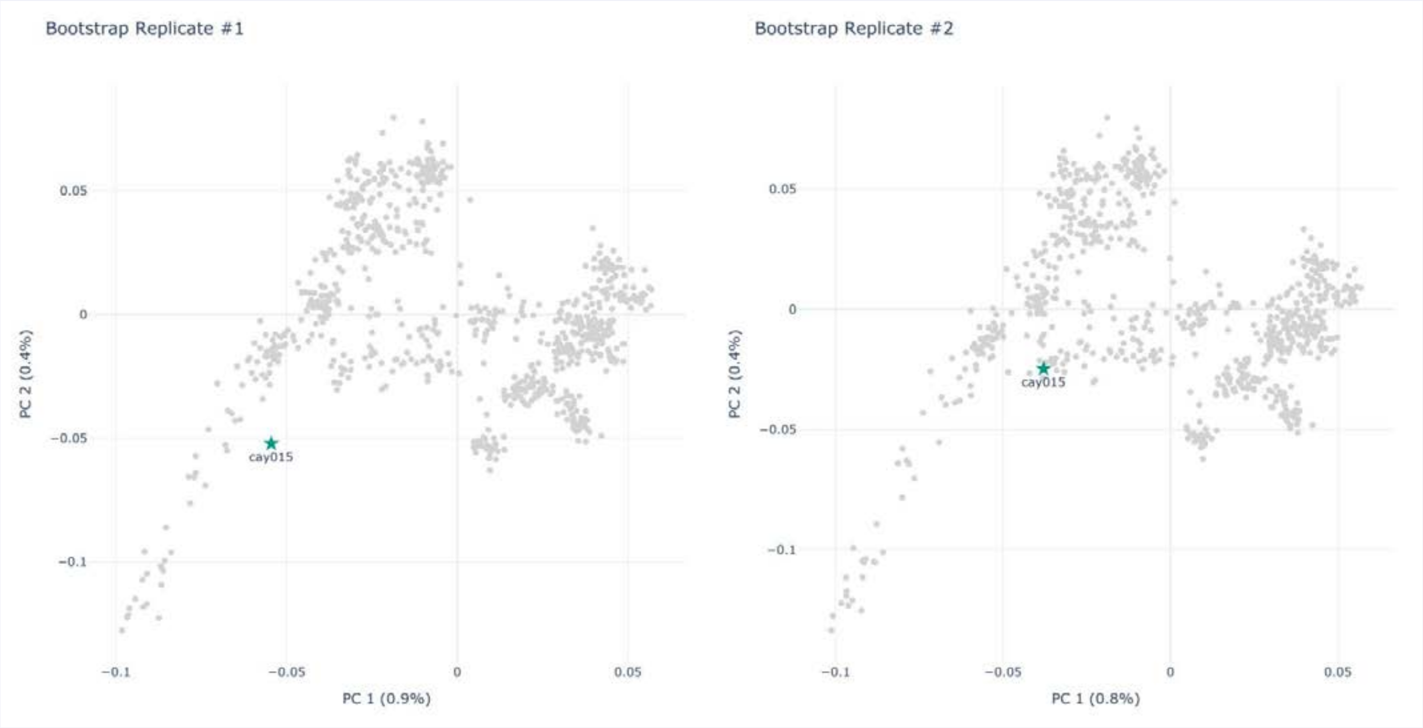


Figure 9
Distinct PCA placements of an ancient Çayönü individual (cay015) in two distinct PCA bootstrap replicates. The dataset was obtained from colleagues in Ankara, Turkey. The ancient cay015 individual receives a low per-individual/per-sample Pandora stability value.

Greek Colonization Through the Lens of Ancient DNA

Moving away from method development and toward ancient data analyses, we recently attempted to shed some light on the foundation of Corinthian colonies (https://www.biorxiv.org/content/10.1101/2025.07.01.662689v1.full.pdf) in collaboration with the Ancient DNA Lab at the Institute of Molecular Biology at the Foundation for Research and Technology Hellas on Crete. Interestingly, the author list comprises former CME visitor Nikos Psonis as well as former CME postdoc Pavlos Pavlidis. For the PCA analysis part of this paper, we already used Pandora (see above).

However, back to the main topic of disentangling the genetic footprint of Corinthian colonization: During Ancient Greek colonization, Corinth established a stable network of economic and political ties by founding colonies that connected southern Greece with the mainland of Epirus (located in the NW part of modern mainland Greece) and that reached as far as the eastern Adriatic coast. Amvrakia – one of the main Corinthian colonies, which was founded during the 7th century BCE – was characterized by its strong dependence on its metropolis. We aimed to investigate the genetic relationships between the Corinthian metropolis and the Amvrakia colony, the contribution of the local population to the founding genetic pool, and the Amvrakian demography in subsequent periods. We found that during its foundation in the Archaic period, Amvrakia appears to have been shaped by genetic influences from at least two distinct sources. The first source migrated from the Corinthian territory, which is represented by the Archaic Tenea population and is supported by an identity-by-descent (IBD) analysis. The second source shows a direct ancestry from Late Bronze Age (LBA) / Iron Age Greece, including a local LBA population represented by the Ammotopos site located close to Amvrakia, as shown by a plethora of independent population genomics analyses. During the subsequent Classical and Hellenistic periods, the population of Amvrakia appears to have slightly differentiated, yet we do observe evidence of genetic continuity over time.

We thus concluded that the migration of Corinthians to Amvrakia contributed to the initial genetic pool of the colony in conjunction with the local genetic pool. This conclusion indicates that the Corinthian colonization included both genetic and cultural transmission between the metropolis and its colony.

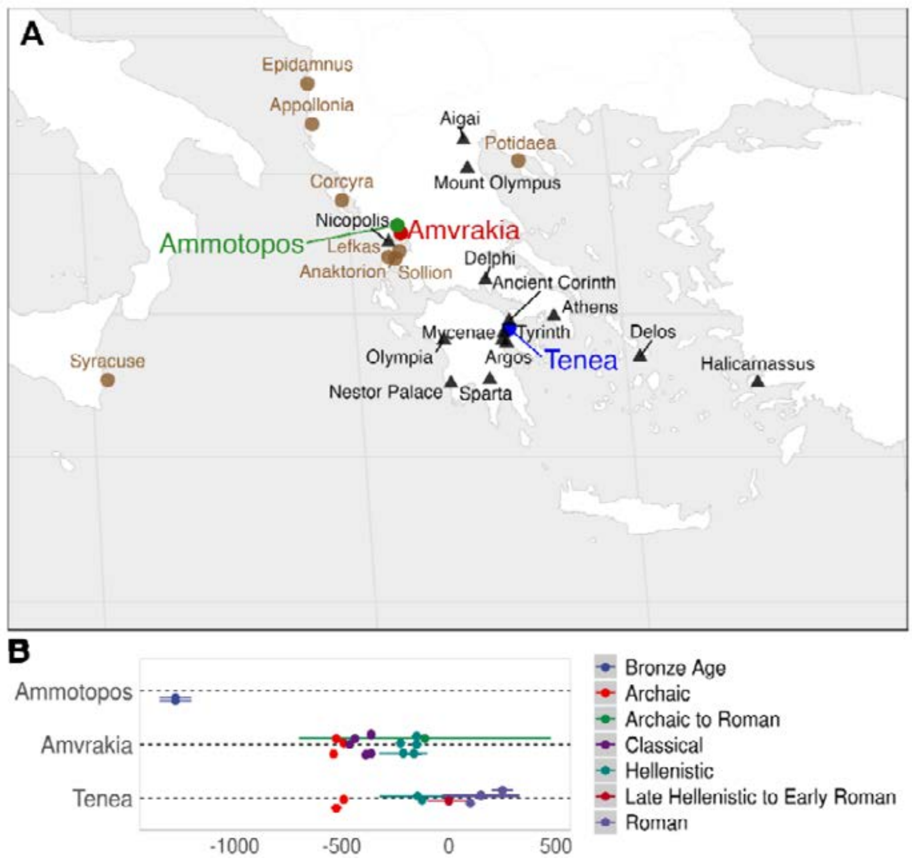
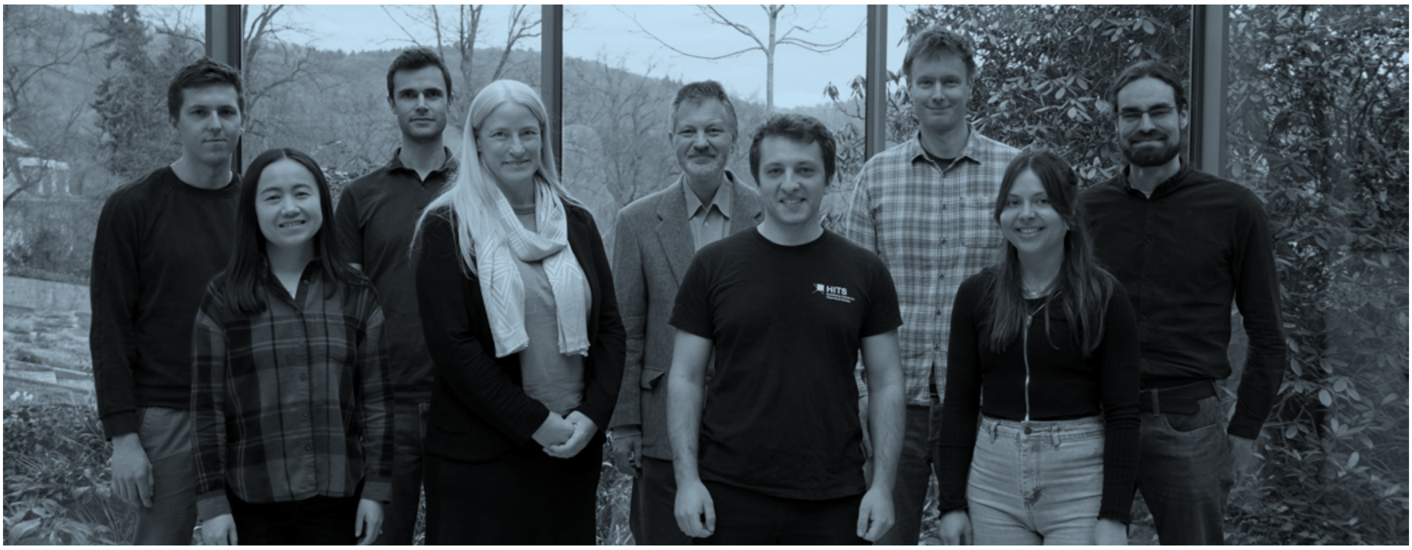


Figure 10
Part A. Location of Ammotopos (green), Amvrakia (red), and Tenea, the latter of which serves as a proxy for Corinth on the contemporary geographic map of Greece. The main Corinthian colonies are displayed in brown. Other important ancient locations are marked with black triangles.

Part B. Temporal coverage and geographic provenance of the 26 new ancient genomes published in conjunction with our study. The x-axis is a time axis in years, with 0 denoting the beginning of the current epoch.

> Corinthian colonization included both genetic and cultural transmission between the metropolis and its colony



Research

2.3 CST

Computational Statistics

Group Leader

Prof. Dr. Tilmann Gneiting

Team

Prof. Dr. Sándor Baran
(visiting scientist; University of Debrecen; July 2025)

Jun.-Prof. Dr. Johannes Bracher
(visiting scientist; Karlsruhe Institute of Technology)

Tenure-Track Prof. Dr. Timo Dimitriadis
(visiting scientist; Goethe University Frankfurt)

Dr. Alexander I. Jordan
(staff scientist)

Kristof Kraus
(PhD student; since March 2025)

Prof. Dr. Sebastian Lerch
(visiting scientist; University of Marburg)

Christoffer Kinttof Øhlenschlæger
(visiting scientist; University of Copenhagen;
since August 2025)

Dr. Marc-Oliver Pohle
(visiting scientist; Karlsruhe Institute of Technology)

Dr. Johannes Resin
(visiting scientist; Goethe University Frankfurt)

Prof. Dr. Melanie Schienle
(senior researcher)

Prof. Dr. Victoria Stodden
(visiting scientist; University of Southern California;
February–November 2025)

Evgeni Ulanov
(visiting scientist; MPIP Mainz)

Dr. Eva-Maria Walz
(postdoc)

Daniel Wolffram
(visiting scientist; Karlsruhe Institute of Technology)

Assistant Prof. Dr. Lu Yang
(visiting scientist; University of Minnesota;
February–May 2025)

Prof. Dr. Johanna Ziegel
(visiting scientist; ETH Zurich)

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 focus of the group’s research lies in the theory and practice of forecasting.

As the future is uncertain, forecasts should be probabilistic in nature, which means that they should take the form of probability distributions over future quantities or events. Accordingly, over the past several decades, we have borne witness to a trans-disciplinary paradigm shift from deterministic (or point) forecasts to probabilistic forecasts. The CST group seeks to provide guidance and leadership in this transition by developing both the theoretical foundations for the science of forecasting and cutting-edge methodologies in statistics and machine learning, notably in connection with applications.

While this year’s report focuses on a joint project with seismologists, weather forecasting and collaborative research with meteorologists continue to be prime examples of our work.

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. Unser diesjähriger Bericht legt jedoch den Schwerpunkt auf ein Projekt zur Evaluierung von Erdbebenvorhersagen.

General News

In 2025, we were delighted to host multiple guests. In February, Victoria Stodden from the University of Southern California joined our research group for much of the year. A highlight was her thought-provoking Colloquium talk in March, in which she discussed how the scientific landscape might change in the advent of AI and how we can continue to verify the correctness of computational results. Also arriving in February and staying for three short but fruitful months, Lu Yang from the University of Minnesota brought new ideas around calibration – one of the core research competencies of the CST group. As in most recent years, Sándor Baran from the University of Debrecen spent the month of July in Heidelberg, which led as always to inspiring discussions in weather forecasting and post-processing. Finally, Christoffer Øhlenschlæger from the University of Copenhagen came for a six-month research stay during his PhD studies, bringing expertise in extreme value theory and connecting to our ongoing research into correlation measures.

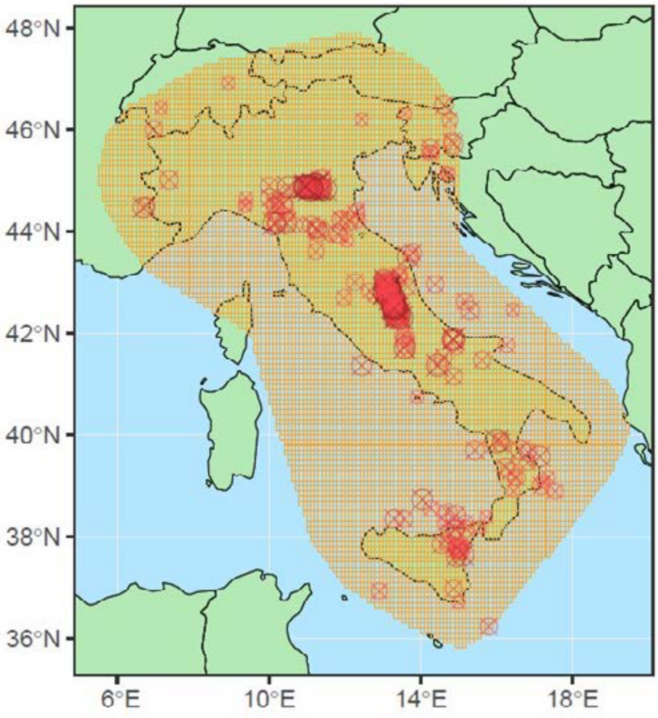
Aside from the daily interactions with our guests and colleagues, we also took a day-long joint trip with the MLI group. After the event had been postponed from its original date in May, we finally spent a somewhat rainy September day on a boat trip from Heidelberg to Neckarsteinach and took a subsequent hike past four castles back to Neckargemünd. Despite the weather, everyone was in good spirits and left the "Alte Scheune" in the evening well fed before catching their respective trains home.

Scientific Report – Earthquake Forecasting in Italy

The earth is constantly moving – not only in space, but also within itself in a process referred to as plate tectonics. In extreme cases, the movements yield earthquakes with substantial impact on the local population as well as on all of society. Therefore, earthquake forecasts are in critical demand. The cover of the 2025 HITS Annual Report stems from a pertinent project [Brehmer et al., 2025], which we report on here in greater detail. The project's aim was to enhance the testing phase proposed by the Collaboratory for the Study of Earthquake Predictability (CSEP). Any model needs to pass evaluation by CSEP standards before inclusion in the operational earthquake forecasting system in Italy (OEF Italy). Our project partners from the University of Naples and members of CSEP – Marcus Herrmann and Warner Marzocchi – were interviewed for a January 2026 piece in Scienza in Rete (an online Italian scientific newspaper born from the collaboration with the Gruppo 2003 of highly influential Italian scientists).

> The cover of the 2025 HITS Annual Report stems from a pertinent project, which we report on here in greater detail

Italy is one of the regions on earth that are disproportionately affected by frequent earthquakes. Along the entire Italian peninsula, the Apennine Mountains demarcate the boundary between the Eurasian and the Adriatic plates, whereas in the south of Italy, the African plate is sliding beneath the Eurasian plate, forming the famous Etna volcano in Sicily. Almost the entire country is regularly subjected to earthquakes, as shown in Figure 11. The figure displays the occurrence of earthquakes whose magnitude surpassed a value of 4.0 (M4+ earthquakes) within the 15-year period from 2005 to 2020.



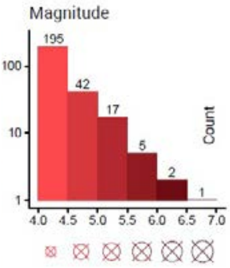


Figure 11
Left: Forecast region of Operational Earthquake Forecasting (OEF) Italy (orange grid, 8,993 grid cells, testing region of the Italian CSEP experiment) and locations of observed M4+ target earthquakes (crossed circles) between 16 April 2005 and 20 May 2020. Right: Logarithmic bar plot of earthquake magnitudes. Source: Adapted from [Brehmer et al., 2025; Figure 1].

Around magnitude 4.0, an earthquake is felt by most people, and minimal damage can be expected that often merely involves fallen household items and rarely building damage. Determining the magnitude of an earthquake is a complex research endeavor in itself. The original “Richter” scale – which was developed by Charles F. Richter in 1935 – measures the maximal amplitude as recorded by a seismograph. Today, the most frequently used magnitude scale for larger earthquakes aims to measure how much seismic energy is released. However, all magnitude scales use a logarithmic scale.

Predictability of Earthquakes

While it is important to investigate the degree to which a certain magnitude correlates with damage to lives and property, the CSEP Italy experiment focuses on the predictability of earthquakes by specifically considering daily predictions of the number of M4+ earthquakes within the subsequent seven-day time frame. Owing to the physical complexity of earthquake occurrence and our inability to precisely determine the state of the tectonic plates, earthquake forecasts are unavoidably probabilistic, meaning they provide all or part of the predictive distribution of future seismicity.

This inherent forecast uncertainty results in challenges to communication between scientists, decision makers, the press, and ultimately also the public. Over four months in 2009, the city of L’Aquila, Italy, experienced many occasional low- and medium-magnitude tremors – that is, seismic swarms that concluded in a major earthquake that led to hundreds of casualties and thousands of damaged buildings. Leading up to the earthquake, anxiety among the residents had grown to the point that a meeting of the Commissione Grandi Rischi (Major Risks Commission) was called, which was followed by a press conference that did not recommend any specific additional measures of protection to the residents. As a major shock to the residents, the tragic earthquake occurred less than one week later.

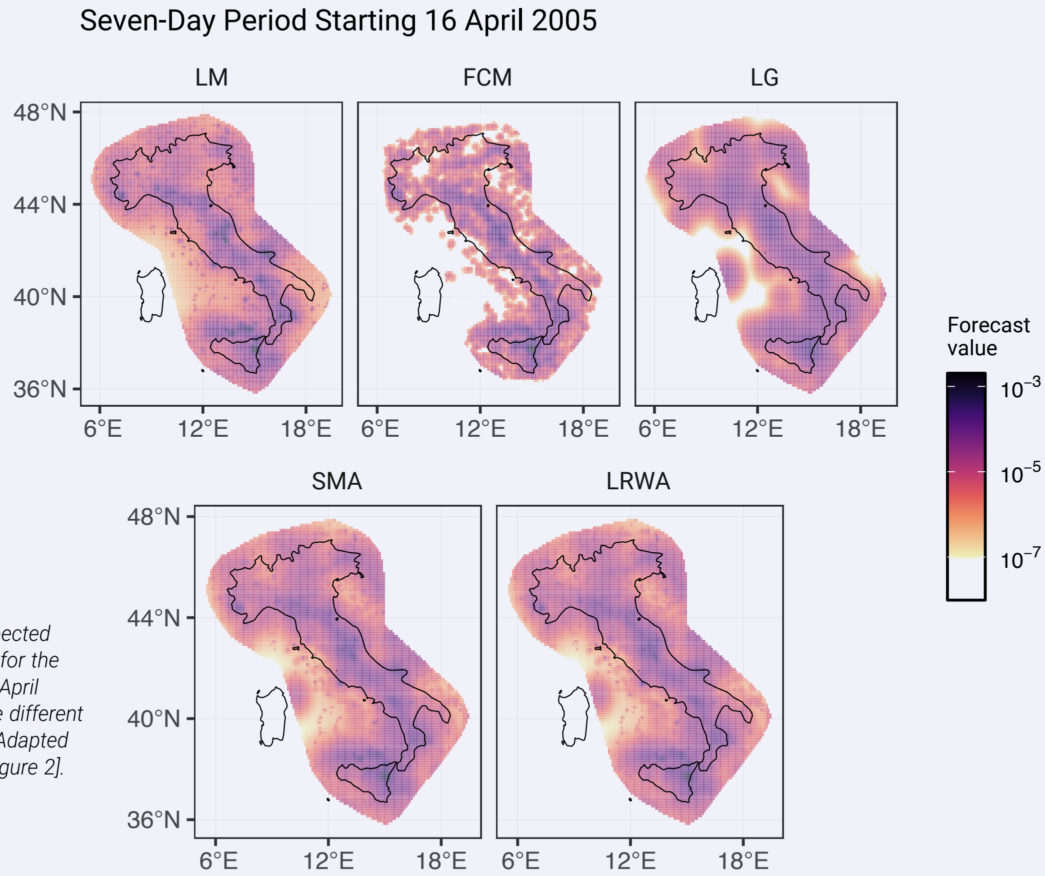


Figure 12
Spatial distribution of the expected number of M4+ earthquakes for the seven-day period starting 16 April 2005 in Italy according to five different forecasting models. Source: Adapted from [Brehmer et al., 2025; Figure 2].

> Short-term earthquake forecasts operate within a low probability – or low-count – environment

Evidently, short-term earthquake forecasts operate within a low probability – or low-count – environment. Figure 12 illustrates the expected number of earthquakes in a seven-day period starting 16 April 2005 in the Italian CSEP testing region. Five different prediction models issue competing forecasts. For this period, the forecast values for a typical location indicate a risk of only one M4+ earthquake in more than 1,000 similar weeks. However, predicting the expected number of earthquakes allows for convenient aggregation over space. For example, Figure 13 displays the sum of the expected counts over the entire CSEP testing region for every seven-day period in the evaluation period. Of course, such aggregation helps little to avoid tragedies, but it can help in planning central resources for disaster recovery that can be deployed nationwide.

In our study, we compared five models, each falling into one of two categories: (1) time-varying stochastic process models and (2) weighted-average ensemble models. The stochastic process models assume cascading earthquakes, in which a random trigger earthquake causes aftershocks whose number and magnitude depend on the magnitude of the initial earthquake and of other previous aftershocks. Three models are of this type: LM, FCM, and LG, all of which are named after the first letters of the proposing authors' surnames and differ in their distinct customized structure and calibration choices. Both of the remaining two models – the score model average (SMA) and the logistic regression–based weighted average (LRWA) – combine the former three models into one weighted-average ensemble model. The SMA weights are proportional to the inverse log-likelihood – that is, they depend on the predictive performance of the individual models – whereas the LWRA weights are estimated based on a logistic regression model.

Comparing Earthquake Forecasting Models

The CST group contributes to the task of earthquake prediction with our expertise in measuring the forecast quality of competing models and identifying systematic errors in forecasting models. The former is done using consistent scoring functions that yield negatively oriented penalties. Subsequent statistical testing informs whether a difference in average scores between forecasts might be due to chance alone.

The forecasting models that we consider predict an expected number of earthquakes over a given seven-day period. Therefore, we use scoring functions that are minimized (on average) when the expectation of a "true" underlying distribution is quoted as the prediction. Famously, the squared error is one such scoring function. The Poisson scoring function is more suitable in the context of earthquake forecasting, where it coincides with the log-likelihood of a Poisson-distributed random variable. However, the connection is purely formal, and the use of the Poisson scoring function neither explicitly nor implicitly requires Poisson distributions from the forecasting model. Our work clarifies that the Poisson score is more broadly applicable than previously thought.

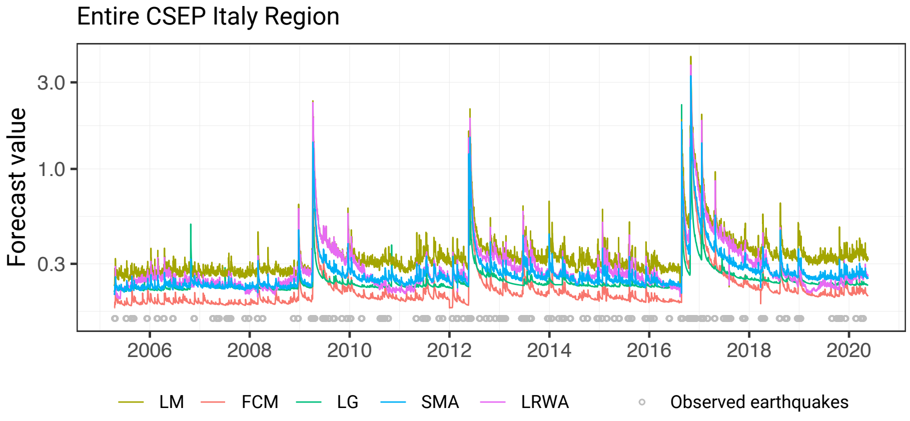


Figure 13
Expected number of M4+ earthquakes in the Italian CSEP testing region from 16 April 2005 to 26 May 2020 according to five different forecasting models. Source: Adapted from [Brehmer et al., 2025; Figure 2].

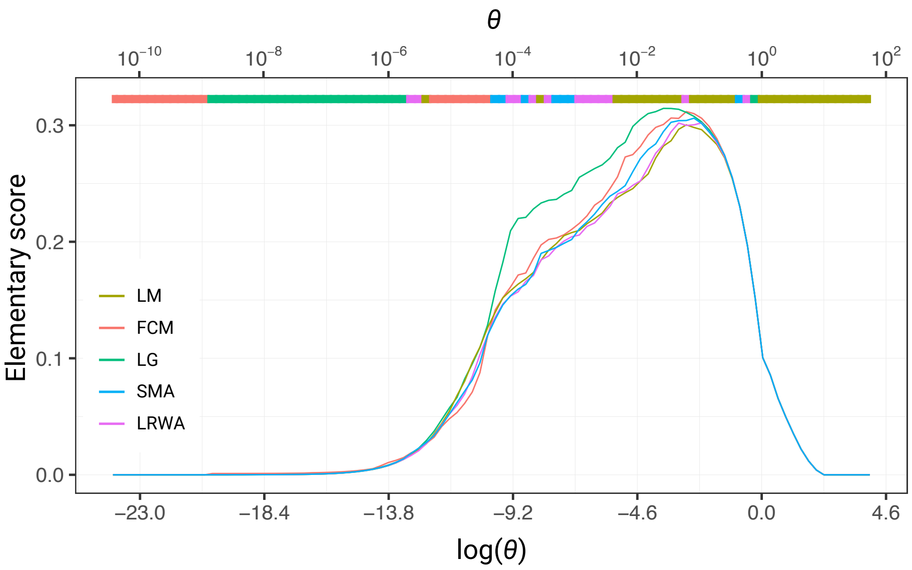


Figure 14

Logarithmic Murphy diagram for the five forecast models. Each curve plots a model's average elementary score against the logarithmic threshold parameter. The average elementary score only penalizes forecasts that are on the opposite side of their corresponding observation relative to the threshold parameter. The colored bar at the top indicates the best model at the respective threshold. The integral under a model's curve equals the average Poisson score. Source: Adapted from [Brehmer et al., 2025; Figure 3].

The Poisson score is just one member of a whole class of consistent scoring functions. In the 2015 Annual Report, we discussed how these scoring functions can be built from the class's elementary members, which constitute a subclass that is small enough to allow for convenient graphs known as Murphy diagrams. Figure 14 depicts such a Murphy diagram that evaluates the five forecasting models. Due to the nature of earthquake data, the horizontal axis is logarithmic, and this choice recovers the average Poisson score as the area under the respective elementary score curve. The visual impression that the lines corresponding to the LM and LRWA models are lowest on average is confirmed by the diagonal entries in Table 1, which report the average Poisson scores.

Model	**LRWA**	**LM**	**SMA**	**FCM**	**LG**
LRWA	**2.69**	1.17	1.64	2.27	2.84
LM	0.12	**2.71**	0.60	1.56	2.45
SMA	0.05	0.27	**2.73**	2.46	3.05
FCM	0.01	0.06	0.01	**2.80**	2.83
LG	0.00	0.01	0.00	0.00	**3.02**

Table 1. Diebold–Mariano test of the null hypothesis of equal predictive ability in terms of the Poisson score for the five forecast models. Source: Adapted from [Brehmer et al., 2025; Table 2].

The off-diagonal entries in Table 1 are related to the Diebold–Mariano test of equal predictive ability, which is a carefully adapted Student's t-test that checks whether the population difference in average Poisson score between two forecasting models could plausibly be 0. The values below the diagonal in Table 1 are one-sided *p*-values that indicate whether the deviation of the average score difference from 0 is statistically significantly in favor of the forecast model with the better Poisson score. The values above the diagonal are the corresponding *z*-statistic values, where large values are interpreted analogously. The bottom row (and the rightmost column) of Table 1 confirms the visual impression in Figure 14 that the LG model is consistently beaten by all other models in terms of predictive performance. Importantly for formal inference, the proposed implementation of the Diebold–Mariano test more accurately accounts for sampling variability than does the current CSEP T-test implementation. With the proposed correction, we can be more confident in the results of the statistical test.

Identifying Model Deficiencies

In addition to comparative evaluation, which answers the question of whether a model is better than its competitors, we also need diagnostic tools with the goal of identifying – and eventually remedying – the deficiencies of any given model. We now turn to the question of calibration and investigate how well a model's forecasts agree with the outcomes. Calibration refers to the statistical consistency between forecasts and observations. When a forecast predicts an expected earthquake count of 0.88, then the number of observed earthquakes following such a forecast should average 0.88.

The extant CSEP methodology puts the focus on formal statistical tests of agreement between forecasts and outcomes. However, under the (questionable) joint hypothesis that the modeled expected counts are correct and that each observed count comes from a Poisson distribution, a perfectly correct model might fail the test when the observed counts do not come from the Poisson distribution.

> We also need diagnostic tools with the goal of identifying – and eventually remedying – the deficiencies of any given model

By contrast, we propose displays that plot the conditional expected count against the corresponding forecast. These displays are called reliability diagrams and identify miscalibration via deviations from the diagonal. In the 2021 Annual Report, we outlined our development of the CORP approach to achieving stable reliability diagrams, which avoids the need for tuning parameters while providing guarantees in the form of asymptotic consistency and optimal finite-sample binning. The approach has since been extended and can now be applied in the present earthquake prediction setting.

Figure 15 displays mean-reliability curves for the five earthquake forecasting models. In practice, mean-reliability curves deviate from the diagonal even when the hypothesis of mean-calibration is true due to sampling variability. We use consistency bands to address the question of whether an observed deviation from the diagonal can be attributed to random fluctuations. Again, note the non-linear axes with tick marks at 0, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, and 0.88 – the largest prediction made by one of the five models. The LM model is well calibrated, as its mean-reliability curve largely remains within the consistency band. By contrast, the calibration curves of the FCM, LG, and SMA models indicate that the forecast values just above 10^{-5} tend to be too large because the curve lies considerably below the consistency band.

The text in this report has been adapted and contains excerpts from [Brehmer et al., 2025].

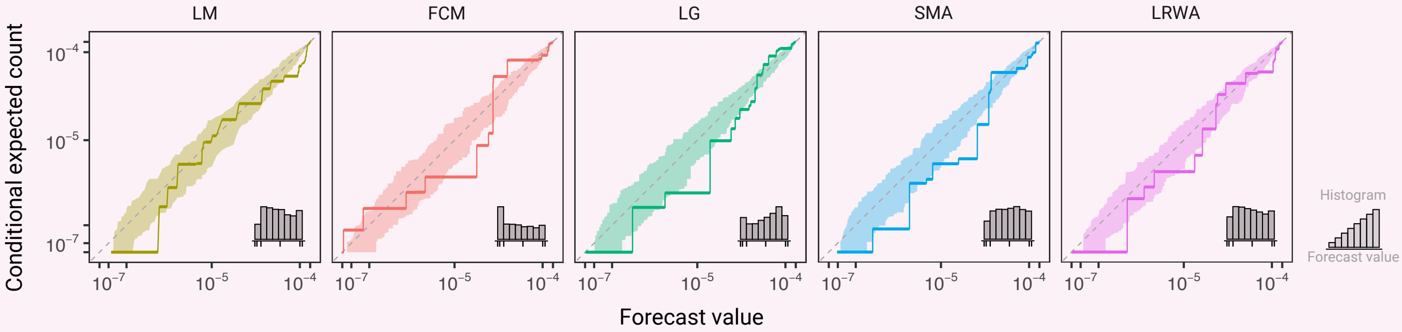
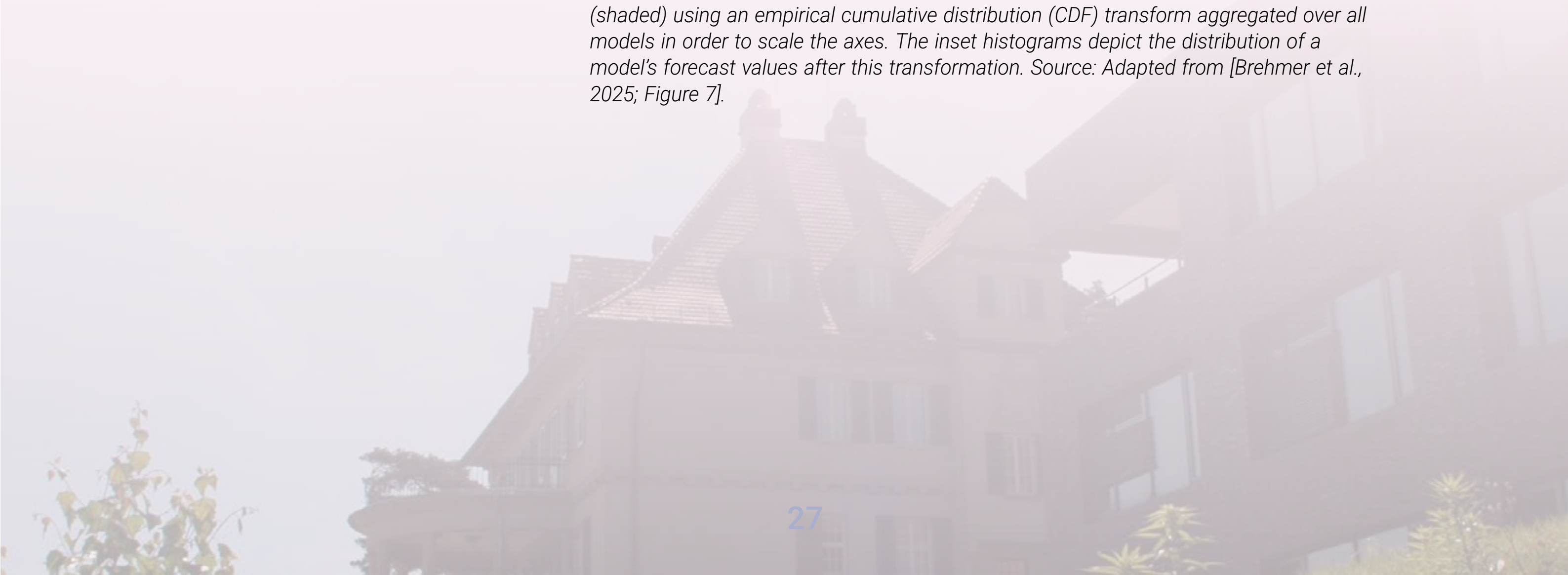


Figure 15

Mean-reliability curves for the five forecast models along with 90% consistency bands (shaded) using an empirical cumulative distribution (CDF) transform aggregated over all models in order to scale the axes. The inset histograms depict the distribution of a model's forecast values after this transformation. Source: Adapted from [Brehmer et al., 2025; Figure 7].



Research

2.4 DMQ

Data Mining and Uncertainty Quantification

Group Leader

Prof. Dr. Vincent Heuveline

Team

Aksel Alpay
(visiting scientist; Heidelberg University)

Riccardo Beccaria
(PhD Student; HITS Lab; Heidelberg University)

Marcus Buchwald
(visiting scientist; Heidelberg University)

Yann Ebling
(student assistant; Heidelberg University; since April 2025)

Ayse Erozan
(visiting scientist; Heidelberg University; until September 2025)

Isabel Gernand
(PhD student)

Maximilian Ludwig
(visiting scientist; Heidelberg University; since May 2025)

Stefan Machmeier
(visiting scientist; Heidelberg University)

Pascal Memmesheimer
(visiting scientist; Heidelberg University)

Jakob Niessner
(visiting scientist; Heidelberg University)

Rimpa Pal
(PhD student; HITS scholarship)

Jonas Roller
(visiting scientist; Heidelberg University)

Valentin Schmid
(PhD student)

Alexander Zeilmann
(visiting scientist; Heidelberg University)

The Data Mining and Uncertainty Quantification group – led by Vincent Heuveline – has been active since May 2013 and operates in close collaboration with the Engineering Mathematics and Computing Lab (EMCL) at Heidelberg University's Interdisciplinary Center for Scientific Computing. The group's research is positioned at the interface of computational mathematics, medical engineering, and information security, with a strong emphasis on uncertainty quantification in data-driven and simulation-based workflows.
A central objective is the extraction of reliable, decision-relevant information from large and heterogeneous datasets, especially in the context of biomedical and medical engineering applications. This includes data generated by computational models and measurement processes, such as in medical imaging and in digital-twin concepts for patient-specific analysis. In this context, uncertainty quantification plays a pivotal role, enabling systematic assessment, propagation, and the reduction of epistemic and aleatory uncertainties that are inherent in both data acquisition and mathematical modeling.

Methodologically, the group combines advanced numerical analysis, stochastic modeling, and machine-learning techniques with high-performance and hardware-aware computing strategies. This integration ensures the robustness, scalability, and reproducibility of computational pipelines, which is essential for safety-critical medical applications and for compliance with stringent IT security and data protection requirements.

Die Data Mining and Uncertainty Quantification Gruppe unter der Leitung von Vincent Heuveline, wurde 2013 am HITS etabliert und arbeitet eng mit dem Engineering Mathematics and Computing Lab (EMCL) der Universität Heidelberg am Interdisziplinären Zentrum für Wissenschaftliches Rechnen zusammen. Die Forschung der Gruppe ist an der Schnittstelle von Computational Mathematics, Medizintechnik und IT-Sicherheit angesiedelt, mit Schwerpunkt auf der Unsicherheitsquantifizierung in datengetriebenen und simulationsbasierten Prozessen. Ein zentrales Ziel der DMQ-Gruppe ist es, verlässliche, entscheidungsrelevante Informationen aus großen und heterogenen Datensätzen zu gewinnen, insbesondere in der biomedizinischen Forschung und in medizintechnischen Anwendungen. Dazu zählen Daten aus numerischen Modellen und Messprozessen, etwa in der medizinischen Bildgebung oder bei patientenspezifischen Digital-Twin-Ansätzen. Dabei kommt der Unsicherheitsquantifizierung eine Schlüsselrolle zu, da sie eine systematische Bewertung, Fortpflanzung und Reduktion sowohl epistemischer als auch aleatorischer Unsicherheiten in Daten und mathematischen Modellen erlaubt.

Methodisch verbindet die Gruppe fortgeschrittene numerische Analysis, stochastische Modellierung und maschinelles Lernen mit leistungsfähigen, hardwarebewussten Hochleistungsrechenkonzepten. Diese Integration gewährleistet robuste, skalierbare und reproduzierbare Rechenpipelines, die für sicherheitskritische medizintechnische Anwendungen sowie für die Einhaltung strenger Anforderungen an IT-Sicherheit und Datenschutz unerlässlich sind.

Mathematical Oncology: Multimodal Digital Twin Modeling for the Diagnosis and Therapy of Prostate Cancer

A digital twin provides a powerful framework for integrating heterogeneous clinical data into patient-specific models

Prostate cancer is the most common form of cancer diagnosed in men. Despite significant advances in imaging, diagnostics, and treatment, clinical decision making remains challenged by both the heterogeneity of disease progression and treatment response across patients.

A digital twin is a computational representation of a physical object that dynamically assimilates object-specific data in order to guide decision making. In the context of prostate cancer, a digital twin provides a powerful framework for integrating heterogeneous clinical data into patient-specific models capable of supporting a diagnosis, a prognosis, and treatment planning (see Figure 17).

Patients diagnosed with localized prostate cancer are commonly stratified into two principal cohorts: active surveillance and surgical intervention. Active surveillance is typically recommended for patients with a low-risk disease and involves regular monitoring through prostate-specific antigen measurements, imaging, and repeat biopsies in order to track disease progression while avoiding or delaying invasive treatment. In contrast, patients with higher-risk or progressive disease are more likely to undergo definitive treatment, most commonly a radical prostatectomy with the goal of tumor removal and long-term disease control.

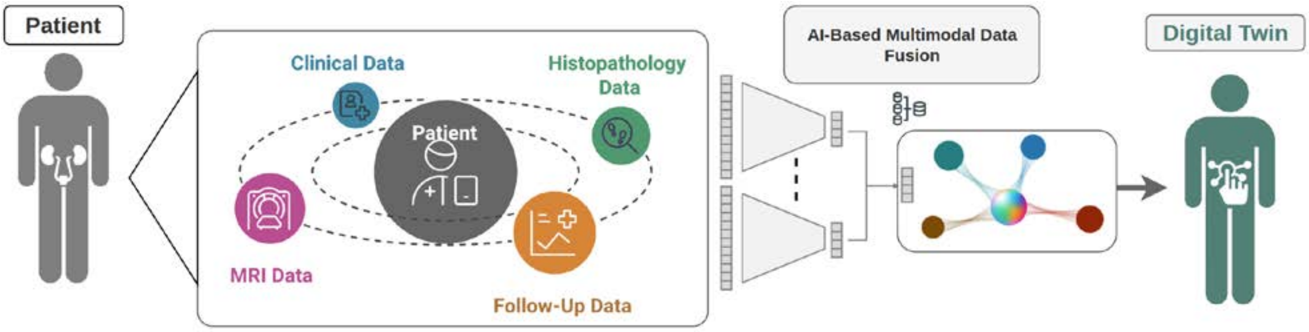


Figure 17
Schematic of data processing, modeling, and prediction workflow.

Robust Prediction of Prostate Cancer Recurrence

Prostate cancer that returns after a radical prostatectomy remains a pressing issue for urologists. Early risk detection can guide more intensive surveillance and timely adjuvant therapy, thereby ultimately improving patient outcomes. Modern diagnostic pathways generate a diversity of heterogeneous information ranging from high-resolution magnetic-resonance images to detailed pathology reports and a suite of clinical measurements. The challenge – and the opportunity – lies in weaving this information together into a single, coherent digital twin of the patient that can reliably forecast a relapse.

In joint collaboration with the DMQ group, the Urology Department at the German Cancer Research Center (DKFZ) assembled three complementary modalities for every patient: structured clinical reports, histology reports that are enriched with digitized slide images, and T2-weighted MRI volumes that indicate the prostate's zonal architecture and capsular integrity. In order to integrate these sources, we first transformed the textual material into dense, context-aware embeddings using a BERT-style model that had been fine-tuned on biomedical literature. Simultaneously, the MRI data were normalized and distilled using a vision model. The core of our fusion strategy is a multi-view autoencoder, which is a deep neural network that learns a shared latent space across the three modalities. In this compressed representation, each patient is described by a multimodal fingerprint that captures cross-modal relationships.

