HITS
Heidelberg Institute for Theoretical Studies

2024
Annual Report
Jahresbericht

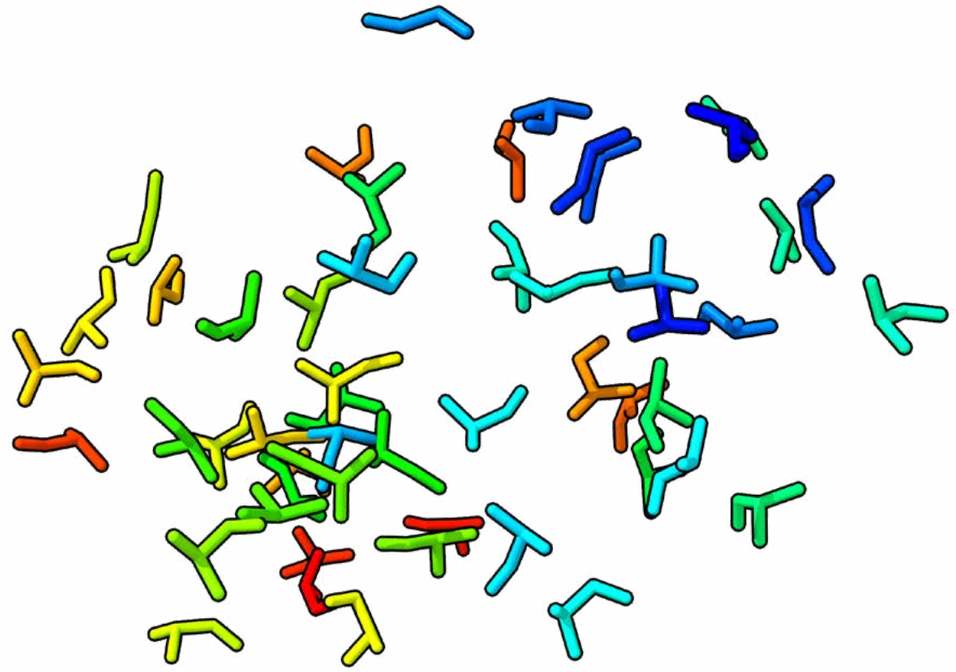
Think beyond the limits!

Proteins are fundamental to the existence and functioning of all living organisms and are responsible for numerous functions in biological systems. Artificially created, custom-designed proteins open up new possibilities for research, enabling their use for drugs, vaccines, or sensors. The Machine Learning and Artificial Intelligence (MLI) group have introduced a generative model for protein backbone design, utilizing geometric products and higher order message passing called GAFL (Geometric Algebra Flow Matching).

The image shows an example for a Flow Matching trajectory of GAFL, where noise is sampled from a Gaussian distribution and transformed to a protein backbone structure. A protein structure can be represented by a set of fragments, each of which represents four atoms from the protein backbone. Protein generation models, like the GAFL model developed at HITS, learn how to re-assemble noise of up to 500 fragments with randomly sampled positions - as portrayed in the picture - into a consistent protein structure. (cf. Chapter 2.6 Machine Learning and Artificial Intelligence, p. 37ff).

Proteine zählen zu den wichtigsten Bausteinen lebender Organismen und sind für zahlreiche Aufgaben in biologischen Systemen zuständig. Künstlich erzeugte, maßgeschneiderte Proteine eröffnen der Forschung neue Möglichkeiten, sie als Medikamente, Impfstoffe oder Sensoren einzusetzen. Die HITS-Gruppe Machine Learning and Artificial Intelligence (MLI) hat ein generatives Modell für das Design von Protein-Backbones eingeführt: das sogenannte Geometric Algebra Flow Matching (GAFL).

Das Bild zeigt ein Beispiel für das Flow-Matching von GAFL, bei der das Rauschen gemäß der Gaußschen Verteilung in eine Proteinrückgratstruktur umgewandelt wird. Eine Proteinstruktur kann durch verschiedene Fragmente dargestellt werden, von denen jedes vier Atome des Proteinrückgrats repräsentiert. Proteingenerierungsmodelle, wie das am HITS entwickelte GAFL-Modell, lernen das Rauschen aus bis zu 500 Fragmenten mit zufällig gewählten Positionen in eine konsistente Proteinstruktur umzuwandeln – so wie im Bild dargestellt. (Mehr dazu siehe Kapitel 2.6 Machine Learning and Artificial Intelligence, S. 37ff).



1 Think beyond the limits!

Dr. Gesa Schönberger
(Managing Director / Geschäftsführerin)

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

Science is a dynamic and constantly growing field. Every discovery and every new theory raises novel questions and opens up new perspectives.

These dynamics were also reflected at HITS in 2024:
Ganna (Anya) Gryn'ova – head of the Computational Carbon Chemistry (CCC) junior group – accepted an offer for the position of Associate Professor of Computational Chemistry at the University of Birmingham in the United Kingdom and left HITS in April 2024 (see Chapter 2.2).

Moreover, after fifteen years of highly successful work as head of the Molecular Biomechanics (MBM) group at HITS, **Frauke Gräter** joined the Max Planck Institute for Polymer Research in Mainz as a director effective January 2025 (see Chapter 2.7).

Furthermore, **Adele Goldberg** and **Gert-Martin Greuel** retired from the HITS Scientific Advisory Board (SAB) after ten years of dedicated service. In return, we were able to welcome our new SAB members **Viola Vogel** – a biophysicist at ETH Zurich, Switzerland – and **Barbara Ercolano** – an astrophysicist at LMU Munich – in 2024 (see Chapter 11).

Last but not least, **Wilfried Juling** retired as director of the HITS Foundation in November 2023, and we organized a special symposium on the theme of AI at HITS on 18 March 2024 in his honor (see Chapter 5.3).

We are extremely grateful for our former colleagues' contributions and commitment and wish them all the best for their future. In fact, the new career paths of the two scientists are illustrative of the exceptionally high quality of the scientific work that is carried out at HITS. At the same time, we are happy to welcome our new distinguished SAB members who will guide us on the scientific paths that will come.

Despite many changes, HITS continues to focus on cutting-edge research in the natural, mathematical, and computer sciences. Here are just a few highlights from 2024:
Frauke Gräter received a Human Frontier Science Program research grant together with Ronen Zaidel-Bar (Tel Aviv University, Israel) and Alexander Dunn (Stanford University, USA) to study the effects of physical force on extracellular collagen in mice and C. elegans.

Tilmann Gneiting received the 2024 Ulf Grenander Prize in Stochastic Theory and Modeling. The prize is awarded every three years by the American Mathematical Society and recognizes seminal work in the areas of probabilistic modeling, statistical inference, or related computational algorithms, especially for the analysis of complex or high-dimensional systems (see Chapter 10.5).

For the ninth year in a row, **Alexandros Stamatakis** was recognized as a Highly Cited Researcher by Clarivate Analytics (see Chapter 10.5).

Guillermo Cabrera-Vives – a computer scientist from the University of Concepción, Chile – visited HITS through the Klaus Tschira Guest Professorship program. Guillermo presented his research within the HITS Colloquium, intensified collaboration with the AIN group, and interacted with many other HITSters (see Chapter 6.1).

Felicitas Mokler – a German science journalist – spent April through September 2024 at HITS via the HITS Journalist in Residence program. During her stay, Felicitas interacted closely with HITSters, gave a public talk at the "Mathematics Informatics Station" (MAINS), and shared her expertise on science journalism and science communications (see Chapter 4).

Every change – big or small – gives us the opportunity to reflect on the past and think of what we want to achieve in the future. Scientific excellence and innovative research have always been our goal, and this goal will not be changed. HITS has a unique combination of diverse research areas, outstanding researchers, and the perfect setting for interdisciplinary exchange. Several new groups are currently being discussed and planned, and we appreciate the opportunity to both expand our knowledge and pursue new lines of research.

We hope you enjoy reading our Annual Report 2024!

Die Wissenschaft ist ein dynamisches und ständig wachsendes Feld. Jede Entdeckung, jede neue Theorie wirft neue Fragen auf und eröffnet neue Perspektiven.

Die Dynamik in der Wissenschaft spiegelte sich 2024 auch am HITS wider: **Ganna (Anya) Gryn'ova**, Leiterin der Juniorgruppe Computational Carbon Chemistry (CCC), nahm das Angebot für eine Stelle als Associate Professor für Computational Chemistry an der Universität Birmingham, Großbritannien, an und verließ das HITS im April 2024 (siehe Kapitel 2.2).

Nach fünfzehn Jahren erfolgreicher Arbeit als Leiterin der Arbeitsgruppe Molekulare Biomechanik (MBM) am HITS wechselte **Frauke Gräter** im Januar 2025 als Direktorin ans Max-Planck-Institut für Polymerforschung in Mainz (siehe Kapitel 2.7).

Adele Goldberg und **Gert-Martin Greuel** schieden nach zehn Jahren engagierter Arbeit aus dem wissenschaftlichen Beirat (SAB) des HITS aus. Im gleichen Jahr durften wir unsere neuen SAB-Mitglieder **Viola Vogel**, Biophysikerin an der ETH Zürich, und **Barbara Ercolano**, Astrophysikerin an der LMU München, begrüßen (siehe Kapitel 11).

Und schließlich ist **Wilfried Juling** im November 2023 als Direktor der HITS-Stiftung zurückgetreten. Ihm zu Ehren haben wir im März 2024 ein Symposium zum Thema Künstliche Intelligenz am HITS veranstaltet (siehe Kapitel 5.3).

Wir sind unseren ehemaligen Kolleg*innen und Gremienmitgliedern sehr dankbar für ihre Leistungen und ihr Engagement für unser Institut und wünschen ihnen alles Gute für ihre Zukunft. Die neuen Positionen der beiden Forscherinnen sprechen für die außergewöhnlich hohe Qualität der wissenschaftlichen Arbeit am HITS. Zugleich freuen wir uns über unsere neuen SAB-Mitglieder, die unsere Forschung mit ihrer Erfahrung und Expertise begleiten.

Allen Veränderungen zum Trotz: Das HITS konzentriert sich weiterhin auf Spitzenforschung in den Natur-, Mathematik- und Computerwissenschaften. Wir stellen hier einige Highlights vor:

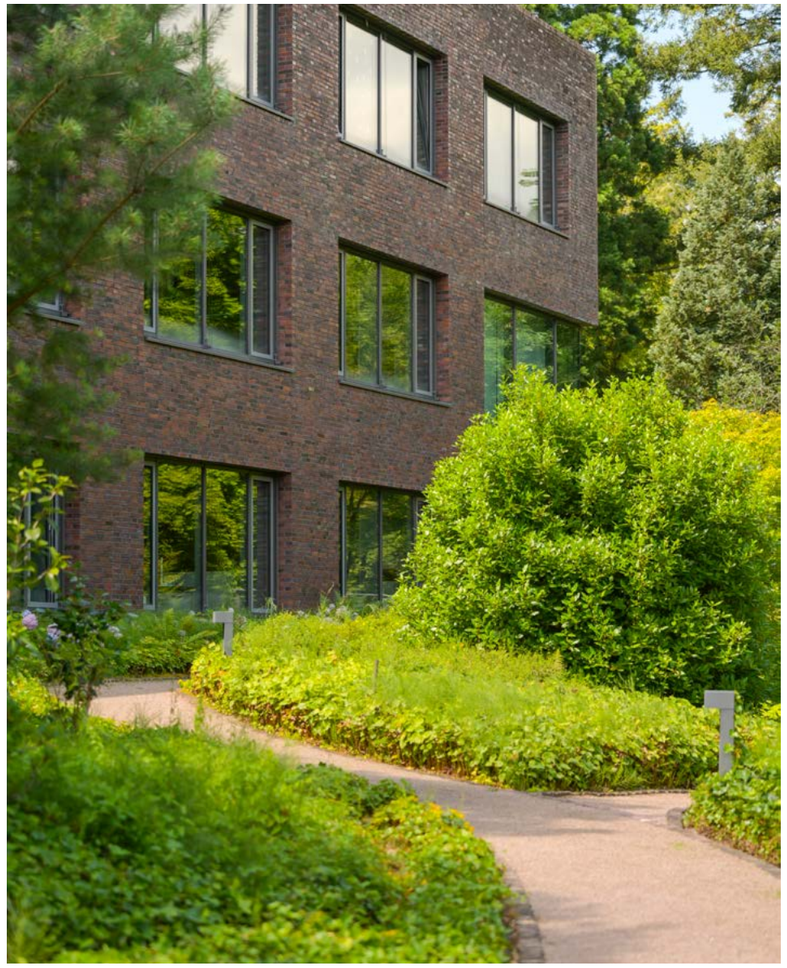
Frauke Gräter erhielt gemeinsam mit Ronen Zaidel-Bar (Tel Aviv University, Israel) und Alexander Dunn (Stanford University, USA) ein Forschungsstipendium des Human Frontier Science Program, um die Auswirkungen von physischer Kraft auf extrazelluläres Kollagen in Mäusen und C. elegans zu untersuchen.

Tilmann Gneiting erhielt den Ulf-Grenander-Preis 2024 für stochastische Theorie und Modellierung. Der Preis wird alle drei Jahre von der American Mathematical Society verliehen und würdigt bahnbrechende Arbeiten in den Bereichen probabilistische Modellierung, statistische Inferenz oder verwandte Computeralgorithmen, insbesondere für die Analyse komplexer oder hochdimensionaler Systeme (siehe Kapitel 10.5).

Zum neunten Mal in Folge wurde **Alexandros Stamatakis** von Clarivate Analytics als „Highly Cited Researcher" ausgezeichnet (siehe Kapitel 10.5).

Guillermo Cabrera-Vives, Informatiker an der Universität Concepción in Chile, war ab September 2024 als fünfter Klaus Tschira Gastprofessor am HITS. Guillermo nutzte die Zeit am Institut, um seine Zusammenarbeit mit der Astroinformatik-Gruppe (AIN) zu intensivieren,

präsentierte seine Forschung im Rahmen eines HITS-Kolloquiums und tauschte sich mit vielen anderen HITStern aus (siehe Kapitel 6.1).



Felicitas Mokler, eine deutsche Wissenschaftsjournalistin, war von April bis September 2024 im Rahmen des HITS Journalist in Residence-Programms am Institut. Während ihres Aufenthalts stand Felicitas in engem Kontakt mit den HITS-Forschenden und hielt einen öffentlichen Vortrag in der Mathematik-Informatik-Station (MAINS) in Heidelberg über Wissenschaftsjournalismus und Wissenschaftskommunikation (siehe Kapitel 4).

Jede Veränderung - ob groß oder klein - ist eine Gelegenheit, aus der Vergangenheit zu lernen und dies für die Zukunft zu nutzen. Dabei werden wir an unseren Kernzielen, der wissenschaftlichen Exzellenz und der innovativen Forschung, festhalten. Zugleich hat sich am HITS die einzigartige Kombination aus unterschiedlichen Forschungsbereichen, exzellenten Wissenschaftler*innen und perfekten Rahmenbedingungen für interdisziplinären Austausch bewährt. Unter dieser Prämisse diskutieren und planen wir aktuell mehrere neue Forschungsgruppen und sind dankbar, damit unsere Themengebiete erweitern und neue Forschungsansätze entwickeln zu können.

Wir wünschen Ihnen beim Lesen des Jahresberichts viel Vergnügen!

2 Research
2.1 Astroinformatics (AIN)

Group leader
Dr. Kai Polsterer

Team
Guillermo Cabrera-Vives (Klaus Tschira Guest Professor; since September 2024)
Catarina Corte-Real (PhD student; since October 2024)
Iliana Isabel Cortés Pérez (PhD student)
Mariia Demianenko (visiting scientist)
Dr. Nikos Gianniotis (staff scientist)
Fenja Schweder (visiting scientist; since February 2024)
Sebastian Müller (visiting scientist; until May 2024)
Dr. Franciso Pozo Nuñez
Johanna Riedel (bachelor's student)
Solomiya Serkiz (bachelor's student)
Dr. Sebastian Trujillo Gomez
Renuka Velu (master's student)

In recent decades, computers have revolutionized astronomy. Advances in technology have given rise to new detectors, complex instruments, and innovative telescope designs. These advances enable today's astronomers to observe more objects than ever before and at higher spatial, spectral, and temporal resolutions. In addition, new, untapped wavelength regimes along with other messengers – such as gravitational waves and astro-particles – are now granting more complete observational access to the Universe than ever before.

The Astroinformatics group deals with the challenges of analyzing and processing complex, heterogeneous, and large datasets. Our scientific focus in astronomy is on evolutionary processes and extreme physics in galaxies, such as those found around active 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 in the next generation of exa-scale simulations.

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

Probabilistic approaches to astronomical inverse problems

Modeling astronomical spectra is a key focus of the AIN group. Many of the challenges we address are inverse problems in which observed photometric data are linked to an unobserved, latent spectrum through an observational or physical mechanism. Taking this mechanism into account is important as it provides us with a guide as to how to design a probabilistic model that – when put into inference mode – produces a density of plausible, physically interpretable solutions to the inverse problem. The spectral model is a crucial component of such a probabilistic model. Below, we briefly report on two approaches concerned with learning a model for spectra from observed data using dimensionality reduction. We also report on recent work that aims to uncover how spectra vary over time given multi-band light curves.

A positive and scale-invariant Gaussian process latent variable model

One way of learning a model for spectra is by casting the problem as a dimensionality reduction problem. We view the high-dimensional spectra as manifestations of a set of abstract coordinates that reside in a space of significantly lower dimensionality. In our work, we employ the Gaussian process latent variable model (GPLVM), which models the high-dimensionality spectra as the images of low-dimensional coordinates under a probabilistic map. This mapping is governed by a Gaussian process (GP), which effectively learns the relationship between these spaces while automatically tuning hyperparameters – such as kernel parameters – and avoiding overfitting.

In our work, we endeavor to take into account two important properties of astronomical spectra: the fact that they are strictly positive-valued data, and the fact that their amplitude is not an intrinsic property and is therefore not relevant to our analysis. Accordingly, we extended the GPLVM to incorporate scale invariance, thereby ensuring that spectra with similar features are projected close together, regardless of differences in amplitude. From a technical perspective, enforcing both scale invariance and positivity results in an intractable log likelihood. We address this challenge using a variational Bayes approach. The variational formulation introduces a large number of free parameters that need to be optimized, but this problem can also be mitigated via amortized inference: Our solution is to parametrize the free parameters as a neural network map of the low-dimensional coordinates.

Figure 1 illustrates our approach applied to a set of high-dimensional astronomical spectra. The blue dots are the inferred two-dimensional coordinates. We notice that spectra group in clusters. We assign a number to each cluster and show three representative spectra of each cluster. We note that even though the spectra may differ in amplitudes, under the proposed extension, the model is scale-invariant and projects spectra close to one another as long as they are similar in terms of shape. In contrast, the standard GPLVM cannot achieve this kind of organization in the latent space because amplitude differences would influence the projection of the spectra.

Ensembles of neural network decoders

While the Gaussian process approach helps prevent overfitting and eliminates the need for dedicated hyperparameter optimization by learning a distribution of mappings between low- and high-dimensional spaces, it comes with a high computational cost when applied to large datasets of observed spectra. In our group, we explore an alternative method based on neural networks. Instead of learning a probabilistic map between the latent and data spaces, we assume that each high-dimensional spectrum is the image of a low-dimensional latent coordinate under a neural network map. Learning this mapping can be formulated as a regression problem in which the latent coordinates serve as inputs and the high-dimensional spectra as targets. However, jointly optimizing both the coordinates and the mapping tends to cause severe overfitting. In order to mitigate this problem, we employ an ensemble of neural networks, which provides strong regularization and stabilizes the learning process. In Figure 2 (next page), we show the reconstructions produced by our model when it is given observed noisy spectra with missing values. By pooling spectra observed at different redshifts, the model learns to

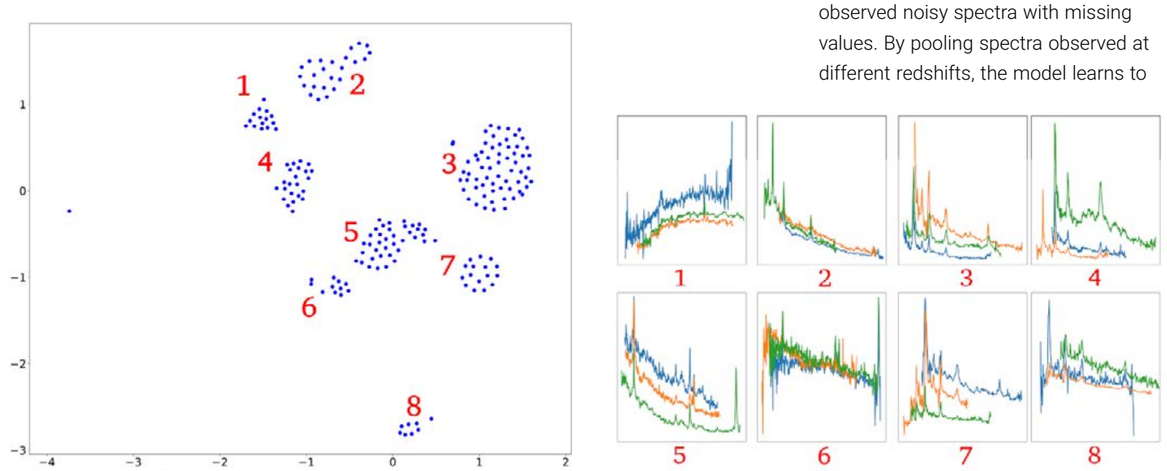


Figure 1: Two-dimensional visualization produced with a positive and scaled-invariant extension of the GPLVM. High-dimensional data organize in clusters. We display representative spectra of each numbered cluster.The fact that the representative spectra vary in amplitude indicates that the model effectively discounts the scale of the spectra.

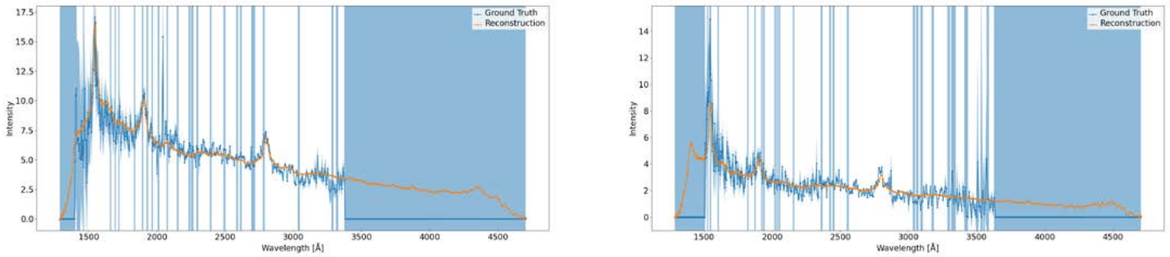


Figure 2: The blue dots correspond to spectral intensity observations, and the lighter blue shade shows the uncertainty of the measurements. The vertical stripes and blocks indicate highly uncertain observations or missing observations. The orange line is the reconstruction obtained via the proposed model. We note that the model can reconstruct the missing portions of the observed spectra.

extrapolate and produce reconstructions that span the entire wavelength range that we are interested in.

Photometric time series analysis of SMBH variability using Gaussian processes

The supermassive black hole (SMBH) activity at the center of active galactic nuclei (AGNs) is thought to be responsible for the observed flux variability over time. However, modeling the underlying mechanisms that drive this variability remains a non-trivial task, and traditional methods – such as reverberation mapping – require extensive observational data. Astronomical observations in the optical range are primarily obtained in two formats: photometry and spectroscopy. While spectroscopy reveals detailed information about the source's composition, it requires significantly longer exposure times than photometry, thereby making it a more resource-intensive technique. As a result, recent and upcoming advancements in time domain astronomy have prioritized the photometric time series of various sources.

Our work aims at disentangling the contributions of the thermal continuum and the non-thermal emission lines in photometric light curves by modeling a latent function (see lower panel of Figure 3) that emulates SMBH activity using a Gaussian process (GP). The GP framework allows us to recover a density of possible solutions, while its flexibility enables us to model AGN variability without imposing strict assumptions on the poorly understood physical processes that drive the complex dynamics of the SMBH. In Figure 3, we present an example in which our model is applied on a synthetic time series: We infer that the observed multi-band time series in the upper panel is driven by the postulated latent signal recovered and shown in the bottom panel. In the middle panel, we display reconstructions of spectra for a few selected points in time.

Physical analysis of galaxies via the spectral reconstruction of deep imaging

We have recently begun a new project that aims to reconstruct the latent spectrum underlying each pixel in multi-band images. The aim of the project is to learn a spatial–spectral model from integral field spectrometry (IFS), which provides detailed spectral information on extended astronomical sources, such as galaxies. Unfortunately, IFS data are not as widely available as are deep imaging data that capture only broadband flux and miss out on spectral detail. With a learned spatial–spectral model, we plan to analyze faint sources observed via deep imaging and to reconstruct the missing spectral information. Since this is clearly an ill-posed inverse problem, we expect that many different reconstructions could be possible solutions (see Figure 4). This situation calls for a probabilistic treatment that captures the set of reconstructions as a posterior density.

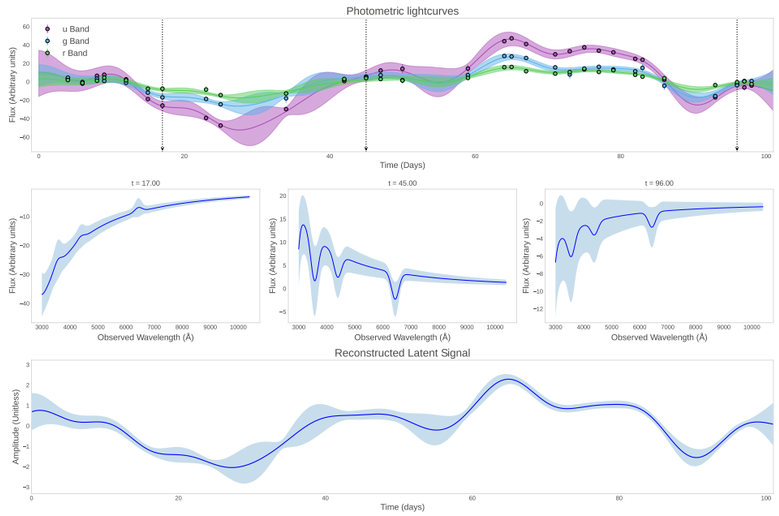


Figure 3: We demonstrate the proposed Gaussian process–based model on a synthetic, multiband time series displayed in the upper panel. The middle panel shows reconstructions of spectra at a few selected points in time, with the darker blue line indicating the expectation of the reconstruction and the lighter blue shade indicating its uncertainty. The bottom panel displays the reconstructed latent signal that drives the observed time series.

Re-evaluating LSST's capability for time delay measurements in quasar accretion disks

The Legacy Survey of Space and Time at the Vera C. Rubin Observatory is poised to observe thousands of quasars using deep drilling fields (DDFs) across six broadband filters over one decade. Understanding quasar accretion disk (AD) time delays is pivotal when it comes to probing the physics of these distant objects. Our recent work in [Pozo Nuñez, F, Bruckmann, C, Deesamutara, S, Czerny, B, Panda, S, Lobban, AP, Pietrzyński, G, and Polsterer, KL (2023). Modeling photometric reverberation mapping data for the next generation of big data surveys. Quasar accretion discs sizes with the LSST. Monthly Notices of the Royal Astronomical Society 522(2):2002–2018] (hereafter abbreviated as PN23) demonstrated the feasibility of recovering AD time delays with accuracies ranging from 5% to 20% depending on the quasar's redshift and time sampling intervals. In [Pozo Nuñez et al. 2024], we re-assessed the potential for measuring AD time delays under the current DDF observing cadence, which is a placeholder until a final cadence is decided. We found that contrary to prior expectations, achieving reliable AD time delay measurements for quasars is significantly more challenging – if not unfeasible – due to the limitations imposed by the current observational strategies.

We adapted the modeling approach outlined in PN23 for light curves, thereby conforming to the observed cadence of the DDFs. As a specific example, we used the baseline DDF cadence 8 from [Czerny, Bozena, et al. "Expectations for time delay measurements in active galactic nuclei with the Vera Rubin Observatory." Astronomy & Astrophysics 675 (2023): A163]. In order to develop the AD transfer functions, four test cases ("source type") were considered that were differentiated by redshift, black hole mass, and accretion rate (see Table 1 in PN23). For each test case, we generated 2000 random light curves spanning 3,700 days, each with a signal-to-noise ratio of 100 in each filter. Time delays relative to the u-band (the driver light curve) were calculated using the ICCF method by Gaskell and Peterson and also using the Von-Neumann method [Chelouche, Doron, Francisco Pozo-Nuñez, and Shay Zucker. "Methods of reverberation mapping. I. Time

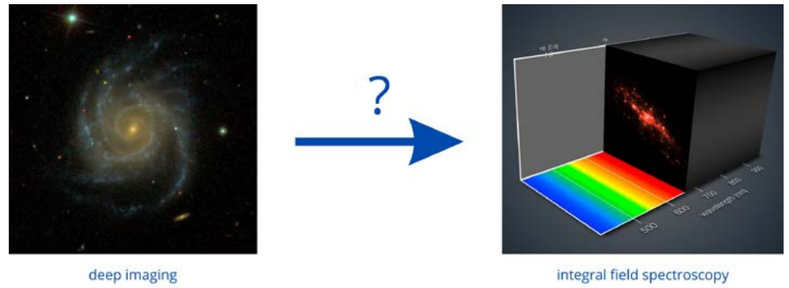


Figure 4: Illustration of the reconstruction of integral field spectroscopy (IFS) data cubes using deep imaging inputs, which is an ill-posed problem.

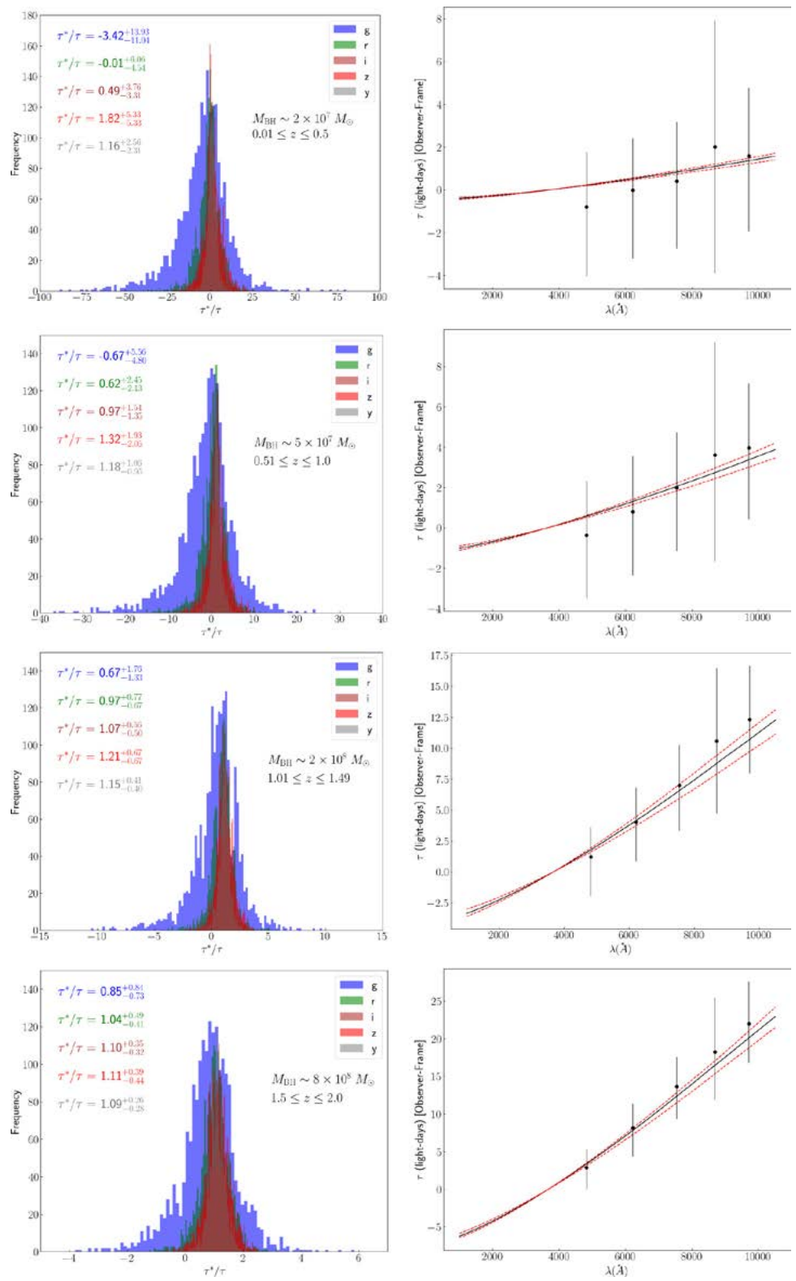


Figure 5: Recovered distributions of delays for different redshift and black hole masses. Right: time delay spectrum as predicted from the response functions. Left: The black circles represent the mean delays and the error bars of the standard deviation expected for a single source measurement. The dotted red lines show the delay spectrum obtained for a black hole mass with 30% uncertainty.

lag determination by measures of randomness." The Astrophysical Journal 844.2 (2017): 146]; however, for brevity, we present only the ICCF results due to the method's widespread use and comparable performance.

Time delays for quasars at redshifts $1.5 < z < 2.0$ can be recovered with an accuracy of between 5% and 10% for a time sampling interval of 2–5 days. This analysis applies to black hole masses in the range of approximately 10^8–10^9 $M_\odot$. However, recovering the time delay for smaller black hole masses – specifically those below 5×10^7 $M_\odot$ – is not feasible with the current baseline DDF cadence. In order to extend this capability to lower mass ranges, a denser time sampling interval of 1–2 days is required for all bands and without significant gaps (in order for a light curve duration of at least twice the time delay to be detected), which could improve the precision of recovered delays to about 10%–20%. We refer to Figure 5 (previous page) for an illustration of our results. Consequently, for effective quasar AD time delay science, we recommend that the DDF focus on shorter monitoring periods – such as a single season of 6 months – with a time sampling interval of 1 day. This recommendation aligns with those presented by the AGN Science Collaboration's accordion cadence note 9. This accordion cadence is proposed to be dense and uniform, but only over a period of 2.5 months in a given year in order to preserve the total depth measurement requirements, while PN23 claims very satisfactory results assuming six months for the duration of the monitoring. As the question of cadence has not yet been resolved, it will later be important to examine in detail the proposition based on new developments in cadence optimization.

Data-driven discovery with the largest surveys and numerical simulations

Astronomers like tradition. Indeed, we still use naming conventions and measurement scales dating back hundreds or even thousands of years. We classify galaxies using the categories that Edwin Hubble developed nearly one century ago. Despite the long list of breakthroughs ushered in by traditional data analysis techniques, astrophysics is currently undergoing a paradigm shift as the data from the largest-ever surveys of the sky begin to pour forth in quantities that no scientist could ever attempt to inspect, let alone analyze. There is vast potential for discovery in this deluge of data, but exploiting the information will require a complete re-thinking of our analysis techniques and workflows.

In the meantime, machine learning (ML) now allows computers to learn and extract useful information from datasets nearing the size of the entire human knowledge. However, ML is primarily concerned with making predictions based exclusively on data, with little concern for providing explanations. Developing explanations (i.e., models) is the main goal of any field of science, which severely limits the applicability of ML to scientific discovery.

Recent developments in generative deep learning (GDL) now allow us to peek inside the black box. Using representation learning techniques, these algorithms can extract valuable insights from large datasets by projecting complex high-dimensional data onto understandable simple representations that humans can easily visualize and understand. Our group is now adopting these techniques in order to develop a novel data-driven analysis workflow for studies using multimodal data (e.g., spectra, images, time series, and data cubes) from the enormous surveys undertaken by observatories like Gaia, Euclid, SKA, Rubin, and Roman across the wavelength spectrum. These surveys aim to understand the structure of the Milky Way by mapping billions of stars, the formation and evolution of billions of galaxies and their components, the evolution of the Universe as a whole, and the elusive nature of dark matter and dark energy.

Our scientific workflow is focused on extracting the maximum amount of useful information directly from the survey data products without the need for scientists (or humans in general) to painstakingly comb through all the data. The framework allows the user to select from various state-of-the-art GDL algorithms in order to train on arbitrarily large survey datasets. The learned compact representation contains all the main features of the data and can be projected onto an interactive hierarchical visualization that displays the entire dataset as well as the properties of individual objects of interest to the user. A key component of the workflow is the automatic learning of the relevant features: In an example from the Spherinator tool, the algorithm finds the most relevant morphological characteristics that define the galaxy population based only on a large set of galaxy images. This effectively improves and automates the process behind Hubble's famous discovery of the tuning fork diagram.

However, the data are only half of the scientific method. In order to provide hypotheses and build theories, sophisticated numerical simulations that match the level of detail of modem observations are now essential. Our new workflow is designed to apply the same dimensionality reduction techniques to the simulations, which have themselves become challenging to interpret due to their size and high dimensionality. The need to avoid moving large data volumes from their storage locations in high-performance computing (HPC) centers and survey databases renders this process extremely challenging, but we are currently incorporating existing solutions for users with different requirements. This step allows for the deployment of the workflow on a variety of architectures ranging from HPC sites to the user's laptop.

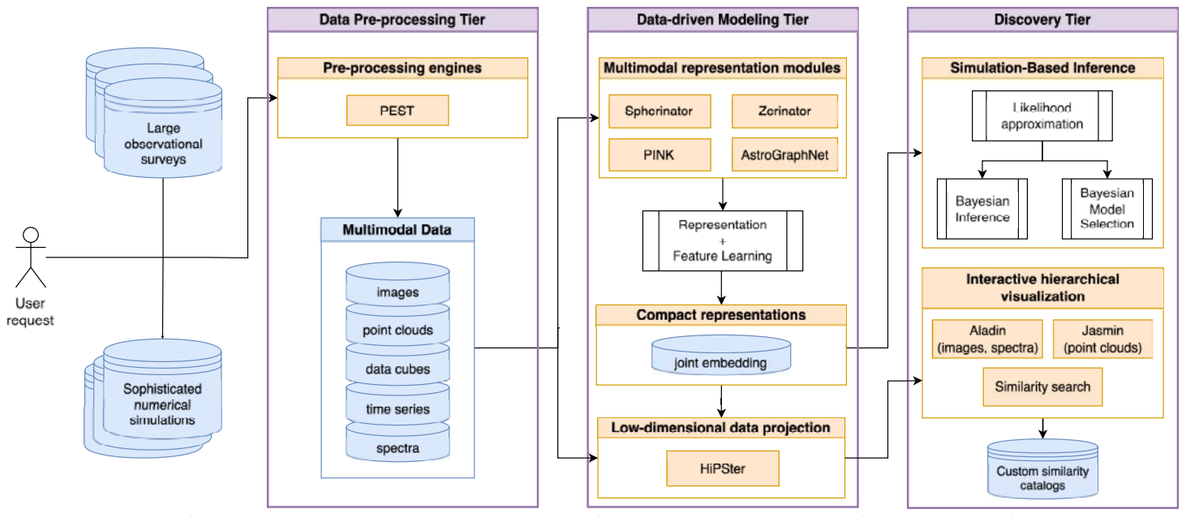


Figure 6: Schematic of our new data-driven astrophysics discovery workflow. The seamless integration of multimodal data from large observational surveys and numerical simulations enables the use of representation learning methods to find an ideal joint compact representation of large datasets. The representation can be used for various discovery tasks, including interactive hierarchical visualization and exploration as well as simulation-based inference.

The last stage of our data-driven workflow consists of learning a detailed representation of both the survey data and the synthetic data from the simulations. This step is the key to enabling statistical inference and hypothesis testing in a Bayesian framework using recently developed simulation-based inference (SBI) techniques. SBI provides a statistically rigorous way to compare observations and simulations by approximating likelihood. Coupled with information-maximizing data representations, SBI enables scientists to answer the most important question: Does my model (or do my models) explain the observational data? Figure 6 displays a schematic of the workflow.

We are also simultaneously developing the first application of these methods to currently available datasets in order to obtain data-driven non-parametric descriptions of the observed structure of galaxies and their theoretically predicted dark matter halos. The well-known "small-scale" challenges of the ΛCDM cosmological model point to significant discrepancies between simulation predictions and observations of low-mass galaxies (see Figure 7). Several of these problems have persisted for decades due to the unknown uncertainties in the simulations and to the inherent difficulty in directly comparing observed and simulated data. The only way to move forward requires Bayesian data-driven comparisons of large multimodal observational datasets with the predictions of the concordance cosmological model.

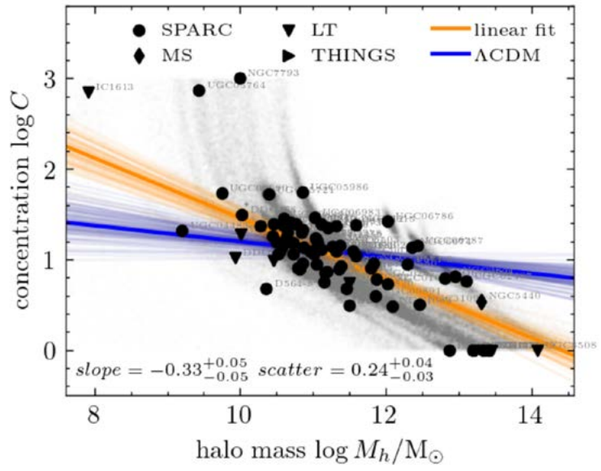


Figure 7: The structure of dark matter halos inferred from currently available galaxy dynamics datasets compared with the predictions of the concordance cosmological model. The statistically significant discrepancy between the predicted and observed relations (colored lines) may indicate a failure of the model. An approach employing non-parametric representations learned directly from the observational and simulated data will be needed to understand the origin of the discrepancy.

In den letzten Jahrzehnten hat der Einsatz von Computern die Astronomie stark beeinflusst. Der technologische Fortschritt ermöglichte den Bau neuer Detektoren und innovativer Instrumente sowie neuartiger Teleskope. Damit können Astronomen nun mehr Objekte als je zuvor mit bisher unerreichtem Detailreichtum, sowohl räumlich, spektral als auch zeitlich aufgelöst beobachten. Hinzu kommen neue Beobachtungsmöglichkeiten durch beispielsweise Astroteilchen sowie Gravitationswellen, die neben bisher nicht beobachtbaren Wellenlängenbereichen ein vollständigeres Bild des Universums bieten.

Die **Astroinformatik Gruppe** beschäftigt sich mit den Herausforderungen, die durch die Analyse und Verarbeitung dieser komplexen, heterogenen und großen Daten entstehen. In der Astronomie beschäftigen uns die Fragestellungen im Bereich der Galaxienentwicklung sowie die extremen physikalischen Vorgänge, wie man sie z.B. in der Umgebung von aktiven supermassereichen schwarzen Löchern in den Zentren von Galaxien findet. Auf diesen Fragestellungen basierend, entwickeln wir neue Methoden und Werkzeuge, die wir frei zur Verfügung stellen. In der Informatik liegt unser Interesse auf der Zeitreihenanalyse, dem Umgang mit spärlichen Daten, der morphologischen Klassifikation, der richtigen Auswertung und dem richtigen Training von Modellen sowie explorativen Forschungsumgebungen. Diese Werkzeuge und Methoden sind eminent wichtig für aktuelle und sich gerade in der Vorbereitung befindende Projekte, wie SKA, Gaia, LSST und Euclid sowie Daten aus exa-scale Simulationen.

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

2.2 Computational Carbon Chemistry (CCC)

Group leader

Dr. Ganna (Anya) Gryn'ova
(until March 2024)

Team

Alessandro Calzolari (PhD student; until March 2024)
Dr. Christopher Ehlert (until May 2024)
Rostislav Fedorov (PhD student)
Lilliana Florido Martins (student intern, Heidelberg University; January–February)
Carina Herrle (student research assistant, Heidelberg University; January–March)
Gregor Lauter (PhD student; until March 2024)
Lukas Lehr (student research assistant, Heidelberg University; January–March)
Stiv Llenga (PhD student; until May 2024)
Tobias Melzer (student intern, Heidelberg University; January–February)
Dominique Ostermayer (PhD student; until March 2024)
Anna Piras (PhD student; until March 2024)
Tergite Zeqiri (student intern, Heidelberg University; January–February)
Levi Ziegenhagen (student intern, Heidelberg University; January–February)

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

In the Computational Carbon Chemistry (CCC) group, we use computational chemistry, physics, materials science, and artificial intelligence to explore and exploit the chemistry of functional organic molecules and materials that are pertinent to applications in sustainable (electro)catalysis, molecular sensing, capture, storage, transport, delivery, and organic electronics. We are particularly interested in the hybrid molecule-material systems and in the role that both shape and topology play in these systems' chemistry. Our work involves constructing databases and multiscale modeling workflows, mining and digitizing literature data, as well as developing diverse tools for chemical machine learning and explainable artificial intelligence. Our ultimate goals are (i) to uncover new fundamental chemical phenomena and (ii) to design new and improved functional molecules and materials.

History and outlook

The Computational Carbon Chemistry group was founded at HITS in April 2019 with the appointment of Ganna (Anya) Gryn'ova as Junior Group Leader. In April 2024, after exactly 5 years, the group moved to the University of Birmingham in the UK, where Anya became Associate Professor of Computational Chemistry. Below, we reflect on the five years of exciting research and lasting collaborations that the group forged at HITS.

The CCC group grew rapidly from a sole PI in 2019 to a cohort comprising 6 PhD students, several postdocs, and multiple student assistants and interns by 2024. Anna Piras was the group's first PhD student to graduate magna cum laude from Heidelberg University in March 2024 after defending her thesis entitled "Predicting Adsorbent and Electronic Properties of Graphene-Based Materials." Anna and many other group alumni went on to continue their research careers as PhD students and postdocs at various institutions in Germany, Switzerland, and Italy, while several others moved on to research and development careers in industry. Gregor Lauter and Stiv Llenga transferred to the University of Birmingham with Anya, whereas Rostislav Fedorov stayed at HITS, was "adopted" by the MCM group, and is due to defend his PhD thesis on the geometric deep learning of redox active materials at Heidelberg University in the fall of 2025.