Several classical machine-learning classifiers were trained to predict recurrence. Their performance is limited by cohort size. In parallel, we evaluated MedGemma, which is a large multimodal medical language model that ingests a single concatenated text comprising the clinical summary, the pathology narrative, and a textual description of the MRI findings (see Figure 18, next page). MedGemma generates a recurrence risk estimate directly, effectively performing clinical reasoning in the same way a seasoned urologist might. Remarkably, MedGemma's predictions matched or exceeded those of the classical classifiers, thereby highlighting the power of foundation models in data-limited settings.

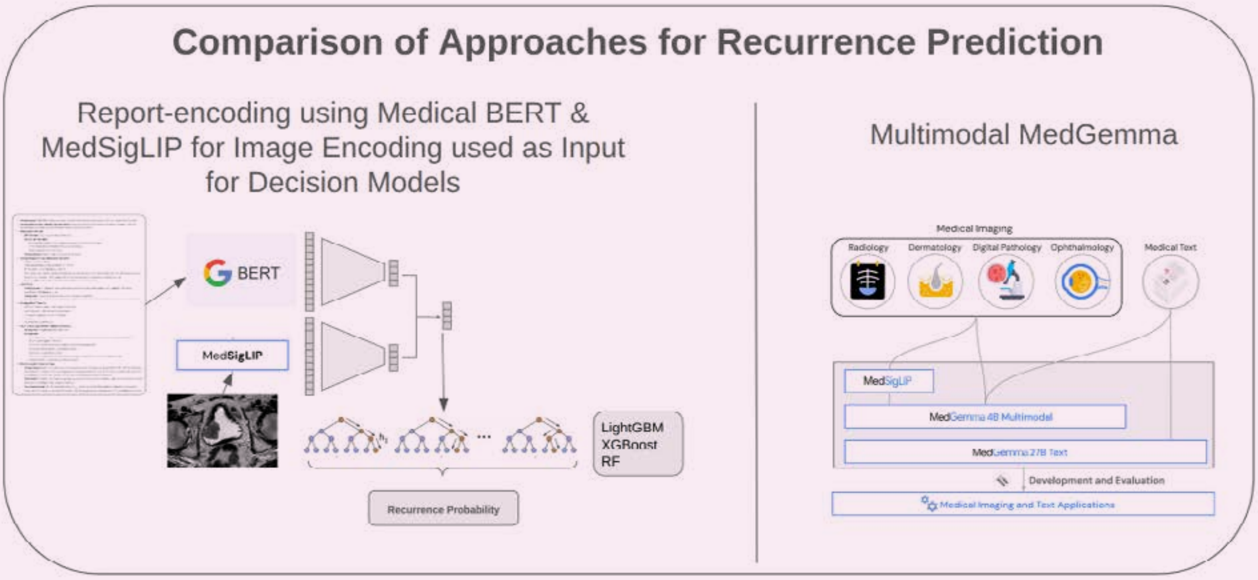


Figure 18
Learning a suitable representation of multimodal data and training a decision model (left) compared with using Foundation Model (MedGemma) with and without fine-tuning (right).

Modeling Organ and Tumor Growth in the Prostate

The second focus of the project addressed patients in the active surveillance cohort, for whom the prostate is often enlarged, either as a result of benign prostatic hyperplasia or due to the presence of low-grade tumors. Clinical management in these patients relies on repeated monitoring over time rather than on immediate treatment, thereby making it crucial to distinguish between normal or benign prostate growth on the one hand and changes that may indicate disease progression on the other hand.

In order to support this goal, longitudinal MRI and clinical reports were analyzed jointly. Automated image-analysis methods were used to segment the prostate from surrounding pelvic structures and to track changes in prostate volume across multiple imaging time points. Suspicious lesions were identified separately in order to complement this organ-level analysis.

Instead of modeling tumor growth directly from the outset, the project first concentrated on modeling the growth of the prostate organ itself (see Figure 19). Such baseline growth can obscure the interpretation of imaging findings and diagnostic trends if it is not explicitly taken into account. In order to capture these dynamics, a mathematical growth model based on partial differential equations was used to describe the gradual expansion of prostate tissue and to quantify patient-specific growth patterns using longitudinal imaging data. This organ-level model provides an essential foundation for future work by accounting for benign and age-related changes.

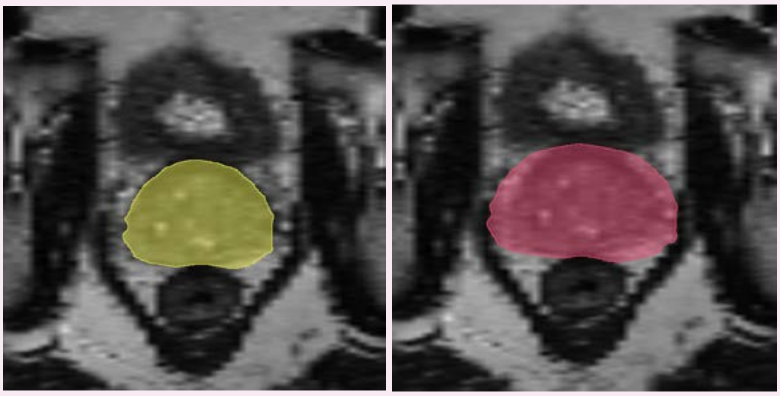


Figure 19
MRI with yellow prostate segmentation (left) and simulated enlargement (red, right).

Reliable MELD: Uncertainty-Aware Forecasting for Transplant Support

In the high-stakes area of organ transplantation, the Model for End-Stage Liver Disease (MELD) score serves as the gold standard for clinical triage. However, its traditional application relies on static, point-in-time assessments that often overlook the inherent volatility and rapid physiological shifts that are characteristic of end-stage hepatic failure. While accurately predicting a patient's trajectory is critical for effective triage, clinical time-series data present a challenge. The complex interplay of therapeutic interventions, comorbid fluctuations, and irregular sampling intervals creates highly non-stationary environments that violate the fundamental “independent and identically distributed” (i.i.d.) assumptions of standard machine learning models and classical uncertainty quantification methods.

In order to address these limitations, we introduced Reliable MELD, a forecasting framework designed to shift the clinical paradigm from point estimation to robust uncertainty quantification. At the core of our framework is the integration of Conformal Prediction, which is specifically adapted to handle the non-exchangeable nature of longitudinal medical records. Unlike traditional "black-box" algorithms that output a single, potentially misleading value, Reliable MELD generates rigorous prediction intervals backed by mathematical coverage guarantees (see Figure 20).

These intervals provide a dual-layered insight: The predicted score indicates the expected severity, while the width of the interval serves as a proxy for clinical stability. A narrow interval signals a predictable, stable trajectory, whereas a widening interval alerts clinicians to high-variance states in which a patient’s condition may be entering a period of acute instability. This integration of statistical rigor and interpretability provides physicians with the necessary transparency to be able to identify the most vulnerable patients earlier, thereby essentially enabling a “glimpse into the future” and ultimately helping to mitigate waitlist mortality through more precise and informed organ allocation.

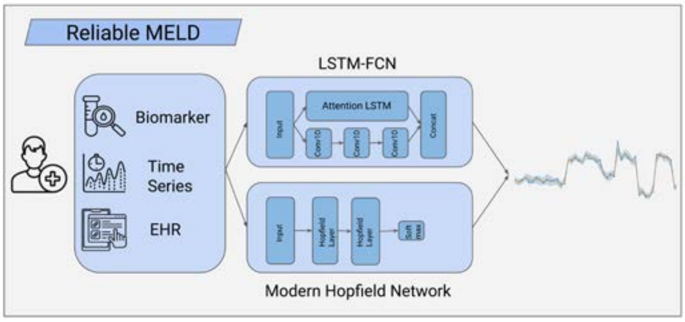


Figure 20
Representation of the Reliable MELD architecture, which integrates clinical data through a hybrid LSTM-FCN and modern Hopfield network in order to provide robust, uncertainty-aware forecasts.

> We introduced a reliable forecasting framework designed to shift the clinical paradigm from point estimation to robust uncertainty quantification

KI-Morph: Advancing Large-Scale 3D Image Analysis

Modern life sciences research generates unprecedented volumes of three-dimensional imaging data. High-throughput microscopes can produce terabytes of image data in a single day, which is far more than any researcher could manually analyze. These images contain valuable information about biological structures, but extracting this information requires identifying and outlining specific features in a process called segmentation. Traditionally, researchers once had to trace these features by hand, which could take weeks or months for a single dataset. KI-Morph provides a web-based platform that combines artificial intelligence with powerful computing resources in order to enable the rapid, accurate segmentation and analysis of massive 3D biomedical datasets. Researchers can access state-of-the-art AI tools through a simple web browser and can thus process datasets that would previously have required specialized hardware and possibly months of manual work.

A core motivation for developing KI-Morph is to open researchers’ access to state-of-the-art analysis tools. Currently, the ability of a research laboratory to use AI depends heavily on financial resources since the laboratory must be able to afford costly hardware, ongoing electricity costs, cooling and maintenance expenses, and the specialized staff needed to manage the system. This combination of financial and technical barriers creates a digital divide in which resource-poor laboratories – particularly those at smaller institutions with limited grant funding – are effectively excluded from modern quantitative image analysis. As the field moves toward data-driven approaches, researchers without access to AI infrastructure cannot participate in cutting-edge research or validate their findings using state-of-the-art methods.

> KI-Morph provides a web-based interface that completely removes the need for researchers to own or maintain expensive computing equipment

KI-Morph bridges this gap by providing a web-based interface that completely removes the need for researchers to own or maintain expensive computing equipment. Researchers can access powerful computing clusters through a standard web browser, thereby enabling advanced analysis on ordinary laptops, tablets, or even devices with limited capabilities (see Figure 21). The platform handles all the technical complexity behind the scenes. The AI systems, data management, and software updates all happen automatically on centralized servers. This means that a researcher working from a basic laptop in a resource-constrained environment has the same access to cutting-edge AI tools as a researcher at a well-funded institution with dedicated computing resources.

KI-Morph addresses several other critical considerations. The platform adopts a "responsibility-by-design" approach in which ethical safeguards are built directly into the system. Even though KI-Morph's AI was trained on over 120 diverse biomedical datasets, hidden biases cannot be completely ruled out. The platform addresses this by requiring human researchers to review and validate all AI-generated results, thereby creating a system of checks and balances. The development of KI-Morph demonstrates that building effective research services requires careful attention to technical excellence, legal compliance, and ethical responsibility. This ensures that scientific progress is driven by ideas and evidence rather than by infrastructure budgets.

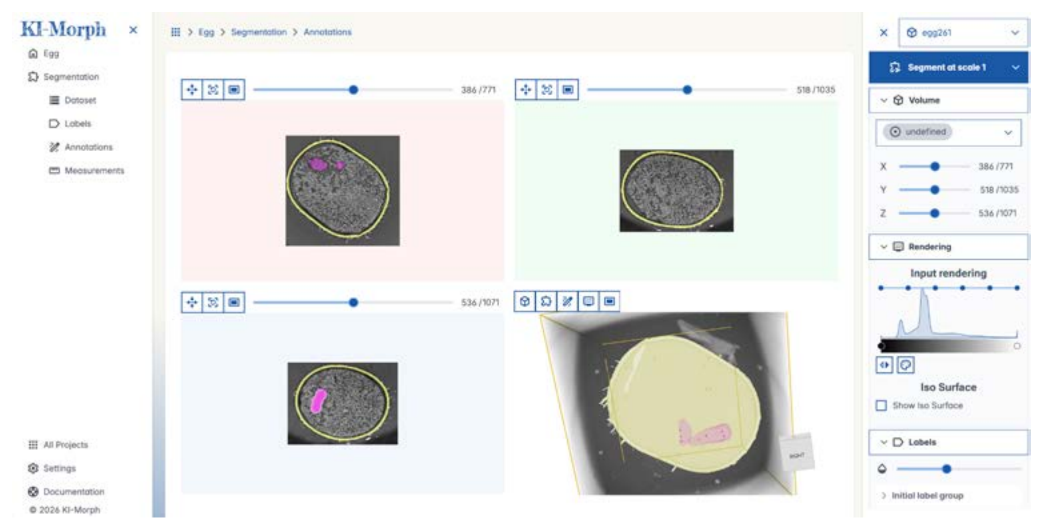


Figure 21
The annotation user interface of KI-Morph, which allows the user to visualize, annotate, and segment large-scale image volumes.

Detecting Suspicious Behavior in Computer Networks Using Machine Learning

Machines on the Internet communicate with specified protocols in order to establish connections or exchange data. One of the most essential protocols for the Internet is the Domain Name System (DNS) protocol, which in essence resolves domain names to IP addresses in order to remove the burden of memorizing numbers for users. However, not all actors on the Internet have good intentions. In fact, the threat landscape for computer networks is growing: Indeed, the 2025 Fortinet Threat Landscape reports a rise in automated scanning and exploitation tools. This change in attacks leaves security operators with little reaction time and shifts the need for automated methods that can counteract these endeavors. Such detection methods are applicable on network interfaces in order to inspect the contents of network communication to find suspicious behavior in an approach known as intrusion detection.

Misusing Protocols in Order to Steal Data

Protocols have intended applications. Taking DNS as an example, attackers misuse protocols in order to establish back doors to restricted network areas and to avoid detection by smuggling data through DNS requests. In essence, attackers encode data in subdomains, such as mail addresses or credit card information. By resolving these domains to an authoritative name server, attackers transfer sensitive data. Figure 22 illustrates an example of such an attack.

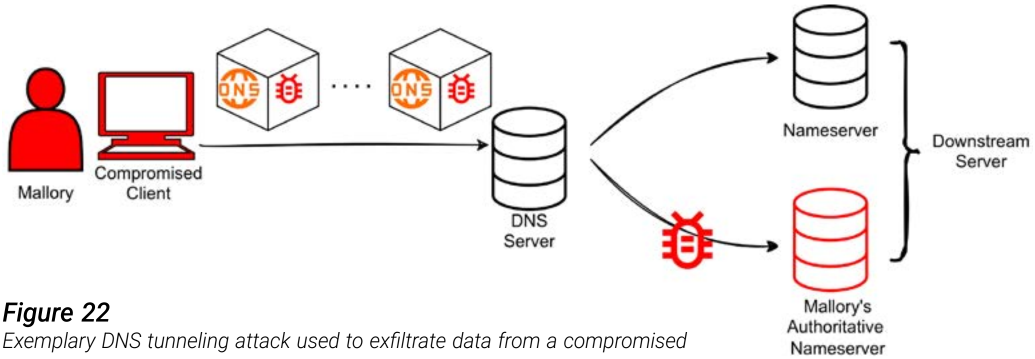


Figure 22
Exemplary DNS tunneling attack used to exfiltrate data from a compromised client to an attacker-controlled nameserver.

Mallory – the attacker – has successfully compromised a client machine within a private network and wants to transfer classified data without the intrusion detection raising alarms. To transfer the data "abc" secretly, he encodes it to the subdomain abc.mallory.com. DNS is a hierarchical protocol, meaning that it asks other DNS servers to fully resolve a domain. It begins with the top-level domain ".com", goes to the second-level domain ".mallory", and tries to resolve "abc". Mallory's quirk is that he holds a DNS server that resolves *.mallory.com requests, which is also known as an authoritative name server. This allows Mallory to act as a legitimate DNS resolver who resolves incoming requests. From a security operator perspective, Mallory's server appears like a real DNS resolver and can be ignored. In addition, this attack is also utilized to establish active communications, which are also named DNS tunnel, or DNS covert channel.

Analyzing DNS Communication

In order to detect compromised clients and avoid continuous damage in private networks, it is important to inspect DNS requests and responses over time. Therefore, we formulate a tunneling activity as a time-series-based analysis in order to disclose similarities of DNS behavior with known malicious tunneling tools. For instance, normal user DNS behavior differs from active tunneling tools in the Shannon entropy value and packet size of requests over time. Tunneling tools send requests in a higher frequency in order to transfer data, they send "keep alive messages" periodically, and they exceed average packet sizes, whereas normal users crawl websites, take breaks, and continue again. The behavior of the tunneling tools is atypical, and packet sizes are constant. In contrast, benign domain names have lower entropy values compared with base64 encoded data, which attackers try to maximize in order to lower the total number of requests. In order to attribute the similarity of a time series, we use the distance metric dynamic time warping (DTW) to quantify the unknown samples to the distribution of malicious requests over time. DTW searches for the temporal alignment that minimizes the Euclidean distance between two time series. Our detector analyzes user behavior over a 15-second time window and clusters the time series using k-nearest neighbor (kNN) with DTW as a distance metric (see Figure 23). Our results indicate that only a small subset of DNS requests is required in order to accurately predict unknown samples. Remarkably, our classifier required only a subset of our training data and achieved low false detection without fine-tuning. We see the application field of our lightweight detector at DNS servers or main egress points of networks to find suspicious behavior in advance.

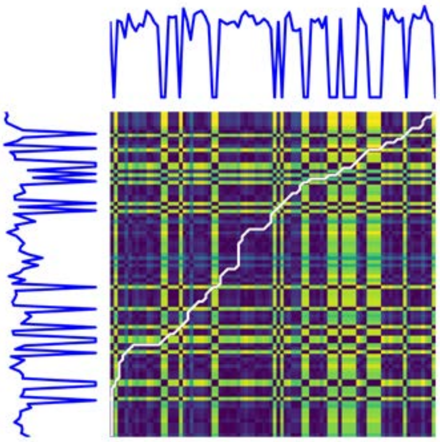


Figure 23
Dynamic time-warping path of two DNS requests, reformulated as time-series data.

Research

2.5 MLI

Machine Learning and Artificial Intelligence

Group Leader

Jun.-Prof. Dr. Jan Stühmer

Team

Torben Berndt
(PhD student)

Mila Coetzee
(PhD student)

Dr. Beni Egressy
(postdoc; until December 2025)

Leif Seute
(PhD student)

Nicolas Wolf
(visiting scientist)

Elyes Farjallah
(student assistant; since October 2025)

Moritz Feik
(master's student; since September 2025)

Elias Huber
(master's student; since October 2025)

Eddie Merbitz
(student intern; until February 2025)

Chaitanya Sampara
(master's student; since September 2025)

David Sichma
(master's student; since September 2025)

Hendrik Stumpf
(student assistant; until March 2025)

Dominik Sucker
(master's student; since December 2025)

The Machine Learning and Artificial Intelligence (MLI) group works on two main research topics: geometric deep learning and structured latent variable models in deep learning.

Geometric deep learning is an emerging paradigm that extends the field of deep learning to complex data structures such as graphs, molecules, and physical systems. In geometric deep learning, models are categorized by the symmetries they preserve within data — a principle that leads to more efficient and effective learning models. In its research, the MLI group utilizes the framework of geometric deep learning to design more efficient architectures for applications in science and engineering. An example of the group's research contributions can be found in protein design: The combination of concepts from geometric deep learning and so-called Clifford algebras resulted in a more well-balanced representation of the secondary structure of a protein: namely beta sheets and alpha helices.

Another focus of the MLI group is on structured latent variable models, with the aim of making deep learning models more understandable and trustworthy. These models combine traditional statistical techniques with the flexibility of deep learning, thereby allowing researchers to uncover meaningful structures in complex data. A key innovation in this area is the concept of identifiability — that is, ensuring that different runs of the model produce consistent and interpretable internal representations. This consistency allows the extracted features to carry semantic meaning, thereby making the model more interpretable by humans. At the same time, identifiability can bridge the gap between the "sub-symbolic" features used in neural networks and "symbolic" reasoning systems, leading to neuro-symbolic architectures.

Die MLI Gruppe beschäftigt sich mit der Entwicklung von neuartigen Algorithmen und Verfahren des maschinellen Lernens. Besondere Schwerpunkte liegen hierbei im Geometric Deep Learning und in den interpretierbaren Repräsentationen.

Geometric Deep Learning bezeichnet einen Ansatz, mit dem Deep Learning auf komplexe Datenstrukturen wie zum Beispiel mathematische Graphen, Moleküle oder physikalische Systeme angewendet werden kann. Im Geometric Deep Learning werden Modelle anhand der Symmetrien, welche in den Daten erhalten bleiben, kategorisiert. Durch diese prinzipielle Betrachtung wird es möglich, effizientere Modellarchitekturen für unterschiedlichste Anwendungen zu entwickeln. In der Forschung der MLI Gruppe wird dieser Ansatz verwendet, um effiziente Architekturen für Anwendungen des maschinellen Lernens in den Naturwissenschaften zu entwickeln. Ein beispielhaftes Anwendungsfeld der Forschung der MLI Gruppe liegt im Proteindesign: Durch die Kombination von Konzepten des Geometric Deep Learning mit sogenannten Cliffordalgebren gelang es, eine ausgewogenere Verteilung von Sekundärstrukturen von Proteinen zu erhalten, den Beta-Sheets und Alpha-Helices.

Ein weiterer Fokus der MLI Gruppe liegt in der Entwicklung von interpretierbaren strukturierten Modellen, mit dem Ziel, Deep Learning Verfahren verständlicher und vertrauenswürdiger zu gestalten. Diese Modelle kombinieren traditionelle Methoden der Statistik mit der Flexibilität des Deep Learning und erlauben es, Einsichten und Zusammenhänge aus komplexen Daten zu gewinnen. Ein zentraler Bestandteil in diesem Zusammenhang ist das Konzept der Identifizierbarkeit, welches besagt, dass man auf denselben Eingabedaten eine konsistente interne Repräsentation des Modells erhält. Diese Konsistenz erlaubt es, den internen Repräsentationen eine semantische Bedeutung zuzuweisen, und macht somit das Modell interpretierbar. Gleichzeitig ermöglicht Identifizierbarkeit, eine Brücke zwischen den „sub-symbolischen“ Repräsentationen eines neuronalen Netzes und den „symbolischen“ Methoden des logischen Schließens schlagen – sogenannte neurosymbolische Ansätze.

Predicting and Designing Protein Motion With Flow Matching

Proteins – the central molecular machinery of life – are not static objects; rather, they are in constant motion. Their dynamics are often a prerequisite for function. Enzymes, for example, rely on flexibility in order to bind substrates and release products, and many signaling proteins operate by switching between different conformations.

Recent advances in geometric deep learning have made it possible to design protein structures with remarkable accuracy. However, current generative models focus on producing a single, stable structure and largely ignore dynamics. In 2025, we introduced two new deep learning models – BBFlow and FliPS – that extend flow-based generative modeling by explicitly addressing protein flexibility and motion.

> Current generative models focus on producing a single, stable structure and largely ignore dynamics

Predicting Protein Dynamics From Backbone Geometry

A natural starting point for designing protein dynamics is the ability to predict them. BBFlow (backbone flow) addresses this task by learning to generate conformational ensembles of proteins based on their backbone geometry [Wolf et al., 2025].

Traditionally, protein dynamics are studied using molecular dynamics (MD) simulations, which approximate the Boltzmann distribution of states by numerically integrating physical equations of motion. MD simulations are computationally expensive and hence impractical for large-scale screening or iterative design. BBFlow is trained to sample states from the same distribution as MD, with a speedup of several orders of magnitude.

Technically, BBFlow is a flow-matching model defined based on the space of protein backbones. Given an equilibrium backbone structure, the model learns a flow that transforms a simple, prior distribution into an ensemble of realistic backbone states as observed in MD. Crucially, both the flow and the prior are conditioned not only on the input sequence, but also on the input structure. Hence, the model relies on backbone geometry rather than on evolutionary sequence information or on pre-trained folding models, which have been used in previous approaches. This makes BBFlow particularly well suited for de novo proteins, in which evolutionary information is sparse or entirely absent. This capability plays a central role in the design framework introduced below.

From Predicting Dynamics to Designing Dynamics

While BBFlow predicts how a given protein moves, the inverse problem is even more challenging: How can a protein structure be designed such that it exhibits a desired pattern of motion? FliPS (Flexibility-conditioned Protein Structure generation) addresses this inverse problem by turning dynamics into an explicit design objective [Viliuga et al, 2025].

FliPS is a flow-based structure-generation model that learns the conditional distribution of protein backbone structures given target per-residue flexibilities. As input into the model, the user can specify how flexible or rigid each region of the designed protein should be. The model then generates backbone structures whose dynamics – as measured by molecular dynamics – match this target profile.

Methodologically, FliPS builds directly on Geometric Algebra Flow Matching (GAFL) – a flow-based generative model for protein backbone design developed in the MLI group. As in GAFL, protein backbones are represented as sets of rigid bodies, each of which represents a residue. The model learns a flow on the space of rigid body collections that transforms noise into structured protein backbones. FliPS extends this framework by conditioning the learned flow on per-residue flexibility, thereby enabling the model to learn systematic relationships between backbone geometry and dynamic behavior.

To train such a conditional model, a dataset of tens of thousands of proteins with annotated flexibility is needed. Molecular dynamics is prohibitively slow at predicting protein dynamics on a large scale; hence, we used BBFlow to annotate the dataset with flexibility.

> Taken together, BBFlow and FliPS form a coherent framework for protein dynamics modeling

A Framework for Designing Dynamical Proteins

Within a protein design pipeline, BBFlow can be used to rank generated structures according to how closely their predicted dynamics match the target. This enables us to generate large numbers of structures with FliPS and to filter the most promising candidates with BBFlow before continuing to the next stage of the design campaign.

Taken together, BBFlow and FliPS form a coherent framework for protein dynamics modeling. BBFlow enables the fast and accurate prediction of backbone motion and provides large-scale dynamic annotations. FliPS inverts this knowledge in order to enable the design of new protein structures with controlled flexibility. This shift from static structures to dynamic ensembles opens a new direction in protein design, thereby bringing the field a step closer to designing proteins in which specific motion is required, such as enzymes or conformational switches in sensors or regulators.

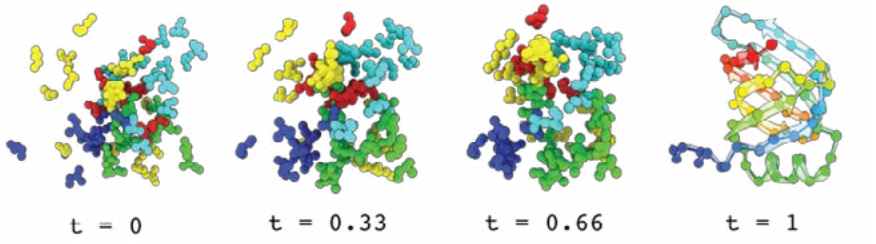


Figure 24
Flow-matching models for protein structure generation such as BBFlow and FliPS learn to transform noise into well-structured backbones.

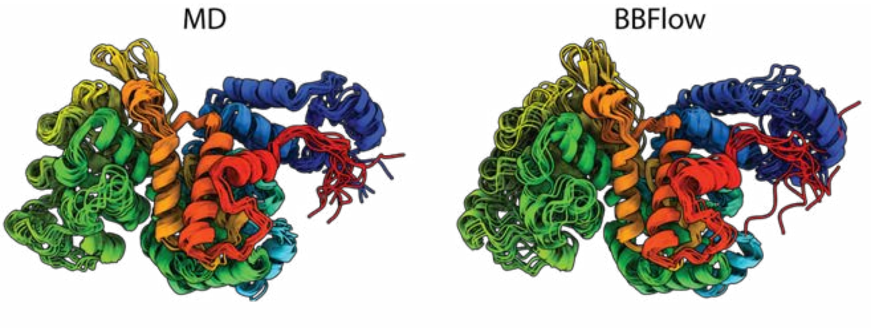


Figure 25
Comparison between a protein backbone ensemble generated by BBFlow and one obtained from MD simulation.

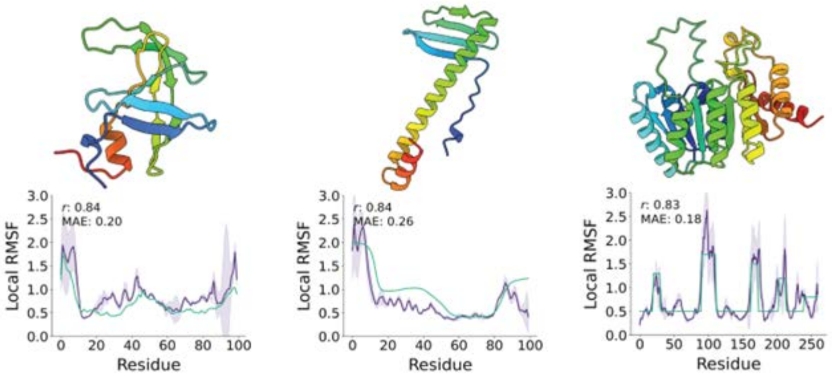


Figure 6
Proteins generated by FliPS, with target flexibility profile (green) and flexibility verified in MD simulation (purple), along with Pearson correlation and MAE between the profiles.

Set-LLM: A Permutation-Invariant Large Language Model

When it comes to the topic of artificial intelligence, most people first think of large language models (LLMs). Despite their impressive performance in a large variety of tasks, LLMs still show sensitivity to astonishingly simple so-called adversarial attacks. One such attack is based on simply re-ordering the choices in a multiple-choice test, which is surprisingly effective at misleading the LLM to give an incorrect answer.

Order Bias in Large Language Models

This weakness – also called order bias – affects an LLM's performance not only in multiple-choice tests, but also in everyday tasks such as the summarization of multiple documents or when answering questions about the content of multiple documents. In order to process the information in multiple documents with an LLM, the documents are concatenated one after another and – together with the question – are given as input to the LLM. The probability of getting the correct answer now largely depends on where in this input stream the information is contained. If the information for answering the question is in the first or last document, the LLM is much more likely to give the correct answer than if it is in a document somewhere in the middle. This is especially true for LLMs with a smaller number of parameters. While these models provide a more efficient runtime performance than are larger models, they display a stronger drop in performance when the information we are looking for is somewhere in the middle of the input stream.

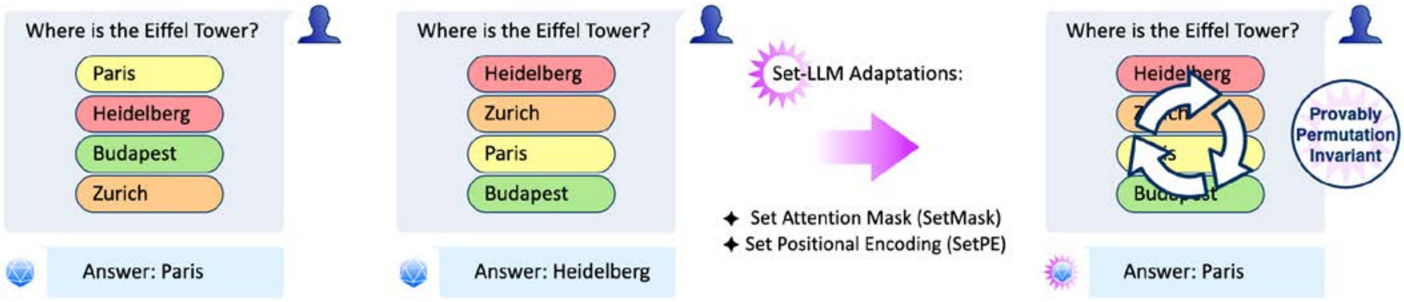


Figure 27
Existing large language model architectures contain a bias toward the order of the input. As an example, we consider a multiple-choice test in which the LLM's response changes by simply reordering the answers (example for illustrative purposes only).

Removing the Order Bias: Modified Attention Matrix and Positional Encoding

To remove this order bias, we approached the problem from the perspective of geometric deep learning, in which invariances and equivariances of deep learning models to transformations of the input are studied. The order bias of large language models can only occur if the model is not invariant to changes to the order of the input, which are also called permutations. On the one hand, order of course matters when reading a sentence, paragraph, or whole text. In some cases, however, as in the above-mentioned multiple-choice test or when processing multiple documents of equal importance, the models' output should not depend on the order of some part of the input. In our recent work "Set-LLM," which we published at the Conference on Neural Information Processing Systems (NeurIPS) in December 2025, we proposed a modification of the transformer architecture that serves as the basis of most large language models. In order to describe our proposed modification, we have to dive a little deeper into the architecture of the transformer model. One of the main components of the transformer architecture is the attention mechanism, which for each layer of the model determines which part of the input is important for processing the information required for predicting the next token. In a standard transformer architecture, a so-called "causal mask" is often used for the attention mechanism, which determines that the current output token can only depend on previous parts of the input. This is particularly useful when training the transformer model on the standard task of next token prediction: In this case, all the information we have is in the preceding part of the text. Therefore, with causal mask attention, future information is masked out, and the current output only depends on the previous text. Another component of the transformer model that we need to modify is the positional encoding, which assigns a label that marks the position of each token within the input. For simplicity, we want to think of this positional encoding as a number that indexes the tokens of the input (see Figure 28, next pages).

Acing the Multiple-Choice Test

Causal attention masks and standard positional encoding work well when the input is a text in sequential order. In our example of a multiple-choice test, however, the input contains two different parts: the question as sequential text, and the answers to the question, for which the order should not matter. We treat the different answers as “set-valued” information, which means that instead of treating the answers as sequential input, we treat each answer as an element of a set (the set of answers) for which the order of the elements in the set should not have any influence on the output of the model. This requires the attention mechanism to be changed such that the current output token does not depend on the order of the answers. We achieve this by treating the tokens of each answer as independent paths (Figure 28) and by assigning the same positional encoding to the elements along each path. This way, the different possible answers form several parallel paths through the attention graph instead of a single consecutive sequence. The result is that the order of the answers has no influence on the output of the model as these items all appear of equal rank to the model. When predicting the next token, the information is processed along each of these paths individually and is then finally fused together to output the answer to the question. The necessary modifications to both the positional encoding and the attention mask are depicted in Figure 28.

With the proposed modifications, the output is independent of the order of the answers, and the model does not show the order bias of the standard transformer architecture. In our experiments, we could show that Set-LLM consistently outperforms the standard transformer architecture in multiple-choice tests when the order of the answers is randomized.

> We could show that Set-LLM consistently outperforms the standard transformer architecture in multiple-choice tests when the order of the answers is randomized

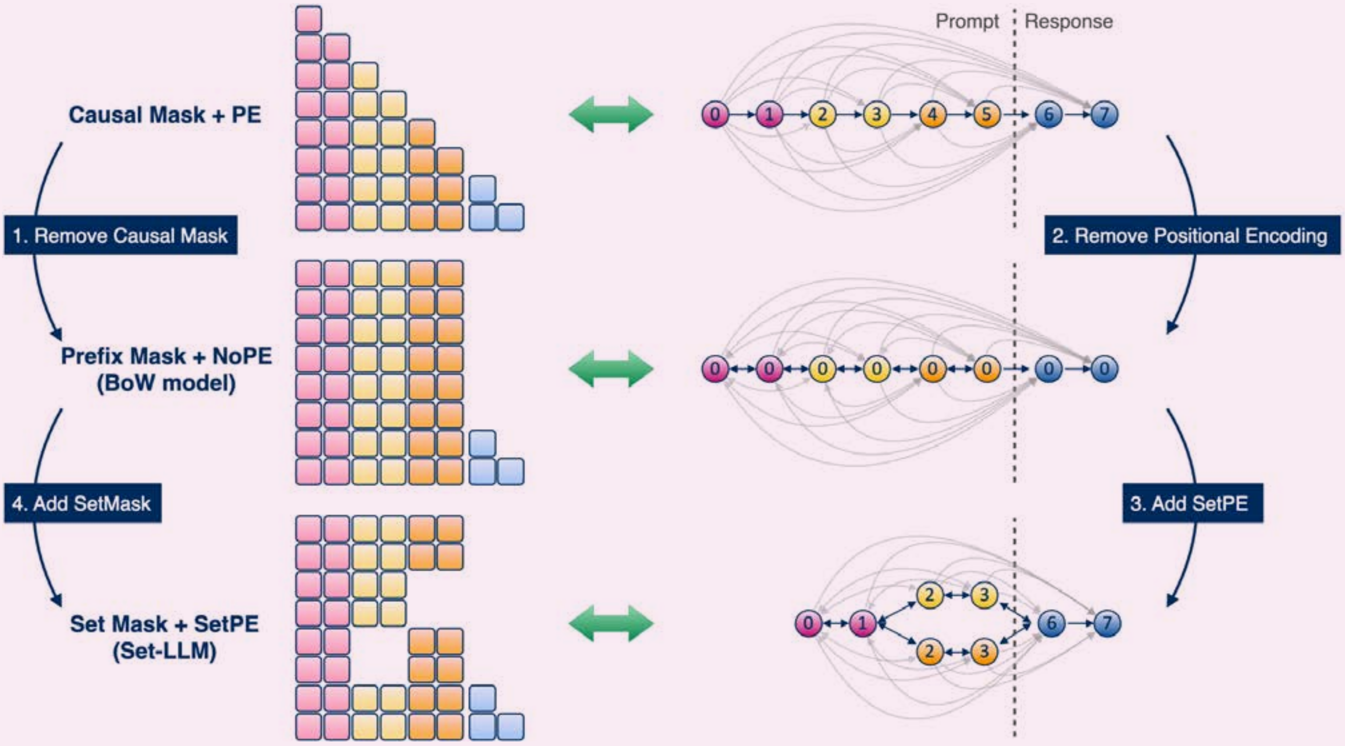


Figure 28
Proposed modifications for removing order bias in large language models. We begin with the standard architecture with causal attention mask and standard positional encoding (PE). First, we remove the causal mask and the positional encoding, which leaves us with a so-called bag-of-words model. While the order of the input does not have any effect, the syntactical structure of the sentence is also removed. The proposed modifications to the Attention Mask (Set Mask) and the set positional encoding (SetPE) are also invariant to permutations of the input but keep the syntactical structure of the sentences intact.

Research

2.6 MCM

Molecular and Cellular Modeling

Group Leader

Prof. Dr. Rebecca Wade

Team

Riccardo Beccaria
(PhD student; HITS Lab)

Tommaso Bartoloni
(PhD student)

Mislav Brajkovic
(PhD student)

Paul Brunner
(student intern; until January 2025)

Vinicius Cabral
(visiting PhD student; CAPES Scholarship; since September 2025)

Zeynep Edizcan
(student worker; since August 2025)

Rostislav Fedorov
(PhD student; until September 2025)

Nico Herb
(master's student; until July 2025)

Tine Limberg
(student intern; October–November 2025)

Levi Miederer
(PhD student; since November 2025)

Dr. Kushal Singh
(postdoc)

Feroz Karim
(PhD student; DAAD PhD Scholarship)

Dr. Stefan Richter
(staff scientist)

Raziye Sari
(student worker; until September 2025)

Izar Scharf
(student intern; January–April 2025)

Dr. Christine Schulz
(visiting scientist; Heidelberg University; since June 2025)

Ieva Sorokina-Ozola
(student intern; since October 2025)

Artem Stetoi
(student intern; September–December 2025)

Darius Szablowski
(Master's student; since March 2025)

Jonathan Teuffel
(PhD student)

Anastasia Warken
(bachelor's student; March–July 2025)

Qingyang Yang
(master's student; June–December 2025)

Molecular recognition, binding, and catalysis are fundamental processes for cell function. The ability to understand how biomacromolecules interact with their binding partners and participate in complex cellular networks is critical to the prediction of macromolecular function as well as to applications such as protein engineering and structure-based drug design.
In the MCM group, we are primarily interested in understanding how biomolecules interact. What determines the specificity and selectivity of a drug–receptor interaction? How can proteins assemble to form a complex? How is the assembly of a complex influenced by the crowded environment of a cell? What makes some binding processes quick and others slow? How do the motions of proteins affect their binding properties? One of our aims is to gain a mechanistic molecular-level understanding of drug interactions along the process that extends from drug delivery to drug–target binding and drug metabolism.
We take an interdisciplinary approach that entails collaboration with experimentalists and that makes concerted use of computational approaches based on physics and bio-/chem-informatics. The broad spectrum of techniques that are developed and employed ranges from interactive, web-based visualization tools to machine-learning methods and atomic-detail molecular simulations.

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 this report, we outline some of the results achieved in 2025. Following a general overview of what the group did last year, we highlight results in three research projects:

(i) Combining Quantum Mechanics and Molecular Dynamics in Order to Compute the Electron Transfer Reorganization Energy in Cytochrome P450-Redox Protein Complexes
(ii) Computing Drug–Protein Association Rate Constants by Combining Molecular and Brownian Dynamics Simulations
(iii) Simulating Neurotrophin Receptors in Order to Investigate Signaling Mechanisms

What Did We Do in 2025?

The second phase of SIMPLAIX (www.simplaix.org, see Chapter 7) began in October 2025 and is planned to run for three years with the support of the Klaus Tschira Foundation and the participating institutions: KIT, Heidelberg University, and (beginning in Phase 2) the Max Planck Institute for Polymer Physics, Mainz. SIMPLAIX is coordinated by Rebecca Wade, who also chaired the 3rd SIMPLAIX workshop (see Chapter 5.1.2), to which Jonathan Teuffel and Rostislav Fedorov from the MCM group contributed as members of the scientific organizing committee. Levi Miederer joined the MCM group in November as a SIMPLAIX doctoral student. Initially, Levi worked together with Jonathan Teuffel on completing our SIMPLAIX Phase 1 project on electron transfer in cytochrome P450 enzymes (see below and [Teuffel et al, 2025]).

In 2025, we obtained funding from the DFG to expand our collaboration with Michael Lanzer and colleagues (Parasitology Department, Center for Infectology, Heidelberg University) on understanding the structural and functional requirements for substrate-induced conformational transitions of the Plasmodium falciparum chloroquine resistance transporter PfCRT. Anastasia Warken conducted her bachelor's thesis research in molecular biotechnology on this topic, working together with Kushal Singh.

In a collaboration of the Heidelberg University Scientific Software Center (SSC) with Rebecca Wade, funding was obtained through the "Scientific Software" call of the Klaus Tschira Stiftung to establish the Bio-Structure Hub (https://ssciwr.github.io/BioStructureHub/) at the SSC. Christine Schulz set up the Bio-Structure Hub in April 2025 and has successfully established it as a valuable resource offering services to Heidelberg researchers in AI-based biomacromolecular structure prediction.

Mislav Brajkovic was awarded an Isabel Rojas Travel Award to support his stay at AstraZeneca, Cambridge, UK, where he worked for three months in spring 2025 on optimizing the MCM group's τRAMD method for computing drug–target residence times for applications in the pharmaceutical industry. Following promising initial results, we have pursued this collaboration further since Mislav's return to Heidelberg.

In further collaborative drug discovery projects, we carried out modeling and simulation studies in order to

- identify lead antibacterial compounds against Streptococcus pyogenes (computational docking and screening by Jonathan Teuffel with experiments in the groups of Tomas Fiedler (University of Rostock) and Herman Schindelin (University of Würzburg)) [Schütt et al, 2025],

- identify a synthetic peptide with activity against heart failure (modeling and simulation by Manuel Glaser, now continued by Tommaso Bartoloni, in order to design modified peptides with improved pharmacokinetic profiles in collaboration with Patrick Most and Julia Ritterhoff (Heidelberg University Hospital)) [Kehr et al, 2025],

- identify heparin analogues without the anticoagulation activity of heparin that have improved activity against SARS-CoV-2 (project led – and simulations conducted – by Giulia Paiardi with experiments in the groups of Marco Rusnati (University of Brescia, Italy) and Petr Chlandla (Bioquant, Heidelberg University)) (Milanesi, M., Urbinati, C., Zimmermann, L. et al. K5 polysaccharides inhibit SARS-CoV-2 infection by preventing spike-proteolytic priming. npj Viruses 4, 3 (2026). https://doi.org/10.1038/s44298-025-00163-4),

- design high-affinity nanobodies that target SARS-CoV-2 spike glycoprotein (simulation and machine learning pipeline developed by Matheus Ferraz with experimental validation by the group of Roberto Lins (Department of Virology, Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, Brazil)) [Ferraz et al, 2025], and

- design 2-aminobenzothiazole derivatives targeting trypanosomatid PTR1 via a multidisciplinary fragment hybridization approach in a study initiated by the EU-supported NMTrypI consortium and led by former MCM postdoc Joanna Panecka-Hofman (University of Warsaw, Poland) [Panecka-Hofman et al, 2025].

Rostislav Fedorov – who stayed on in the MCM group after the departure of the CCC group (see Annual Report 2024) – successfully defended his doctoral thesis in September. Moreover, former group members Abraham Muniz-Chicharro and Alexandros Tsengenes both defended their doctoral theses. In addition, Nico Herb completed his master's thesis research in molecular bioscience. Furthermore, Paul Brunner, Izar Scharf, and Tine Limberg (Molecular Biotechnology) completed internships in the group as part of their master's studies at Heidelberg University.

Rebecca Wade was a co-organizer (with Giovanni Bottegoni (Univ. Urbino), Ariane Nunes-Alves (TU Berlin), and Stefan Wolf (Univ. Freiburg)) of the CECAM flagship workshop on "Predicting and understanding drug–target binding kinetics via molecular simulations," held in Lausanne, Switzerland, in March 2025. As one outcome of the workshop, we wrote a paper on the current state of drug–target binding kinetics with a roadmap to their establishment in drug discovery campaigns (Wolf et al, chemrxiv: https://chemrxiv.org/doi/full/10.26434/chemrxiv.15000435/v1).