The CCC group's research also grew in its breadth, with areas ranging from the high-level modeling of graphene-based sensors to the machine learning–assisted exploration of the vast chemical space (Figure 8). Michelle Ernst joined the group in 2020 as an Early Postdoc. Mobility fellow funded by the Swiss National Science Foundation, bringing her expertise in organic framework materials. Stiv Llenga and Rostislav Fedorov became the group's experts on chemical machine learning, while Christopher Ehlert led the group's research on method benchmarking and theoretical chemistry. In its five years at HITS, the CCC group developed computational workflows for simulating molecular adsorption on graphene-based materials, elucidated the design guidelines for bespoke organic framework materials, and put forward a new quantum-inspired molecular representation. These efforts were greatly supported by (i) funding from the Klaus Tschira Stiftung (Klaus Tschira Foundation) through both the SIMPLAIX Strategic Research Initiative for Data-enhanced Multiscale Simulation and the HITS Lab, (ii) the German Research Foundation (DFG) through Collaborative Research Centre SFB1249 "N-Heteropolycycles as Functional Materials," and (iii) the European Research Council through Anya's 2021 Starting Grant, "PATTERNCHEM: Shape and Topology as Descriptors of Chemical and Physical Properties in Functional Organic Materials." Within and beyond these networks, the CCC group established fruitful collaborations with the groups of Frauke Graeter (HITS) and Lutz Greb (Heidelberg University), among others.

At the University of Birmingham, the group continues to pursue these research avenues with the goal of gaining

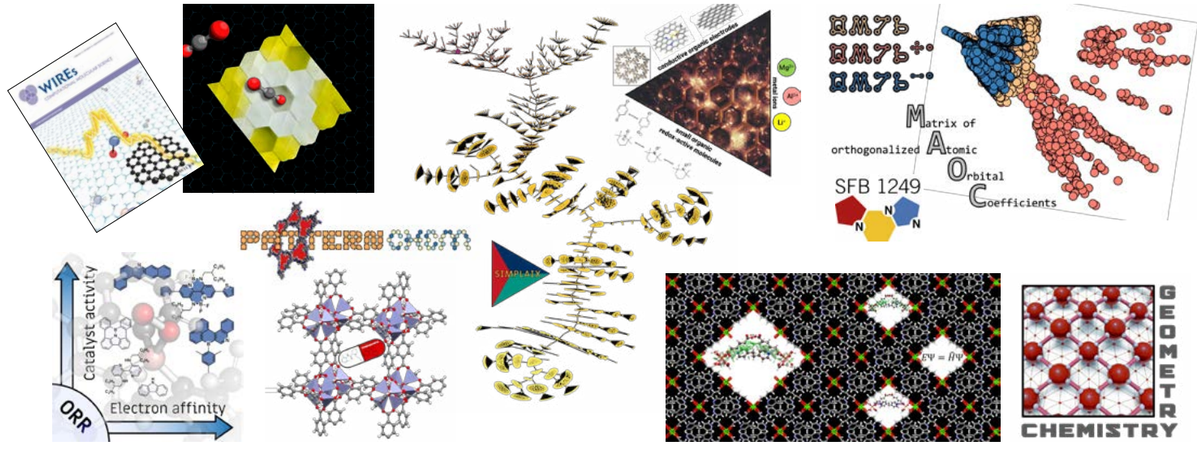


Figure 8: Highlights of the CCC research from 2019–2024.

a better understanding of – and reliable design strategies for – topologically complex functional organic materials (Figure 9). Lukrecia Mertova – a PhD student in the SDBV group at HITS – visited the CCC group in Birmingham for a 6-month research stay, igniting new projects and collaborations on chemical data mining, digitalization, and contextualization. Building on the successes at HITS, the CCC group continues to make its mark in the fields of computational materials chemistry and chemical machine learning, with Anya co-organizing a successful first edition of the Chemical Compound Space Conference (CCSC) in Heidelberg in May 2024 and joining the editorial board of the npj Computational Materials in late 2024.

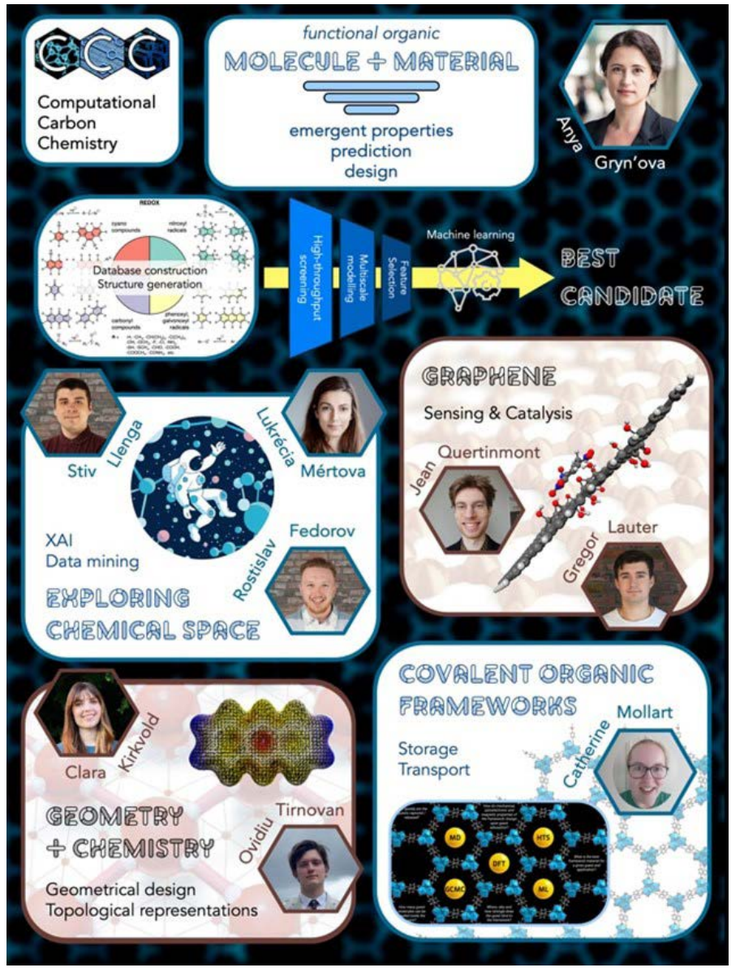
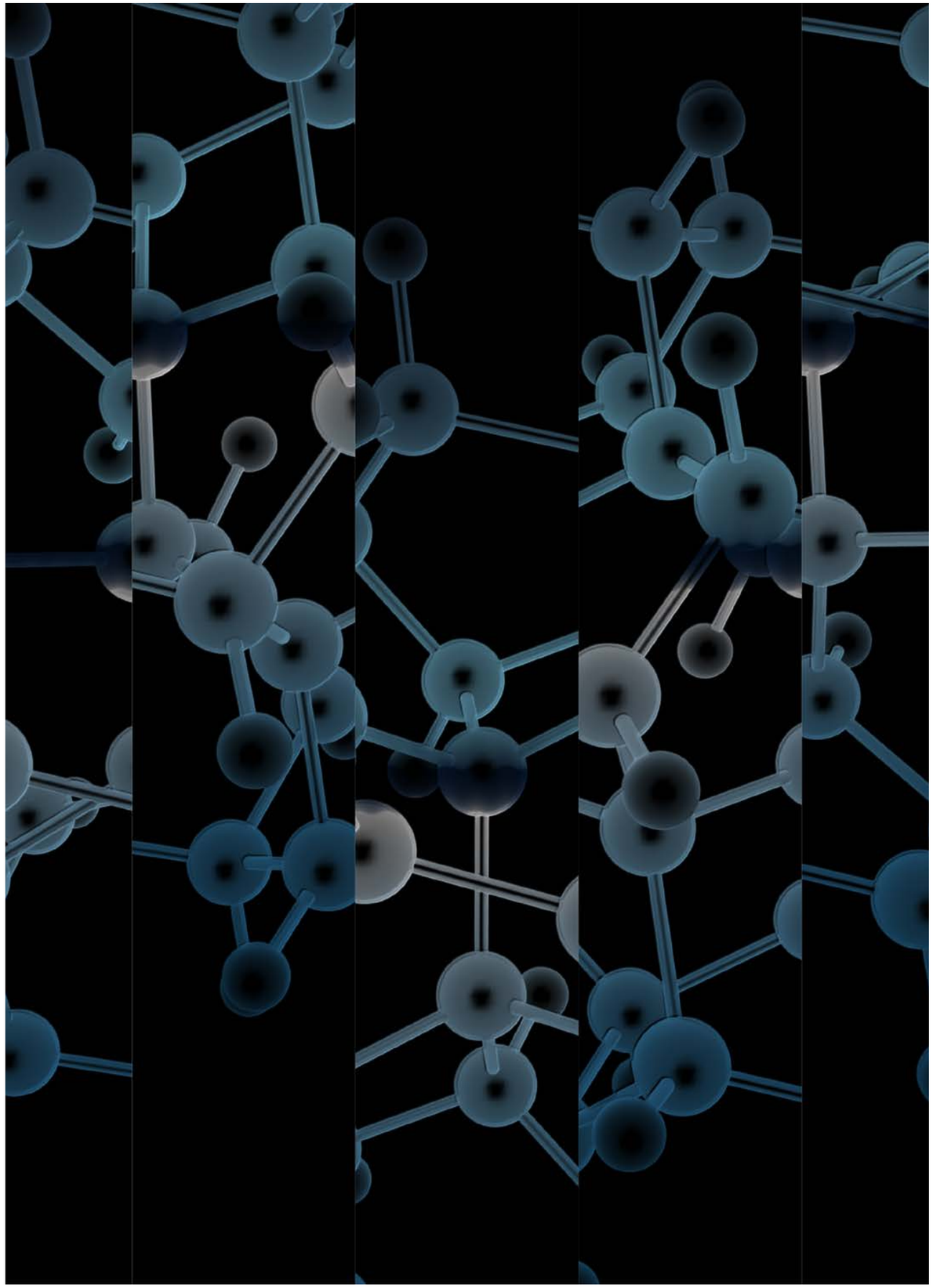


Figure 9: Current composition of – and key research themes in – the CCC group in Birmingham.

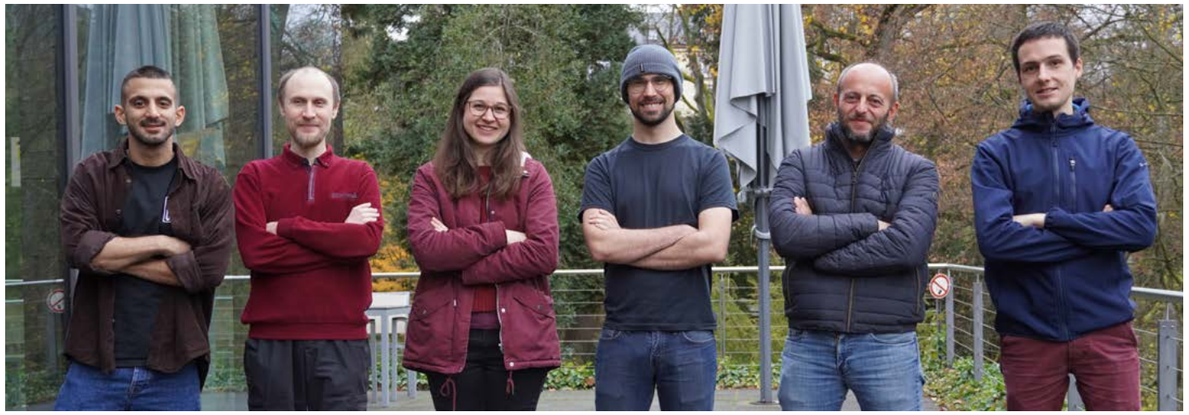
Moderne Funktionsmaterialien, wie beispielsweise Graphen und organische Gerüststrukturen, kombinieren strukturelle Komplexität mit zielgerichteter Performance und werden in verschiedenen Bereichen von Industrie und Forschung eingesetzt, von der Nanoelektronik bis hin zur Massenfertigung. Theoretische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

In der Forschungsgruppe **Computational Carbon Chemistry (CCC)** nutzen wir computergestützte Chemie, Physik und Materialwissenschaft in Kombination mit maschinellem Lernen, um die Chemie funktionaler organischer Moleküle und Materialien zu untersuchen und auszuwerten. Dies ist für die Anwendungen in der nachhaltigen (Elektro-)Katalyse, der molekularen Sensorik, der Erfassung, Speicherung, dem Transport, der Freisetzung sowie der organischen Elektronik relevant. Wir sind besonders an hybriden Molekül-Material-Systemen interessiert sowie am Einfluss von Form und Topologie auf deren chemische Eigenschaften. Unsere Arbeit umfasst den Aufbau von Datenbanken und Multiskalenmodellierung-Workflows, das Extrahieren und Digitalisieren von Literaturdaten sowie die Entwicklung verschiedener Werkzeuge für chemisches maschinelles Lernen und erklärbare künstliche Intelligenz. Unsere übergeordneten Ziele sind (i) die Entdeckung neuer fundamentaler chemischer Phänomene und (ii) das Design neuer und verbesserter funktionaler Moleküle und Materialien.

Die Gruppe wurde im April 2019 gegründet, mit der Verpflichtung von Ganna (Anya) Gryn'ova als Juniorgruppenleiterin. Im April 2024 nahm Anya einen Ruf als Associate Professor an der University of Birmingham/UK an und verließ zusammen mit ihrer Gruppe das HITS – nach genau fünf erfolgreichen Jahren am Institut.



2.3 Computational Molecular Evolution (CME)



Group leader

Prof. Dr. Alexandros ("Alexis") Stamatakis

Team

Mattis Bodynek (master's student; since December 2024)
Erik Borker (master's student, until September 2024)
Julia Haag (PhD student; HITS Scholarship)
Luise Häuser (PhD student; HITS Scholarship)
Johannes Hengstler (PhD student; HITS scholarship; since October 2024)
Dimitri Höhler (PhD student)
Lukas Hübner (visiting scientist from KIT)
Lukas Knirsch (student, until September 2024)
Dr. Alexey Kozlov (staff scientist)
Eric Laudemann (master's student)
Luc Mercatoris (master's student, until february 2024)
Dr. Benoit Morel (postdoc, until February 2024)
Dominik Siebelt (bachelor's student)
Christoph Stelz (student)
Alex Suhrkamp (master's student)
Anastasis Togkousidis (PhD student; HITS Scholarship)
Julius Wiegert (master's student, until March 2024)
Julian Wu (master's student; since December 2024)

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.

Secondary research interests include
- emerging parallel architectures,
- discrete algorithms on trees,
- ancient DNA analysis, and
- population genetics.

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

In other words, we strive to support research. One aim of evolutionary biology is to infer evolutionary relationships between species and the properties of individuals within populations of the same species. In modern biology, evolution is a widely accepted fact that can be analyzed, observed, and tracked at the DNA level.

As evolutionary biologist Theodosius Dobzhansky's famous and widely quoted dictum states, "Nothing in biology makes sense except in the light of evolution."

What happened in the lab in 2024?

Alexis completed the second 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 has been fully staffed (with 3 postdocs and 3 PhD students) since early 2024 and is closely linked to the CME group at HITS. For instance, former CME master's student Noah Wahl joined the BCG as a PhD student, and former CME PhD student Ben Bettisworth joined the BCG as a postdoc. Within this context, many exchanges have already taken place. CME staff scientist Alexey Kozlov and former CME PhD student Lucas Czech visited the BCG in October. In addition, CME PhD student Julia Haag attended the "Machine Learning for Evolutionary Genomics Data" conference that Alexis organized on Crete and also gave two talks. Furthermore, BCG members Georgios Koutsovoulos, Ben Bettisworth, Noah Wahl, and Lucia Martin Fernandez were or are actively involved in supervising several students and master's theses at the Karlsruhe Institute of Technology (KIT).

In the winter of 2023/24, we launched a new teaching endeavor with the joint 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 an all-time record high in exam registrations (40 at KIT and 11 at UoC). Approximately half of the live lectures took place at KIT, and the other half took place at UoC, with all lectures being streamed via Zoom to the other university.

Because the summer semesters at UoC and KIT mostly do not overlap, Alexis also taught a new seminar on "Reproducibility in Bioinformatics" during the UoC spring term. Finally, during the summer semester of 2024, we again taught our main seminar, "Hot Topics in Bioinformatics," at KIT.

Johannes Hengstler – our master's student from the Department of Computer Science at KIT – joined the lab as a PhD student in the fall of 2024 and is initially focusing on using tertiary protein structure data to infer phylogenies. In addition, the 2023 trend of supervising a large number of bachelor's and master's theses at KIT continued: Last year, 1 bachelor's and 6 master's theses were completed, and a total of 3 student projects are currently ongoing, whereas beginning with the establishment of the CME in 2010 until about 2022, there used to be only 2 or 3 theses per year. Finally, our staff scientist Alexey Kozlov continued his new 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 2024 in Hinxton, UK, for the 14th time (see Chapter 5.1). Dimitri supported the summer school as a teaching assistant, and Alexis served as a lecturer and co-organizer. The next iteration of our summer school is planned for May 2025, this time on Crete again, and will also feature a satellite workshop.

Additionally in 2024, Alexis was listed on the Clarivate Analytics list of highly cited researchers for the ninth year in a row. For this reason, Alexis was also invited to the President's Evening of Honor in November 2024, where the Executive Board of KIT honored the scientific achievements and social commitment of KIT researchers (see Chapter 10.5).

The year additionally consisted of a plethora of public outreach activities. For instance, Alexey presented his tool for energy-efficient computing at the HITS Open House (see Chapter 5.4), and Alexis also taught the basics of air traffic control to two separate classes of elementary school children in Southern Crete, with the teaching material having been developed by Alexis in collaboration with a Greek air traffic controller. Moreover, BCG member Giorgos Koutsovoulos also gave an outreach talk on the first results of our DNA-based insect biodiversity monitoring project in the village of Listaros (where the monitoring stations are set up). The evening was organized by Alexis in the village square.

Figure 10: CME and BCG members at HITS.

Figure 11: BCG biodiversity excursion in Crete.

Another highlight of 2024 was the organization of the first iteration of the "Machine Learning for Evolutionary Genomics Data" conference on Crete, for which Alexis served as the main organizer. The conference attracted approximately 80 researchers from across Europe and the US to Crete.

In sum, 2024 was dominated by the completion of the BCG setup on Crete and by the further tightening of the connections between the two groups, while all activities at HITS and KIT in Germany continued as in the years before.

INTRODUCTION

The term "computational molecular evolution" refers to computer-based methods of reconstructing evolutionary trees from DNA or – for example – from protein data or morphological data.

The term also refers to the design of programs that estimate statistical properties of populations – that is, programs that disentangle evolutionary events within a single species.
The very first evolutionary trees were inferred manually by comparing the morphological characteristics (i.e., traits) of the species under study. Today, in the age of the molecular data avalanche, manually reconstructing trees is no longer feasible. Evolutionary biologists thus have to rely on computers and algorithms for phylogenetic and population genetic analyses.

Ever since the introduction of so-called "short-read sequencing machines" (i.e., machines that biologists use in the wet lab to extract DNA data from organisms), scientists have been able to generate over one billion short DNA fragments (each containing between 30 and 400 DNA characters) as well as continuous improvements in sequencing technology that also allow substantially longer reads to be generated. As a result, the community as a whole now faces novel challenges. One key problem that needs to be addressed is the fact that the volume of molecular data available in public databases is growing at a significantly faster rate than the computers that are capable of analyzing the data can keep up with. In addition, the costs of sequencing a genome are decreasing at a faster rate than are the costs of computation, although the curve seems to have been flattening out in the last 3–4 years (see https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost).

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

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

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 Julius Wiegert, we managed to show that we can predict statistical support values that reflect the degree of certainty we have in specific branches of an evolutionary/phylogenetic tree using machine learning. The prediction is highly accurate and is 10 times faster than conventional mechanistic models.

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 geographical, 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. Our PhD student Luise Häuser won the Best Paper Award at a conference with peer-reviewed full paper contributions for work on the use of sounds to infer language histories.

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.

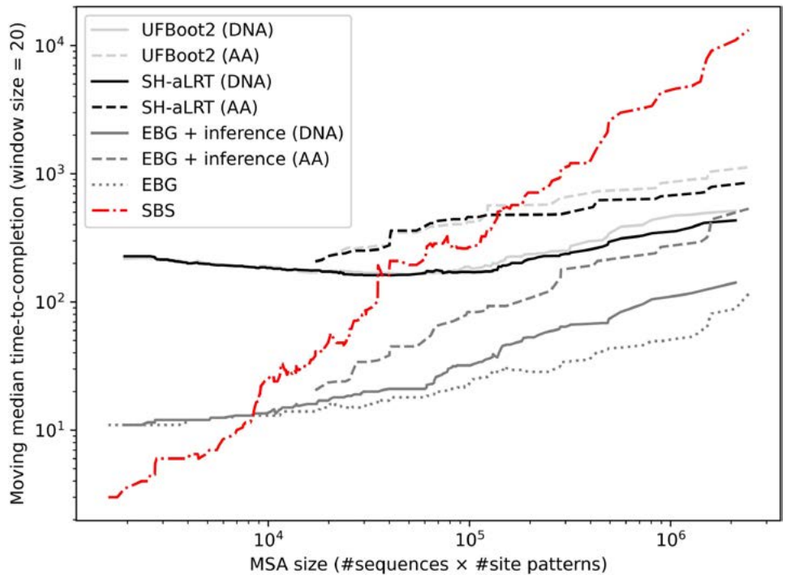


Figure 12: Support value calculation times (y-axis) over input dataset size (x-axis) for distinct support value calculation methods. Note the logarithmic scale on both axes! EBG is our new machine learning–based tool; SH-aLRT and UFBoot2 are competing fast approximate methods for computing support values; and SBS refers to the standard, slow Felsenstein bootstrap. We show different graphs for DNA and protein input data.

Rapid Prediction of Phylogenetic Support Values via Machine Learning

An important operation in phylogenetic inference is the computation of branch support values – that is, statistical support measures that reflect the confidence we have in each inner branch/split of the phylogenetic tree we have reconstructed. These support values are traditionally computed via the famous Felsenstein Bootstrap procedure, which is extremely compute-intensive and largely dominates phylogenetic inference times. Over the years, researchers have therefore attempted to devise computationally more efficient approximations of the Felsenstein bootstrap in order to keep pace with the aforementioned molecular data avalanche. In the field, attempts are currently being made to approximate bootstrap support values via machine learning methods: In fact, with our outstanding KIT master's student Julius and the excellent supervision provided by CME PhD students Julia and Dimitri, we have achieved exactly that with the development of the open-source tool we call Educated Bootstrap Guesser (EBG; paper: https://academic.oup.com/mbe/article/41/10/msae215/7825466). Now, we can predict bootstrap support values with high accuracy while being one order of magnitude faster than current approximate methods and two orders of magnitude faster than the traditional Felsenstein Bootstrap (see Figure 12) as implemented in our flagship tool for phylogenetic inference, RAxML-NG. While Julius has unfortunately left the CME, we are still collaborating with him on integrating EBG into the next major RAxML-NG release. This integration has already resulted in a substantial improvement in runtimes for feature calculation.

Removing Subjectivity and Labor-Intensive Tasks from Linguistics

Along our new line of work on computational linguistics that is being driven by CME PhD student Luise Häuser, and specifically regarding the reconstruction of phylogenies for disentangling the evolutionary history of natural languages, we stumbled across the problem of synonym selection in dataset assembly. Language datasets typically contain very basic concepts that are present in every language, such as "man," "women," "hand," and so forth, and these datasets group together languages that comprise words of the same origin. Consider the concept of "big," for example: In English, at least two synonymous words for this concept exist: "big" and "great." The problem now is which word to select to represent the concept of "big" in English because choosing "great" will induce a relationship with German ("groß") and Dutch ("groot"),

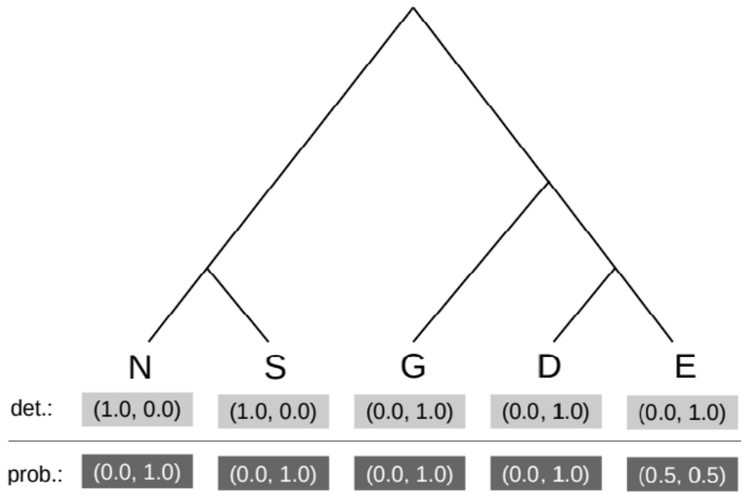


Figure 13: Deterministic (det. traditional) subjective approach to synonym modeling and new probabilistic (prob.) objective approach to synonym modeling for the concept of "big" in Norwegian (N), Swedish (S), German (G), Dutch (D), and English (E). In English (bottom right), we assign an equal probability of being the most common word of 0.5 to both "great" and "big," thereby integrating out subjectivity.

whereas choosing "big" will not induce any relationship with these words. Linguists traditionally select the synonym that they perceive as being more commonly used, which is presumably "big" in the current case.

For us computer scientists, there are two things we do not like about this approach: First, it is extremely labor intensive (can the procedure be automated?), and second, it is extremely subjective (can the subjectivity be eliminated?) and also potentially misleading. By using a broad set of linguistic datasets, we showed that this subjective or random synonym choice can yield language phylogenies that are highly incongruent with what is considered among linguists to be the gold standard reference tree (https://aclanthology.org/2024.scil-1.16.pdf). In addition, we managed to show that by making an appropriate modification of the likelihood function that can seamlessly take into account the uncertainty that is induced by synonym choice, we can obtain accurate trees with respect to the gold standard without needing to conduct an explicit synonym selection (see Figure 13).

A New Line of Work: Research Policy

Based on the specific role of ERA chairs as specified by the EU – that is, "Encouraging institutional reforms in research institutions and in the national R&I system in widening countries" – Alexis has become active in the area of shaping research policy. More specifically, with his colleague Melina Tamiolaki, he initially published an article in the monthly journal of the German Association of University Professors entitled "Substantially Underfunded – Challenges in the Greek Research Landscape" ("Substanziell unterfinanziert - Probleme der griechischen Forschungslandschaft," PDF at https://www.biocomp.gr/assets/downloads/Unterfinanziert.pdf). With Melina and his colleague Panagiotis Tsakalides (both of whom are coordinators of ERA chair programs), Alexis then wrote an extended version of this article and managed to publish his first-ever peer-reviewed political science paper on "Necessary reforms in the Greek academic system" in Frontiers in Political Science (open access link: https://www.frontiersin.org/journals/political-science/articles/10.3389/fpos.2024.1471002/full). The paper covers a broad range of topics and respective policy proposals, including non-competitive salaries, insufficient national research funding, gender equality, intellectual property rights, and so forth. For each topic, we proposed solutions that were classified by the associated costs (zero, low, high). The paper was highly acclaimed by colleagues in Greece and also made it to the popular press (https://www.dnews.gr/eidhseis/science/500516/oi-erevnites-sti-xora-mas-zoun-mia-peripeteia).

Another policy-related activity consisted of defining and popularizing the term "specimen drain," which was invented by Alexis and his colleague Nikos Poulakakis, the director of the Natural History Museum of Crete. The term describes the phenomenon of poor countries' exportation of important specimens (e.g., biodiversity specimens or archeological specimens) to research centers in the Global North because these poor countries do not have the necessary equipment or funding to analyze the samples and thereby lose the lead (i.e., the first/last author position) of the analyses and respective publications. This topic also made its way to the popular press (https://www.dnews.gr/eidhseis/science/500237/meta-to-brain-drain-i-ellada-vionei-kai-ti-di-arroi-deigmaton-gia-epistimoniki-erevna).

Finally, Alexis continued his efforts to establish a "National Female Professors and Researchers Program" in Greece and led an initiative to draft a corresponding proposal. In December of 2024, he had a meeting with policymakers in Athens to present the proposal (see Figure 14).

Figure 14: Together with Maria Klapa, his dear colleague and president of the Gender Equality Committee at FORTH, Alexis met with the Deputy Minister of Equality and Human Rights and her staff to present the concept of the "National Female Researchers and Professors Program," which specifically aims to increase the representation of women in academic management. The proposal is supported by the gender equality committees of all Greek universities and all research centers. Photo: Maria and Alexis in the corridor of the ministry in front of the Greek word for equality

Die Gruppe **rechnerbasierte Molekulare Evolution (CME)** beschäftigt sich mit Algorithmen, Modellen und dem Hochleistungsrechnen für die Bioinformatik.

Unsere Hauptforschungsgebiete sind:
- Rechnerbasierte molekulare Stammbaumrekonstruktion
- Analyse großer evolutionsbiologischer Datensätze
- Hochleistungsrechnen
- Quantifizierung von Biodiversität
- Analysen von „Next-Generation" Sequenzdaten
- Qualität & Verifikation wissenschaftlicher Software.

Sekundäre Forschungsgebiete sind unter anderem:
- Neue parallele Rechnerarchitekturen
- Diskrete Algorithmen auf Bäumen
- Analyse von Ancient DNA-Daten
- Methoden der Populationsgenetik.

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

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

2 Research

2.4 Computational Statistics (CST)

Group leader

Prof. Dr. Tilmann Gneiting

Team

Prof. Dr. Sándor Baran (visiting scientist; University of Debrecen; July)
Jun.-Prof. Dr. Johannes Bracher (visiting scientist; Karlsruhe Institute of Technology)
Jun.-Prof. Dr. Timo Dimitriadis (visiting scientist; Heidelberg University)
Dr. Alexander I. Jordan (staff scientist)
Kristof Kraus (student worker)
Dr. Sebastian Lerch (visiting scientist; Karlsruhe Institute of Technology)
Dr. Marc-Oliver Pohle
Dr. Johannes Resin (visiting scientist; Heidelberg University)
Prof. Dr. Melanie Schienle (senior researcher)
Evgeni Ulanov
Dr. Eva-Maria Walz
Daniel Wolffram
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 at HITS and Professor of Computational Statistics at the Karlsruhe Institute of Technology (KIT). The group's research focuses on the theory and practice of forecasting.

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

Weather forecasting and collaborative research with meteorologists continue to represent prime examples of our work.

Introduction

Forecasting precipitation in the tropics remains a great challenge despite the enormous potential to create socioeconomic benefits in sectors such as food and energy production. In this year's Annual Report, we illustrate how methods for prediction and evaluation that were developed in the CST group can create new avenues to more accurate and resource-efficient operational precipitation forecasts in the Global South.
In [Walz et al., 2024b], we studied the performance of a comprehensive suite of state-of-the-art methods used to predict the occurrence and accumulation of precipitation over northern tropical Africa. Two factors that make precipitation forecasting challenging are (1) the multiple physical processes by which precipitation occurs and (2) the lack of simple parametric tools for uncertainty quantification.

Tropical Africa exhibits an exceptionally high degree of convective precipitation, which occurs when warm air rises vertically in the atmosphere and cooler altitudes reduce the amount of moisture that can be maintained in gaseous form, thereby leading to clouds and rain. This type of precipitation tends to occur in showers, which are intense and more limited in region and time. In contrast, when moist air is forced upward diagonally by larger-scale winds and atmospheric dynamics over other air masses or rising landforms, the occurrence of precipitation is more stable. Numerical weather prediction (NWP) models seem to work better in the latter circumstances, whereas convective precipitation has proven particularly difficult to predict.

Ideally, a forecaster quotes their confidence for every possible outcome rather than providing a singular best guess (the latter of which is known as a deterministic forecast). This approach is particularly difficult for precipitation accumulation. Often, there is no precipitation, but if precipitation does occur, it is a continuous amount. In other words, precipitation accumulation follows a mixture distribution with a point mass at zero (the probability of no precipitation) and a continuous part in the positive real numbers. In contrast, when modeling predictive uncertainty for temperature or pressure, it is often suitable to simply use a Gaussian distribution. We address this issue in precipitation forecasting by using Easy Uncertainty Quantification (EasyUQ; [Walz et al. 2024a]), which produces non-parametric predictive distributions from real-valued point predictions.
The horizon for the precipitation forecasts in [Walz et al., 2024b] was 24 hours ahead for a study area in Africa that comprised 19 x 61 grid boxes spanning from 0° to 18 °N in latitude and from 25 °W to 35 °E in longitude. Only grid boxes over land were considered in the evaluation that was performed for the years 2011–2019. For this report's brief overview, we also restrict ourselves to the major rainy season in the Sahel region from July to September of each year and to the following subset of prediction methods:

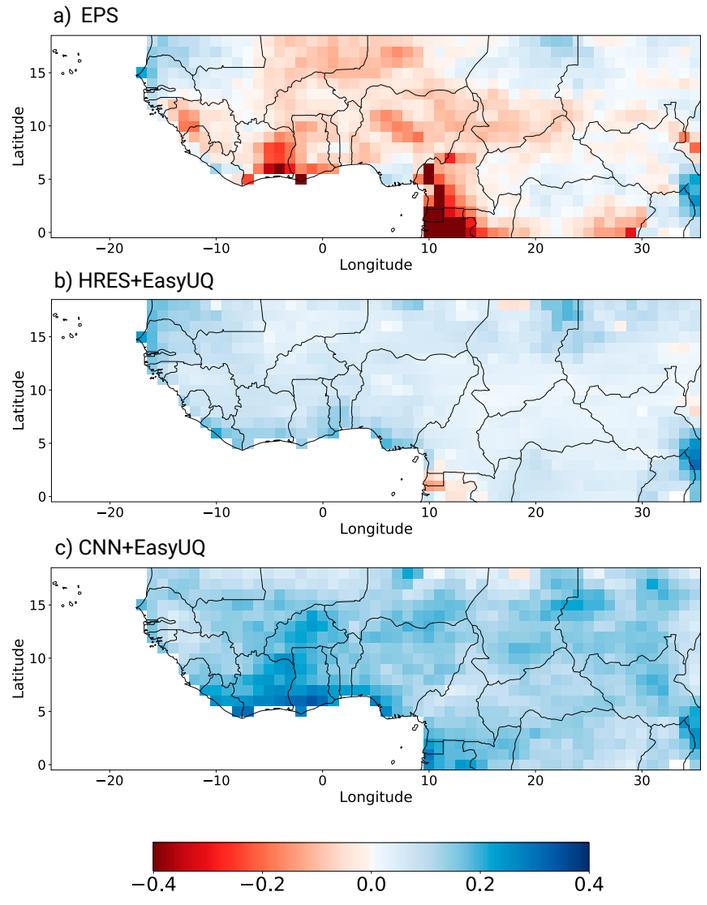


Figure 15: Spatial structure of the CRPS skill score (larger is better) relative to the MPC baseline for probabilistic forecasts of precipitation accumulation from the (a) EPS, (b) HRES+EasyUQ, and (c) CNN+EasyUQ prediction methods. Source: adapted from [Walz et al., 2024b; Figure 10].

1. EPS: The 51-member ensemble prediction system by the European Center of Medium-Range Weather Forecasts (ECMWF) is the result of the meteorologists' classical approach to uncertainty quantification. NWP models are run multiple times with different initial conditions and different parameterizations in order to obtain a collection of point predictions.
2. HRES+EasyUQ: Relying on point forecasts as input, the EasyUQ method by [Walz et al, 2024a] can be applied to the leading NWP point forecasts of the ECMWF HRES model. This HRES model runs at a particularly high resolution and is initialized with the most accurate analysis product for the atmospheric state.
3. CNN+EasyUQ: In a purely data-driven approach, EasyUQ can be applied to the point forecasts of a convolutional neural network (CNN). This network operates directly on the two-dimensional pseudo images of size 19 x 61 in order to learn spatial relations from the data without the need to extract spatial information beforehand.
4. MPC: As a baseline reference prediction, we take monthly probabilistic climatology. A prediction date is associated with its corresponding month, and the resulting prediction is simply the collection of observations from the same month in the previous years going back to 2001.

Figure 15 (previous page) shows the spatial structure of the predictive performance for these models over the chosen study area. Red areas indicate that a forecast model performed worse than the MPC baseline reference as measured by a so-called skill score, which is normalized to 0 for a predictive model performance that is equal to that of the baseline model with a maximum value of 1. Over the relatively dry areas along the Guinea Coast as well as over Gabon and southern Cameroon, the EPS shows negative skill relative to MPC, with little to be gained over the remaining study area from using the NWP ensemble approach. Instead, using EasyUQ on top of the single best NWP run improves the predictive performance considerably and lifts the performance of the HRES+EasyUQ above the MPC baseline at almost all grid boxes. Moving to the purely data-driven CNN+EasyUQ, we see another considerable jump in performance, which reaches up to 40% improvement relative to the climatological benchmark. At this point, we emphasize the forecast horizon of 24 hours because data-driven approaches work particularly well for short horizons. Nevertheless, NWP models exhibit issues when presented with conditions that are prevalent in the study region during the peak monsoon season.

Easy Uncertainty Quantification

In [Walz et al., 2024a], we introduced EasyUQ as an easy-to-implement method that transforms real-valued deterministic model output into calibrated statistical distributions. EasyUQ is trained on pairs of deterministic forecasts and corresponding outcomes and is thus independent of the type of model used to generate the single-valued forecasts. In particular, EasyUQ can be applied to the output of any NWP, statistical, or machine learning model that generates deterministic forecasts. The EasyUQ forecast distributions are discrete and have mass exclusively at the observation values in the training set. In the context of precipitation accumulation, rainfall amounts are typically reported in small but fixed increments, which means that the EasyUQ forecast distributions adapt naturally to the level of discretization in the observation values without any need for tuning. In fact, this automatic adaptability allows the quantification of uncertainty in very general settings.

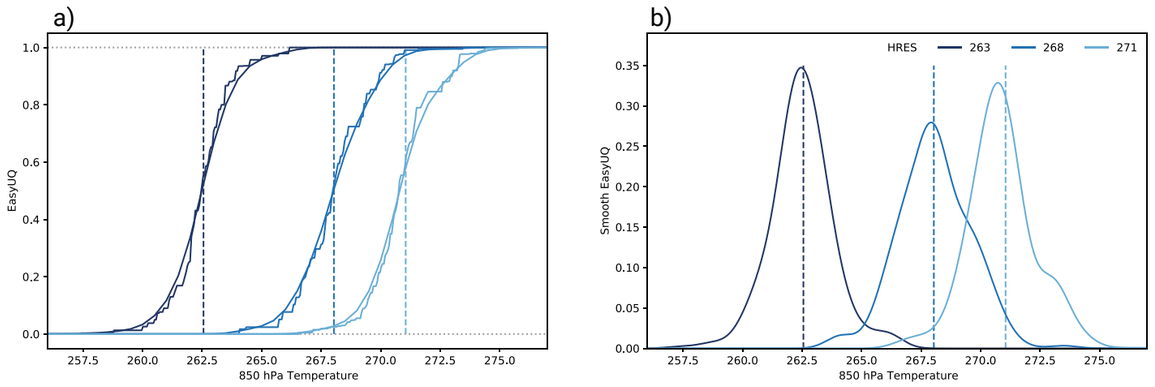


Figure 16: EasyUQ illustrated on three-days-ahead single-valued HRES model forecasts of upper air temperature. (a) Basic and smooth EasyUQ predictive CDFs and (b) smooth EasyUQ predictive densities at selected values of the HRES forecast. Source: adapted from [Walz et al., 2024a; Figure 1].

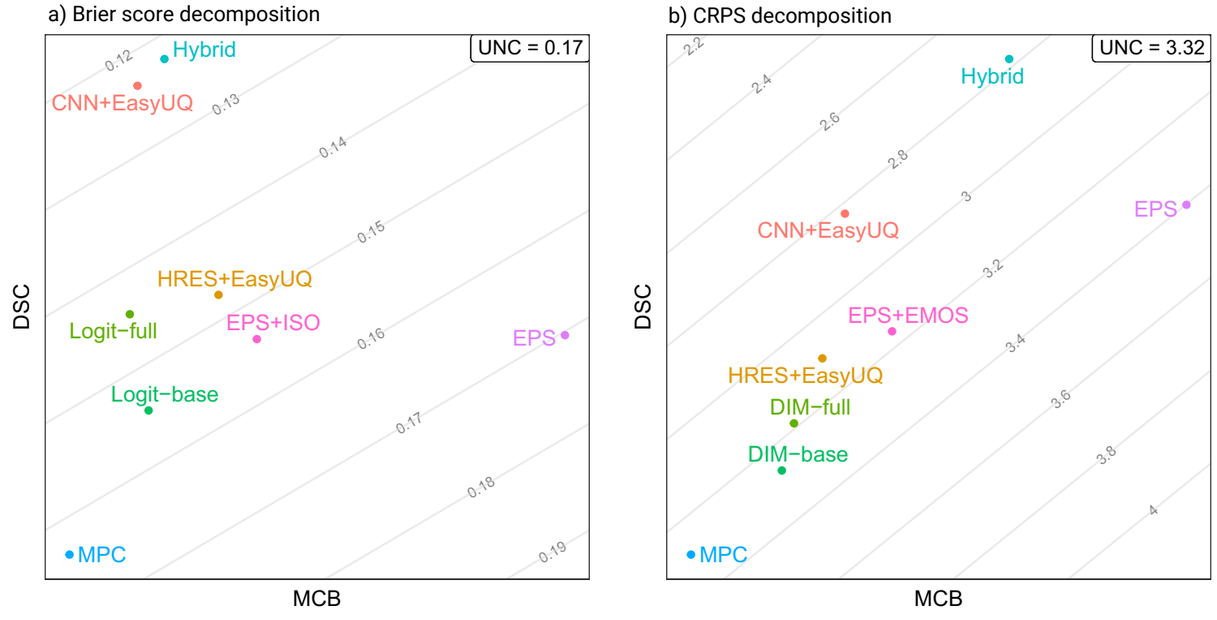


Figure 17: MCB (smaller is better), DSC (larger is better), and UNC component of (a) the mean Brier score for probability forecasts of precipitation occurrence and (b) the mean CRPS for probabilistic forecasts of precipitation accumulation in millimeters. Parallel lines correspond to equal mean scores. Source: [Walz et al., 2024b; Figure 14].

In its basic form, EasyUQ is a special case of isotonic distributional regression (IDR) – a technique we covered in the 2021 Annual Report. The fundamental assumption is one of isotonicity – that is, whenever the deterministic forecast that serves as input for EasyUQ increases, so should the predictive distribution that EasyUQ returns. Specifically, the predictive distribution should increase in stochastic order, or in plain words, the probability that the outcome will exceed any threshold should increase with the increasing value of the deterministic forecast. The left panel of Figure 16 illustrates this increase in a data example of three-days-ahead temperature forecasts: As the vertical dashed line corresponding to a single-valued HRES forecast increases, the EasyUQ predictive cumulative distribution function (CDF) moves to the right without ever intersecting. Note that this move is not simply a translation.

Figure 16 also shows a smooth version of EasyUQ. For a genuinely continuous outcome, it is preferable to use a continuous predictive distribution. To achieve this, Smooth EasyUQ applies kernel smoothing while preserving the stochastic ordering of the basic EasyUQ predictions. One important distinction between the two variants is the requirement to tune parameters with smooth EasyUQ, whereas the basic variant can be employed in a fully automated workflow.

Evaluating overall score, miscalibration, and discrimination

The skill scores in Figure 15 – which are used to compare the improvement of the prediction methods over a baseline – are based on the average continuous ranked probability score (CRPS). Belonging to the class of proper scoring rules, the CRPS assigns a score to a forecast–observation pair, while the minimum expectation is only achieved by correctly specifying the distribution of the outcome. This property is essential when it comes to comparing and ranking models' predictive ability, but the reduction to a single value obscures other aspects of forecast performance. In [Arnold et al., 2024], we developed a decomposition of the average CRPS into three interpretable components: (1) a term that measures miscalibration (MCB) or reliability, (2) a term that measures discrimination ability (DSC), and (3) a term that quantifies the overall uncertainty (UNC) of the outcome. Calibration refers to the statistical consistency between a forecast and the observations, such as whether there is a systematic bias in location or spread, whereas discrimination refers to the ability to identify forecast regimes, such as whether it is monsoon season, which would lead to more precipitation.

Historically, the first score decomposition was introduced by Allan H. Murphy in 1973 for the mean Brier score (BS). Also a proper scoring rule, the Brier score is closely related to loss functions for point predictions in that it applies to probability forecasts of binary events rather than to the more involved predictive distributions of real-valued outcomes. However, the underlying idea remains in which a recalibrated version of a forecast is used to determine the excess score due to miscalibration that could have been avoided. Simultaneously, the score difference between the recalibrated forecast and a climatological forecast represents the value of the information that was incorporated into the prediction model over an uninformative baseline. In the 2021 Annual Report, we highlighted work in which we argued that the Brier score decomposition should rely on an isotonically recalibrated forecast, which is straightforward to achieve for probability forecasts of binary events. In [Arnold et al., 2024], we leveraged the IDR technique to the same effect when dealing with probabilistic forecasts for real-valued outcomes that are evaluated with the CRPS.

These decompositions facilitate plots that can be particularly useful when evaluating the trade-offs between competing models. Returning to the data example of precipitation forecasts over northern tropical Africa, Figure 17 (previous page) reveals that methods of similar rank can perform differently when it comes to miscalibration and discrimination. The left panel shows the performance measured by the Brier score for probability forecasts of precipitation occurrence, whereas the right panel shows the performance in terms of the CRPS for probabilistic forecasts of precipitation accumulation. In addition to the performance of the EPS, HRES+EasyUQ, CNN+EasyUQ, and MPC methods, the figure contains the results of a few more models that illustrate the usefulness of the MCB-DSC plots when comparing a larger selection of models. For details on these models, we refer to [Walz et al., 2024b].

In MCB-DSC plots, as in Figure 17, the better models tend toward the top-left corner, which corresponds to a low level of miscalibration and a high level of discrimination ability. In addition, parallel lines show levels of equal mean scores. We can draw similar conclusions as from Figure 15: namely that the EPS is slightly worse than MPC overall, that the HRES+EasyUQ shows improvements over both EPC and MPC, and that the CNN+EasyUQ exhibits a further substantial increase in predictive performance. Notably, the EPS and MPC are in very different areas of the plot, with the EPS demonstrating a competitive level of discrimination together with the worst miscalibration and MPC showing no signs of miscalibration but also no discrimination ability.

In [Dimitriadis et al., 2024], we presented a further diagnostic tool for in-depth investigations of overall predictive performance, miscalibration, and discrimination when focusing on a small selection of methods for predicting binary events. Drawing on methodology that we had developed in the prior decade, we used Murphy curves to show a different decomposition of the Brier score that can be used to reconstruct any proper scoring rule for probability forecasts, and CORP reliability curves can be used to identify whether and when miscalibration is the result of under- or overprediction. Specifically, Murphy curves display mean scores for an entire class of linearly parameterized elementary proper scoring rules, where the classical misclassification rate shows up for a threshold value of ½ on the horizontal axis and the area under a Murphy curve equals the Brier score. Reliability curves show an estimate of the event probability as a function of the forecast value and should therefore remain close to the diagonal (the identity function) for a well-calibrated prediction method. Discrimination ability can be further inspected using concave ROC curves computed from isotonically recalibrated forecasts. An example of a so-called triptych (comprising Murphy, reliability, and ROC curves) in the context of solar flare predictions is given in Figure 18. For a more in-depth discussion of the individual diagrams, we refer to the Annual Reports from 2015 for Murphy curves, from 2021 for CORP reliability diagrams, and from 2022 for ROC curves.

The text in this report has been adapted and contains excerpts from the following publications: [Arnold et al., 2024], [Dimitriadis et al., 2024], [Walz et al., 2024a], and [Walz et al., 2024b].

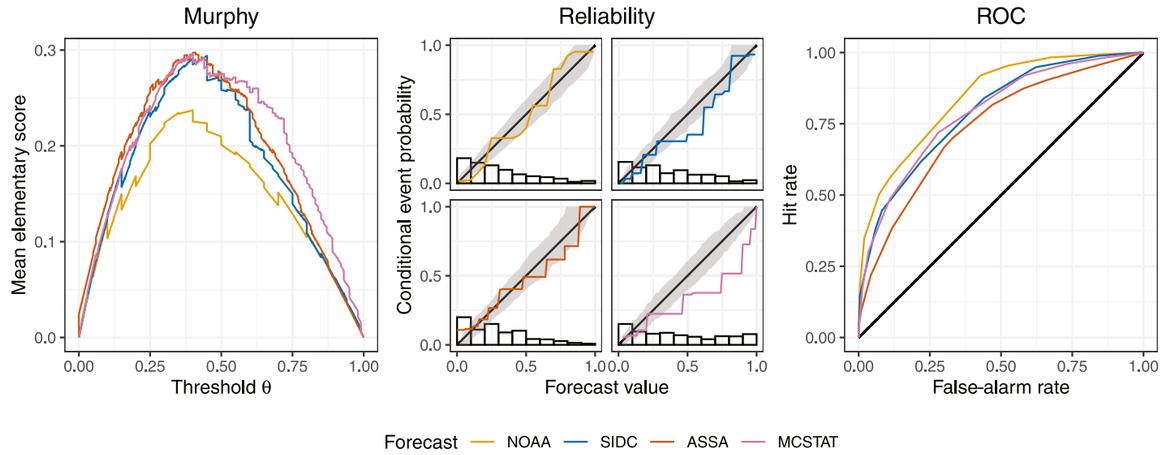


Figure 18: Triptych of diagnostic graphics used for evaluating and comparing the probability forecasts of class C1.0+ solar flares: Murphy curves (lower is better), reliability curves (close to the diagonal is preferred) with 90% consistency bands, and ROC curves (higher is better). Source: [Dimitriadis et al., 2024; Figure 1].

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.

2.5 Data Mining and Uncertainty Quantification (DMQ)

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)
Ayse Erozan (visiting scientist; Heidelberg University)
Isabel Gernand (PhD student)
Dr. Saskia Haupt (visiting scientist; Heidelberg University; until May 2024)
Alejandra Jayme (visiting scientist; Heidelberg University; until April 2024)
David Lehmann (visiting scientist; Heidelberg University)
Stefan Machmeier (visiting scientist; Heidelberg University)
Pascal Memmesheimer (visiting scientist; Heidelberg University; since January 2024)
Jakob Niessner (visiting scientist; Heidelberg University; since April 2024)
Rimpa Pal (PhD student; Heidelberg University; since October 2024)
Jonas Roller (visiting scientist; Heidelberg University)
Valentin Schmid (PhD student)
Elaine Zaunseder (visiting scientist; Heidelberg University)
Alexander Zeilmann (visiting scientist; Heidelberg University)

The Data Mining and Uncertainty Quantification (DMQ) group, headed by Vincent Heuveline, began its research in May 2013. The group works in close collaboration with Heidelberg University's Engineering Mathematics and Computing Lab (EMCL) at the Interdisciplinary Center for Scientific Computing (IWR), which is also headed by Vincent Heuveline. The DMQ group's research focus lies in gaining knowledge from extremely large and complex datasets, with a particular concentration on computational fluid dynamics (CFD) and biomedical research. Both fields – data mining and uncertainty quantification – require a decidedly interdisciplinary approach to mathematical modeling, numerical simulation, hardware-aware computing, high-performance computing, and scientific visualization. In 2024, the group's research mainly involved developing robust and efficient machine learning methods for uncertainty quantification as well as for the analysis of the acquired data.

Improving tumor diagnostics with mathematical modeling

In a world in which cancer is one of the most common causes of death, diagnosing and treating cancer diseases traditionally plays a major role in medical research. One of the most prevalent types of cancer is colorectal cancer in which cancer cells begin to develop in the colon or the rectum. Colorectal cancer is often caused by mutations in oncogenes or in tumor suppressor genes as well as by deficiencies in DNA repair mechanisms and can be either sporadic or hereditary. The combination of mathematical modeling techniques, data analysis and medical science can be especially powerful for gaining valuable insights into biological processes and improving existing diagnostic tools for colorectal tumors.

As part of the interdisciplinary project "Mathematics in Oncology" (MathOnco) in cooperation with partners at Applied Tumor Biology at the University Hospital Heidelberg, we aim to improve the diagnosis of microsatellite instability, which is both a genetic predisposition to mutations caused by deficient DNA mismatch repair and a frequent cause of colorectal cancers. The medical diagnosis of the condition is usually performed by analyzing the distribution of lengths of DNA fragments that have been extracted from tissue and that have been subsequently amplified via polymerase chain reaction (PCR, see Figure 19). The considered DNA markers are so-called "microsatellites", that is short repetitive sequences that are prone to mutations specifically in tumors with DNA mismatch repair deficiency.

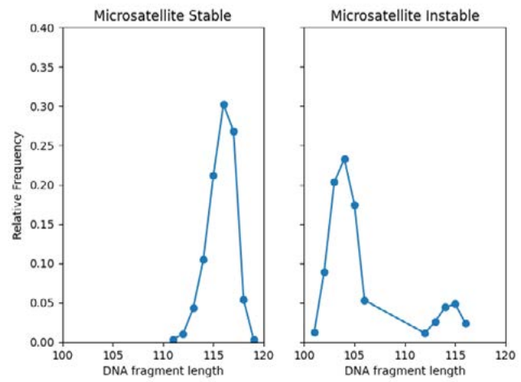


Figure 19: PCR data depicting the distribution of lengths of microsatellites. The left curve shows a microsatellite-stable (MSS) sample, while the right curve corresponds to a microsatellite-instable (MSI) sample.

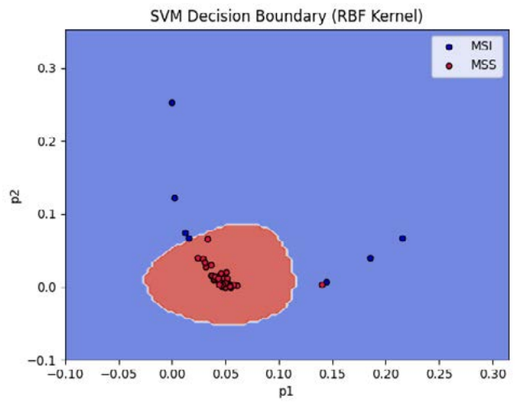


Figure 20: Classification in a two-dimensional parameter space obtained via parameter estimation with a probabilistic model and a heuristic optimization scheme. Red dots correspond to MSS samples, blue dots to MSI samples.

In our work, we aim to improve this framework by including a probabilistic model of the propagated mutations that occur in the form of changes to the lengths of DNA fragments that have been amplified over multiple cell cycles. The mathematical model is applied to PCR data from approximately 80 patients and identifies variables of interest such as mutation probabilities or unknown DNA marker lengths. The entire calibration process is an optimization problem with a possibly non-smooth objective function and is predominantly solved with heuristic optimization algorithms. The resulting model enables valuable insights to be made into the quality of different DNA markers and into the biological processes that occur during the PCR. Furthermore, gained parameters are used to perform a classification task in the resulting low-dimensional parameter space: Classifiers based on support vector machines and tree-based methods enable us to determine whether a patient shows signs of microsatellite instability (see Figure 20). This determination is especially relevant when it comes to understanding the differences in available patient data in order to develop effective diagnostic tools.

Mathematical modeling of tumor invasion

Tumor development is a complex process characterized by uncontrolled cell growth, which can lead to the invasion of surrounding tissues and can disrupt normal tissue architecture. Traditional experimental approaches often provide limited information on the spatial dynamics of tumor growth, thereby highlighting the need for mathematical modeling to simulate and analyze tumor behavior. In this context, partial differential equations (PDEs) offer a powerful tool for capturing the spatio-temporal evolution of tumors. A PDE is a type of continuous model that describes how a quantity – such as tumor cell density – changes over space and time. The equations provide a framework for modeling complex biological processes, including cell proliferation, migration, and

interaction with the extracellular matrix. Accurately modeling tumor growth requires addressing the dynamic changes in the tumor boundary as it interacts with surrounding tissues. This is inherently a moving interface problem in which both the shape and size of the tumor evolve over time. Mathematically, the problem involves solving PDEs in a domain whose boundary is an unknown variable that must be determined alongside the solution. As an example, the images (Figure 21 and 22) below depict a simulation of tumor growth within colonic crypts, which are tubular invaginations in the colonic epithelium. The simulation illustrates two crypts that showcase how a tumor can develop and interact with these structures.

In order to solve these complex problems on given geometries, we use the finite element method (FEM), which is a numerical technique that discretizes the spatial domain and approximates the solution using a finite dimensional function space. This approach enables the simulation of tumor growth in complex geometries while providing a reasonable approximation of the underlying biological processes.

As part of our ongoing research collaboration with clinicians and biologists from the Applied Tumor Biology (ATB) department at University Hospital Heidelberg, we aim to develop more accurate and predictive models of tumor growth. Our goal is to provide clinicians with a valuable tool for both better understanding tumor growth dynamics and making predictions about disease progression.

A novel machine learning framework for personalized medicine: Conformal prediction and optimal therapy design

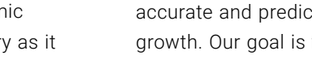
In today's evolving healthcare landscape, patient-centric care is of the utmost importance. Traditional models often prioritize efficiency over individual needs, but as the rate of chronic diseases continues to increase and technology advances, a shift toward personalized, patient-driven care is essential.

> "If you want to know how to improve patient care, ask the patients."
>
> – Regina Holliday

Focusing on shared decision making, improved communication, and holistic well-being enhances outcomes. Moreover, patient-centric healthcare reduces costs and fosters trust in addition to addressing disparities, thereby ensuring equitable and culturally sensitive care. In an increasingly complex system, prioritizing patients is vital when it comes to a more effective and sustainable healthcare future.

Our research introduces an innovative approach to personalized medicine by developing a machine learning framework that quantifies uncertainty in individual treatment effect estimation. Our model recognizes that some patients can tolerate certain side effects while others cannot, and it tailors treatment recommendations accordingly.

The system uses patient-reported outcome (PRO) data collected over time to identify potential side effects and when they might appear, and it then adapts therapy variants in order to minimize the probability of treatment discontinuation. This process ensures that patients receive therapies that are aligned with their specific tolerances and needs, thereby enhancing adherence and overall well-being. Our framework integrates conformal prediction methods with optimal experimental design, thereby providing reliable, interpretable, and actionable insights for medical decision making (see Figure 23).



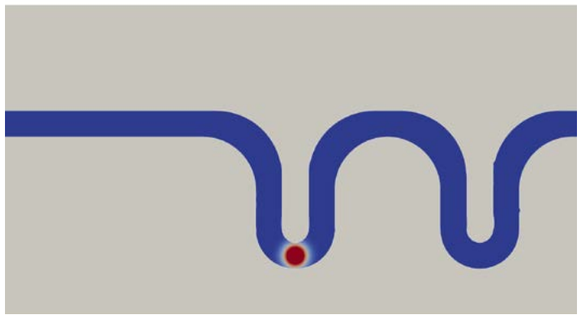
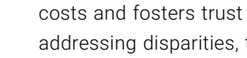


Figure 21: Example of a tumor growth simulation in a geometry in which the epithelium contains two crypts. An initial higher tumor cell density can be seen in red.



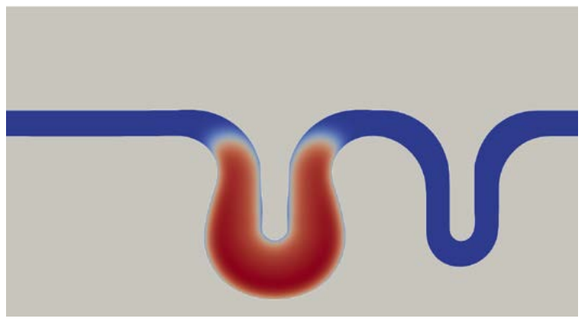


Figure 22: In this image, the simulation shows the spread of the tumor at a later point in time. The structure of the crypts has changed as a result of the tumor growth.

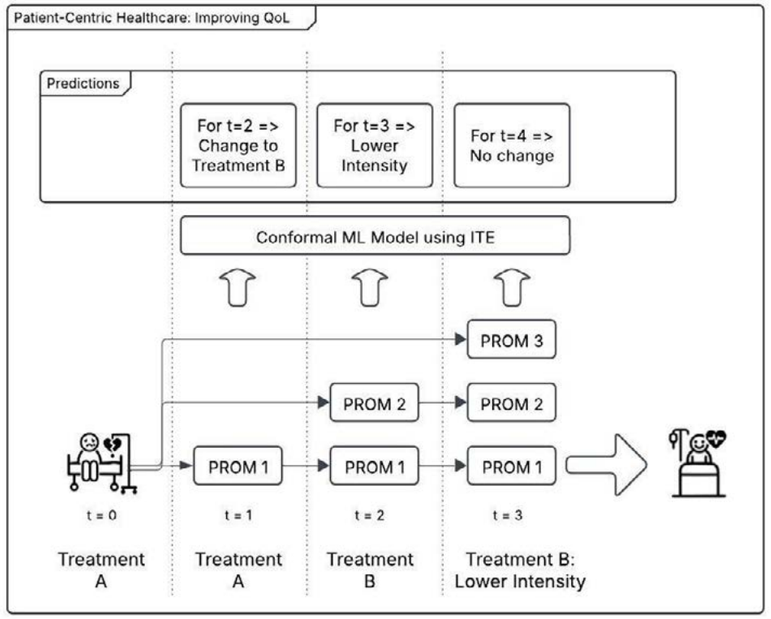


Figure 23: Patient-centric healthcare guided by a conformal ML model using ITE and PROMs to personalize treatment decisions over time and improve quality of life.

KI-Morph: Enabling AI-powered image analysis for researchers

Artificial intelligence (AI) is revolutionizing many fields – including medical research, biology, and engineering – by helping scientists analyze large sets of images quickly and efficiently. However, using AI for complex image analysis often requires expertise in both AI and high-performance computing (HPC), which can be a major challenge for researchers who are not computer specialists.

In order to bridge this gap, researchers from the DMQ group at HITS developed KI-Morph – a user-friendly platform that allows scientists to harness the power of AI for large-scale image analysis without the need for deep technical knowledge. KI-Morph runs on bwHPC, which is a high-performance computing system in Heidelberg, Germany, thereby making it possible to process huge 3D image datasets efficiently.

This efficiency is particularly important for fields such as medicine and biology, in which researchers rely on 3D imaging techniques such as MRI scans and microscopy images to study complex structures. These images are often hundreds of gigabytes in size, thereby making manual analysis nearly impossible.

KI-Morph overcomes these challenges by offering an easy-to-use, web-based platform that integrates seamlessly with HPC systems. Researchers can upload their image data to a secure storage system and use AI-powered tools to identify patterns, classify structures, and create detailed 3D visualizations (see Figure 24). The platform also provides intuitive annotation tools that allow users to refine their results interactively. Complex AI tasks – such as the training of deep-learning models – are automated; therefore, researchers do not need coding skills.

Because KI-Morph is web-based, it works on any device with a browser, thereby eliminating the need for complex software installations. The platform additionally ensures data privacy by storing all information securely within academic research facilities. By making AI-powered image analysis more accessible, KI-Morph enables researchers to focus on their discoveries rather than on technical challenges. As AI

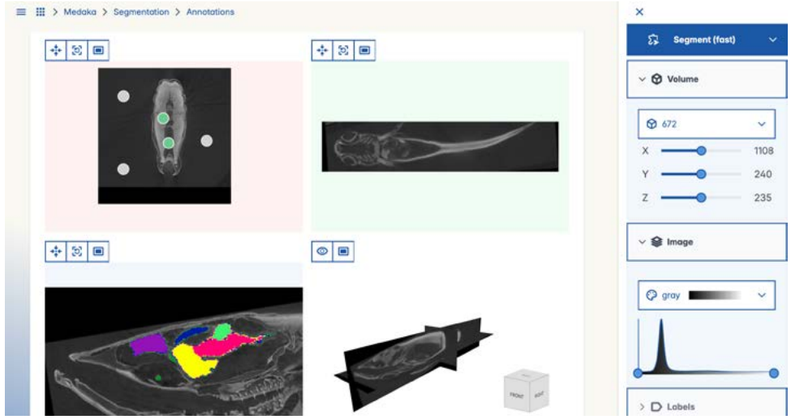


Figure 24: The current version of the user interface can visualize image data using three 2D slices along the major axis directions as well as using a 3D image viewer. The data can be annotated using point annotations (see top left panel). After sending the annotations to the high-performance cluster to perform the data analysis, the results – for example, segmentations (see lower left panel) – are visualized directly alongside the data.

continues to evolve, platforms such as KI-Morph will play a crucial role in advancing scientific research, thereby making high-powered computing tools available to experts in medicine, biology, and beyond.

Numerical study of enhanced radial heat transfer by dielectrophoretic force

The study of fluids that are contained between differentially heated walls has a great number of applications in engineering, for example, in solar heaters. The transfer of heat in such devices can be enhanced by convective processes. The improvement of heat convection by the dielectrophoretic (DEP) force – which arises when a dielectric fluid is subjected to an inhomogeneous electric field – has been established as an efficient and sustainable approach in several basic research works. The dynamic process under consideration is then referred to as thermal electrohydrodynamical (TEHD) convection. Our recently published paper on this subject matter is based on cooperation with Christoph Egbers' group at BTU Cottbus, with funding provided by DFG. In the paper, we investigate the impact of high voltage on the dynamics of a particular type of silicone oil contained between two concentric, horizontally aligned cylinders. The inner cylinder is heated while the outer cylinder is cooled and the voltage is applied to the inner cylinder, with the outer cylinder being connected to the ground. The DMQ group's contribution to this research consists of numerical simulations of the TEHD model equations, parameter estimation via uncertainty quantification (UQ) techniques, and post-processing in order to obtain several quantities of interest. The task of our partners in Cottbus is to design the physical apparatus as well as to execute and postprocess the real-world experiments.

Our numerical approach to the TEHD equations – which are a set of partial differential equations (PDE) – is to use the finite element method (FEM). This type of discretization has been shown to converge to the exact solution of the TEHD equations by way of a priori error estimates. In earlier works, however, a considerable quantitative discrepancy was found between the numerical solutions and the corresponding experimental data. We therefore introduced a physically reasonable non-constant thermal boundary condition on the outer cylinder wall (see Figure 25). This boundary condition introduces an additional parameter into the system that has a quantitative influence over the model solution. As we did not have measurements or theoretical estimates for this parameter, we instead estimated it via inverse UQ methods based on surrogate modeling using generalized polynomial chaos (gPC) expansion. For a wide range of experimental configurations, the calibrated model exhibits a stronger quantitative agreement with the experiment than does the original version. Results computed by FEM approximation of the calibrated model lead to multiple interesting conclusions. For a low-to-intermediate voltage, the initially steady flow does not destabilize; rather, it exhibits an increase in both fluid velocity and heat transfer. Through a series of simulations, we were able to show that this flow is dominated by the azimuthal and radial components and is thus two-dimensional. At sufficiently high voltages, the flow destabilizes with a stronger enhancement of heat transfer. By comparing two-dimensional simulations with three-dimensional simulations, we were additionally able to numerically demonstrate that the flow becomes fully three-dimensional after the instability is triggered (see Figure 26). Ongoing work will apply the same experimental and numerical setup to a vertical configuration with a rotating inner cylinder.

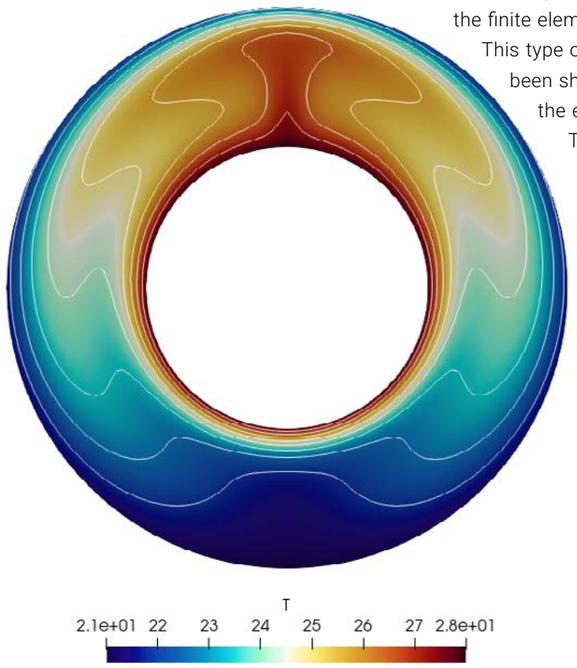


Figure 25: Snapshot of the numerical temperature solution in the mid-length position of the cylindrical annulus, with the inner cylinder being ΔT = 7K warmer than outer cylinder. A voltage of 7kV is applied.

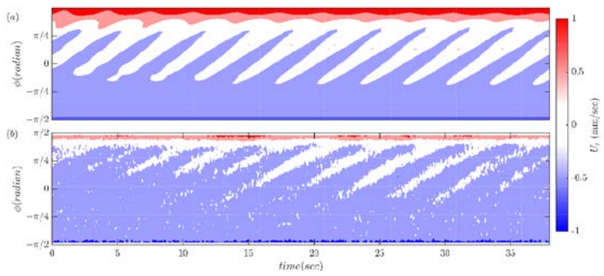


Figure 26: Space–time plot (Hovmöller diagram) for the spatially averaged radial velocity for ΔT = 9K and a peak voltage of 17kV. The x-axis represents time, while the y-axis denotes the azimuth. The azimuth is equal to π/2 at the top of the apparatus and to −π/2 at the bottom. There is clearly a strong agreement between the experimental and numerical results.

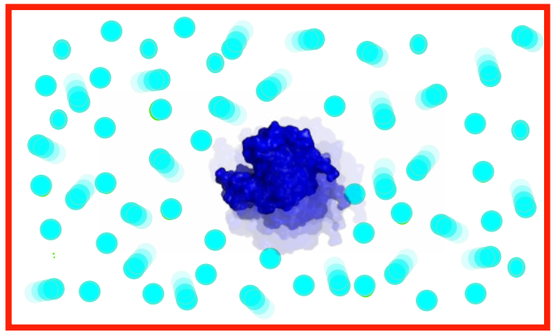


Figure 27: Illustration of protein surrounded by thermally moving water molecules

Protein electrostatics in ionic solutions

As we all know from experience, when we rub a balloon against our hair, the hair sticks to the balloon and does not want to come off when we try to remove it. This phenomenon is generally known as "electrostatics" and is one of the four fundamental forces in the Universe. Consequently, it comes as no surprise that this force plays a crucial role in the interaction between bio-essential molecules, such as proteins and enzymes.

In collaboration with the Molecular and Cellular Modeling (MCM) group, we have been working on a scheme to compute electrostatic forces both more accurately and faster during simulations of such molecules. The simulation scheme that the project targets is called "Brownian dynamics." This name comes from a physical phenomenon known as "Brownian motion," which was discovered by Scottish physician and botanist Robert Brown. While observing pollen of the species Clarkia pulchella in a drop of water, Brown found that very small particles (i.e., those the size of 6–8 micrometers) underwent sudden and irregular movement. It was not until eighty years later that Albert Einstein and Marian Smoluchowski managed to explain and mathematically quantify this observation with the aid of the thermal motion of the surrounding water molecules (see Figure 27). A heavy Brownian particle is constantly exposed to collisions with quickly moving water molecules due to thermal motion. The particle then performs a random walk both rotationally and translationally. Mathematically, this process can be described by the following two stochastic differential equations:

$$dw_i = \frac{D_{ij}}{k_b T} T_j dt + W(t)$$

For translational motion and

$$dr_i = \frac{D_{ij}}{k_b T} F_j dt + R(t)$$

for rotational motion.

In these equations, dr and dw are infinitesimally small steps, D is the diffusion tensor, T and F are the torque and the force, respectively, and R and W are random contributions that follow a Gaussian distribution. As mentioned above, one of the most important contributions to the force is the electrostatic force that acts between the molecules. This force can be quantified via electrostatic potential. In the case of distilled water, this equation is very simple and easy to solve. However, as soon as minerals and salts are introduced to the surrounding medium, the problem becomes significantly more challenging. When salts and minerals are dissolved in water, they loosen their bonds, and their atoms become freely mobile ions (electrically charged atoms). These ions then try to adjust to the electric fields. The electric field in this situation is described by the Poisson–Boltzmann equation, which is an elliptic non-linear partial differential equation. Hence, accurate simulation requires accurately and quickly solving this equation (see Figure 28).

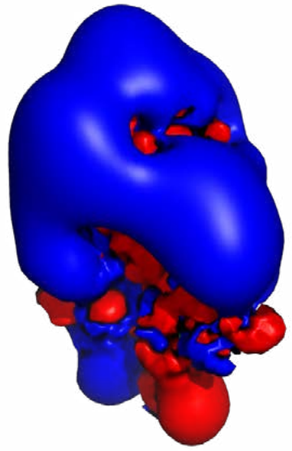


Figure 28: Level sets at +/− 1 k_b T/e of the electrostatic potential according to the Poisson–Boltzmann equation for bovine pancreatic trypsin.

Die Gruppe **Data Mining and Uncertainty Quantification** (DMQ) unter der Leitung von Vincent Heuveline begann ihre Forschung im Mai 2013. Die Gruppe arbeitet eng mit dem Engineering Mathematics and Computing Lab (EMCL) am Interdisziplinären Zentrum für Wissenschaftliches Rechnen (IWR) an der Universität Heidelberg zusammen, das ebenfalls von Vincent Heuveline geleitet wird. Im Fokus der Forschungsarbeit steht ein zuverlässiger und strukturierter Erkenntnisgewinn aus großen und komplexen Datensätzen mit Schwerpunkt auf Computational Fluid Dynamics (CFD) und biomedizinischer Forschung. Beide Bereiche – Data Mining und Uncertainty Quantification – erfordern einen interdisziplinären Ansatz für mathematische Modellierung, numerische Simulation, hardwarenahe Programmierung, Hochleistungsrechnen und wissenschaftliche Visualisierung. Im Jahr 2024 hat die DMQ Gruppe besonders die Entwicklung robuster und effizienter maschineller Lernmethoden für UQ sowie die Analyse der erfassten Daten weiterentwickelt.

2.6 Machine Learning and Artificial Intelligence (MLI)

Group leader

Jun.-Prof. Dr. Jan Stühmer

Team

Benjamin Barth (student; until March 2024)
Torben Berndt (PhD student; since October 2024)
Mila Coetzee (PhD student)
Beni Egressy (postdoc; since November 2024)
Elyes Farjallah (student internship; since November 2024)
Moritz Feik (student assistant; until August 2024)
Eddie Merbitz (student internship; since December 2024)
Steven Schürstedt (student; until March 2024)
Leif Seute (PhD student)
Hendrik Stumpf (student assistant; since November 2024)
Simon Wagner (student; until October 2024)
Nicolas Wolf (guest scientist; since November 2024)

The Machine Learning and Artificial Intelligence (MLI) group works on two main research topics: geometric deep learning – which is a theoretical framework used to analyze and develop deep learning architectures from first principles – 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. Originally developed to study the capabilities of graph neural networks, geometric deep learning has evolved into a versatile framework for analyzing a wide range of machine learning architectures. 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. Building on this success, the group now explores context-specific protein generation, in which a protein is designed to bind to a particular molecule (ligand) or to include a desired binding site. This approach holds great promise for personalized cancer therapies and next-generation drug design.
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.

The MLI group's research on interpretable deep learning also contributes to the emerging field of neuro-symbolic methods, which bridge the gap between the "sub-symbolic" features used in neural networks and "symbolic" reasoning systems. By translating neural network outputs into symbolic representations, these methods open the door to advanced reasoning tasks, thereby combining the strengths of both AI paradigms.

Generating novel protein structures with geometric algebra flow matching

Nature uses proteins as versatile molecular tools – for example, as signal substances, as tissue components, or to facilitate chemical reactions. With recent advances in deep learning, it is now possible to design proteins with specific functional properties, thereby leading to a revolution in protein science, as was recognized by the 2024 Nobel Prize in Chemistry. Proteins with certain chemical reactivities, binding propensities, or durability are of particular interest in biotechnological applications, in the development of nanomaterials, and even in sustainability problems, such as plastic degradation.

These applications rely on two key developments: advances in generative modeling that enable complex probability distributions to be learned, and the construction of powerful equivariant architectures that are capable of describing the intricate geometry of protein structures. In 2024, we worked on the latter topic and proposed a new model architecture for protein structure based on projective geometric algebra. We employed the architecture within a flow-matching framework and developed a state-of-the-art protein structure design model [Wagner et al., 2024].

Flow matching
The aim of generative modeling is to learn to sample from probability distributions – for example, the distribution of realistic images or language. For distributions on continuous data manifolds, such as images, diffusion models achieved a breakthrough in the early 2020s (Ho et al., Denoising Diffusion Probabilistic Models, NeurIPS, 2020). Recently, flow matching has been proposed as a generalization of diffusion models (Lipman et al., Flow Matching for Generative Modeling, ICLR, 2023). In flow matching, a transformation on the data domain – that is, a flow – is learned such that samples from a simple distribution (i.e., noise) are mapped to samples of the target probability distribution. The key idea is to express this transformation as the integral of a vector field that is parametrized by a neural network (see Figure 29). It is possible to show that regressing on tangent vectors along curves that connect noise samples to data samples during training indeed approximates a flow that has the desired property of mapping the noise distribution to the target distribution.

The main advantage of flow matching over diffusion is that the trajectories from noise to data can be controlled and are not constrained to diffusion paths. This ability enables optimal transport solutions to be learned, thereby reducing the number of timesteps – or numerical integration steps in the case of flow matching – that are required in order to accurately generate samples.

Generating protein structures
One of the reasons that proteins are interesting tools for both nature and technological applications is that they are encoded in a simple one-dimensional representation – that is, the sequence of amino acids. Through both entropic and enthalpic effects, this chain of amino acid residues folds into a specific structure. For decades, the problem of predicting this folded structure from the protein's sequence remained unsolved – that is, until recently, when the geometric deep learning model AlphaFold achieved unprecedented accuracy at this task (Jumper et al., Highly accurate protein structure prediction with AlphaFold, Nature 596, 583–589, 2021). In contrast to structure prediction, protein design aims at the generation of

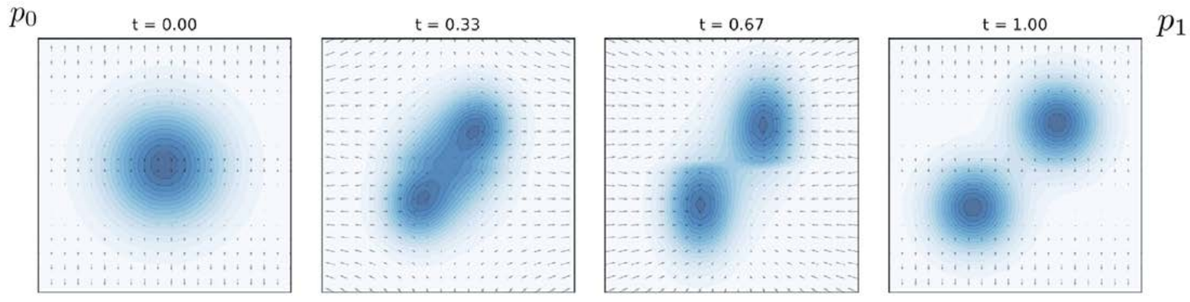


Figure 29: In flow matching, a transformation mapping between two probability distributions (blue) is expressed via a learnable flow vector field (black).

new proteins with a given function. Since a protein's function is fundamentally related to its structure, developing methods for generating biophysically consistent, diverse, and novel structures without a given input sequence is of great interest.

Following AlphaFold, we represent a protein's backbone structure via a set of translations and rotations – that is, one for each residue (Figure 30). In so doing, we make use of the constant bond distances and angles within a backbone fragment, thereby making each residue a rigid body with three translational and three rotational degrees of freedom. This data representation introduces a geometric inductive bias for machine learning models, which usually leads to improved generalization capability and data efficiency.

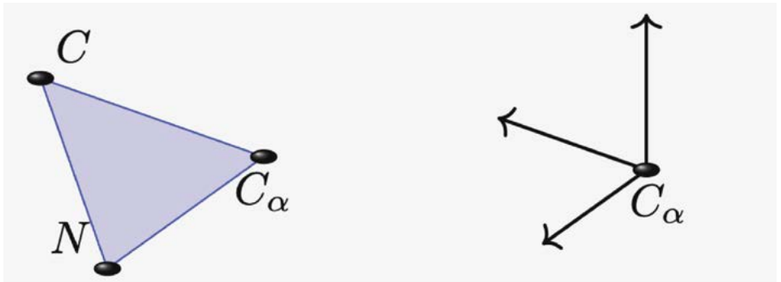


Figure 30: Due to constant angles and bonds, a backbone fragment can be described via a single rigid body rotation and translation, thereby defining a local coordinate frame.

However, the frame representation also creates a difficulty within the generative modeling framework: The group of rotations is a non-Euclidean manifold, and curvature effects thus have to be considered. Intuitively speaking, we cannot simply add and subtract noise as is common in diffusion models because addition is not defined per se on a Riemannian manifold. In order to understand this issue intuitively, we can imagine a sphere embedded in three-dimensional space: If two points are added by simply adding their Euclidean coordinates, the resulting point will not necessarily lie on the sphere! While diffusion can also be formulated in curved spaces with some theoretical effort, flow matching naturally extends to Riemannian manifolds. The learnable flow vector field is then defined on the tangent space and induces a flow on the manifold. Usually, the flow trajectories are chosen as lines of shortest distance – that is, geodesics – between noise and data samples.

Clifford frame attention

In order to learn the flow vector field, we require a machine learning model that maps from a point on the manifold to a vector in the tangent space at the respective point. As discussed above, in our case, the manifold is the group of rotations and translations – or equivalently, the set of coordinate frames that represent the backbone fragments. Due to a peculiarity of the group of rotations, it is possible to parametrize the vector field via a point on the manifold; thus, this point suffices for constructing a model that maps from one set of coordinate frames to another.

For this mapping, current state-of-the-art models rely on AlphaFold's structure module, in which geometric information is processed by the invariant point attention (IPA) architecture. Learnable features are point-valued – that is, they represent points that are defined in the respective local coordinate frame of the residue. During message passing, the points are transformed from one coordinate frame into the other, thereby both encoding the relative orientation of the two frames and ensuring equivariance under global rotations and translations of the protein. The latter item is an important geometric bias because the learned flow process should be independent of rotations and translations that are applied to both noise and data.

In our work [Wagner et al., 2024], we extended this point-based architecture by introducing features that not only are point-valued, but also can represent planes and lines. We hypothesized that these geometric instances should be important for expressing residue membership in secondary structures, such as helices and strands (Figure 31). In order to model these more complex geometries, we used projective geometric algebra (PGA), which is a Clifford algebra in which the elements represent scalars, points, lines, and planes. The bilinear operation of the algebra – that is, the geometric product – can be used to calculate structurally relevant quantities, such as angles between lines and planes or the line of intersection of two planes.

In particular, we introduced both PGA-valued features as a learnable bilinear layer that relies on the geometric product between the PGA features of two nodes and higher-order messages that represent not only pairwise relationships, but also properties of three-tuples of residues. We termed the resulting architecture "Clifford frame attention" (CFA).

Geometric algebra flow matching for protein structure

We employed the new CFA architecture within a flow matching framework for protein structure generation and called the resulting structure design model "geometric algebra flow matching" (GAFL). We trained the model on 25,000 monomeric structures from the protein data bank (PDB) for 15 days on two

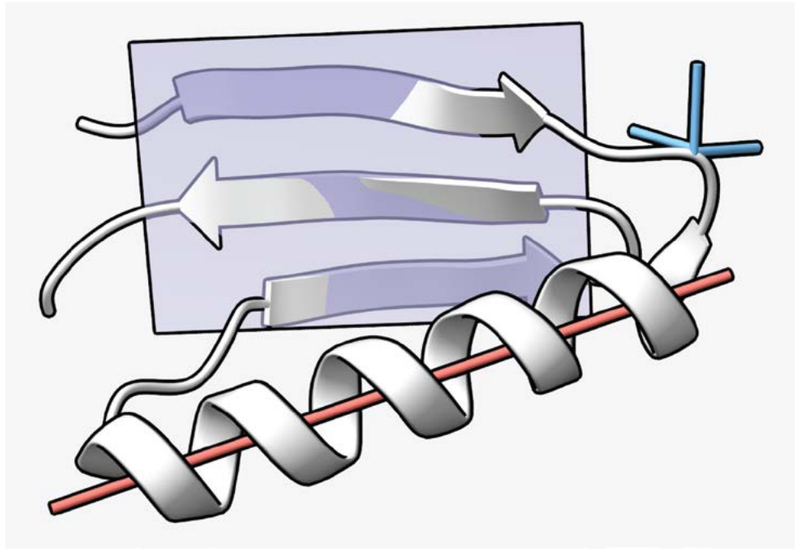


Figure 31: Lines, planes, and frames can be represented by PGA elements, thereby allowing the model to express geometric relationships between residues that are relevant for secondary structures, such as helices and strands.

GPUs. Figure 32 displays an example of a flow-matching trajectory of GAFL in which noise was sampled from a Gaussian distribution for the translations and from a uniform distribution for the rotations and was then transformed into a protein backbone structure.

In order to benchmark the model, we generated structures for proteins with lengths of between 100 and 300, for which we display examples in Figure 33. Typically, the performance of protein structure generation models is assessed by testing whether a sequence can be found that folds into the sampled structure – a proxy for the physical consistency of the sample (Figure 34, next page). Additionally, structural similarity to the training set (novelty) and between generated samples (diversity) is measured. Since we observed an over-representation of helices in the structures generated by current state-of-the-art models, we also included secondary structure content in our metrics. The aim was to have content similar to that observed in natural proteins if the probability distribution is learned accurately.

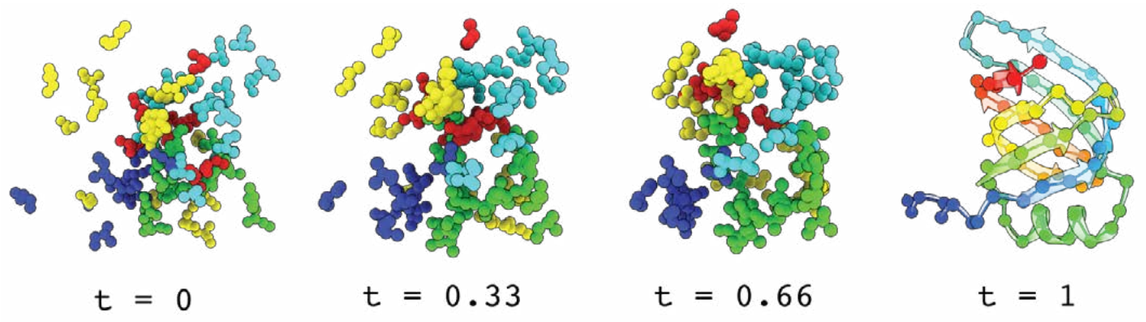


Figure 32: The learned flow transformed backbone fragments that were sampled from a noise distribution into a protein structure.

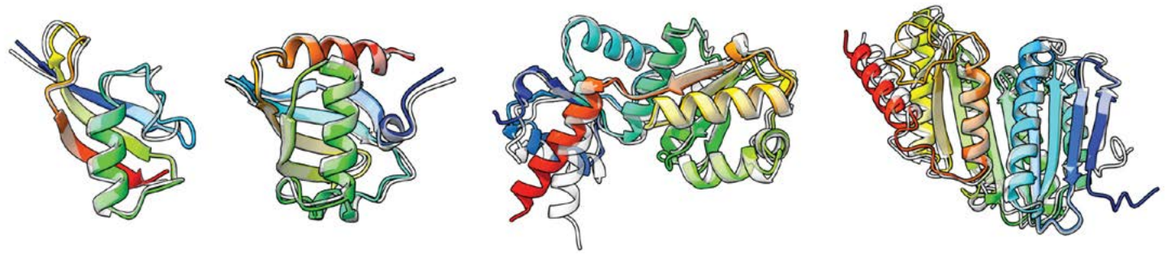


Figure 33: Protein backbone structures generated with GAFL, and their re-folded states in grey.

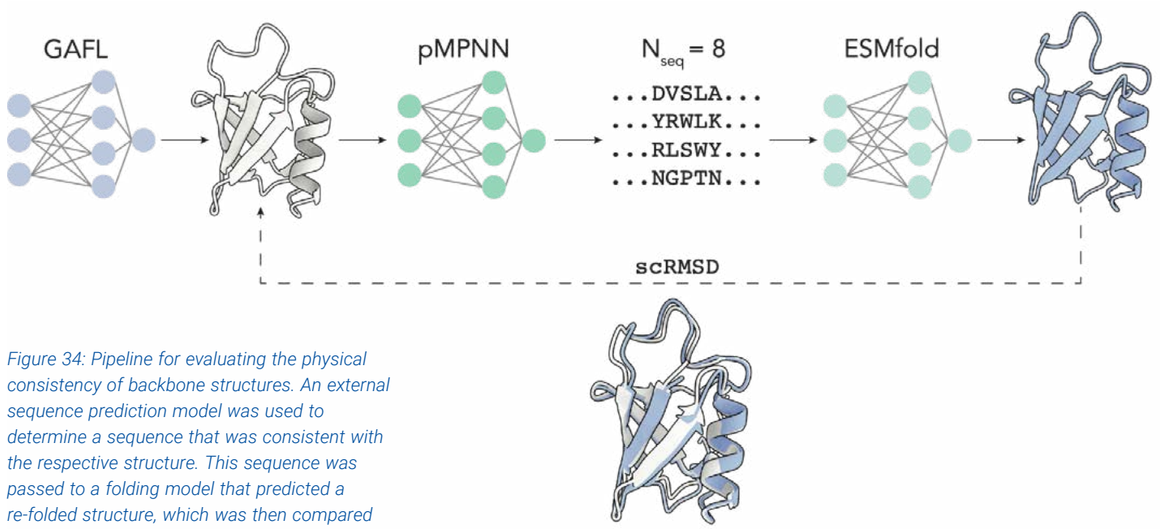


Figure 34: Pipeline for evaluating the physical consistency of backbone structures. An external sequence prediction model was used to determine a sequence that was consistent with the respective structure. This sequence was passed to a folding model that predicted a re-folded structure, which was then compared with the original generated structure.

In an ablation study, we trained models with both our proposed CFA architecture and the established IPA. We found that CFA consistently performs better at the tradeoff between designability and balanced secondary structure content (Figure 35).

We compared GAFL with the current state-of-the-art model for protein structure generation: namely RFdiffusion (Watson et al., De novo design of protein structure and function with RFdiffusion, Nature 620, 1089–1100, 2023), which relies on weights from a folding model that are trained on a dataset that is much larger than the monomeric PDB. In so doing, we found that at a speed-up by a factor of three, GAFL achieved the same designability in addition to slightly increasing diversity and novelty. The secondary structure content of RFdiffusion was strongly length-dependent: For small, domain-sized proteins of up to 150 residues, beta strands were strongly under-represented. GAFL, on the other hand, generated strands to a similar extent as observed in natural proteins, however, it over-represented helices for large structures (Figure 36).

With these results, our work not only demonstrated that protein structure generation models can achieve high performance without relying on weights from pre-trained folding models, but also provided a promising tool for the design of domain-sized proteins – an interesting space of proteins due to the increased cost of expressing large proteins in the lab. We achieved this feat by introducing CFA – a PGA-based extension of the widely used IPA architecture that demonstrates the benefits of geometric inductive biases and that could enable improvements in protein structure models for other domains in the future.