Rebecca Wade (as QCMS Scientific Chair) organized the first QSAR, Chemoinformatics, and Modeling Society (QCMS) online lecture, which was given by Harel Weinstein (Cornell University, New York, USA) in November 2025. This new lecture series is now being continued with several events per year organized by members of QCMS (see https://www.youtube.com/@QCMSOCIETY).

Riccardo Beccaria, Stefan Richter, and Rebecca Wade (with Rommie Amaro (UCSD) and Chia-en Chang (UC Riverside)) organized another edition of the Biological Diffusion and Brownian Dynamics Brainstorm (BDBDB) series – BDBDB6 – which took place online in December 2025 (https://bdbdb6.h-its.org/). This workshop provided a valuable opportunity for catching up with the latest developments in the field, presenting recent results, and discussing collaboration on methods and software (see Chapter 5.1.5).

Combining Quantum Mechanics and Molecular Dynamics in Order to Compute the Electron Transfer Reorganization Energy in Cytochrome P450-Redox Protein Complexes

Cytochrome P450 (CYP) enzymes – ubiquitous proteins with key roles in processes such as drug metabolism and steroidogenesis – require the supply of two electrons for catalysis. The transfer of electrons to CYPs from partner redox proteins often limits the rate of catalysis. Within the framework of a SIMPLAIX project in collaboration with Marcus Elstner (Karlsruhe Institute for Technology, KIT), we are employing molecular simulation and machine learning approaches in order to compute electron transfer (ET) rates to CYPs with the aim both of understanding the factors influencing ET rates and of using this information to predict catalysis as well as to guide drug design and protein engineering.

> This study is the first application of the QM+MM method to protein–protein assemblies in a membrane environment

In work published in 2025 [Teuffel et al, 2025], we focused on a crucial parameter in the calculation of ET rates: the reorganization energy, λ. We revealed how a combination of molecular dynamics simulations and quantum mechanics computations can be used to compute λ for reducing cytochrome P450 17A1 (CYP17A1, steroid-17-alpha-hydroxylase) by the secondary redox protein cytochrome b5 (CYb5) – a critical step in the regulation of steroidogenesis for which experimental measurements of λ are not available. CYP17A1 is a drug target, for example, for prostate cancer treatment. The enzyme catalyzes two reactions with different products, with the second reaction being modulated by CYb5. Whether CYb5 exerts its effects via ET or allosteric effects is an open question whose answer would facilitate the design of CYP17A1 inhibitors with fewer side effects than existent active-site inhibitors that affect both reactions. This study is the first application of the QM+MM method to protein–protein assemblies in a membrane environment. Our results indicate that a quantum mechanical treatment of the redox-active cofactors is necessary even though the surrounding protein and solvent – which are modeled classically – contribute most to λ. We find that the λ values computed for the individual soluble globular domains of the two proteins sum to approximately the λ values computed for the membrane-bound CYP17A1-CYb5 complex, thereby indicating that additivity can be invoked in a computationally efficient approach to estimating λ values for such protein–protein complexes (see Figure 29). In further ongoing work, we are using the computed λ values to investigate the ET mechanism of CYP17A1 via molecular dynamics simulation, and we are also using the QM+MM method to compute λ values for other CYP-redox protein complexes.

This study was supported by the Klaus Tschira Foundation (SIMPLAIX Project 3), and computing resources were provided by the state of Baden-Württemberg through bwHPC, by the German Research Foundation (DFG) through Grant Nos. INST 35/1597-1 (Helix cluster) and INST 40/575-1 FUGG (JUSTUS2 cluster), and by the High Performance Computing Center, Stuttgart, Germany (HLRS; Project Dynathor).

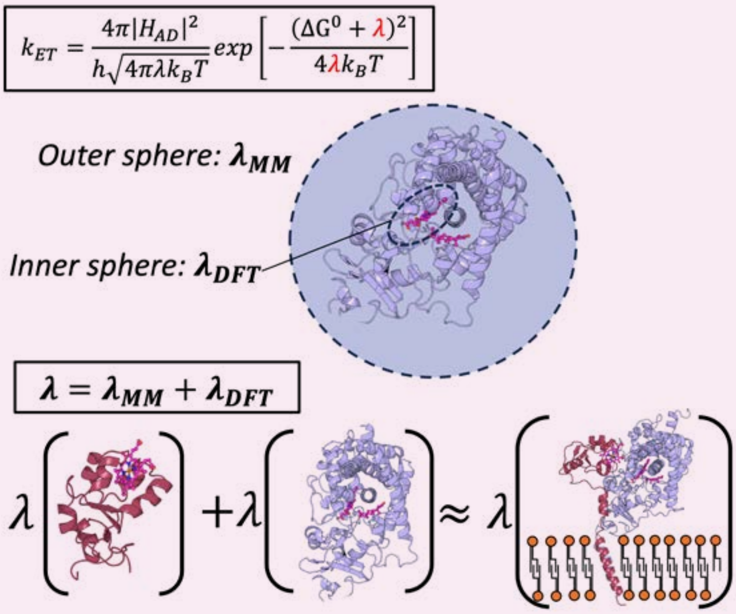


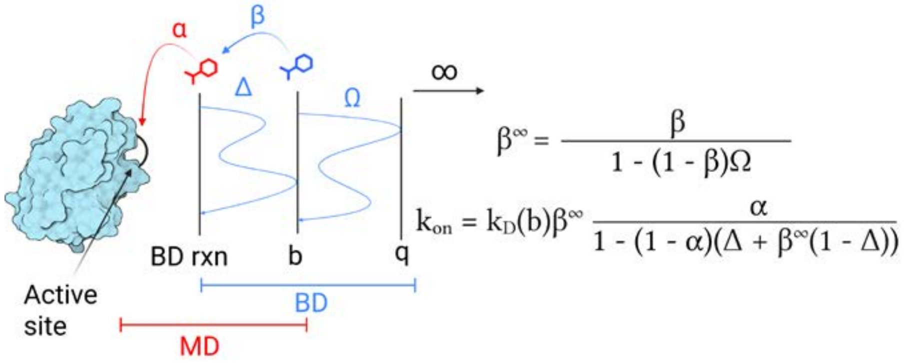
Figure 29
Schematic figure displaying how the electron transfer reorganization energy, λ, is computed using a QM+MM (quantum mechanics + molecular mechanics) approach for a cytochrome P450 enzyme and its redox partner protein, cytochrome b5 [Teuffel et al, 2025]. According to Marcus theory, the rate of electron transfer, kET, is dependent on λ. In order to compute λ, the protein system is divided into inner-sphere and outer-sphere regions, for which more detailed quantum mechanics and less detailed molecular mechanics calculations, respectively, are carried out. A key finding is that summing λ values for the individual proteins in solution roughly equals the λ value for the protein–protein complex in a membrane environment, which is much more computationally demanding to compute. Thus, this work provides a viable path to computing λ for other redox protein complexes.

Computing Drug–Protein Association Rate Constants by Combining Molecular and Brownian Dynamics Simulations

While drug design traditionally involves the search for compounds that bind their target with high affinity – that is, with the greatest binding free energies – drug–protein binding kinetic parameters are often found to correlate with drug efficacy in vivo. The MCM group is developing methods to compute drug–protein binding kinetic parameters: namely association and dissociation rates, which are often costly and time-consuming to measure experimentally.

> Due to the much lower computational cost of BD simulations and the reduced MD simulation time, the approach is computationally efficient while also preserving accuracy

In 2025, we published a multiscale simulation approach for computing protein–ligand association rate constants by combining Brownian dynamics (BD) and molecular dynamics (MD) simulations [Muniz-Chicharro et al, 2025]. BD – which employs a rigid-body model – is used for simulating long-range diffusion and diffusional encounter complex formation, whereas MD captures the subsequent formation of the bound complex, thereby providing a detailed treatment of short-range interactions and molecular flexibility. While existing methods that employ this approach have successfully yielded estimated association rate constants, they often require extensive computational resources. Therefore, we developed a multiscale pipeline that achieves improved computational efficiency by optimizing the sampling via BD simulation in order to generate an ensemble of diffusional encounter complexes in which the ligand comes very close to its protein binding site and then uses these complexes as starting structures for MD simulation (see Figure 30, next page). Due to the much lower computational cost of BD simulations and the reduced MD simulation time, the approach is computationally efficient while also preserving accuracy. The pipeline has been validated for a diverse set of protein–ligand complexes that vary in size, flexibility, and binding properties, thereby yielding values of bimolecular association rate constants that align well with experimental measurements in addition to providing insights into the physical determinants of association rates and the association pathways (see Figure 31, next page). The SDAMD workflow – available in the SDA 7.4.0 release (https://mcm.h-its.org/sda7/) – was developed as part of the pipeline to enable the setup and execution of the BD and MD simulations followed by their analysis in order to compute association rate constant values. This pipeline and the SDAMD workflow enable post-diffusional induce fit effects to be accounted for in computing bimolecular association rate constants. In further work, we are extending this approach to account for conformational gating and macromolecular crowding, both of which can affect the diffusional association of drugs to their macromolecular targets (Muniz-Chicharro, PhD thesis, 2025, see Chapter 9).



$$\beta^{\infty} = \frac{\beta}{1 - (1 - \beta)\Omega}$$

$$k_{on} = k_D(b)\beta^{\infty}\frac{\alpha}{1 - (1 - \alpha)(\Delta + \beta^{\infty}(1 - \Delta))}$$

Figure 30
Schematic illustration of the multiscale pipeline used to compute the drug–protein association rate constant (kon) by calculating probabilities α and β from molecular dynamics (MD) and Brownian dynamics (BD) simulations, respectively. In order to ensure efficient sampling during the simulations, the threshold distances for determining the probabilities are computed from the simulation results using predefined criteria and therefore vary according to the interaction properties of the molecules. From [Muniz-Chicharro et al, 2025].

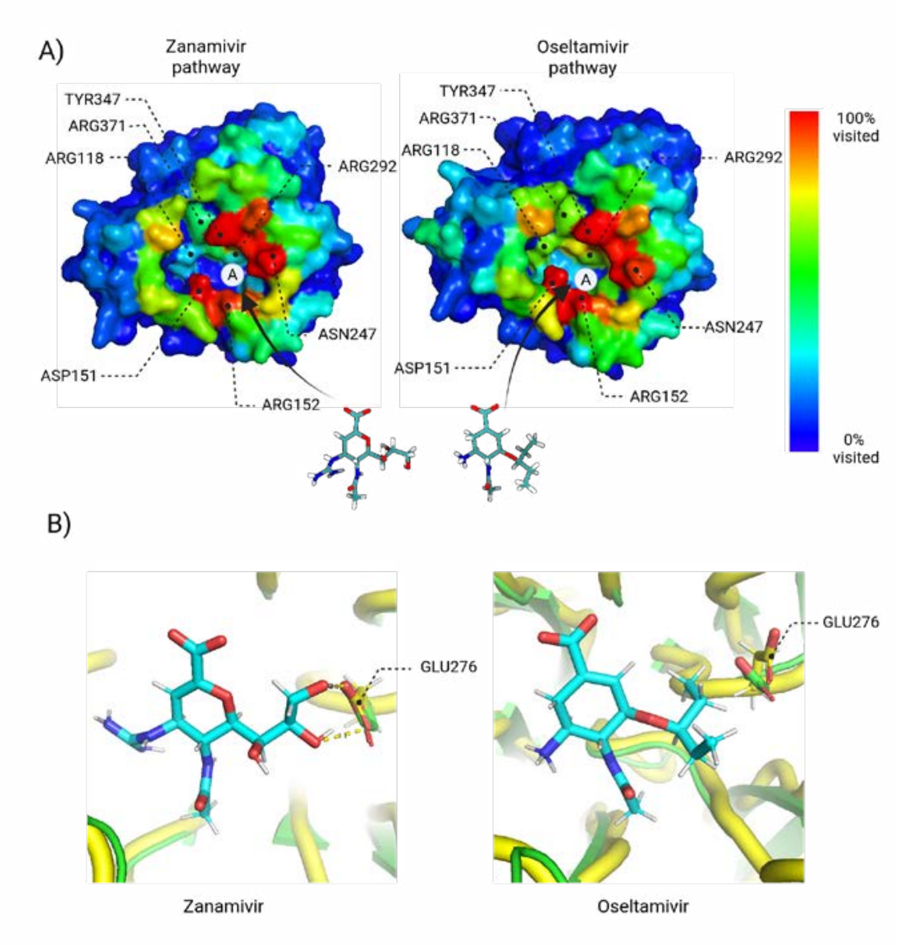


Figure 31
Pathways of two influenza drugs diffusing into their binding sites on the influenza virus neuraminidase N1 protein during BD simulations. The binding pathway of zanamivir (left) and oseltamivir (right) to the enzyme active site (indicated by the letter A) in BD simulations is represented by the coloring of the molecular surface of neuraminidase, spanning from red (areas visited in 100% of the reactive trajectories) to blue (areas never visited). During the MD simulations following the BD simulations, the molecules undergo induced fit conformational changes. This is illustrated by comparing the orientation of glutamic acid 276 (shown in stick representation) in neuraminidase due to interactions with the ligand during MD simulations: Zanamivir (left) forms attractive hydrogen-bonding interactions with the carboxylate group of residue 276 through its hydroxyl groups, whereas oseltamivir tends to repel the sidechain of residue 276 due to the steric and hydrophobic effects of its pentyl group. Upper part (A) of the figure from [Muniz-Chicharro et al, 2025].

Simulating Neurotrophin Receptors in Order to Investigate Signaling Mechanisms

Neurotrophin proteins are neurotrophic factors that promote neuronal survival. They have key roles in brain development and function as well as in neurodegenerative diseases. We participated in the EU-supported EuroNeurotrophin international doctoral training network on the discovery of neurotrophin small-molecule mimetics for treating neurodegeneration and neuroinflammation. Neurotrophins exert their actions by binding extracellularly to various transmembrane receptor proteins, thereby resulting in intracellular signal transmission. The neurotrophins form complexes with receptor dimers whose single-helix transmembrane domains interact and thereby transmit signals through the membrane. However, the mechanisms by which the transmembrane helices mediate receptor activation remain elusive. We therefore modeled three neurotrophin receptors – p75, TrkA, and TrkB – and carried out molecular dynamics simulations in order to investigate their functional dynamics. Due to the large and dynamic nature of the neurotrophin receptors, we first modeled and simulated parts of the receptors before building complete models of the neurotrophin receptors.

> These findings have implications both for mechanistic studies of neurotrophin receptor signaling and for the design of neuroprotective drugs

In 2025, we published our simulations of the transmembrane helical domains for these three receptors [Athanasiou et al, 2025]. We carried out unbiased and enhanced sampling molecular dynamics simulations of the transmembrane domain dimers of the wild-type p75, TrkA, and TrkB neurotrophin receptors and selected mutants in micelle and bilayer lipid environments at atomistic and coarse-grained levels of representation (see Figure 32). The coarse-grained simulations enabled the exploration of multiple states of the transmembrane domain dimers and revealed the influence of the lipid environment on the transmembrane helix arrangements. From the simulations, we identified active and inactive transmembrane helix arrangements of the p75 and TrkA receptors that are supported by experimental data and suggest that the two receptors have different signaling mechanisms through the C-terminal regions of the transmembrane helices. For TrkB, a single dominant but less energetically stable arrangement of the transmembrane domain dimer was observed. These findings have implications both for mechanistic studies of neurotrophin receptor signaling and for the design of neuroprotective drugs that can stabilize specific states of the transmembrane domain of the receptors.

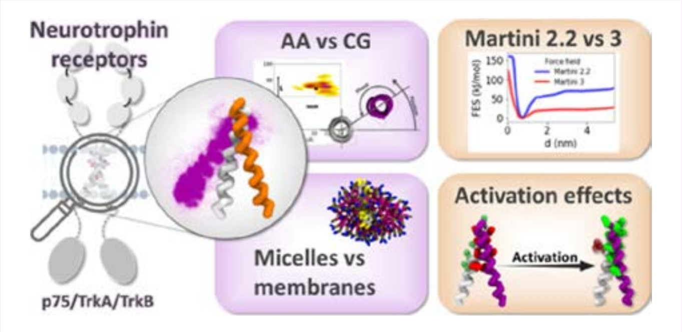


Figure 32
Illustration of the simulation of neurotrophin receptor transmembrane helix dimers, which transmit the signal due to the extracellular binding of neurotrophin into the cell interior. Simulations were carried out with the coarse-grain (CG) Martini model in order to achieve sufficient sampling of inactive and active arrangements of the helices. A modified version of Martini 3 was able to reproduce experimental data for micelles and was then used to study the transmembrane domains in phospholipid bilayers. The simulations indicated different rearrangements of the helices upon activation in the three studied neurotrophin receptors. From [Athanasiou et al, 2025].

Next, we simulated the extracellular domains of the neurotrophin receptors. While crystal structures of these domains are available, the glycan chains of these glycosylated receptors are not visible in the structures determined by X-ray crystallography. Glycosylation of some neurotrophin receptors has been reported to affect their localization and function, but the mechanism is unclear. We therefore built models of glycosylated forms of the receptors and carried out atomistic molecular dynamics simulations of both unglycosylated and glycosylated states in order to investigate the effects of glycosylation (see Figure 33). The degree and effects of glycosylation differ between the three studied receptors and go beyond solely shielding the protein surface. The TrkA and TrkB extracellular domains carry multiple N-glycans that dynamically shield the protein and – in the case of TrkB – increase the contact area between the receptor and the neurotrophin.

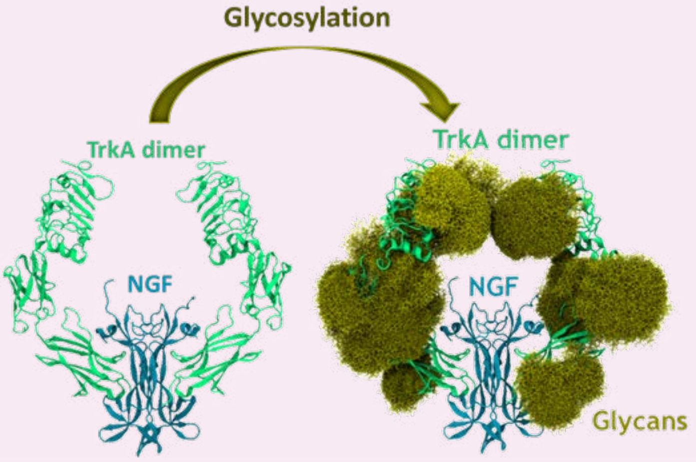


Figure 33
Glycosylation of the extracellular domain of the TrkA neurotrophin receptor shields the protein surface and promotes a more extended conformation of the receptor. A ribbon representation of the crystal structure of the extracellular domain of the TrkA receptor is shown. Glycan clouds are shown as superpositions of snapshots from atomic-detail molecular dynamics simulations of glycan chains modeled at glycosylation sites on the protein. Adapted from Tsengenes, biorxiv. https://www.biorxiv.org/content/10.64898/2025.11.29.691319v1 (2025).

Glycosylation also promotes more extended conformations of TrkA and TrkB that better accommodate neurotrophin binding and thereby support signal transduction.

In ongoing work, we are simulating the dynamics of full-length neurotrophin receptors in which the extracellular, intracellular, and transmembrane domains are connected by flexible linker domains and the receptors are embedded in a model neuronal membrane.

Research

2.7 NLP

Natural Language Processing

Group Leader

Prof. Dr. Michael Strube

Team

Souvik Banerjee
(PhD student)

Yi Fan
(PhD student)

Dr. Vagrant Gautam
(HITS Independent Postdoc; since September 2025)

Wei Liu
(PhD student; until August 2025)

Dr. Julius Steuer
(postdoc; since October 2025)

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.

2025 turned out to be a successful year for the NLP group, with three publications at *ACL conferences. In August 2025, Wei Liu submitted his PhD thesis, which he defended with distinction in early December. He took up a position in industry but will return to academia as a postdoc in early 2026. Since October, the NLP group has had a new postdoc, Julius Steuer, who completed his graduate studies at Saarland University. The group also works closely with the new HITS Independent Postdoc, Vagrant Gautam, who also earned their PhD from Saarland University. Vagrant arrived at HITS in September 2025 (see Chapter 2.12).

In 2025, Michael Strube co-chaired the ACL Test-of-Time Paper Award, which highlights papers published both 25 and ten years ago that have withstood the test of time. The award was presented in August 2025 at the ACL conference in Vienna, Austria. Michael Strube additionally co-chaired the “6th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI 2025),” which was co-located with the EMNLP conference in November 2025 in Suzhou, China. The year ended with a highlight in the form of the "Future of NLP" Symposium (see Chapter 5.4).

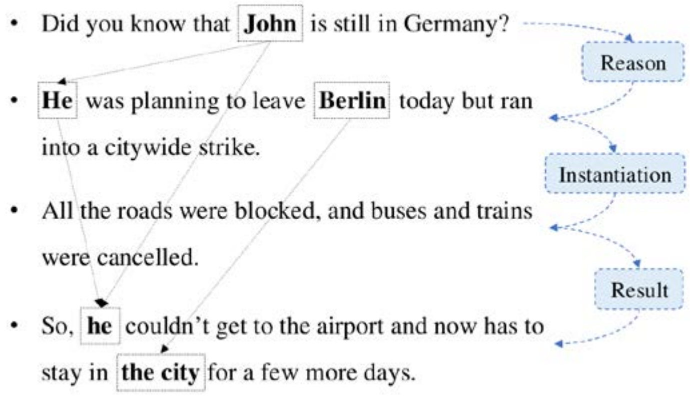
Natural Language Processing (NLP) ist ein interdisziplinäres Forschungsgebiet, das mit Methoden der Informatik linguistische Fragestellungen bearbeitet. Die NLP Gruppe entwickelt Methoden, Algorithmen und Tools zur automatischen Analyse von Sprache. Sie konzentriert sich auf die Diskursverarbeitung und verwandte Anwendungen.

*Das Jahr 2025 war ein wissenschaftlich erfolgreiches Jahr für die NLP Gruppe mit drei Veröffentlichungen bei *ACL-Konferenzen. Im August reichte Wei Liu seine Doktorarbeit ein, die er Anfang Dezember mit Auszeichnung verteidigte. Vorübergehend arbeitete er in der Industrie, wird aber Anfang 2026 eine Stelle als Postdoktorand antreten. Seit Oktober 2025 hat die NLP Gruppe einen neuen Postdoktoranden, Julius Steuer, der Doktorand an der Universität der Saarlandes in Saarbrücken war. Die NLP Gruppe arbeitet ebenfalls eng zusammen mit Vagrant Gautam, HITS Independent Postdoc, zuvor ebenfalls an der Universität des Saarlandes. Vagrant ist seit September 2025 am Institut (siehe Kapitel 2.12).*

Michael Strube war Co-Chair des ACL Test-of-Time Paper Awards, der Publikationen auszeichnet, die vor 25 und zehn Jahren publiziert wurden und ihre Bedeutung seitdem nicht verloren haben. Der Award wurde im August 2025 auf der ACL Konferenz in Wien verliehen. Michael Strube war erneut Co-Chair des “6th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI 2025)”, der im November 2025 während der EMNLP Konferenz in Suzhou ausgerichtet wurde. Das Jahr endete mit einem weiteren Highlight in Form des “The Future of NLP” Symposiums (siehe Kapitel 5.4).

Joint Modeling of Entities and Discourse Relations for Coherence Assessment (Wei Liu)

We investigate text coherence assessment, which is a task in computational linguistics that aims to distinguish coherent text from incoherent text using linguistic cues. In linguistics, coherence arises from various sources, notably from the continuity of entities across sentences and from logical or rhetorical relations (discourse relations) between sentences. Prior work focused on only one of these – that is, either on entity transitions (i.e., tracking shared nouns or coreferent mentions) or on discourse relations (e.g., cause or result) – but never combined both. Real texts, however, often rely on both cues interacting in complementary ways. We thus explore whether the joint modeling of entities and discourse relations improves coherence evaluation. For an illustration, see the example in Figure 34.



- Did you know that **John** is still in Germany?
- **He** was planning to leave **Berlin** today but ran into a citywide strike.
- All the roads were blocked, and buses and trains were cancelled.
- So, **he** couldn't get to the airport and now has to stay in **the city** for a few more days.

Figure 34
An example of a coherent text. It is coherent by means of both entities and discourse relations. We mark the interlinked entities in the text in bold and display the discourse relations between sentences.

In this sequence, entity-based signals alone fail to link all adjacent sentences: There are no overlapping entities between the second and third sentences; thus, a bridging relation between "citywide" and "roads" must be inferred. However, bridging is a complex phenomenon. Meanwhile, discourse relations – such as instantiation and result – establish coherence between these adjacent sentences. This example underscores the need to consider both entity continuity and discourse relations jointly.

First, we construct a graph that links entities and discourse relations to their respective sentences (see Figure 35). Then, we propose two modeling approaches that integrate entities and discourse relations. The first – Method I (Fusion) – organizes sentences, identifies entities and discourse relations in a flat sequence, and processes them with a fusion transformer. This model uses position-aware attention and a visible matrix to focus attention on relevant connections among sentences, entities, and relations, preserving structural information while building on a transformer architecture (see Figure 36, next page).

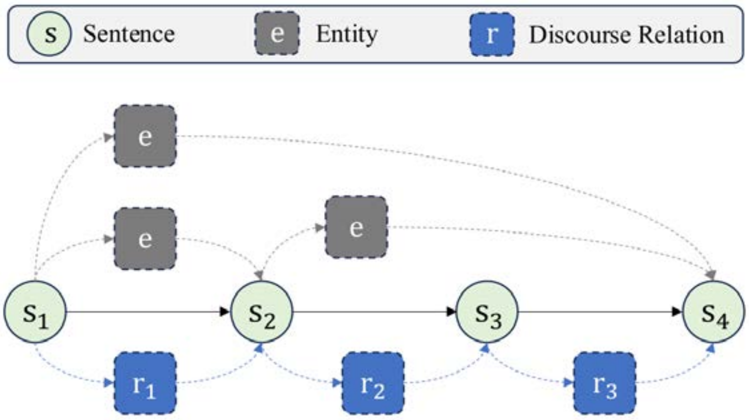


Figure 35
Sentences from Figure 34 linked by entities and discourse relations.

Joint modeling of entities and discourse relations improves coherence evaluation

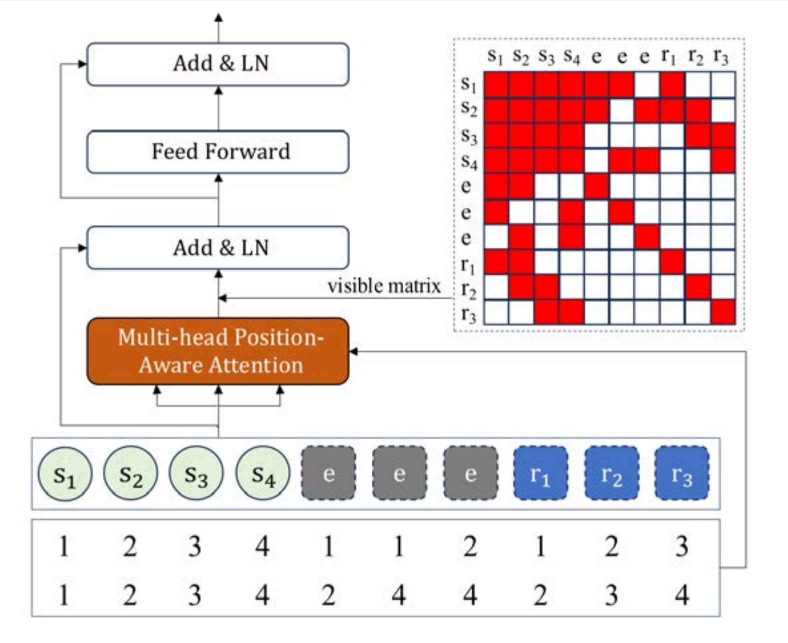


Figure 36
Method I: The sentences, entities, and discourse relations in Figure 35 are put into a flat structure in which each element is assigned a two-dimensional position that indicates its start and end position in the original sentence sequence. This flat input is then processed by a fusion transformer.

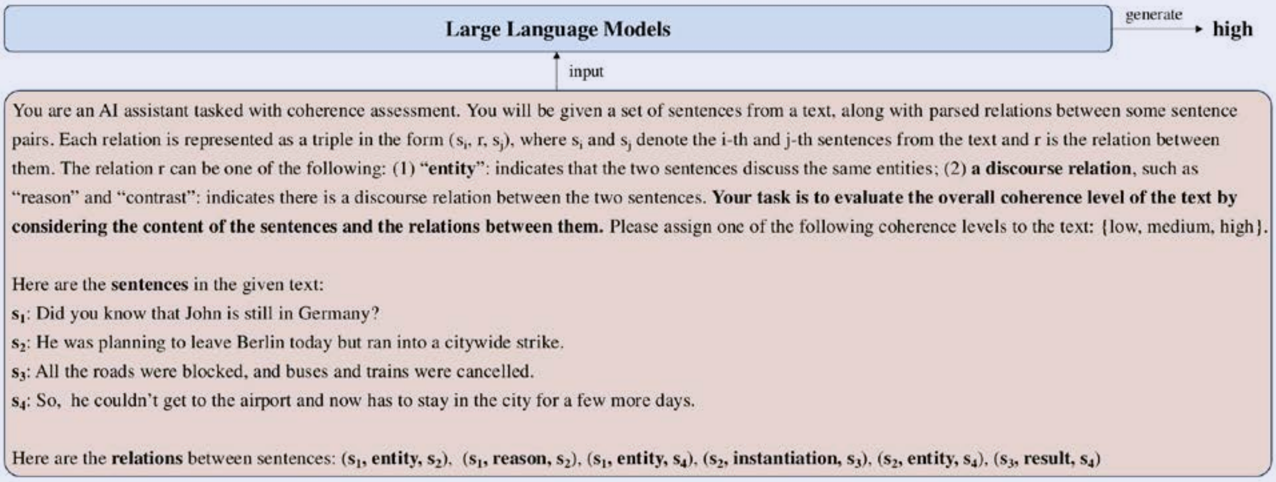
The second approach – Method II (prompting) – encodes entity and discourse relation patterns in natural language and supplies them as part of a prompt to a large language model (LLM). By converting the connection graph into triples (e.g., sentence–entity–sentence or sentence–relation–sentence), the LLM is instructed to consider these cues when judging coherence (see Figure 37).

Experiments on three benchmark datasets — GCDC (annotated for discourse coherence), CoheSentia (GPT-3 generated stories with human coherence ratings), and TOEFL (essay scoring) — show that models that integrate both entities and discourse relations outperform baselines that use only text, only entities, or only discourse relations. In the Fusion setting, the joint model consistently yields greater accuracy than ablations across multiple domains. In the prompt setting, the explicit inclusion of both entities and discourse relations improves performance in zero-shot and fine-tuned scenarios. Furthermore, we show that combining signals enhances robustness to imbalanced label distributions and improves domain transfer.

We provide empirical evidence that entity continuity and discourse relations complement each other in modeling text coherence and that joint modeling yields more reliable coherence assessment than does using either entities or discourse relations in isolation.

> Entity continuity and discourse relations complement each other in modeling text coherence

Figure 37
Method II: We describe the graph from Figure 35 in natural language. We then instruct LLMs to assess the coherence of the input document.



You are an AI assistant tasked with coherence assessment. You will be given a set of sentences from a text, along with parsed relations between some sentence pairs. Each relation is represented as a triple in the form (s_i, r, s_j), where s_i and s_j denote the i-th and j-th sentences from the text and r is the relation between them. The relation r can be one of the following: (1) **"entity"**: indicates that the two sentences discuss the same entities; (2) **a discourse relation**, such as "reason" and "contrast": indicates there is a discourse relation between the two sentences. **Your task is to evaluate the overall coherence level of the text by considering the content of the sentences and the relations between them.** Please assign one of the following coherence levels to the text: {low, medium, high}.

Here are the **sentences** in the given text:
s_1: Did you know that John is still in Germany?
s_2: He was planning to leave Berlin today but ran into a citywide strike.
s_3: All the roads were blocked, and buses and trains were cancelled.
s_4: So, he couldn't get to the airport and now has to stay in the city for a few more days.

Here are the **relations** between sentences: (**s_1, entity, s_2**), (**s_1, reason, s_2**), (**s_1, entity, s_4**), (**s_2, instantiation, s_3**), (**s_2, entity, s_4**), (**s_3, result, s_4**)

Consistent LLM-Based Discourse-Level Temporal Relation Extraction (Yi Fan)

We address the task of temporal relation extraction, which aims to determine how events in a text relate in time (e.g., before, after). Extracting these relations across entire documents is important for building coherent temporal structures that are useful in summarization, timeline construction, and comprehension tasks. Although large language models (LLMs) have displayed strong general capabilities, they perform poorly in zero-shot and few-shot temporal relation tasks compared with smaller fine-tuned models and often produce inconsistent outputs when predicting inter-event relations. We highlight three challenges in applying LLMs to discourse-level temporal relation extraction: (1) selecting the relevant context from long documents without introducing noise, (2) improving performance to be competitive with fine-tuned smaller models, and (3) ensuring logical consistency in the set of predicted temporal relations since inconsistencies undermine coherent temporal understanding.

To address these challenges, we propose a three-step framework. First, a context selection strategy extracts only event-relevant text segments in order to reduce noise and help the model focus on informative content. Two techniques are used: discourse segmentation based on entity coherence and LLM-guided sentence selection tailored to specific event pairs. Second, inspired by Allen's interval algebra, we design structured prompts that lead the model to reason explicitly about the start and end times of events (a form of chain-of-thought prompting) before determining their temporal relation, thereby encouraging systematic inference rather than direct classification. Third, we introduce a reflection-based consistency learning approach: After an initial round of fine-tuning, the model's predictions are checked for logically inconsistent triples of events; these inconsistencies are then used to generate new training inputs without revealing true labels, and the model is fine-tuned again so that it can self-correct and internalize temporal coherence (see Figure 38).

We evaluate our approach on discourse-level corpora annotated with temporal relations. The experiments reveal that our approach improves both the relation extraction performance (e.g., F1 scores) and the consistency of temporal graphs compared with baseline models. We perform ablation studies that indicate that context selection and structured prompting contribute significantly to the performance of our approach. However, there are limitations: Even with these enhancements, LLMs still struggle with short-distance temporal dependencies and with accurately capturing event duration, especially for stative verbs. We additionally observe that simply increasing the model size does not substantially improve temporal understanding. Hence, there is a need for reasoning and the induction of implicit temporal information.

In summary, our work highlights the importance of input design, reasoning scaffolding, and iterative consistency learning when it comes to adapting large language models to discourse-level temporal relation extraction, and it also suggests directions for future research to further enhance temporal reasoning capabilities.

> Challenges in applying LLMs to discourse-level temporal relation extraction

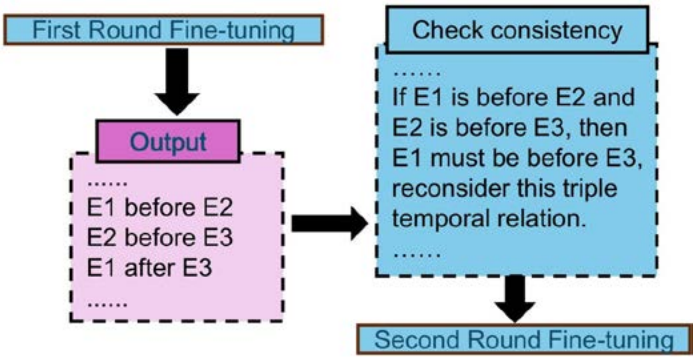


Figure 38
We check the consistency of the prediction from the initial round. If we find inconsistent triples, we construct the new input only for the test samples that are inconsistent. In this process, the new input does not include the true label. Then, our model fine-tunes again with the new input in the second round in order to improve consistency.

The Future of NLP

The year came to a close with an NLP highlight: The NLP group organized the first-ever "HITS Future of … Symposium" from 24 – 25 November 2025 on "The Future of NLP." This two-day event featured a keynote speech by Mirella Lapata (University of Edinburgh) on "Compositional Intelligence: Coordinating Multiple LLMs for Complex Tasks." The symposium program was divided into sections on applications, theory, and NLP and society. The first day ended with panel discussions chaired by Michael Strube, the second with a panel discussion chaired by Vagrant Gautam. The speakers were encouraged to ponder the future of NLP, including what our field will look like in five, ten, and 25 years (for more information, see Chapter 5.4).

Research

2.8 PSO

Physics of Stellar Objects

Group Leader

Prof. Dr. Friedrich Röpke

Team

Veronica Agaeva
(visiting scientist; Heidelberg University)

Dr. Róbert Andrássy
(visiting scientist; Heidelberg University)

Korinna Riccarda Bayer
(master's student, since October 2025)

De. Ke-Jung Chen
(visiting scientist; AvH Scholarship, July–September 2025)

Wun-Yi Chen
(visiting scientist; DAAD & NSTC Scholarship; July–September 2025)

Paul Christians
(visiting scientist; TU Darmstadt)

Christina Fakiola
(visiting scientist; Heidelberg University)

Dr. Damien Gagnier
(visiting scientist; Heidelberg University)

Prof. Dr. Raphael Hirschi
(Klaus Tschira Guest Professor; April–December 2025)

Alexander Holas
(visiting scientist; Heidelberg University)

Dr. Mike Lau
(visiting scientist; Heidelberg University)

Dr. Giovanni Leidi
(postdoc)

Dr. Georgios Lioutas
(postdoc)

Dr. Javier Moran Fraile
(postdoc; until May 2025)

Vijayalakshmi Vijayakumaran Nair
(master's student; until May 2025)

Jonas Reichert
(master's student; since September 2025)

Dr. Federico Rizzuti
(postdoc; AvH Scholarship; since March 2025)

Arkaprabha Roy
(master's student; since April 2025)

Marco Vetter
(visiting scientist; Heidelberg University)

Kristián Vitovský
(PhD student)

Stars shine across the electromagnetic spectrum and are the fundamental building blocks of galaxies and larger cosmological structures. With the help of numerical simulations, the Physics of Stellar Objects group seeks to understand the processes that take place in stars and stellar explosions. New numerical techniques and the ever-increasing power of supercomputers facilitate the modeling of stellar objects in unprecedented detail and precision.

Classical astrophysical theory describes stars as one-dimensional objects in hydrostatic equilibrium. However, their simplifying assumptions limit the predictive power of such models. New numerical tools explore 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.

Another goal is to model the thermo-nuclear explosions of white dwarf stars that lead to Type Ia supernovae. These supernovae are the main source of iron in the Universe and serve 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.

Sterne leuchten über das elektromagnetische Spektrum hinweg und sind die grundlegenden Bausteine von Galaxien und größeren kosmologischen Strukturen. Mithilfe numerischer Simulationen versucht die Forschungsgruppe, die Vorgänge in Sternen und bei Sternexplosionen zu verstehen. Neue numerische Verfahren und die stetig wachsende Rechenleistung von Supercomputern ermöglichen die Modellierung stellarer Objekte in bisher unerreichter Detailgenauigkeit und Präzision.

Die klassische astrophysikalische Theorie beschreibt Sterne als eindimensionale Objekte im hydrostatischen Gleichgewicht. Ihre vereinfachenden Annahmen schränken jedoch die Vorhersagekraft solcher Modelle ein. Neue numerische Werkzeuge untersuchen dynamische Phasen der Sternentwicklung mittels dreidimensionaler Simulationen. Unser Ziel ist es, eine neue Generation von Sternmodellen zu entwickeln, die auf einer verbesserten Beschreibung der in Sternen ablaufenden physikalischen Prozesse basiert.

Ein weiteres Ziel unserer Gruppe ist es, die thermonuklearen Explosionen von Weißen Zwergen zu modellieren, die zu Typ Ia Supernovae führen. Diese Supernovae sind die Hauptquelle für Eisen im Universum und dienen in der Kosmologie als Entfernungsindikatoren, was zu der spektakulären Entdeckung der beschleunigten Expansion des Universums geführt hat. Mehrdimensionale strömungsdynamische Simulationen in Verbindung mit Berechnungen zur Nukleosynthese und Strahlungstransportmodellen liefern ein detailliertes Bild der physikalischen Prozesse, die in Typ Ia Supernovae ablaufen, und werden in der PSO-Gruppe auch auf andere Arten kosmischer Explosionen angewendet.

A New Perspective on How the Universe Built Its Rarest Constituents

In the vast tapestry of the Universe, the origins of heavy elements – such as gold, platinum, and uranium – have long puzzled scientists. These elements – which are heavier than iron – are not forged in the hearts of ordinary stars through simple fusion; instead, they arise from extreme cosmic events involving rapid neutron capture, which is known as the r-process. A groundbreaking study conducted in part at our institute explores a novel way of understanding how these heavy elements formed and were distributed across the cosmos. Entitled “Distribution of Heavy-Element Abundances Generated by Decay from a Quasi-Equilibrium State” [Röpke G, Blaschke, Röpke FK, 2025], this research proposes a “freeze-out” model that mirrors processes seen in laboratory experiments and applies the model to astronomical scales. By identifying a universal starting point in the early Universe or explosive stellar events, the work sheds light on why the patterns of heavy elements in our Solar System and distant stars look so remarkably similar.

At its core, the study draws inspiration from physics analogies. Imagine nuclear matter expanding rapidly, much like in heavy-ion collisions at particle accelerators or in the rare “ternary fission,” in which a heavy nucleus splits into three parts. In these scenarios, nuclei form in a hot, dense soup and “freeze out” their composition because the system cools and expands too quickly for further reactions to maintain equilibrium; only the decay of excited states occurs. The researchers – that is, Gerd Röpke from the University of Rostock, David Blaschke from both the University of Wroclaw and the Helmholtz-Zentrum Dresden-Rossendorf, and Friedrich K. Röpke from both HITS and Heidelberg University – extend this idea to cosmic nucleosynthesis. The researchers account for quantum effects such as “Pauli blocking,” in which particles cannot occupy the same state, and the "Mott effect," which dissolves nuclei at high densities. This work is not merely theoretical; rather, it is grounded in real data from nuclear physics experiments in laboratories.

To model the formation of heavy elements in astrophysical scenarios, the team uses a sophisticated framework called the Nonequilibrium Statistical Operator, which was developed by Dmitri Zubarev and collaborators. The framework treats nuclear matter as a grand canonical ensemble, with the state defined by parameters that are the non-equilibrium generalizations of temperature (T), baryon density (nB), and proton fraction (Yp). The team searches for an initial quasi-equilibrium state in which heavy, neutron-rich nuclei abound and then let this state evolve through radioactive decay – such as beta decay, alpha emission, and fission – in order to match the abundances observed today. After crunching the numbers, they pinpoint a sweet spot of T = 5.266 million electron volts (ca. 61 billion Kelvin), nB = 0.013 particles per cubic femtometer (a density just below nuclear saturation), and Yp = 0.13 (i.e., mostly neutrons).

In this initial state, the mass fractions of elements extend to superheavy nuclei around a mass of A=600, with peaks near “magic numbers” – in which nuclei are particularly stable – and even a hypothetical double-magic 358Sn. This gigantic nucleus – far beyond the limits of the known elements – would be neutron-rich and unstable, eventually fissioning into fragments that fill in the observed abundance peaks at A around 128–138 and 197–208. Post–freeze-out, processes such as neutron evaporation (i.e., shedding excess neutrons) shift the distribution, while alpha decays and fissions redistribute mass to the lighter end of the distribution of heavy elements. As shown in Figure 39, the model beautifully reproduces the coarse-grained solar abundances from A=76 onward, where “coarse-grained” means

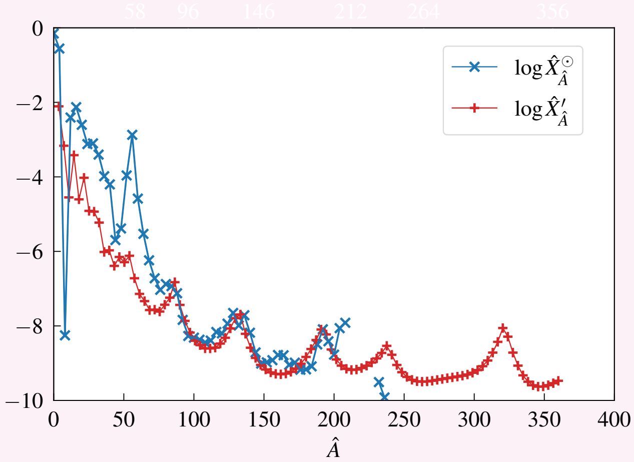


Figure 39

Accumulated mass fraction according to the new model (red) compared with the solar values (blue) (figure taken from [Röpke G, Blaschke, Röpke FK, 2025].