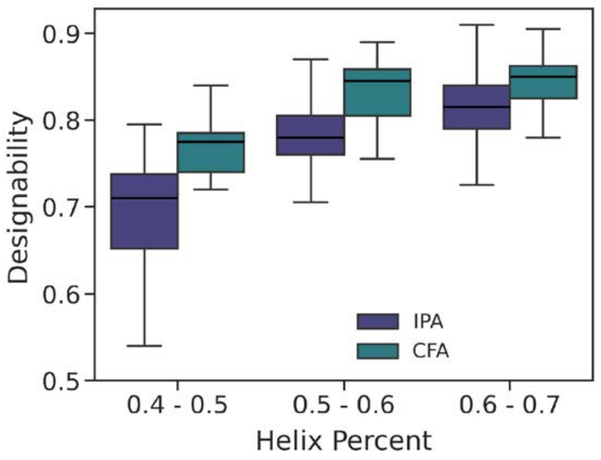


Figure 35: Designability as a function of the helix content for several snapshots of structure generation models that rely on IPA and CFA, respectively. CFA achieved higher designability for the same helix content.

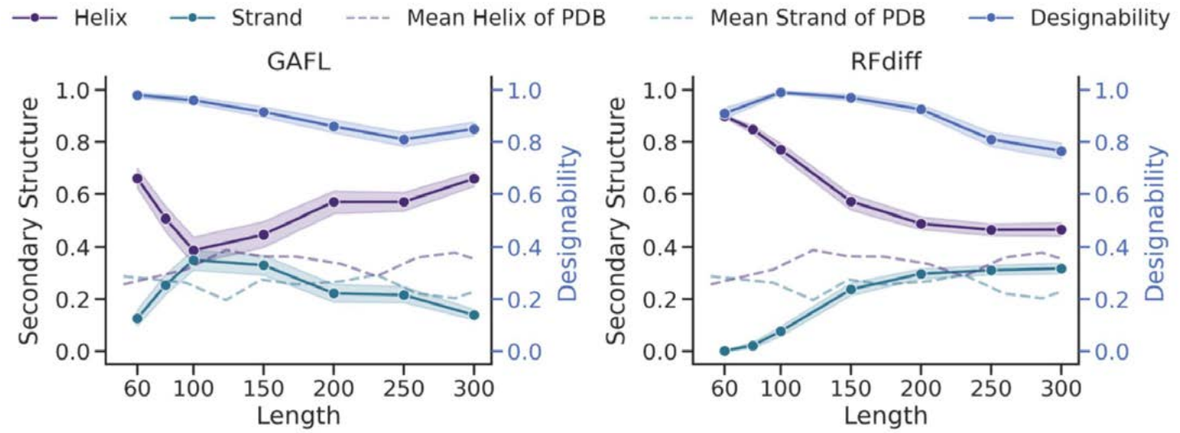


Figure 36: Designability and secondary structure content of GAFL and RFdiffusion as a function of protein length.

Die **MLI Gruppe** beschäftigt sich mit der Entwicklung von neuartigen Algorithmen und Verfahren des maschinellen Lernens. Besondere Schwerpunkte sind hierbei im Bereich des Geometric Deep Learning und der 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. Ursprünglich wurde dieser Ansatz entwickelt, um die Eigenschaften von Graph Neural Networks zu untersuchen, doch mittlerweile findet er breite Anwendung zur Beschreibung und Entwicklung von unterschiedlichsten Modellarchitekturen des Maschinellen Lernens. 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 Clifford-Algebren gelang es, eine ausgewogenere Verteilung von Sekundärstrukturen von Proteinen zu erhalten, den Beta-Sheets und Alpha-Helices. Ausgehend von diesen positiven Resultaten untersucht die Forschungsgruppe nun Kontext-spezifische Methoden des Proteindesigns, bei denen ein Protein mit einer für einen Ligand passenden Bindungsstelle gesucht wird. Dieser Ansatz ermöglicht vielversprechende Anwendungen zum Beispiel in der personalisierten Krebstherapie.
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. Einen 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.

Die Forschung der MLI Gruppe zu interpretierbaren Modellen trägt zu dem wachsenden Forschungsfeld der neurosymbolischen Methoden bei, welche eine Brücke zwischen den „sub-symbolischen" Repräsentationen eines neuronalen Netzes und den „symbolischen" Methoden des logischen Schließens schlagen. Indem die internen Repräsentationen eines neuronalen Netzes in symbolische Repräsentationen übersetzt werden, können beide Ansätze der KI miteinander kombiniert werden.

2 Research

2.7 Molecular Biomechanics (MBM)

Group leader

Prof. Dr. Frauke Gräter

Team

Dr. Camilo Aponte-Santamaria (staff scientist and acting group leader)
Dr. Saber Boushehri (until October 2024)
Matthias Brosz
Johanna Buck
Jannik Buhr
Dr. Svenja de Buhr
Guido Giannetti (visiting scientist; University of Vienna)
Christina Goß
Paul Greicius
Eric Hartmann
Chiara Keil
Dr. Markus Kurth (until February 2024)
Eddie Merbitz (student; February–April 2024)
Dr. Erik Poppleton
Dr. Kai Riedmiller (until October 2024)
Andrea Sassoli
Boris Schüpp
Leif Seute
Daniel Sucerquia
Evgeni Ulanov
Aysecan Ünal
Nuriza Suleimenova
Viliuga Vsevolod (until September 2024)
Nicolas Wolf

For the past 15 years, our mission at HITS has been to unravel how and why Nature senses mechanical stress and is amazingly resilient against it but eventually fails under excessive mechanical loads. These years have been full of exciting research, unexpected discoveries, and fun.

With the transition of the Molecular Biomechanics group to the Max Planck Institute for Polymer Research in Mainz as of January 2025, it is time for us to bid farewell. This move marks the beginning of a new chapter, but we would like to take this opportunity to once again express our deepest gratitude to the former and current members of HITS, to Klaus Tschira, and to the Klaus Tschira Foundation for making our research possible; to our collaborators near and far; to our friends and colleagues at Heidelberg University and KIT; and to all our funding partners: Thank you for your unwavering support and inspiration. While we embark on this new journey, we are committed to maintaining strong connections with HITS through the HITS Lab, SIMPLAIX, and beyond. We will also remain closely associated with Heidelberg University and are immensely thankful for the university's continued support and the close relationship they have fostered with HITS on so many levels. These partnerships and collaborations will undoubtedly continue to shape and enrich our work in the years to come.

In this, our final contribution to the HITS annual report, we are happy to report on our new insights into lubricin (a natural lubricant), into why collagen networks can shrink and swell, and into how biological samples can be used in 3D printing. We also present Grappa, our new machine learning method trained on quantum mechanical data that is used to predict molecular dynamics parameters. We look forward to many further scientific and personal interactions with the HITS community for many years to come.

O-glycans expand lubricin and attenuate its viscosity

Lubricin is a protein that helps joints move smoothly and stay functional. However, the molecular basis for this function was hitherto unknown. We therefore investigated the impact of glycosylation on the structure and viscosity of lubricin through extensive MD simulations and continuum simulations. In so doing, we gained valuable insights into the role of O-glycans in influencing lubricin's properties.
On the one hand, our findings revealed that glycosylated disordered lubricin solutions exhibit lower viscosities at zero shear and thus also lower friction when compared with nonglycosylated proteins at the same mass density. On the other hand, glycosylation was found to attenuate the pronounced loss of viscosities. According to our results, the molecular basis of this effect is the presence of O-glycans, which promote the dispersion of lubricin, thereby preventing the clustering of these O-glycans. Our continuum model showcases how quantitative viscosity relations from nanometer-scale molecular simulations can be harnessed in order to draw conclusions on much larger length scales and for more complex boundary conditions, such as rugged cartilage surfaces sheared against one another. The insights from this work contribute to our understanding at the molecular level of how lubricin's conformational dynamics tailor its lubricating function in synovial joints.

3-dimensional (3D) printing at the interface of biological materials

3D bioprinting makes use of 3D printing and related technologies to manipulate and arrange cells and their scaffolds. Applications range from single cell research to the fabrication of biomedical parts that imitate natural tissues and organs.

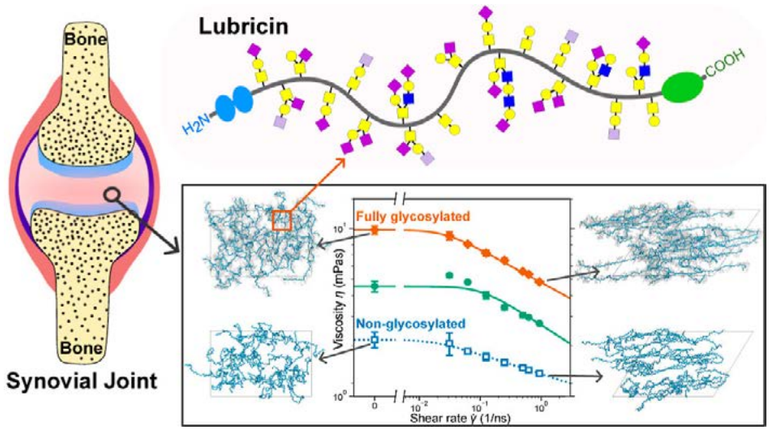


Figure 37: Structure and shear thinning behavior of lubricin.

The selection of cytocompatible inks that do not exert toxicity in the interior of cells is critical to the success of printing. In laser-based 3D bioprinting, photoresist molecules are typically used as bioinks. However, whether these molecules permeate across biological membranes and potentially induce cytotoxic effects has remained largely unknown. By combining simulations and experiments, we were able to classify several inks that are commonly used for 3D bioprinting according to their residence in biological lipid membranes and their permeation through these membranes. In addition, we also identified the main molecular factors that govern the permeation of such light-sensitive molecules across lipid bilayers (Figure 37). This information allowed us to classify photoresists as highly permeable and thus as being potentially implicated in cytotoxicity or as impermeable and thus as being suitable for 3D printing at the cell boundary. This information will help to guide the selection of cytocompatible photoresists for 3D bioprinting applications in the future [Diedrich et al, 2024].

A key component in the mechanical stability of scaffolds generated by 3D printing is collagen. A micro-printed collagen structure has recently been observed to reversibly shrink upon heating rather than to expand as intuition would

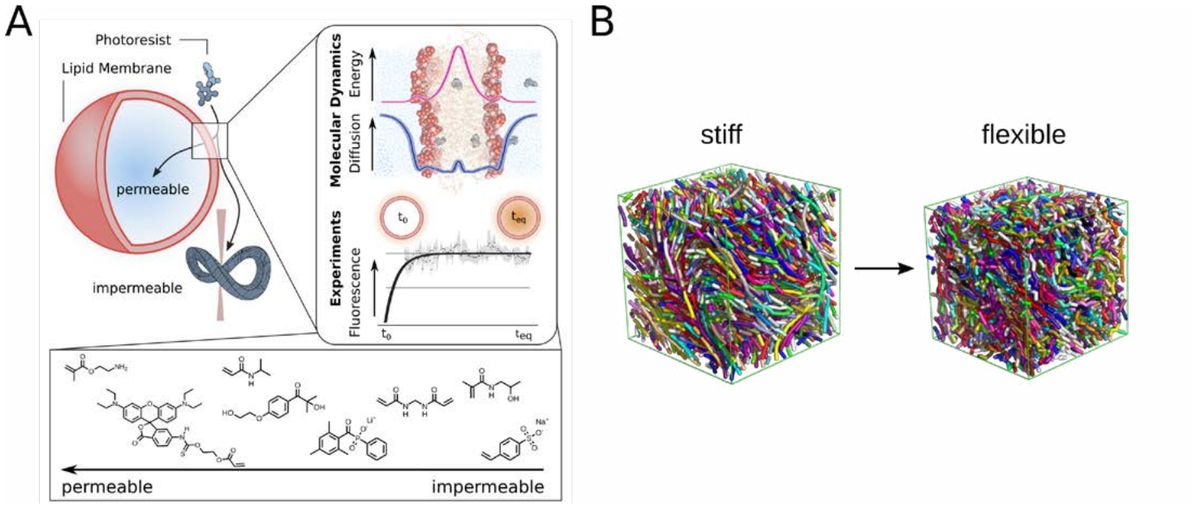


Figure 38: 3-dimensional (3D) printing at the interface of biological materials. (A) For laser-based 3D bioprinting at the interface of synthetic cells, inks used for the printing process should not permeate across the lipid membrane. By combining simulations and experiments, we were able to classify inks that are commonly used for 3D bioprinting according to their residence in biological lipid membranes and their permeation through these membranes. (B) A mixture of cross-linked semi-flexible polymers shrinks by making the polymer chains flexible. This effect mimics the shrinking that collagen structures undergo upon denaturation induced by heating.

suggest. Simulations of a cross-linked mixture of polymers that mimic such collagen structures suggest that this effect is due to the denaturation of the collagen fibers, which makes them change from a stiff configuration into a more flexible one (Figure 38). This interesting property opens new perspectives for designing novel responsive biomaterials for 3D printing [Mainik et al, 2024].

Grappa – A machine-learned molecular mechanics force field

Simulating large molecular systems over long timescales requires force fields that are both accurate and efficient. In recent years, E(3)-equivariant neural networks have lifted the tension between the computational efficiency and accuracy of force fields, but they are still several orders of magnitude more expensive than are established molecular mechanics (MM) force fields. We therefore introduced Grappa, which is a machine learning framework used to predict MM parameters from the molecular graph by employing a graph attentional neural network and a transformer with symmetry-preserving positional encoding. The resulting Grappa force field outperforms traditional and machine-learned MM force fields in terms of accuracy at the same computational efficiency and can be used in existing molecular dynamics software. Grappa predicts energies and forces of small molecules, peptides, and RNA at state-of-the-art MM accuracy while also reproducing experimentally measured values for J-couplings. We demonstrated Grappa's transferability to macromolecules in MD simulations from a small fast-folding protein up to a whole virus particle. Our force field sets the stage for biomolecular simulations that are closer to chemical accuracy but with the same computational cost as established protein force fields (Figure 39).

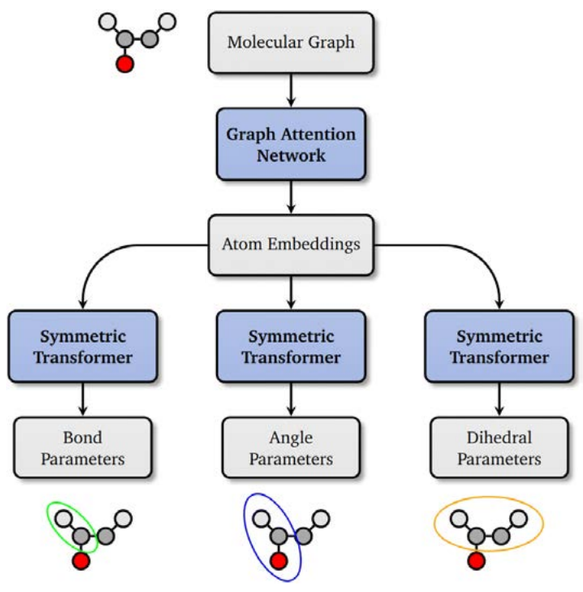


Figure 39: Architecture of the Grappa framework.

Precise rupture of DNA under ultrasound

Ultrasound with sufficient energy density has long been known to be capable of damaging DNA or even rupturing it into smaller fragments. In collaboration with the experimental group of Andreas Herrmann from Aachen, we redirected the previously unspecific fragmentation of DNA to dedicated weak points that were determined by partially cleaved DNA phosphate backbones, which are known as "nicks."

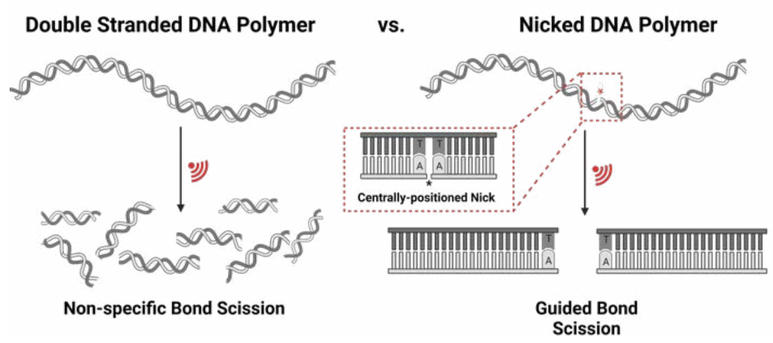


Figure 40: The mechanism of this novel DNA fragmentation through ultrasound.

The mechanism of this novel DNA fragmentation through ultrasound was investigated and resolved through the analysis of next-generation sequencing (NGS) data and molecular dynamics simulations conducted by our group. Using MD simulations, we were able to show that the base pairing in the base pairs adjacent to the nick is broken under mechanical stress, thereby leading to a phenomenon called "fraying," which exposes a single-stranded region opposite the nick. Fragmentation of nicked DNA then takes place within the exposed single-stranded region, which explains the experimentally observed guided bond scission in the vicinity of the nick. In order to support this hypothesis, a modulation of the central point and the width of the observed breaking distribution was achieved experimentally by strengthening or weakening the base pairing through DNA sequence variation (see Figure 40).

Unser Ziel ist es – und das schon seit 15 Jahren am HITS – zu enträtseln, wie die Natur mechanische Belastungen wahrnimmt, gleichzeitig erstaunlich widerstandsfähig dagegen ist, aber schließlich bei übermäßiger mechanischer Belastung versagt. Diese Jahre waren voll von spannender Forschung, unerwarteten Entdeckungen und Spaß.

Nun heißt es Abschied nehmen, denn die **Gruppe Molekulare Biomechanik** wechselt ab Januar 2025 an das Max-Planck-Institut für Polymerforschung in Mainz. Dieser Umzug markiert den Beginn eines neuen Kapitels, aber wir nutzen diese Gelegenheit, um noch einmal unseren tiefsten Dank auszusprechen. An die ehemaligen und aktuellen Mitglieder des HITS, an Klaus Tschira und die Klaus Tschira Stiftung für die Ermöglichung unserer Forschung, an unsere Mitarbeitenden in der näheren und weiteren Umgebung, an unsere Freund*innen und Kolleg*innen an der Universität Heidelberg und am KIT sowie an alle unsere Finanzierungspartner - vielen Dank für Ihre unermüdliche Unterstützung und Inspiration. Während wir uns auf diesen neuen Weg begeben, werden wir durch das HITS Lab, SIMPLAIX (siehe Kapitel 7) und darüber hinaus enge Verbindungen zu HITS aufrechterhalten. Wir sind auch weiterhin eng mit der Universität Heidelberg verbunden und sind sehr dankbar für die anhaltende Unterstützung der Universität und die enge Beziehung, die sie auf so vielen Ebenen mit HITS aufgebaut hat. Diese Partnerschaften und Kooperationen werden zweifellos auch in den kommenden Jahren unsere Arbeit prägen und bereichern.

In diesem letzten Beitrag zum HITS-Jahresbericht berichten wir über unsere neuen Erkenntnisse über das natürliche Schmiermittel Lubricin, darüber, warum Kollagennetzwerke schrumpfen und anschwellen können, und darüber, wie biologische Proben für 3D-Druck verwendet werden können. Außerdem stellen wir Grappa vor, unsere neue Machine Learning-Methode, die auf quantenmechanischen Daten trainiert wird, um Molekulardynamikparameter vorherzusagen. Wir freuen uns auf viele wissenschaftliche und persönliche Interaktionen mit der HITS-Familie in den kommenden Jahren.

2.8 Molecular and Cellular Modeling (MCM)

Group leader

Prof. Dr. Rebecca Wade

Team

Riccardo Beccaria
Tommaso Bartoloni
Sophia Ber (until March 2024)
Mislav Brajkovic
Elisa Brost (July–October 2024)
Paul Brunner (since October 2024)
Dr. Giulia D'Arrigo (until February 2024)
Selina Ernst (May–June 2024)
Yassine Faik (January–April 2024)
Giorgi Georgiev (November–December 2024)
Manuel Glaser (until June 2024)
Edin Hepgülüm (September–December 2024)
Nico Herb (since May 2024)
Feroz Karim (since December 2024)
Jonathan Kunz (February–June 2024)
Fabian Mulder (March–April 2024)
Abraham Muniz Chicharro (until March 2024)
Jakob Niessner (until April 2024)
Dr. Giulia Paiardi
Dr. Stefan Richter
Raziye Sari (since December 2024)
Dr. Kushal Singh (since October 2024)
Jonathan Teuffel
Congcong Xu (until March 2024)

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 and to applications such as protein engineering and structure-based drug design.

In the MCM group, we are primarily interested in understanding how biomolecules interact. What determines the specificity and selectivity of a drug–receptor interaction? How can proteins assemble to form a complex? How is the assembly of a complex influenced by the crowded environment of a cell? What makes some binding processes quick and others slow? How do the motions of proteins affect their binding properties? One of our aims is to gain a mechanistic molecular-level understanding of drug interactions along the process that extends from drug delivery to drug–target binding and further to drug metabolism.

We take an interdisciplinary approach that entails collaboration with experimentalists and that makes concerted use of computational approaches based on physics and bio-/chem-informatics. The broad spectrum of techniques developed and employed ranges from interactive, web-based visualization tools to machine learning methods and atomic-detail molecular simulations.

In this report, we outline some of the results achieved in 2024. Following a general overview of what happened in the group last year, we highlight results in three research projects:

i. the interplay between conformational variability and functional interactions in a model drug-metabolizing cytochrome P450 as revealed via multiresolution molecular dynamics simulations,
ii. the computational screening of the effects of mutations on the residence times and dissociation mechanisms of protein–protein complexes, and
iii. SARS-CoV-2 spike-cell receptor binding: the effects of heparan sulfates, N-glycans, heparin, and heparin analogues.

What happened in the lab in 2024?

Last year, we completed the Bridges stage of our project in the Informatics4Life (I4L) consortium with the support of the Klaus Tschira Foundation (see also section 7). Within the project, which is carried out in cooperation with Patrick Most and Julia Ritter (Molecular and Translational Cardiology Section, Department of Cardiology, Angiology and Pneumology, Heidelberg University Hospital), we investigate the synthetic peptide S100A1ct. This peptide is derived from the terminal region of the protein S100A1, which is an important controller of cardiac function. The S100A1ct peptide provides a promising new approach to the therapy of acute heart muscle weakness [Kehr, 2024]. At HITS, we developed and validated a computational pipeline for docking peptides to membrane-exposed surfaces of transmembrane proteins, such as predicted molecular effectors in the diseased heart cells. We applied this pipeline in order to investigate the mechanism by which S100A1ct exerts its effects by modeling peptide SERCA2a Ca^{2+}-ATPase complexes [Kehr, 2024]. In ongoing work, we are currently designing variants and peptidomimetics of the S100A1ct peptide and are carrying out simulations in order to investigate their mechanisms.

EBRAINS 2.0 is an EU-supported project designed to support the development of the services of EBRAINS (European Brain Research Infrastructures). EBRAINS is a collaborative research platform (ebrains.eu) that is designed to advance neuroscience and brain health. In EBRAINS 2.0, we are working on developing use cases for molecular-level tools. We have validated our τRAMD approach for computing the relative residence times of protein–protein complexes [D'Arrigo, 2024] (see below for more details) and are now applying this approach to interleukin-based signaling using a combination of our τRAMD and SDA approaches for computing binding kinetic parameters with the mathematical modeling of subcellular signaling networks.

The MCM group participates in two SIMPLAIX research projects (www.simplaix.org) as well as in the organization of SIMPLAIX activities, such as the second workshop, which was held in May (see section 7, SIMPLAIX). In one research project, the aim is to decipher the determinants of electron transfer in membrane-bound cytochrome P450-redox protein complexes. As part of this endeavor, we carried out multiresolution molecular dynamics simulations in order to reveal the interplay between conformational variability in a model drug-metabolizing cytochrome P450 and its functional interactions [Han, 2024] (see below). In ongoing work, we are currently combining quantum mechanical calculations with atomistic and coarse-grained molecular dynamics simulations in order to compute inter-protein electron transfer rates. In the second project, the aim is to harness the power of machine learning for exploring peptide–protein complexes and for peptide design. With Ullrich Köthe (Heidelberg University), we investigated novel machine learning approaches for this purpose. Furthermore, we built computational pipelines to exploit AI-based structure prediction algorithms and molecular simulations for peptide design. For example, with the support of an AI Health Innovation Cluster grant and in collaboration with Elke Burgermeister (University Hospital Mannheim), Giulia Paiardi applied such a pipeline to the design of a peptide to tackle gastric cancer.

We published our simulations of how the interactions of SARS-CoV-2 spike with its host receptor are affected by heparan sulfate proteoglycan co-receptors [Paiardi, 2024] (see below). Building on our simulations, we identified heparin analogues with improved antiviral activity compared with heparin. In collaboration with Petr Chlandla (Bioquant, Heidelberg University) in the CoVLP project – which is supported by the Engineering Molecular Systems Flagship Initiative of Heidelberg University – and Marco Rusnati (University of Brescia, Italy), these heparin analogues were characterized experimentally and showed antiviral activity without the anticoagulation activity of heparin.

In 2024 postdocs Giulia D'Arrigo and Giulia Paiardi left the group, with both transitioning to jobs in companies in Mannheim. Abraham Muniz Chicharro and Manuel Glaser also completed their terms in the group. Following their master's studies in chemistry and physics, respectively, Sophia Ber went on to PhD studies in the UK, and Jakob Niessner began his PhD studies in Vincent Heuveline's group at

Heidelberg University, where he is currently developing an efficient solution to the Poisson–Boltzmann equation – an important component of the HITS Lab project between the DMQ and MCM groups that Riccardo Beccaria (a PhD student in the MCM group) is currently working on. In the fall of 2024, Kushal Singh joined the group as a postdoc to work on molecular dynamics simulations in order to investigate resistance mutations in a malaria drug transporter protein in collaboration with Michael Lanzer (Center of Infectious Diseases, Parasitology, Heidelberg University Hospital). Toward the end of the year, Feroz Karim joined the group on a DAAD scholarship for doctoral studies on the machine learning of histone protein interactions. In addition, Jonathan Kunz successfully completed his bachelor's thesis research in biochemistry. Furthermore, Edin Hepgülüm (molecular biosciences, molecular and cellular biology), Selina Ernst, Elisa Brost, Yassine Faik, Giorgi Georgiev (all molecular biotechnology), and Fabian Mulder (molecular systems biology) completed internships in the group as part of their master's studies at Heidelberg University.

The 24th iteration of the European Symposium on Quantitative Structure Activity Relationships (EuroQSAR) – which was held in Barcelona, Spain, in September – was co-organized by Rebecca Wade as scientific chair of the QSAR, Chemoinformatics and Modeling Society (QCMS) and Jordi Mestres (University of Girona), who was the symposium chair. With about 400 participants and an excellent speaker line-up, the symposium proved highly successful. Several members of the MCM group presented their work at the meeting, and Giulia Paiardi co-organized a career workshop at the symposium.

Interplay between conformational variability and functional interactions in a model drug-metabolizing cytochrome P450 revealed by multiresolution molecular dynamics simulations

Cytochrome P450 2B4 (CYP 2B4) is one of the best-characterized members of the cytochrome P450 (CYP) enzyme family and serves as a key model system for understanding the mechanisms of microsomal class II CYPs, which metabolize most known drugs. The highly flexible nature of CYP 2B4 is apparent from crystal structures that show the active site with either a closed or a wide-open heme-binding cavity (see Figure 41). We investigated the conformational ensemble of the full-length CYP 2B4 in a phospholipid bilayer using multiresolution molecular dynamics simulations. Coarse-grained molecular dynamics simulations revealed two predominant orientations of CYP 2B4's globular domain with respect to the bilayer. Their refinement via an atomistic resolution molecular dynamics simulation showed adaptation of the enzyme's interaction with the lipid bilayer, thereby leading to open configurations that facilitate ligand access to the heme-binding cavity. Analysis of enzyme tunnels, water routes, and random acceleration molecular dynamics (RAMD) simulations of ligand dissociation support the conformation-dependent passage of molecules between the CYP2B4 active site and the protein surroundings. Furthermore, the simulation of the re-entry of an inhibitor – the antifungal drug bifonazole – into the open conformation of CYP 2B4 resulted in binding at a transient hydrophobic pocket within the active site cavity that may play a role in substrate binding or in allosteric regulation. Together, these results reveal how the open conformation of CYP 2B4 facilitates the binding of substrates from – and the release of products into – the membrane, whereas the closed conformation prolongs the residence time of substrates or inhibitors and selectively allows the passage of smaller reactants via the solvent and water channels. As CYP 2B4 is a representative of the class of cytochrome P450 enzymes that metabolize most known drugs, the insights from this study are pertinent to the prediction of drug metabolism and the design of CYP inhibitors as therapeutics.

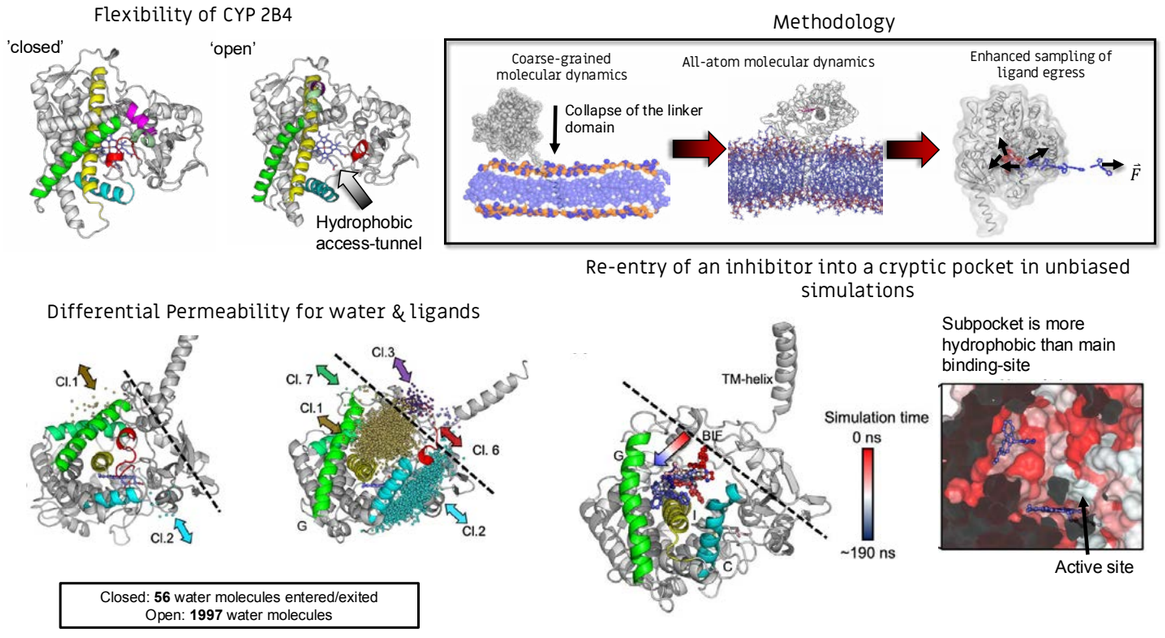


Figure 41:. Illustration of the multiresolution molecular dynamics simulation study that reveals the interplay between the conformational variability and functional interactions of cytochrome P450 2B4 (CYP 2B4), a model drug-metabolizing enzyme. Top left: Crystal structures show that CYP 2B4 can adopt a wide range of conformations, including those with a closed or an open active site cavity. CYP 2B4 is shown in cartoon representation with key helices colored. Top right: The multiresolution simulation pipeline that is used to model and simulate membrane-bound CYP 2B4. Bottom left: The passage of water molecules to and from the active site of CYP 2B4 in conventional atomistic molecular dynamics trajectories of the closed (left) and open (right) conformations. The exit points of water routes are displayed as spheres colored by cluster number, with Cl. 1 being the largest. The directions of the routes followed by the clusters are indicated by arrows. The approximate position of the surface of the membrane is shown by a dashed line. Bottom right: Following dissociation in RAMD simulations, conventional atomistic molecular dynamics simulations captured the re-entry of the antifungal drug bifonazole (BIF, colored according to time) into CYP 2B4, resulting in its binding at an additional transient subpocket. The close-up view displays the surface of the cavity inside the protein colored by hydrophobicity, which shows that the transient subpocket is more hydrophobic than the main substrate-binding site, thereby supporting its ability to bind to compounds such as bifonazole. Adapted from [Han, 2024].

Computational screening of the effects of mutations on the residence times and dissociation mechanisms of protein–protein complexes

The dissociation rate – or its reciprocal, residence time (τ) – is a crucial parameter for delineating the duration and biological impact of biomolecular interactions. Accurate prediction of τ is essential to understanding protein–protein interactions and to identifying potential drug targets or modulators for tackling diseases. Conventional molecular dynamics simulation techniques are inherently constrained by their limited timescale, which is typically microseconds for proteins, thereby making it challenging to estimate residence times, which can range from minutes to hours. Having previously developed τ-random acceleration molecular dynamics (τRAMD) for computing the relative residence times of protein–small molecule complexes and having also successfully applied the technique to a diverse range of systems, in this study, we investigated the use of τRAMD for estimating the residence times of protein–protein complexes [D'Arrigo, 2024]. τRAMD enables the observation of the dissociation of two proteins in simulations on the nanosecond timescale, thereby facilitating the rapid and efficient computation of relative residence times.

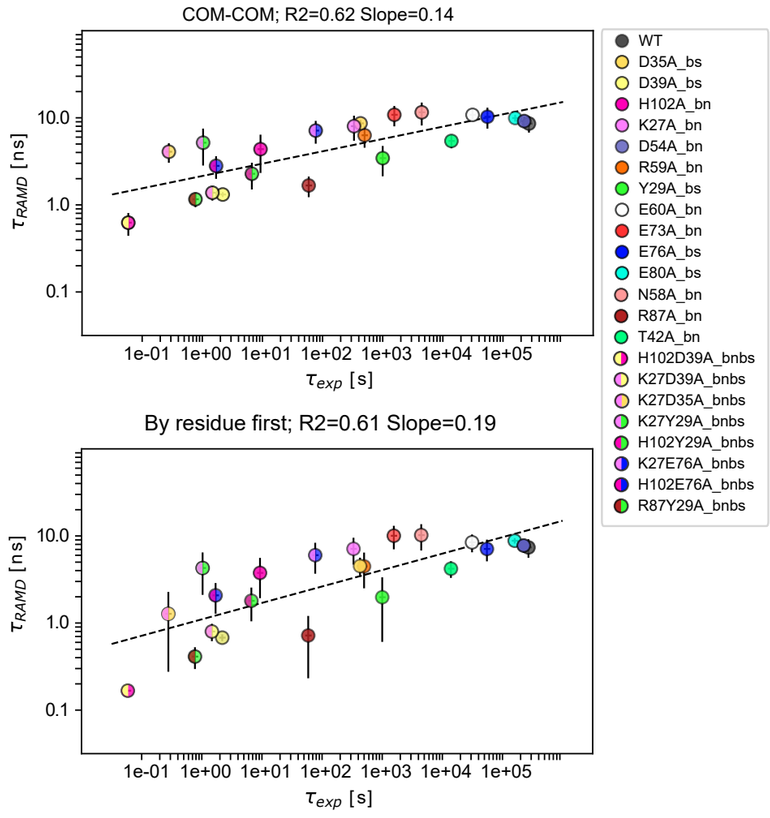


Figure 42: Example of the correlation of computed RAMD residence times (τRAMD) with experimental residence time (τexp, inverse of koff) for protein–protein complexes obtained with the τRAMD procedure. Results are shown for the enzyme-inhibitor complexes between barnase (Bn) and barstar (Bs) and their mutants. The mutants have single or double mutations and are denoted by color. Results are shown for two of the definitions of computed residence time that were evaluated, both of which give a similar correlation (R2 ~ 0.6 with dashed line shows the straight-line fit). From [D'Arrigo, 2024].

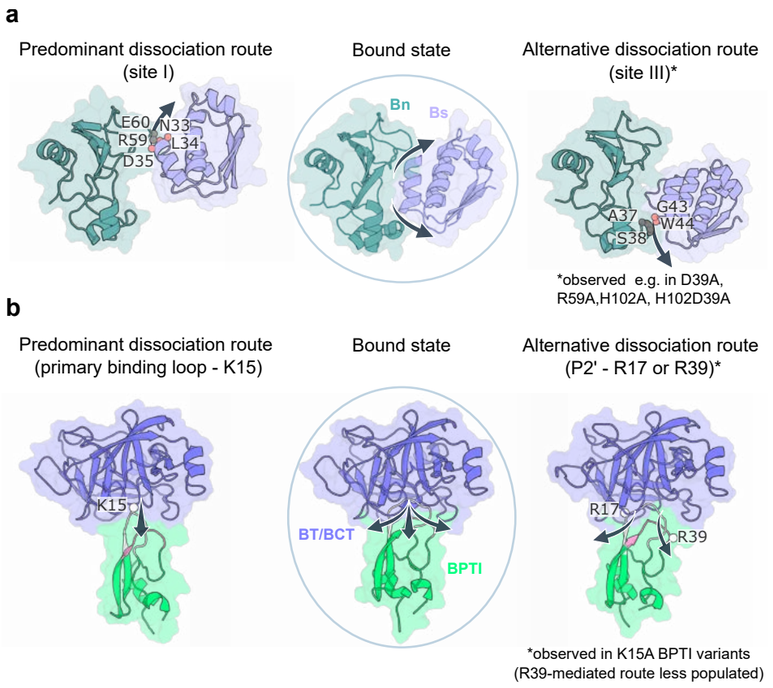


Figure 43: Representative protein–protein dissociation pathways derived from RAMD simulations of the dissociation of protein–protein complexes. Comparison of the bound state and the predominant and alternative dissociation routes for wild-type barnase (Bn)-barstar (Bs) complex and its mutants (a) and for trypsin/chymotrypsin (BT/BCT)-bovine pancreatic trypsin inhibitor (BPTI) complexes and their mutants (b). Key interacting residues are labeled. The predominant dissociation routes are found for the complexes with longer residence times, whereas mutations that lead to shorter residence times result in increased dissociation via the alternative routes. From [D'Arrigo, 2024].

We evaluated the τRAMD methodology for three protein–protein complexes and their extensive mutant datasets, achieving good agreement between computed and experimental data (Figure 41). By combining τRAMD with MD–IFP (Molecular Dynamics – Interaction Fingerprint) analysis, dissociation mechanisms were characterized, and their sensitivity to mutations were investigated, thereby enabling the identification of molecular hotspots for the selective modulation of dissociation kinetics (Figure 42). Our findings underscore the versatility of τRAMD as a simple and computationally efficient approach that breaks new ground by enabling the in silico screening of sets of mutations with respect to protein–protein residence times and dissociation mechanisms. τRAMD can thus be used to aid in the design of modulators of protein–protein interactions. We are now applying τRAMD to other types of protein–protein complexes, including spike-ACE2 complexes (see below).

SARS-CoV-2 spike-cell receptor binding: The effects of heparan sulfates, N-glycans, heparin, and heparin analogues

SARS-CoV-2 – the virus responsible for COVID-19 – infects cells by attaching its spike glycoprotein to the ACE2 receptor on the surface of human cells. However, this binding is not as simple as a lock and key. While the roles of heparan sulfate and N-glycans in the interaction between the viral spike and the host cell receptor ACE2 have been unclear, their variable expression in different cells and individuals appears to correlate with the virus's ability to infect these cells. We therefore built molecular models and carried out atomic-detail molecular dynamics simulations in order to gain a deeper understanding of how heparan sulfates and N-glycans are involved in the SARS-CoV-2 infection mechanism and how they relate to host cell susceptibility [Paiardi, 2024]. The simulations enabled us to analyze the molecular details and to find that heparan sulfate and N-glycans on the spike and ACE2 receptor synergistically strengthen the spike-ACE2 complex (see Figure 44). This effect results from structural rearrangements and from the propagation of forces in the spike upon ACE2 and heparan sulfate binding. RAMD simulations of the dissociation of ACE2 and spike in the presence of glycosaminoglycan (GAG) chains showed the mechanisms by which interfacial glycans and the GAG chains stabilize the ACE2-spike complex. Furthermore, force distribution analysis showed how the binding of ACE2 allosterically influences functional sites on spike. Together, the simulations revealed how host cell heparan sulphate proteoglycan co-receptors can act as accomplices to the ACE2 receptors, thereby facilitating viral infection (seeFigure 45).

These findings unlock potential new therapeutic approaches to combating SARS-CoV-2 by developing heparan sulfate mimetics that prevent the virus from binding to and entering human cells. We previously showed through molecular dynamics simulations and experiments that heparin could exert its antiviral effect against SARS-CoV-2 via direct and allosteric mechanisms by binding to the viral spike [Paiardi, 2022]. Using information from our simulations of heparin-spike complexes in the presence and absence of the ACE2 receptor, we are currently assessing heparin analogues with variable levels of sulfation for their ability to show improved antiviral activity without the anticoagulant activity of heparin. This work is being conducted in collaboration with Marco Rusnati (University of Brescia, Italy) and Petr Chlandla (Heidelberg University).

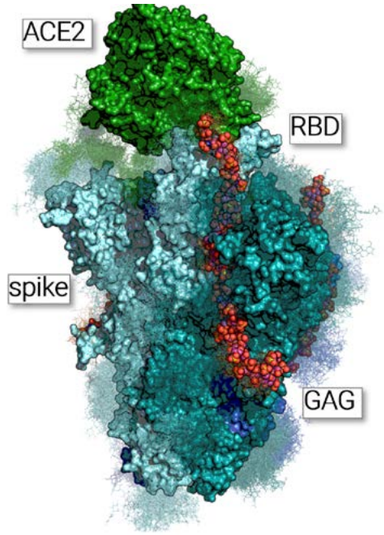


Figure 44: Simulated structural model of the open, active spike homotrimer, with the ACE2 receptor–binding domain and three glycosaminoglycan (GAG) chains bound. The stabilizing interactions of the GAG chains and N-glycans with spike and ACE2 can be observed. The three subunits of spike (shades of blue) and ACE2 (green) are shown as molecular surfaces. N-glycans that are covalently attached to the spike and ACE2 are shown in line representation and are colored according to the protein to which they are attached. A total of 40 superimposed frames of the N-glycan structures that were collected at intervals of 25 ns from a molecular dynamics simulation are shown. The 31mer GAG chains are shown as spheres colored by element with magenta carbons. One GAG chain spans from the spike's up-receptor-binding domain (RBD) to its S1/S2 multifunctional domain on the adjacent subunit. The other two chains are partially occluded from view and follow a similar path, simultaneously binding the spike's down-RBD as well as the N-terminal and S1/S2 domains of an adjacent subunit. Adapted from [Paiardi, 2024].

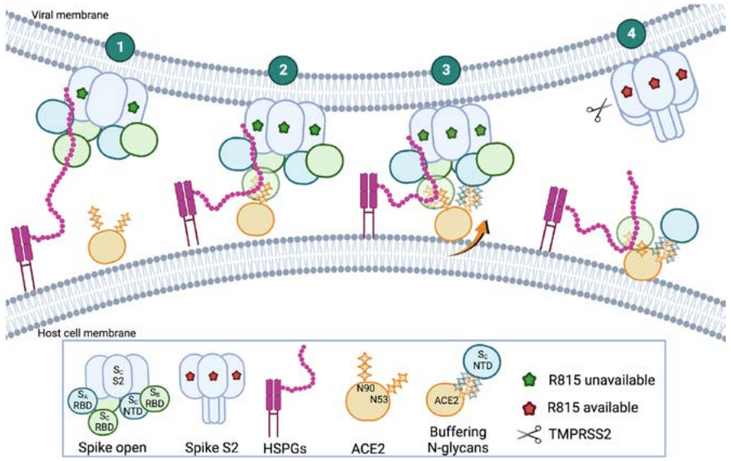


Figure 45: Illustration of the proposed role of heparan sulfate proteoglycans (HSPGs) in facilitating the encounter of spike SARS-CoV-2 with ACE2 (1), the binding of the two proteins mediated by heparan sulfates (GAGs) and N-glycans (2), the subsequent reorientation of the spike (3), and the activation and cleavage of the viral protein (4). Illustration created with BioRender.com. This illustration has been reproduced with permission from [Paiardi, 2024]. Copyright 2024 National Academy of Sciences (PNAS).

In der **Molecular and Cellular Modeling Gruppe** (MCM) sind wir in erster Linie daran interessiert zu verstehen, wie Moleküle interagieren. Was bestimmt die spezifische und selektive Wirkung beim Zusammenspiel von Wirkstoff und Rezeptor? Wie werden Proteinkomplexe gebildet und welche Formen können sie annehmen? Welche Wirkung hat die beengte Zellumgebung auf die Bildung eines Proteinkomplexes? Warum verlaufen einige Bindungsprozesse schnell und andere langsam? Welche Auswirkungen haben Proteinbewegungen auf ihre Bindungseigenschaften?