> The study explores a novel way of understanding how heavy elements such as gold, platinum, and uranium are formed

grouping every four mass numbers in order to smooth out fine details from slower processes such as neutron-capture networks.

For light elements, the approach also aligns with helium-4 abundances of ca. 24.5% by mass, but with tweaks for in-medium effects. At high densities, helium clusters dissolve due to the Mott effect, thereby preventing overproduction. Comparisons with lab data from heavy-ion collisions (at energies from Fermi to LHC scales) and from the ternary fission of isotopes such as 252Cf validate the freeze-out concept. For instance, fission yields match equilibrium predictions only when quantum corrections are included, with freeze-out temperatures around 1–10 MeV and densities far below saturation.

The implications are profound. This universal initial state could explain r-process universality without relying solely on rare events such as neutron star mergers. Instead, the universal initial state points to the early Universe itself via inhomogeneous Big Bang nucleosynthesis (IBBN). In IBBN, primordial density clumps – perhaps seeded by quantum fluctuations or primordial black holes – briefly reach nuclear densities, thereby allowing heavy-element freeze-out before dilution. This scenario would seed the Universe with metals from the very beginning, thus explaining why even the oldest stars show traces of r-process elements and why Population III stars might be mythical. The proposed nucleosynthesis process also ties into supernova and merger scenarios, in which similar conditions arise in ejected matter.

In summary, this study demystifies heavy-element origins by proposing a freeze-out from a hot, neutron-rich quasi-equilibrium that, after decay, matches what we see today. It thus bridges lab physics with cosmology, offering a fresh perspective on the Universe's chemical evolution and underscoring the power of interdisciplinary science.

A Snapshot of the Ignition of a Supernova Explosion

Explosions known as Type Ia supernovae light up the cosmos like brilliant fireworks. These stellar blasts are crucial for scientists because they help measure distances across space, thereby aiding our understanding of dark energy – the mysterious force accelerating the Universe's expansion. These special supernovae also forge much of the iron in our Galaxy, thus contributing to the building blocks of planets and life. While the exact origins of these supernovae have puzzled astronomers for years, a recent study published in Nature Astronomy [Das et al, 2025] sheds new light on this enigma by examining the remnants of one such explosion, thereby offering evidence of a specific type of stellar demise.

The specific class of supernovae considered here is explained by the thermonuclear explosion of a white dwarf – a dense, Earth-sized star with the mass of our Sun. White dwarfs are stable unless they gain extra mass from a companion star, pushing them toward instability, which sets in when they approach 1.4 times the Sun's mass – the well-known Chandrasekhar mass limit for white dwarf stars. Traditionally, scientists thought that growth in mass toward this limit would ultimately lead to the ignition of a thermonuclear explosion resulting in a Type Ia supernova. However, a number of arguments based on detailed observations made in recent years have called this model into question. New studies have found better agreement with observed supernovae if the ignition instead occurs in “sub–Chandrasekhar-mass” white dwarfs long before they reach the limit of stability.

> Our study provides direct evidence that double detonations in lighter white dwarfs are a real pathway for Type Ia supernovae

But why would such objects explode in the first place? A possible mechanism for this explosion is the "double detonation" scenario for these lighter white dwarfs. Imagine a white dwarf with a carbon–oxygen core that accumulates a thin layer of helium from a companion, such as a helium-burning star or another white dwarf. The helium layer ignites first in a detonation – a rapid, shock-wave-driven burn. This initial blast then shocks the core into a second detonation that burns the carbon–oxygen material and gives rise to a bright supernova explosion.

How can this hypothesized mechanism for the ignition of the supernova explosion be tested observationally? When observed near their maximum brightness, distant supernovae are point sources that cannot be spatially resolved, even with the most powerful telescopes. However, as time passes and the explosion ejecta expand, a large supernova remnant forms that can be resolved. Consequently, the research published by guest scientists from Australia at HITS together with the PSO group focuses on a young supernova remnant called SNR 0509-67.5, which is located in the Large Magellanic Cloud, a neighboring galaxy to our Milky Way. This remnant is the leftover shell from a Type Ia supernova that occurred about 300–350 years ago.

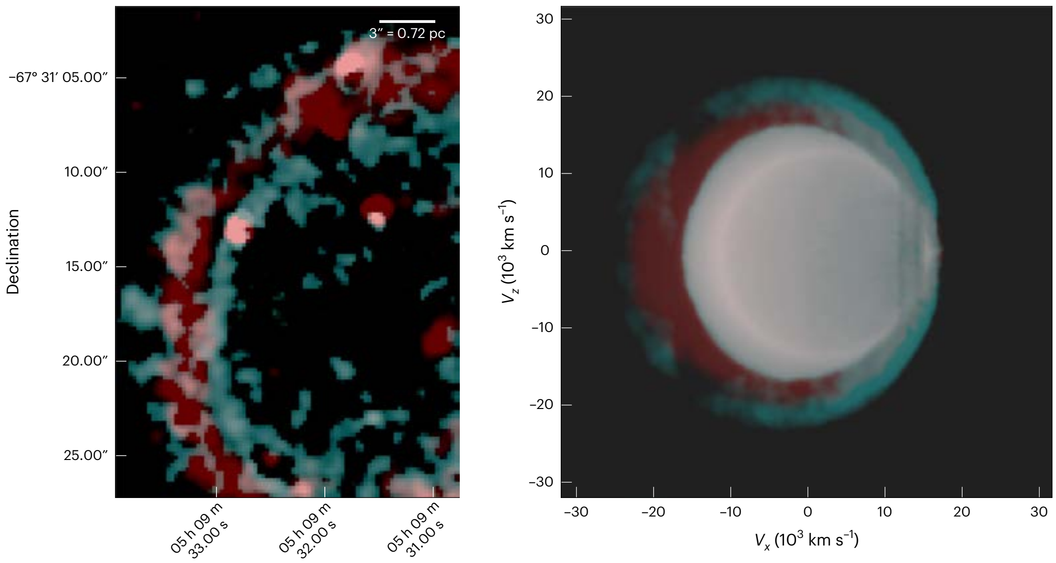


Figure 40
Comparison of observed composition of the supernova remnant (left) with an explosion model (right). The two shells – which are composed mainly of calcium – are shown in cyan, while sulfur is colored in red (figure taken from [Das et al, 2025]..

What makes this study special is its use of advanced technology: namely high-resolution integral field spectroscopy from the MUSE instrument on the European Southern Observatory's Very Large Telescope. Over many nights of observation totaling around 29 hours, the team captured detailed data on the remnant's structure, revealing hidden layers of elements created during the explosion. In SNR 0509-67.5, the team discovered a striking double-shell structure in highly ionized calcium, which was detected through its emission lines. There is an outer shell of calcium at ca. 2.06 parsecs (one parsec is roughly 3.26 light years) from the center and an inner shell at 1.73 parsecs. Between them lies a single shell of sulfur – another element produced in the explosion.

This configuration matches computer simulations of a double detonation in a sub–Chandrasekhar-mass white dwarf with a core of ca. 1 solar mass and a thin helium shell of ca. 0.03 solar masses. The outer calcium comes from the first detonation in the helium layer, where low densities favor producing intermediate-mass elements, such as calcium. The inner calcium arises from the second detonation in the outer parts of the carbon–oxygen core, where burning is not complete enough to make heavier iron-group elements. Sulfur – being lighter – forms in the density sweet spot between the two calcium layers. The double-shell calcium structure with sulfur sandwiched in between is therefore a fingerprint of the "double-detonation" mechanism of a Type Ia supernova explosion.

The progenitor system – the setup before the explosion – could be a single white dwarf accreting helium from a companion or a double degenerate system in which two white dwarfs merge. In mergers, simulations suggest even more complex "quadruple detonations," but the basic double-detonation fingerprint holds. The study rules out alternatives such as clumping and projection effects as the sulfur's position and consistent velocities support layered abundances from the explosion itself.

In a sense, supernova remnants act like time capsules, preserving the ejecta's structure long after the initial light has faded. In the early phases of a supernova, opacity hides inner layers, but remnants allow "tomography" – that is, slicing through the debris with spectroscopy. This approach can reveal explosion mechanisms that light curves and spectra alone cannot.

The findings emphasize the value of studying young remnants in nearby galaxies. Future work could apply similar techniques to other remnants and use advanced models that simulate ionization and excitation in order to match observations. This approach could help us classify supernova subtypes and trace their evolution, thereby ultimately enabling us to paint a fuller picture of how these cosmic beacons form and influence the Universe.

Overall, our study provides direct evidence that double detonations in lighter white dwarfs are a real pathway for Type Ia supernovae. This finding represents a step toward resolving long-standing debates and reveals that nature's explosions are more varied and ingenious than we once thought.

2.9 SDBV

Scientific Databases and Visualization

Group Leader

PD Dr. Wolfgang Müller

Team

Dr. Susan Eckerle

Arno Fricke

Martin Golebiewski

Anne Elin Heggland

Xiaoming Hu

Matthew Johnstone
(since July 2025)

Lukrécia Mertová

Ghadeer Mobasher
(until January 2025)

Dr. Maja Rey

Dr. Andreas Weidemann
(until March 2025)

Dr. Ulrike Wittig

Students

Elisa Brost

Tobias Martiné

Maria Paula Schröder

Fabian Springer

The SDBV group continues both to make data FAIR (findable, accessible, interoperable, and re-usable) and to enable others to make data FAIR. While this mission has remained, our projects have grown, and our surroundings have evolved. The advent of large language models (LLMs) and AlphaFold has clearly changed the perception of FAIR data and their usefulness as well as the way in which we generate these data.

Some may think that the problem of how to extract and curate data from a paper has been solved. However, while LLMs are certainly pushing the boundaries of what is doable, as of now, they do not provide a turnkey solution to data curation problems. Key problems include hallucinations (i.e., the LLM fabricates a result that is disconnected from the truth), repeatability, and verifiability.

These challenges motivate continued work on reliable FAIR data. In 2025, we continued work on standardization (within ISO and NFDI4Health), on within-project data sharing (de.NBI, ELIXIR, and LiSyM-Cancer), and on professionally curated data (de.NBI and SABIO-RK).

For more than 10 years now, the group has been involved in ISO standardization activities. We report on ISO TC276/WG5 below. The second, longer section below deals with the re-write of SABIO-RK, our database for reaction kinetics data.

Die SDBV macht weiter Daten FAIR (findable, accessible, interoperable, reusable) und hilft anderen, sie FAIR zu machen. Während diese Mission konstant bleibt, wachsen Projekte und die Umgebung verändert sich. Large Language Models (LLMs) und AlphaFold haben beide die Sicht auf FAIRe Daten und ihren Nutzen verändert, ebenso wie die Sicht darauf, wie man FAIRe Daten erzeugt.

Man denkt vielleicht, dass die Extraktion und Kuratierung von Daten aus Veröffentlichungen jetzt gelöst sei. Jedoch verschieben LLMs zwar die Grenzen des automatisch Machbaren, aber aktuell sind sie keine unmittelbare Lösung des Kuratierungsproblems. Wichtige Hindernisse sind z.B. "Halluzinationen" (also vom LLM Erfundenes) und die Wiederholbarkeit und Nachprüfbarkeit von Resultaten.

Diese Herausforderungen motivieren weitere Arbeit an verlässlichen FAIRen Daten. Im letzten Jahr haben wir weiter an Standardisierung (ISO & NFDI4Health), Daten-Sharing in Projekten (de.NBI, ELIXIR, LiSyM-Cancer) und professionell kuratierten Daten (de.NBI, SABIO-RK) gearbeitet.

Seit mehr als 10 Jahren ist die Gruppe an ISO-Standardisierungsaktivitäten beteiligt. Wir berichten über die ISO TC276/WG5. Zu guter Letzt heben wir die Neufassung von SABIO-RK, unserer Datenbank für Reaktionskinetik, hervor.

Emerging Standards for In Silico and Personalized Medicine

Many different types of data are used to set up computational models for simulations and predictions in personalized medicine, for example, for in silico clinical trials, disease prognosis, treatment predictions, and patient stratification. These data need to be interoperable in order for researchers to be able to integrate them into models or to use them for model validation. Thus, standards are crucial for structuring, describing, and associating models and data as well as their respective parts, corresponding interactions, graphical visualization, and applied methods [Golebiewski & Mayer: Standardization landscape, needs and gaps for the virtual human twin (https://doi.org/10.5281/zenodo.10492795), Golebiewski & Mayer: Data Formats for Systems Biology, Systems Medicine and Computational Modeling (https://doi.org/10.1016/B978-0-323-95502-7.00164-0)].

Standards are defined by the scientific communities, such as the COMBINE network and the Research Data Alliance, and by standard defining organizations (SDOs), such as the International Organization for Standardization (ISO). SDBV group member Martin Golebiewski is the convenor of the ISO technical committee working group ISO/TC 276/WG 5 Data Processing and Integration, which has published several important standards in the domain in the past few years.

ISO 20691 provides a framework for the development and application of domain-specific and interoperable (meta-)data standards in the life sciences. It defines requirements and rules for formatting, describing, and documenting data as well as for designing interoperable data formats and terminologies. There is also an interactive annex available online. The ISO 23494 series “Provenance information model for biological material and data” describes how to document the provenance information in the life sciences in order to be able to trace it back to its original sources after several data processing steps.

ISO 9491 “Recommendations and requirements for predictive computational models in personalized medicine research” – initiated by the EU-STANDS4PM project that the SDBV was part of – provides guidelines for modeling in personalized medicine. It has two parts. The first part is ISO 9491-1 "Guidelines for constructing, verifying, and validating models," which was first published in 2023 and has recently been revised. It was included as a requirement in funding calls of the European Commission on Virtual Human Twin projects. The second part is ISO TS 9491-2 "Guidelines for implementing computational models in clinical integrated decision support systems," which is currently under publication.

To complement these activities, a new joint working group was formed that consists of the ISO and the International Electrotechnical Commission IEC and that has the goal of standardizing the credibility assessment of computational modeling for medical devices through verification, validation, and uncertainty quantification. This work will build on the national US standard ASME V&V 40:2018 and extend it to an international standard. In this emerging standardization committee, the SDBV also plays a leading role, with Martin Golebiewski as designated co-convenor.

SABIO-RK, New Generation

> SABIO-RK currently contains approximately 75,000 entries

Biochemical reactions – that is, metabolic, transport, and signaling reactions – are essential to keeping all organisms alive by providing energy, producing and repairing the building blocks of the body, and keeping balance inside a cell.

Models help us understand, predict, and control how metabolism works in living systems. In order to be able to create models of these processes, information about reaction kinetics is required. While certain parameters can be estimated theoretically, models are more reliable when they use experimentally measured data.

The old version of our publicly accessible database SABIO-RK (https://sabiork.h-its.org/) provided modelers with such reaction kinetic information. These kinetic parameters – together with their contextual information – were manually curated from the scientific literature and were organized into structured tables, where they were linked to ontologies and additional data resources. In this old version of software, which was online and continuously refined beginning in 2006, users could explore the SABIO-RK using a wide range of search options. They could either perform a free-text search across all fields, set filters for some attributes (wildtype/mutant, temperature, etc.), or select particular attributes that the query was to be restricted to (e.g., compounds).

SABIO-RK offered distinct result views, including two types of tables that displayed key attributes, such as the biochemical reaction, catalyzing enzymes, the originating organism, and tissues. Selected entries could be exported in CSV or SBML (Systems Biology Markup Language) formats. Other result views were more visual-based, including the clickable bar charts, for which the result set was sorted by some prominent attributes (e.g., organisms, tissues, cell locations, EC numbers), and the Visual-Search developed in the SABIO-VIS project (see the 2019 Annual Report), which allowed fine-grained visualization and interactive exploration of the multi-dimensional data.

Previous Curation Workflow

Curators uploaded references to suitable scientific publications into the legacy input interface, where the publications were distributed as tasks and assigned to student workers, who extracted the relevant data and entered it into structured forms connected to a MySQL database in the backend. Subsequently, curators reviewed the same publications, validated and annotated the extracted data, and finally submitted the entries to the input database (PostgreSQL) before these entries were transferred to the final output database (also PostgreSQL).

Limitations of the Legacy System

This separation into three databases that are not seamlessly integrated emerged as a major limitation of the historically grown system. Transferring new data from the input to the output database required multiple supervised steps, including manual intervention, which made the process time-consuming and error-prone. In addition, post-transfer corrections often had to be applied directly to the frontend database because entries could not simply be overwritten from the input database. This process inevitably led to discrepancies between the input and output databases.

Further limitations of our old system resulted mainly from its outdated architecture, including restricted API access, the inconsistent naming of identical attributes, the absence of both data provenance and change tracking, and the lack of a JSON-based representation of the data entries.

Re-Implementation of SABIO-RK

SABIO-RK is funded by the BMFTR (Federal Ministry of Research, Technology and Space) and the MWK BW (Ministry of Science, Research and the Arts of the State of Baden-Württemberg) via de.NBI (German Network for Bioinformatics Infrastructure), which has the mission of providing life scientists in Germany with comprehensive, high-quality bioinformatics services, training, and infrastructure in order to support data analysis, management, and reproducible research.

While we had long wanted to re-architect SABIO-RK, our resources did not suffice to keep SABIO-RK online and rewrite it at the same time. In 2025, de.NBI residual funds enabled us to contract an external company in order to re-implement SABIO-RK. Our objective was to use an open-source code base in order to prevent long-term licensing expenses. Additionally, we aimed to implement the code with fewer components in one single framework.

A visual redesign of the website was also considered a beneficial improvement. It was important that the full functionality of the existing input and output interfaces be retained within a unified platform.

SABIO-RK currently contains approximately 75,000 entries, which means that the data are not large by computing standards, but they are complex and accurately curated.

Two external companies presented their ideas for implementing our requirements and ultimately made us an offer for the re-implementation of SABIO-RK, while a third company declined to make an offer for this complex endeavor.

We were pleased to contract the highly motivated company Klug & Milke for the re-implementation project, which built on our previous positive collaboration with their former company, klug newmedia (KNM), in which the company had managed the official website design and maintenance for the well-established consortium LiSyM-Cancer.

First, our external partners finalized their functional specification document, which detailed the system architecture, the new JSON-based data model for entries, the core functionalities (i.e., search, views, editing, workflow, roles, API), and the user concept. The subsequent implementation phase was characterized by a close and productive collaboration – including weekly online meetings and joint GitHub issue tracking – in order to meet all user requirements. After more than one year of intensive development and testing, a new, modern SABIO-RK was finally created with the features outlined below.

Open-Source Technology Stack

The new system is based entirely on open-source technologies with permissive licenses, thereby ensuring long-term maintainability and control for the SDBV Group. The backend – built with .NET 10.0, ASP.NET Core, and Entity Framework – handles data processing, business logic, and persistence, providing a type-safe interface to a relational PostgreSQL database. Apache Solr enables fast full-text searches across entries. The frontend – developed with SolidJS and TypeScript – provides an interactive web interface for browsing, searching, editing, and exporting curated data. The system is fully containerized with Docker, thereby ensuring reproducible deployment and easy maintenance.

Data Entry Page

The new interface for data input – with its clear dashboard on the left – guides the user smoothly through the curation workflow: First, the task management allows for the (bulk) import of PubMed text files and SBML files as well as for the assignment of tasks to student workers. The assigned tasks are automatically moved from the "Unassigned" to "Input Pending" status. After the data input – that is, the creation of database entries from a publication – has taken place, the tasks are automatically marked as “Pending Curation” and await both review by the curators and the processing of new entities by adding metadata such as ontology identifiers or links to other resources. Subsequently, the tasks are marked as "Completed Tasks - Not Public," which allows for a re-review of all entries – for example, from the same publication – before they are submitted to the public interface by clicking the "Publish" button. This automatically moves the task to the "Completed Tasks - Public" section.

Similar to the old SABIO-RK, the new system supports various user roles with different permissions, including regular users, editors, curators, administrators, and developers. Access to functions and resources can be controlled down to the level of individual users, operators, and records.

All functionalities that had been integrated into the old input interface over the years in order to simplify data entry and promote standardization were also implemented in the new interface and were often improved, such as selection lists and warnings when incorrect or incomplete information has been entered. Now, the curator is guided to the detail pages of newly added entities during the curation process, where they can immediately add additional metadata – such as In-ChI, SMILES, and PubChem IDs – to newly entered compounds. This has the significant advantage that the curator is already familiar with the topic and therefore does not have to delve into it a second time.

> The new system is based entirely on open-source technologies with permissive licenses, thereby ensuring long-term maintainability and control for the SDBV Group

A completely new feature is the option to amend already-published database entries – that is, to directly correct mistakes reported by users. For example, if a compound is subsequently corrected, all database entries containing this compound are automatically updated. This is not only significantly faster than the previous process of making changes at the database level, but above all, much safer because automation prevents manual errors. In addition, all data changes are logged, thereby providing a complete history, enabling version comparisons, and allowing the user to revert to previous states. Such version control was completely lacking in the old system.

The New Public Interface

First, the newly implemented "Browse" function provides a good initial overview of the available data, with the various attributes sorted into color-coded topic groups . The same color coding is maintained all over the input and user interface. Within an attribute group (e.g., organism), entities with the most entries are at the top. In order to get an overview of available data, for example, for a specific organism, you can click directly on it and be taken to its detail page (e.g., Homo sapiens) (Figure 41).

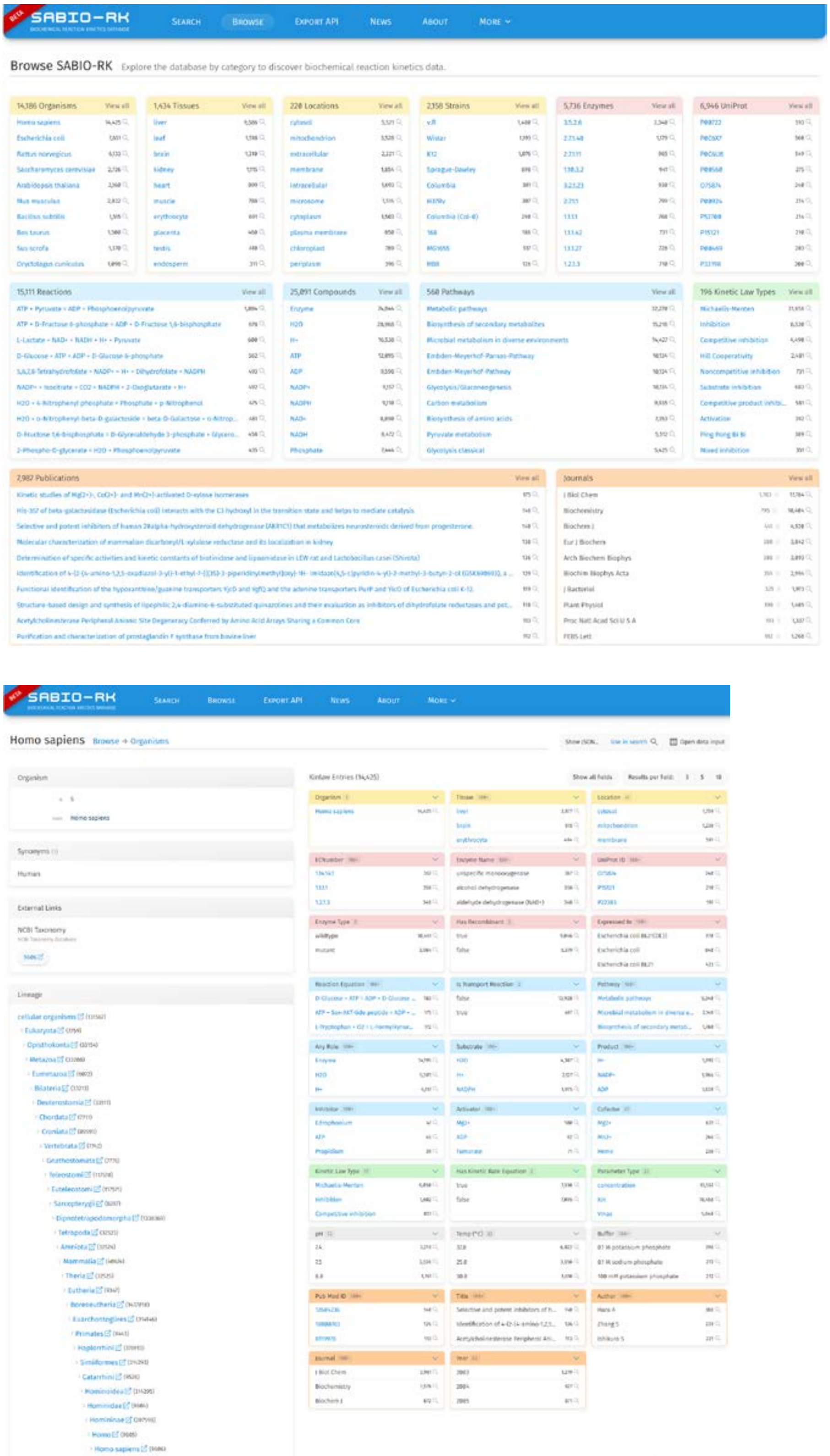
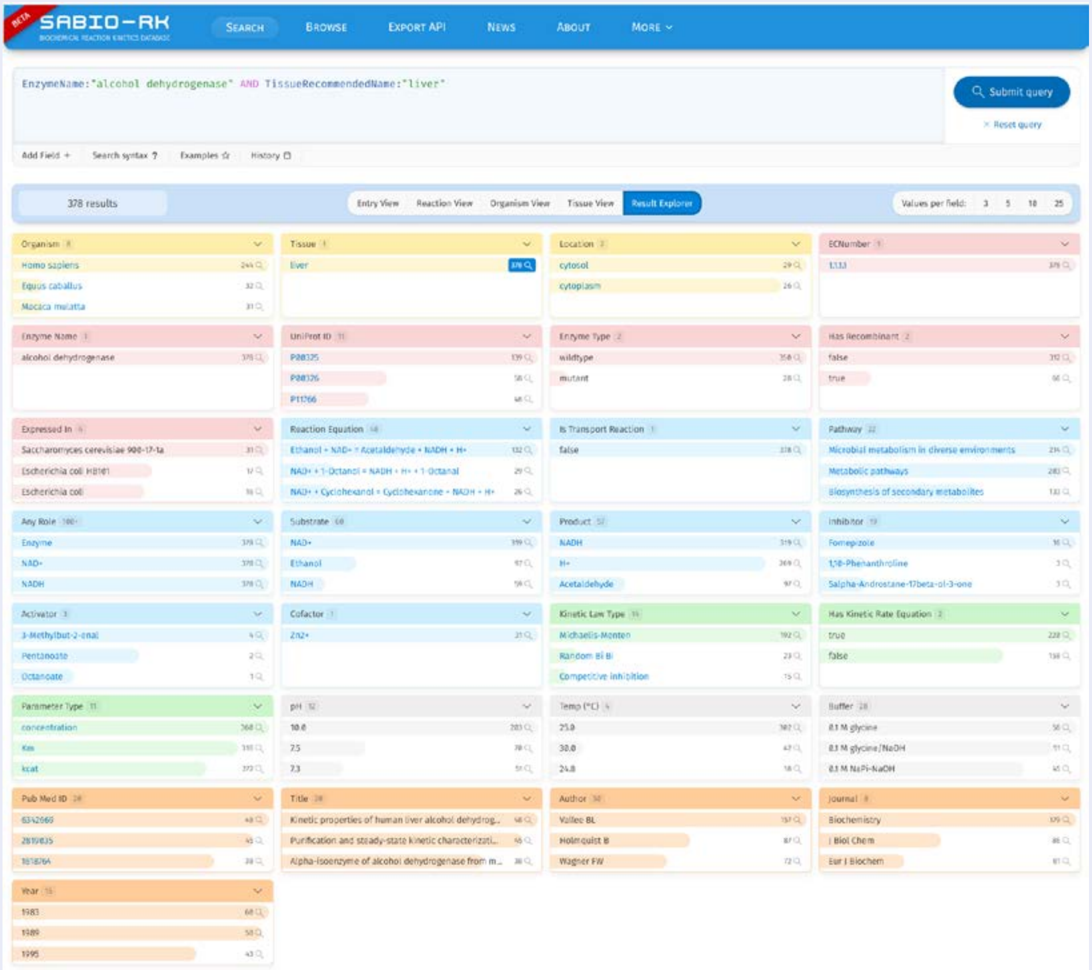


Figure 41
Top: "Browse" provides a simple initial overview of the SABIO-RK data. Clicking on a blue entity (e.g., "Homo sapiens") opens the corresponding details page (bottom).



EnzymeName:"alcohol dehydrogenase" AND TissueRecommendedName:"liver"

Figure 42
The "Result Explorer" facilitates the narrowing of the search by enabling the user to click on the number of entries next to a specific entity.

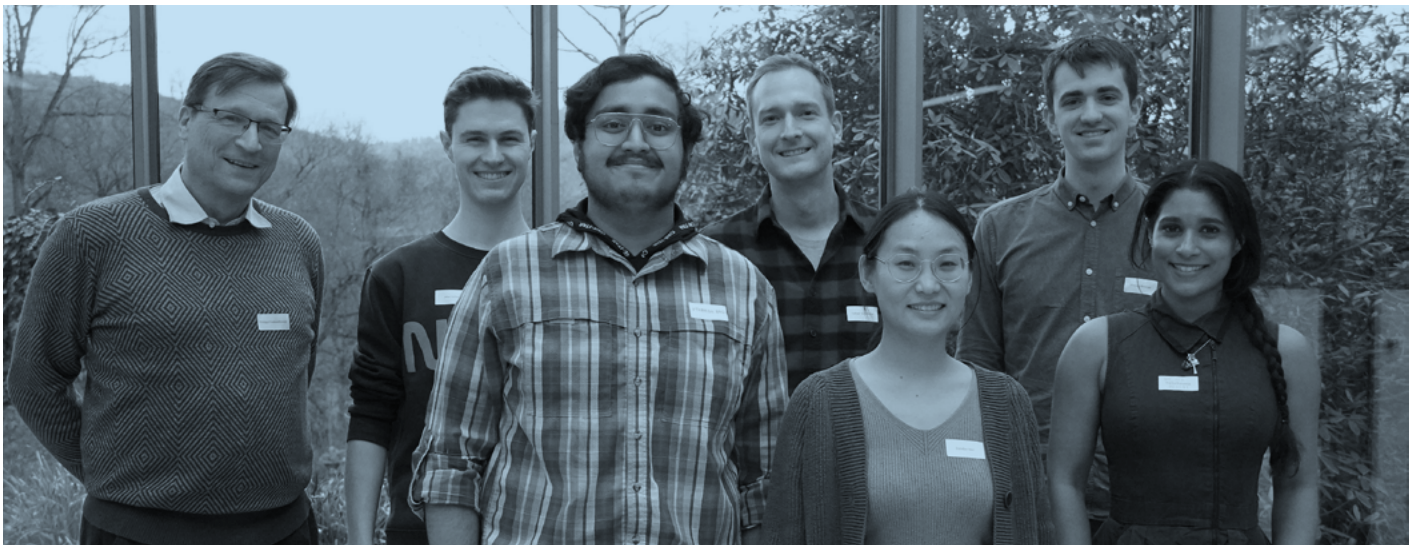
The "Search" function is very user-friendly as it offers several ways to refine a search. Similar to in the former user interface, both free text and advanced search are enabled. Examples of SOLR search syntax make it easier for users to get started with formulating search queries. Suggestion lists that appear while typing help the user to select or narrow the exact search term.

The search query history is new. Here, users can not only find the last 100 search queries under "Recent," but also define their own, individually named search queries that are "Saved" long-term. Exporting and importing search queries is additionally possible.

By default, the Search result is first shown in the "Results Explorer," which offers an easy way of refining a search without typing specific search terms. For example, if a user searches for EnzymeName:"alcohol dehydrogenase," they will immediately see how many entries exist in SABIO-RK for this enzyme in combination with specific organisms, tissues, reactions, etc. By clicking on a number next to a specific entity (e.g., liver), the user can easily further narrow their query results and repeat this process until they have included all their search terms (Figure 42).

As in the old SABIO-RK, the query results can also be displayed in several table views. The Entry View lists all database entries matching the search and allows the user to open the single database entries or to select entries for export in JSON, SBML, or other SBFC formats. In these table views again, reaction, EC number, organism, and tissue are directly linked to their corresponding detail pages.

As before, SABIO-RK can be accessed programmatically via a redesigned API that is based on REST principles with corresponding HTTP methods and status codes. In order to ensure security and performance, the request rate is limited (60 requests / minute / IP address), and IP-based partitioning prevents denial-of-service attacks. ApiLab – which provides a standalone OpenAPI explorer and a test user interface – was used for testing.



Research

2.10 SET

Stellar Evolution Theory

Group Leader

Dr. Fabian Schneider

Team

Utkarsh Basu
(master's student)

Vincent Bronner
(PhD student)

Max Marius Heller
(student assistant; until February 2025)

Dr. Jan Henneco
(PhD student; until February 2025)

Utkarsh Jain
(master's student; since November 2025)

Dr. Rajika Kuruwita
(HITS Independent Postdoc; until September 2025))

Qingxuan Li
(master's student; since December 2025)

Aina Sophie Nambena
(student assistant; March & May 2025)

Prof. Dr. Philipp Podsiadlowski
(visiting scientist)

Pranavandhan Upendranath
(master's student; since April 2025)

Rui Zhan
(master's student; since October 2025)

Stars are the basic building blocks of the visible Universe and produce almost all the chemical elements heavier than helium. Understanding how stars have transformed the pristine Universe into the one we live in today lies at the heart of astrophysics research.

Massive stars are cosmic powerhouses. They can be several million times more luminous than the Sun and explode as powerful supernovae. They helped to re-illuminate the Universe after the Cosmic Dark Ages and leave behind some of the most exotic forms of matter: neutron stars and black holes. Mergers of such objects are now routinely observed thanks to gravitational-wave observatories, thereby opening a new window into the Universe.

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 form highly magnetized neutron stars during their terminal supernova explosions. These magnetic neutron stars – called magnetars – are the strongest-known magnets in the Universe.

Sterne sind die elementaren Bausteine des sichtbaren Universums und produzieren nahezu alle chemischen Elemente, die schwerer als Helium sind. Zu verstehen, wie Sterne das Universum nach dem Urknall in seine heutige Form verwandelt haben, steht seit jeher im Zentrum der astrophysikalischen Forschung.

Massereiche Sterne sind kosmische Kraftwerke. Sie können mehrere Millionen Mal heller sein als die Sonne und explodieren in gewaltigen Supernovae. Dank dieser Eigenschaften haben sie dazu beigetragen, unser Universum nach den kosmischen „Dark Ages" wieder zu erhellen. Sie hinterlassen außerdem einige der exotischsten Formen von Materie: Neutronensterne und Schwarze Löcher. Verschmelzungen solcher Objekte werden heute dank Gravitationswellen-Observatorien routinemäßig beobachtet und eröffnen dadurch ein neues Fenster zum Universum.

Die Stellar Evolution Theory (SET) Gruppe untersucht das turbulente und explosive Leben massereicher Sterne. Im Fokus stehen massereiche Doppelsternsysteme, deren Verschmelzungsprozesse und die Frage, welche Sterne Schwarze Löcher bilden. Sternverschmelzungen erzeugen starke Magnetfelder und können zu besonders stark magnetisierten Neutronensternen führen. Diese als Magnetare bekannten Neutronensterne sind die stärksten Magnete im Universum.

Group News

In February 2025, our very first doctoral student, Jan Henneco, graduated with magna cum laude from Heidelberg University and has since moved to Newcastle University for postdoctoral research. We wish Jan all the best on his next steps and look forward to continuing our fruitful collaboration. Jan's leaving signaled that the ERC Starting Grant – which mostly funds the SET group – is approaching its planned end in October 2026. A new generation of master's students joined in the autumn, and we look forward to working with Utkarsh Jain, Qingxuan Li, and Rui Zhan on the structures of the progenitors of core collapse supernovae, black hole formation in massive stars, and stellar mergers. In July, the SET group was evaluated extremely positively and will thus be extended until the end of 2027.

Black Holes and Their Gravitational Wave Signals

Massive stars – that is, those more than ten times the mass of our Sun – live fast and die young. They form some of the most exotic and enigmatic objects in the Universe when they collapse at the end of their lives: neutron stars and black holes. But exactly how massive stars evolve and what determines the properties of the compact objects they produce depends, for example, on many detailed aspects of stellar and nuclear physics as well as on supernova theory. Understanding these processes has become a major topic and driving force in physics thanks to the detection of gravitational waves from the mergers of neutron stars and black holes.

Studying the evolution of massive single and binary stars toward the end of their lives and understanding the likely outcome of their core collapse is a central theme of the research that is conducted in the SET group. Over the last decades, a non-monotonic landscape has emerged in which a star's explodability and ultimate fate are determined not by its mass, but by other properties of the star's cores. In Figure 43 [Laplace et al, 2025], some of the commonly used summary variables of the core structures that indicate the ultimate fates of stars are shown as a function of the stars' initial mass, Mi: core compactness $\xi_{2.5}$, central entropy s_c, iron core mass $M_{\text{Fe-core}}$, and envelope binding energy E_B. Generally, sufficiently large values indicate black hole formation, while low values suggest successful supernova explosions and neutron-star formation. In [Laplace et al, 2025], we finally revealed the physical origin of this non-monotonic landscape. The features in Figure 43 are due to a complex interplay between nuclear fusion, gravitational contraction, and neutrino losses during the advanced core and shell burning phases of carbon and heavier elements.

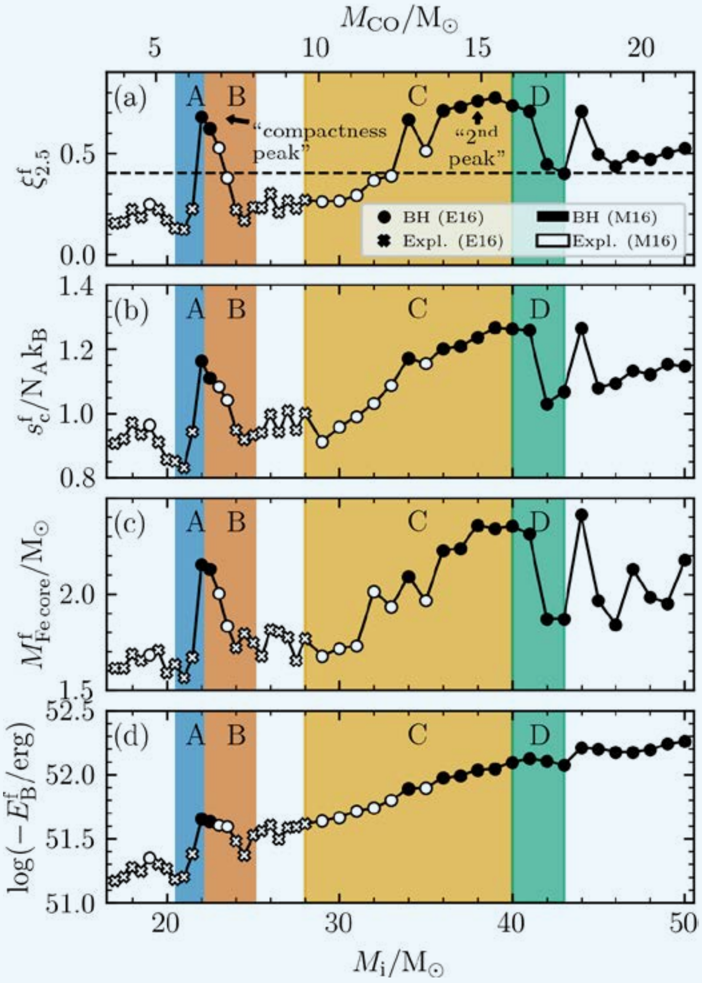


Figure 43
Core compactness $\xi_{2.5}$, central entropy sc, iron core mass $M_{\text{Fe-core}}$, and envelope binding energy E_B as a function of initial mass M_i and carbon–oxygen core mass M_{CO} for massive single stars. Two explosion models – E16 and M16 – were applied to the models in order to differentiate between likely supernova explosions and black hole formation.

Thanks to these insights and a large library of single and binary star evolution models computed in the SET group, HITS Lab doctoral student Kiril Maltsev developed a simple prescription for forecasting the likelihood of black hole formation in single and binary-stripped stars [Maltsev et al, 2025]. Binary-stripped stars have lost their outer envelopes in a mass-transfer event and are thought to be the progenitors of the black holes seen in gravitational wave merger events. The new prescriptions enabled us to implement our findings into population synthesis models in order to predict the properties of binary black hole mergers, which are nowadays routinely observed via their gravitational wave emission.

To that end, we teamed up with postdoc Reinhold Willcox from KU Leuven, Belgium, who implemented our new prescription for black hole formation in stars and made population synthesis predictions for the distribution of the so-called chirp masses of binary black hole mergers ([Willcox et al, 2025], see Figure 44, next page). The chirp mass – $M_S = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ – is the combination of the masses m_1 and m_2 of the two merging compact objects and can be directly inferred from the gravitational-wave merger signal. It is one of the best-measured quantities in gravitational-wave astronomy for $M_S \lesssim 30$ solar masses. As our new black hole formation model predicts a bimodal structure that is preserved through cosmic history (i.e., redshift z and metallicity Z), the chirp mass distribution is predicted to exhibit three characteristic features (Figure 44) at chirp masses of about 9, 13, and >15 solar masses.

> The features would directly constrain many uncertain processes in stellar, nuclear, neutrino, and supernova physics

At the end of August 2025, the LIGO-VIRGO-KAGRA collaboration released the first results of their fourth observing run (O4a), which roughly doubled the number of gravitational-wave merger events. We were thrilled to find that the observed events display features in the chirp mass distribution of binary black hole mergers that are qualitatively in agreement with our predictions. Given the current statistics, the observed features are not yet significant at a 5σ limit and are thus not designated a definite detection (i.e., there is a greater than 1-in-3.5-million chance that the features are random fluctuations). Future data releases will settle this question, and we are eagerly awaiting them. If the features continue to persist and can indeed be linked to our prediction of a bimodal black hole mass distribution from binary-stripped stars, this will have far-reaching implications. The features would directly constrain many uncertain processes in stellar, nuclear, neutrino, and supernova physics. They could even be used as “standard sirens” for precision cosmology in order to measure cosmic expansion and structure formation.

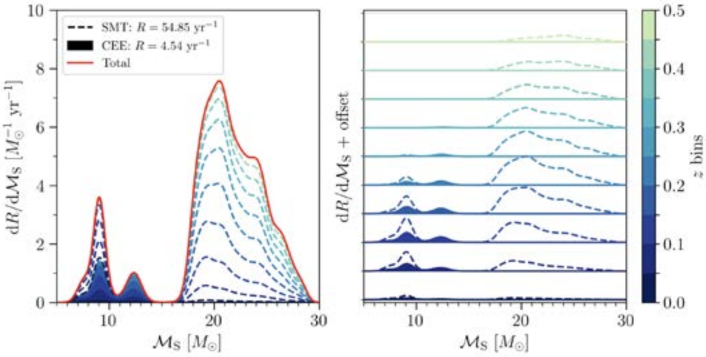


Figure 44
Predicted chirp mass – M_S – distribution of binary black hole mergers using our new formation model of black holes. The distributions are displayed as a function of redshift z in order to show the universality of the main features across cosmic evolution. Figure taken from [Willcox et al, 2025].

Binarity at LOw Metallicity (BLOeM)

The BLOeM (Binarity at LOw Metallicity) project is a major observational survey that uses the Very Large Telescope to monitor nearly 1,000 massive stars in the Small Magellanic Cloud (Figure 45) – a nearby galaxy with low heavy-element content similar to that of the early Universe. Its goal is to measure how common binary and multiple star systems are in such metal-poor environments and to determine the orbital properties of these massive stars – including finding systems with inert black holes – by collecting 25 epochs of spectroscopy. This will create a rich dataset revealing how massive stars form, evolve, and interact under conditions unlike those in our own Milky Way, thereby improving our understanding of both stellar evolution and the origins of exotic objects such as black hole binaries.

The SET group is an integral part of this project. BLOeM is the successor program of a similar observing campaign in the Large Magellanic Cloud called the VLT-FLAMES Tarantula Survey (VFTS). BLOeM also uses the FLAMES spectrograph on the Very Large Telescope of the European Southern Observatory in Chile, which is one of the world’s largest telescopes. FLAMES has 132 fiber optics, each of which can be directed at a different star and then be observed simultaneously. Several group members are directly involved in this project, and we have just published the first big data release and accompanying papers. For example, doctoral student Vincent Bronner of the SET group obtained the first estimates of the initial masses and ages of the target stars (published in [Bestenlehner et al, 2025]), which will be crucial both for many of the upcoming studies and for making sense of this treasure trove of data.