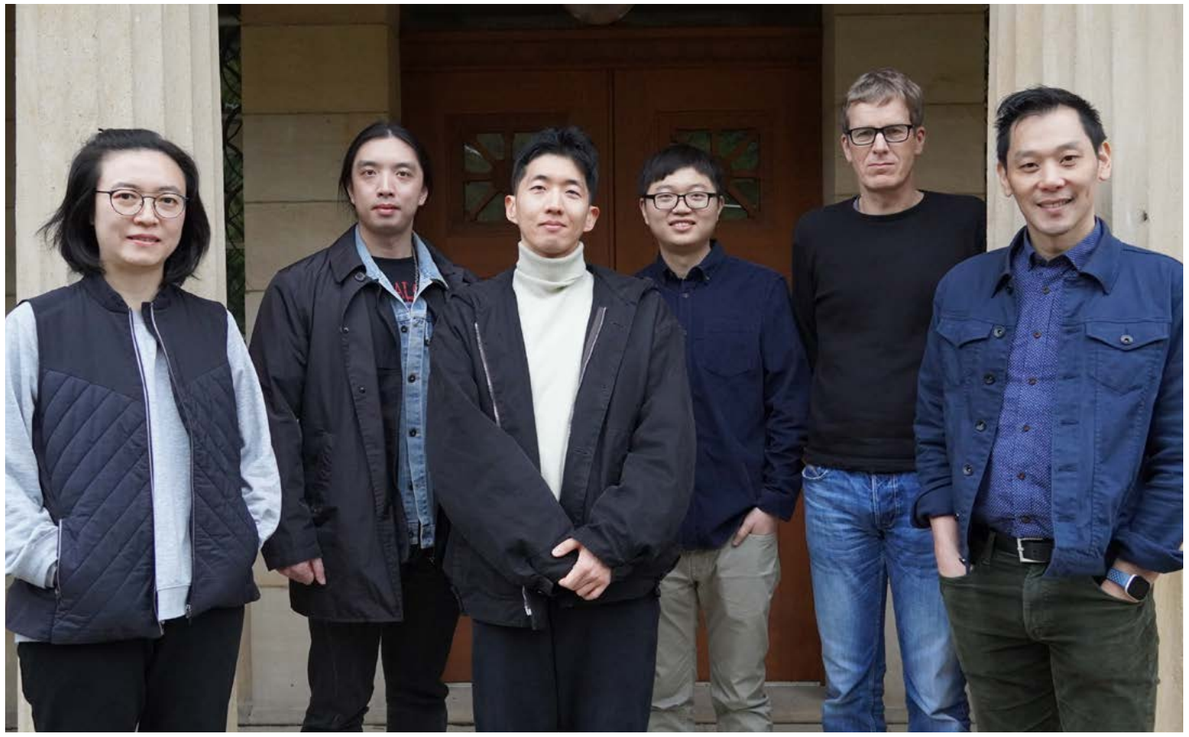
Eines unserer Ziele besteht darin, die Mechanismen besser zu verstehen, die bei Wechselwirkung von Medikamenten auf der molekularen Ebene ablaufen, von der Freisetzung des Wirkstoffs über die Bindung zum Rezeptor bis hin zum Metabolismus des Medikaments. In einem interdisziplinären Ansatz kooperieren wir mit experimentell arbeitenden Forschenden und verwenden gemeinsam rechnerische Methoden aus den Bereichen der Physik-, Bio- und Cheminformatik. Das breite Spektrum der Techniken, die wir entwickeln und einsetzen, reicht dabei von interaktiven web-basierten Visualisierungswerkzeugen bis hin zu Molekularsimulationen auf atomarer Ebene.

In diesem Bericht beschreiben wir einige der Ergebnisse aus dem Jahr 2024. Nach einem allgemeinen Überblick über Neuigkeiten in der Gruppe konzentriert sich der Bericht auf Projekte zu:

i. Untersuchung des Zusammenspiels von struktureller Flexibilität und funktionalen Interaktionen in einem beispielhaften Cytochrom P450 des Medikamentenabbaus durch multiskalare Molekulardynamiksimulationen
ii. Computergestütztes Screening der Auswirkungen von Mutationen auf die Verweildauer und die Dissoziationsmechanismen von Protein-Protein-Komplexen
iii. SARS-CoV-2-Spike-Zellrezeptor-Bindung: Die Auswirkungen von Heparansulfaten, N-Glykanen, Heparin und Heparinanaloga.

2 Research

2.9 Natural Language Processing (NLP)



Group leader

Prof. Dr. Michael Strube

Team

Souvik Banerjee (since July 2024)
Haixia Chai (HITS Scholarship holder; until March 2024)
Yi Fan
Sungho Jeon (HITS Scholarship holder; until March 2024)
Wei Liu
Dr. Stephen Wan (visiting scientist; CSIRO Award; until February 2024)

Natural Language Processing (NLP) is an interdisciplinary research area that lies at the intersection of computer science and linguistics. The NLP group develops methods, algorithms, and tools for automatically analyzing natural language. The group focuses on discourse processing and related applications, such as automatic summarization and readability assessment.

The NLP group hosted Stephen Wan – a team leader and research scientist at CSIRO, Sydney, Australia – as a guest scientist from September 2023 to February 2024. His research centers around information extraction from scientific documents, which aligns nicely with the NLP group's research focus on discourse processing. Stephen worked together with members of the NLP group – in particular, Wei Liu – on discourse relation classification.

Scientifically, 2024 turned out to be a very successful year for the NLP group, with two publications at *ACL conferences and two further publications in prestigious journals on machine learning (JMLR and TMLR). In March 2024, both Haixia Chai and Sungho Jeon submitted their PhD theses. Haixia subsequently took up a lecturer position in the Computer Science Department at the University of Aberdeen, Scotland, and Sungho now works as a research engineer at Samsung Visual Display in South Korea. Mehwish Fatima – who graduated in 2023 – became an assistant professor in the School of Electrical Engineering and Computer Science at the National University of Sciences and Technology in Islamabad, Pakistan. In July 2024, Souvik Banerjee joined the NLP group as a PhD student. He has a master's in computer science from IIIT Hyderabad, India. In August 2024, PhD student Wei Liu joined Amazon Research, Berlin, for an internship. There is also good news from NLP Alumna Yufang Hou (PhD at HITS in 2016): In 2024, she was appointed Full Professor at IT:U University in Linz, Austria. Yufang decided to go back to academia after eight years in industrial research at IBM Research, Dublin, Ireland.

Also last year, Michael Strube was the General Chair of EACL 2024, the European flagship conference in Computational Linguistics and Natural Language Processing, which was held in Malta in March. Michael's tasks ranged from choosing the venue and selecting all sorts of committees and chairs to tasting the food for the social events. In the end, the conference attracted about 800 in-person participants and 400 online participants, all of whom attended the main conference, six tutorials, and about 20 workshops on different aspects of computational linguistics. One of these workshops was the "Fifth Workshop on Computational Approaches to Discourse," which was also co-chaired by Michael Strube.

What causes the failure of using explicit discourse relations to recognize implicit ones? (Liu Wei)

Discourse relations – such as contrast and cause – describe the logical relationship between two text spans (i.e., arguments). They can either be signaled explicitly with connectives, as in (1), or be expressed implicitly, as in (2):

1. [They may feel emotionally secure now]Arg1 because [they are not heavily involved in the stock market.]Arg2
2. [He has not changed, but those around him have.]Arg1 [Many of his views on the protection of wilderness areas are now embraced by the mainstream.]Arg2

Many studies have attempted to use explicit examples to classify implicit relations due to the easy annotation of the explicit corpus. However, these studies have found that classifiers trained on explicit examples perform poorly in real implicit relations. Recent works have focused on enhancing the transfer performance from explicit to implicit discourse relations. However, little attention has been paid to the underlying causes of the poor results. We showed that one cause for the poor transfer performance is the occurrence of label shift during the construction of the implicit-like corpus. In other words, removing connectives from explicit examples affects the discourse relations they originally expressed. To verify this hypothesis, we manually annotated 100 explicit examples with the connective removed and compared the annotated relation with and without a connective. In so doing, we found that 37 out of 100 examples expressed completely different relations after removing the connective, which indicated the occurrence of a label shift in these explicit examples. We further designed an experiment that

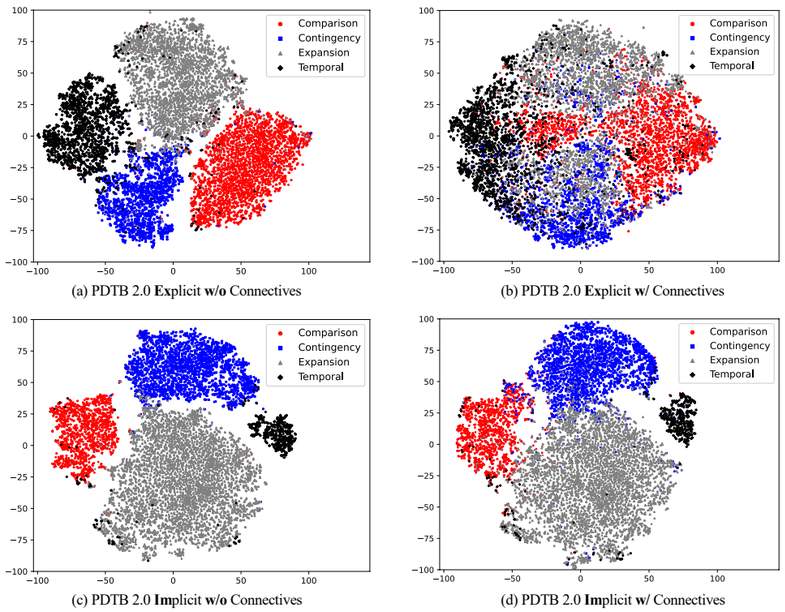


Figure 46: Representation visualization of explicit and implicit examples when containing or not containing a connective.

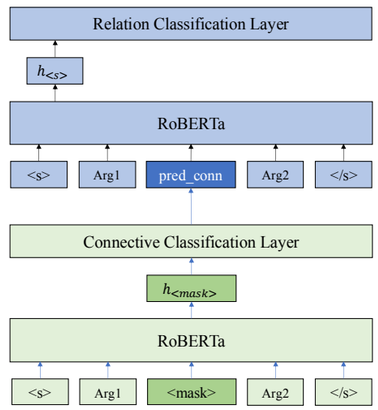


Figure 47: The joint training model for alleviating the label shift issue, which first recovers a connective between two arguments and then predicts the discourse relation.

compared the representations of explicit and implicit examples with and without a connective. Figure 46 (previous page) shows the visualized results. Implicit examples showed generally unchanged results with and without a connective. By contrast, explicit examples underwent significant changes after a connective had been inserted between two arguments. These results demonstrate that a label shift occurs at the corpus level in explicit examples.

We designed two strategies to mitigate the label shift: filtering out noisy data, and joint learning with connectives (see Figure 47, previous page). Experiments on PDTB 2.0, PDTB 3.0, and the GUM dataset demonstrated that classifiers that had been trained with our strategies outperformed strong baselines.

Consistent temporal relation extraction (Yi Fan)

Temporal relations describe the interaction between events along the temporal dimension. The objective of temporal relation extraction is to identify the temporal relations between all event pairs within a given document, thereby constructing a temporal structure graph. Researchers in the field once preferred to use the traditional machine learning approach combined with linguistic features to solve the problem before the widespread use of neural networks such as transformers, but linguistic-based features have been silently discarded in recent research. Moreover, consistency is a crucial aspect that should not be overlooked in this task. In recent years, research efforts have primarily focused on designing models that achieve high F1 scores, often at the expense of addressing consistency. In the extraction of temporal event relations, consistency refers to the logical coherence and non-contradiction among the extracted temporal relationships. Global consistency is crucial as it ensures that the predicted temporal relations are logically sound. Any inconsistencies among the predicted temporal relations impair the system's ability to produce a complete temporal structure of the input document, thereby ultimately leading to errors in the overall understanding of the text (see Figure 48 for an example). Although introducing consistency constraints could lead to a decline in the performance of neural network–based methods, the importance of consistency in producing a complete and accurate temporal graph cannot be ignored.

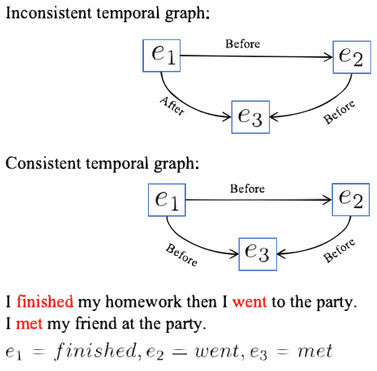


Figure 48: The above example illustrates an inconsistent temporal graph and a consistent temporal graph extracted from the same text. In the text, e1 comes before e2, and e2 comes before e3. Logically, if we know that e1 comes before e2 and that e2 comes before e3, then we can easily infer that e1 comes before e3. If the system produces an inconsistent temporal graph with e1 after e3, this means the system does not understand the input text well.

Inspired by previous work, we used current pre-trained language models combined with linguistic features to generate a linguistically enhanced event representation in order to solve this task. Our approach employed integer linear programming (ILP) to ensure the extraction of a consistent temporal structure from the input text. Our model is shown in Figure 49.

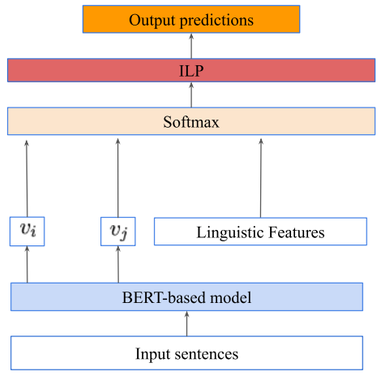


Figure 49: An overview of our model.

We evaluated our method on TDD-Man and TDD-Auto. Our experimental results show that our model can produce a consistent temporal structure while maintaining a high F1 Score.

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

Die NLP Gruppe wurde von September 2023 bis Februar 2024 von Stephen Wan besucht. Stephen ist Team Leader und Research Scientist bei der Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Sydney, Australia. Seine Forschungsinteresse ist Informationsextraktion aus wissenschaftlichen Texten, was die Forschung der NLP Gruppe sehr gut ergänzt. Bei HITS arbeitete Stephen in erster Linie zusammen mit Wei Liu über die Klassifikation von Diskursrelationen.

Das Jahr 2024 war ein wissenschaftlich sehr erfolgreiches Jahr für die NLP Gruppe mit zwei Veröffentlichungen bei *ACL-Konferenzen und zwei Publikationen in angesehenen Zeitschriften im Bereich Maschinelles Lernen (JMLR und TMLR). Im März 2024 reichten sowohl Haixia Chai als auch Sungho Jeon ihre Dissertationen ein und verließen HITS. Haixia Chai ist jetzt Lecturer im Computer Science Department and University of Aberdeen, Schottland. Sungho Jeon arbeitet als Research Engineer bei Samsung Visual Display in Südkorea. Mehwish Fatima, die 2023 promoviert wurde, wurde Assistant Professor and der National University of Sciences and Technology in Islamabad, Pakistan. Im Juli 2024 stieß Souvik Banerjee als neuer Doktorand zu uns. Er hat einen Master-Abschluss in Computer Science vom IIIT Hyderabad, Indien. Im August 2024 begann Wei Liu ein Praktikum bei Amazon Research, Berlin. Es gibt auch gute Nachrichten von NLP Alumna Yufang Hou (PhD 2016 bei HITS). 2024 wurde sie zur Professorin an der neugegründeten IT:U University in Linz, Österreich, berufen, wo sie den Bereich Natural Language Processing vertritt. Nach acht Jahren in der industriellen Forschung bei IBM Research in Dublin, Irland, entschied sie sich, in die akademische Forschung zurückzukehren.

Michael Strube war General Chair der Konferenz EACL 2024, der wichtigsten europäischen Konferenz im Bereich Computerlinguistik und automatische Sprachverarbeitung, die im März 2024 in Malta stattfand. Michael Strubes Aufgaben reichten vom Bestimmen des Konferenzortes und Hotels über die Auswahl zahlreicher Komitees bis hin zum Verkosten von Häppchen für die Social Events. Die Konferenz wurde von 800 Teilnehmern vor Ort besucht. Dazu kamen etwa 400 Online-Teilnehmer. Neben der dreitägigen Hauptkonferenz gab es an zwei weiteren Tagen sechs Tutorials und etwa 20 Workshops zu verschiedensten Themen der Computerlinguistik. Einer dieser Workshops war der “Fifth Workshop on Computational Approaches to Discourse,” der ebenfalls von Michael Strube (ko-)geleitet wurde.

2 Research

2.10 Physics of Stellar Objects (PSO)

Group leader

Prof. Dr. Friedrich Röpke

Team

Veronica Agaeva (visiting scientist; Heidelberg University, since October 2024)
Dr. Róbert Andrássy (visiting scientist; Heidelberg University)
Paul Christians (visiting scientist; GSI Darmstadt)
Priyam Das (visiting scientist; University of New South Wales (UNSW), Canberra, Australia; Oct-Dec 2024)
Lucas Eekhof (student; April–June 2024)
Christina Fakiola (visiting scientist; Heidelberg University; since July 2024)
Prof. Dr. Robert Fisher (visiting scientist; University of Massachusetts, Dartmouth; until July 2024)
Dr. Javier Morán Fraile
Dr. Damien Gagnier (visiting scientist; Heidelberg University; since September 2024)
Leon Görgen (student; May–September 2024)
Alexander Holas (PhD student)
Dr. Mike Lau (Croucher Fellow)
Dr. Giovanni Leidi
Dr. Georgios Lioutas (postdoc; since October 2024)
Dr. Kiril Maltsev (HITS Lab)
Vijayalakshmi Vijayakumaran Nair (student; since August 2024)
Evans Kojo Owusu (visiting scientist; UNSW, Canberra, Australia; August–November 2024)
Dr. Ashley Ruiter (visiting scientist; UNSW, Canberra, Australia; June–December 2024)
Prof. Dr. Ivo Seitenzahl (visiting scientist; UNSW, Canberra, Australia; June–December 2024)
Marco Vetter (visiting scientist; Heidelberg University)
Kristián Vitovský (PhD student; since July 2024)
Freyja Walberg (student)

"We are stardust." Indeed, the very matter we are made of is largely the result of processing the primordial material that formed during the Big Bang, whereas heavier elements originate from nucleosynthesis in stars and in gigantic stellar explosions. Discovering how this material formed and how it is distributed throughout the Universe are fundamental concerns for astrophysicists. At the same time, stellar objects make the Universe accessible to us by way of astronomical observations. Stars shine in optical and other parts of the electromagnetic spectrum and are the fundamental building blocks of galaxies and larger cosmological structures.
With the help of extensive numerical simulations, the Physics of Stellar Objects research group seeks to understand the processes that take place in stars and stellar explosions. Newly developed numerical techniques and the ever-increasing power of supercomputers facilitate the modeling of stellar objects in unprecedented detail and with unparalleled precision.
One of our group's primary goals is to model the thermo-nuclear explosions of white dwarf stars that lead to the astronomical phenomenon known as Type Ia supernovae. These supernovae are the main source of iron in the Universe and have been instrumental as distance indicators in cosmology, which has led to the spectacular discovery of the accelerating expansion of the Universe. Multi-dimensional fluid dynamic simulations in combination with nucleosynthesis calculations and radiative transfer modeling provide a detailed picture of the physical processes that take place in Type Ia supernovae and are also applied in the PSO group to other kinds of cosmic explosions.
Classical astrophysical theory describes stars as one-dimensional objects in hydrostatic equilibrium – an approach that has proven extremely successful and that explains why stars are observed in different configurations while also providing a qualitative understanding of stellar evolution. However, simplifying assumptions limit the predictive power of such models. Using newly developed numerical tools, our group explores dynamic phases in stellar evolution via three-dimensional simulations. Our aim is to construct a new generation of stellar models based on an improved description of the physical processes that take place in stars.

How stars mix their interiors: A new look at convective penetrations

Stars are vast, luminous spheres of hot gas, with their internal structure determining how they evolve, shine, and eventually die. Inside many stars, particularly those larger than about 1.2 times the Sun's mass, the core is not a static ball of nuclear fire; rather, it is a convective region in which hot gas rises and cooler gas sinks, much like in a pot of boiling water. This constant churning helps mix elements inside the star, which affects how long the star will burn and how it will evolve.
For decades, astronomers have noticed something puzzling: Observations show that these convective cores are significantly larger than traditional models predict. Whether measured by observing binary star systems in which one star's light eclipses another or by using asteroseismology (a method that studies sound waves traveling through a star's interior), the evidence is clear: Something is extending the size of these cores beyond theoretical expectations.
Members of the PSO group have explored one possible explanation: convective penetration, which is a process in which convective mixing extends beyond the formally defined boundary and alters the star's structure. The authors use cutting-edge computer simulations to study this effect, thereby refining our understanding of how large convective cores really are (see Figure 50).
The size of a star's convective core is not just an academic detail; rather, it has profound effects on how long a star will live, what elements it will produce, and how it will end its life:

1. Stars with larger convective cores live longer. If mixing extends beyond the expected boundary, fresh hydrogen is brought into the nuclear-burning core, thereby delaying the star's exhaustion of fuel. This means that some massive stars could shine for millions of years longer than expected before they evolve into red giants or supernovae.
2. The extended mixing affects element formation: Elements such as carbon, oxygen, and nitrogen are produced deep in the core and then spread throughout the star due to convective mixing. More mixing causes different ratios of elements to be created and later expelled into space, thereby influencing the chemical makeup of future stars and planets.
3. Massive stars end their lives in spectacular explosions, but the timing and type of supernova depend on how much fuel remains in the core. Understanding convective penetration helps refine our models of when a star will explode and whether it will leave behind a neutron star or a black hole.

Since real stars evolve over millions to billions of years, scientists cannot simply watch one change in real-time; instead, they use computer simulations to model stellar interiors and to speed up time artificially. The study carried out by members of the PSO group focused on stars of 15 times the mass of the Sun at the start of its hydrogen-burning phase, which is called the zero-age main sequence (ZAMS). The group members simulated the behavior of this star's convective core under different conditions using the Seven-League Hydro (SLH) code developed in the PSO group at HITS.
One challenge in simulating stars lies in the fact that the process of convective penetration take place on an extremely long timescale. In order to make the simulations more manageable, the authors artificially increased the star's energy production and reduced its opacity (i.e., how much light it absorbs) by factors ranging from 1,000 to 1,000,000 times. This technique speeds up

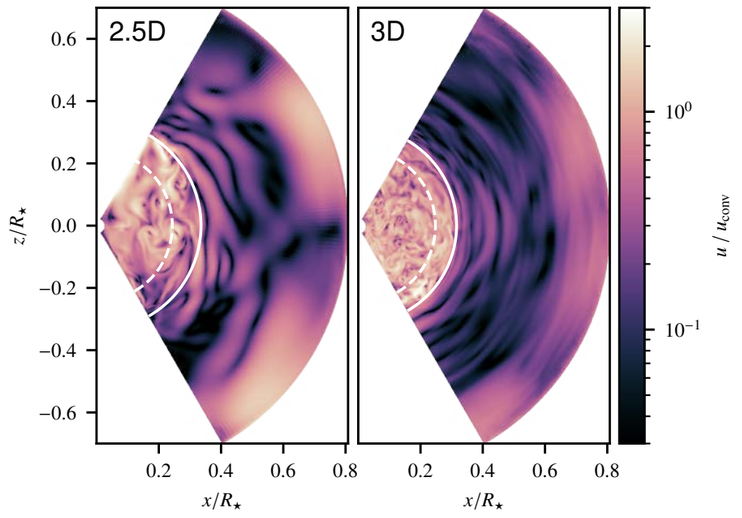


Figure 50: Velocity fields in simulations carried out under the assumption of axial symmetry (left) and in a fully three-dimensional setup (right; here, a slice is shown). The dashed line shows the expected location of the convective boundary according to a stability analysis as used in traditional stellar evolution theory, whereas the solid line marks the actual boundary of the convective core in the simulated star (Figure taken from [Andrassy et al., 2024]).

the star's evolution while preserving its fundamental physics.

The PSO group's study produced several important findings:

1. A well-defined penetration layer forms at the convective boundary: Once the simulation reaches thermal equilibrium, the penetration layer stops growing. As expected, convective mixing does not continue indefinitely.
2. The structure of this layer closely resembles the "step overshoot" model. This model assumes that the penetrative region is fully mixed and follows a temperature profile determined by constant entropy before sharply transitioning to the surrounding stable layers.
3. The penetration layer is only weakly dependent on the artificial energy boost.

The penetration distances found in the PSO group's study are similar to values inferred from real observations of stars. The study confirmed that convective penetration is real and significant in massive stars. This finding extends the mixed core region beyond what classical models predict, thereby helping to explain why observed stars have larger cores than expected. By refining our understanding of stellar interiors, this research improves our ability to predict the lifetimes, chemical yields, and ultimate fates of massive stars, including whether they end up as neutron stars, black holes, or supernovae.

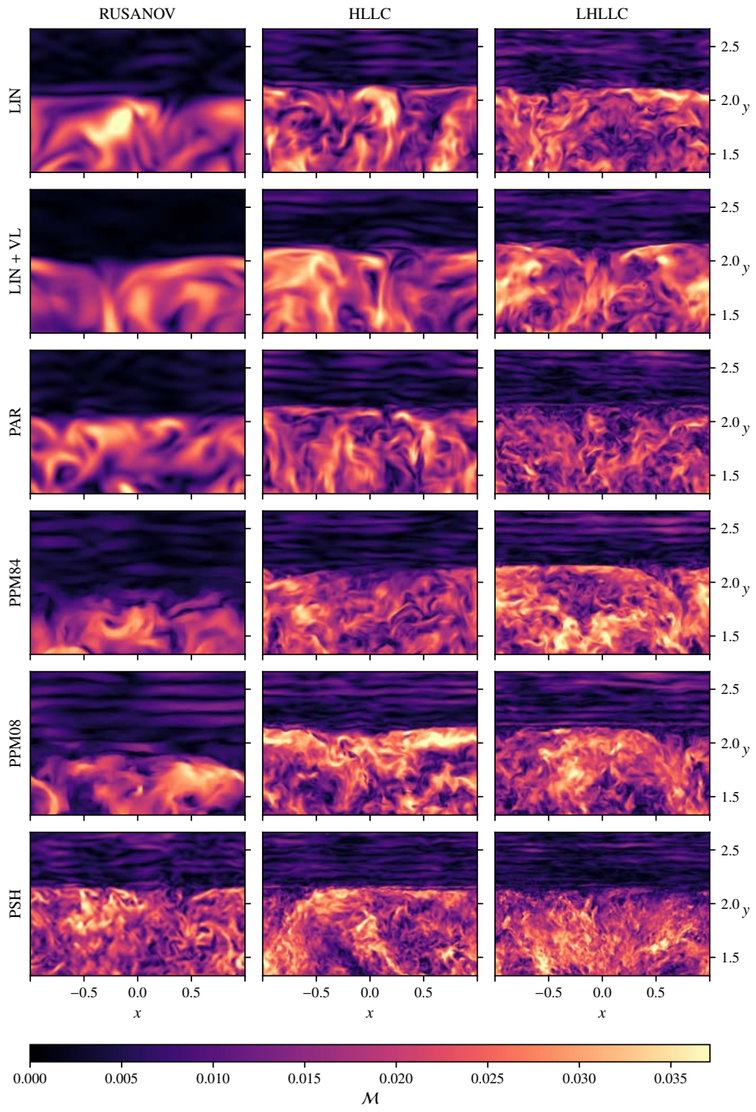


Figure 51: Mach number of flows in simulations of a convectively unstable region underneath a stable layer with different numerical methods. Strong differences in the ability of the applied methods to resolve finer features of the convective flows and are apparent (figure taken from [Leidi et al., 2024]).

Simulating the flow of gas in stars

As exemplified above, understanding how gas moves inside stars is a crucial part of astrophysics. The movement of gas within stars – be it through turbulence, convection, or waves – affects how these stars evolve, shine, and eventually die. In order to study these processes, scientists use computer simulations to model gas flow. However, accurately simulating these movements is a challenge, especially when the flow is very slow compared with the speed of sound, as is often the case in stars.

Members of the PSO group [Leidi et al., 2024] examined Godunov-type methods, which are a family of high-precision numerical techniques commonly used in astrophysics. These methods solve the hydrodynamic equations that describe how fluids move, but different versions of these methods vary in their accuracy and computational efficiency. The goal of the study was thus to determine which numerical methods perform best for slow gas flows, which are common in stars and planetary atmospheres.

Many astrophysical processes involve slow-moving gas. Inside stars, hot gas rises, and cool gas sinks, thereby creating convection currents similar to boiling water. Understanding this process of convection is key to predicting a star's lifetime and energy output. The mixing of different layers in a star determines how elements such as carbon, oxygen, and iron form. Internal gravity waves travel through layers of different densities inside stars and can be observed as periodic oscillations of these stars' brightness. Correctly interpreting these waves allows us to reconstruct the internal structure of stars, which is otherwise inaccessible to direct observation.

Given the importance of these processes, scientists need efficient and accurate methods to simulate them, especially when dealing with low Mach number–flows, in which gas moves at a small fraction of the speed of sound.

The members of the PSO group used the hydrodynamic simulation code called Seven-League Hydro (SLH; developed in the PSO group at HITS) to test 18 different combinations of numerical methods. They evaluated these methods in terms of their accuracy and computational efficiency using two different test cases:

1. Kelvin–Helmholtz instability (2D simulation): This is a well-known fluid instability that occurs when two layers of gas move at different speeds, thereby creating swirling vortices. This test checks how well different methods handle turbulence and small-scale structures.
2. Turbulent convection in a 3D Star: This is a simulation of gas convecting inside a star, which generates internal gravity waves. Simulations using different numerical methods are shown in Figure 51. This test evaluates how well the methods model realistic stellar conditions.

Each method was tested at different resolutions (i.e., grid sizes) in order to see how quickly they converged to a correct solution. The study produced several important insights:

1. Some methods were up to 10,000 times more efficient than others. The best method provided accurate results while also being computationally affordable.
2. Higher-order methods were found to reduce numerical diffusion. In simulations, numerical diffusion can blur fine details. The best methods minimized this issue and were thus able to accurately capture sharp gradients and small-scale turbulence.
3. The best-performing combination was found to be a low-dissipation Riemann solver combined with a high-order reconstruction method. This combination provided the best balance of accuracy and speed.
4. Low-dissipation solvers outperformed standard methods: These solvers introduced less artificial smoothing, thereby making them better suited for slow flows.
5. Some numerical methods introduced artifacts: Slope-limiting methods were found to cause unwanted acoustic waves that are unrealistic and that can affect the accuracy of stellar simulations.

This study provides useful guidance when it comes to determining which numerical methods should be used in future astrophysical simulations. More accurate models of convection in stars can improve our understanding of stellar evolution and structure. In order to create such models, efficient numerical methods are required that allow us to run larger, more complex simulations with available computing power.

„Wir sind Sternenstaub" – die Materie, aus der wir geformt sind, ist zum großen Teil das Ergebnis von Prozessierung des primordialen Materials aus dem Urknall. Alle schwereren Elemente stammen aus der Nukleosynthese in Sternen und gigantischen stellaren Explosionen. Wie dieses Material gebildet wurde und wie es sich im Universum verteilt, stellen für Astrophysiker fundamentale Fragen dar. Sterne sind fundamentale Bausteine von Galaxien und aller größeren kosmologischen Strukturen. Gleichzeitig machen stellare Objekte das Universum für uns in astronomischen Beobachtungen überhaupt erst sichtbar. Sterne scheinen im optischen und anderen Teilen des elektromagnetischen Spektrums. Am Ende ihrer Entwicklung kollabieren massereiche Sterne zu Neutronensternen oder Schwarzen Löcher. Eine Verschmelzung solcher kompakten Objekte wurde kürzlich mit Hilfe von Gravitationswellen beobachtet, die ein neues Fenster für astronomische Beobachtungen des Universums öffnen.

Unsere Forschungsgruppe **Physik stellarer Objekte** strebt mit Hilfe von aufwendigen numerischen Simulationen ein Verständnis der Prozesse in Sternen und stellaren Explosionen an. Neu entwickelte numerische Techniken und die stetig wachsende Leistungsfähigkeit von Supercomputern ermöglichen eine Modellierung stellarer Objekte in bisher nicht erreichtem Detailreichtum und mit großer Genauigkeit.

Die klassische astrophysikalische Theorie beschreibt Sterne als eindimensionale Objekte im hydrostatischen Gleichgewicht. Dieser Ansatz ist extrem erfolgreich. Er erklärt, warum wir Sterne in verschiedenen Konfigurationen beobachten, und liefert ein qualitatives Verständnis der Sternentwicklung. Die hierbei verwendeten vereinfachenden Annahmen schränken jedoch die Vorhersagekraft solcher Modelle stark ein. Mit neu entwickelten numerischen Hilfsmitteln untersucht unsere Gruppe dynamische Phasen der Sternentwicklung in dreidimensionalen Simulationen. Unser Ziel ist es, eine neue Generation von Sternmodellen zu schaffen, die auf einer verbesserten Beschreibung der in ihnen ablaufenden physikalischen Prozesse basiert.

Eine weitere Komplikation, die in klassischen Sternentwicklungsmodellen nur sehr grob angenähert werden kann, ist die Binarität. Wohl wegen des Beispiels unserer Sonne tendieren wir oft dazu, Sterne als isolierte Objekte zu sehen; tatsächlich findet man die meisten von ihnen jedoch in Systemen mit zwei oder sogar mehr Sternen. Einige von diesen wechselwirken miteinander, und das hat weitreichende Auswirkungen auf ihre weitere Entwicklung. Solche Interaktionen sind inhärent mehrdimensional und können in klassischen Modellen nicht konsistent behandelt werden. Die PSO-Gruppe führt dreidimensionale Simulationen zu stellaren Wechselwirkungen durch, um neue Einsichten in diese entscheidenden Phasen der Entwicklung von Sternsystemen zu gewinnen.

Das dritte Forschungsfeld der PSO Gruppe ist die Modellierung von thermonuklearen Explosionen Weißer Zwergsterne, die zum astronomischen Phänomen der Supernovae vom Typ Ia führen. Diese sind die Hauptquelle des Eisens im Universum und wurden als Abstandsindikatoren in der Kosmologie eingesetzt, was zur spektakulären Entdeckung der beschleunigten Expansion des Universums führte. Mehrdimensionale strömungsdynamische Simulationen kombiniert mit Nukleosyntheserechnungen und Modellierung des Strahlungstransports ergeben ein detailliertes Bild der physikalischen Prozesse in Typ Ia Supernovae, werden aber auch auf andere Arten von kosmischen Explosionen angewendet.

2 Research

2.11 Scientific Databases and Visualization (SDBV)

Group leader

PD Dr. Wolfgang Müller

Team

Dr. Haitham Abaza (until June 2024)
Dr. Ina Biermayer (until June 2024)
Elisa Brost (student; from April 2024)
Dr. Susan Eckerle
Arno Fricke (from October 2024)
Martin Golebiewski
Anne Elin Heggland (from February 2024)
Xiaoming Hu
Tobias Martiné (student)
Gerhard Mayer
Dr. Olga Krebs
Lukrécia Mertová
Ghadeer Mobasher
Dr. Maja Rey
Maria Paula Schröder (student)
Fabian Springer (student)
Dr. Andreas Weidemann
Dr. Ulrike Wittig
Qi Wu (February–October 2024)

Making scientific data more useful is the center of our work. Much of this work revolves around structuring what computer scientists call "unstructured data." The goal is to render the data FAIR – that is, Findable, Accessible, Interoperable, Reusable. Most of the time, this entails structuring the data as much as possible and annotating it to ontologies. Structuring data means changing the locations of data items and grouping them such that additional, machine-readable meaning is bestowed on the data.

SABIO-RK – our database for reaction kinetics – is an example of exactly that: Go to sabio.h-its.org, run a query, and then select an entry. A SABIO-RK entry groups information about a reaction and its kinetics under given conditions. The information is grouped such that it can be read much more easily than on paper. The entry can also be downloaded as machine-readable SBML.

The FAIRDOMHub is less stringent in its FAIRness requirements, but it encourages scientists to simplify access to scientific data by opening and structuring their data. The SEEK system that underlies the FAIRDOMHub is also used in many other projects, including in LiSyM-Cancer. We additionally provide support for scientists in actually using the possibilities of SEEK and the connected data management systems that are utilized in the project. In addition, since Phase 1 of the project, we have largely merged the program management and data management teams, with weekly meetings ensuring close collaboration. The team at HITS is led by program director Beat Müllhaupt from Zürich.

LiSyM-Cancer Phase 2 was approved in May 2024, which gives us the opportunity in this annual report to highlight what we have thus far for the project. The second highlight that we present here centers around the BeeProject, which illustrates how curation support inspires research.

The LiSyM-Cancer network

For more than 15 years, the BMBF has been funding projects aimed at demonstrating the benefits of interdisciplinary network research for solving complex challenges in systems biology.

LiSyM-Cancer is an integrative network of molecular and cell biologists, clinical researchers, and experts in mathematical modeling who are jointly conducting research to investigate the development of liver cancer from pre-existing conditions, such as non-alcoholic fatty liver (NAFLD, now called MASLD = metabolic dysfunction-associated steatotic liver disease) and liver cirrhosis. The scientists take an integrative systems medicine approach to improving the early detection and prevention of liver cancer.

The research network builds on over 15 years of systems biology funding in the field of liver research and is based on preliminary work and findings from the related funding programs HepatoSys I + II, Virtual Liver Network (VLN), and Liver Systems Medicine (LiSyM, I+II).

LiSyM-Cancer consists of the three consortia SMART-NAFLD, C-TIP-HCC, and DEEP-HCC as well as the program and data management (PM-DM) team.

In order to achieve the network's goals, we approach its activities using a matrix structure that – in addition to the coordination of the three consortia and data management – includes a combination of clinical, experimental, and modeling activities as overarching topics across the individual consortia.

While each of the three networks has its own coordination team and coordinators at the helm (Ursula Klingmüller, DKFZ; Steven Dooley, Universitätsmedizin Mannheim; and Jochen Hampe, Uniklinikum Dresden), the coordination of overarching activities as well as the communication both within the network and with the outside world is carried out by a common program directorate that is situated at HITS and that is led remotely by Beat Müllhaupt from Zürich. The team consists of the program director, a project manager at HITS (who is simultaneously the C-TIP-HCC project coordinator), a communications manager at HITS, the DEEP-HCC project coordinator in Dresden, and a data management coordinator.

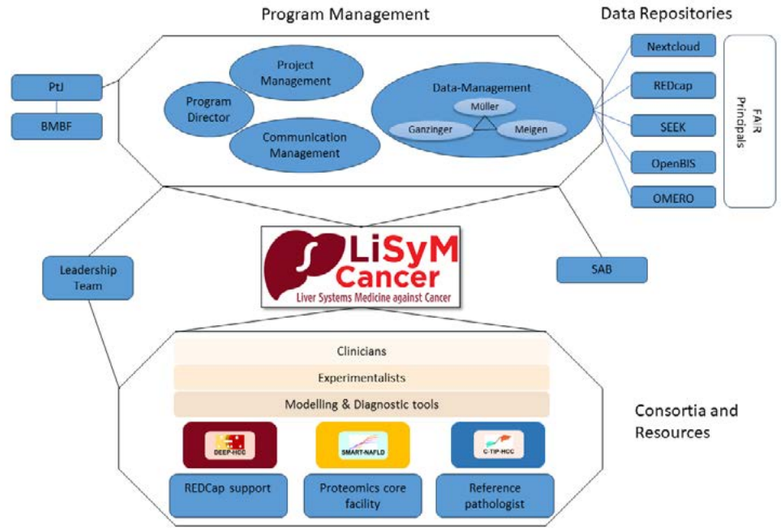


Figure 52: The LiSyM-Cancer Network consists of the three consortia C-TIP-HCC, DEEP-HCC and SMART-NAFLD and the Program Management (PM) with Project, Communications and Data Management (DM) as central unit. The PM is headed by the Program Director and is the primary contact for the PtJ and BMBF, the Scientific Advisory Board and the Leadership Team. The DM runs, improves and services the Data Repositories used in the network by FAIR Principals. Across the consortia, Clinicans, Experimentalists and Modellers cooperate to harmonise protocols and exchange know-how to forward and align standardisation and tool developent. As central resources for the LiSyM-Cancer network, DEEP-HCC hosts the REDCap data repository, SMART-NAFLD provides the proteomics core facility, and the reference pathologist is associated with C-TIP-HCC.

Program and data management are connected at HITS

Program and data management are combined for optimal cooperation at HITS, with the two areas working in close collaboration in order to facilitate the use of data infrastructure across the network.

The main goal of the program management is to create a good basis for cross-network consortia cooperation and to use the existing synergies in the sense of the research mission in a targeted manner. The data management team facilitates and organizes data exchange within the network in accordance with FAIR criteria and in compliance with data protection regulations.

The main objectives of the program and data management include

1. Reaching the LiSyM-Cancer milestones.
2. Successfully communicating progress within and beyond the network.
3. Reporting FAIR data management.

The Ministry of Education and Research (BMBF) sees the importance of coordinating offices such as ours and therefore initiated an exchange between these types of offices at a BMBF-led meeting in Berlin. The first meeting of the coordination offices of various research networks was supervised by the BMBF and facilitated information exchange; thus, it is to be continued in the second funding period as a best practice of sharing and could possibly be extended to other projects with similar challenges.

Facilitating communication within the network

A project such as LiSyM is a collaboration of equals who are each on a journey of discovery. While there are initial, aligned plans, research necessitates frequent changes. The functions of the program director include knowing what happens in the consortium and bringing people together who should talk with one another. In order to fulfill these functions, the program directorate visits project partners, typically grouping visits to partners who are physically close to one another. This activity complements the biweekly leadership team Zoom discussions, the annual status seminar and scientist retreats, and the SBMC conference.

Organizing the SBMC conference

While meetings are held twice per year that are generally internal (i.e., the status seminar and the scientist retreat), the organization of the SBMC conference is different. Being held once every two years, the SBMC conference is an international event that revolves around the systems biology of mammalian cells. Since its foundation, the focus has broadened to the multi-scale systems biology of (mostly) mammals. The organization of the event is a special and challenging task, which we learned first-hand when we undertook the organization of the 9th SBMC conference. This event took place from 13–15 May 2024 in the Paulinum auditorium of Leipzig University. It was organized by the program management team in collaboration with the LiSyM-Cancer leadership team, with a total of 170 international scientists taking part. The budget responsibility for the SBMC conference lay entirely with HITS. This challenge was only able to be successfully met thanks to the excellent and constructive cooperation with the HITS controlling team (Harald Haas, Yashasvini Balachandra) and the support of the whole administration led by Gesa Schönberger. Thanks!

Second funding period

In January/February 2024, the PM-DM team prepared and successfully submitted the applications for the 2nd funding period of LiSyM-Cancer for both the PM-DM team and LiSyM-Cancer as a whole (together with the coordinators of the consortia). With the start of the 2nd funding phase on 1 July 2024, partners Matthias Ganzinger (University Hospital Heidelberg) and Christof Meigen (University Hospital Leipzig) joined the PM-DM team. Ganzinger and Meigen will be responsible for linking the OpenBIS and REDCap databases with SEEK and the LiSyM-Cancer NextCloud at HITS.

In close and constructive cooperation with Iris Bussian (KTA), the PM team assumed the management of the complex contract creation and coordination for the four consortium agreements as well as the joint controller agreement (JCA), thereby bringing 40 institutional partners together. These contracts are the legal foundation for the shared work within the network, which is a necessary precondition for exchanging person-related data and specimens. In addition to facilitating the research as such, the program management is also responsible for providing life science research with the legal and data protection that it requires.

Data management example

To identify biomarkers for the transition – or "tipping point" – from MASLD or liver cirrhosis to hepatocellular carcinoma (HCC), the researchers use various approaches. These incluse 3D modeling of healthy and diseased liver tissue, along with experimental data from the metabolome, proteome, and genome.

At present, no one knows how the shift occurs from (reversible) MASLD – which goes away with an improved diet – to irreversible damage to the liver. Understanding this process is one challenge in LiSyM-Cancer.

Therefore, the clinicians involved in the project provide either samples from these patients collected during surgery or blood samples obtained during patient follow-up visits. In addition to collecting these samples, detailed information on their origin must also be collected in a structured and secure manner since this is highly sensitive data.

The experimentalists and clinicians agreed on the information to be collected and came up with more than 400 data fields that can be filled in on a form in the REDCap system, which is situated in Leipzig. Not only is it time-consuming to collect all of these data and to extract them from various sources, such as doctors' letters and measuring devices, but the data must also be transmitted from clinical information systems (CIS) to the network's REDCap database.

We worked with scientists in Leipzig and Düsseldorf as pilots on a generic piece of software that allows patient data to be uploaded directly from CIS into REDCap using a REDCap import template instead of having to enter it manually via the input mask as before.

The individual scientists in the project use complex Excel templates that are adapted to long-term, local needs in order to manage their patient data. The templates' structure greatly differs from the format of the REDCap import template; therefore, the corresponding attributes have to be mapped to one another, and the associated values have to be converted if necessary (e.g., words from selection lists are always represented as whole numbers; therefore, when choosing, e.g., the gender "w", "m", or "d," the "w" becomes a "1," the "m" a "2," and the "d" a "3"). Carrying out this conversion by hand is complex and error prone. In addition, checking for errors takes just as long as the actual implementation of the format. Systematic errors are also highly expensive to correct – in fact, they are just as expensive to correct as are a multitude of independent small errors. All of these issues make quality control highly difficult and time-consuming.

We address this problem with Excel tools, such as a new – yet still-unnamed – Excel transformation tool in addition other tools, including OpenRefine and ReStoRunT, the latter of which was also developed at HITS. The aim here is to be able to build solutions for the relevant partners as quickly as possible that are both useful and reproducible. Ever-changing requirements make the work challenging and time-consuming, but also worthwhile.

We are therefore always on the search for a set of tools that can adapt to quickly changing requirements without sacrificing accuracy. With our constantly evolving tool, we continuously both upload real data for real use in a real database and improve our concepts regarding which type of Excel tooling provides the FAIRest data with the least effort.

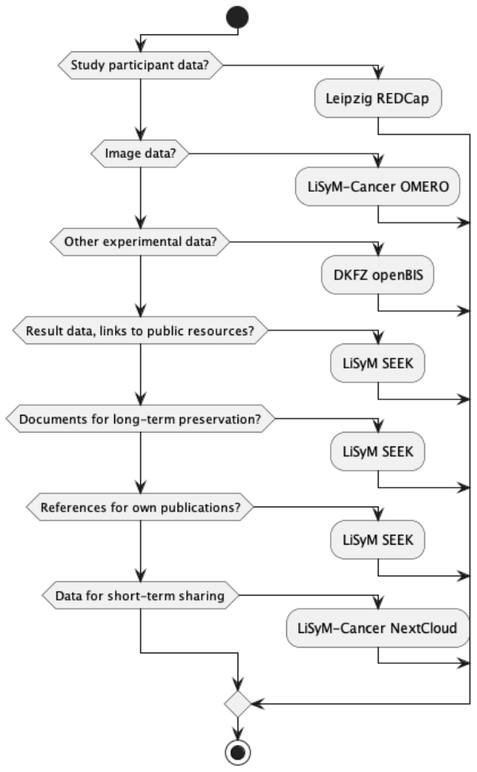


Figure 53: What system is used for what data in LiSyM?