Most importantly, we have already been able to demonstrate that massive stars in this metal-poor environment commonly live in close pairs, much like their counterparts in our own Milky Way galaxy. By measuring changes in the stars’ velocities over time, we found that more than 70% show signs of acceleration and deceleration caused by the gravitational tug of a nearby companion, thereby indicating that binary systems are common even in

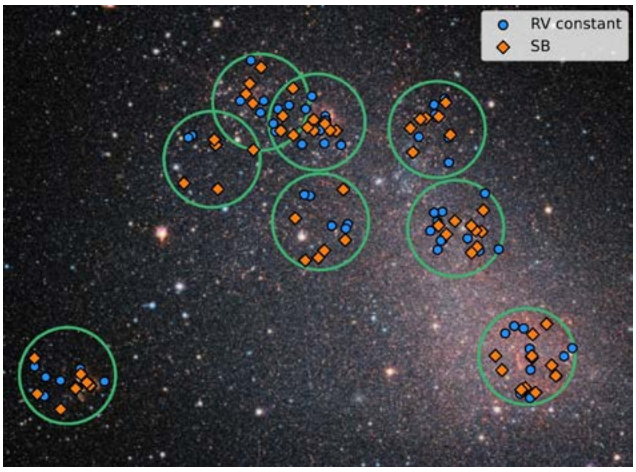


Figure 45
Massive stars in the Small Magellanic Cloud. Of the stars studied, 70% (i.e., the red diamonds) appear to accelerate and decelerate. This indicates the presence of a partner. Image credit: ESO/Sana et al.

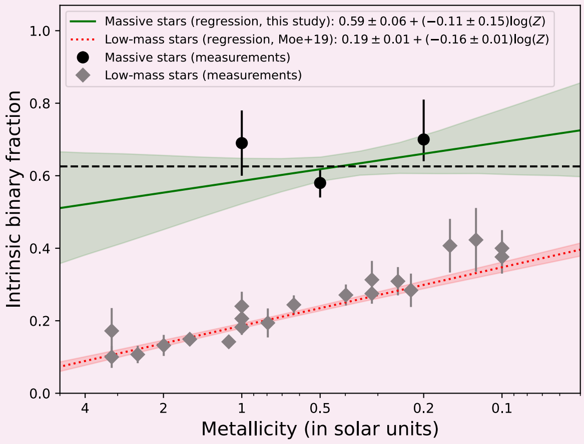


Figure 46
Observed binary fractions in low- and high-mass stars at different metallicities. The high-mass data points at metallicities of 0.5 and 0.2 of the Sun were determined by the VFTS and BLOeM projects, respectively.

metal-poor galaxies (Figure 46). This study was published in the journal Nature Astronomy [Sana et al, 2025].

This result is highly significant because the Small Magellanic Cloud's chemical makeup resembles that of galaxies from when the Universe was young; therefore, the finding suggests that massive stars early in cosmic history also likely had close partners. The presence of these binary companions can strongly affect how stars evolve and die, thereby impacting supernova explosions and the formation of exotic remnants such as black holes as well as paving the way to understanding the observed gravitational wave sources described above.

All observations within the BLOeM project have now been taken, and the collaboration will enter a new phase of analyzing the data over the next few years in order to precisely map the stars' orbits, determine the orbital parameters and masses of the stars, and better understand how such binary systems form and evolve in such a low-metallicity region. If lucky, we may also discover a hidden, inert black hole or other exotica.

> The presence of these binary companions can strongly affect how stars evolve and die

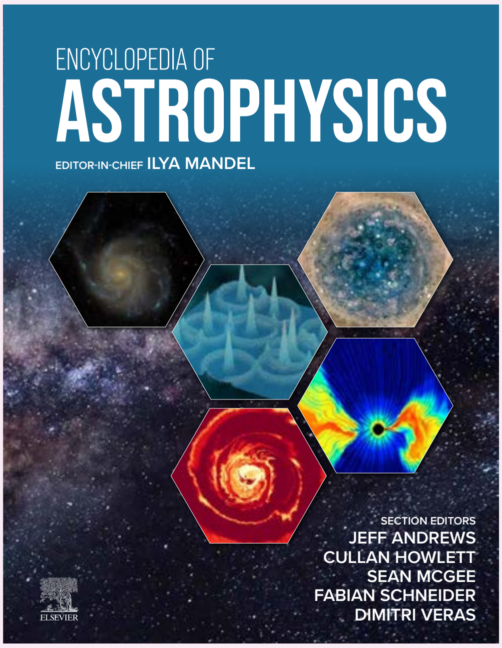


Figure 47
Cover of the Encyclopedia of Astrophysics.

Encyclopedia of Astrophysics

Since 2023, group leader Fabian Schneider has been the editor of the "Stars" section of the five-volume Encyclopedia of Astrophysics (Figure 47), which was published by Elsevier as a Major Reference Work in December 2025. The Encyclopedia of Astrophysics is a comprehensive, authoritative reference that brings together expert-written articles covering the full scope of modern astrophysics ranging from cosmology, galaxies, and stars to planetary science, high-energy phenomena, observational techniques, and theoretical foundations. Designed as an introductory text and scholarly resource, the work provides clear, self-contained entries that summarize key concepts, methods, and discoveries, thereby making it useful for students, researchers, and professionals seeking reliable overviews and context across the discipline. Several HITS researchers contributed chapters. For instance, former HITS Independent Postdoc Rajika Kuruwita wrote a chapter on star formation, while Fabian Schneider, Mike Lau, and Friedrich Röpke contributed a chapter on stellar mergers and common-envelope evolution.

Research

2.11 TOS

Theory and Observations of Stars

Group Leader

Prof. Dr. ir. Saskia Hekker

Team

Dr. Felix Ahlborn
(postdoc)

Dr. Michaël Bazot
(staff scientist)

Beatriz Bordadágua
(PhD student)

Bjela Böttcher
(bachelor's student; until September 2025)

Dr. Lynn Buchele
(PhD student; until July 2025)

Jeong Yun Choi
(PhD student)

Dr. Quentin Coppée
(PhD student; until October 2025)

Lucas Eekhof
(master's student; until July 2025)

Francisca Espinoza-Rojas
(PhD student)

Jonas Granzow
(bachelor's student; until July 2025)

Dr. Jan Henneco
(PhD student; until February 2025)

Gregory Jung
(master's student; since May 2025)

Inmaculada Moyano-Rejano
(PhD student; since October 2025)

Jonas Müller
(PhD student)

Emeric Poiraud
(bachelor's student; until August 2025)

Adrian Thessmann
(bachelor's student; until August 2025)

Dr. Jordan Van Beek
(postdoc)

Tobias van Lier
(PhD student; since December 2025)

Kristin Weidner
(team assistant)

Stars are an important source of electromagnetic radiation in the Universe that allow for studies of many phenomena ranging from distant galaxies to extra-solar planets. However, due to their opacity, it was once said that "at first sight it would seem that the deep interior of the Sun and stars is less accessible to scientific investigation than any other region of the universe" (Sir Arthur Eddington, 1926). Now, through modern mathematical techniques and high-quality data, it has become possible to probe and study the internal stellar structure directly through global stellar oscillations: a method known as asteroseismology.

In asteroseismology, the properties of waves are used to trace the internal conditions of stars. Oscillations that impact upon the whole star reveal information that is hidden by the opaque surface. This asteroseismic information provides insights into both the stellar structure and the physical processes that take place in stars.

Understanding the physical processes that take place in stars and how these processes change as a function of stellar evolution is the ultimate goal of the Theory and Observations of Stars (TOS) group at HITS.

Sterne sind eine wichtige Quelle elektromagnetischer Strahlung im Universum, mit der viele Phänomene untersucht werden können, von fernen Galaxien 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 die Eigenschaften von Wellen, 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 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 ist die Untersuchung dieser physikalischen Prozesse, die in Sternen ablaufen, und wie sich diese in Abhängigkeit von der Sternentwicklung verändern.

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 through turbulence in the near-surface convection layer of a star. These oscillations are sound waves that are expected to be present in all stars with convective outer layers. A convective envelope is typically present in low-mass main-sequence stars, subgiants, and red-giant stars with surface temperatures below ~6,700 °K.

The stellar structure is imprinted in the global oscillation modes of a star. An oscillation mode is uniquely determined by the properties of the matter through which it travels and is described by its frequency (or period) and mode identification – that is, by its radial order (i.e., the number of nodal lines in the radial direction), its spherical degree (i.e., the number of nodal lines on the surface), and its azimuthal order (i.e., the number of nodal lines that cross the spin axis). 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 spherical degree of 1) have sensitivity to both the deep interior and the outer layers – that is, the oscillations resonate in an inner (gravity) and an outer (acoustic) cavity separated by an evanescent zone (i.e., the area between the cavities where oscillations cannot propagate and decay exponentially). From the resulting mixed pressure–gravity oscillations, the coupling between the two oscillating cavities and the phases of the waves in each cavity can be derived and provide information on the physical conditions in the evanescent region. Furthermore, the difference in period between pure gravity dipole modes with consecutive radial orders (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.

Scientific Highlights:

The Angular Momentum Transport of Red-Giant Stars

Red-giant stars are stars that have exhausted the hydrogen in their cores. This results both in a dense helium core surrounded by a thin shell in which hydrogen burning occurs and in an extended hydrogen envelope. As the gravitational force is no longer balanced by nuclear burning pressure, the core contracts over time. Assuming conservation of angular momentum, the core spins up while the expanding envelope slows down.

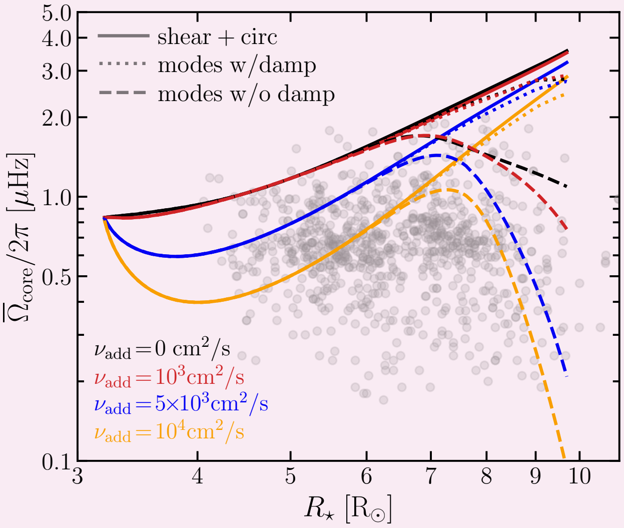


Figure 48

Core rotation rate as a function of the stellar radius for a 1.3-solar-mass stellar evolution track. The different colors indicate models computed with different viscosity values for an additional unknown diffusion transport process. The grey circles indicate the observed core rotation rates from Gehan, C., Mosser, B., Michel, E., Samadi, R., and Kallinger, T., "Core rotation braking on the red giant branch for various mass ranges," A&A 616, A24 (2018).

> When oscillations from two stars are detected in a single light curve, we refer to them as asteroseismic binaries

Measurements of core rotation rates obtained through asteroseismic observations conflict with our current predictions – that is, the cores of red giants appear to rotate two orders of magnitude more slowly than expected. This suggests the existence of a physical mechanism that counteracts the spin-up due to core contraction and redistributes angular momentum to the envelope. Several mechanisms have been proposed to explain how angular momentum is transported in these stars; however, there is currently no general consensus.

In the work "The efficiency of mixed modes for angular momentum transport" [Bordadágua et al, 2025], we incorporated a mechanism into stellar models that we refer to as the "mixed-mode mechanism." This mechanism relies on the mixed pressure–gravity modes themselves – which propagate in both the core and the envelope of these stars – in order to transport angular momentum from the core to the outer layers.

Contrary to predictions, our results showcase a localized spin-down in the hydrogen-burning shell that is not efficiently redistributed by known hydrodynamic transport processes (see the black solid track labeled shear + circ in Figure 48). Hence, an additional transport mechanism must be considered. For example, a diffusion process with sufficiently high viscosity would be enough to redistribute the localized spin-down (see the orange solid track in Figure 48). Furthermore, the efficiency of the mixed-mode mechanism is significantly reduced when radiative losses in the core are accounted for during evolution (see the dotted tracks labeled modes w/damp in Figure 48). This highlights the importance of not neglecting these losses in future work. These results ultimately lead to the conclusion that the mixed-mode mechanism alone is insufficient to explain the observed slowdown of red-giant cores.

An Analytical Expression for Damped Mixed Modes

The oscillations in red-giant stars (and solar-like oscillators in general) are both excited and damped. The observable amplitude of the modes is determined by a balance between the excitation and damping processes that occur in the star. The efficiency of these processes is not uniform and is usually limited to a certain region of the star. At the stellar surface, the interaction of oscillations with convection can act both as a source of stochastic excitation (see background) and as a source of damping. In the stellar core, heat losses due to fluctuations in the temperature gradient dampen the mixed-mode oscillations. In addition to these processes, observational evidence suggests the existence of further damping processes in the core that are orders of magnitude more efficient. The physical mechanism responsible for this strong damping is still being intensively investigated and could be caused by a strong internal magnetic field.

In order to compare the theoretical efficiency of damping processes with the relative amplitudes of the observed oscillation modes, a tool is needed that links the two. Thus far, this has only been possible under the assumption that the damping processes are not very efficient, which is not the case for a significant subset of red-giant stars. The reason for this is that these stars exhibit mixed oscillation modes with very low amplitudes, which means that they are strongly damped. In an article accepted by Astronomy & Astrophysics [Müller et al. “Asymptotic power spectra and visibilities of damped mixed modes,” (2026)], we developed an analytical approach that links the observable properties of the oscillations with the theoretical efficiencies of damping processes. This method is a flexible and easy-to-use tool as it does not require the additional use of numerical stellar evolution or oscillation codes.

Our approach makes it possible to predict the observational signatures caused by different damping processes and to compare them with actual observations. We have found that stars with a strong damping process in their cores can appear in observations as if the efficiency of their corresponding damping process were infinite, which has been observed in a number of stars. Furthermore, we have been able to quantify the efficiency of the damping process in the cores of stars with very low multipole mode amplitudes, which can now be compared with theoretical estimates. As a next step, we will apply our method to numerical stellar models in order to derive the relative amplitudes of the oscillation modes during the evolution of the stars.

Red-Giant Asteroseismic Binaries

In the Universe, binary stars are as common as single stars. Both components of a binary share a common formation history, thereby leading to the same initial chemical composition and age. This fact allows us to constrain stellar model inputs and explore diverse evolutionary paths depending on mass, luminosity, and the interaction of the two elements. Moreover, spectroscopic observations and light curve analysis of eclipsing systems enable us to directly measure the masses and radii of the binary components through dynamical modeling.

Asteroseismology has become a potent method for inferring stellar parameters through scaling relations, particularly for stars with a convective surface layer. By observing their solar-like oscillations, we can gain insights into the internal structure of stars. From the Kepler mission, solar-like oscillations have been detected in about 30,000 red giants.

Among the identified binary systems with oscillating components, solar-like oscillators are particularly valuable. The dynamical masses derived from orbital analysis can constrain the asteroseismic scaling relation and further improve stellar modeling. When oscillations from two stars are detected in a single light curve (see Figure 49), we refer to them as asteroseismic binaries (ABs), regardless as to whether the two components are gravitationally

> Measurements of core rotation rates obtained through asteroseismic observations conflict with our current predictions – that is, the cores of red giants appear to rotate two orders of magnitude more slowly than expected

bound. However, only a few asteroseismic binaries have been discovered in which both stars exhibit solar-like oscillations.

In Kepler's long-cadence data, several studies(e.g., [Miglio, A., Chaplin, W. J., Farmer, R., et al: Prospects for detecting asteroseismic binaries, in Kepler Data. ApJ, 784, L3, 2014]) predicted the presence of at least 200 ABs. These ABs were expected to be gravitationally bound red giants that both exhibit detectable oscillations. However, [Espinoza-Rojas et al, 2025] presented 40 red-giant ABs that show two distinct power excesses in their power density spectra (PDS) and found only two wide-binary candidates, with most cases being chance alignments.

In order to distinguish between wide-binary candidates and chance alignments, we used Gaia DR3 astrometry and binarity indicators to check Keplerian constraints on relative stellar motions. This approach combined Gaia positions, parallaxes, proper motions, and – when available – radial velocities in order to estimate the total orbital velocity difference between pairs of stars. Gravitationally bound systems were expected to exhibit similar kinematics on the sky, whereas large velocity discrepancies above a defined threshold were expected to indicate chance alignments. For the two identified wide-binary candidates, we revealed that the components of these systems share consistent evolutionary stages and asteroseismic mass estimates, in line with a common formation history.

While we focused in [Espinoza-Rojas et al, 2025] on stars in which the oscillations of the two stars in the AB were present at different frequencies (see Figure 49), it may actually be more likely to find ABs with similar frequencies. This is a result of stellar evolution, with stars in their core-helium burning phase being relatively long lived and exhibiting oscillations in a similar frequency range, with their frequency of maximum oscillation power (vmax) typically being around 20–50 µHz. ABs consisting of two stars with oscillations in the same frequency range may show oscillation patterns that look different from those of single stars.

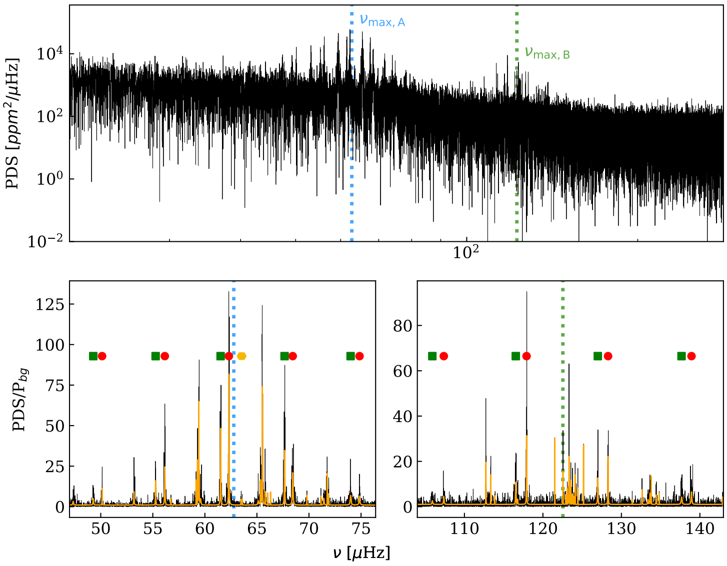


Figure 49

Power density spectrum and peak-bagging results of KIC 6501237. In the top panel, the dotted blue and green lines indicate the frequency of maximum oscillation power (vmax) for Star A and Star B, respectively. The lower panels display the fitted oscillation modes of both solar-like oscillators in orange. Radial, quadrupole, and octupole modes are marked by red circles, green squares, and yellow hexagons.

Due to the difficulties in identifying these systems, the characteristics of red-giant ABs with overlapping oscillation patterns remain unexplored. Choi et al. 2025 created 5,000 artificial asteroseismic binary (AAB) systems from two red giants oscillating in similar frequency ranges. The study used Shannon entropy to quantify the complexity of the PDS shape and combined it with the signal-to-noise (S/N) ratio to investigate different PDS morphologies. The resulting variety of PDS morphologies can be considered a template for identifying potential asteroseismic binary candidates in the observed data.

Indeed, among 5,000 AABs, most consisted of two core helium burning stars and exhibited increased Shannon entropy and decreased maximum S/N compared with their individual components. PDS morphologies of AABs in which the component stars had similar brightness displayed complex oscillation patterns (see the more complex pattern in the bottom panel of Figure 50 compared with in the two top panels). These results suggest that observed stars exhibiting unusually complex oscillation patterns with low oscillation power could be explained by these asteroseismic binaries. Furthermore, these results will be useful for mining the large datasets provided by modern space-based photometry – such as Kepler, TESS, and the forthcoming PLATO mission – in order to detect unresolved binary systems with overlapping oscillation power excesses.

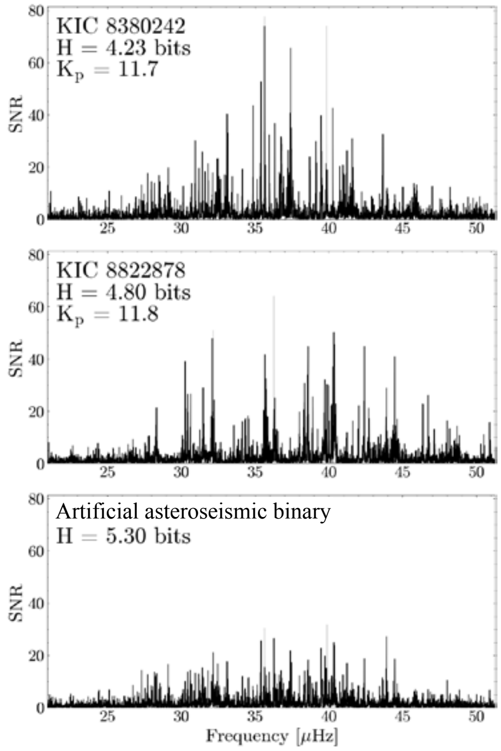


Figure 50

PDS of two core helium–burning stars in the top two panels and an artificial asteroseismic binary (AAB) in the bottom panel. We indicate Kepler Input Catalog (KIC) IDs, Shannon entropy (H), and Kepler magnitude (Kp) in the legend.

2.12 HITS Independent Postdoc Research

HITS Independent Postdocs

Dr. Vagrant Gautam

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.

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.

Since 2022, there have been three HITS Independent Postdocs: Astrophysicist Rajika Kuruwita (Sri Lanka; 2022–2025), biologist Fabian Grünewald (Germany; since 2023), and computational linguist Vagrant Gautam (Canada; since 2025).

*Das HITS-Independent-Postdoc-Programm bietet Doktorand*innen eine großartige Chance beim Übergang zur Leitung von Nachwuchsgruppen. Es unterstützt junge Forschende dabei, eigene ambitionierte Ideen zu erforschen und neue Hypothesen zu testen. Die ausgewählten Postdocs arbeiten mit HITS-Gruppen zusammen, während sie ihre unabhängigen Forschungsprojekte entwickeln und verfolgen.*

*Das Programm läuft zwei Jahre, mit der Option auf eine einjährige Verlängerung nach positiver Evaluierung. 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.*

Seit 2022 kamen bislang drei junge Forschende ans HITS: Die Astrophysikerin Rajika Kuruwita (Sri Lanka, 2022-2025), der Biologe Fabian Grünewald (Deutschland, seit 2023) und Vagrant Gautam (Kanada, Computerlinguistik, seit 2025).

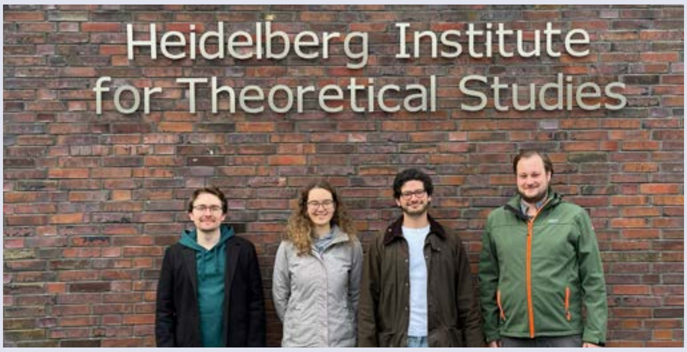
Fabian Grünewald: Accelerating the Design of Nanomedicine Technologies Through Simulations

Fabian Grünewald develops and applies simulation and machine learning methods in order to uncover the molecular organization and function of nanomedicine technologies. Nanomedicine has the potential to transform modern healthcare, as exemplified by the unprecedented impact of mRNA vaccines during the COVID-19 pandemic. In order to further push the boundaries of nanomedicine, Fabian's research reveals how these devices engage with cellular components on the molecular scale. In particular, he focuses on hybrid membranes and polymer nanoparticles for the delivery of therapeutics.

> Fabian's research reveals how nanomedical devices engage with cellular components on the molecular scale

Hybrid polymer–lipid membranes (HM) are formed by the co-self-assembly of amphiphilic block copolymers and lipid molecules, which are an integral part of natural cell membranes. Their enhanced stability and ability to functionally reconstitute natural membrane proteins make them a versatile platform for nanotechnology and synthetic biology. Previous work by Fabian in collaboration with the group of Prof. Maglia (University of Groningen) revealed that the nanoscale organization of these membranes is key to their successful functioning. Expanding on this work, Fabian and his student Clara (center left) conducted a systematic study of proteins and different membranes, revealing how phase behavior and defect formation cooperatively remodel hybrid membranes, thereby regulating local composition and protein interaction landscapes. This mechanistic framework explains why HMs robustly reconstitute integral membrane proteins and provides design principles for constructing synthetic membranes that couple long-term stability with functional biomimicry.

Furthermore, Fabian has continued his collaboration with the (previous) HITS MBM group, resulting in one publication that presents a coarse-grained simulation model of collagen. Part of the collaboration on data-efficient physics-informed machine learning models with the MLI group was published in 2025. Together with two bachelor's students – Sebastian Segnitz and Maximilian Fidlin – he is currently investigating how polymer nanoparticles for gene therapy form and interact with cell membranes.



Fabian and bachelor's students from left to right: Sebastian Segnitz, Clara Schlitter, Maximilian Fidlin, Dr. Fabian Grünewald.

Fabian continues to serve as a core member of the Martini Force Field Initiative (https://www.cgmartini.nl), which develops, maintains, and oversees the widely used Martini coarse-grained simulation method. In this capacity, he contributed lectures and tutorials at the 2025 Martini Workshop in Groningen (NL).

In 2025, Fabian lectured again in the Molecular Biotechnology study program at Heidelberg University.

Vagrant Gautam: Trustworthy Natural Language Processing

Vagrant Gautam works on social and technical aspects of trustworthy natural language processing (NLP) – that is, they are interested in developing NLP systems that process everyone's language fairly and that are faithful to both context and user needs. Large language models are one type of NLP system that has been rapidly and widely adopted but that remains poorly understood. Trained with massive volumes of data, such models pick up, replicate, and even amplify stereotypes and harmful social biases while simultaneously showing remarkable generalization to novel tasks and linguistic phenomena. At HITS, Vagrant will investigate the interplay of such biases and generalization in new contexts, including in the retrieval of factual knowledge and pronominal reasoning. They will additionally examine the broader societal context and ethical implications of the current language-model-focused paradigm of NLP, which is now led by major players in industry rather than by universities and independent research institutions.

> They will examine the broader societal context and ethical implications of the current language-model-focused paradigm of NLP

After joining HITS in September 2025, Vagrant gave talks at the University of Edinburgh and Heidelberg University. At the 2025 Conference on Language Modeling in Montreal, they presented work that had been completed during their PhD on evaluating misgendering with large language models, and they co-organized the first Workshop on the Interplay of Model Behavior and Model Internals. In November, they moderated the panel NLP and Its Impact on Society at the Future of NLP Symposium. Beyond their research pursuits, they were also invited to speak on a panel about Navigating the STEM Landscape as a QTIPOC Professional, and they co-organized the Queer in AI social associated with the Conference on Language Modeling.

3 Centralized Services

Group Leader

Dr. Gesa Schönberger

Team

Yashasvini Balachandra
(Controlling)

Christina Blach
(Front Office)

Frauke Bley
(Human Resources)

Benedicta Frech
(Front Office; until December 2025)

Silvia Galbusera
(Human Resources)

Harald Haas
(Controlling; Team Lead Grant Administration)

Jessica Herbert
(Accounting)

Ingrid Kräling
(Controlling)

Dr. Barbara Port
(Scientific Manager)

Rebekka Riehl
(Assistant to the Managing Director; Human Resources; Team Lead Front Office)

Darija Stein
(Front Office; since October 2025)

Jason Vay-Disterhöft
(Team Assistant Human Resources)

Irina Zaichenko
(Accounting)

3.1 Administration

The HITS Administration team manages key administrative functions, such as personnel management, office and building operations, procurement, and financial oversight. It ensures that all institute procedures comply with legal standards and that legal issues are handled appropriately. In addition, team members join the Events team to coordinate scientific conferences, workshops, and public outreach activities.

International orientation, mutual respect, family-friendly policies, and diversity are cornerstones of our institute and are essential for excellent science. In addition, HITS is characterized by a constant stream of people coming and going due to the large number of temporary contracts on various projects, many of which are funded by third parties.

In 2025, 63 people joined the institute, while 59 people left, thereby bringing the total to about 160 people. The international character of HITS is especially pronounced, with 31 nationalities represented (as of December 2025). This situation requires consistent and reliable support from our Human Resources team.

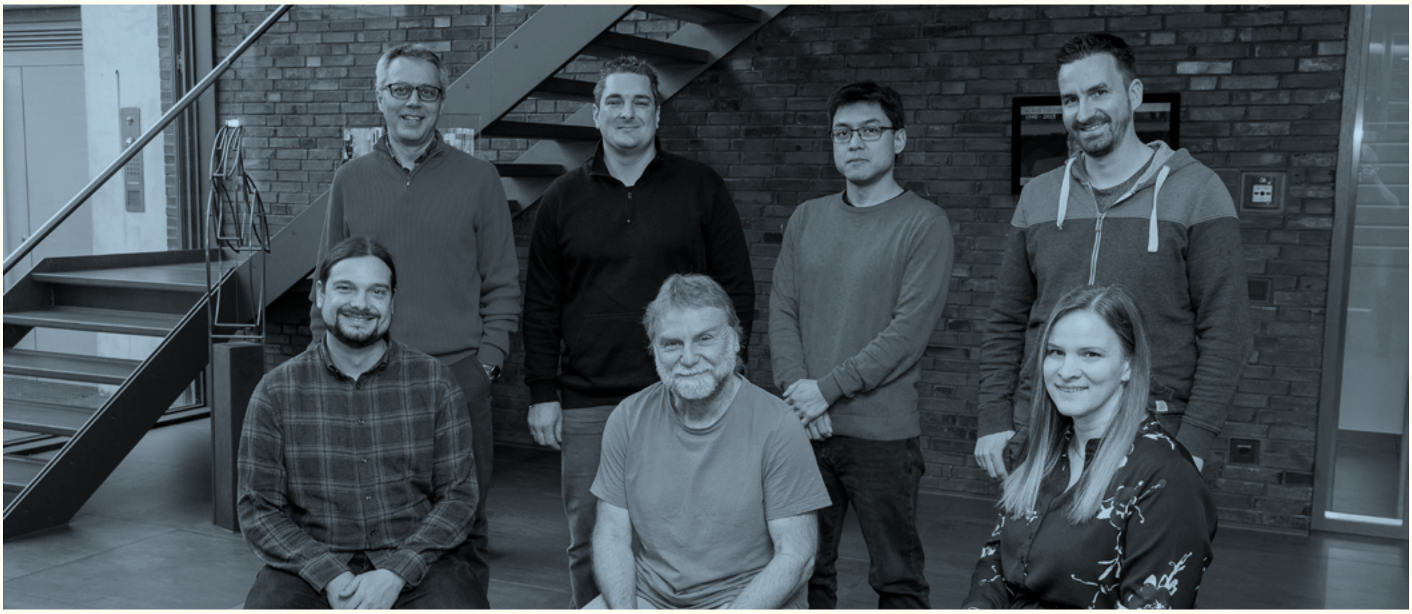
Due to our constant analysis of what we can improve, we revised several HR processes last year, such as the off boarding survey, in which we invite departing staff to share their experience and make suggestions. This feedback is both invaluable and forward looking.

As a result of earlier surveys, a “Good Bye Germany” checklist was set up in 2025 for everyone leaving the country. The checklist provides a systematic guide to many necessary formalities.

A new booking system that was implemented at the end of 2025 streamlined the process of accessing available meeting rooms, parking spots, and desks for all people working at HITS.

Last year was also marked by significant personnel transitions in key roles of the service groups. Angela Michel – Deputy Head of Communications – was systematically integrated in order to succeed Peter Saueressig as Head of Communications in 2026. Furthermore, the IT Services group saw the departure of Group Leader Bogdan Costescu alongside team members Norbert Rabes and Simon Kreuzer. Recruitment for these vacancies remained ongoing at year’s end. In the Front Office team, Benedicta Frech left, and Darija Stein took over supporting our scientists in central assistance matters as well as in supporting the SIMPLAIX project (see Chapter 7).

Thus, coming and going – like breathing in and out – is part of the HITS organism. It generates a large number of alumni, whom we are happy to invite to our biannual HITS Alumni Meeting (see Chapter 5.3).



3.2 IT Services

As the IT Services group, our primary goal is to support our researchers in all IT-related matters. We operate and manage the HPC infrastructure and essential IT services around it, including monitoring, user management, storage, and backup solutions. Furthermore, we provide scientific software development, the maintenance of clients and workstations, and procurement of hardware and software.

Last year we modernized several core services to strengthen the support of our users. For instance, we replaced our legacy ticketing system with a modern, open-source solution and introduced an open-source project management tool in order to enhance our team collaboration and improve visibility across all ongoing projects. In addition, we successfully migrated all virtual machines to a new all-flash storage system on our organization-wide virtualization platform. Furthermore, we completed a major upgrade of our monitoring system and implemented a centralized Time Machine backup solution for our Mac users.

In scientific software development, we concentrated on the EU Project SPACE, which aims to classify data from cosmological simulations using machine learning techniques. A notable highlight was the general assembly held in Barcelona in May, where we presented our machine learning workflow and discussed advancements involving ONNX and Apache Parquet. In June, we participated in the Helmholtz AI Conference at KIT, followed by the ML4ASTRO unconference, where we discussed AI developments in scientific research. Additionally, we showcased our work at the ADASS conference in Görlitz, presenting results from nearly 220 million spectra that had been obtained by the ESA's Gaia mission. In the last quarter of the year, we continued our GROMACS-RAMD project in collaboration with the MCM group. We introduced new features and aimed to contribute to the official GROMACS repository.

Our HPC infrastructure was upgraded at the HITS and URZ data centers by deploying smart PDUs for detailed power monitoring with additional energy savings by automatically powering down idle compute nodes, which was realized by upgrading the Slurm job scheduler to the latest release. At URZ, a new all-flash Lustre storage system was introduced in order to support I/O-intensive AI and ML workloads. The system was connected via NDR InfiniBand and enables GPUDirect Storage (GDS) for improved performance. Additionally, for the 2026 GPU cluster expansion, we planned a dual-phase direct-on-chip liquid cooling system. This technology reduces the cooling energy demand, enables higher outlet water temperatures in order to improve heat reuse and free cooling, and prepares the infrastructure for future high-density CPUs and GPUs, thereby supporting efficient long-term cluster operation.

Group Leader

Dr. Ion Bogdan Costescu
(until December 2025)

Dr. Kai Polsterer
(since October 2025)

Team

Dr. Felix Bangerter
(System Administrator; since September 2025)

Dr. Bernd Doser
(Senior Software Developer)

Dr. Simon Kreuzer
(System Administrator; until August 2025)

Christiane Luttermann
(Team Assistant)

Norbert Rabes
(System Administrator; until August 2025)

Andreas Ulrich
(System Administrator)

Martin Wendt
(System Administrator)

Taufan Zimmer
(System Administrator)

4 Communication and Outreach

Head of Communications

Dr. Peter Saueressig

Team

Marisa Almeida Stegnos

Helena Camilleri
(communication assistant; since April 2025)

Anna Cap
(student; until June 2025)

Arno Fricke
(since June 2025)

Angela Michel
(Deputy Head of Communications)

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 aim to spark enthusiasm for science among high-school and university students and the general public alike through our outreach activities. For the 15th anniversary of HITS in 2025, we organized and contributed to many events, including the 30th anniversary of the Klaus Tschira Foundation.

From Discovery to Discourse – Driving Interest Through Research Success

A research institute's communication is highly dependent on its researchers as well as on their scientific success, without which communicators do not have much to say. In 2025, the Communications team was again pleased to announce several success stories to the public.

HITS researchers were successful twice with their applications for Clusters of Excellence within the Excellence Strategy of the federal and state governments of Germany. As part of "Dynaverse," Kai Polsterer (Astroinformatics, Chapter 2.1) and his team will work toward understanding the dynamic processes in the Universe across a wide range of time scales. Biophysicist Rebecca Wade (Molecular and Cellular Modeling, Chapter 2.6) runs a sub-project at "SynthImmune", entitled "Engineering Immune Function with Synthetic Biology" and provides expertise in both molecular modeling and computer simulation.

A team of researchers from HITS and the Max Planck Institute for Polymer Research (MPIP) have developed a model that learns how to generate proteins whose structures are highly flexible, even with patterns that are uncommon in natural proteins (see Chapter 2.5). Their work – presented at the International Conference on Machine Learning (ICML) – marks a step toward the goal of designing new proteins for applications in biotechnology, therapeutics, and environmental research.

According to this year's Highly Cited Researchers list from Clarivate, group leader Alexandros Stamatakis was named one of the most cited researchers worldwide for the tenth year in a row (see Chapter 2.2). The ranking is an important indicator of the impact of a researcher's scientific publications.

Later in 2025, the Institute of Mathematical Statistics (IMS) announced that HITS group leader Tilmann Gneiting would receive the 2026 Wald Memorial Award and deliver the Wald Lecture at the IMS Annual Meeting in Salzburg in July 2026. The Wald Memorial Award and Lecture is a prestigious recognition given by the IMS to a person whose contributions have been fundamental to the development of statistics or probability.

Girls' Day participants became asteroseismologists for a day.

Workshops, Outreach, and HITS Communication "On Tour" Again

2025 was once again a busy year full of events that had to be organized or curated by the Communications team, including scientific workshops at the Studio Villa Bosch as well as outreach activities both on the premises and outside the institute. These events contributed to increasing the visibility of HITS among both the scientific and general public in Heidelberg and beyond.

First was the "Girls' Day" on 3 April, with the topic "Music of the Stars," conceived by the TOS group, which hosted around 20 girls from Heidelberg and the surrounding towns. The visitors engaged in hands-on experiments to learn more about asteroseismology. Female scientists from the institute presented their personal journey towards a scientific career in order to spark the participants' imaginations.

Biochemistry students from Heidelberg University visited HITS on 24 April to learn about the different fields of scientific research conducted at the institute. Current PhD students from several groups presented their work and highlighted how the students could potentially join HITS to work on their theses.

In early May, the 3rd SIMPLAIX Workshop on "Machine Learning for Multiscale Molecular Modeling" took place (see Chapter 5.1.2). The workshop was a successful platform for exchange for the roughly 100 participating scientists.

Between 25 and 29 June, HITS once again participated in the Explore Science festival at Herzogenriedpark in Mannheim. The five days were filled to the brim with activities that were both educational and fun. Despite the heat, around 25,000 people attended the event.

At the HITS tent, the overall theme was "Experiment, tinker, do-it-yourself – Energy for you and your environment!" The topic of energy was presented in a twofold manner: In the first part,

Scientific discussion during a visit of biochemistry students from Heidelberg University.

children were invited to test how wind and solar power work by using green energy sets, and they had the chance to craft their own pinwheels. The second part was dedicated to the liver, with a fascinating station set up by the LiSyM-Cancer project that included a microscope under which parts of the human liver could be viewed in a liver tissue section. We also had a setup with papers depicting an outline of a child's body onto which children could stamp the liver and then color everything in. The CME, MCM, and SDBV groups – as well as members of the Administrative Services – worked hard to make the event a success.

In July, we celebrated the highlight of our 15th anniversary: The HITS Alumni Meeting (see Chapter 5.3). Only four days later, we organized the 2nd ML4Astro workshop, which had 65 participants from several astronomy institutions in Heidelberg and beyond (see Chapter 5.1.4).

Shortly afterward, the institute again took part in the International Summer Science School Heidelberg (ISH) and welcomed one student each from Heidelberg's twin cities of Rehovot (Israel) and Jelenia Gora (Poland). As in the year before, members of the AIN group volunteered to give an instructive workshop on AI in astronomy for these two students. HITS also hosted the final presentations of the ISH students at the Studio Villa Bosch, where they shared their insights and learnings during the program.

As every year, in September, HITS participated in the Heidelberg Laureate Forum and took advantage of its booth in the reception area. Our Scientific Director, Kai Polsterer, gave a short talk within the "Science Hub Showcase" format in the lecture hall (see Chapter 7).

Examining liver cancer under the microscope: School students visit the HITS booth at Explore Science.

In October, the Klaus Tschira Foundation organized an internal event called "Verbundtreffen" ("Network Meeting") with talks, workshops, and more. As part of the exhibition, the Communications team set up a booth with a mood board that presented different aspects of HITS to the other network institutions (see Chapter 5.5).

> These events contributed to increasing the visibility of HITS among both the scientific and general public in Heidelberg and beyond

Lucky Number 13: 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. In 2012, HITS established the Journalist in Residence program, which it has continued to refine since then. The program 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 as well as new research topics without the pressures of the daily grind.

Our 13th Journalist in Residence was multi-award-winning science journalist Jackson Ryan from Adelaide, Australia. His stay turned out to be extremely successful and included an intense internal seminar with the HITSters, a fruitful visit to the Science Media Center in Cologne, a talk at the IWR of Heidelberg University on the topic of scientific misconduct and the role of journalists, and finally, a splendid public talk on "Antarctica and good journalism" at the German-American Institute (DAI) within the framework of the "Geist" Science Festival and with an audience of almost 200 people.

In the summer of 2025, HITS announced the next call for applications, with

ISH students Alicja Gruszczyńska (front row, far right) and Ilay Ben Shimol (second row, far right) with their tutors from the AIN group.

candidates from six continents applying. A committee of science journalists and scientists selected Monika Mondal (India) as the next HITS Journalist in Residence. She will come to the institute in February 2026 for a six-month stay.

In December, Peter Saueressig attended the World Conference of Science Journalists (WCSJ) in Pretoria, South Africa, as a speaker in a panel on residencies. Former HITS Journalist in Residence Michele Catanzaro moderated the panel, which was entitled "Stepping Out of the Office to Learn More," where Peter was able to share insights from our longstanding program. At the conference, five former HITS Journalists in Residence were present as speakers, session producers, or panelists, thereby demonstrating both the success and sustainability of the program.

Peter Saueressig on stage at the WCSJ in Pretoria.

Over the years, the Journalist in Residence program has inspired several new initiatives that are modeled after the HITS residency. These initiatives range from projects in Germany and Austria to the large European Research Council (ERC) "Frontiers" program, which has just chosen the residencies for its third iteration. These developments set a spark of hope for science journalism – an increasingly endangered field that is desperately needed by both science and society.

13th Journalist in Residence Jackson Ryan during his talk at the "Geist" Science Festival.

> Over the years, the Journalist in Residence program has inspired several new initiatives that are modeled after the HITS residency

5 Events

5.1 Conferences, Workshops & Courses

5.1.1 Main–Neckar Martini User Meeting 2025

Studio Villa Bosch, Heidelberg, 13–14 February 2025

The Martini coarse-grained force field is one of the most popular simulation methods in bio- and soft-matter physics. It is developed and maintained by the Martini Force-Field Initiative, which is an international collaboration of researchers (see https://cgmartini.nl/).

The user meeting brought together over 45 participants from leading groups in the Main–Neckar area to discuss common challenges and solutions when using Martini. The meeting featured sessions on lipid nanoparticle designs, high-throughput simulations, proteins, and lipid membranes. Group discussions and poster sessions facilitated an exchange between the participants on current and future developments.

HITS Independent Postdoc Fabian Grünewald (Chapter 2.12) organized the meeting together with Sebastian Thallmair (Frankfurt Institute for Advanced Studies).

5.1.2 Third SIMPLAIX Workshop on “Machine Learning for Multiscale Molecular Modeling”

Studio Villa Bosch, Heidelberg, 7-9 May 20255

SIMPLAIX is a three-way inter-institutional cooperation between the Heidelberg Institute for Theoretical Studies (HITS), the Karlsruhe Institute of Technology (KIT), and Heidelberg University. The project aims to bridge scales in modeling and simulating molecular systems ranging from (bio)molecules to molecular materials by using multiscale simulation and machine learning. The 3rd SIMPLAIX workshop – which ran from 7–9 May 2025 – was jointly organized by SIMPLAIX and the DFG-funded Research Training Group (RTG 2450) of KIT. The workshop was designed to bring together experts in the field in order to share their research and discuss current challenges (see http://www.simplaix.org and Section 7 of the Annual Report).

About 100 participants attended the three-day workshop, which featured exciting presentations by 12 invited speakers and vivid discussions with active contributions from both early-career and established scientists. The covered topics included machine learning in photochemistry, physics-based approaches and large language models in drug design, quantum chemistry–based force fields, and protein structure prediction for the human proteome. The broad range of discussed applications highlighted the impact of recent advances in machine learning approaches on the modeling and simulation of (bio-)molecules and materials in terms of scope, accuracy, and computational efficiency.

SIMPLAIX is supported by the Klaus Tschira Foundation and the participating institutions (see Chapter 7). Scientific Organizing Committee: Rebecca Wade, Rostislav Fedorov, Jonathan Teuffel, Daniel Sucerquia (all from HITS), Pascal Friederich, Marcus Elstner, David Hoffmann (all from KIT), and Tristan Bereau (Heidelberg University).

5.1.3 Practical Course on Computational Molecular Evolution

Heraklion, Greece, 11–22 May 2025

The practical course on "Computational Molecular Evolution" has been held biennially since 2009 (except in 2020 and 2021 due to the pandemic) in both Heraklion, Greece, and Hinxton, UK. In 2025, the 15th installation of the event took place again at the Hellenic Centre for Marine Research (HCMR) on the island of Crete.

As the volume of genomic data is growing exponentially, appropriate tools and skills for conducting adequate biological data analyses are required. To address these needs, the course aims to provide early career researchers with theoretical knowledge and practical skills to carry out phylogenetic and molecular evolutionary analyses of sequence data. The hands-on course covers a broad spectrum of techniques, such as alignment methods, phylogenetic inference, hypothesis testing, detecting selection, divergence dating, and coalescent analysis. The course includes sessions on the analysis of both protein and nucleotide sequences.

Moreover, this EMBO Practical Course additionally offers a unique opportunity for direct interaction with some of the world-leading scientists and authors of widely used phylogenetic analysis tools (Maria Anisimova, Bastien Bousseau, Bruce Rannala, Alexandros Stamatakis, Benjamin Redelings, Ziheng Yang, Rachel Warnock, etc.).

CME group leader Alexandros Stamatakis was the main organizer of this event together with Aglaia ("Cilia") Antoniou (HCMR, Crete, Greece), Ziheng Yang (University College London, UK), George Kotronakis (HCMR, Crete, Greece), and Adam Leaché (University of Washington, USA). Among the training team of 17 researchers were two HITS Alumni: Ben Bettisworth (postdoc at the Institute of Computer Science, Foundation for Research and Technology Hellas, Greece) and Paschalia Kapli (now a research leader at the Natural History Museum, London, UK).

5.1.4 ML4ASTRO Workshop

Studio Villa Bosch, Heidelberg, 8 July 2025

The Machine Learning for Astronomy (ML4Astro) workshop was held on 8 July 2025 at the Studio Villa Bosch in Heidelberg. The event was jointly organized by the Interdisciplinary Center for Scientific Computing (IWR), the Max Planck Institute for Astronomy (MPIA), the Center for Astronomy of Heidelberg University (ZAH), and the HITS Astroinformatics group. About 65 people from over 15 countries with a shared background and career in astronomy joined the workshop, which focused on the application of machine learning techniques to address complex astrophysics problems.

The program began with a series of presentations covering topics such as physics-suggested neural nets, machine learning in microlensing, and chemical structure models of stars. These sessions were followed by unconference discussions that fostered an interactive and collaborative atmosphere.

The primary objective of the one-day workshop was to facilitate knowledge exchange among the participants regarding the application of machine learning techniques for solving similar challenges across various astrophysics contexts. Additionally, the workshop also provided a platform for exploring potential new collaborations and joint research initiatives.

5.1.5 Sixth Biological Diffusion and Brownian Dynamics Brainstorm (BDBDB6)

Online, 11–12 December 2025

The Biological Diffusion and Brownian Dynamics Brainstorm is a free online workshop series with a long tradition dating back to 2007. At its core, the workshop is dedicated to exploring and informally discussing the latest experimental and theoretical advances in studying biological diffusion. It focuses on the Brownian dynamics molecular simulation method and its applications in simulating biological macromolecules.

Over 90 participants signed up for the workshop, which had 45 attendees at any one time who spanned various time zones, thereby connecting scientists on both sides of the Atlantic. Five inspiring keynote lectures were supplemented by talks as well as by lively discussions on recent software developments and roadmaps for future developments in the field.

Scientific Organizing Committee: Rebecca Wade, Riccardo Beccaria, Stefan Richter (all from HITS), Rommie Amaro (University of California, San Diego), and Chia-en Chang (University of California, Riverside).

5.1.6 19th Winter Workshop on Stellar Astrophysics

Studio Villa Bosch, Heidelberg, 15–17 December 2025

Once again, the annual Workshop on Stellar Astrophysics brought together over 45 astrophysicists from various national and international institutions at Studio Villa Bosch from 15–17 December 2025.

The agenda included talks by Christine Collins (Trinity College Dublin, Ireland), Tamara Rogers (Newcastle University, UK), Dominic Bowman (Newcastle University, UK), Rüdiger Pakmor (Max Planck Institute for Astrophysics, Garching, Germany), and Christian Klingenberg (University of Würzburg, Germany) as well as 2025 Klaus Tschira Guest Professor Raphael Hirschi (Keele University, UK).

The workshop is a recurring opportunity to reunite with peers in order to foster connections and convert shared ideas into future collaborations.

5.2 HITS Colloquia

Simon Olsson

Chalmers University of Technology, Gothenburg, Sweden

27 January 2025: How Can Generative AI Help Understand the Physics of Molecules?

Cecilia Clementi

Freie Universität Berlin, Department of Physics, Germany

10 February 2025: Modeling Protein Dynamics with Machine Learning and Molecular Simulation

Victoria Stodden

Daniel J. Epstein Department of Industrial & Systems Engineering,
University of Southern California, USA

24 March 2025: Verifying Correctness in AI-enabled Scientific Research – a New Frontier

Christina Elmer

TU Dortmund University, Institute & School of Journalism, Germany

28 April 2025: Synthetic truthfulness? Learning systems as a challenge for communication and a transformative force for journalism

Massimiliano Bonomi

Institut Pasteur, Paris, France

21 July 2025: Structural and Dynamic Biology with Integrative Approaches

Raphael Hirschi

Klaus Tschira Guest Professor, Keele University, Staffordshire, UK

20 October 2025: Predicting the Impact of Massive Stars in the Universe and how to Store Renewable Energy From the Sun on Earth

Jackson Ryan

Journalist in Residence, Australia

22 October 2025: Lessons from Antarctica: How Great Science Journalism Is Made (and why we need more of it!), at the "Geist" Science Festival, German-American Institute (DAI), Heidelberg

Rommie Amaro

University of California, San Diego, USA

2 December 2025: Multiscale Computational Microscopy for In Situ Molecular Dynamics

5.3 Alumni Meeting

HITS Campus, Heidelberg, 4 July 2025

Over the course of the last 15 years, HITS has grown from its early days – when it succeeded its predecessors, EML and EML Research – into an institute that now hosts 11 working groups and about 130 scientists from more than 30 nations. Many people have passed through the institute and will always be connected by their shared experiences here. At our biennial Alumni Meeting, we came together to embrace networking and exchange between alumni and current HITSters. And this year, the celebration of 15 years of HITS was bigger than ever.

The welcome address by HITS Managing Director Gesa Schönberger and HITS Scientific Director Kai Polsterer set the stage for the event. They began with an icebreaker to determine the longest-serving HITSters in the audience, some of whom predated the founding of the institute itself. Anne Schreiter (GSO* – Guidance, Skills, & Opportunities) delivered the subsequent keynote speech, which was centered on "the lucky mindset." She began by quoting Louis Pasteur: "Chance favors only the prepared mind," underlining the notion that a proactive approach to career planning can help in navigating a seemingly unpredictable academic career path. Building a network of peers and mentors, for example, is a necessary foundation when it comes to being exposed to new opportunities.

Anne Schreiter (GSO) during her talk.*

On the career panel (f.r.t.l.): Nick Michelarakis, Isabel Kemmer, Anna Piras and moderator Alexander Jordan.

This year's career panel was chaired by Alexander Jordan, who is a staff scientist in the CST group. Isabel Kemmer and Nick Michelarakis (both formerly MBM) as well as Anna Piras (CCC) spoke about their lives before and during HITS and provided some insights into what they are doing now. Isabel stated that "one presentation can sometimes be the first step [in the] next career direction." Having begun at HITS as a research assistant in database curation, Isabel was introduced to the FAIR principles of sharing data. After completing her PhD in molecular simulation and having communicated with the public at multiple institute events, today, she combines all these experiences as a FAIR Image Data Steward.

Nick transitioned to the pharmaceutical industry, while Anna stayed in academia to join the National Research Council of Italy. The diversity across the panel highlighted the role of HITS as a career springboard.

The event was rounded off by a surprise photo challenge and our traditional "sit and sizzle" barbeque. The mixture of alumni, seasoned HITSters, and new members provided a good opportunity to share highlights of the institute's 15-year history and yielded insights into how HITS has been shaped into the institute of today. We look forward to many more successful years to come.

A meeting in high temperatures and high spirits: Maja Rey, Ivan Savora, Olga Krebs and Ulrike Wittig (f.l.t.r.).

5.4 The Future of NLP

Studio Villa Bosch, Heidelberg, 24–25 November 2025

"Think beyond the limits!" – The HITS motto also serves as an imperative for the institute to keep an eye out for and explore future trends before they become mainstream research.

Consequently, HITS introduced a new event format called "The Future of…" beginning with the Natural Language Processing (NLP) group. This research field develops methods, algorithms, and tools for automatically analyzing natural language. The two-day symposium in the Studio Villa Bosch focused on questions related to applications and theories in NLP while also addressing the field of machine learning as well as large language models and their general impact on society.

HITS Scientific Director Kai Polsterer and KTS Managing Director Rafael Lang welcomed the more than 60 participants. Mirella Lapata (University of Edinburgh, UK) delivered a keynote talk on "Compositional Intelligence: Coordinating Multiple LLMs for Complex Tasks." She illustrated the multi-LLM approach through two case studies: narrative story generation and visual question answering. In nine talks and two panels, speakers from Germany, Switzerland, Austria, the UK, and Canada elaborated on and discussed different applications, theoretical approaches, and the manifold relationships between NLP and society.

The event was organized by Marisa Almeida Stegnos, Helena Camilleri, Angela Michel (Communications), and Christina Blach (Administration).

More than 60 people attended the event.

A warm welcome by HITS Scientific Director Kai Polsterer (left) and Rafael Lang (KTS Managing Director).

5.5 30th Anniversary of the Klaus Tschira Foundation

Gutshof Ladenburg, 16 October 2025

Enabling science, education, and science communication in the natural sciences, mathematics, and computer science has always lain at the core of the Klaus Tschira Foundation (KTS). Since 1995, the non-profit foundation has launched many successful endeavors. As a research institute of the KTS, HITS is one such endeavor that exists in a so-called "Verbund" ("network") with other organizations such as the Heidelberg Laureate Forum Foundation (HLFF) and the Science Media Center Germany (SMC). The work is complemented by partner projects such as "Make your school" – in which school children learn problem-oriented digital and technical skills – and the SIMPLAIX research project on data-driven molecular research. Over the last three decades, the commitment has created a long-lasting impact, which it will continue to do going forward.

In October, the foundation celebrated its 30th anniversary with a festive event at the Gutshof Ladenburg. Around 400 guests from the fields of science, education, politics, culture, and society attended, including companions and employees from the KTS "Verbund". The "Future Dialogue" highlighted the current transformation process of the foundation toward a more impactful, transparent, and connected funding management. The entire network came together in the exhibition hall to present their latest work and exchange ideas.

At the exhibition, the HITS Communications team presented current developments in their projects.

At the HITS booth, current developments such as the Annual Report redesign, the “Future of …” symposium series, and the Journalist in Residence program were highlighted via a mood board.

The subsequent science slam revealed how cutting-edge science can be communicated in an engaging and fun way. The event was rounded out by an evening of music, food, shared stories about certain projects, and casual interviews with some of Klaus Tschira’s companions, including former HITS Managing Director Andreas Reuter and SMC Managing Director Volker Stollorz, who had met Klaus Tschira during Stollorz’s stay as the first HITS Journalist in Residence in 2012. In his interview, HITS Scientific Director Kai Polsterer reminded the participants of the founder’s motto: “Think beyond the limits!”

HITS Scientific Director Kai Polsterer being interviewed during the evening gala event (pictures: © Klaus Tschira Foundation/Annette Mück).

5.6 A HITS Year-in-Sports

In and around Heidelberg, Throughout 2025

From the start, HITSters have always been very active in different fields of sports that are as diverse as the people themselves – from running, biking, and football to swimming, climbing, and bouldering. The 2025 season was particularly successful, with broad participation across events.

In April, solo and relay runners joined the ATOS Half Marathon and Henkel Team Run. The 21.1-km course led up and down many roads and hills in Heidelberg and even passed right in front of the HITS institute, where a celebration booth cheered for all the runners on the final stretch of the race.

HITS' "Alpha Centauri" team at the finish line of the Heidelberg Triathlon.

Every morning, the HITS garage is always full of bikes belonging to commuter members. Therefore, it was a natural fit to join the City Cycling Competition, where a total of 2,690 km was cycled between 23 June and 13 July. Not only is the competition a fun way to track everyone's cycling distance, but the collected data are also used to improve the cycling infrastructure in Heidelberg.

From 11 to 13 July, HITS runners joined the NCT cancer charity event alongside 5,000 other participants across the globe. Over three days, HITSters ran a total of 266.32 km, and HITS sponsored their efforts with €500.

HITS' own "Alpha Centauri" team joined the Heidelberg Triathlon in July. Francisca Espinoza (TOS) swam 1.6 km, Vincent Bronner (SET) rode his bicycle 36 km, and finally, Beatriz Bordadagua (TOS) ran 10 km, all in just over 2 hours and 42 minutes, thereby securing them 18th place among their age group.

On 10 October, the HITS football team – coordinated by Harald Haas (Administrative Services) and reinforced by HITS Alumni and Christian Roos (Head of KTS IT) – joined a tournament at the Technical University in Kaiserslautern. Although there was an element of friendly competition, the emphasis was on coming together to play football and having fun at the BBQ afterward.

The HITS football team in Kaiserslautern.

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 field of natural, mathematical, and computer science. To that end, HITS invites internationally renowned scientists for sabbaticals or extended research stays ranging from three weeks to six months. Invited guest professors collaborate with scientists at the institute and potentially develop joint research projects. In addition, these guests 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.

Raphael Hirschi during his HITS colloquium talk on the impact of massive stars in the Universe.

This year again, HITS welcomed a Klaus Tschira Guest Professor.

Applying Astrophysics for the Benefit of Society: Raphael Hirschi

Raphael Hirschi is a Professor of Stellar Hydrodynamics and Nuclear Astrophysics at the University of Keele, UK. He came to HITS in April 2025 and stayed until December 2025.

In October, Raphael held a HITS colloquium talk on “Predicting the impact of massive stars in the Universe and how to store renewable energy from the Sun on Earth” (see Chapter 5.2). His talk highlighted the application range of skills and tools in astrophysics. When he became the head of the Digital Society Institute at Keele, Raphael wanted to expand his research and found ways to explore renewable energy as a field in which to apply his astrophysics knowledge for the benefit of society.

Thinking beyond the limits of his field made fruitful collaborations with a variety of groups at HITS possible. Raphael’s main point of contact was with Friedrich Röpke (Physics of Stellar Objects) in working on 3D magnetohydrodynamic (MHD) simulations of massive

> His curiosity for stars that he could see from his backyard led him to pursue physics

stars. After joining HITS in April, Raphael set goals early on with Friedrich, and they began analyzing their first results toward the end of Raphael’s stay. Moreover, Raphael also regularly discussed binary stars with Fabian Schneider (SET). With Saskia Hekker (TOS), he also connected theory and asteroseismic observations. Raphael additionally explored the work of Kai Polsterer’s group, stating that “it's inspiring to see how other researchers apply machine learning in astrophysics.” The foundation for shared future projects has thus been laid.

Raphael’s personal journey in science has not been straightforward. “People often think you’re born a scientist,” he stated in an interview with the HITS newsletter "The Charts". Growing up on a farm in Switzerland, a scientific career once seemed a very far reach away for Raphael. However, his curiosity for stars that he could see from his backyard led him to pursue physics. With powerful tools such as ESO’s Extremely Large Telescope and the James Webb Space Telescope at hand, Raphael expects great scientific advancements in the coming years: “What was once limited to our galaxy will be extended to the entire Universe.”

6.2 HITS Independent Postdoc Program

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

> … a wonderful opportunity for highly talented young scientists who wish to transition from PhD students to junior group leaders

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 the HITS shareholders of Heidelberg University and the Karlsruhe Institute of Technology (KIT). In addition, successful candidates benefit from outstanding computing resources as well as from the various courses offered at HITS.

Candidates for the program must hold a doctoral degree or an equivalent academic qualification at the beginning of the fellowship. Application is open to candidates up to three years after the completion of their PhD by the time of the application deadline. This deadline can be extended in the case of documented career breaks, for example, due to parental leave. Candidates are not allowed to have conducted previous research at HITS except during brief visits, and the main thread of this research is not allowed to have been in collaboration with a HITS group leader.

The first HITS Independent Postdoc was 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). After three years of successful research (and outreach work for the “Habitable” game), Rajika left HITS in September 2025.

In October 2023, biologist **Fabian Grünewald** (University of Groningen, the Netherlands) joined the institute as the second HITS Independent Postdoc. Born in Germany, Fabian received his PhD with distinction in physical/computational chemistry at the University of Groningen. His research interests lie in the in silico design and understanding of polymeric materials at the interface of biology and traditional material science by means of computer simulations. In his project, Fabian focuses on RNA and aims to develop both computational methods and software that can help in designing RNA nanoparticle therapeutics (see Chapter 2.12). He is also a core member of the Martini Force Field Initiative community, which develops a widely used simulation method. In 2025, Fabian organized the first “Main-Neckar Martini User Meeting” at the Studio Villa Bosch – an event that addressed challenges and solutions when using the Martini course-grained force field method (see Chapter 5.1.1).

Computer scientist **Vagrant Gautam** joined HITS as an Independent Postdoc in September 2025. Born in Canada, Vagrant was raised in India and Singapore and returned to Canada to study computational linguistics at Simon Fraser University. Vagrant earned a PhD at the Faculty of Mathematics and Computer Science at Saarland University, Germany, with a dissertation on the “Fair and Faithful Processing of Referring Expressions in English.” At HITS, Vagrant collaborates closely with the Natural Language Processing (NLP) group.

7 Collaborations

SIMPLAIX

SIMPLAIX is a 3-way inter-institutional collaboration between HITS, Heidelberg University, and the Karlsruhe Institute of Technology (KIT) that aims to pool the expertise of the three partner institutes with the goal of addressing the challenge of bridging scales in studies of molecules and molecular materials by using multiscale simulations and machine learning.

> SIMPLAIX aims to pool the expertise of the three partner institutes

In SIMPLAIX, these methods are developed and employed to study a set of challenging problems in computational studies of biomacromolecules and molecular materials within 8 multidisciplinary, inter-institutional research projects.

SIMPLAIX is coordinated by HITS group leaders Rebecca Wade (MCM) and Jan Stühmer (MLI). Throughout the course of 2025, eleven doctoral students worked on the 8 SIMPLAIX projects. SIMPLAIX is funded by the Klaus Tschira Foundation and is supported by in-kind contributions from KIT and Heidelberg University.

The third SIMPLAIX workshop attracted about 100 participants.

From 7–9 May 2025, the third SIMPLAIX Workshop on "Machine Learning for Multiscale Molecular Modeling" took place at the Studio Villa Bosch in Heidelberg (see Chapter 5.1.2). The aim of the event was to bring together scientists working in the field in order to enable them to share their research and discuss current challenges. In this on-site meeting, about 100 participants attended the eight sessions, in which twelve invited speakers gave talks on current developments in the use of machine learning for chemistry and molecular biology in fields ranging from photochemistry to drug design.

Rommie Amaro (University of California, San Diego, USA) during her HITS–SIMPLAIX colloquium talk in December 2025.

Over the year, three internal project meetings were held – one each at the Studio Villa Bosch, Heidelberg University, and KIT. At these meetings, the SIMPLAIX researchers engaged in several journal session discussions. Moreover, four HITS–SIMPLAIX colloquia (see Chapter 5.2) and scientific talks with external speakers were organized.

Plans for 2026 include the fourth international SIMPLAIX workshop in May as well as several joint colloquia and the start of the second SIMPLAIX project phase.

Heidelberg Laureate Forum

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

Representing HITS at the HLF

From 14–19 September, the 12th HLF returned to the Heidelberg University premises. The program included laureate lectures, master classes, lightning talks, discussions, and various interactive program elements.

Once again, HITS had the chance to be visible at the event with a booth in the reception area. The communications and human resources team – supported by students Leif Seute, Mila Coetzee, Torben Berndt (all MLI), and Alexander Jordan (CST) – prepared a short presentation and giveaways, including stickers and magnets with the HITS logo. Located at the bottom of the stairs to the HLF sessions (and close to the coffee machines), the booth attracted many participants as they passed by.

Ready for questions: The human resources team and scientists of the MLI group at the HITS booth.

The “Science Hub Showcase”

On Wednesday morning, the HLF young researchers were given the opportunity to learn about some of the scientific institutes and science-based companies in Heidelberg and the surrounding region. The goal of this “Science Hub Showcase” was to help young researchers expand their professional network and learn more about the wide range of employment opportunities for mathematicians and computer scientists. Thirteen organizations participated in the introduction round, which took place in the university auditorium. For HITS, Scientific Director Kai Polsterer hit the stage and outlined the most important facts about the institute to the audience. “With 130 researchers from more than 30 countries working in 11 groups, HITS is an interdisciplinary, international, and inspiring research institute.”

> HITS is an interdisciplinary, international, and inspiring research institute

HITS Scientific Director Kai Polsterer during his “elevator pitch” talk at the “Science Hub Showcase.” (Picture: © Heidelberg Laureate Forum Foundation (HLFF)).

After the presentations, the HITS booth in the reception area was crowded with interested young researchers. The HITS Communications team distributed flyers and stickers and collected business cards, especially from young researchers interested in the institute’s research fields. The science reporters covering the event were also interested in the Journalist in Residence Program.

8 Publications

Abac A, Abramo R, Albanesi S, Albertini A, Agapito A, Agathos M, Albertus C, Andersson N, Andrade T, Andreoni I, Angeloni F, Antonelli M, Antoniadis J, Antonini F, Sedda MA, Artale MC, Ascenzi S, Auclair P, Bachetti M, Badger C, Banerjee B, Barba-Gonzalez D, Barta D, Bartolo N, Bauswein A, Begnoni A, Beirnaert F, Bejger M, Belgacem E, Bellomo N, Bernard L, Bernardini MG, Bernuzzi S, Berry CPL, Berti E, Bertone G, Bettoni D, Bezares M, Bhagwat S, Bisero S, Bizouard MA, Blanco-Pillado JJ, Blasi S, Bonino A, Borghese A, Borhanian S, Bortolas E, Botticella MT, Branchesi M, Breschi M, Brito R, Brocato E, Broekgaarden FS, Bulik T, Buonanno A, Burgio F, Burrows A, Calcagni G, Canevarolo S, Cappellaro E, Capurri G, Carbone C, Casadio R, Cayuso R, Cerda-Duran P, Char P, Chaty S, Chiarusi T, Chruslinska M, Cireddu F, Cole P, Colombo A, Colpi M, Compere G, Contaldi C, Corman M, Crescimbeni F, Cristallo S, Cuoco E, Cusin G, Canton TD, Dalya G, D'Avanzo P, Davari N, De Luca V, De Renzis V, Della Valle M, Del Pozzo W, De Santi F, De Santis AL, Dietrich T, Dimastrogiovanni E, Domenech G, Doneva D, Drago M, Dupletsa U, Duval H, Dvorkin I, Elias-Rosa N, Ente J, Fairhurst S, Fantina AF, Fasiello M, Fays M, Fender R, Fischer T, Foucart F, Fragos T, Foffa S, Franciolini G, Fumagalli J, Gair J, Gamba R, Garcia-Bellido J, Garcia-Quiros C, Gergely LA, Ghirlanda G, Ghosh A, Giacomazzo B, Gittins F, Giudice IF, Goncharov B, Gonzalez A, Goriely S, Graziani L, Greco G, Gualtieri L, Guidi GM, Gupta I, Haney M, Hannam M, Harms J, Harutyunyan A, Haskell B, Haungs A, Hazra N, Hemming G, Heng IS, Hinderer T, van der Horst A, Hu Q, Husa S, Iacovelli F, Illuminati G, Inguglia G, Villalba DI, Janquart J, Janssens K, Jenkins AC, Jones I, Kacskovics B, Klessen RS, Kokkotas K, Kuan H-J, Kumar S, Kuroyanagi S, Laghi D, Lamberts A, Lambiase G, Larrouturou F, Leaci P, Lenzi M, Levan A, Li TGF, Li Y, Liang D, Limongi M, Liu B, Llanes-Estrada FJ, Loffredo E, Long O, Lope-Oter E, Lukes-Gerakopoulos G, Maggio E, Maggiore M, Mancarella M, Mapelli M, Marchant P, Margiotta A, Mariotti A, Marriott-Best A, Marsat S, Martinez-Pinedo G, Maselli A, Mastrogiovanni S, Matos I, Melandri A, Mendes RFP, de Souza JMS, Mentasti G, Mezcua M, Mosta P, Mondal C, Moresco M, Mukherjee T, Muttoni N, Nagar A, Narola H, Nava L, Moreno PN, Nelemans G, Nielsen AB, Nissanke S, Obergaulinger M, Oertel M, Oganesyan G, Onori F, Pacilio C, Pagliaroli G, Palomba C, Pang PTH, Pani P, Papalini L, Patricelli B, Patruno A, Pedrotti A, Perego A, Perez-Garcia MA, Perigois C, Perna G, Peroux C, Perret J, Perrodin D, Pesci A, Pfeiffer HP, Piccinni OJ, Pieroni M, Piranomonte S, Pompili L, Porter EK, Porto RA, Pound A, Powell J, Puech M, Pratten G, Puecher A, Pujolas O, Quartin M, Raduta AR, Ramos-Buades A, Rase A, Razzano M, Rea N, Regimbau T, Renzini A, Rettegno P, Ricciardone A, Riotto A, Romero-Rodriguez A, Ronchini S, Rosinska D, Rossi A, Roy S, Rubiera-Garcia D, Rubio J, Ruiz-Lapuente P, Sagun V, Sakellariadou M, Salafia OS, Samajdar A, Sanchis-Gual N, Sanna A, Santoliquido F, Sathyaprakash B, Schmidt P, Schmidt S, Schneider FRN, Schneider R, Sedrakian A, Servant G, Sevrin A, Shao L, Silva HO, Simakachorn P, Smartt S, Sotiriou TP, Spera M, Stamerra A, Steer DA, Steinhoff J, Stergioulas N, Sturani R, Suarez D, Suresh J, Swain S, Tagliazucchi M, Tamanini N, Tasinato G, Tauris TM, Tissino J, Tomaselli GM, Toonen S, Torres-Forne A, Turski C, Ugolini C, Vagenas EC, Dall'Armi LV, Valenti E, Valiante R, Broeck CVD, van de Meent M, van Son LAC, Vanvlasselaer M, Vaglio M, Varma V, Veitch J, Vaskonen V, Vergani SD, Wils M, Witek H, Wong ICF, Yazadjiev S, Yim G, Acernese F, Ahn H, Allocca A, Amato A, Andres-Carcasona M, Avallone G, Bachlechner M, Baer P, Bagnasco S, Balbi G, Barone F, Benedetti E, Benning C, Bini S, Salcedo JLB, Bozza V, Bruno M, Butz T, Califano M, Calloni E, Carapella G, Cardini A, Subrahmanya SC, Chiadini F, Chiummo A, Cianetti S, Ciani G, Coccia E, Contu A, Cornelissen R, Cozzumbo A, Croney L, Crosta M, D'Agostino R, Danilishin S, D'Antonio S, De Bolle J, Degallaix J, De Laurentis M, della Monica R, De Marco F, de Martino I, De Rosa R, De Salvo R, De Simone R, Detavernier C, Diaferia G, Di Cesare M, Di Fiore L, Di Giovanni M, Di Pace S, Docherty J, D'Urso D, Mecherfi OE, Errico L, Fabrizi F, Fafone V, Fanti V, Fittipaldi R, Fiumara V, Freise A, Funk S, Gaedtke M, Garufi F, Gerberding O, Giangrandi E, Giunchi C, Graham V, Granata M, Granata V, Green A, Haughian K, Heisenberg L, Hennig M, Hild S, Hoang VL, Holland N, Iannone G, Isleif K-S, Joppe R, Kim C-H, Kim C, Kim K, Korb E, Korobko M, Kranzhoff L, Kuhlbusch T, Lacaille G, Lartaux-Vollard A, Lavezzi L, Laycock P, Lee S, Lee S, Lee S, Losurdo G, Lucchesi L, Luck H, Macquet A, Majorana E, Mangano V, Martelli F, Martin I, Martinez M, Masoni A, Massaro L, Melini D, Mercurio A, Mereni L, Miller AL, Mirasola L, Mitchell A, Molinari I, Montani M, Mow-Lowry C, Murgia R, Murray PG, Muscas G, Naticchioni L, Nela A, Nery M, Niggemann T, Nippe N, Novak J, Numic A, Olivieri M, Orsini M, Park JG, Pascucci D, Perreca A, Piergiovanni F, Pierro V, Pinard L, Pinto I, Punturo M, Puppo P, Quochi F, Rading RO, Rapagnani P, Ricci M, Rodrigues D, Romano R, Rozza D, Saffarieh P, Santucci F, Schramm S, Schwab B, Sequino V, Neto LS, Silenzi L, Sintes AM, Sopuerta CF, Spencer A, Stahl A, Steinlechner J, Steinlechner S, Szabo R, Thummler T, Tofani E, Torniamenti S, Travaglini R, Trozzo L, Vaccaro MP, Valentini M, Van P, van Dongen J, van Heijningen J, van Ranst Z, Vardaro M, Verdier P, Vernieri D, Wagner N, Woehler J, Wolf J, Zavattini G, Zink A, Zmija A (2025) The Science of the Einstein Telescope. arXiv preprint arXiv:2503.12263

Abaza H, Shutsko A, Klopfenstein SAI, Vorisek CN, Schmidt CO, Brünings-Kuppe C, Clemens V, Darms J, Hanß S, Intemann T, Jannasch F, Kasbohm E, Lindstädt B, Löbe M, Nimptsch K, Nöthlings U, Ocanto MG, Osei TB, Perrar I, Peters M, Pischon T, Sax U, Schulze MB, Schwarz F, Schwedhelm C, Thun S, Waltemath D, Zeleke AA, Müller W, Golebiewski M (2025) Toward a Domain-Overarching Metadata Schema for Making Health Research Studies FAIR (Findable, Accessible, Interoperable, and Reusable): Development of the NFDI4Health Metadata Schema. JMIR Med Inform 13:e63906

Ahlborn F, Bellinger EP, Hekker S, Basu S, Mokrytska D (2025). The robustness of inferred envelope and core rotation rates of red giant stars from asteroseismology. A&A 693:A274

Ahlborn F, Joel Ong JM, Van Beeck J, Bellinger EP, Hekker S, Basu S (2025). Impact of near-degeneracy effects on linear rotational inversions for red giant stars. A&A 704:A230

Ajith P, Seoane PA, Sedda MA, Arcodia R, Badaracco F, Banerjee B, Belgacem E, Benetti G, Benetti S, Bobrick A, Bonforte A, Bortolas E, Braito V, Branchesi M, Burrows A, Cappellaro E, Ceca RD, Chakraborty C, Subrahmanya SC, Coughlin MW, Covino S, Derdzinski A, Doshi A, Falanga M, Foffa S, Franchini A, Frigeri A, Futaana Y, Gerberding O, Gill K, Giovanni MD, Giudice IF, Giustini M, Gläser P, Harms J, van Heijningen J, Iacovelli F, Kavanagh BJ, Kawamura T, Kenath A, Keppler E-A, Kobayashi C, Komatsu G, Korol V, Krishnendu NV, Kumar P, Longo F, Maggiore M, Mancarella M, Maselli A, Mastrobuono-Battisti A, Mazzarini F, Melandri A, Melini D, Menina S, Miniutti G, Mitra D, Morán-Fraile J, Mukherjee S, Muttoni N, Olivieri M, Onori F, Papa MA, Patat F, Perali A, Piran T, Piranomonte S, Pol AR, Pookkillath MC, Prasad R, Prasad V, Rosa AD, Chowdhury SR, Serafinelli R, Sesana A, Severgnini P, Stallone A, Tissino J, Tkalčić H, Tomasella L, Toscani M, Vartanyan D, Vignali C, Zaccarelli L, Zeoli M, Zuccarello L (2025). The Lunar Gravitational-wave Antenna: mission studies and science case. J. Cosmol. Astropart. Phys. (01):108

Alpay A, Heuveline V (2025). Adaptivity in AdaptiveCpp: Optimizing Performance by Leveraging Runtime Information During JIT-Compilation. In: Proceedings of the 13th International Workshop on OpenCL and SYCL. ACM, Heidelberg Germany, pp 1-12

Alper P, D'Anna F, Droesbeke B, Andrabi M, Andrade Buono R, Bianchini F, Bösl K, Chandramouliswaran I, Cook M, Faria D, Fatima N, Hooft R, Jareborg N, Jetten M, Pilvar D, Poires-Oliveira G, Popleteeva M, Portell-Silva L, Slifka J, Suchánek M, van Gelder C, Welter D, Wittig U, Coppens F, Goble C (2025). RDMkit: A research data management toolkit for life sciences. Patterns:101345

Amaral AVR, Wolffram D, Moraga P, Bracher J (2025). Post-processing and weighted combination of infectious disease nowcasts. PLOS Computational Biology, 21(3):e1012836

Amaro RE, Åqvist J, Bahar I, Battistini F, Bellaiche A, Beltran D, Biggin PC, Bonomi M, Bowman GR, Bryce RA, Bussi G, Carloni P, Case DA, Cavalli A, Chang C-EA, Cheatham TE, Cheung MS, Chipot C, Chong LT, Choudhary P, Cisneros GA, Clementi C, Collepardo-Guevara R, Coveney P, Covino R, Crawford TD, Dal Peraro M, De Groot BL, Delemotte L, De Vivo M, Essex JW, Fraternali F, Gao J, Gelpí JLl, Gervasio FL, González-Nilo FD, Grubmüller H, Guenza MG, Guzman HV, Harris S, Head-Gordon T, Hernandez R, Hospital A, Huang N, Huang X, Hummer G, Iglesias-Fernández J, Jensen JH, Jha S, Jiao W, Jorgensen WL, Kamerlin SCL, Khalid S, Laughton C, Levitt M, Limongelli V, Lindahl E, Lindorff-Larsen K, Loverde S, Lundborg M, Luo YL, Luque FJ, Lynch CI, MacKerell AD, Magistrato A, Marrink SJ, Martin H, McCammon JA, Merz K, Moliner V, Mulholland AJ, Murad S, Naganathan AN, Nangia S, Noe F, Noy A, Oláh J, O'Mara ML, Ondrechen MJ, Onuchic JN, Onufriev A, Osuna S, Palermo G, Panchenko AR, Pantano S, Parish C, Parrinello M, Perez A, Perez-Acle T, Perilla JR, Pettitt BM, Pietropaolo A, Piquemal J-P, Poma AB, Praprotnik M, Ramos MJ, Ren P, Reuter N, Roitberg A, Rosta E, Rovira C, Roux B, Rothlisberger U, Sanbonmatsu KY, Schlick T, Shaytan AK, Simmerling C, Smith JC, Sugita Y, Świderek K, Taiji M, Tao P, Tieleman DP, Tikhonova IG, Tirado-Rives J, Tuñón I, Van Der Kamp MW, Van Der Spoel D, Velankar S, Voth GA, Wade R, Warshel A, Welborn VV, Wetmore SD, Wheeler TJ, Wong CF, Yang L-W, Zacharias M, Orozco M (2025) The need to implement FAIR principles in biomolecular simulations. Nat Methods 22(4):641–645

Athanasiou C, Claveras Cabezudo A, Tsengenes A, Wade RC (2025). Simulation of Neurotrophin Receptor Transmembrane Helix Interactions Reveals Active States and Distinct Signaling Mechanisms. JACS Au 5(5):2186-2200

Banerjee S, Fan Y, Strube M (2025). HITS at DISRPT 2025: Discourse segmentation, connective detection, and relation classification. In Proceedings of the 4th Shared Task on Discourse Relation Parsing and Treebanking (DISRPT) at CODI, Suzhou, China, 9 November, pp.63-78

Battino U, Keegans JD, Allen M, Röpke FK, Herwig F, Best A, Hirschi R, Piersanti L, Straniero O, Sim SA, Travaglio C, Denissenkov PA (2025). Trans-Fe elements from type Ia supernovae: I. Heavy element nucleosynthesis during the formation of near-Chandrasekhar white dwarfs. A&A 703:A44

Beier S, Arend D, Bauer D, Brandizi M, Brilhaus D, Chadwick E, Clemens V, Crusoe MR, Feser M, Grieb J, Hu X, Miller A, Mühlhaus T, Owen S, Rey M, Schneider G, Schneider J, Schneider K, Weil HL, Wetzels F (2025). BioHackEU24 report: Creating user benefit from ARC-ISA RO-Crate machine-actionability & Increasing FAIRness of digital agrosystem resources by extending Bioschemas. osf.io/x9d6b_v1, Preprint

Berndt, T, Walker, B, Qin, T, Stühmer, J, & Kormilitzin, A (2025). Permutation Equivariant Neural Controlled Differential Equations for Dynamic Graph Representation Learning. In The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS) 22484

Bestenlehner JM, Crowther PA, Bronner VA, Simón-Díaz S, Lennon DJ, Bodensteiner J, Langer N, Marchant P, Sana H, Schneider FRN, Shenar T (2025). Binarity at LOw Metallicity (BLOeM): pipeline-determined physical properties of OB stars. Monthly Notices of the Royal Astronomical Society 540(4):3523-3548

Bettisworth B, Stamatakis A (2025). bigrig: A range simulator for the DEC[+J] model. Preprint. https://doi.org/10.1101/.11.24.690345

Bettisworth B, Psonis N, Poulakakis N, Pavlidis P, Stamatakis A (2025). Read Length Dominates Phylogenetic Placement Accuracy of Ancient DNA Reads. Molecular Biology and Evolution 42(2), msaf006

Blaschke D, Röpke FK, Röpke G (2025). Universality and variability of the heavy r-process element abundance pattern from a nonequilibrium approach. Front. Astron. Space Sci. 12:1733496

Bodensteiner J, Shenar T, Sana H, Britavskiy N, Crowther PA, Langer N, Lennon DJ, Mahy L, Patrick LR, Villaseñor JI, Abdul-Masih M, Bowman DM, de Koter A, de Mink SE, Deshmukh K, Fabry M, Gilkis A, Götberg Y, Holgado G, Izzard RG, Janssens S, Kalari VM, Keszthelyi Z, Kubát J, Mandel I, Maravelias G, Oskinova LM, Pauli D, Ramachandran V, Rocha DF, Renzo M, Sander AAC, Schneider FRN, Schootemeijer A, Sen K, Stoop M, Toonen S, van Loon JTh, Valli R, Vigna-Gómez A, Vink JS, Wang C, Xu X-T (2025). Binarity at LOw Metallicity (BLOeM): Multiplicity properties of Oe and Be stars. A&A 698:A38

Bordadágua B, Ahlborn F, Coppée Q, Marques JP, Belkacem K, Hekker S (2025). The efficiency of mixed modes for angular momentum transport. A&A 699:A310

Bracher J, Littek JM (2025). A statistical assessment of influenza intensity thresholds from the moving epidemic and WHO methods. Journal of the Royal Statistical Society Series A: Statistics in Society, 188(4):1203–1230

Bracher J, Němcová B (2025). A unifying class of compound Poisson integer-valued ARMA and GARCH models. Scandinavian Journal of Statistics, 52(3):1176–1205

Brehmer JR, Kraus K, Gneiting T, Herrmann M, Marzocchi W (2025). Enhancing the statistical evaluation of earthquake forecasts—An application to Italy. Seismological Research Letters, 96(3):1966–1988

Britavskiy N, Mahy L, Lennon DJ, Patrick LR, Sana H, Villaseñor JI, Shenar T, Bodensteiner J, Bernini-Peron M, Berlanas SR, Bowman DM, Crowther PA, de Mink SE, Evans CJ, Götberg Y, Holgado G, Johnston C, Keszthelyi Z, Klencki J, Langer N, Mandel I, Menon A, Moe M, Oskinova LM, Pauli D, Pawlak M, Ramachandran V, Renzo M, Sander AAC, Schneider FRN, Schootemeijer A, Sen K, Simón-Díaz S, van Loon JTh, Vink JS (2025). Binarity at LOw Metallicity (BLOeM): Multiplicity of early B-type supergiants in the Small Magellanic Cloud. A&A 698:A40

Bronner VA, Laplace E, Schneider FRN, Podsiadlowski Ph (2025). Explosions of pulsating red supergiants: A natural pathway for the diversity of Type II-P/L supernovae. A&A 703:A61

Buchele L, Bellinger EP, Hekker S, Basu S (2025). Asteroseismic Structure Inversions of Main-sequence Solar-like Oscillators with Convective Cores. ApJ 987(1):97

Buchwald M, Memmesheimer P, Moghadam AD, Tuschner I, Suarez LS, Itzel T, Antoni C, Daza J, Gerhards C, Neumaier M, Brochhausen C, Galle PR, Ebert M, Weinmann A, Hesser J, Heuveline V, Teufel A (2025). Comparative evaluation of standard machine learning models for liver fibrosis detection. Z Gastroenterol 63(01):e13-e13

Callan FP, Collins CE, Sim SA, Shingles LJ, Pakmor R, Srivastav S, Pollin JM, Gronow S, Röpke FK, Seitenzahl IR (2025). Exploring the range of impacts of helium in the spectra of double detonation models for Type Ia supernovae. Monthly Notices of the Royal Astronomical Society 539(2):1404-1413

Callan FP, Holas A, Morán-Fraile J, Sim SA, Collins CE, Shingles LJ, Pollin JM, Röpke FK, Pakmor R, Schneider FRN (2025). NLTE spectral modelling for a carbon-oxygen and helium white dwarf merger as a Ca-rich transient candidate. A&A 702:A29

Chadwick E, Brouwer M, Schneider K, Arend D, Bacall F, Bardet E, Beier S, Brilhaus D, Hu X, Le Roy Pardonche E, Mühlhaus T, Owen S, Pommier C, Weil HL (2025). BioHackEU25 report: Towards a Robust Validation Service for Data and Metadata in ARC RO-Crates. osf.io/zah28_v1, Preprint

Chen J, Höhlein K, Lerch S (2025). Learning low-dimensional representations of ensemble forecast fields using autoencoder-based methods. Journal of Geophysical Research: Machine Learning and Computation, 2(4):eJH000819

Choi JY, Espinoza-Rojas F, Coppée Q, Hekker S (2025). Power density spectra morphologies of seismically unresolved red-giant asteroseismic binaries. A&A 699:A180

Collins CE, Shingles LJ, Sim SA, Callan FP, Gronow S, Hillebrandt W, Kromer M, Pakmor R, Röpke FK (2025). Non-LTE radiative transfer simulations: improved agreement of the double detonation with normal Type Ia supernovae. Monthly Notices of the Royal Astronomical Society 538(3):1289-1300

Das P, Seitenzahl IR, Ruiter AJ, Röpke FK, Pakmor R, Vogt FPA, Collins CE, Ghavamian P, Sim SA, Williams BJ, Taubenberger S, Laming JM, Suherli J, Sutherland R, Rodríguez-Segovia N (2025). Calcium in a supernova remnant as a fingerprint of a sub-Chandrasekhar-mass explosion. Nat Astron 9(9):1356-1365

Davín AA, Woodcroft BJ, Soo RM, Morel B, Murali R, Schrempf D, Clark JW, Álvarez-Carretero S, Boussau B, Moody ERR, Szánthó LL, Richy E, Pisani D, Hemp J, Fischer WW, Donoghue PCJ, Spang A, Hugenholtz P, Williams TA, Szöllősi GJ (2025). A geological timescale for bacterial evolution and oxygen adaptation. Science 388(6742),eadp1853

Deka M, Ahlborn F, Braun TAM, Weiss A (2025). Implications of a turbulent convection model for classical Cepheids. A&A 699:A351

Demianenko M, Afanasiev A, Rubtsov E, Toptun V, Pott J-U, Belinski A, Bauer FE, Chilingarian IV, Grishin KA, Burlak M, Ikonnikova N (2025). Broad line region echo from highly accreting intermediate-mass black hole candidate SDSS J144850.08+160803.1:First probe of intra-night variability and reverberation mapping. A&A 702:A202

Doser B, Polsterer KL, Fehlner A, Trujillo-Gomez S (2025) Machine Learning Workflow for Morphological Classification of Galaxies. arXiv preprint arXiv:2505.04676

Egressy, B, & Stühmer, J (2025). Set-LLM: A permutation-invariant LLM. In The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS) 2247