This area of research has been a focus of the SDBV group for quite some time now. Our approaches changed during the last 10 years, as do those of the community, and there is still room for improvement – so stay tuned!

The BeeProject

Modern experimental science benefits from repeatability. The same experiment can be performed with new measurement setups and sensors, yielding refined, expanded data. However, long-term environmental observations are fundamentally different. For instance, the long-term measurements of the evolution of the Earth and the environment are unique and unreproducible. Because we cannot redo e.g. temperature measurements in May 1800, it is crucial to reuse old measurement data, even if they are hard to digitize. Ensuring that these historical records remain accessible requires converting them into machine-readable formats. We therefore developed an algorithm for automated table extraction (structured data) from historical handwritten tables. We applied this approach to a dataset provided by the Julius Kühn Institute (JKI) in collaboration with Oleg Lewkowski and Severin Polreich. The dataset consists of hive-scale measurements collected by beekeeper associations in the German states of Lower Saxony, Hesse, Mecklenburg–Western Pomerania, Thuringia, and Brandenburg. The data contribute to the collaborative research project MonViA, which aims to assess the impact of climate change on bee vitality. These records document daily beehive weight fluctuations alongside meteorological data, with hive weight serving as a proxy for colony size and with weather conditions influencing foraging activity.

The dataset consists of 14,738 handwritten scans, with 3,819 scans having been manually digitized for validation. In order to address the inefficiencies of manual transcription, Lukrécia Mertová at HITS led the software development and designed an automated solution for streamlining the digitization process.

Traditional optical character recognition (OCR) tools fail to handle handwritten tables effectively, frequently ignoring the structured layout of the tables, let alone properly handling the unclarities that occur due to problems in handwriting. This situation therefore required a new solution: Our new modular system integrates state-of-the-art OCR and table extraction techniques that fit the character of a wide range of tabular forms and historical collections. The following observations were key to the development of our method:

1. There are many reasonable solutions for most of what we want to achieve; however, our modular approach helps re-use and isolate the parts that we do ourselves.
2. High-performing OCR models already exist for continuous text; therefore, we structure our method to segment tables into smaller fragments that can be processed as continuous text, thereby using high-quality SOTA.
3. The report tables that we examine adhere to a standard format. We can discover the structure by isolating the form that stays the same across reports.

Our approach follows three key steps: (i) form classification (optional), (ii) table structure recognition, and (iii) table content recognition. Since the dataset is grouped during scanning, classification is performed manually. The structural recognition step exploits the fact that the document format is consistent across many records and isolates recurring table elements from handwritten content. This activity requires many preprocessing steps, including pen elimination (i.e.

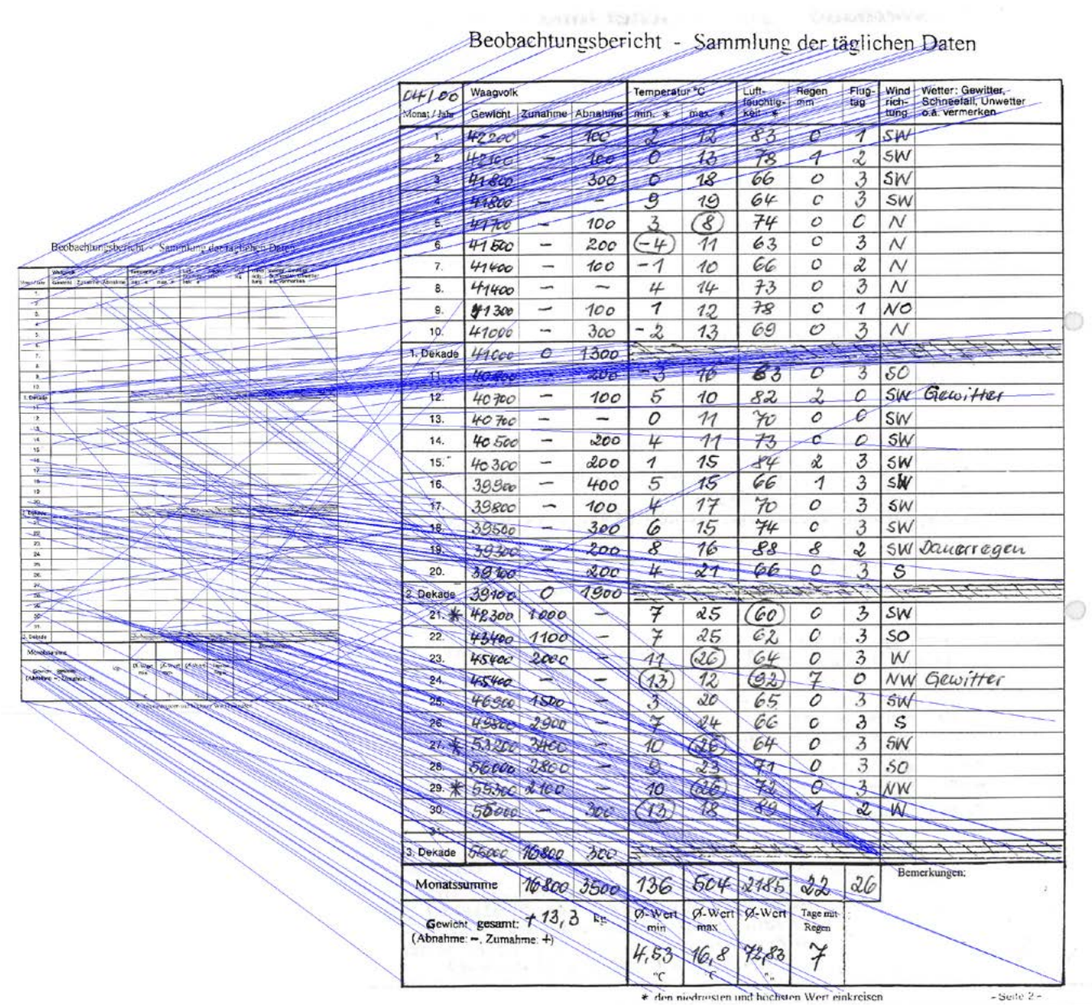


Figure 54: Mapping the template (small document) onto a filled form (big document). The lines show salient points extracted by the SIFT algorithms and how they correspond to each other. Lines crossing each other mean that the matching of salient points is wrong. One can see that most of the points get correctly mapped.

removing the written text) via the classification of pixels), noise elimination, and threshold binarization. Furthermore, we correct alignment errors in scanned documents, which is achieved through scale-invariant feature transform (SIFT) – a method that is used to detect salient points in images.
Salient points are points in which something "happens" in the image – for example, a dent somewhere or a point where lines meet. These points are discovered by determining the texture of images in different scales and directions. We then determine such salient points for two scanned forms and find out how to rotate or incline (i.e., how to affinely transform the image) such that most of the salient points in both images match. By aligning multiple scans to a reference template, we produce an averaged image in which handwritten entries are suppressed, thereby revealing a clean form structure. We can then analyze the lines in order to determine the width of the table grid and thus make each line and column of the scan accessible. This process enables the precise detection of gridlines and column widths, thereby making each cell accessible for content extraction.

Once the table structure has been identified, we use state-of-the-art OCR models for table content recognition. As a part of this processing step, we subtract the template from the new scan, thereby obtaining a clean image that contains only relevant handwritten text that is ready for the OCR. We use Amazon Azure, TrOCR, and Google Vision. Postprocessing stages require aligning and assigning each text position to the cell, finding the best-fitting model, and returning the constructed table to the user.
The evaluation involves comparing OCR output with manually transcribed ground-truth data, which involves a large amount of human work. The correctness of the algorithm is calculated in three steps: alignment, human error, and overall precision. The correctness of the alignment algorithm (SIFT and ORB) is divided into three classes: fully correct, partly correct, and incorrect. Our system achieves 95% fully correct accuracy, with up to 98% near-correct cases depending on the method used.
The degree of human error is evaluated by aligning OCR-extracted data with manually entered records. We developed a correction tool to streamline this process that visually maps extracted text to the corresponding scanned regions, thereby enabling rapid error detection and validation. Based on a random sample evaluation, 98.5% of cells were found to be correctly recognized by humans after a single correction pass, with manual processing taking 5–10 minutes per sheet, including verification.
The content recognition algorithm reached 85% accuracy, which was lower than the 98.5% accuracy obtained through human double-checking, but the work was completed in just 10 seconds per image and without manual intervention. This means that obtaining additional data is much less costly. Scanning is thus the dominant factor. Moreover, domain knowledge and the detection of biologically implausible changes could filter many recognition errors in postprocessing. For example, as the beehive is unlikely to become 20% heavier in a day, surprising changes can be weeded out.
In our work, we developed a comprehensive approach to extracting structured data from historical measurement records, thereby enabling the digitization of handwritten forms with greater efficiency and precision. The extracted data will contribute to biodiversity research by making long-term ecological observations more accessible for analysis.
This work aligns with the NFDI4Biodiversity initiative, which focuses on storing and mobilizing biodiversity research data at an early stage. Additionally, the work is part of our broader efforts to develop tools that support data curation. While this project differs in scope from platforms such as SABIO-RK and SEEK, it addresses a fundamental challenge in field research: Despite advances in digital tools, handwritten data collection remains prevalent. Whether near a beehive or in the middle of a remote field, researchers often rely on pencil and paper to record observations. By bridging this gap, our approach facilitates the transition from analogue to machine-readable formats, thereby ensuring that valuable ecological data can be preserved and used effectively.

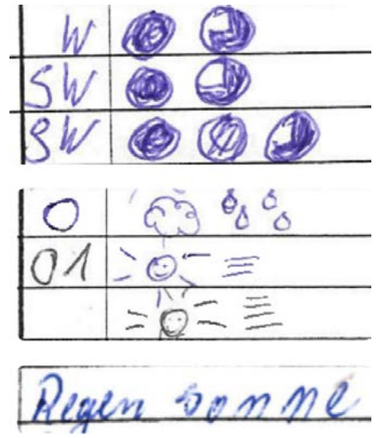


Figure 55: One example of creative form filling. Here: Three ways of writing about the weather.

Das Bestreben, wissenschaftliche Daten nützlicher zu machen, ist das Herzensanliegen der **Scientific Databases and Visualization Gruppe** (SDBV). Ein großer Teil unserer Arbeit befasst sich mit der Strukturierung von so genannten "unstrukturierten Daten", wie Informatiker sie nennen. Wir wollen diese Daten FAIR machen, Findable (auffindbar), Accessible (zugreifbar), Interoperable (vollständig kompatibel), Reusable (wiederverwendbar). Meistens bedeutet dies auch, die Daten so viel wie möglich zu "strukturieren" und mit Ontologien zu verknüpfen. Strukturieren der Daten bedeutet hierbei, die Daten so anzuordnen und zu gruppieren, dass dies die Daten mit maschinenlesbarer Bedeutung ausstattet.

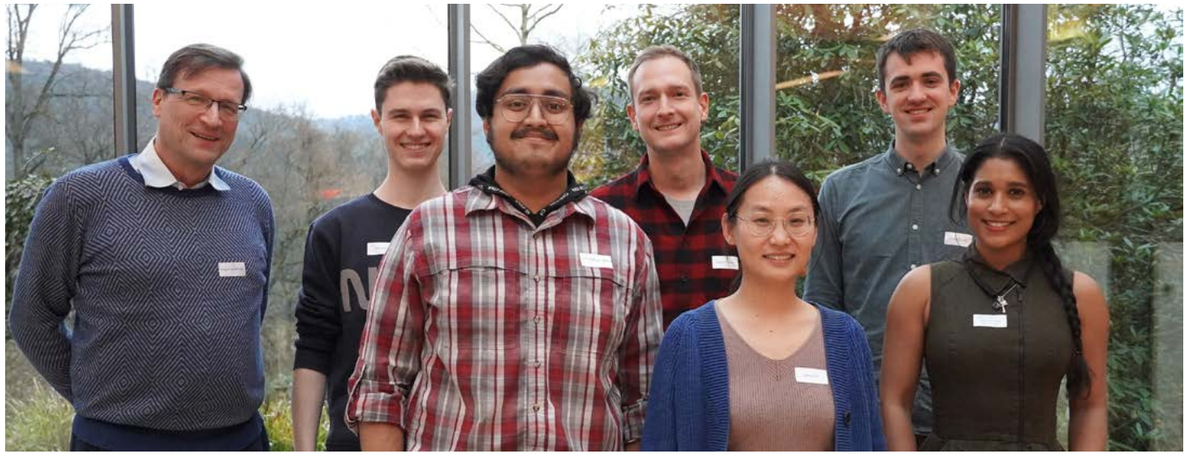
SABIO-RK, unsere Datenbank für Reaktionskinetische Daten, ist ein Beispiel dafür. Gehen Sie zu http://sabiork.h-its.org, starten Sie eine Abfrage, wählen Sie einen so genannten Entry (Eintrag) aus. Die Information ist dort so angeordnet, dass sie sehr viel leichter lesbar ist als eine wissenschaftliche Veröffentlichung. Dieser Eintrag kann auch als maschinenlesbares SBML ausgelesen werden.

Der FAIRDOMHub ist weniger strikt in seinen Anforderungen an FAIRness, aber er ermutigt Forschende dazu, den Zugriff auf wissenschaftliche Daten durch Öffnung und Strukturierung der Daten zu vereinfachen. Das SEEK-System, das dem FAIRDOMHub zugrunde liegt, wird auch in vielen anderen Projekten verwendet. Unter ihnen ist LiSyM-Cancer. Hier bieten wir den Forschenden zusätzlich Unterstützung, die Möglichkeiten von SEEK und der angeschlossenen weiteren Datenmanagement-Systeme zu nutzen. Zusätzlich haben wir seit Phase 1 des Projektes das Programm-Management und das Datenmanagement-Team zusammengelegt. Wöchentliche Meetings sichern enge Zusammenarbeit. Das Team am HITS wird von Programmdirektor Beat Müllhaupt aus Zürich geleitet.

Die Phase 2 von LiSyM-Cancer wurde im Mai 2024 bewilligt. Dies gibt uns die Gelegenheit, im Jahresbericht zu zeigen, was wir für das Projekt tun. Das zweite Highlight, das wir präsentieren, ist das BeeProjekt, das illustriert, wie Kuratierung die Forschung beeinflusst.

2 Research

2.12 Stellar Evolution Theory (SET)



Group leader

Dr. Fabian Schneider (since January 2021)

Team

Utkarsh Basu (student intern; October 2024 – February 2025)
Vincent Bronner
Max Marius Heller (master's student; since August 2023)
Jan Henneco
Dr. Rajika Kuruwita (HITS Independent Postdoc; since October 2022)
Dr. Eva Laplace
Aina Sophie Nambena (student assistant; August–November 2024)
Prof. Dr. Philipp Podsiadlowski (visiting scientist)
Dr. Dandan Wei (visiting scientist)

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

Group News

In 2024, our master's student, Max Heller, successfully completed his thesis on modeling the complex process of merging stars with a one-dimensional code. Max made several important findings that he is currently writing about for publication. In early 2025, he will join the PSO group as a doctoral student, and we wish him all the best in his future endeavors.
We were also very glad to welcome back our former bachelor's student Aina Nambena for a few months last year to extend her bachelor's project with new data received from our colleagues on Tenerife. The new data applied to Aina's Bayesian framework yielded some remarkable results on the birth mass distribution of black holes that we will be following up on in 2025.
Sadly, we also had to say goodbye to Eva Laplace, who was a fantastic colleague and friend. Eva will join KU Leuven as an assistant professor, and we are extremely proud of her.
While several people left our group in 2024, we were also very happy to welcome Utkarsh Basu. Utkarsh began to work on the population modeling of binary stars during an internship and will launch his master's thesis work in our group in 2025.
Lastly, we secured funding from the Deutscher Akademischer Austauschdienst (DAAD, Germany) and the National Institute of Natural Sciences (NINS, Japan) to initiate the "Stellar Bridge" collaboration. Thanks to generous funding over two years, an exchange program will be initiated to "bridge" our theoretical modeling of stars to the supernova stage and thereby connect Heidelberg with Tokyo.

Merger seismology

A considerable fraction of stars are born and live together in pairs. During their lives, these "binary stars" can exchange mass, which can result in the two stars merging into one. These "merger products" can have unusual characteristics and are often used to explain observations that are inexplicable with our theory of how single stars evolve. Despite these peculiar properties, it is often difficult to distinguish merger products from genuine single stars (i.e., stars born alone) based on their surface properties alone. Ignoring the potential merger origin of observed single stars can lead to incorrect interpretations of the observations and can mislead unsuspecting astronomers.
Not only do merger products and genuine single stars have different surface properties, but their interiors are also expected to differ due to the large-scale mixing processes and shock waves that take place during the merger phase. Therefore, we explored whether asteroseismology – that is, the study of stellar interiors through stellar pulsations – can help us distinguish merger products from genuine single stars.
For a variety of reasons and during different phases of their lives, stars pulsate (or oscillate), and these pulsations – if they reach the stellar surface with enough energy – lead to observable deformations of the stellar surface (see Figure 56 for an exaggerated representation of this effect). By observing stars for extended periods of time, we can capture these deformations as well as the brightness variations they cause.
From these observed oscillation "modes," asteroseismologists can infer the internal structure of stars since the modes carry information about the regions within the star where they propagate. This is much like how earthquakes have been used to characterize Earth's interior.
Specifically, we computed models of merger products and genuine single stars that look like blue supergiant stars with the same temperature and luminosity (the star's power output), and we predicted their pulsation modes. These blue supergiants are particularly interesting because the theory of single stellar evolution predicts that they are rare, yet many have been observed. One possible solution to this "blue supergiant problem" is that many of these observed stars may be merger products since they are predicted to retain blue supergiant characteristics for longer than are genuine single stars.
In this prediction, we focused on a particular type of pulsation mode: gravity modes. The propagation of these gravity modes through the stellar interior is facilitated by the buoyancy force (Archimedes' principle), which means that they can only propagate in regions of the star in which energy transport happens via radiation. In regions in which convection transports energy, the buoyancy force drives large-scale turbulent boiling motions. Hence, this force cannot facilitate gravity-mode

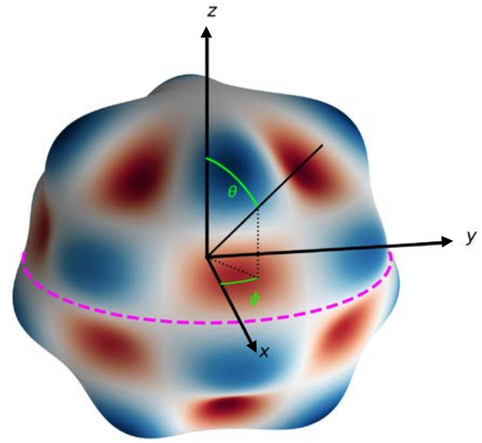


Figure 56: Illustration of the deformation of the spherical stellar surface caused by a stellar pulsation. The blue regions (hills) are moving outward, whereas the red regions (depressions) are moving inward. The white lines indicate the positions with no displacement with respect to the unit sphere (nodal lines). The dashed magenta line shows the equator. The amplitude of the displacement with respect to the unperturbed stellar surface is exaggerated compared with the amplitudes of real stellar oscillations.

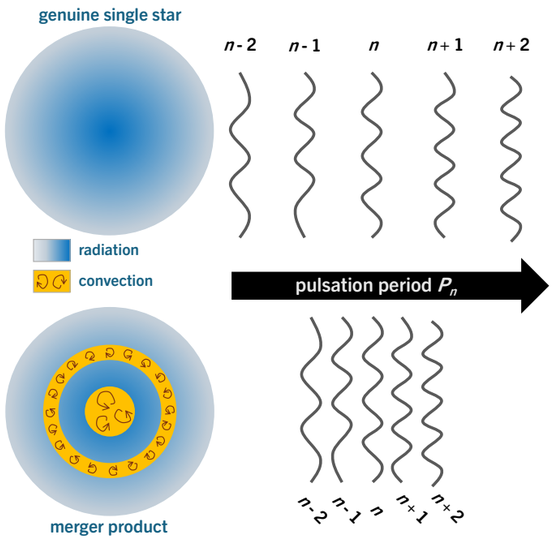


Figure 57: Sketches showing the internal structure of the genuine single star and merger product models (left) as well as the spacings between their gravity-mode pulsation periods P_n with different overtones n (right).

oscillations, and the oscillation amplitudes decay exponentially.
Gravity modes are thus restricted to the radiative regions of the star – that is, to their so-called "mode cavity" – and are sensitive to the conditions there. More specifically, the spacing between the pulsation periods P_n of gravity modes with different overtones n (like the overtones in music theory) depends on the size and shape of the mode cavity. Looking back at the models of the interior structure of the merger product and the genuine single star, we see that they have considerably different interior structures, with the genuine single star being fully radiative and the merger product having two large convective regions (see Figure 57, left). While gravity modes in the genuine single star can propagate in one large mode cavity, the gravity modes in the merger product are confined to two smaller mode cavities (radiative zones). This characteristic is also reflected in the aforementioned period spacings (Fig. 2.12.2, right). We find that the gravity modes of the merger product are spaced almost twice as close as in the genuine single stars. Moreover, we find that in the merger product, some of the gravity modes from the outer mode cavity tunnel through the barrier caused by the convection zone (similar to quantum tunnelling) and interact with the gravity modes in the inner cavity, which causes additional features in the period spacings.
With this proof-of-concept study, we demonstrated that it is in principle possible to distinguish merger products from genuine single stars if we make good observations of these stars' pulsations. This is the first of many steps into "merger seismology," which will be an exciting new field to explore further, especially if we want to better understand stars – both single and binary.

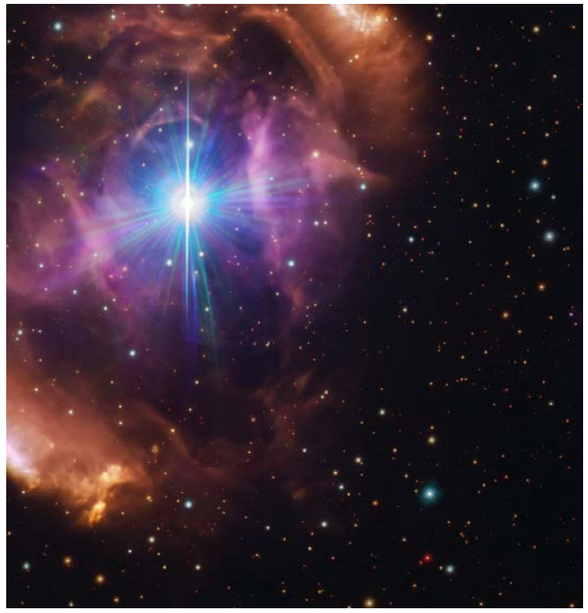


Figure 58: Dragon's Egg nebula NGC 6164/6165. The nebula is a cloud of gas and dust surrounding a pair of stars called HD 148937. This image was taken with the VLT Survey Telescope. Credit: ESO/VPHAS+ team. Acknowledgement: CASU.

A clear view of magnetic stars through a beautiful nebula

Astronomers and astrophysicists are used to thinking in long timescales ranging from millions to billions of years. Sometimes, however, they are surprised to observe phenomena that occurred only several thousand years ago, which is just the blink of an eye in the evolution of stars! This rare chance came when we spotted a pair of stars at the heart of the Dragon's Egg nebula – a stunning cloud of gas and dust (Figure 58). One star appears younger and – unlike the other – is magnetic. Moreover, the nebula is only 7,500 years old and contains large amounts of nitrogen, carbon, and oxygen, which is a telltale sign of a recent dramatic event.
This very special system known as HD 148937 is located about 3,800 light-years away from Earth in the direction of the Norma constellation. In order to unravel the mystery, we assembled nine years' worth of data from instruments on ESO's Very Large Telescope Interferometer (VLTI) – which is located in Chile's Atacama Desert – and the FEROS instrument located at ESO's La Silla Observatory. After a detailed analysis, we were able to demonstrate that the more massive star appears much younger than its companion, which doesn't make any sense since both stars should have formed at the same time. The age difference — with one star appearing to be at least 1.5 million years younger than the other — suggests that something must have rejuvenated the more massive star. The hypothesis is that there were originally three stars in the system. The two inner stars may have merged in a violent manner, thereby creating a magnetic star and throwing out some material, which created the nebula. The more distant, tertiary star likely formed a new orbit with the newly merged, now-magnetic star, thereby creating the binary we see today at the center of the nebula.
The study helped solve a long-standing mystery in astronomy: How do some ~10% of massive stars get their strong surface magnetic fields? While magnetic fields are a common feature of low-mass stars, such as our Sun, more massive stars cannot sustain magnetic fields in the same way. However, some massive stars are indeed magnetic.

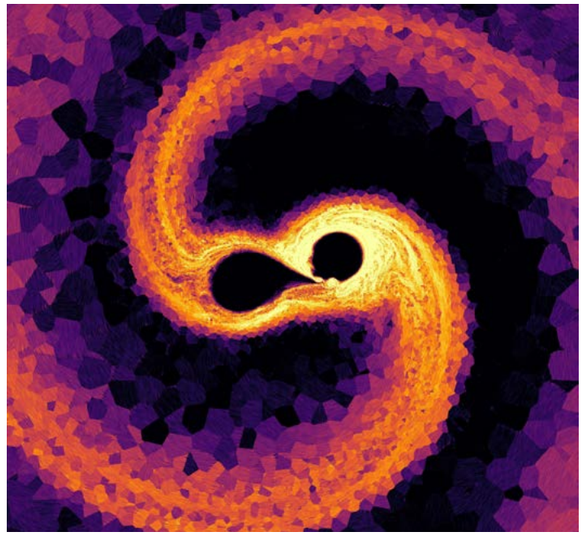


Figure 59: "The birth of a magnetic star." The simulation marks the birth of a magnetic star. The image is a cut through the orbital plane, with the coloring indicating the strength of the magnetic field and with the hatching representing its field lines (picture: Ohlmann/ Schneider/Röpke).

Strong magnetic fields have been suggested to possibly be produced when two stars collide. In 2019, we indeed succeeded in testing this hypothesis (Figure 59). Using a novel simulation code on a large compute cluster, we were able to simulate the merger of two massive stars, and we published our findings in the journal Nature. The simulations showed that the generated magnetic fields might even be sufficient to explain the exceptionally strong magnetic fields that are inferred to exist in magnetars. In the collision of two stars, merger debris is scattered around the merger remnant and should form a nebula that is visible for a short time. Such a nebula around an apparently overly young and magnetic star is thus a smoking-gun signal for the merger hypothesis – and it seems that such a system has finally been found.

Unveiling the rhythms of dying stars: How pulsations shape supernova light curves

Massive stars live for millions of years, but their lives end in a dramatic explosion called a supernova. This explosion happens when a star runs out of nuclear fuel and its fusion can no longer support its own weight, thereby causing its core to collapse while its outer layers are blasted into space. One common type of supernova is called a Type II supernova and forms from a red supergiant, which is a huge star that is hundreds to one thousand times larger than our Sun and that glows with a reddish color. Betelgeuse – a well-known star in the constellation Orion – is an example of a red supergiant that will one day explode. Astronomers have long noticed that not all Type II supernovae look the same. Some fade steadily over time, whereas others stay at a constant brightness for months before dimming. Scientists have struggled to understand why this happens. One possibility is that some stars may lose a lot of material before they explode, thereby changing the way the supernova looks. However, our new study suggests a different explanation: The star's own natural pulsations before it explodes can determine what kind of supernova we see.
Red supergiants are not stable in their final years; instead, they expand and contract in a slow, periodic pattern. Our study (Bronner et al., in prep.) used detailed computer models to simulate a red supergiant with initially 15 times the mass of the Sun in its last stages before exploding. We found that the star pulsated with a cycle of about 800 days because such red supergiants are extremely large and luminous, thereby making their envelopes dynamically unstable to pulsations. The process is amplified by the periodic release of energy during the recombination of hydrogen and helium — that is, the process that occurs when electrons reunite with ions to form neutral atoms — within the star's envelope (Figure 60, next page).
The key discovery in this study is that the timing of the explosion within this pulsation cycle has a huge impact on the appearance of the supernova (see Figure 61, next page). If the star explodes while it is small and compact, the supernova remains at a constant brightness for months before fading, which is a pattern known as a Type II-P supernova (where the "P" stands for "plateau," which refers to the steady brightness period). Howev-

er, if the star explodes while it is large and expanded, the supernova fades steadily, thereby producing what astronomers call a Type II-L supernova (where the "L" stands for "linear," which means a steady decline in brightness). This means that Type II-P and Type II-L supernovae may not be fundamentally different and may instead simply be the result of whether the red supergiant exploded in a compact or expanded state.

This observation might help to explain why supernovae that come from similar types of stars can look so different. This also means that when astronomers observe red supergiants before they explode, their brightness and temperature may not be reliable indicators of their true size or stage of evolution because these red supergiants might have been caught at different points in their pulsation cycle. Finally, this research suggests that red supergiants do not need to lose large amounts of mass before exploding to create a Type II-L supernova, which simplifies our understanding of how these cosmic events unfold.

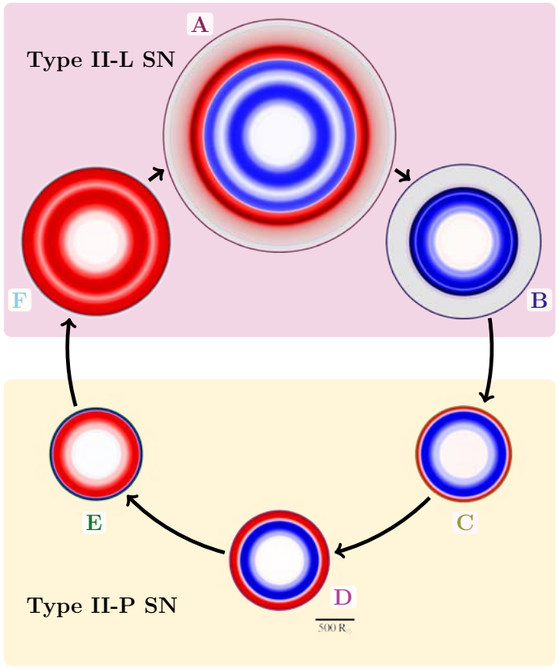


Figure 60: One full pulsation cycle of the red supergiant (to scale). Each state shows a slice through the star, with the color indicating the recombination (red) and ionization (blue) of hydrogen and helium. The star expands and contracts as it pulsates. During the expansion, the star cools down, and recombination energy is released. During the contraction, the star heats up and ionizes hydrogen and helium.

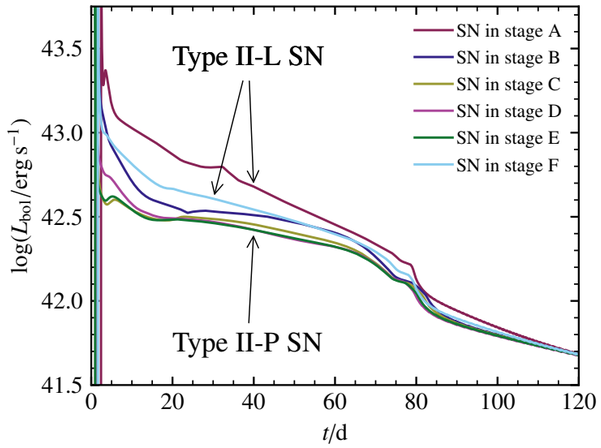


Figure 61: Comparison of the light curves from supernovae with different explosion timings. The Type II-L supernovae show a linear decline in brightness (bolometric luminosity, Lbol), whereas the Type II-P supernovae show a plateau phase followed by a decline but are generally less luminous than Type II-L supernovae.

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

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

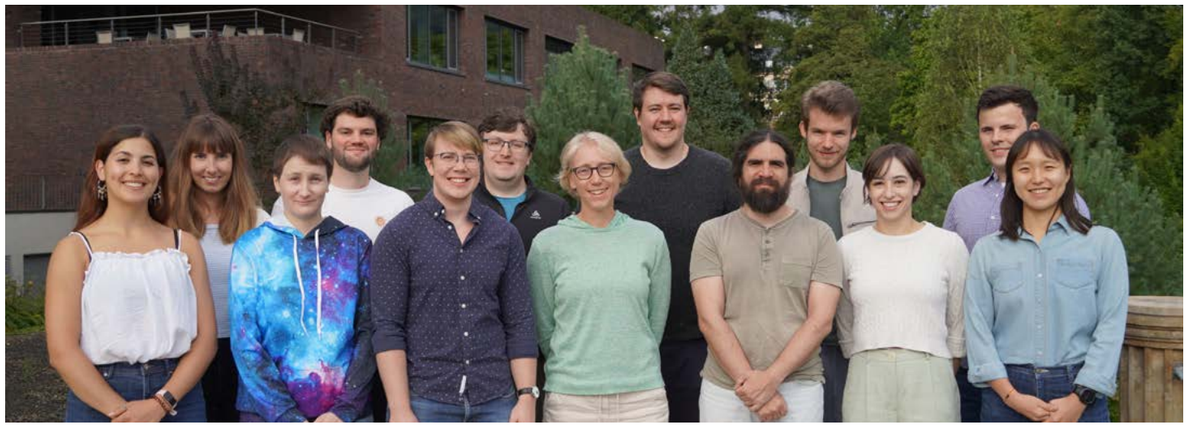
Am Ende ihres Lebens hinterlassen massereiche Sterne einige der exotischsten Formen von Materie: Neutronensterne und Schwarze Löcher. Die Untersuchung dieser Überbleibsel ermöglicht Einblicke in Materieformen, die so auf der Erde nicht verfügbar sind. Die Verschmelzungen von Neutronensternen und Schwarzen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

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

Die **Stellar-Evolution-Theory (SET)** Gruppe untersucht das turbulente und explosive Leben massereicher Sterne. Derzeit konzentriert sich die Gruppe auf massereiche Doppelsternsysteme, deren Verschmelzungsprozesse und die Frage, welche Sterne als schwarze Löcher enden. Sternverschmelzungen erzeugen starke Magnetfelder und können zu stark magnetisierten Neutronensternen führen. Diese als Magnetare bekannten magnetischen Neutronensterne sind die stärksten Magnete im Universum.

2 Research

2.13 Theory and Observations of Stars (TOS)



Group leader

Prof.dr.ir. Saskia Hekker

Team

Dr. Felix Ahlborn (Postdoc)
Dr. Michaël Bazot (Staff Scientist)
Beatriz Bordadagua (PhD student)
Lynn Buchele (PhD student)
Jeong Yun Choi (PhD student)
Quentin Coppèe (PhD student)
Lucas Eekhof (MSc student; since April 2024)
Francisca Espinoza-Rojas (PhD student)
Yann Girodo (MSc student; November and December 2024)
Jan Henneco (PhD student; partly in TOS and partly in SET)
Anastasiya Kapinskaya (MSc student; until September 2024)
Jonas Müller (PhD student)
Dr. Anthony Noll (Postdoc; until February 2024)
Nicholas Proietti (MSc student; June, July, and August 2024)
Tobias van Lier (MSc student; since October 2024)
Kristin Weidner (assistant; since June 2024)

Stars are an important source of electromagnetic radiation in the Universe that allow for studies of many phenomena ranging from distant galaxies to the interstellar medium and 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). However, 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 in a method known as "asteroseismology."

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

Understanding the physical processes 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. We focus on – but do not limit ourselves to – low-mass main-sequence stars, subgiants, and red giants. These stars are interesting as they go through a series of internal structure changes. Furthermore, they are potential hosts of planets and serve as standard candles for Galactic studies (core helium burning red-giant stars); hence, exoplanet studies as well as Galactic archaeology also benefit from an increased understanding of these stars.

Background

In the TOS group, we focus on stars with oscillations similar to those 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. The oscillations are sound waves that are expected to be present in all stars with convective outer layers. A convective envelope is typically present in low-mass main-sequence stars, subgiants, and red-giant stars with surface temperatures below ~6,700 K.

The stellar structure is imprinted in the global oscillation modes of a star. An oscillation mode is uniquely determined by the properties of the matter through which it travels and is described by its frequency (or period) and mode identification – that is, by its radial order (the number of nodal lines in the radial direction), spherical degree (the number of nodal lines on the surface), and azimuthal order (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 (spherical degree of 1) modes have sensitivity both to the deep interior and to the outer layers – that is, the oscillations resonate in an inner (gravity) and an outer (acoustic) cavity separated by an evanescent zone (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, thereby providing information on the physical conditions in the evanescent region. Furthermore, the difference in period between pure gravity dipole modes with consecutive radial orders (so-called period spacing, which can be extracted from mixed dipole modes) provides a measure of the extent of the gravity mode cavity and thus of the properties of the stellar core. Determining these values and understanding the physical processes in these deep parts of stars is one of the aims of the TOS group.

Scientific highlights

Internal rotation rates of red-giant stars
As described above, red-giant stars are the evolved counterparts of our Sun. When the hydrogen in the center of the star that powers the luminosity over the largest fraction of its life is exhausted, the central region contracts while the outer layers start to expand. As a response to these changes, the core spins faster, and the envelope slows down. The governing principle – known as angular momentum conservation – can be illustrated with the well-known example of the figure skater who pulls in or extends their arms and thereby changes their rotation speed.

However, multiple asteroseismic observations have shown that contrary to the above-described picture, the cores of red giants do not spin as fast as would be expected from angular momentum conservation. This finding implies that angular momentum from the core is transported to other layers of the star. To date, several mechanisms have been proposed for achieving this angular momentum transport, though no study has yet reached a conclusion. In the work "The robustness of inferred envelope and core rotation rates of red giant stars from asteroseismology," which was published in the journal Astronomy and Astrophysics, we aimed to increase the accuracy as well as to understand the systematics of the estimated internal rotation rates of red giants. In

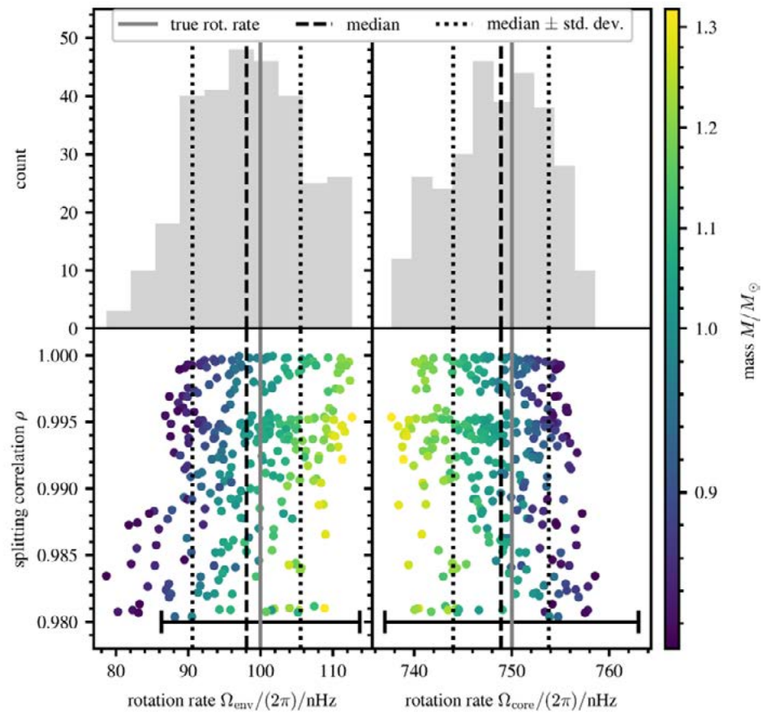


Figure 62: Estimated core and envelope rotation rates for the ensemble of reference models together with the associated correlation coefficients.

order to characterize the rotation profiles in red-giant stars, we used a method called rotational inversion, which allows us to reconstruct the rotation rate in different layers of a red-giant star. Due to the presence of oscillation modes in red-giant stars that are sensitive to both the core and the envelope of the star, we can measure an average core and envelope rotation rate.

The main idea of rotational inversions is to use the measurements of several non-radial oscillation modes and linearly combine them into an estimate of the internal rotation rates. In the presence of rotation, individual non-radial oscillation modes are observed to have multiple frequencies as compared with a single frequency in the non-rotating case. The differences between the frequencies are called "rotational splittings." These rotational splittings originate from an integrated effect of the rotation rates in many different layers of the star. In order to use this integrated signal and make the above-mentioned linear combination, a theoretical model of the stellar interior is needed. Due to the comparably fast evolution of red-giant stars, the construction of an accurate stellar model is a highly challenging task. In this context, we explore which properties of the observed star need to be reproduced with high accuracy in order to reconstruct internal rotation rates, and which aspects might be neglected.

In the paper "The robustness of inferred envelope and core rotation rates of red-giant stars from asteroseismology," we suggest a novel approach to selecting stellar reference models. The cores of red giants have already been well established to rotate faster than the envelopes, which means that oscillation modes that are more sensitive to the core have larger rotational splittings than modes that are more sensitive to the envelope. Stellar models allow us to determine which layers the modes are most sensitive to. In order to select a reference stellar model, we combined these two aspects by searching for models in which the sensitivity of the oscillation modes most closely reproduced the values of the rotational splittings. This was done by computing a correlation coefficient between the observed rotational splittings and the sensitivities of the stellar models. Furthermore, we demonstrated that it is not necessary to identify a single reference model that perfectly reproduces all the star's properties. Instead, a set of reference models that is selected based on the aforementioned correlation can – on average – accurately reconstruct the underlying rotation rates. We also showed that the errors introduced by a mismatch between the observed star and the reference model's structure are comparable with – or smaller than – the random uncertainties caused by measurement errors in the observed frequencies (see Figure 62, previous page).

Our results additionally suggest that other fundamental parameters – such as the mass or the composition of the star – do not play a central role in recovering the underlying rotation rates. In the future, we will be able to use this method to determine accurate rotation rates of red-giant stars in different stages along the red-giant evolution, which will enable us to put better constraints on the angular momentum transport processes.

Additional damping in suppressed dipole-mode stars

Within the framework of the ERC consolidator grant "Dipolar Sound," we investigated the seismic properties of red giants that show less power in their dipole modes (i.e., the suppressed dipole-mode stars). Other than the reduced power, these stars have very similar global surface properties (e.g., temperature, metallicity, mass, radius). In order to identify stars with suppressed dipole modes, we compared the total energy in the dipole modes with that of the radial modes. Typical red giants have about 50% more energy in their dipole modes than in their radial modes. The stars with suppressed dipole modes, on the other hand, have significantly less energy in their dipole modes compared with in their radial modes (i.e., 0–90% of the energy of the radial modes; see Figure 63). As the energy in the modes is set by the balance between the excitation and damping processes that the modes go through, the observed mode suppression is the result of either an inhibited excitation or an additional damping source (which results in a decrease in the total energy of the mode). We can probe the damping processes through the width of the modes in the frequency power density spectrum, and we can probe the total mode energy through the mode amplitude. As these properties cannot be observed in a consistent way for the dipole modes, we decided to investigate the mode energy and damping rates of radial modes in order to link the suppression to the lack of excitation or additional damping. We determined these mode properties using our in-house code TACO (Tools for the Automated Characterisation of Oscillations).

In order to detect any changes pertaining to the effect of the suppression, we selected a control sample of stars with similar global properties as the suppressed dipole-mode stars and with typical power in their dipole modes. Upon comparing the amplitudes and widths of the radial modes of the stars in both samples, we found that the suppression does not significantly affect the mode properties of the radial modes. In other words, the mode energy and damping rate of the radial modes are left unaffected by the suppression. Since the mode energy

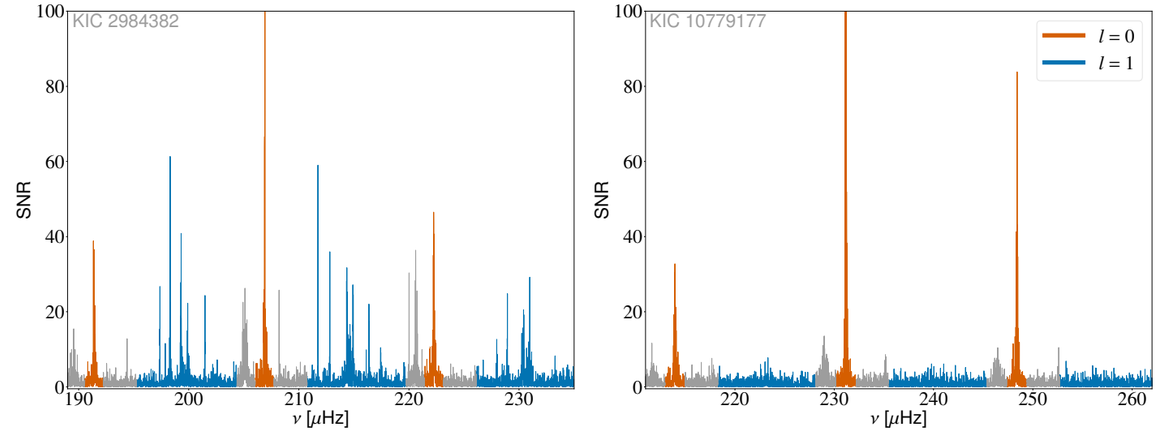


Figure 63: Frequency spectra of a star with typical dipole modes (left) and of a suppressed dipole-mode star (right). The expected frequency ranges of the radial and dipole modes are highlighted in orange and blue, respectively.

is set by the balance of excitation and the damping processes, we inferred from our results that the excitation of the radial modes is also unaffected. In red giants, we expect the excitation processes (i.e., turbulent motion due to convection) to be similar for all modes (independent of spherical degree). We therefore expect the excitation of the dipole modes to not be significantly affected by suppression. This also means that additional damping is causing the observed mode suppression.