Erozan A, Zeilmann A, Chatzimpinou A, Annesha FM, Heuveline V, Weinhardt V (2025) Spatially-Informed Multi-Stage Active Learning for Segmentation of Nucleus in Soft X-ray Tomography. In: 2025 IEEE Symposium on Computational Intelligence in Image, Signal Processing and Synthetic Media (CISM). IEEE, Trondheim, Norway, pp 1–7

Espinoza-Rojas F, Themeßl N, Hekker S (2025). Red giant asteroseismic binaries in the Kepler field. A&A 703:A66

Fan Y, Strube M (2025). Consistent discourse-level temporal relation extraction using large language models. In: Findings of the Association for Computational Linguistics: EMNLP, Suzhou, China, 5-9 November , pp.18605-18622

Ferrand G, Pakmor R, Fujimaru Y, Lee S-H, Safi-Harb S, Nagataki S, Röpke FK, Decourchelle A, Seitenzahl IR, Patnaude D (2025). The Role of the Secondary White Dwarf in a Double-degenerate Double-detonation Explosion, in the Supernova Remnant Phase. ApJ 995(1):85

Ferraz MVF, Adan WCS, Lima TE, Santos AJC, de Paula SO, Dhalia R, Wallau GL, Wade RC, Viana IFT, Lins RD (2025). Design of nanobody targeting SARS-CoV-2 spike glycoprotein using CDR-grafting assisted by molecular simulation and machine learning. PLoS Comput Biol 21(4):e1012921

Fluck J, Darms J, Golebiewski M, Schmidt CO, Vorisek CN, Löbe M, Sax U, Pigeot I (2025) Challenges and opportunities of integrating NFDI4Health into the European Health Data Space. Zenodo. https://zenodo.org/doi/10.5281/zenodo.16736285

Gagnier D, Pejcha O (2025). Journey to the center of the common envelope evolution. A&A 697:A68

Gavriilidou A, Stamatakis A, Kupczok A, Bista I, Jiggins CD, Fernández R, Skourtanioti E, Amoutzias G, Delneri D, Kyrpides N, Nikolaou C, Pittis AA, Manousaki T, Vakirlis N (2025). Advances and challenges in understanding evolution through genome comparison: meeting report of the European Molecular Biology Organization (EMBO) lecture course "Evolutionary and Comparative Genomics". Bioinformatics Advances 5(1),vbaf223

Gilkis A, Laplace E, Arcavi I, Shenar T, Schneider FRN (2025). The landscape of binary core-collapse supernova progenitors and the late emergence of Wolf–Rayet winds. Monthly Notices of the Royal Astronomical Society 540(4):3094-3120

Golebiewski M, Mayer G (2025). Data Formats for Systems Biology, Systems Medicine and Computational Modeling. In: Encyclopedia of Bioinformatics and Computational Biology (Second Edition). Elsevier, p. 321–333

Golebiewski M, Fürst J, Kasbohm E, Lindstädt B, Prause G, Schmidt C, Vorisek C, NFDI4Health (2025). How to make your study data FAIR in the health domain: Standardisation and quality recommendations. PUBLISSO. https://repository.publisso.de/resource/frl:6505767

Grisoni V, Rizzuti F, Cescutti G (2025). Fluorine evolution in the Galactic halo. A&A 704:A45

Gull M, Weisz DR, El-Badry K, Henneco J, Savino A, Durbin M, Choi Y, Cohen RE, Cole AA, Correnti M, Dalcanton JJ, Gilbert KM, Goldman SR, Guhathakurta P, McQuinn KBW, Newman MJB, Skillman ED, Williams BF (2025). A Low Metallicity Massive Contact Binary Star System Candidate in WLM Identified by Hubble and James Webb Space Telescope Imaging. ApJ 986(1):25

Haag J, Stamatakis A (2025). Pythia 2.0: New Data, New Prediction Model, New Features. biorxiv;.03.25.645182v1, Preprint

Haag J, Jordan AI, Stamatakis A (2025). Pandora: a tool to estimate dimensionality reduction stability of genotype data. Bioinformatics Advances, 5(1):vbaf040

Henneco J, Schneider FRN, Laplace E (2025). Contact tracing of binary stars: Pathways to stellar mergers (Corrigendum). A&A 702:C4

Henneco J, Schneider F, Heller M, Hekker S, Aerts C (2025) Massive Merger Products on the Main Sequence: Do Their Pulsations Betray Them? Zenodo. https://doi.org/10.5281/ZENODO.17379594

Henneco J, Schneider FRN, Heller M, Hekker S, Aerts C (2025). Asteroseismic predictions for a massive main-sequence merger product. A&A 698:A49

Hoffmann DS, Dohmen PM, Sokolov M, Kleinekathöfer U, Elstner M (2025). Exciton Transfer Simulations in a Light-Harvesting 2 Complex Reveal the Transient Delocalization Mechanism. J. Phys. Chem. B 129(13):3345-3365

Holas A, Röpke FK, Pakmor R, Callan FP, Pollin J, Sim SA, Collins CE, Shingles LJ, Moran-Fraile J (2025). The asymmetry of white dwarf double detonations and the observed scatter around the Phillips relation. A&A 698:A269

Horat N, Klerings S, Lerch S (2025). Improving model chain approaches for probabilistic solar energy forecasting through post-processing and machine learning. Advances in Atmospheric Sciences, 42(2):297–312

Hu X, Owen S, Meineke F, Bacall F, Goble C, Müller W, Golebiewski M (2025) Flexible Metadata Structuring for Research Data Management Through the FAIRDOM-SEEK Platform - Implementing Tailored and Complex Metadata Schemes in SEEK. Zenodo. https://zenodo.org/doi/10.5281/zenodo.16736322

Huang W-C, Dombrowski N, Mahendrarajah TA, Wahl N, Stamatakis A, Szollosi GJ, Williams TA, Spang A (2025). Phylogenetic reconciliation supports a methanogenic ancestor of the Archaea and a derived origin for host-associated lineages. Preprint https://doi.org/10.1101/.11.11.687807

Häuser L, List J-M (2025) Lexibench: towards an improved collection of benchmark data for computational historical linguistics. https://doi.org/10.15475/CALCIP.2025.1.2

Häuser L, Stamatakis A (2025) The Cognate Data Bottleneck in Language Phylogenetics. arXiv preprint arXiv:2507.00911

Jeon S, Strube M (2025). Entity tracking in small language models: An attention-based study of parameter-efficient fine-tuning. Entity tracking in small language models: An attention-based study of parameter-efficient fine-tuning. In Proceedings of the 4th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences, Suzhou, China, 9 November, pp.42-53

Kasbohm E, Inau ET, Zeleke AA, Shiferaw KB, Abaza H, Clemens V, Golebiewski M, Gonzalez-Ocanto M, Klopfenstein SAI, Vorisek CN, Waltemath D, Schmidt CO (2025). Lessons learned from evaluating the NFDI4Health metadata schema for health studies using the Research Data Alliance's FAIR data maturity model. Preprint. https://doi.org/10.1101/.09.26.25336637

Kehr D, Ritterhoff J, Glaser M, Jarosch L, Salazar RE, Spaich K, Varadi K, Birkenstock J, Egger M, Gao E, Koch WJ, Sauter M, Freichel M, Katus HA, Frey N, Jungmann A, Busch C, Mather PJ, Ruhparwar A, Busch M, Völkers M, Wade RC, Most P (2025). S100A1ct: A Synthetic Peptide Derived From S100A1 Protein Improves Cardiac Performance and Survival in Preclinical Heart Failure Models. Circulation 151(8):548-565

Kollmeier JA, Rix H-W, Aerts C, Aird J, Vera Alfaro P, Almeida A, Anderson SF, Arseneau SM, Assef RJ, Aviram S, Aydar C, Badenes C, Bandyopadhyay A, Barger K, Barkhouser RH, Bauer FE, Behmard A, Bender C, Besser F, Bhattarai B, Bilgi P, Bird J, Bizyaev D, Blanc GA, Blanton MR, Bochanski J, Bovy J, Brandon C, Brandt WN, Brownstein JR, Buchner J, Burchett JN, Carlberg J, Casey AR, Castaneda-Carlos L, Chakraborty P, Chanamé J, Chandra V, Cherinka B, Chilingarian I, Comparat J, Cosens M, Covey K, Crane JD, Crumpler NR, Cruz-Gonzalez I, Cunha K, Cunningham T, Dai X, Darling J, Davidson JW Jr, Davis MC, De Lee N, Deacon N, Méndez Delgado JE, Demasi S, Demianenko M, Derwent M, D'Onghia E, Di Mille F, Dias B, Donor J, Dow PN, Drory N, Dwelly T, Egorov O, Egorova E, El-Badry K, Engelman M, Eracleous M, Fan X, Farr E, Fries L, Frinchaboy P, Froning CS, Gänsicke BT, García P, Gelfand J, Gentile Fusillo NP, Glover S, Grabowski K, Grebel EK, Green PJ, Grier C, Gupta P, Gray AC, Häberle M, Hall PB, Hammond RP, Hawkins K, Harding AC, Hegedűs V, Herbst T, Hermes JJ, Rodríguez Hidalgo P, Hilder T, Hogg DW, Holtzman JA, Horta D, Huang Y, Hwang H-C, Ibarra-Medel HJ, Imig J, Inight K, Jana A, Ji AP, Jiménez-Arranz Ó, Jofre P, Johns M, Johnson J, Johnson JW, Johnston EJ, Jones AM, Katkov I, Knapp GR, Koekemoer AM, Kounkel M, Kreckel K, Krishnarao D, Krumpe M, Kumari N, Kupfer T, Lacerna I, Laporte C, Lepine S, Li J, Liu X, Loebman S, Long K, Roman-Lopes A, Lu Y, Majewski SR, Maoz D, McKinnon KA, Medan I, Merloni A, Minniti D, Morrison S, Myers N, Mészáros S, Nandra K, Nayak PK, Ness MK, Nidever DL, O'Brien T, Oeur M, Oravetz A, Oravetz D, Otto J, Pallathadka GA, Palunas P, Pan K, Pappalardo D, Pandey R, Negrete Peñaloza CA, Pinsonneault MH, Pogge RW, Taghizadeh Popp M, Price-Whelan AM, Pulatova N, Qiu D, Ramirez S, Rankine A, Ricci C, Runnoe JC, Sanchez S, Salvato M, Sarbadhicary SK, Sattler N, Saydjari AK, Sayres C, Schinnerer E, Schlaufman KC, Schneider DP, Schreiber MR, Schwope A, Serna J, Shen Y, Sifón C, Singh A, Sinha A, Smee S, Song Y-Y, Souto D, Stassun KG, Steinmetz M, Stone-Martinez A, Stringfellow G, Stutz A, Sánchez-Gallego J, Tan JC, Tayar J, Thai R, Thakar A, Ting Y-S, Tkachenko A, Tovmassian G, Trakhtenbrot B, Fernández-Trincado JG, Troup N, Trump J, Tuttle S, van der Marel RP, Villanova S, Villaseñor J, Wachter S, Way Z, Weijmans A-M, Weinberg D, Wheeler A, Wiggins AI, Wilson J, Wofford A, Wong T, Wu Q, Wylezalek D, Wyse RFG, Xue X-X, Yang Q, Ybarra J, Zakamska N, Zari E, Zasowski G, Zeltyn G, Zucker C, Román-Zúñiga CG, de J. Zermeño R (2025). Sloan Digital Sky Survey. V. Pioneering Panoptic Spectroscopy. AJ 171(1):52

Laplace E, Schneider FRN, Podsiadlowski Ph (2025). It's written in the massive stars: The role of stellar physics in the formation of black holes. A&A 695:A71

Larsen JR, Rørsted JL, Aguirre Børsen-Koch V, Lundkvist MS, Christensen-Dalsgaard J, Winther ML, Stokholm A, Li Y, Slumstrup D, Kjeldsen H, Corsaro E, Benomar O, Dhanpal S, Weiss A, Mosser B, Hekker S, Stello D, Korn AJ, Jendreieck A, Elsworth Y, Handberg R, Kallinger T, Jiang C, Ruchti G (2025). Pushing the boundaries of asteroseismic individual frequency modelling: Unveiling two evolved very low-metallicity red giants. A&A 697:A153

Lau MYM, Hirai R, Price DJ, Mandel I, Bate MR (2025) Common envelopes in massive stars: III. The obstructive role of radiation transport in envelope ejection. A&A 699:A274

Lau MYM, Cantiello M, Jermyn AS, MacLeod M, Mandel I, Price DJ (2025). Hot Jupiter engulfment by an early red giant in 3D hydrodynamics. A&A 694:A264

van Lier T, Müller J, Hekker S (2025). Weak or strong: Coupling of mixed oscillation modes on the red giant branch. A&A 700:A1

Liu W, Strube M (2025). Joint modeling of entities and discourse relations for coherence assessment. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Suzhou, China, 5-9 November, pp.21910-21926

Liu W, Strube M (2025). Discourse relation-enhanced neural coherence modeling. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vienna, Austria, July 27 to August 1, pp.4748-4762

Liu W, Trenous S, Ribeiro LF, Byrne B, Hieber F (2025). XRAG: Cross-lingual retrieval-augmented generation. In: Findings of the Association for Computational Linguistics: EMNLP, Suzhou, China, 5-9 November ,pp.15669-15690

Ludwig M, Machmeier S, Heuveline V (2025) Cross-Study Comparability for IDS: A Call for Clarity. In: 2025 9th Cyber Security in Networking Conference (CSNet). IEEE, Abu Dhabi, United Arab Emirates, pp 1–8

Löbe M, Golebiewski M, Sax U (2025) Profiling W3C DCAT for use in specific domains - Experience with the HealthDCAT-AP draft standard for the European Health Data Space (EHDS). Zenodo. https://zenodo.org/doi/10.5281/zenodo.16736359

Machmeier S, Ludwig M, Fuchs M, Heuveline V (2025). heiDGAF: A Novel Open-Source CIDS Solution for DGA Detection

Maltsev K, Schneider FRN, Mandel I, Müller B, Heger A, Röpke FK, Laplace E (2025). Explodability criteria for the neutrino-driven supernova mechanism. A&A 700:A20

Mandal AK, Pozo Nuñez F, Jaiswal VK, Naddaf MH, Czerny B, Panda S, Karczmarek P, Pietrzyński G, Pandey S, Peterson BM, Zajaček M, Dovčiak M, Karas V, Narloch W, Kicia M, Górski M, Kałuszyński M, Hajdu G, Wielgórski P, Zgirski B, Gałan C, Pych W, Smolec R, Bąkowska K, Gieren W, Kervella P (2025). HALO. I. Photometric continuum reverberation mapping of Fairall 9. A&A

Mandel I, Riley J, Boesky A, Brček A, Hirai R, Kapil V, Lau MYM, Merritt J, Rodríguez-Segovia N, Romero-Shaw I, Song Y, Stevenson S, Vajpeyi A, van Son LAC, Vigna-Gómez A, Willcox R (2025). Rapid Stellar and Binary Population Synthesis with COMPAS: Methods Paper II. ApJS 280(1):43

Meineke F, Hu X, Golebiewski M, Löbe M (2025). Roll-Out of the NFDI4health Local Data Hubs for Medical Research Data Management. In: Intelligent Health Systems – From Technology to Data and Knowledge. IOS Press

Meineke F, Golebiewski M, Hu X, Hänsel R, Abedi M, Müller W, Löbe M (2025) Rollout of the NFDI4Health Local Data Hubs - First Results from Call 2024/25. Zenodo. https://zenodo.org/doi/10.5281/zenodo.16735781

Mertová L, Polreich S, Lewkowski O, Müller W (2025). The BeeProject: Advanced Digitisation and Creation of a Dataset for the Monitoring of Beehives. In The 2024 ACM/IEEE Joint Conference on Digital Libraries (JCDL '24), December 16–20, 2024, Hong Kong, China. ACM, New York, NY, USA, 11 pages

Muñiz-Chicharro A, Ganotra GK, Wade RC (2025). A Multiscale Simulation Approach to Compute Protein–Ligand Association Rate Constants by Combining Brownian Dynamics and Molecular Dynamics. J. Chem. Inf. Model. 65(20):11215-11231

Müller J, Coppée Q, Hekker S (2025). Oscillations of red giant stars with magnetic damping in the core. A&A 696:A134

Nag P, Hong Y, Abdulah S, Qadir GA, Genton MG, Sun Y (2025). Efficient large-scale nonstationary spatial covariance function estimation using convolutional neural networks. Journal of Computational and Graphical Statistics, 34(2):683–696

Nguyen CT, Cescutti G, Matteucci F, Rizzuti F, Mucciarelli A, Romano D, Magrini L, Korn AJ, Bressan A, Girardi L (2025). Lithium evolution in the Galactic thin disc from main-sequence and early red-giant-branch stars. A&A 703:A204

Noll A, Basu S, Hekker S (2025). Ensemble seismic study of the properties of the core of red clump stars. A&A 704:A188

Pakmor R, Shen KJ, Bhat A, Rajamuthukumar AS, Collins CE, O'Donnell C, Bauer EB, Callan FP, Röpke FK, Pollin JM, Maguire K, Kwok LA, Seth R, Taubenberger S, Justham S (2025). Violent mergers revisited: The origin of the fastest stars in the Galaxy. arXiv preprint arXiv:2510.11781

Panecka-Hofman J, Linciano P, Pöhner I, Dyguda-Kazimierowicz E, Jedwabny W, Landi G, Santarem N, Witt G, Ellinger B, Kuzikov M, Luciani R, Ferrari S, Aiello D, Mangani S, Pozzi C, Cordeiro-da-Silva A, Gul S, Costi MP, Wade RC (2025). Design of 2-Aminobenzothiazole Derivatives Targeting Trypanosomatid PTR1 by a Multidisciplinary Fragment Hybridization Approach. J. Med. Chem. 68(19):20595-20618

Patrick LR, Lennon DJ, Najarro F, Shenar T, Bodensteiner J, Sana H, Crowther PA, Britavskiy N, Langer N, Schootemeijer A, Evans CJ, Mahy L, Götberg Y, de Mink SE, Schneider FRN, O'Grady AJG, Villaseñor JI, Bernini-Peron M, Bowman DM, de Koter A, Deshmukh K, Gilkis A, González-Torà G, Kalari VM, Keszthelyi Z, Mandel I, Menon A, Moe M, Oskinova LM, Pauli D, Renzo M, Sander AAC, Sen K, Stoop M, van Loon JTh, Toonen S, Tramper F, Vink JS, Wang C (2025). Binarity at LOw Metallicity (BLOeM): The multiplicity properties and evolution of BAF-type supergiants. A&A 698:A39

Pinsonneault MH, Zinn JC, Tayar J, Serenelli A, García RA, Mathur S, Vrard M, Elsworth YP, Mosser B, Stello D, Bell KJ, Bugnet L, Corsaro E, Gaulme P, Hekker S, Hon M, Huber D, Kallinger T, Cao K, Johnson JA, Liagre B, Patton RA, Santos ÂRG, Basu S, Beck PG, Beers TC, Chaplin WJ, Cunha K, Frinchaboy PM, Girardi L, Godoy-Rivera D, Holtzman JA, Jönsson H, Mészáros S, Reyes C, Rix H-W, Shetrone M, Smith VV, Spoo T, Stassun KG, Wang J (2025). APOKASC-3: The Third Joint Spectroscopic and Asteroseismic Catalog for Evolved Stars in the Kepler Fields. ApJS 276(2):69

Psonis N, Tabakaki E, Vassou D, Papadantonakis S, Souleles A, Nafplioti A, Kousis Tsampazis G, Papadopoulou A, Xanthopoulos K, Panailidis P, Georgiadou A, Papakosta D, Koursioti S, Evangelinou M, Papadopoulou V, Evaggeloglou P, Korka E, Christidis I, Ioannou M, Kontogianni T, Arkoumanis A, Stamatakis A, Poulakakis N, Papageorgopoulou C, Pavlidis P (2025). Genetic affinities between an ancient Greek colony and its metropolis: the case of Amvrakia in western Greece. biorxiv;.07.01.662689v1, Preprint

Pulatova NG, Rubtsov E, Chilingarian IV, Rix H-W, Demianenko M, Grishin KA, Katkov IYu, Schneider DP, Aydar C, Buchner J, Salvato M, Merloni A, Koekemoer AM, Assef RJ, Ricci C, Wylezalek D, Gasymov D, Nielsen Brandt W, Negrete Peñaloza CA, Morrison S, Anderson SF, Bauer FE, Ibarra-Medel HJ, Wu Q (2025). Optical emission line properties of eROSITA-selected SDSS-V galaxies. A&A 702:A67

Qadir GA, Sun Y (2025). Modeling and predicting spatio-temporal dynamics of PM_2.5 concentrations through time-evolving covariance models. Statistica Sinica, 35:1649–1669

Reinhardt M, Bruce NJ, Kokh DB, Wade RC (2025). Correction to "Brownian Dynamics Simulations of Proteins in the Presence of Surfaces: Long-Range Electrostatics and Mean-Field Hydrodynamics". J. Chem. Theory Comput. 21(15):7724-7725

Remme R, Kaczun T, Ebert T, Gehrig CA, Geng D, Gerhartz G, Ickler MK, Klockow MV, Lippmann P, Schmidt JS, Wagner S, Dreuw A, Hamprecht FA (2025). Stable and Accurate Orbital-Free Density Functional Theory Powered by Machine Learning. J. Am. Chem. Soc. 147(32):28851-28859

Röpke G, Blaschke D, Röpke FK (2025). Distribution of Heavy-Element Abundances Generated by Decay from a Quasi-Equilibrium State. Universe 11(10):323

Ruiter AJ, Seitenzahl IR (2025). Type Ia supernova progenitors: a contemporary view of a long-standing puzzle. Astron Astrophys Rev 33(1),1

Sana H, Shenar T, Bodensteiner J, Britavskiy N, Langer N, Lennon DJ, Mahy L, Mandel I, de Mink SE, Patrick LR, Villaseñor JI, Dirickx M, Abdul-Masih M, Almeida LA, Backs F, Berlanas SR, Bernini-Peron M,

Bowman DM, Bronner VA, Crowther PA, Deshmukh K, Evans CJ, Fabry M, Gieles M, Gilkis A, González-Torà G, Gräfener G, Götberg Y, Hawcroft C, Hénault-Brunet V, Herrero A, Holgado G, Izzard RG, de Koter A, Janssens S, Johnston C, Josiek J, Justham S, Kalari VM, Klencki J, Kubát J, Kubátová B, Lefever RR, van Loon JTh, Ludwig B, Mackey J, Maíz Apellániz J, Maravelias G, Marchant P, Mazeh T, Menon A, Moe M, Najarro F, Oskinova LM, Ovadia R, Pauli D, Pawlak M, Ramachandran V, Renzo M, Rocha DF, Sander AAC, Schneider FRN, Schootemeijer A, Schösser EC, Schürmann C, Sen K, Shahaf S, Simón-Díaz S, van Son LAC, Stoop M, Toonen S, Tramper F, Valli R, Vigna-Gómez A, Vink JS, Wang C, Willcox R (2025). A high fraction of close massive binary stars at low metallicity. Nat Astron 9(9):1337-1346

Schiappacasse-Ulloa J, Magrini L, Lucatello S, Randich S, Bragaglia A, Carretta E, Cescutti G, Rizzuti F, Worley C, Lucertini F, Berni L (2025). Neutron-capture element signatures in globular clusters:Insights from the Gaia-ESO Survey. A&A 699:A41

Schneider FRN (2025). Theory, Simulations and Observations of Stellar Mergers. arXiv preprint arXiv:2509.18421

Schneider FRN, Laplace E, Podsiadlowski Ph (2025). Supernovae from stellar mergers and accretors of binary mass transfer: Implications for Type IIP, 1987A-like and interacting supernovae. A&A 700:A253

Schopmans H, Friederich P (2025). Temperature-Annealed Boltzmann Generators. arXiv preprint arXiv:2501.19077

Schweder F, Trujillo-Gomez S, Polsterer K (2025). JAvaScript Multimodal INformation Explorer. arXiv preprint arXiv: 2504.21393

Schütt I, Einwohlt P, Wandinger A-M, Teuffel J, Wirsing R, Hlawatschke BH, Fehlauer HL, Bothe S, Bader N, Monaci E, Kreikemeyer B, Schindelin H, Wade RC, Fiedler T (2025). Inhibitors of GapN-dependent NADPH supply as potential lead compounds for novel therapeutics against Streptococcus pyogenes. Virulence 17(1):2609393

Shukla N, Romeo A, Caravita C, Riha L, Vysocky O, Strakos P, Jaros M, Barbosa J, Vavrik R, Mignone A, Rossazza M, Truzzi S, Berta V, Colonnelli I, Medić D, Boella E, Gregori D, Sciacca E, Tornatore L, Taffoni G, J. Deka P, Bacchini F, Wilhelm R-P, Doulis G, Pierre K, Rezzolla L, Colman T, Commerçon B, Bouizi O, Kuhn M, Raffin E, Sergent M, Wissing R, Marin G, Dolag K, S. Karademir G, Perna G, Zanotti M, Trujillo-Gomez S (2025). EuroHPC SPACE CoE: Redesigning Scalable Parallel Astrophysical Codes for Exascale. Invited Paper. In: Proceedings of the 22nd ACM International Conference on Computing Frontiers: Workshops and Special Sessions. ACM, Cagliari Italy, pp 177-184

Sitnova TM, Lombardo L, Mashonkina LI, Rizzuti F, Cescutti G, Hansen CJ, Bonifacio P, Caffau E, Koch-Hansen A, Meynet G, Fernandes de Melo R (2025). Unlocking the mystery of strontium synthesis in the early Galaxy through analysis of barium isotopes in very metal-poor stars. A&A 699:A262

Soultanis T, Maltsev K, Bauswein A, Chatziioannou K, Röpke FK, Stergioulas N (2025). Gravitational-wave model for neutron star merger remnants with supervised learning. Phys. Rev. D 111(2),023002

Stelz C, Hübner L, Stamatakis A (2025). Bit-Reproducible Phylogenetic Tree Inference under Varying Core-Counts via Reproducible Parallel Reduction Operators. biorxiv;.06.02.656320v1, Preprint

Strube M, Braud C, Hardmeier C, Li JJ, Loaciga S, Zeldes A, Li C (2025). Proceedings of the 6th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI), Suzhou, China, 9 November

Teuffel J, Mukherjee G, Han SB, Elstner M, Wade RC (2025). On the determinants of electron transfer reorganization energy in a cytochrome P450: cytochrome b5 complex. A combined quantum mechanics and molecular dynamics simulation study. The Journal of Chemical Physics 162(19):195101

Togkousidis A, Stamatakis A, Gascuel O (2025). Accelerating Maximum Likelihood Phylogenetic Inference via Early Stopping to Evade (Over-) optimization. Systematic Biology, 74(6), 1020-1028

Togkousidis A, Gascuel O, Stamatakis A (2025). A Systematic Investigation of Overfitting in Maximum Likelihood Phylogenetic Inference. biorxiv;.10.07.680876v1, Preprint

Ulanov E, Qadir GA, Riedmiller K, Friederich P, Gräter F (2025). Predicting hydrogen atom transfer energy barriers using Gaussian process regression. Digital Discovery, 4(2):513–522

Uttarwar SB, Lerch S, Avesani D, Majone B (2025). Performance assessment of neural network models for seasonal weather forecast postprocessing in the Alpine region. Advances in Water Resources, 204:105061

Vetter M, Röpke FK, Schneider FRN, Pakmor R, Ohlmann S, Morán-Fraile J, Lau MYM, Leidi G, Gagnier D, Andrassy R (2025). Magnetically driven outflows in the 3D common envelope evolution of massive stars. A&A 698:A133

Viliuga, V, Seute, L, Wolf, N, Wagner, S, Elofsson, A, Stühmer, J, & Gräter, F (2025). Flexibility-conditioned protein structure design with flow matching. Proceedings of the 42nd International Conference on Machine Learning (ICML) 2250

Villaseñor JI, Sana H, Mahy L, Shenar T, Bodensteiner J, Britavskiy N, Lennon DJ, Moe M, Patrick LR, Pawlak M, Bowman DM, Crowther PA, de Mink SE, Deshmukh K, Evans CJ, Fabry M, Fouesneau M, Holgado G, Langer N, Maíz Apellániz J, Mandel I, Oskinova LM, Pauli D, Ramachandran V, Renzo M, Rix H-W, Rocha DF, Sander AAC, Schneider FRN, Sen K, Simón-Díaz S, van Loon JTh, Toonen S, Vink JS (2025). Binarity at LOw Metallicity (BLOeM): Enhanced multiplicity of early B-type dwarfs and giants at $Z = 0.2\ Z_\odot$. A&A 698:A41

Vitovský K, Brož M (2025). Hydrodynamical models of the β Lyr A circumstellar disc. A&A 704:A189

Vorisek CN, Hölter T, Saß J, Shutsko A, Kasbohm E, Golebiewski M, Schmidt CO, Thun S, Klopfenstein S (2025). Standardizing Metadata in NFDI4Health: A FHIR Implementation Guide for Interdisciplinary Health Research. German Medical Science GMS Publishing House. https://conferences.publisso.de/en/conferences/gmds/25gmds110

Vrard M, Pinsonneault MH, Elsworth Y, Hon M, Kallinger T, Kuszlewicz J, Mosser B, García RA, Tayar J, Bennett R, Cao K, Hekker S, Loyer L, Mathur S, Stello D (2025). Red giant evolutionary status determination: The complete Kepler catalog. A&A, 697:A165

Wahl NA, Koutsovoulos G, Bettisworth B, Stamatakis A (2025) raxtax : a k -mer-based non-Bayesian taxonomic classifier. Bioinformatics 41(12):btaf620

Walter LJ, Bereau T (2025). Navigating chemical space: multi-level Bayesian optimization with hierarchical coarse-graining. Chem. Sci.,10.1039.D5SC03855C

Wan S, Liu W, Strube M (2025). On the role of context for discourse relation classification in scientific writing. In Proceedings of the 4th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences, Suzhou, China, 9 November, pp.96-106

Willcox R, Schneider FRN, Laplace E, Podsiadlowski P, Maltsev K, Mandel I, Marchant P, Sana H, Li T, Hertog T (2025). New gravitational-wave data support a bimodal black-hole mass distribution. arXiv preprint arXiv:2508.20787

Wolf, N, Seute, L, Viliuga, V, Wagner, S, Stühmer, J, & Gräter, F (2025). Learning conformational ensembles of proteins based on backbone geometry. In The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS) 2249

Yu J, Gehan C, Hekker S, Bazot M, Cameron RH, Gaulme P, Bedding TR, Murphy SJ, Han Z, Ting Y-S, Tayar J, Chen Y, Gizon L, Nordhaus J, Bi S (2025). Enhanced magnetic activity in rapidly rotating binary stars. Nat Astron 9, 1045–1052. https://doi.org/10.1038/s41550-025-02562-2

Zeilmann A, Schnetter E, Heuveline V (2025). Interactive AI on bwHPC -- Lessons Learned from Building a Large-Scale Image Analysis Platform. Proceedings of the 11th bwHPC Symposium

Zhang Y, Kurtz A, Zhao H, Wang L, Hao J, Golebiewski M, Stacey GN, Zhao T (2025). Development of Stem Cell Data Systems and Associated Data Standards for Cell Therapy. Adv Exp Med Biol.;1486:363-374

9 Teaching

9.1 Degrees

Anne Beck:
"Evaluating Defense Strategies for Adversarial Attacks against Face Recognition Systems", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Mattis Bodynek:
"Quantifying Uncertainty in Protein Multiple Sequence Alignment", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2025).

Bjela Böttcher:
"On the Influence of Stellar Model Parameters on the Sound-Speed Profile", Bachelor's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Saskia Hekker (2025).

Anna Carsi:
"Classification of Large Image Data using Multiscale Resolution Adjustment", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Nam Do:
"Extending Probabilistic k-mer Methods for Taxonomic Binning", Bachelor's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2025).

Lucas Eekhof:
"Investigating the mass dependent morphology of the red-giant branch bump", Master's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Saskia Hekker (2025).

Rostislav Fedorov:
"Machine Learning for Supported Organic Electrode Materials", Ph.D. thesis, Combined Faculty of Mathematics, Engineering and Natural Sciences, Heidelberg University and HITS: Ganna Gryn'ova (2025).

Maximilian Fidlin:
"Coarse-grained Molecular Dynamics simulations of polymethacrylate-based polyplexes", Bachelor's thesis, Faculty of Engineering Sciences, Heidelberg University and HITS: Fabian Grünewald (2025).

Manuel Fuchs:
"Benchmarking an Open-Source distributed Intrusion Detection System for DNS Attacks", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Clemens Glomb:
"Evaluation von Detektionsverfahren für DNS-basierte Angriffe", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Jonas Granzow:
"Non-adiabatic solar-like oscillations", Bachelor's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Saskia Hekker (2025).

Jan Henneco:
"The Progenitors and Products of Stellar Mergers", Ph.D. thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Fabian Schneider (2025).

Nico Herb:
"Prediction of Binding Kinetics of the Interleukin-13 – Interleukin-13 Receptor α1 Complex", Master's thesis, Molecular and Cellular Biology, Faculty of Biosciences, Heidelberg University and HITS: Mislav Brajkovic and Rebecca C. Wade (2025).

Marvin Isiksal:
"Deployment and Analysis of Mid- and High-Interaction Honeypots in a Restricted Network", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Jan Jakob:
"Operator Learning Using Random Features for Phase-Field Tumor Models", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Wei Liu:
"Linguistically-Inspired Neural Coherence Modeling", Ph.D. thesis, Heidelberg University and HITS: Michael Strube (2025).

Lukrécia Mertová:
"Towards semiautomatic curation of multi- modal data in life science publications", Ph.D. thesis, Faculty Information Systems and Applied Computer Sciences, Otto-Friedrich-Universität Bamberg: Andreas Henrich and Wolfgang Müller (HITS) (2025).

Ghadeer Mobasher:
"WeLT: Weighted Loss Trainer for Biomedical Joint Entity and Relation Extraction", Ph.D. thesis, Combined Faculty of Mathematics, Engineering and Natural Sciences, Heidelberg University and HITS: Wolfgang Müller (2025).

Philipp Müller-Roosen:
"Detecting QUIC attacks using machine learning", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Abraham Muniz Chicharro:
"Prediction of drug-protein binding kinetics", Ph.D. thesis, Combined Faculty of Mathematics, Engineering and Natural Sciences, Heidelberg University and HITS: Rebecca C. Wade (2025).

Vijayalakshmi Vijayakumaran Nair:
"Machine Learning-based Emulators for Equation of State Used in Astrophysical Simulations", Master's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Friedrich Röpke (2025).

David Nguyen:
"3D Gaussian Splatting für volumetrische Tomographie-Daten", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Bogdan Patenko:
"Neue Ansätze zur Effizienzoptimierung von heiDPI", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Filippa Piazolo:
"Galaxy Image Generation conditioned on Morphological Measurements using Latent Diffusion Models", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Ruppert Quentin:
"Validation of Public Proxy Servers using Internet-Wide-Scanning", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Henri Claude Rebitzky:
"Evaluation of Image-Based intrusion detection in industrial control systems", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Leila Said:
"Evaluating Datasets for Clustering Methods to Identify Anomalies in Network Traffic", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Clara Schlitter:
"Lipid accumulation around proteins in hybrid lipid polymer membranes", Bachelor's thesis, Faculty of Engineering Sciences, Heidelberg University and HITS: Fabian Grünewald (2025).

Bálint Soproni:
"Evaluating and Optimizing GROMACS by Leveraging the AdaptiveCpp SSCP Compiler", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Christoph Stelz:
"Advanced Substitution Model Selection Methods in RAxML-NG", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2025).

Adrian Thessmann:
"Asteroseismic Analysis and Stellar Modeling of KIC 6603624", Bachelor's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Saskia Hekker (2025).

Manuel Trageser:
"Designing and Evaluationg Autonomous PenTest Frameworks Driven by LLM Agents", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Alexandros Tsengenes:
"Computer-Aided Drug Design of Small-Molecule Neurotrophin Mimetics", Ph.D. thesis, Combined Faculty of Mathematics, Engineering and Natural Sciences, Heidelberg University and HITS: Rebecca C. Wade (2025).

Tobias van Lier:
"Stellar Harmony: Asymptotic Oscillation Parameters for Red-Giant Branch Stars", Master's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Saskia Hekker (2025).

Renuka Velu:
"Rotational Invariance for Galaxy Morphology Classification", Master's thesis, Interdisciplinary Center for Scientific Computing (IWR Heidelberg), Heidelberg University and HITS: Kai Polsterer (2025).

Freyja Walberg:
"KYann: an accelerated nuclear network for stellar evolution", Master's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Friedrich Röpke (2025).

Binwu Wang:
"Machine Learning-Based Multimodal Hepatocellular Carcinoma Data Analysis", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2025).

Anastasia Warken:
"Computational Investigation of Quinoline-Based Compounds and their Interaction with the Plasmodium Falciparum Drug Resistance Transporter", Bachelor's thesis, Molecular and Cellular Biology, Faculty of Biosciences, Heidelberg University and HITS: Kushal Singh and Rebecca C. Wade (2025).

9.2 Courses, Lectures and Seminars

Aksel Alpay:
Practical, *"Advanced: Towards completing the offloading-aware C++ standard algorithms library in AdaptiveCpp"*, Heidelberg University, summer semester 2025.

Torben Berndt, Mila Coetzee, Beni Egressy, Jan Stühmer:
Seminar, *"Interpretability and Causality in Machine Learning"*, Heidelberg University, summer semester 2025.

Torben Berndt, Jan Stühmer:
Lecture, *"Geometric Deep Learning"*, Karlsruhe Institute of Technology, winter semester 2025/2026.

Lynn Buchele:
Invited seminar talk, *"Testing our Understanding of Stellar Interiors with Asteroseismology"*, Wichita State University, Fairmount, USA, 2025.

Marcus Buchwald:
Workshop, *"YoKI-Workshop: Der richtige Umgang mit Open-Source Sprachmodellen am Beispiel der YoKI Plattform"*, Computing Center of Heidelberg Unviersity, 24 June 2025.

Isabel Gernand:
Teaching assistant, *"Einführung in die numerische Mathematik"*, Heidelberg University, winter semester 2024/2025.

Teaching assistant, *"Numerik"*, Heidelberg University, summer semester 2025.

Teaching assistant, *"Finite Elements"*, Heidelberg University, winter semester 2025/2026.

Tilmann Gneiting, Kristof Kraus:
Lecture, *"Forecasting: Theory and Practice II"*, Karlsruhe Institute of Technology, summer semester 2025.

Seminar, *"Forecasting: Theory and Practice"*, Karlsruhe Institute of Technology, winter semester 2025/2026.

Martin Golebiewski:
Tutorial, *"FAIR Provenance for Health Data Spaces"*, 16th International SWAT4HCLS 2025 Conference - Semantic Web Applications and Technologies for Life Science and Health Care, Barcelona, Spain, 24-27 February 2025.

Jonas Granzow:
Teaching assistant, *"Übungen zum Mathematischen Vorkurs zum Kompaktkurs der Experimentalphysik und Einführung in die Auswertung von Versuchen"*, Heidelberg University, summer semester 2025.

Frauke Gräter, Camilo Aponte-Santamaria (MPI Mainz), Rebecca Wade, Stefan Richter, Riccardo Beccaria:
Lecture and Practical Course, *"Computational Molecular Biophysics"*, Heidelberg University, winter semester 2025/2026.

Fabian Grünewald:
Lecture, *"Thermodynamics for Life Sciences"*, Molecular Biotechnology Bachelor's program, Heidelberg University, summer semester 2025.

Saskia Hekker:
Lecture, *"Asteroseismology"*, Heidelberg University, winter semester 2024/2025.

Seminar, *"Applications of Asteroseismology"*, Heidelberg University, summer semester 2025.

Lecture, *"Asteroseismology"*, Heidelberg University, winter semester 2025/2026.

Lecture course, *"Scientific Writing Basics for IMPRS PhD Students"*, Heidelberg University, winter semester 2024/2025.

Max Heller:
Tutorial, *"Introduction to Astronomy and Astrophysics I"*, Heidelberg University, winter semester 2025/2026.

Vincent Heuveline:
Lecture, *"IT-Sicherheit 1"*, Heidelberg University, winter semester 2024/2025.

Lecture, *"Einführung in die numerische Mathematik"*, Heidelberg University, winter semester 2024/2025.

Seminar, *"IT-Security"*, Heidelberg University, winter semester 2024/2025.

Lecture, *"Numerik"*, Heidelberg University, summer semester 2025.

Lecture, *"IT-Sicherheit 2"*, Heidelberg University, summer semester 2025.

Seminar, *"IT-Security"*, Heidelberg University, summer semester 2025.

Lecture, *"Finite Elements"*, Heidelberg University, winter semester 2025/2026.

Lecture, *"IT-Sicherheit 1"*, Heidelberg University, winter semester 2025/2026.

Seminar, *"AI Safety and Biomedicine"*, Marsilius Bridge Seminar at Heidelberg University, winter semester 2025/2026.

Maximilian Ludwig:
Teaching assistant, *"IT-Security"*, Heidelberg University, summer semester 2025.

Teaching assistant, *"IT-Sicherheit 2"*, Heidelberg University, summer semester 2025.

Practical, *"Beginner: IT-Sicherheit"*, Heidelberg University, summer semester 2025.

Practical, *"Advanced: IT-Sicherheit"*, Heidelberg University, summer semester 2025.

Teaching assistant, *"IT-Sicherheit 1"*, Heidelberg University, winter semester 2025/2026.

Stefan Machmeier:
Practical, *"IT-Sicherheit"*, Heidelberg University, winter semester 2024/2025.

Teaching assistant, *"IT-Security"*, Heidelberg University, winter semester 2024/2025.

Practical, *"Beginner: IT-Sicherheit"*, Heidelberg University, summer semester 2025.

Practical, *"IT-Sicherheit"*, Heidelberg University, summer semester 2025.

Teaching assistant, *"IT-Sicherheit 2"*, Heidelberg University, summer semester 2025.

Teaching assistant, *"IT-Security"*, Heidelberg University, summer semester 2025.

Teaching assistant, *"IT-Sicherheit 1"*, Heidelberg University, winter semester 2025/2026.

Wolfgang Müller:
Course, *"OpenRefine + Rightfield in Action"*, NFDI4Biodiversity, Hochschule Anhalt, Köthen, Germany, 6 - 27 March 2025.

Seasonal School, *"Three useful tools to get data in shape"*, NFDI4Biodiversity iDiv & HeFDI Seasonal School 2025, online, 6 - 10 October 2025.

Jonas Roller:
Teaching assistant, *"Einführung in die numerische Mathematik"*, Heidelberg University, winter semester 2024/2025.

Teaching assistant, "*Numerik*", Heidelberg University, summer semester 2025.

Friedrich Röpke:
Lecture, *"Computational Astrophysics"*, Heidelberg University, summer semester 2025.

Seminar, *"Physics of Stellar Objects"*, Heidelberg University, winter semester 2024/2025.

Seminar, *"Physics of Stellar Objects"*, Heidelberg University, summer semester 2025.

Seminar, *"Physics of Stellar Objects"*, Heidelberg University, winter semester 2025/2026.

Ulrich Sax, Matthias Löbe, Caroline Bönisch, Xiaoming Hu:
Workshop, *"FAIR Interfaces for improved Research Data Sharing and Interoperability"*, Medical Informatics Europe 2025, Glasgow, UK, 19 May 2025.

Valentin Schmid:
Teaching assistant, *"Einführung in die numerische Mathematik"*, Heidelberg University, winter semester 2024/2025.

Teaching assistant, *"Numerik"*, Heidelberg University, summer semester 2025.

Teaching assistant, *"Finite Elements"*, Heidelberg University, winter semester 2025/2026.

Fabian Schneider:
Lecture course, *"Stars Squared: Evolution of Binary Stars"*, Heidelberg University, summer semester 2025.

Fabian Schneider, Friedrich Röpke:
Lecture course, *"The Stellar Cookbook: A practical guide to the theory of stars"*, Heidelberg University, winter semester 2024/2025.

Lecture course, *"The Stellar Cookbook: A practical guide to the theory of stars"*, Heidelberg University, winter semester 2025/2026.

Kushal Singh, Carl Herrman (Univ. Heidelberg):
"Meet-EU Heidelberg-team", European Bioinformatics Master Network, winter semester 2025/2026.

Alexandros Stamatakis:
Seminar, *"Reproducibility in Bioinformatics"*, Computer science Master's program at the University of Crete, Greece, spring semester 2025.

Summer School, *"Computational Molecular Evolution"*, Hellenic Center for Marine Research, Crete, Greece, May 2025.

Guest Lecture, *"Phylogenetic Inference"*, University of Thrace, Alexandroupoli, Greece, March 2025.

Alexandros Stamatakis, Julia Haag, Alexey Kozlov:
Lecture, *"Introduction to Bioinformatics for Computer Scientists"*, Computer science Master's program at Karlsruhe Institute of Technology and the University of Crete, winter semester 2024/2025.