Our results for the radial modes and our conclusions may seem paradoxical because the radial-mode damping was unaffected by the suppression, yet we deduced that the suppression is in fact additional damping. However, the radial modes were mostly sensitive to the conditions of the outer layers of the star, whereas the dipole modes – with a mixed character – are sensitive to both the inner and the outer layers. We furthermore concluded that the additional damping mainly dominates the inner regions of the star.

This work has been published in Astronomy & Astrophysics (see [Coppèe et al, 2024]).

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

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

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

2.14 HITS Independent Postdoc Research

HITS Independent Postdocs
Dr. Rajika Kuruwita
Dr. Fabian Grünewald

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

The fellowship is awarded for two years, with an option for a one-year extension following a positive evaluation. It offers a vibrant research community as well as a highly interdisciplinary and international working environment with close links to HITS shareholders Heidelberg University and the Karlsruhe Institute of Technology (KIT). In addition, successful candidates benefit from outstanding computing resources and the various courses offered at HITS.

Astrophysicist Rajika Kuruwita is the first HITS Independent Postdoc. Born in Sri Lanka, Rajika completed her PhD at the Australian National University, was an EU Interactions Fellow at the University of Copenhagen, and joined HITS in September 2022. She collaborates closely with the SET group (see Chapter 2.12).

The second HITS Independent Postdoc is Fabian Grünewald. Born in Germany, Fabian studied chemistry at the University of Groningen, the Netherlands, and received his PhD with honors in physical/computational chemistry from the University of Groningen. He joined HITS in October 2023 and collaborates closely with both the MBM group (see Chapter 2.7) and the MLI group (see Chapter 2.6).

Rajika Kuruwita: The Formation of Multiple Star Systems

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

After the last annual report, Rajika published her work on the relationship between binary/multiple stars and the formation of fast rotators (see [Kuruwita et al., 2024b].). Fast rotators are a class of low-mass stars that have rotation periods of less than 2 days (for context, our Sun has a rotation period of 25 days), and many of these stars are found in binary star systems. We found that the conditions that can cause stars to rotate faster also likely lead to the formation of another star, thereby supporting this observed relationship among fast rotators, which tend to exist in binary star systems (see Figure 64). This work was supported by the internal Isabel Rojas Travel Award, which enabled Rajika to collaborate on her work with Christoph Federrath at the Australian National University.

In 2024, Rajika also continued to support her master's students at the University of Copenhagen in publishing the results of their theses. These students worked on an ambitious project using (1) zoom-in simulations to quantify how protoplanetary disc sizes and masses evolve in binary star systems, (2) radiative transfer to produce synthetic observations, and (3) machine learning to match observations of young binary star systems to synthetic observations. This work was published in a series of two papers (see [Tuhtan et al., 2024], [Al-Belmpeisi et al., 2024]).

Rajika's expertise in binary and multiple star formation has been recognised through invitations to give a review talk at the Annual Scientific Meeting of the European Astronomical Society as well as through invited colloquia at Uppsala University, the University of Stockholm, and the University of Duisburg-Essen. Rajika was also invited to write the Star Formation chapter for Elsevier's Encyclopedia for Astrophysics (see [Kuruwita et al., 2024b]). She has additionally continued to contribute her expertise to observers' studying binary and multiple star formation(see [Li et al, 2024], [Evans et al., 2024]).

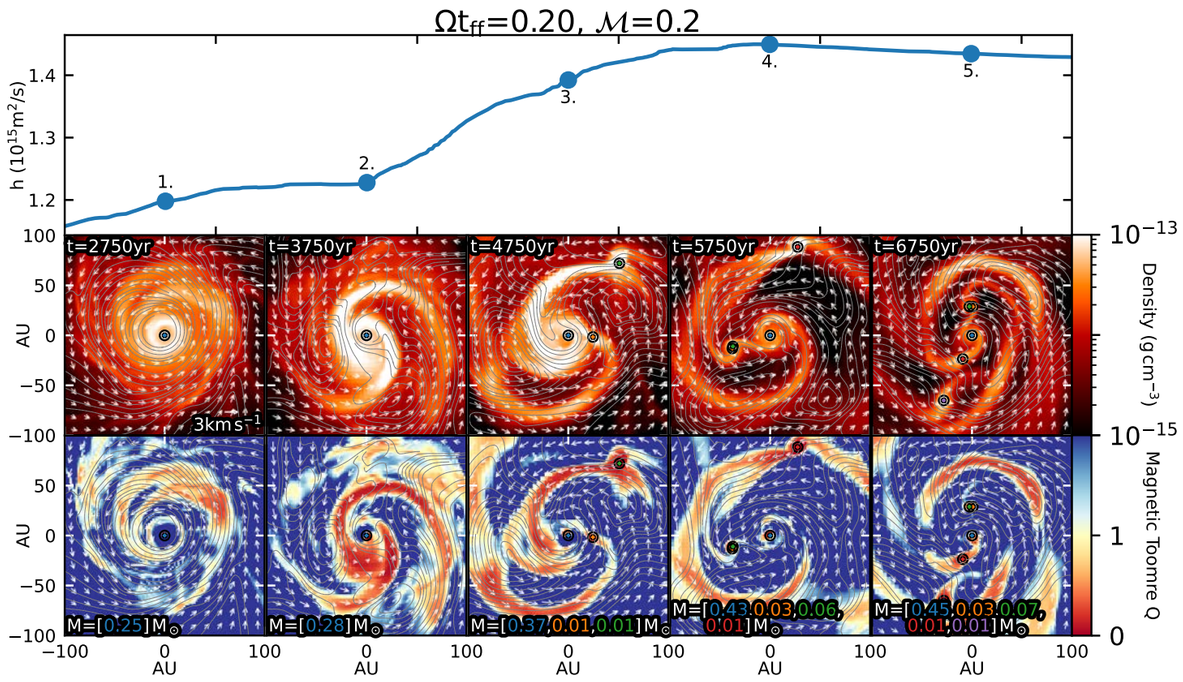


Figure 64: Top: Spin of the primary star over the course of a strong spin-up event. The points indicate the times when the images below were taken. Middle: Projections of the gas density in the disc around the primary star. Bottom: Projections of the magnetic Toomre Q parameter, which is a measure of stability. Blue shows stable regions, and red shows unstable regions. We see that an instability grows in the disc, which triggers the spin-up event in the primary star. It is from this instability that stellar companions form.

Fabian Grünewald: Accelerating the Design of Therapeutics Through Simulations

Fabian Grünewald develops and applies simulation methods in order to decipher structure–function relationships in biomaterials that push the boundaries of (bio)-polymer design at the interface of material science and nanomedicine. In 2024, Fabian collaborated with experimentalists from Prof. Maglia's group on stabilizing nanopore proteins in synthetic polymer and hybrid polymer-lipid membranes for potential applications in medical sensing. While nanopores naturally occur in lipid membranes, these membranes lack the stability needed for medical applications, particularly when they are exposed to blood. This work demonstrated that polymer membranes offer greater stability but compromise sensing capabilities. However, the incorporation of lipids to form hybrid membranes leads to the recovery of the sensing capabilities. The simulations revealed that the proteins in hybrid membranes are located in lipid islands that mimic their native environments, while the surrounding polymer matrix enhances overall stability. Inspired by these findings, the collaboration continues to optimize nanopores in polymer membranes and to investigate the generalizability of these observations (see Figure 65).

Furthermore, Fabian continued the development of high-throughput coarse-grained simulation methods for studying therapeutics. In particular, he contributed to a computational toolkit for studying the formation and structure of lipid nanoparticles, which play a key role in vaccine development. In collaboration with the MLI group, he developed a novel string notation (CGsmiles) to represent molecules at different resolutions. This notation simplifies the handling of large datasets of coarse-grained simulation models and facilitates the automatization of the high-throughput pipeline.

The recognition of Fabian's work is reflected in invited talks at the Frontiers of Coarse-Grained Models workshop of at the Centre Européen de Calcul Atomique et Moléculaire (CECAM) and at the Giersch International Conference 2024 as well as in a contributed talk at the 2024 ACS Fall Meeting. In 2024, Fabian was appointed a core member of the Martini Force Field Initiative (https:www.cgmartini.nl), which develops, maintains, and oversees the development of the widely used Martini coarse-grained simulation method.

Beginning in 2024, Fabian took on lecturing responsibilities at Heidelberg University, where he supervises student internships. He also mentored the computational subgroup of the University of Heidelberg's student team, which won the iGEM synthetic biology competition.

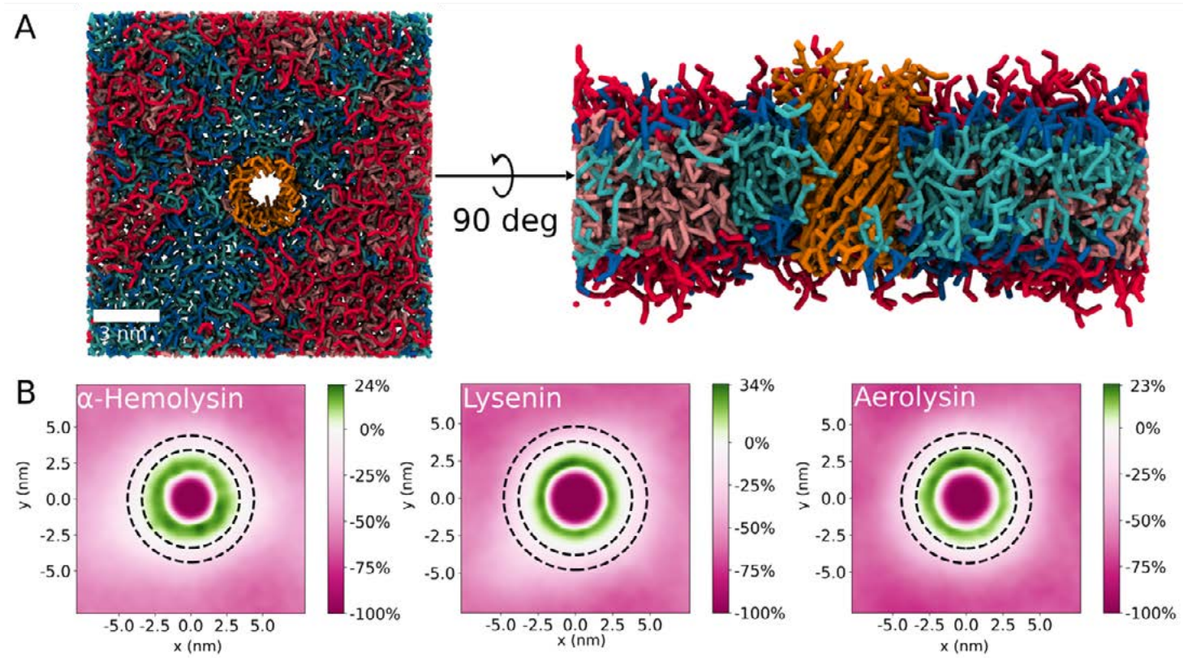


Figure 65: (A) Structure of a hybrid membrane composed of lipid molecules (blue) and polymers (red) with a nanopore protein (orange) embedded in it. (B) Enrichment in % of lipid molecules around different nanopore proteins, which is a measure of stability. Blue shows stable regions, and red shows unstable regions. (Images taken from: Vreeker E et al: Nanopore-Functionalized Hybrid Lipid-Block Copolymer Membranes Allow Efficient Single-Molecule Sampling and Stable Sensing of Human Serum. Advanced Materials, 4 March 2025).

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

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

Die Astrophysikerin **Rajika Kuruwita** ist die erste Wissenschaftlerin im „HITS Independent Postdoc" Programm. Sie stammt aus Sri Lanka, promovierte an der Australian National University und war danach Fellow an der Universität Kopenhagen. Sie kam im September 2022 ans HITS und arbeitet eng mit der SET-Gruppe (siehe Kapitel 2.12) zusammen.

Der zweite „HITS Independent Postdoc" ist der gebürtige Deutsche **Fabian Grünewald**, der seine Promotion in physikalischer und computergestützter Chemie an der Universität Groningen / Niederlande Anfang 2023 mit Auszeichnung abschloss und im Oktober 2023 ans HITS kam. Er arbeitet eng mit der MBM-Gruppe (siehe Kapitel 2.7) und der MLI-Gruppe (siehe Kapitel 2.6) zusammen.

3 Centralized Services

3.1 Administrative Services

Group leader
Dr. Gesa Schönberger

Team
Yashasvini Balachandra (Controlling)
Christina Blach (Front Office)
Frauke Bley (Human Resources)
Benedicta Frech (Front Office)
Silvia Galbusera (Human Resources)
Harald Haas (Team Lead Grant Administration)
Jessica Herbert (Accounting)
Ingrid Kräling (Controlling)
Dr. Barbara Port (Scientific Manager)
Rebekka Riehl (Human Resources; Assistant to the Managing Director)
Jason Vay-Disterhöft (Human Resources)
Irina Zaichenko (Accounting)

The HITS administration team is responsible for various essential administrative processes, including personnel management, operational office and building management, procurement, and financial management. Ultimately, it ensures that legal issues are dealt with and that all processes at the institute comply with legal requirements. In addition to these core functions, the administration team also supports the communications team in organizing events.

Service groups at HITS are specifically designed to relieve the scientists of formal tasks so that they can concentrate fully on their research. In addition to its day-to-day business, in 2024 the HITS team supported the implementation of new Rules for Safeguarding Good Scientific Practice; revised and published an update of the HITS Gender Equality Plan, and established guidelines for the use of AI applications.

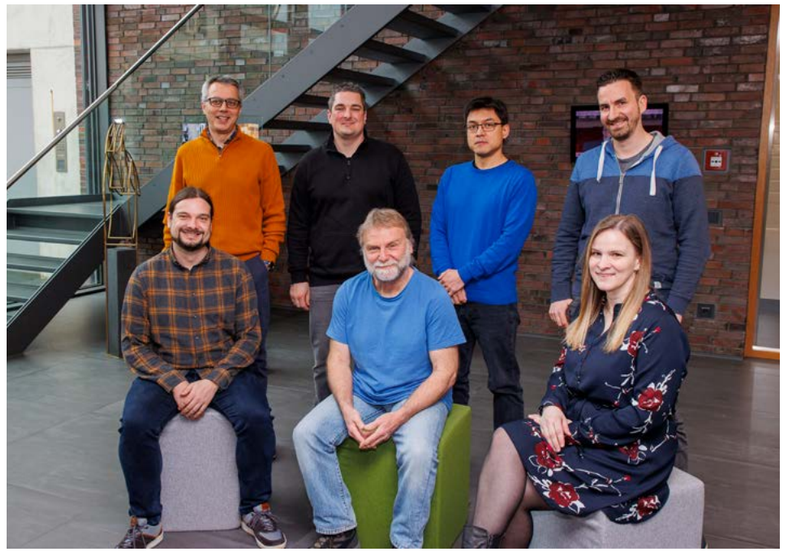
Two issues related to the administrative workload of our scientists were of particular concern to us in 2024.

On the one hand, general legal requirements have significantly increased, thereby resulting in more forms and rules, which have affected the administration of both the institute as well as the individual research groups. Digitization offers a potential solution to this challenge by enhancing the manageability and comprehensibility of work processes. The HITS administration team has been addressing this issue for some time now. However, digitization also goes hand in hand with certain disadvantages, such as the necessity to introduce standard procedures, which potentially reduces flexibility in certain processes. In addition, with digitization, numerous formal activities are transferred from the administration to the scientists, who then require adequate training.

On the other hand, scientific competition for funding is becoming increasingly challenging when it comes to scientists' time and proficiency, particularly when the scientists participate in large consortia as non-university partners. In the current German Excellence Strategy, HITS contributed to nine proposals, with three having progressed to the second phase. A comprehensive analysis of the return on investment would be a valuable effort once the results are available.

While these issues are challenging for the Institute, the administration's goal of supporting scientists so that they can concentrate on their research has not changed - and it will not change in the future.

3.2 IT Infrastructure and Network



Group leader
Dr. Ion Bogdan Costescu

Team
Dr. Bernd Doser (Senior Software Developer)
Dr. Simon Kreuzer (System Administrator)
Christiane Luttermann (Team Assistant)
Norbert Rabes (System Administrator)
Andreas Ulrich (System Administrator)
Martin Wendt (System Administrator; since January 2024)
Taufan Zimmer (System Administrator)

The IT landscape of the institute contains many virtual and physical servers, most of which used CentOS 7. As updates for this operating system stopped being delivered in June 2024, we reinstalled and partly reorganized the servers. For the vast majority of the new installations, we used an automated mechanism in order to simplify long-term maintenance and facilitate another such migration to a new operating system. Through reorganization, some services were consolidated and logically separated, thereby improving their security and preparing for an even more secure setup, which is planned for 2025.

Also in 2024, the Wi-Fi setup was completely replaced. The new access points support a larger number of simultaneously connected devices and a higher bandwidth, and they are centrally managed by our KTA IT colleagues. The authentication in eduroam – which is used by HITSters and by visiting scientists alike – was further optimized and secured.

Our HPC environment also underwent many changes. On the HITS side, the storage system was extended with 2 new servers, thereby boosting its capacity to around 4.1PB. The overall capacity available in our clusters was thus increased to more than 8.6PB.

As the demand for specialized computational resources continued to grow, several compute nodes were added to our cluster located in the Computing Centre of Heidelberg University (URZ). Some of them are CPU-only with many cores, whereas others feature NVIDIA A100 GPUs and large RAM sizes.

Later in the year, the HITS cluster at URZ was relocated to a neighboring server room. As part of this move, most of the CPU-only compute nodes – which had been in operation since 2016 – were retired and were replaced by 45 new nodes featuring dual AMD EPYC 9554 CPUs, 384GB RAM, and NDR InfiniBand connections. Their configuration was chosen after careful consideration of both performance and sustainability aspects. At the same time, the existing GPU nodes and the storage servers were also upgraded to NDR InfiniBand, thereby bringing the whole cluster to a new performance level.

At the very end of the year, two further compute nodes were added to the HITS cluster at URZ. Each of them features dual AMD EPYC 9554 CPUs, 8 NVIDIA H200 CPUs, and more than 2TB RAM. Alongside the A100 GPUs added earlier in the year, these nodes significantly increased the computational resources available to HITSters for developing and running machine learning applications.

In summary, 2024 was a year of large-scale changes in our IT landscape. We invested significant time in planning and executing these changes while trying to keep their impact on the scientific activities at HITS to a minimum.

4 Communication and Outreach

Head of Communications

Dr. Peter Saueressig

Team

Marisa de Sá Almeida
Angela Michel (Deputy Head of Communications)
Dr. Michael Stitz (June–October 2024)
Anna Cap (student)

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. In 2024, we were involved in a plethora of events and other activities.

Success in research – The prerequisite for public interest

A research institute's communication is highly dependent on its researchers and on their scientific success, without which, communicators do not have much to say. In 2024, the Communications team was again pleased to announce several success stories to the public.

HITS group leader Frauke Gräter (Molecular Biomechanics, Chapter 2.7) received an HFSP Research Grant Award for a project on a novel form of mechanosensing. Together with colleagues from Israel and the USA, Frauke is investigating the effects of physical force on the collagen protein in two different animal model systems. The goal is to measure the effects of mechanoradicals on both the integrity of the tissue and the well-being of the organism, including the impact of these mechanoradicals on health and aging.

According to this year's Highly Cited Researchers list from Clarivate, group leader Alexandros Stamatakis (Computational Molecular Evolution, Chapter 2.3) was named one of the most-cited researchers worldwide for the ninth year in a row. The ranking is an important indicator of the impact of a researcher's scientific publications.

HITS researchers were also busy last year publishing papers that gained public attention: Members of the Data Mining and Uncertainty Quantification group (DMQ, Chapter 2.5) published a paper on digital babies that helps us better understand infants' health in their critical first 180 days of life. Together with colleagues at the University of Galway (Ireland) and Heidelberg University Hospital, the authors developed 360 advanced computer models that simulated the unique metabolic processes of each baby (for the project, see the HITS Annual Report 2023, Chapter 2.5). The study was published in the journal Cell Metabolism and resulted in a remarkable international media response.

DMQ researchers Elaine Zaunseder (left) and Vincent Heuveline translated the metabolic processes of infants into mathematical concepts.

An important aspect of research at HITS is the institute's focus on long-term projects. Back in 2014, the journal "Science" featured an article on the avian tree of life that was co-authored by Alexandros Stamatakis. The paper mentioned the essential role of algorithms and supercomputers, which enable modern research in evolutionary biology for all types of living beings. One decade and one giant leap in tool development later, part of the team that coordinated the computer analyses at that time co-authored another paper – this time, in "Nature" – on the complexity of avian evolution.

Workshops, outreach, and HITS Comms "on tour"

2024 was 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.

The year began with the "Neujahrsfest" (New Year's Festival) on 21 January, which was organized by the city of Heidelberg. The event showcased the diverse activities that are undertaken by clubs, associations, and organizations in Heidelberg, including the city's scientific institutes. The HITS communications team rose to the occasion by preparing a poster and a presentation. Moreover, the team used the opportunity to reach out to the broader public and advertised the open house event.

In early May 2024, the 2nd SIMPLAIX Workshop on "Machine Learning for Multiscale Molecular Modeling" took place (see Chapter 5.1.2). Only one week later, HITS scientists Debora Monego (Molecular Biomechanics, MBM), Jeong Yun Choi (Theory and Observations of Stars, TOS), and Leif Seute (Machine Learning and Artificial Intelligence, MLI) were invited to talk about their research on stars, proteins, and machine learning at a SciencePub in the Café Leitstelle that was organized by the JuFORUM e.V. The talks had been arranged by Marisa de Sá Almeida, who gave a short introduction to HITS at the beginning of the event.

In June, HITS again participated in the Explore Science festival at Herzogenried-

Playing the dam game: The CST group at Explore Science.

park in Mannheim, which welcomed more than 30,000 visitors in 2024 year with the topic "Mathematics." Researchers from the CST and SET groups developed two hands-on activities: a dam game and the "Habitable" game (see below). These researchers also joined forces with colleagues from the Heidelberg Institute for Geoinformation Technology (HeiGIT), who offered a hands-on station on (thawing) permafrost, thereby making the shared tent even more diverse.

In July, HITS opened its doors once again to the general public (see Chapter 5.4). 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 Montpellier (France) and Jelenia Gora (Poland). Members of the AIN group gave a workshop on AI in astronomy for these two students.

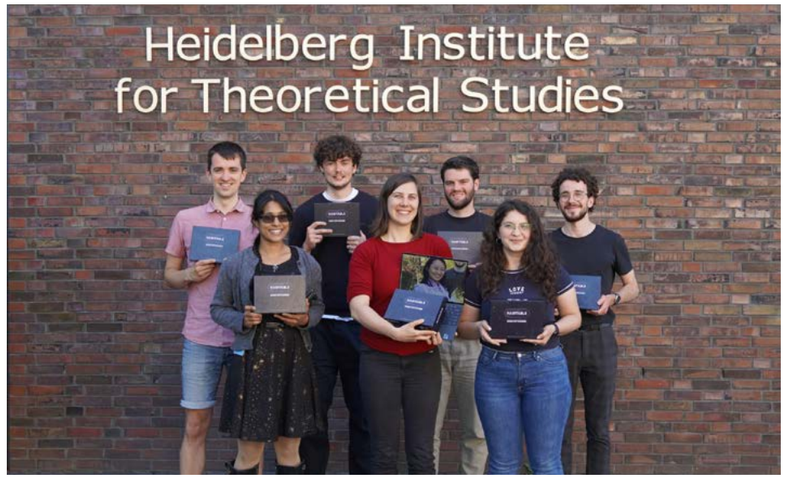
As every year, in September, HITS participated in the Heidelberg Laureate Forum. However, this event turned out to be a special one: HITS took over a booth in the reception area and provided a short talk within the new "Science Hub Showcase" format (see Chapter 7).

In late October, a group of female high-school students visited HITS for a workshop within the framework of the Zukunfts-Orientierungs-Akademie (Future Orientation Academy; short: ZOrA). Christina Goss and Aysecan Ünal (both MBM) as well as Mila Coetzee and Leif Seute (both MLI) showcased examples of their current work and had a coffee chat afterward with the participants, offering them personal insights into career opportunities in science.

Two days later, the HITS Communications team headed for Heidelberg University to assemble its booth for the "Wissenswerte" ("Worth Knowing") conference – the largest event for science journalists in the DACH region (Germany, Switzerland, and Austria). There, the team had the chance to present the interdisciplinary research that had been conducted at HITS to an important target group and to network for new contacts. Moreover, Leif Seute (MLI) again gave a "Science Insight" talk on the use of machine learning in protein design, which gained a lot of feedback.

By the end of 2024, the Events team (Communications plus Christina Blach and Benedicta Frech from the HITS Administration) had managed to organize many scientific events intra et extra muros, including a number of colloquium talks that had to be organized in a hybrid format, streamed, recorded, and broadcast via the HITS YouTube channel (see Chapter 5.2).

The HITS – ISH interns Amelia Niklas (from Jelenia Gora, left) and Zacharie Cotton (Montpellier, right) with their tutors Iliana Isabel Cortés Pérez (2nd right) and Kai Polsterer (2nd left).



The "Habitable" team presenting the printed version of the game at the final showing at HITS.

All that's fit to print: "Habitable" – The success story continues

What makes a planet habitable? How can it be kept habitable, how can life be developed, and what strategic decisions put this life in danger? Members of the Stellar Evolution Theory (SET) group playfully tackled these existential questions in an effort to test the habitability of planets. The result is a board game that combines astronomy and the climate crisis. In a competition as part of the German Science Year "Our Universe," the international team was awarded a prize for their idea in early 2023 and received funding of €10,000 to implement the game. In less than one year, they had managed to develop "Habitable" step by step with game testing events at HITS and elsewhere. Finally, the team published an online version and produced a prototype of the board game (see HITS Annual Report 2023).

In 2024, the project continued to flourish with the arrival of the printed version, a final showing at HITS, and a presentation at Explore Science. The Communications team supported the international team in organizing the testing events and in spreading the news, which resulted in remarkable public awareness.

A full dozen and a view to the future: 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 12th Journalist in Residence was German science journalist and author Felicitas Mokler. She stayed at HITS for six months, gave an internal seminar on "Science Journalism and Science Communication," and delivered a public talk on "Journalismus und Künstliche Intelligenz" (Journalism and Artificial Intelligence). The talk was held again at the MAINS (Mathematics and Informatics Station) in downtown Heidelberg. Moreover, Felicitas used her stay to participate in the Heidelberg Laureate Forum, to visit several institutes in the region and beyond, and to give talks at universities in Germany.

In the summer of 2024, HITS announced the next call for applications, with candidates from six continents applying. A committee of science journalists and scientists selected science journalist Jackson Ryan (Australia) as the next HITS Journalist in Residence. Jackson will come to the institute in April 2024 for a six-month stay.

Two HITS Journalists in Residence meet at the Heidelberg Laureate Forum: Felicitas Mokler with science journalist TV Padma (Journalist in Residence 2017).

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 now begun its third opening round. These developments set a spark of hope for science journalism – an increasingly endangered species that is desperately needed in both science and society.

5 Events

5.1 Conferences, Workshops & Courses

5.1.1 9th International Conference on Systems Biology of Mammalian Cells (SBMC)

Leipzig University, Germany, 13–15 May 2024
The 9th International SBMC Conference was hosted by the HITS Scientific Databases and Visualization (SDBV) group as well as by the LiSyM Cancer team, the latter of which consisted of Thomas Berg (Chair), Stefan Höhme, Ursula Klingmüller, Jens Timmer, Johannes Bode, Steven Dooley, Jochen Hampe, and Beat Müllhaupt (LiSyM Cancer Network Director). The conference ran from 13–15 May 2024 at Leipzig University's Paulinum. Under the theme "Translating Systems Medicine into the Clinics," the event brought together 170 participants from 18 countries to discuss cutting-edge research in systems biology.

The program encompassed several sessions on single-cell analysis, AI in diagnosis and therapy, and the future of systems biology in addition to two poster sessions in which early career researchers had the opportunity to present their work. Moreover, the "Monika and Thomas Zimmermann" (MTZ) Award for outstanding dissertations in systems biology was bestowed upon three young researchers, one of whom was HITS member and LiSyM Cancer Program Manager Ina Biermann (SDBV). In their keynote speeches, Fabian Theis (Helmholtz Munich, Germany) and Ron Heeren (Maastricht University, the Netherlands) gave insights into single-cell dynamics and spatial biology. The conference concluded with a round table discussion on translating systems biology into clinical practice that emphasized patient involvement and biobanks.

The event was funded by the Federal Ministry of Education and Research (BMBF). The next SBMC conference will be held in Mannheim in 2026.

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

Studio Villa Bosch, Heidelberg, 15–17 May 2024
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 molecular systems ranging from small molecules to molecular materials by combining multiscale simulation and machine learning approaches. The 2nd SIMPLAIX workshop – which ran from 15–17 May 2024 – was designed to bring together experts in the field in order to share their research and discuss current challenges (see www.simplaix.org and Section 7 of the Annual Report).

About 100 participants attended the workshop, which featured exciting presentations by 13 invited speakers and vivid discussions with active contributions from both early-career and established scientists. The covered topics included machine-learned force fields across different scales, large language models for chemistry, and simulations of transport processes and dynamics in biomolecular systems and materials. 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 computational efficiency, accuracy, and scope.

SIMPLAIX is supported by the Klaus Tschira Foundation and the participating institutions. For more details, see https://simplaix-workshop2024.h-its.org/.

Scientific Organizing Committee: Rebecca Wade, Rostislav Fedorov, and Daniel Sucerquia (all from HITS) as well as Pascal Friederich, Marcus Elstner, and David Hoffmann (all from KIT).

5.1.3 Legend2024: Machine Learning for Evolutionary Genomics Data

Heraklion, Greece, 13–15 May 2024
This was the inaugural conference of an emerging conference series that took place at the Foundation for Research and Technology Hellas in Heraklion, Crete (the 2025 version will take place in the French Alps). HITS group leader Alexandros Stamatakis (Computational Molecular Evolution) served as the main organizer (https://legend2024.sciencesconf.org/). CME PHD students gave two oral presentations at the conference.

5.1.4 Practical Course on Computational Molecular Evolution

Hinxton, UK, 18–29 May 2024
The practical course on "Computational Molecular Evolution" has been run biennially in Heraklion, Greece, and Hinxton, UK, since 2009. In 2024, the 14th installation of the event took place again at the Wellcome Genome Campus in Hinxton.

The aim of the course was to provide early-career-stage researchers with the theoretical knowledge and practical skills needed to carry out molecular evolutionary analyses on sequence data. The overall objective was to raise the standard of research that uses statistical methods and software programs in the field of molecular evolution and phylogenetics by providing high-quality, in-depth, face-to-face training on the topic for as many researchers as possible.

The course offered an overview of phylogenetic terminology, methods and analyses, phylogeny reconstruction (including maximum likelihood and Bayesian methods), alignment techniques, and hypothesis testing and computer simulation in addition to the analysis of genomic data as well as of protein and nucleotide sequences.

Moreover, the participants had the opportunity to interact with leading scientists in evolutionary bioinformatics and authors of famous analysis tools (Maria Anisimova, Bastien Bousseau, Nick Goldman, Adam Leaché, Michael May, Joana Meier, Bruce Rannala, Alexandros Stamatakis, Benjamin Redelings, Ziheng Yang, Jeffrey Thorne, Rachel Warnock, etc.).

CME group leader Alexandros Stamatakis was a co-organizer of the event together with Aglaia (Cilia) Antoniou (IMBCC, Heraklion, Greece) and Adam Leaché (University of Washington, USA). Among the training team of 15 researchers were CME group member Dimitri Höhler and HITS Alumna Paschalia Kapli (now a research leader at the Natural History Museum, London, UK).

5.1.5 ML4ASTRO Workshop

Studio Villa Bosch, Heidelberg, 25 June 2024
The Machine Learning for Astronomy (ML4Astro) workshop was held on 25 June 2024 at 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 60 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 astrophysical problems.

The program began with a series of presentations covering topics such as machine learning for stellar spectra, Bayesian statistics, and clustering methods in astrophysics. These sessions were followed by informal group discussions, which fostered an interactive and collaborative atmosphere.
The primary objective of the one-day workshop was to facilitate knowledge exchange among participants regarding the application of machine learning techniques for solving similar challenges across various astrophysical contexts. Additionally, the workshop provided a platform to explore potential new collaborations and joint research initiatives.

5.1.6 18th "Würzburg" Winter Workshop on Stellar Astrophysics

Studio Villa Bosch, Heidelberg, 9–11 December 2024
Bringing together over 40 astrophysicists from institutions across Germany and around the world, the annual "Würzburg" Workshop on Stellar Astrophysics took place at Studio Villa Bosch from 9–11 December 2024. Participants discussed topics such as kilonovae and supernovae, common envelopes, and stellar mergers as well as machine learning in the respective contexts.
The agenda included talks by Fiona McNeill (Queen's University Belfast, Northern Ireland, United Kingdom), Federico Rizzuti (University of Trieste, Italy), Rajika Kuruwita (HITS), and Christian Klingenberg (University of Würzburg, Germany).
The workshop aimed not only to unite familiar peers in the field, but also to foster connections and spark future collaborations. In keeping with the season, traditional German Christmas customs were an integral part of the event, as they are every year.

5.2 HITS Colloquia

Meik Bittkowski
Science Media Center, Cologne, Germany
22 January 2024: Augmenting Science Journalism – The development of technology-based support for science journalism at the Science Media Center Germany (Hybrid)

Pascal Friederich
KIT, Karlsruhe, Germany
18 March 2024: Machine Learning for materials science. In: AI at HITS – Symposium in honor of Wilfried Juling (Hybrid)

Ganna Gryn'ova
HITS, Heidelberg, Germany
18 March 2024: Exploring Chemical Space with Machine Learning. In: AI at HITS – Symposium in honor of Wilfried Juling (Hybrid)

Sebastian Lerch
KIT, Karlsruhe and HITS, Heidelberg, Germany
18 March 2024: Exploring Chemical Space with Machine Learning. In: AI at HITS – Symposium in honor of Wilfried Juling (Hybrid)

Michele Parrinello
Italian Institute of Technology, Genova, Italy
22 April 2024: HITS-SIMPLAIX Joint Colloquium – Do we really understand catalysis? (Hybrid)

Volker Springel
MPI for Astrophysics, Garching, Germany
24 June 2024: Interfacing galaxy formation with precision cosmology (Hybrid)

Felicitas Mokler
Journalist in Residence, Germany
17 October 2024: Journalismus und Künstliche Intelligenz - passt das zusammen? (at MAINS, Heidelberg)

Guillermo Cabrera-Vives
Klaus Tschira Guest Professor, Department of Computer Science, Universidad de Concepción, Chile
21 October 2024: Machine Learning & Astronomy: Unique Challenges for a Unique Field (Hybrid)

Markus Meuwly
Department of Chemistry, University of Basel, Switzerland
16 December 2024: HITS-SIMPLAIX Joint Colloquium – Computational Protein Dynamics in the Era of Machine-Learned Atomistic Simulations (Hybrid)

5.3 AI at HITS: A symposium in honor of Wilfried Juling

On 18 March 2024, HITS organized a scientific symposium in honor of Wilfried Juling, a longtime supporter of HITS and member of the board of directors of the HITS-Stiftung (HITS Foundation) from 2016–2023.

The symposium comprised three scientific talks on different aspects of artificial intelligence ranging from chemistry and materials research to weather prediction. It also included a laudation by Carsten Könneker, Director of the HITS-Stiftung. After the introduction by Managing Director Gesa Schönberger and Scientific Director Tilmann Gneiting, HITS researcher Sebastian Lerch (CST) talked about the alleged "AI revolution of weather prediction," followed by a talk by HITS group leader and principal investigator (PI) in SIMPLAIX Ganna (Anya) Gryn'ova (CCC) on "exploring chemical space with machine learning." Subsequently, Pascal Friederich from KIT – also a PI in SIMPLAIX – talked about "machine learning for materials science."

In his laudatory speech, Carsten Könneker elaborated on Wilfried Juling's career and multiple achievements in both research and science management. A mathematician by training, Juling held professorships at the Universities of Rostock and Karlsruhe before becoming Director of the Steinbuch Centre for Computing at the then-newly formed KIT in 2008. As a representative of KIT, Wilfried Juling was involved in the process of establishing HITS, closely collaborating with Klaus Tschira and Andreas Reuter, the latter of whom is the former Scientific and Managing Director of the HITS predecessor, EML Research. After Klaus Tschira's untimely death in 2015, Wilfried Juling was appointed his successor as Director of the HITS-Stiftung – a position he held until the end of 2023. During this time period, Juling eagerly worked on the establishment of a triangular initiative between HITS and its university partner. The outcome was the SIMPLAIX initiative in 2021, which has successfully contributed to the visibility of AI in the field of simulations (see Chapter 7).

Former and current members of the boards of directors of the HITS Stiftung (HITS Foundation) (f.l.t.r.): Dieter Kranzlmüller, Carsten Könneker, Andreas Reuter, Wilfried Juling, and Rafael Lang.

Ganna (Anya) Gryn'ova during her talk.

5.4 HITS Open House Event

13 July 2024

After two years, HITS opened its doors once again to the general public, thereby providing an opportunity to explore the institute's premises and learn about its research activities. The event featured a diverse program that catered to all age groups and included face painting for children, engaging hands-on stations, and informative talks.

"Weather, viruses and the Universe – AI for the world of tomorrow" was the motto of the event, which attracted more than 500 visitors to the HITS campus. A highlight was the "flood experiment" by the Computational Statistics group, for which participants built dams and tested their stability with simulated water surges determined by dice rolls. Visitors enjoyed a glimpse of the institute's workspace during the guided garden tours, while in the Studio Villa Bosch, three talks showcased the diverse research conducted at HITS.

The first talk – which was led by junior group leader Fabian Schneider from the Stellar Evolution Theory (SET) group – delved into the intriguing phenomenon of black hole sounds. This was followed by a talk from Debora Monego – a postdoc with the Molecular Biomechanics (MBM) group – who demonstrated the use of computer simulations to solve complex biological puzzles. Debora's session included an interactive component in which attendees were invited to fold an origami protein from paper. The final presentation – which was held by junior group leader Jan Stühmer of the Machine Learning and Artificial Intelligence (MLI) group – explored the mechanisms through which large language models such as ChatGPT store factual knowledge. In his talk, Jan addressed the reasons these models occasionally provide incorrect information.

Throughout the campus, visitors participated in interactive activities, such as card sorting, experiments on solar energy, and wind tunnel tests. At the photo station, guests posed in research-themed scenarios and were provided with keepsake photos.

The interactive nature of the event resonated strongly with families, who appreciated the variety of activities designed to engage children. The event succeeded in offering a glance into the dynamic research environment at HITS while also stimulating curiosity among the attendees.
The next HITS open day is scheduled for summer 2026.

Scan the QR code for more impressions of the event:

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.

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

Machine Learning in an interdisciplinary environment: Guillermo Cabrera-Vives
Guillermo Cabrera-Vives is an astronomer and computer scientist from the University of Concepción, Chile. He came to HITS in September 2024 and will stay until February 2025.

At the beginning of his stay, Guillermo participated in the Heidelberg Laureate Forum (see Chapter 7), delving into the vibrant atmosphere of this special event and meeting laureates as well as young researchers from computer science and mathematics. In October, he held a HITS colloquium talk on machine learning and astronomy at the Studio Villa Bosch (see Chapter 5.2).

Guillermo's research focus is on machine learning in different application fields, including astronomy data, genomics, biomedical imaging, and satellite images. His main point of contact was with group leader Kai Polsterer (Astroinformatics). The two have known each other ever since the 2018 Astroinformatics Conference – an international event organized by Kai that also took place at the Studio Villa Bosch.

Meeting and discussing science at the Heidelberg Laureate Forum (f.r.t.l.): Kai Polsterer, Guillermo Cabrera-Vives, and U.S. science journalist Zack Savitsky.

Guillermo additionally met with Tilmann Gneiting (Computational Statistics) and Jan Stühmer (Machine Learning and Artificial Intelligence) – and with their group members – to discuss applications, theoretical aspects, and problems in machine learning. "I learned a lot of techniques how I can calibrate my models and measure their uncertainty," Guillermo stated in an interview with the HITS newsletter, "The Charts."

As an open-minded, interdisciplinary-oriented scientist, Guillermo is also interested in other fields, for instance, in how machine learning can help in cancer treatment. He is thus currently pursuing a project that uses AI algorithms to segment tumors in order to help radiologists monitor cancer treatment effectively. With members of the Data Mining and Uncertainty Quantification group, Guillermo discussed the methods they use to segment fossils, which are similar to his own methods.

Guillermo has also enjoyed the atmosphere at HITS: "I admire the way how research is done here. I like the collaborative work and the spirit: You can feel it." In a resume of his stay, he stated, "Scientists usually think about the next paper. Here, people talk about the field: not about short-term, but long-term projects. They have the bigger picture in mind."

Guillermo Cabrera-Vives during his HITS colloquium talk on machine learning and astronomy in October 2024.

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 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 independent research projects.

The fellowship is awarded for two years, with an option for a one-year extension following a positive evaluation. It offers a vibrant research community and a highly interdisciplinary and international working environment with close links to the HITS shareholders of Heidelberg University and the Karlsruhe Institute of Technology (KIT). In addition, successful candidates 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 at the application deadline. This deadline can be extended in the case of documented career breaks, for example, due to parental leave. Candidates must not have conducted research at HITS previously except for brief visits, and the main thread of their research must not have been in collaboration with a HITS group leader.

The first HITS Independent Postdoc is astrophysicist Rajika Kuruwita. Born in Sri Lanka, Rajika completed her PhD at the Australian National University, was a fellow at the University of Copenhagen, and joined HITS in September 2022 (see Annual Report 2022). In 2024, she published several papers, one of which was a study on the relationship between binary/multiple stars in the Astronomy & Astrophysics journal (see Chapter 2.14). Rajika was invited to write the Star Formation chapter in Elsevier's Encyclopedia of Astrophysics. Moreover, she was also an invited speaker at international scientific meetings in Sweden and Germany (see Chapter 2.14).

In October 2023, biologist Fabian Grünewald (University of Groningen, the Netherlands) joined the institute as the second HITS Independent Postdoc (see also Chapter 2.14). Born in Germany, Fabian studied chemistry at the University of Groningen, the Netherlands. He received his PhD with distinction in physical/computational chemistry at the University of Groningen, where he stayed as a researcher in the Molecular Dynamics group before joining HITS. Fabian's 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, he focuses on RNA and aims to develop computational methods and software that can help to design RNA nanoparticle therapeutics.

In 2024, Fabian worked with experimentalists from Groningen on proteins in hybrid membranes. He also collaborated closely with colleagues from the Machine Learning and Artificial Intelligence group at HITS, developing a novel string notation for use in representing molecules at different resolutions. Moreover, Fabian held invited talks at conferences in Switzerland, the USA, and Germany. Last but not least, he was appointed a core member of the Martini Force Field Initiative community, which develops a widely used simulation method (see Chapter 2.14).

The next opening for the program will be in 2025.

7 Collaborations

SIMPLAIX

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

In SIMPLAIX, these methods are developed and employed to study a set of challenging problems in 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 Frauke Gräter (MBM). Ganna Gryn'ova (CCC, see Chapter 2.2) was also among its Principal Investigators (PI). In 2024, Jan Stühmer (MLI) became the 10th SIMPLAIX PI. In the course of the year, eleven doctoral students worked on the 8 projects. SIMPLAIX is funded by the Klaus Tschira Foundation and is supported by in-kind contributions from KIT and Heidelberg University.

From 15-17 May 2024, the second 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 to share their research and discuss current challenges. In this on-site meeting, about 100 participants attended the eight sessions, in which thirteen invited speakers gave talks on current developments in machine learning for chemistry and molecular biology, from metal chemistry to drug design.

Shortly after the workshop, the Chemical Compound Space Conference (CCSC2024), which was co-organized by two SIMPLAIX PIs, took place at the Heidelberg Congress Center. With both events providing potential material and authors, several SIMPLAIX PIs co-edited a virtual special issue of the Journal of Chemical Information and Modeling, published by the American Chemical Society (see [Gryn'ova, Bereau et al, 2024], Chapter 8). In the course of the year, four internal project meetings were held at Studio Villa Bosch, at Heidelberg University and at KIT, and the SIMPLAIX researchers engaged in several journal session discussions. Moreover, several joint HITS–SIMPLAIX colloquia (see Chapter 5.2) and scientific talks with external speakers were organized.

The second SIMPLAIX workshop attracted about 100 participants.

Michele Parinello (Italian Institute of Technology, Genova, Italy) during his HITS-SIMPLAIX colloquium talk in April 2024.

Plans for 2025 include the third international SIMPLAIX workshop in May, and several joint colloquia and project meetings are also in the offing.

Heidelberg Laureate Forum

The Heidelberg Laureate Forum (HLF) is a networking conference at which 200 carefully selected young researchers in mathematics and computer science spend a week interacting with laureates from the same two disciplines, including recipients of the Abel Prize, the ACM A.M. Turing Award, the ACM Prize in Computing, the Fields Medal, and the Nevanlinna Prize. Established in 2013, the HLF is held annually by the Heidelberg Laureate Forum Foundation (HLFF). HITS has been a scientific partner of the HLF since 2016.

For the first time: A HITS booth at the HLF

From 22–27 September, the HLF again took place on the Heidelberg University premises. The program included laureate lectures, master classes, lightning talks, discussions, and various interactive program elements.

For the first time, HITS had the chance to be visible at the event with a booth in the reception area. The communications team – supported by student Johanna Riedel (AIN) – prepared a short presentation and

At the HITS booth: Marisa de Sá Almeida (HITS Communications) talking to young researchers.

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.

A new format: 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. Ten organizations participated in the introduction round, which took place in the university auditorium. For HITS, Deputy Scientific Director Kai Polsterer hit the stage and outlined the most important facts about the institute to the audience. Afterward, 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.

HITS Deputy Scientific Director Kai Polsterer during his "elevator pitch" talk at the "Science Hub Showcase".