Alexandros Stamatakis, Luise Häuser, Alexey Kozlov:
Lecture, *"Introduction to Bioinformatics for Computer Scientists"*, Computer science Master's program at Karlsruhe Institute of Technology and the University of Crete, winter semester 2024/2025.

Alexandros Stamatakis, Anastasis Togkousidis, Alexey Kozlov, Luise Häuser, Lukas Hübner, Dimitri Höhler, Johannes Hengstler:
Seminar, *"Hot Topics in Bioinformatics"*, Computer science Master's program at Karlsruhe Institute of Technology, summer semester 2025.

Michael Strube:
Colloquium, *"Computational Linguistics Colloquium"*, Heidelberg University, winter semester 2024/2025.

Colloquium, *"Computational Linguistics Colloquium"*, Heidelberg University, summer semester 2025.

Jan Stühmer:
Lecture, *"Geometric Deep Learning"*, Karlsruhe Institute of Technology, winter semester 2024/2025.

Lecture, *"Geometric Deep Learning"*, Heidelberg University, summer semester 2025.

Jonathan Teuffel, Riccardo Beccaria, Rebecca Wade:
Tutorial, *"τRAMD calculations of protein-ligand dissociation rates and SDA (Simulation of Diffusional Association)"*, EBRAINS user meeting, Heidelberg University, 12 March 2025.

Tobias van Lier:
Teaching assistant, *"Physik A (Physik für Nebenfächler)"*, Heidelberg University, winter semester 2024/2025.

Teaching assistant, *"Physik B (Physik für Nebenfächler)"*, Heidelberg University, summer semester 2025.

Kristián Vitovský:
Tutorial, *"Theoretical Astrophysics"*, Heidelberg University, winter semester 2024/2025.

Tutorial, *"Computational Astrophysics"*, Heidelberg University, winter semester 2024/2025.

Tutorial, *"Theoretical Astrophysics"*, Heidelberg University, winter semester 2025/2026.

Rebecca Wade:
Lecture contributions: B. Sc. Molecular Biotechnology *"Biophysics"* (winter semester, 2024/25), M. Sc. Molecular Biotechnology *"MoBi4All"* (winter semester, 2024/25), M. Sc. Biochemistry *"Computational Biochemistry"* course (winter semester, 2024/25), M.Sc. Molecular & Cellular Biology Module 3 on *"Protein Modelling"* (summer semester, 2025).

Rebecca Wade, Stefan Richter, Jonathan Teuffel:
Lecture and Practical Course, *"Bioinformatics"*, Heidelberg University, winter semester 2024/2025.

Rebecca Wade, Feroz Karim, Leif Seute, Evgeni Ulanov, Frauke Gräter (MPIP Mainz):
Seminar Course, *"Machine Learning for the Biomolecular World"*, Heidelberg University, summer semester 2025.

Rebecca Wade, Stefan Richter, Riccardo Beccaria:
Lecture and Practical Course, *"Bioinformatics"*, Heidelberg University, summer semester 2025.

Rebecca Wade, Stefan Richter, Mislav Brajkovic, Christine Schulz:
Lecture and Practical Course, *"Computational prediction of protein structures and binding properties"*, Heidelberg University, winter semester 2025/2026.

Ulrike Wittig:
Workshop, *"COST Action CA21111 OneHealthDrugs"*, FAIRDOMHub Hands-on training, online, 14 May 2025.

Workshop, *"Cloud meets RDM: Unlocking Research Data Management in the Cloud"*, FAIRDOMHub / FAIRDOM-SEEK Hands-on training, Gatersleben, Germany, 8-9 September 2025.

Software Demonstration, *"Datenbanken + Datenmanagement"*, SETAC/ GDCh-Postgradualkurs *"Entwicklung von Alternativmethoden"*, online, 3-7 November 2025.

Alexander Zeilmann:
Practical, "*Advanced: IT-Sicherheit*", Heidelberg University, winter semester 2024/2025.

10 Miscellaneous

10.1 Talks and Presentations

Felix Ahlborn:
"Estimating internal rotation rates of red giants in the presence of interacting mixed modes", poster, TASC9/KASC16 Workshop 2025, Vienna, Austria, July 2025.

Róbert Andrássy:
"Convective penetration in upper main sequence stars", talk, Stellar Hydro Days VI, Victoria, BC, Canada, 13 May 2025. *"Radiation transport for simulations of common envelope evolution"*, invited talk, Max Planck Institute for Astronomy, Heidelberg, Germany, 27 November 2025; talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025.

Tommaso Bartoloni:
"Simulations to investigate the membrane interactions of a peptide with activity against heart failure", poster, Biophysical Society Annual Meeting (BPS), Los Angeles, USA, 15-19 February 2025; 3rd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany, 7-9 May 2025.

Utkarsh Basu:
"Products and Efficiency of Mass Transfer", talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025.

Michaël Bazot:
"Saturated stars PLATO WG", talk, PLATO Week #16, Paris, France, October 2025.

Riccardo Beccaria:
"Protein-Ligand Association in Crowded Media: a Multiscale Brownian Dynamics Simulation Approach", poster, 3rd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany, 7-9 May 2026; talk, MaSBiC Symposium, Ancona, Italy, 10-12 September 2025; Sixth Biological Diffusion and Brownian Dynamics Brainstorm (BDBDB6), On-line meeting, 11-12 December 2025.

Torben Berndt:
"Permutation Equivariant Neural Controlled Differential Equations for Dynamic Graph Representation Learning", poster, NeurIPS 2025, Vienna, Austria, 6 December 2025.

Beatriz Bordadágua:
"Combining the effects of mixed modes and magnetic instabilities in angular momentum transport", poster, TASC9/KASC16 Workshop 2025, Vienna, Austria, 9 July 2025.

Mislav Brajkovic:
"Investigation of the effect of loop motion on the residence times of histamine-1-receptor antagonists by τRAMD", poster, CECAM Flagship Workshop on *"Predicting and understanding drug-target binding kinetics via molecular simulations"*, Lausanne, Switzerland, 10-12 March 2025.

Vincent Bronner:
"How pulsations shape supernova light curves", talk, Annual Meeting of the European Astronomical Society, Cork, Ireland, 26 June 2025. *"When stars go dynamic",* invited talk, Institute Seminar, Institute of Astronomy, KU Leuven, Belgium, 17 July 2025. *"Pulsating red supergiants and their supernovae"*, invited talk, ESO Star-Planet Exchange, European Southern Observatory, Garching, 9 December 2025. *"Pulsation-Driven Diversity in Type II Supernova Light Curves"*, invited talk, Seminar on Stellar Astrophysics, Max Planck Institute for Astrophysics, Garching, 10 December 2025; talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025. *"Pulsating Red Supergiants: A New Perspective on Type II Supernova Light Curve Diversity"*, talk, VLT, Supernovae, Bruno @65 & H0 @?, Garching, Germany, 10 April 2025; poster, One hundred years of supernova science, Stockholm, Sweden, 20 August 2025. *"Comparing stellar evolution grids with BONNSAI",* talk, BLOeM collaboration meeting, Amsterdam, Netherlands, 16 October 2025.

Lynn Buchele:
"Applicability of current structure inversion techniques to subgiant stars", talk, TASC9/KASC16 Workshop 2025, Vienna, Austria, 8 July 2025.

Romain Chazotte:
"Improved O(2) Equivariance via Embedding in Zernike Polynomial Space", poster, Astronomical Data Analysis Software and Systems, Görlitz, Germany, 9-13 November 2025. *"Improved O(2) Equivariance via Embedding in Zernike Polynomial Space"*, talk, Annual Meeting of the Astronomische Gesellschaft, Görlitz, Germany, 15-19 September 2025.

Jeong Yun Choi:
"Power density spectra morphologies of seismically unresolved red-giant asteroseismic binaries", poster, TASC9/KASC16 Workshop 2025, Vienna, Austria, July 2025. *"Power density spectra morphologies of red-giant asteroseismic binaries with overlapping oscillations"*, invited talk, Europe-Korea Conference on Science and Technology 2025, TU Vienna, Austria, 25-28 August 2025.

Paul Christians:
"Multidimensional Hydrodynamical Simulations of Thermonuclear Ignition in Oxygen-Neon-Carbon Cores", talk, DPG-Frühjahrstagung 2025, Girona, Spain, 10 - 14 March 2025. *"Thermonuclear Ignition in Degenerate Oxygen-Neon Cores: The Impact of Electron Captures and Convection"*, poster, International Symposium on Nuclei in the Cosmos XVIII, Heidelberg, Germany, 15 - 20 June 2025. *"Hydrodynamical Simulations of Oxygen Ignition in Oxygen-Neon Carbon Cores"*, talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025.

Quentin Coppée:
"Magnetic mode depression and frequency shifts in quadrupole modes of red giants", poster, TASC9/KASC16 Workshop 2025, Vienna, Austria, July 2025.

Iliana Isabel Cortés Pérez:
"Exploring SMBH Activity in AGNs through Photometric Variability", talk, IMPRS Second year retreat, Ebernburg, Germany, 28 - 30 April 2025. *"Photometric decomposition of AGN variability via Machine Learning: Spectral reconstruction and lag characterization",* talk, Annual Meeting of the Astronomische Gesellschaft, Görlitz, Germany, 15-19 September 2025; ZihuAGN Redemption: central kpc region, Zihuatanejo, Mexico, 27- 31 October 2025.

Francisca Espinoza:
"Red giant asteroseismic binaries in the Kepler Field: Identifying gravitationally bound systems", poster, Binary stars in the Space Era, Keele, UK, 3 July 2025; talk, TASC9/KASC16 Workshop 2025, Vienna, Austria, 9 July 2025;

Christina Fakiola:
"On the isotopic yields of thermonuclear explosions in non-accreting progenitors", poster, Multi-messenger nuclear astrophysics in the 21st century, Stockholm, Sweden, 1 -7 March 2025; One hundred years of supernova science, Belfast, United Kingdom, 20 August 2025. *"Simulations of thermonuclear electron-capture supernovae"*, talk, Queen's University Belfast Astronomical Seminar, Heidelberg, Germany, November 2025.

Rostislav Fedorov:
"Exploration of Chemical Space of Target Reduction Potential", poster, 3rd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany, 7-9 May 2025. *"WallpaperNet: A p6mm-Equivariant Graph Neural Network for Molecule Adsorption on Graphene"*, poster, Eastern European Machine Learning Summer School, Sarajevo, Bosnia and Herzegovina, 21-26 July 2025.

Juliane Fluck, Darms Johannes, Martin Golebiewski, Carsten Oliver Schmidt, Carina Nina Vorisek, Matthias Löbe, Ulrich Sax and Iris Pigeot:
"Challenges and opportunities of integrating NFDI4Health into the European Health Data Space (EHDS)", poster, 2nd Conference on Research Data Infrastructure (CoRDI), Heidelberg, Germany, 26-28 August 2025.

Damien Gagnier:
"3D MHD simulations of common envelope evolution", talk, Annual Meeting of the European Astronomical Society, Cork, Ireland, 26 June 2025; Binary Stars in a New Era, Lijiang, China, 25 August 2025. *"Local simulations of common envelope dynamical inspiral"*, talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025.

Vagrant Gautam:
"Evaluating Reference in NLP", invited talk, The University of Edinburgh, Edinburgh, Scotland, September 2025; talk, Heidelberg University, Heidelberg, Germany, October 2025. *"Agree to Disagree? A Meta-Evaluation of LLM Misgendering"*, poster, Conference on Language Modeling, Montreal, Canada, 7-10 October 2025.

Isabel Gernand:
"Data-driven modelling and optimization for microsatellite-unstable tumors in colorectal cancer", poster, European Conference on Numerical Mathematics and Advanced Applications (ENUMATH), Heidelberg, Germany, 1 - 5 September 2025.

Tilmann Gneiting:
"Isotonic Distributional Regression", invited talk, Heidelberg–Paris Workshop on Mathematical Statistics, Heidelberg, Germany, 27 January 2025; Bocconi University, Milano, Italy, 20 March 2025. *"Assessing Monotone Dependence: Area Under the Curve Meets Rank Correlation"*, invited talk, ETH Zürich, Zürich, Switzerland, 23 October 2025; IMS International Conference on Statistics and Data Science (ICSDS), Seville, Spain, 15 December 2025.

Martin Golebiewski:
"ISO/TC 276/WG 5 Data Processing and Integration - Standards for Life Science Data", invited talk, DigSustain2: Workshop on Digital Data Standards Sustainability in the Chemical Sciences, Leuven, Belgium, 3 - 4 April 2025. *"The role of standards in defining an ecosystem for virtual human twins"*, talk, HARMONY 2025, Jena, Germany, 15-18 April 2025. *"Data Sharing Rights in FAIRDOM SEEK and NFDI4Health Local Data Hubs"*, talk, 70. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Jena, Germany, 7-11 September 2025. *"Metadata in Data Usage Agreements"*, talk, 70. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Max Planck Society, Ringberg, Germany, 7-11 September 2025.

Fabian Grünewald:
"From CGsmiles to multi-resolution GNNs for chemical space exploration", talk, 3rd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany, 7-9 May 2025. *"Complex System Building"*, invited talk, Martini Workshop 2025, University of Groningen, Groningen, The Netherlands, 11-15 August 2025. *"Tuning the interplay of polymers with facilitating simulations cellular matter in the digital lab to advance of macromolecules biomaterials"*, invited talk, Symposium on *"Interdisciplinary Micro Nano and Molecular Systems"*, Max Planck Society, Ringberg, Germany, 2-4 October 2025.

Saskia Hekker:
"The power of Asteroseismology the internal structures of stars", invited talk, Astronomisches Rechen-Institut (ARI), 2025. *"Curvature of large frequency separation"*, poster, TASC9/KASC16 Workshop 2025, Vienna, Austria, July 2025.

Jan Henneco:
"Massive Merger Products on the Main Sequence: Do Their Pulsations Betray Them?", talk, TASC9/KASC16, ISTA, Vienna, Austria, 9 July 2025. *"Merger Seismology: Identifying Stellar Merger Products with Asteroseismology"*, talk, IAUS 402, Ensenada, Mexico, 17 September 2025.

Alexander Holas:
"Electron-capture supernovae – Thermonuclear explosion or gravitational collapse? – The fate of intermediate mass stars on a knife's edge", talk, Los Alamos National Laboratory Astrophysics Seminar, Los Alamos, USA, 9 January 2025; One hundred years of supernova science, Stockholm, Sweden, 20 August 2025. *"Drawing the line between explosion and collapse in electron-capture supernovae"*, talk, Queen's University Belfast Astronomical Seminar, Belfast, United Kingdom, November 2025.

Xiaoming Hu, Owen Stuart, Frank Meineke, Finn Bacall, Carole Goble, Wolfgang Müller , Martin Golebiewski:
"Flexible Metadata Structuring for Research Data Management Through the FAIRDOM-SEEK Platform - Implementing Tailored and Complex Metadata Schemes in SEEK", poster, 2nd Conference on Research Data Infrastructure (CoRDI), Padova, Italy (online), 26-28 August 2025.

Alexey Kozlov:
"Green Computing", invited talk, EDUCADO training school, Padova, Italy (online), February 2025. *"One-pedal green computing: no (workload) shifting required"*, talk, EcoCompute 2025 Conference, Berlin, Germany, November 2025.

Kristof Kraus:
"How to compare physics-based numerical and data-driven AI models (for weather prediction) in a fair way?", talk, 12th HKMetrics-Workshop, Mannheim, Germany, 6 June 2025. *"Towards fair comparisons of AIWP and NWP model output"*, talk, Institute of Statistics Inaugural Workshop, Karlsruhe, Germany, 11 December 2025.

Kushal Singh:
"HYDRA Based Pipeline for Modelling Novel Peptide Inhibitors", talk, SIMPLAIX Project Meeting, Mathematikon, Heidelberg University, 27 February 2025.

Mike Lau:
"Radiation hydrodynamic simulations of common-envelope evolution", invited talk, 2nd European Phantom code family users workshop, IPAG, Grenoble, France, 2 - 6 June 2025. *"Common-envelope evolution and stellar mergers"*, invited talk, Annual Meeting of the European Astronomical Society, Cork, Ireland, 26 June 2025. *"The obstructive role of radiation transport in common-envelope evolution"*, invited talk, Binary Stars in a New Era, Lijiang, China, 25 August 2025. XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025. *"Common-envelope evolution in binary stars"*, invited talk, Tsung Dao Lee Institute Astronomy Seminar, Shanghai, China, 5 September 2025. *"What happens when a star "eats" another star or planet?"*, talk, Hong Kong Laureate Forum, Hong Kong, 5-8 November 2025.

Georgios Lioutas:
"Impact of hyperons on neutron star mergers", talk, HEAVYMETAL Consortium Meeting Spring 2025, Ystad, Sweden, 4 - 9 May 2025. *"Impact of hyperons on neutron star mergers: Gravitational waves, mass ejection, and black hole formation"*, talk, Workshop on anti-matter, hyper-matter and exotica production, Salerno, Italy, 10 - 14 November 2025.

Matthias Löbe, Martin Golebiewski, Uli Sax:
"Profiling W3C DCAT for use in specific domains - Experience with the HealthDCAT-AP draft standard for the European Health Data Space (EHDS)", talk, 2nd Conference on Research Data Infrastructure (CoRDI), Abu Dhabi, 26-28 August 2025.

Maximilian Ludwig:
"Cross-Study Comparability for IDS: A Call for Clarity", talk, 9th Cyber Security in Networking Conference (CSNet), Abu Dhabi, 21 October 2025.

Frank Meineke, Matthias Löbe, René Hänsel, Masoud Abedi, Xiaoming Hu, Martin Golebiewski:
"Using NFDI4Health Local Data Hub and Metadata-Schema for clinical research infor-mation management", poster, 16th International SWAT4HCLS 2025 Conference - Semantic Web Applications and Technologies for Life Science and Health Care, Glasgow, Scotland, UK, 24 - 27 February 2025.

Frank Meineke, Xiaoming Hu, Martin Golebiewski, Matthias Löbe:
"Roll-Out of the NFDI4health Local Data Hubs for Medical Research Data Management", poster, 35th Medical Informatics Europe Conference - MIE 2025, Aachen, Germany, 19 - 20 May 2025.

Frank Meineke, Golebiewski Martin, Xiaoming Hu, René Hänsel, Masoud Abedi, Wolfgang Müller, Matthias Löbe:
"Rollout of the NFDI4Health Local Data Hubs - First Results", poster, 2nd Conference on Research Data Infrastructure (CoRDI), Vienna, Austria, 26-28 August 2025.

Jonas Müller:
"Visibilities of damped mixed modes", poster, TASC9/KASC16 Workshop 2025, Vienna, Austria, July 2025.

Wolfgang Müller:
"FAIRDOM SEEK software for data management and some use cases", invited talk, Workshop: RDM Solutions for OMICS, Heidelberg, Germany, 5 June 2025. *"Projektdatenmanagement: Motivation, Methoden Erfolge und Herausforderungen"*, talk, SciNetPro 2025, Thessaloniki, Greece, 1 October 2025.

Stuart Owen, Andrabi Munazah, Finn Bacall, Phil Reed, Xiaoming Hu, Ulrike Wittig, Maja Rey, Martin Golebiewski, Flora D'Anna, Kevin De Pelseneer, Jacky Snoep, Wolfgang Müller, Carole Goble:
"FAIRDOM-SEEK - Platform for FAIR data and research outcome management", poster, ELIXIR All Hands Meeting, Thiruvananthapuram, India, 2 - 5 June 2025.

Rimpal Pal:
"Tackling Tumor Growth Based on Multilevel Approach", poster, Latest Advances in Computational and Applied Mathematics (LACAM), Santiago, Chile, 8 - 11 December 2025.

Francisco Pozo Nunez:
"Mapping High-z Accretion Disks from La Silla to Rubin: Medium-Band PRM Results and LSST Prospects", talk, Observations and Physics of Narrow-Line Seyfert 1 Galaxies: AGN at their Extreme, Santiago, Chile, 26-28 November 2025. *"Probing Quasar Accretion Disks Across Cosmic Time with Medium-Band Reverberation Mapping"*, talk, Highly Accreting Supermassive Black Holes Across all Cosmic Times: from the Local Universe to Cosmic Dawn, Santiago, Chile, 1-5 December 2025.

Jonas Reichert:
"MHD Simulations of Dipole Mode Suppression in Red Giant Stars", talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025.

Maja Rey, Anne Elin Heggland, Wolfgang Müller, Nicola Kunz, Matthias Ganzinger, Christof Meigen:
"Data Management in LiSyM-Cancer", poster, LiSyM-Cancer Status Seminar, Berlin, Germany, 8 - 9 April 2025.

Maja Rey, Ulrike Wittig, Xiaoming Hu, Wolfgang Müller:
"de.NBI-SysBio: Research data management and curation of systems biology applications", poster, ELIXIR-DE All Hands Meeting, Eisenach, Germany, 19 - 21 November 2025.

Maja Rey, Heggland Anne Elin, Matthew Johnstone, Sebastian Kaletta, Matthias Ganzinger, Christof Meigen, Wolfgang Müller:
"Data Management for LiSyM-Cancer II", poster, Scientist Retreat of LiSyM-Cancer II, Victoria, BC, Canada, 3 - 5 December 2025.

Federico Rizzuti:
"The occurrence of shell mergers in massive stars", talk, Stellar Hydro Days VI Workshop, Victoria, Canada, 12-16 May 2025. *"The production of heavy elements from rotating massive stars in the Galaxy"*, talk, sirEN Conference, Giulianova, Italy, 8-13 June 2025. *"The nucleosythesis of massive stars: a bridge between stellar modelling and galactic archaeology"*, talk, STARS II Conference, Bologna, Italy, 19 June 2025. *"Understanding convection in massive stars: a bridge between 1D and 3D models"*, talk, Workshop Stellar Populations and their Explosions: Bridging the Gap, Ringberg Castle, Germany, 10-14 November 2025. *"Simulating a massive-star core: convection, magnetic fields, and nuclear burning"*, talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, 15 - 17 December 2025.

Jonas Roller:
"Numerical Simulation of Taylor-Couette Flow Under Dielectrophoretic Force", talk, International Couette-Taylor Workshop 2025, Durham, UK, 14 - 16 July 2025. *"Optimal Control of Dielectrophoretic Force-Driven Flows"*, talk, 2nd European Fluid Dynamics Conference, Dublin, Ireland, 26 - 29 August 2025.

Friedrich Röpke:
"Type Ia supernova explosion scenarios and their associated progenitor systems", invited talk, *"An extraordinary journey into the transient sky"* conference, Padova, Italy, 3 April 2025. *"The SLH code and the quest for an optimal finite volume scheme"*, talk, Stellar Hydro Days VI, Victoria, Canada, 12 May 2025. *"Multi-messenger astronomy for white dwarf mergers – a theorist's perspective"*, invited talk, LGWA workshop, San Benedetto del Tronto, Italy (online), 16 September 2025.

Arkaprabha Roy:
"Convective Penetration in Massive Stars with and without a Magnetic Dynamo", talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025.

Valentin Schmid:
"PDE Framework for Tumor Invasion and Basement Membrane Dynamics: Application to Colorectal Cancer", talk, GAMM Annual Meeting 2025, Poznań, Poland, 7 - 11 April 2025.

Fabian Schneider:
"Pre-supernova evolution and final fates of binary stars", talk, Workshop Stellar Populations and their Explosions: Bridging the Gap, Ringberg Castle, Germany, 10 November 2025. *"Black Holes Near and Far"*, invited talk, ARI Institute Colloquium, University of Heidelberg, Germany, 18 December 2025.

Leif Seute:
"Learning the Boltzmann distribution of protein backbones", talk, Hünfeld Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, 8 March 2025. *"Learning conformational ensembles of proteins based on backbone geometry"*, talk, 76th Mosbacher Kolloquium, Mosbach, Germany, 20 March 2025. *"Learning conformational ensembles of proteins based on backbone geometry"*, talk, SIMPLAIX Workshop 2025, 7 May 2025. *"Flexibility-Conditioned Protein Structure Design with Flow Matching"*, poster, ICML 2025, Vancouver, Canada, 16 July 2025. *"Learning conformational ensembles of proteins based on backbone geometry"*, poster, NeurIPS 2025, San Diego, USA, 10 December 2025.

Kushal Singh:
"Tracking Conformational Transitions in Malarial PfCRT using Enhanced Sampling Simulations", poster, Gordon Research Conference on Mechanisms of Membrane Transport, Les Diablerets, Switzerland, 28 May 2025.

Alexandros Stamatakis:
"The Biodiversity Computing Group After 34 Project Months: From Brain Gain to Brain Re-Drain", talk, 3rd Greek ERA chair meeting, FORTH, Crete, Greece, October 2025.

Jan Stühmer:
"Geometric Deep Learning as Design Principle for Modern Deep Learning", invited talk, CHAIR Workshop on Structured Learning, Chalmers University, Gothenburg, Sweden, 28 October 2025. *"Equivariance as Design Principle for Modern Deep Learning"*, invited talk, Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany, 11 November 2025.

Jonathan Teuffel:
"Complexation of cytochrome P450 17A1 with its redox proteins in a membrane: Insights from molecular dynamics simulation", poster, Biophysical Society Meeting, Los Angeles, USA, 15-19 February 2025; talk, 24th International conference on cytochrome P450, Bern, Switzerland, 30 June – 4 July 2025. *"On the determinants of electron transfer reorganization energy in a cytochrome P450: cytochrome b5 complex. A combined quantum mechanics and molecular dynamics simulation study"*, poster, 3rd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany, 7-9 May 2026. *"Multiscale simulations to*

study electron transfer complex formation of membrane-bound cytochrome P450 17A1 and its redox proteins", poster, Perspectives in Bioenergetics 2025, Rauischholzhausen, Germany, 29 September–1 October 2025.

Anastasis Togkousidis:
"Investigating Overfitting in Maximum Likelihood Phylogenetic Inference: A Systematic Approach", talk, Mathematical and Computational Evolutionary Biology (MCEB), Seville, Spain, May 2025.

Sebastian Trujillo Gomez:
"Scientific discovery using Representation Learning to interpret the largest cosmological simulations", talk, SPACE Visualization Workshop, Barcelona, Spain, 14-15 May 2025. *"The unknown unknowns: Representation Learning for simulation-based inference in galaxy formation and cosmology"*, talk, SBI2025 Workshop, Bristol, UK, 27-30 May 2025. *"UDGs, galaxy formation, cosmology, and rejecting hypotheses"*, talk, 10 Years of Ultra-diffuse Galaxies' conference, Benasque, Spain, 17-23 August 2025. *"From data to scientific breakthroughs with tools powered by Representation Learning"*, talk, Annual Meeting of the Astronomische Gesellschaft, Görlitz, Germany, 15-19 September 2025. *"From data to scientific breakthroughs using Representation Learning"*, poster, Astronomical Data Analysis Software and Systems, Görlitz, Germany, 9-13 November 2025. *"From simulation outputs to scientific breakthroughs using Representation Learning"*, talk, The Universe in a Chip: Final ICSC Spoke-3 Meeting on HPC & Big Data in Astrophysics, Sexten, Italy, 14-19 December 2025.

Jordan van Beek:
"Resonant Amplitude Equations for Gravito-Inertial Modes: Non-linear mode coupling in a slowly pulsating B star", talk, TASC9/KASC16 Workshop 2025, Vienna, Austria, July 2025.

Tobias van Lier:
"No intermediate coupling prescription needed on the red-giant branch", poster, TASC9/KASC16 Workshop 2025, Vienna, Austria, July 2025.

Marco Vetter:
"Impact of magnetic fields on the common-envelope phase", poster, Binary Stars in a New Era, Lijiang, China, 25 August 2025. *"The role of magnetic fields in the common-envelope phase"*, talk, "*Asymmetrical Post-main-sequence Nebulae 9. The Art of Stellar Wind Sculpting"* conference, Granada, Spain, 22 - 26 September 2025. *"Common-envelope evolution"*, talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025.

Kristián Vitovský:
"Internal gravity waves excited in the interiors of Blue Supergiants", talk, Stellar Hydro Days VI, Victoria, Canada, 13 May 2025. *"Origin of the Stochastic Low Frequency Variability in Bluesupergiant Stars"*, talk, XIX Winter Workshop on Stellar Astrophysics, Heidelberg, Germany, 15 - 17 December 2025.

Carina Nina Vorisek, Hölter Thimo, Julian Saß, Aliaksandra Shutsko, Elisa Kasbohm, Martin Golebiewski, Carsten Oliver Schmidt, Sylvia Thun, Sophie Klopfenstein:
"Standardizing Metadata in NFDI4Health: A FHIR Implementation Guide for Interdisciplinary Health Research", poster, 70. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Studio Villa Bosch, Heidelberg, 07-11 September 2025.

Rebecca Wade:
"Scaling protein-water interactions in the Martini 3 coarse-grained force field to simulate transmembrane helix dimers in different lipid environments", invited talk, Main-Neckar Martini User Meeting, Studio Villa Bosch, Heidelberg, 13-14 February 2025. *"Some challenges in computing binding kinetics"*, invited talk, CECAM Flagship Workshop on *"Predicting and understanding drug-target binding kinetics via molecular simulations"*, Lausanne, Switzerland, 10-12 March 2025. *"Exploring drug-target interactions in crowded environments by Brownian and molecular dynamics simulation"*, invited talk, American Chemical Society Spring Meeting, San Diego, USA, 23-27 March 2025. *"Towards peptide- and polysaccharide-based therapeutic agents via computer simulation"*, invited talk, American Chemical Society Spring Meeting, San Diego, USA, 23-27 March 2025. *"Multi-resolution molecular simulations to investigate the interplay between cytochrome P450 interactions, conformational variability and function"*, invited talk, 6th Manchester Multiscale Meeting, CCPBioSim/CCP5 Annual Meeting, Manchester, U.K., 31 March - 2 April 2025. *"Focusing the computational microscope on the dynamics of mammalian cytochrome P450s and their redox partners"*, invited talk, 24th International conference on cytochrome P450, Bern, Switzerland, 30 June – 4 July 2025. *"Glycans, GAGs and proteins: with molecular simulations towards therapeutics"*, invited talk, CECAM workshop on *"The Dual Nature of Glycans in Protein-Protein Interactions"*, Cagliari, Sardinia, 7-10 September 2025. *"Everything starts with water"*, invited talk, Symposium on *"Decoding Chemical Reactions: Celebrating Bill Jorgensen's half century of service",* Temple University Campus, Rome, Italy, 26-27 September 2025. *"The dynamics of protein binding: Insights from molecular simulation"*, invited talk, BNMRZ Symposium on Ultra-Highfield NMR and 4D Structural Biology – From Mechanisms to Therapies, TU München, Munich, Germany, 6-7 November 2025. *"Adding the 4th dimension to structure-based drug design"*, invited talk, Nostrum Biodiscovery (NBD), Barcelona, Spain, 21 November 2025. *"Brownian dynamics simulation to investigate molecular diffusion and protein complex formation"*, talk, SIMPLAIX Project Meeting, Mathematikon, Heidelberg University, 27 February 2025. *"BDBDB6: Update on SDA"*, talk, Sixth Biological Diffusion and Brownian Dynamics Brainstorm (BDBDB6), On-line meeting, 11-12 December 2025.

Ulrike Wittig:
"Community Collaboration in ELIXIR", invited talk, DigSustain2: Workshop on Digital Data Standards Sustainability in the Chemical Sciences, Gatersleben, Germany, 3 - 4 April 2025. *"From Publications to FAIR Data: Enabling Reusable Biological Data through Curation and FAIR Data Management"*, invited talk, 19th International Symposium on Integrative Bioinformatics, Rüdesheim, Germany, 10-12 September 2025. *"Future perspectives on curating and exchanging enzyme kinetics data with SABIO-RK and EnzymeML"*, invited talk, 6th EnzymeML Workshop, Tübingen, Germany, 29 September - 2 October 2025.

Alexander Zeilmann:
"Interactive AI on bwHPC – Lessons Learned from Building a Large-Scale Image Analysis Platform", talk, 11th bwHPC Symposium, Tübingen, Germany, 23 September 2025.

10.2 Memberships

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

Susan Eckerle:
Leadership Team of Liver Systems Medicine - Cancer II.

Tilmann Gneiting:
Affiliate Professor at the Department of Statistics, University of Washington, Seattle, Washington, United States. Member of the Committee on Fellows at Institute for Mathematical Statistics.

Martin Golebiewski:
Convenor (chair) of ISO/TC 276/WG 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 Scientific Advisory Board of the European INVENTS project

(Innovative designs, extrapolation, simulation methods and evidence-tools for rare diseases addressing regulatory needs). 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 DIN Standards Committee Health Technologies (NA 176) - Section for biotechnology, German Institute for Standardization (DIN). Member of the IEC/TC 62 Medical equipment, software and systems, International Electrotechnical Commission (IEC). Member of the Virtual Physiological Human Institute for Integrative Biomedical Research (VPH Institute).

Fabian Grünewald:
Core Member of the Martini Force-Field Initiative (https://cgmartini.nl).

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

Vincent Heuveline:
Member of the University Council of Universität Hamburg. Spokesman of the Heads of University Computer Centres / State Baden-Württemberg. Senator of the Heidelberg University. Member of the Scientific Advisory Board (SAB) Potsdam Institute for Climate Impact Research. PI of HIDSS4Health Graduate School. Conference Chair eScience-Tage (take place in Heidelberg every 2 years). Founder of the Competence Centre for Research Data, Heidelberg University. Head of the oneAPI Center of Excellence with Intel.

Kai Polsterer:
President of International Astro-Informatics Association. Member of the Standing Committee on Science Priorities of the International Virtual Observatory Alliance. Member of the Astronomische Gesellschaft. Member of the Deutsche Physikalische Gesellschaft. AKPIG Member of the *„Verein für datenintensive Radioastronomie“* (VdR). Member of the International Astrostatistics Association.

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

Alexandros Stamatakis:
Member of the advisory board of the SciLifeLab & Wallenberg National Program for Data-Driven Life Science program in Sweden.

Jan Stühmer:
Member of ELLIS - the European Laboratory for Learning and Intelligent Systems. Member of the Editorial Board of KI Journal – the Journal of the section for Artificial Intelligence (FBKI) in the Gesellschaft für Informatik e.V. (GI).

Jonathan Teuffel:
PhD representatives, SIMPLAIX Scientific Steering Committee.

Kristián Vitovský:
IMPRS-HD student representative 2024/2025. IMPRS-HD deputy student representative 2025/2026.

Rebecca Wade:
Editor-in-Chief: Journal of Molecular Recognition. Academic Editor: PLOS Computational Biology. Editorial Board: Biophysical Journal, 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 International Advisory Board, ‘Complexity in Chemistry’ C2 Programme, Department of Chemical Sciences, University of Padova, Italy. Member of Scientific Advisory Committee, Centre Europeen de Calcul Atomique et Moleculaire (CECAM), Lausanne, Switzerland. Member of Human Frontier Science Program Organization (HFSP) Review Committee for Research Grants, Strasbourg, France. Member of Steering Committee, Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities, Garching, Germany. 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, Scientific Software Center Advisory Board.

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. Working Group (WG1) leader of COST Action CA21111 - One Health drugs against parasitic vector borne diseases in Europe and beyond (OneHealthdrugs). Member of Scientific Advisory Board of the MALDIBANK project (Multi-domain Open MALDI spectra archive for identification of microorganisms).

10.3 Contributions to the Scientific Community

Aksel Alpay:
Program chair for the conference International Workshop on OpenCL and SYCL (IWOCL) '25. Local chair/host for the conference International Workshop on OpenCL and SYCL (IWOCL) '25. Reviewer for the conference International Workshop on OpenCL and SYCL (IWOCL) '25. Instructor: AdaptiveCpp Best Practices, Tutorial @ IWOCL '25.

Beatriz Bordadágua:
Teaching assistant, MESA Summer School, Leuven, 2025.

Susan Eckerle:
Scientific Organizing Committee member for the 10th Conference on Systems Biology of Mammalian Cells (SBMC2026), Mannheim, Germany, 4 -6 May 2026.

Vagrant Gautam:
Organizer, First Workshop on the Interplay of Model Behavior and Model Internals, Montreal, Canada, 10 October 2025.

Nikos Gianniotis:
Tutor, Second EDUCADO Training School, University of La Laguna, Tenerife, 16 July 2025.

Martin Golebiewski:
Organizing Committee Member of HARMONY 2025, Leuven, Belgium, 15 - 18 April 2025. Organizing Committee Member of COMBINE 2025 - Computational Modeling in Biology, Madison, WI, USA, 20 - 23 October 2025. Workshop organizer "Maschinenlesbare Abbildung von Datenzugriffs- und Datennutzungsrestriktionen” at the 70. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epi-demiologie (GMDS), Jena, Germany, 07-11 September 2025. Workshop organizer for ISO and COMBINE standards at HARMONY

2025, Leuven, Belgium, 15-18 April 2026. Chair of the ISO/TC 276 Biotechnology working group WG5 "Data Processing and Integra-tion", Cairns, Queensland, Australia, 31 May - 7 June 2025. Workshop organizer for FAIR interfaces for improved research data sharing and interoperability at the 35th Medical In-formatics Europe Conference - MIE 2025, Glasgow, Scotland, UK, 19-21 May 2025.

Saskia Hekker:
Teacher, MESA Summer School, Leuven 2025. AFP-Chair of European Space Science Committee, 2025. IMPRS-HD Spokesperson.

Wolfgang Müller:
Scientific Organizing Committee member for the 10th Conference on Systems Biology of Mammalian Cells (SBMC2026), Mannheim, Germany, 4 -6 May 2025.

Kai Polsterer:
Co-organizer, E-Science & Virtual Observatory Splinter at Annual Meeting of the Astronomische Gesellschaft, Görlitz, Germany, 15-19 September 2025. IVOA Interoperability Meeting, Görlitz, Germany 14-16 November 2025.

Friedrich Röpke:
Member of the Scientific Organizing Committee, *"Stellar Hydro Days VI"* meeting, University of Victoria, BC, Canada, 12 – 16 May 2025. Member of the Scientific Organizing Committee, Symposium S5 *"Binary Stellar Interactions and Their Outcomes"* at the Annual Meeting of the European Astronomical Society, Cork, Ireland, 25 – 26 June 2025. Member of the Scientific Organizing Committee, *"One hundred years of supernova science"* conference, Stockholm, Sweden, 18 – 22 August 2025.

Alexandros Stamatakis:
Organizer of 2025 Computational Molecular Evolution Summer School, Hellenic Center for Marine Research, Crete, Greece. Co-organizer of legend2025: Machine Learning for Evolutionary Genomics Data conference, Centre Paul Langevin, Aussois, France.

Michael Strube:
Co-Chair of Test-of-Time Paper Award at ACL 2025, The 63rd Annual Meeting of the Association for Computational Linguistics, Vienna, Austria, July 27 – August 1, 2025. Program Co-Chair of CODI 2025, The Fifth Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI 2025) at EMNLP 2025, Suzhou, China, November 9, 2025.

Jonathan Teuffel:
Presentation to visiting Biochemistry students from Heidelberg University, HITS, 24 April 2025.

Kristián Vitovský:
Organization of the IMPRS-HD 20th generation 1st year doctoral retreat, Padova & Asiago, Italy, March 2025.

Rebecca Wade:
with Giovanni Bottegoni (Univ. Urbino), Ariane Nunes-Alves (TU Berlin) and Stefan Wolf (Univ. Freiburg), CECAM Flagship Workshop on "Predicting and understanding drug-target binding kinetics via molecular simulations", Lausanne, Switzerland, 10-12 March 2025; (as QCMS Scientific Chair) with Giovanni Bottegoni, Andrea Cavalli, QSAR, chemoinformatics and modeling society (QCMS) first online lecture given by Prof. Harel Weinstein (Cornell University, New York, USA) and annual general meeting, online, 21 November 2025; with Riccardo Beccaria, Stefan Richter (HITS), Rommie Amaro (UCSD) and Chia-en Chang (UC Riverside), Sixth Biological Diffusion and Brownian Dynamics Brainstorm, On-line meeting, 11-12 December 2025..

Rebecca Wade (Chair):
with Rostislav Fedorov, Daniel Sucerquia, Jonathan Teuffel (HITS), and Pascal Friederich, Marcus Elstner, David Hoffmann (KIT), and Tristan Bereau (Univ. Heidelberg), 3rd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Studio Villa Bosch, Heidelberg, 15-17 May 2025.

Ulrike Wittig:
Member of the Program Committee for the 19th International Symposium on Integrative Bioinformatics, Gatersleben, Germany, 10 - 12 September 2025.

Other Contributions

Mila Coetzee:
"How AI Breaks Down Complex Video into Meaningful Parts", talk at Heidelberg University Students' Visit to HITS, 24 April 2025.

Peter Saueressig:
"The HITS Journalist in Residence program", in: "Stepping Out Of the Office to Learn More"; Session at the World Conference of Science Journalists WCSJ 2025, Pretoria, South Africa, 4 December 2025.

Leif Seute:
"Machine Learning for Protein Design", talk at Heidelberg University Students' Visit to HITS, 24 April 2025.

Jonathan Teuffel:
"Biomolecular Modeling and Simulation", talk at Heidelberg University Students' Visit to HITS, 24 April 2025.

10.4 Awards

Tommaso Bartoloni:
Scholarship for doctoral students for a short research stay in Singapore, (DAAD, 2025).

Lynn Buchele:
Best Postdoc Talk at TASC9/KASC16 Workshop 2025, Vienna, Austria, 2025.

Alexandros Stamatakis:
Highly Cited Researcher, Clarivate Analytics, 2025.

Jonathan Teuffel:
Poster Award, Runner-up, 24th International conference on cytochrome P450, Bern, Switzerland, 4 July 2025.

The HITS Scientific Advisory Board and the HITS management (October 2025). From left to right: Barbara Wohlmuth, Viola Vogel, Jeffrey Brock, Alex Szalay, Thomas Lengauer, Rebecca Wade (HITS Deputy Scientific Director), Gesa Schönberger (HITS Managing Director), Kai Polsterer (HITS Scientific Director).

11 Boards and Management

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 2025, the board consisted of the following members:

Prof. Dr. Jeffrey Brock

Zhao and Ji Professor of Mathematics at Yale University, USA

Prof. Dr. Barbara Ercolano

Theoretical Astrophysics at Ludwig-Maximilians University, Munich (LMU), Germany

Prof. Dr. Thomas Lengauer

Max Planck Institute for Computer Science, Saarbrücken, Germany (Chair, SAB)

Prof. Dr. Victoria Stodden

School of Information Sciences, University of Illinois at Urbana-Champaign, USA (Vice Chair, SAB)

Prof. Dr. Alex Szalay

Johns Hopkins University, USA

Prof. Dr. Viola Vogel

Department of Health Sciences and Technology at the ETH Zürich, Switzerland

Prof. Dr. Barbara Wohlmuth

Chair of Numerical Mathematics at the Technical University of Munich (TUM), Germany

Shareholders’ Board

HITS-Stiftung (HITS Foundation)

Prof. Dr. Carsten Könneker

Member of the Board of Directors (until July 2025)

© Mück/Klaus Tschira Stiftung

Prof. Dr. Dieter Kranzlmüller

Member of the Board of Directors

© Alessandro Podo

Prof. Dr. Rafael Lang

Member of the Board of Directors

© Mück/Klaus Tschira Stiftung

Heidelberg University

Prof. Dr. Andreas Dreuw

Vice President of Research and Structure

© Tobias Schwerdt

Karlsruhe Institute of Technology (KIT)

Prof. Dr. Michael Decker

Head of division II "Informatics, Economics, and Society" (until May 2025)

Prof. Dr. Jan Hesthaven

President (since June 2025)

© KIT/Markus Breig

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:

Dr. Kai Polsterer

(2025–2027)

Deputy Scientific Director:

Prof. Dr. Rebecca Wade

(2025–2027)

HITS

HITS, the Heidelberg Institute for Theoretical Studies, was established in 2010 by physicist and SAP co-founder Klaus Tschira (1940-2015) and the Klaus Tschira Foundation as a private, non-profit research institute. HITS conducts basic research in the natural, mathematical, and computer sciences. Major research directions include complex simulations across scales, making sense of data, and enabling science via computational research. Application areas range from molecular biology to astrophysics.

An essential characteristic of the Institute is interdisciplinarity, implemented in numerous cross-group and cross-disciplinary projects. The base funding of HITS is provided by the Klaus Tschira Foundation.

HITS gGmbH
Schloss-Wolfsbrunnenweg 35
D-69118 Heidelberg

Editor
Dr. Peter Saueressig
Deputy 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 | © 2026 HITS gGmbH

Bluesky:	hitsters.bsky.social
Facebook:	/HITStudies
Youtube:	/TheHITSters
LinkedIn:	company/the-heidelberg-institute-for-theoretical-studies
Instagram:	the_hitsters