Informatics4Life

Despite remarkable progress in the diagnosis and treatment of acute and chronic cardiovascular diseases in recent years, these illnesses remain the leading cause of death and hospitalization for all people worldwide. Informatics4life was a collaborative initiative funded by the Klaus Tschira Foundation that focused on cardiovascular research and brought together experts in clinical research and computational methods. The initiative's patient-centric environment – a completely novel approach – was critical to translating research into clinical application. As Heidelberg is a "hot spot" for health research and medicine with internationally highly competitive institutes and researchers, it is the ideal location for this pioneering initiative.

The project represented an interdisciplinary alliance between cardiovascular physicians and computer scientists from Heidelberg University, University Hospital Heidelberg, and HITS. It consisted of several subprojects in which HITS researchers were involved as principal investigators and contributors.

HITS involvement: From structure-based design to machine learning

Rebecca Wade (MCM) was a Principal Investigator of the subproject "Structure-based design of peptide-based pharmaceuticals against striated muscle disorders" (see Chapter 2.8); Wolfgang Müller (SDBV) was Collaborating Principal Investigator of the subproject "Research data warehouse"; and Vincent Heuveline (DMQ) was Principal Investigator of the subproject "Cognition and uncertainty quantification for numerical heart simulation." Moreover, Vincent was also Co-Coordinator of Informatics4life in cooperation with Hugo Katus and Benjamin Meder (both from Heidelberg University Hospital).

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

8 Publications

Al-Belmpeisi R, Tuhtan V, Christensen MB, Kuruwita R, Haugbølle T (2024). Simulated analogues II: a new methodology for non-parametric matching of models to observations. Monthly Notices of the Royal Astronomical Society 534(4):3194-3210

Alpay A, Heuveline V (2024). AdaptiveCpp Stdpar: C++ Standard Parallelism Integrated Into a SYCL Compiler. Proceedings of the 12th International Workshop on OpenCL and SYCL, pp.1-12, ACM

Alsulami R, Einecke S, Rowell G, McGee P, Filipović MD, Seitenzahl IR, Stupar M, Collins TA, Fukui Y, Sano H (2024). Investigating unusual Hα features towards the Scutum Supershell, Publ. Astron. Soc. Aust. 41,e098

Andrassy R, Leidi G, Higl J, Edelmann PVF, Schneider FRN, Röpke FK (2024). Towards a self-consistent model of the convective core boundary in upper main sequence stars. A&A 683:A97

Arend D, Conte AD, Feser M, Gadiya Y, Gaignard A, Castro LJ, Mičetić I, Moretti S, Neumann S, Rayya N, Tsueng G, Willighagen E, Wittig U (2024). Bioschemas Resource Index for Chem and Plants. OSF; 2024. doi:10.37044/osf.io/yxunp.

Arnold S, Walz E-M, Ziegel J, Gneiting T (2024). Decompositions of the mean continuous ranked probability score. Electronic Journal of Statistics, 18(2):4992–5044

Back S, Aspuru-Guzik A, Ceriotti M, Gryn'ova G, Grzybowski B, Gu GH, Hein J, Hippalgaonkar K, Hormázabal R, Jung Y, Kim S, Kim WY, Moosavi SM, Noh J, Park C, Schrier J, Schwaller P, Tsuda K, Vegge T, von Lilienfeld OA, Walsh A (2024). Accelerated chemical science with AI. Digital Discovery 3(1):23-33

Bermúdez-Bustamante LC, De Marco O, Siess L, Price DJ, González-Bolívar M, Lau MYM, Mu C, Hirai R, Danilovich T, Kasliwal MM (2024). Dust formation in common envelope binary interactions – II: 3D simulations with self-consistent dust formation. Monthly Notices of the Royal Astronomical Society, stae1841

Boushehri S, Holey H, Brosz M, Gumbsch P, Pastewka L, Aponte-Santamaría C, Gräter F (2024). O-glycans Expand Lubricin and Attenuate Its Viscosity and Shear Thinning. Biomacromolecules,25(7):3893-3908

Bracher J, Koster N, Krüger F, Lerch S (2024). Learning to forecast: The Probabilistic Time Series Forecasting Challenge. The American Statistician, 78(1):115–127

Braun TAM, Ahlborn F, Weiss A (2024). Testing a non-local 1-equation turbulent convection model: A solar model. A&A 689:A292

Brehmer JR, Gneiting T, Herrmann M, Marzocchi W, Schlather M, Strokorb K (2024). Comparative evaluation of point process forecasts. Annals of the Institute of Statistical Mathematics, 76:47–71

Bronner VA, Schneider FRN, Podsiadlowski Ph, Röpke FK (2024). Going from 3D to 1D: A 1D approach to common-envelope evolution. A&A 683:A65

Bronner VA, Schneider FRN, Podsiadlowski P, Röpke FK (2024). Going from 3D Common-envelope Simulations to Fast 1D Simulations. Bull. Soc. R. Sci. Liege:230-243

Buchele L, Bellinger EP, Hekker S, Basu S, Ball W, Christensen-Dalsgaard J (2024). Asteroseismic Inversions for Internal Sound Speed Profiles of Main-sequence Stars with Radiative Cores. ApJ 961(2):198

Callan FP, Collins CE, Sim SA, Shingles LJ, Pakmor R, Srivastav S, Pollin JM, Gronow S, Roepke FK, Seitenzahl IR (2024). Exploring the range of impacts of helium in the spectra of double detonation models for Type Ia supernovae. arXiv:2408.03048

Callan FP, Sim SA, Collins CE, Shingles LJ, Lach F, Röpke FK, Pakmor R, Kromer M, Srivastav S (2024). Including a luminous central remnant in radiative transfer simulations for Type Iax supernovae. Monthly Notices of the Royal Astronomical Society 530(2):1457-1473

Cassell TL, Deakin T, Alpay A, Heuveline V, Gadeschi GB (2024). Efficient Tree-based Parallel Algorithms for N-Body Simulations Using C++ Standard Parallelism. SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis, pp.708-717, IEEE

Cetin M, Pinamonti V, Schmid T, Boschert T, Mellado Fuentes A, Kromer K, Lerner T, Zhang J, Herzig Y, Ehlert C, Hernandez-Hernandez M, Samaras G, Torres CM, Fisch L, Dragan V, Kouwenhoven A, Van Schoubroeck B, Wils H, Van Hove C, Platten M, Green EW, Stevenaert F, Felix NJ, Lindner JM (2024). T-FINDER: A highly sensitive, pan-HLA platform for functional T cell receptor and ligand discovery. Sci. Adv. 10(5), eadk3060

Charou D, Rogdakis T, Latorrata A, Valcarcel M, Papadogiannis V, Athanasiou C, Tsengenes A, Papadopoulou MA, Lypitkas D, Lavigne MD, Katsila T, Wade RC, Cader MZ, Calogeropoulou T, Gravanis A, Charalampopoulos I (2024). Comprehensive characterization of the neurogenic and neuroprotective action of a novel TrkB agonist using mouse and human stem cell models of Alzheimer's disease. Stem Cell Res Ther 15(1), 200

Coppée Q, Müller J, Bazot M, Hekker S (2024). The radial modes of stars with suppressed dipole modes. A&A 690:A324

Crisci L, Carpentieri L, Thoman P, Alpay A, Heuveline V, Cosenza B (2024). SYCL-Bench 2020: Benchmarking SYCL 2020 on AMD, Intel, and NVIDIA GPUs. Proceedings of the 12th International Workshop on OpenCL and SYCL, pp.1-12, ACM

D'Anna F, Jareborg N, Jetten M, Ahokas M, Alper P, Andrews R, Bösl K, D'Altri T, Faria D, Fatima N, Fuchs S, Garrard C, Gu W, Heil KF, Kallberg Y, Licciulli F, Lübke N-C, Melo AMP, Mičetić I, Oliveira J, Oulas A, Palagi PM, Poterlowicz K, Perez-Sitja X, Ruch P, Sansone S-A, Schnitzer H, van Gelder C, Vergoulis T, Wibberg D, Wittig U, Leskošek B, Vondrasek J, Andrabi M (2024). A research data management (RDM) community for ELIXIR. F1000Res 13:230

De Lucia G, Kruijssen JMD, Trujillo-Gomez S, Hirschmann M, Xie L (2024). On the origin of globular clusters in a hierarchical universe. Monthly Notices of the Royal Astronomical Society 530(3):2760-2777

Diedrich L, Brosz M, Abele T, Steinke S, Gräter F, Göpfrich K, Aponte-Santamaría C (2024). Energetics and kinetics of membrane permeation of photoresists for bioprinting. Advanced Functional Materials, Volume 34, Issue 20.

Dimitriadis T, Gneiting T, Jordan AI, Vogel P (2024). Evaluating probabilistic classifiers: The triptych. International Journal of Forecasting, 40(3):1101–1122

Dufton PL, Langer N, Lennon DJ, Schneider FRN, Evans CJ, Sana H, Taylor WD (2024). Nitrogen abundances of the Be-type stars in 30 Doradus. Monthly Notices of the Royal Astronomical Society 527(3):5155-5165

D'Arrigo G, Kokh DB, Nunes-Alves A, Wade RC (2024). Computational screening of the effects of mutations on protein-protein off-rates and dissociation mechanisms by τRAMD. Commun Biol 7(1),1159

Elias-Rosa N, Brennan SJ, Benetti S, Cappellaro E, Pastorello A, Kozyreva A, Lundqvist P, Fraser M, Anderson JP, Cai Y-Z, Chen T-W, Dennefeld M, Gromadzki M, Gutiérrez CP, Ihanec N, Inserra C, Kankare E, Kotak R, Mattila S, Moran S, Müller-Bravo TE, Pessi PJ, Pignata G, Reguitti A, Reynolds TM, Smartt SJ, Smith K, Tartaglia L, Valerin G, de Boer T, Chambers K, Gal-Yam A, Gao H, Geier S, Mazzali PA, Nicholl M, Ragosta F, Rest A, Yaron O, Young DR (2024). SN 2020pvb: A Type IIn-P supernova with a precursor outburst. A&A 686:A13

Elter M, Brosz M, Sucerquia D, Kuzhelev A, Kiesewetter DC, Kurth M, Dreuw A, Prisner TF, Freudenberg J, Bunz UHF, Gräter F (2024). Breaking Strong Alkynyl-Phenyl Bonds: Poly(para-phenylene ethynylene)s under Mechanical Stress. Journal of the American Chemical Society

Evans EL, Dupuy TJ, Sullivan K, Kraus AL, Huber D, Ireland MJ, Ansdell M, Kuruwita RL, Martinez RA, Wood ML (2024). Orbital architectures of planet-hosting binaries – III. Testing mutual inclinations of stellar and planetary orbits in triple-star systems. Monthly Notices of the Royal Astronomical Society 534(1):575-607

Feeney T, Petry J, Torche A, Hauschild D, Hacene B, Wansorra C, Diercks A, Ernst M, Weinhardt L, Heske C, Gryn'ova G, Paetzold UW, Fassl P (2024). Understanding and exploiting interfacial interactions between phosphonic acid functional groups and co-evaporated perovskites. Matter, S2590238524000675

Feik M, Lerch S, Stühmer J (2024). Graph Neural Networks and Spatial Information Learning for Post-Processing Ensemble Weather Forecasts. In ICML Machine Learning for Earth System Modeling Workshop. arXiv:2407.11050

Fosten J, Gutknecht D, Pohle M-O (2024). Testing quantile forecast optimality. Journal of Business & Economic Statistics, 42(4):1367–1378

Frost AJ, Sana H, Mahy L, Wade G, Barron J, Le Bouquin J-B, Mérand A, Schneider FRN, Shenar T, Barbá RH, Bowman DM, Fabry M, Farhang A, Marchant P, Morrell NI, Smoker JV (2024). An interaction and merger in a massive multiple system create a magnetic field in a massive star. Bull. Soc. R. Sci. Liege:85-102

Frost AJ, Sana H, Mahy L, Wade G, Barron J, Le Bouquin J-B, Mérand A, Schneider FRN, Shenar T, Barbá RH, Bowman DM, Fabry M, Farhang A, Marchant P, Morrell NI, Smoker JV (2024). A magnetic massive star has experienced a stellar merger. Science 384(6692):214-217

Gagliano A, Berger E, Villar VA, Hiramatsu D, Kessler R, Matsumoto T, Gilkis A, Laplace E (2024). Finding the Fuse: Prospects for the Detection and Characterization of Hydrogen-rich Core-collapse Supernova Precursor Emission with the LSST. ApJ 978(1):110

Gensior J, Feldmann R, Reina-Campos M, Trujillo-Gomez S, Mayer L, Keller BW, Wetzel A, Kruijssen JMD, Hopkins PF, Moreno J (2024). H i discs of L* galaxies as probes of the baryonic physics of galaxy evolution. Monthly Notices of the Royal Astronomical Society 531(1):1158-1178

Ghalami F, Dohmen PM, Krämer M, Elstner M, Xie W (2024). Nonadiabatic Simulation of Exciton Dynamics in Organic Semiconductors Using Neural Network-Based Frenkel Hamiltonian and Gradients. J. Chem. Theory Comput. 20,14, 6160–6174

Gianniotis N, Polsterer KL, Cortés Pérez II (2024). Positive and Scale Invariant Gaussian Process Latent Variable Model for Astronomical Spectra. ESANN proceesdings, pp.315-320, Ciaco - i6doc.com

Gryn'ova G, Bereau T, Müller C, Friederich P, Wade RC, Nunes-Alves A, Soares TA, Merz K Jr (2024). EDITORIAL: Chemical Compound Space Exploration by Multiscale High-Throughput Screening and Machine Learning. J. Chem. Inf. Model. 64(15):5737-5738

Guerra J, Belleri M, Paiardi G, Tobia C, Capoferri D, Corli M, Scalvini E, Ghirimoldi M, Manfredi M, Wade RC, Presta M, Mignani L (2024). Impact of an irreversible β-galactosylceramidase inhibitor on the lipid profile of zebrafish embryos. Computational and Structural Biotechnology Journal 23:1397-1407

Haag J, Jordan AI, Stamatakis A (2024) Pandora: A Tool to Estimate Dimensionality Reduction Stability of Genotype Data. Biorxiv. https://doi.org/10.1101/2024.03.14.584962

Hamede MH, Roller J, Meyer A, Heuveline V, Egbers C (2024). Dielectrophoretic force-enhanced thermal convection within a horizontal cylindrical annulus. Physics of Fluids 36(12),124106

Han SB, Teuffel J, Mukherjee G, Wade RC (2024). Multiresolution molecular dynamics simulations reveal the interplay between conformational variability and functional interactions in membrane-bound cytochrome P450 2B4. Protein Science 33(10), e5165

Henneco J, Schneider FRN, Hekker S, Aerts C (2024). Merger seismology: Distinguishing massive merger products from genuine single stars using asteroseismology. A&A 690: A65

Henneco J, Schneider FRN, Hekker S, Aerts C (2024). Merger seismology: Asteroseismic properties of massive merger products. In 8th TESS/15th Kepler Asteroseismic Science Consortium Workshop, p. 32

Henneco J, Schneider FRN, Laplace E (2024). Contact tracing of binary stars: Pathways to stellar mergers. A&A 682:A169

Hespe D, Hübner L, Kurpicz F, Sanders P, Schimek M, Seemaier D, Uhl TN (2024). Brief Announcement: (Near) Zero-Overhead C++ Bindings for MPI. Proceedings of the 36th ACM Symposium on Parallelism in Algorithms and Architectures, pp.289-291, ACM

Heuveline V, Zeilmann A (2024). KI-Morph – User-friendly large-scale image analysis & AI on bwHPC systems. Proceedings of 10th bwHPC Symposium

Horat N, Klerings S, Lerch S (2024). Improving model chain approaches for probabilistic solar energy forecasting through post-processing and machine learning. Advances in Atmospheric Sciences, 42(2):297–312

Horat N, Lerch S (2024). Deep learning for postprocessing global probabilistic forecasts on subseasonal time scales. Monthly Weather Review, 152(3):667–687

Hu X, Abaza H, Hänsel R, Abedi M, Golebiewski M, Müller W, Meineke F (2024). NFDI4Health Local Data Hubs Implementing a Tailored Metadata Schema for Health Data. Stud Health Technol Inform. 317:115-122

Häuser L, Jäger G, Rama T, List J-M, Stamatakis A (2024). Are Sounds Sound for Phylogenetic Reconstruction? Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP, pp.78–87, St. Julian's, Malta. Association for Computational Linguistics

Häuser L, Jäger G, Stamatakis A (2024). Computational Approaches for Integrating out Subjectivity in Cognate Synonym Selection. Proceedings of the Society for Computation in Linguistics, pp.162–172, Irvine, CA. Association for Computational Linguistics

Höhlein K, Schulz B, Westermann R, Lerch S (2024). Postprocessing of ensemble weather forecasts using permutation-invariant neural networks. Artificial Intelligence for the Earth Systems, 3(1):e230070

Hübner L, Stamatakis A (2024). Memoization on Shared Subtrees Accelerates Computations on Genealogical Forests. Biorxiv. https://doi.org/10.1101/2024.05.23.595533

Jeon S, Yeh C, Inan H, Hsu W, Rungta R, Mehdad Y, Bikel D (2024). Attention or convolution: Transformer encoders in audio language models for inference efficiency. Proceedings of the Workshop on Self-supervision in Audio, Speech and Beyond at IEEE ICASSP, Seoul, South Korea, 14 April

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 (2024). S100A1ct: A Synthetic Peptide Derived From S100A1 Protein Improves Cardiac Performance and Survival in Preclinical Heart Failure Models. Circulation, CIRCULATIONAHA.123.066961

Korol V, Buscicchio R, Pakmor R, Morán-Fraile J, Moore CJ, de Mink SE (2024). Expected insights into Type Ia supernovae from LISA's gravitational wave observations. A&A 691:A44

Kozlov OM, Stamatakis A (2024). EcoFreq: Compute with Cheaper, Cleaner Energy via Carbon-Aware Power Scaling. ISC High Performance Research Paper Proceedings (39th International Conference), pp.1-12, IEEE

Kozyreva A, Shingles L, Baklanov P, Mironov A, Schneider FRN (2024). Primary and secondary source of energy in the superluminous supernova 2018ibb. A&A 689:A60

Kozyreva A, Morán-Fraile J, Holas A, Bronner VA, Röpke FK, Pavlyuk N, Mironov A, Tsvetkov D (2024). Thermonuclear explosions as Type II supernovae. A&A 684:A97

Kuruwita RL, Federrath C, Kounkel M (2024). Protostellar spin-up and fast rotator formation through binary star formation. A&A 690:A272

Lau MYM, Hirai R, Mandel I, Tout CA (2024). Expansion of Accreting Main-sequence Stars during Rapid Mass Transfer. ApJL 966(1):L7

Lehmann DH, Gomes B, Vetter N, Braun O, Amr A, Hilbel T, Müller J, Köthe U, Reich C, Kayvanpour E, Sedaghat-Hamedani F, Meder M, Haas J, Ashley E, Rottbauer W, Felbel D, Bekeredjian R, Mahrholdt H, Keller A, Ong P, Seitz A, Hund H, Geis N, André F, Engelhardt S, Katus HA, Frey N, Heuveline V, Meder B (2024). Prediction of diagnosis and diastolic filling pressure by AI-enhanced cardiac MRI: a modelling study of hospital data. The Lancet Digital Health 6(6):e407-e417

Leidi G, Andrassy R, Barsukow W, Higl J, Edelmann PVF, Röpke FK (2024). Performance of high-order Godunov-type methods in simulations of astrophysical low Mach number flows. A&A 686:A34

Leiter C, Lertvittayakumjorn P, Fomicheva M, Zhao W, Gao Y, Eger S (2024). Towards explainable evaluation metrics for machine translation. Journal of Machine Learning Research, 25(75), pp.1-49

Li S, Sanhueza P, Beuther H, Chen H-RV, Kuiper R, Olguin FA, Pudritz RE, Stephens IW, Zhang Q, Nakamura F, Lu X, Kuruwita RL, Sakai T, Henning T, Taniguchi K, Li F (2024). Observations of high-order multiplicity in a high-mass stellar protocluster. Nat Astron 8, 472–481

Lioutas G, Bauswein A, Soultanis T, Pakmor R, Springel V, Röpke FK (2024). General relativistic moving-mesh hydrodynamic simulations with AREPO and applications to neutron star mergers. Monthly Notices of the Royal Astronomical Society 528(2):1906-1929

Liu W, Wan S, Strube M (2024). What causes the failure of explicit to implicit discourse relation recognition? Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pp.2738–2753, Mexico City, Mexico. Association for Computational Linguistics

Lopez VK, Cramer EY, Pagano R, Drake JM, O'Dea EB, Adee M, Ayer T, Chhatwal J, Dalgic OO, Ladd MA, Linas BP, Mueller PP, Xiao J, Bracher J, Castro Rivadeneira AJ, Gerding A, Gneiting T, Huang Y, Jayawardena D, Kanji AH, Le K, Mühlemann A, Niemi J, Ray EL, Stark A, Wang Y, Wattanachit N, Zorn MW, Pei S, Shaman J, Yamana TK, Tarasewicz SR, Wilson DJ, Baccam S, Gurung H, Stage S, Suchoski B, Gao L, Gu Z, Kim M, Li X, Wang G, Wang L, Wang Y, Yu S, Gardner L, Jindal S, Marshall M, Nixon K, Dent J, Hill AL, Kaminsky J, Lee EC, Lemaitre JC, Lessler J, Smith CP, Truelove S, Kinsey M, Mullany LC, Rainwater-Lovett K, Shin L, Tallaksen K, Wilson S, Karlen D, Castro L, Fairchild G, Michaud I, Osthus D, Bian J, Cao W, Gao Z, Lavista Ferres J, Li C, Liu T-Y, Xie X, Zhang S, Zheng S, Chinazzi M, Davis JT, Mu K,

Pastore y Piontti A, Vespignani A, Xiong X, Walraven R, Chen J, Gu Q, Wang L, Xu P, Zhang W, Zou D, Gibson GC, Sheldon D, Srivastava A, Adiga A, Hurt B, Kaur G, Lewis B, Marathe M, Peddireddy AS, Porebski P, Venkatramanan S, Wang L, Prasad PV, Walker JW, Webber AE, Slayton RB, Biggerstaff M, Reich NG, Johansson MA (2024). Challenges of COVID-19 case forecasting in the US, 2020–2021. PLOS Computational Biology, 20(5):e1011200

Löbe M, Hu X, Klopfenstein SAI (2024). Using Health IT Standards to Facilitate Metadata Exchange Between Research Data Management Systems. Stud Health Technol Inform. 316:1413-1417

Ma X, Strube M, Zhao W (2024). Graph-based clustering for detecting semantic change across time and languages. Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics, St. Julians, Malta,17-22 March, pp.1542-1561

Philipp Mainik, Camilo Aponte-Santamaría, Magdalena Fladung, Ronald Ernest Curticean, Irene Wacker, Götz Hofhaus, Martin Bastmeyer, Rasmus R. Schröder, Frauke Gräter, Eva Blasco (2024). Responsive 3D Printed Microstructures Based on Collagen Folding and Unfolding. Small, Volume 21, issue 3, 2408597

Maltsev K, Schneider FRN, Röpke FK, Jordan AI, Qadir GA, Kerzendorf WE, Riedmiller K, van der Smagt P (2024). Scalable stellar evolution forecasting. Deep learning emulation versus hierarchical nearest-neighbor interpolation. A&A 681:A86

Marra R, Churchill CW, Kacprzak GG, Nielsen NM, Trujillo-Gomez S, Lewis EA (2024). Examining quasar absorption-line analysis methods: the tension between simulations and observational assumptions key to modelling clouds. Monthly Notices of the Royal Astronomical Society 527(4):10522-10537

Mayer G, Golebiewski M (2024). Standardization landscape, needs and gaps for the virtual human twin (VHT). Zenodo. https://doi.org/10.5281/zenodo.14386310

Mayer G, Golebiewski M (2024). EDITH standards implementation guide (IG). Zenodo. https://doi.org/10.5281/zenodo.14386453

Memmesheimer P, Machmeier S, Heuveline V (2024). Increasing Detection Rate for Imbalanced Malicious Traffic using Generative Adversarial Networks. European Interdisciplinary Cybersecurity Conference, pp.74-81, ACM

Miller JM, Holladay DA, Peterson JH, Mauney CM, Berger R, Graham AP, Tsai KC, Barker B, Holas A, Mattsson AE, Gogilashvili M, Dolence JC, Meyer CD, Swaminarayan S, Junghans C (2024). Singularity-EOS: Performance Portable Equations of State and Mixed Cell Closures. JOSS 9(103):6805

Mockert F, Grams CM, Lerch S, Osman M, Quinting J (2024). Multivariate post-processing of probabilistic sub-seasonal weather regime forecasts. Quarterly Journal of the Royal Meteorological Society,150(765):4771–4787

Morán-Fraile J, Röpke FK, Pakmor R, Aloy MA, Ohlmann ST, Schneider FRN, Leidi G, Lioutas G (2024). Self-consistent magnetohydrodynamic simulation of jet launching in a neutron star - white dwarf merger. A&A 681:A41

Morán-Fraile J, Holas A, Röpke FK, Pakmor R, Schneider FRN (2024). Faint calcium-rich transient from a double detonation of a 0.6 M⊙ carbon-oxygen white dwarf star. A&A 683:A44

Noll A, Basu S, Hekker S (2024). Effect of nuclear reactions rates and core boundary mixing on the seismology of red clump stars. A&A 683:A189

Pozo Nuñez F, Czerny B, Panda S, Kovacevic A, Brandt W, Horne K (2024). Reevaluating LSST's Capability for Time Delay Measurements in Quasar Accretion Disks. Res. Notes AAS 8(2):47

Ono M, Nozawa T, Nagataki S, Kozyreva A, Orlando S, Miceli M, Chen K-J (2024). The impact of effective matter mixing based on three-dimensional hydrodynamical models on the molecule formation in the ejecta of SN 1987A. ApJS 271:33

Owusu EK, Buder S, Ruiter AJ, Seitenzahl IR, Rodriguez-Segovia N (2024). At the same age, metallicity, and alpha-enhancement, sodium is a more effective tracer of the young and old sequences of the Milky Way disc, Publ. Astron. Soc. Aust. 41,e092

Paiardi G, Milanesi M, Ferraz M, Zimmermann L, Urbinati C, Oreste P, Caccuri F, Chlanda P, Rusnati M, Wade RC (2024). The role of heparin in spike SARS-CoV-2 infection: From a model for heparan sulfates to a starting structure for antivirals. Biophysical Journal 123(3):157a-158a

Paiardi G, Ferraz M, Rusnati M, Wade RC (2024). The accomplices: Heparan sulfates and N-glycans foster SARS-CoV-2 spike:ACE2 receptor binding and virus priming. Proc. Natl. Acad. Sci. U.S.A. 121(43),e2404892121

Pakmor R, Pelisoli I, Justham S, Rajamuthukumar AS, Röpke FK, Schneider FRN, de Mink SE, Ohlmann ST, Podsiadlowski P, Morán-Fraile J, Vetter M, Andrassy R (2024). Large-scale ordered magnetic fields generated in mergers of helium white dwarfs. A&A 691:A179

Pakmor R, Seitenzahl IR, Ruiter AJ, Sim SA, Röpke FK, Taubenberger S, Bieri R, Blondin S (2024). Type Ia supernova explosion models are inherently multidimensional. A&A 686:A227

Pigeot I, Ahrens W, Darms J, Fluck J, Golebiewski M, Hahn HK, Hu X, Intemann T, Kasbohm E, Kirsten T, Klammt S, Klopfenstein SAI, Lassen-Schmidt B, Peters M, Sax U, Waltemath D, Schmidt CO (2024). Making Epidemiological and Clinical Studies FAIR Using the Example of COVID-19. Datenbank Spektrum 24(2):117-128

Pollin JM, Sim SA, Pakmor R, Callan FP, Collins CE, Shingles LJ, Röpke FK, Srivastav S (2024). On the fate of the secondary white dwarf in double-degenerate double-detonation Type Ia supernovae - II. 3D synthetic observables. Monthly Notices of the Royal Astronomical Society 533(3):3036-3052

Polsterer K, Doser B, Fehlner A, Trujillo-Gomez S (2024). Spherinator and HiPSter: Representation Learning for Unbiased Knowledge Discovery from Simulations. arXiv:2406.03810

Prokhorov DA, Lee S-H, Nagataki S, Vink J, Ellison D, Ferrand G, Patnaude D, Roepke F, Seitenzahl I, Slane P, Cta, Abe K, Abe S, Acharyya A, Adam R, Aguasca-Cabot A, Agudo I, Alfaro J, Alvarez-Crespo N, Alves Batista R, Amans J-P, Amato E, Ambrosino F, Angüner EO, Antonelli LA, Aramo C, Arcaro C, Arrabito L, Asano K, Aschersleben J, Ashkar H, Augusto Stuani L, Baack D, Backes M, Balazs C, Balbo M,

Baquero Larriva A, Barbosa Martins V, Barres De Almeida U, Barrio JA, Bastieri D, Batista PI, Batković I, Batzofin RW, Baxter JR, Beck G, Becker Tjus J, Beiske L, Belardinelli D, Benbow W, Bernardini E, Bernete Medrano J, Bernlöhr K, Berti A, Beshley V, Bhattacharjee P, Bhattacharyya S, Bi B, Biederbeck N, Biland A, Bissaldi E, Blanch O, Blazek J, Boisson C, Bolmont J, Bonnoli G, Bordas P, Bosnjak Z, Bradascio F, Braiding C, Bronzini E, Brose R, Brown AM, Brun F, Brunelli G, Bulgarelli A, Burelli I, Burmistrov L, Burton M, Bylund T, Calisse PG, Campoy-Ordaz A, Cantlay BK, Capalbi M, Caproni A, Capuzzo-Dolcetta R, Carlile C, Caroff S, Carosi A, Carosi R, Carrasco M-S, Cascone E, Cassol F, Castrejon N, Catalani F, Cerasole D, Cerruti M, Chaty S, Chen AW, Chernyakova M, Chiavassa A, Chudoba J, Coimbra Araujo CH, Conforti V, Conte F, Contreras JL, Cossou C, Costa A, Costantini H, Cristofari P, Cuevas O, Curtis-Ginsberg Z, D'Amico G, D'Ammando F, Dadina M, Dalchenko M, David L, Davids ID, Dazzi F, De Angelis A, De Bony De Lavergne M, De Caprio V, De Cesare G, De Gouveia Dal Pino EM, De Lotto B, De Lucia M, De Menezes R, De Naurois M, De Oña Wilhelmi E, De Simone N, De Souza V, Del Peral L, Del Valle MV, Delagnes E, Delgado Giler AG, Delgado C, Dell'aiera M, Della Ceca R, Della Valle M, Della Volpe D, Depaoli D, Dettlaff A, Di Girolamo T, Di Piano A, Di Pierro F, Di Tria R, Di Venere L, Díaz-Bahamondes C, Dib C, Diebold S, Dima R, Dinesh A, Djannati-Ataï A, Djuvsland JI, Dominguez A, Dominik RM, Donini A, Dorner D, Dörner J, Doro M, Dos Anjos RDC, Dournaux J-L, Dravins D, Duangchan C, Dubos C, Ducci L, Dwarkadas VV, Ebr J, Eckner C, Egberts K, Einecke S, Elsässer D, Emery G, Escobar Godoy M, Escudero J, Esposito P, Falceta-Gonçalves D, Fallah Ramazani V, Faure A, Fedorova E, Fegan S, Feijen K, Feng Q, Ferrarotto F, Fiandrini E, Fiasson A, Fioretti V, Foffano L, Font Guiteras L, Fontaine G, Fröse S, Fukami S, Fukui Y, Funk S, Gaggero D, Galanti G, Galaz G, Gallant YA, Gallozzi S, Gammaldi V, Gasbarra C, Gaug M, Ghalumyan A, Gianotti F, Giarrusso M, Giglietto N, Giordano F, Giuliani A, Glicenstein J-F, Glombitza J, Goldoni P, González JM, González MM, Goulart Coelho J, Granot J, Grasso D, Grau Haro R, Green D, Green JG, Greenshaw T, Grolleron G, Grube J, Gueta O, Gunji S, Hadasch D, Hamal P, Hanlon W, Hara S, Harvey VM, Hashiyama K, Hassan T, Heller M, Hernández Cadena S, Hie J, Hiroshima N, Hnatyk B, Hnatyk R, Hoffmann D, Hofmann W, Holler M, Horan D, Horvath P, Hovatta T, Hrupec D, Hussain S, Iarlori M, Inada T, Incardona F, Inome Y, Inoue S, Iocco F, Ishio K, Jamrozy M, Janecek P, Jankowsky F, Jarnot C, Jean P, Jiménez Martínez I, Jin W, Jocou L, Juramy-Gilles C, Jurysek J, Kalekin O, Kantzas D, Karas V, Kaufmann S, Kerszberg D, Khelifi B, Kieda DB, Kleiner TK, Kluźniak W, Kobayashi Y, Kohri K, Komin N, Kornecki P, Kosack K, Kubo H, Kushida J, La Barbera A, La Palombara N, Lainez M, Lamastra A, Lapington JS, Lazarevic S, Lazendic-Galloway J, Leach S, Lemoine-Goumard M, Lenain J-P, Leto G, Leuschner F, Lindfors E, Linhoff M, Liodakis I, Loïc L, Lombardi S, Longo F, López-Coto R, López-Moya M, López-Oramas A, Loporchio S, Lozano Bahilo J, Luque-Escamilla PL, Macias O, Maier G, Majumdar P, Malyshev D, Malyshev D, Mandat D, Manicò G, Marinos PD, Markoff S, Márquez I, Marquez P, Marsella G, Martí J, Martin P, Martínez GA, Martínez M, Martinez O, Marty C, Mas-Aguilar A, Mastropietro M, Maurin G, Max-Moerbeck W, Mazin D, Melkumyan D, Menchiari S, Mestre E, Meunier J-L, Meyer DM-A, Miceli D, Michailidis M, Michałowski J, Miener T, Miranda JM, Mitchell A, Mizote M, Mizuno T, Moderski R, Mohrmann L, Molero M, Molfese C, Molina E, Montaruli T, Moralejo A, Morcuende D, Morik K, Morselli A, Moulin E, Moya Zamanillo V, Mukherjee R, Munari K, Muraczewski A, Muraishi H, Nakamori T, Nava L, Nayak A, Nemmen R, Nickel L, Niemiec J, Nieto D, Nievas Rosillo M, Nikołajuk M, Nishijima K, Noda K, Nosek D, Novosyadlyj B, Novotný V, Nozaki S, O'Brien P, Ohishi M, Ohtani Y, Okumura A, Olive J-F, Olmi B, Ong RA, Orienti M, Orito R, Orlandini M, Orlando E, Ostrowski M, Otte N, Oya I, Pagano I, Pagliaro A, Palatiello M, Panebianco G, Paredes JM, Parmiggiani N, Patel SR, Patricelli B, Pavlović D, Pe'er A, Pech M, Pecimotika M, Peresano M, Pérez-Romero J, Peron G, Persic M, Petrucci P-O, Petruk O, Pfeifle F, Pintore F, Pirola G, Pittori C, Plard C, Podobnik F, Pohl M, Pons E, Prandini E, Prast J, Principe G, Priyadarshi C, Produit N, Pueschel E, Pühlhofer G, Pumo ML, Punch M, Quirrenbach A, Raino S, Randazzo N, Rando R, Ravel T, Razzaque S, Regeard M, Reichherzer P, Reimer A, Reimer O, Reisenegger A, Reposeur T, Reville B, Rhode W, Ribó M, Richtler T, Rieger F, Roache E, Rodriguez Fernandez G, Rodríguez Frías MD, Rodríguez-Vázquez JJ, Romano P, Romeo G, Rosado J, Rowell GP, Rudak B, Ruiter AJ, Rulten CB, Russo F, Sadeh I, Saha L, Saito T, Sakurai S, Salzmann H, Sanchez D, Sanchez-Conde M, Sangiorgi P, Sano H, Santander M, Santangelo A, Santos-Lima R, Sanuy A, Šarić T, Sarkar A, Sarkar S, Saturni FG, Savchenko V, Scherer A, Schipani P, Schleicher B, Schovanek P, Schubert JL, Schüssler F, Schwanke U, Schwefer G, Scuderi S, Seglar Arroyo M, Sergijenko O, Sguera V, Shang RY, Sharma P, Sidibe GDS, Sidoli L, Siejkowski H, Siqueira C, Sizun P, Sliusar V, Slowikowska A, Sol H, Specovius A, Spencer ST, Spiga D, Stamerra A, Stanič S, Starecki T, Starling R, Steppa C, Stolarczyk T, Strišković J, Strzys MC, Suda Y, Suomijarvi T, Tak D, Takahashi M, Takeishi R, Tam P-HT, Tanaka SJ, Tanaka T, Terauchi K, Testa V, Tibaldo L, Tibolla O, Torradeflot F, Torres DF, Torresi E, Tothill N, Toussenel F, Touzard V, Tramacere A, Travnicek P, Tripodo G, Truzzi S, Tsiahina A, Tutone A, Vacula M, Vallage B, Vallania P, Vallés R, Van Eldik C, Van Scherpenberg J, Vandenbroucke J, Vassiliev V, Venault P, Ventura S, Vercellone S, Verna G, Viana A, Viaux N, Vigliano A, Vignatti J, Vigorito CF, Vitale V, Vodeb V, Voisin V, Vorobiov S, Voutsinas GG, Vovk I, Waegebaert V, Wagner SJ, Walter R, Ward M, Wechakama M, White R, Wierzcholska A, Will M, Williams DA, Wohlleben F, Wolter A, Yamamoto T, Yamazaki R, Yang L, Yoshida T, Yoshikoshi T, Zacharias M, Zanmar Sanchez R, Zavrtanik D, Zavrtanik M, Zdziarski AA, Zech A, Zhdanov VI, Ziętara K, Živec M, Zuriaga-Puig J (2024) Potential of the Cherenkov Telescope Array for studying the young nearby supernova remnants. In 38th International Cosmic Ray Conference, p. 824

Prust LJ, Glanz H, Bildsten L, Perets HB, Röpke FK (2024). Morphology and Mach Number Dependence of Subsonic Bondi–Hoyle Accretion. ApJ 966(1):103

Richter S, Jin F, Ritterhoff T, Fergin A, Maurer E, Frank A, Hajnal A, Klevit R, Gräter F, Flotho A, Melchior F (2024). SUMO›s intrinsically disordered N-terminus is an intramolecular inhibitor of SUMO - SIM interactions. Biorxiv. https://doi.org/10.1101/2024.01.04.574143

Riedmiller K, Reiser P, Bobkova E, Maltsev K, Gryn'ova G, Friederich P, Gräter F (2024). Substituting density functional theory in reaction barrier calculations for hydrogen atom transfer in proteins. Chem. Sci. 15(7):2518-2527

Roberts JD, Pinsonneault MH, Johnson JA, Zinn JC, Weinberg DH, Vrard M, Tayar J, Stello D, Mosser B, Johnson JW, Cao K, Stassun KG, Stringfellow GS, Serenelli A, Mathur S, Hekker S, García RA, Elsworth YP, Corsaro E (2024). Nature versus nurture: distinguishing effects from stellar processing and chemical evolution on carbon and nitrogen in red giant stars. Monthly Notices of the Royal Astronomical Society 530(1):149-166

Roessner R, Michelarakis N, Gräter F, Aponte-Santamaria C (2024). Mechanical control of the adhesion of malaria-infected erythrocytes to the placenta. Biophysical Journal,123(3):197a

Roller J, Schmid V, Gerstner P, Niessner J, Heuveline V (2024). Using CutFEM for solving the linearized Poisson-Boltzmann equation: Introductory application in HiFlow3

Ruff R, Reiser P, Stühmer J, Friederich P (2024). Connectivity optimized nested line graph networks for crystal structures. Digital Discovery 3(3):594–601

Röpke F (2024). Jet formation in stellar interactions. In EAS, European Astronomical Society Annual Meeting, p.2451

Schimunek J, Seidl P, Elez K, Hempel T, Le T, Noé F, Olsson S, Raich L, Winter R, Gokcan H, Gusev F, Gutkin EM, Isayev O, Kurnikova MG, Narangoda CH, Zubatyuk R, Bosko IP, Furs KV, Karpenko AD, Kornoushenko YV, Shuldau M, Yushkevich A, Benabderrahmane MB, Bousquet-Melou P, Bureau R, Charton B, Cirou BC, Gil G, Allen WJ, Sirimulla S, Watowich S, Antonopoulos N, Epitropakis N, Krasoulis A, Pitsikalis V, Theodorakis S, Kozlovskii I, Maliutin A, Medvedev A, Popov P, Zaretckii M, Eghbal-Zadeh H, Halmich C, Hochreiter S, Mayr A, Ruch P, Widrich M, Berenger F, Kumar A, Yamanishi Y, Zhang KYJ, Bengio E, Bengio Y, Jain MJ, Korablyov M, Liu C, Marcou G, Glaab E, Barnsley K, Iyengar SM, Ondrechen MJ, Haupt VJ, Kaiser F, Schroeder M, Pugliese L, Albani S, Athanasiou C, Beccari A, Carloni P, D'Arrigo G, Gianquinto E, Goßen J, Hanke A, Joseph BP, Kokh DB, Kovachka S, Manelfi C, Mukherjee G, Muñiz-Chicharro A, Musiani F, Nunes-Alves A, Paiardi G, Rossetti G, Sadiq SK, Spyrakis F, Talarico C, Tsengenes A, Wade RC, Copeland C, Gaiser J, Olson DR, Roy A, Venkatraman V, Wheeler TJ, Arthanari H, Blaschitz K, Cespugli M, Durmaz V, Fackeldey K, Fischer PD, Gorgulla C, Gruber C, Gruber K, Hetmann M, Kinney JE, Padmanabha Das KM, Pandita S, Singh A, Steinkellner G, Tesseyre G, Wagner G, Wang Z, Yust RJ, Druzhilovskiy DS, Filimonov DA, Pogodin PV, Poroikov V, Rudik AV, Stolbov LA, Veselovsky AV, De Rosa M, De Simone G, Gulotta MR, Lombino J, Mekni N, Perricone U, Casini A, Embree A, Gordon DB, Lei D, Pratt K, Voigt CA, Chen K, Jacob Y, Krischuns T, Lafaye P, Zettor A, Rodríguez ML, White KM, Fearon D, Von Delft F, Walsh MA, Horvath D, Brooks CL III, Falsafi B, Ford B, García-Sastre A, Yup Lee S, Naffakh N, Varnek A, Klambauer G, Hermans TM (2024). A community effort in SARS-CoV-2 drug discovery. Molecular Informatics 43(1), e202300262

Schneider FRN, Podsiadlowski Ph, Laplace E (2024). Pre-supernova evolution and final fate of stellar mergers and accretors of binary mass transfer. A&A 686:A45

Schröder M, Machmeier S, Maeng S, Heuveline V (2024). Validating CESU-8 Encoded Text Utilising SIMD Instructions. Proceedings of the 13th International Conference on Software and Computer Applications, pp.102-111, ACM

Schulze S, Fransson C, Kozyreva A, Chen T-W, Yaron O, Jerkstrand A, Gal-Yam A, Sollerman J, Yan L, Kangas T, Leloudas G, Omand CMB, Smartt SJ, Yang Y, Nicholl M, Sarin N, Yao Y, Brink TG, Sharon A, Rossi A, Chen P, Chen Z, Cikota A, De K, Drake AJ, Filippenko AV, Fremling C, Fréour L, Fynbo JPU, Ho AYQ, Inserra C, Irani I, Kuncarayakti H, Lunnan R, Mazzali P, Ofek EO, Palazzi E, Perley DA, Pursiainen M, Rothberg B, Shingles LJ, Smith K, Taggart K, Tartaglia L, Zheng W, Anderson JP, Cassara L, Christensen E, George Djorgovski S, Galbany L, Gkini A, Graham MJ, Gromadzki M, Groom SL, Hiramatsu D, Andrew Howell D, Kasliwal MM, McCully C, Müller-Bravo TE, Paiano S, Paraskeva E, Pessi PJ, Polishook D, Rau A, Rigault M, Rusholme B (2024). 1100 days in the life of the supernova 2018ibb. The best pair-instability supernova candidate, to date. A&A 683:A223

Seute L, Hartmann E, Stühmer J, Gräter F (2024). Grappa – A Machine Learned Molecular Mechanics Force Field. In ICML AI for Science Workshop

Shenar T, Bodensteiner J, Sana H, Crowther PA, Lennon DJ, Abdul-Masih M, Almeida LA, Backs F, Berlanas SR, Bernini-Peron M, Bestenlehner JM, Bowman DM, Bronner VA, Britavskiy N, de Koter A, de Mink SE, 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, Janssens S, Johnston C, Josiek J, Justham S, Kalari VM, Katabi ZZ, Keszthelyi Z, Klencki J, Kubát J, Kubátová B, Langer N, Lefever RR, Ludwig B, Mackey J, Mahy L, Maíz Apellániz J, Mandel I, Maravelias G, Marchant P, Menon A, Najarro F, Oskinova LM, O'Grady AJG, Ovadia R, Patrick LR, Pauli D, Pawlak M, Ramachandran V, Renzo M, Rocha DF, Sander AAC, Sayada T, Schneider FRN, Schootemeijer A, Schösser EC, Schürmann C, Sen K, Shahaf S, Simón-Díaz S, Stoop M, Toonen S, Tramper F, van Loon JTh, Valli R, van Son LAC, Vigna-Gómez A, Villaseñor JI, Vink JS, Wang C, Willcox R (2024). Binarity at LOw Metallicity (BLOeM): A spectroscopic VLT monitoring survey of massive stars in the SMC. A&A 690:A289

Siess L, Bermúdez-Bustamante LC, De Marco O, Price DJ, González-Bolívar M, Mu C, Lau MYM, Hirai R, Danilovich T (2024). Dusty Common Envelope Evolution. Galaxies 12(6):82

Sirazitdinov A, Buchwald M, Heuveline V, Hesser J (2024). Graph Neural Networks for Individual Treatment Effect Estimation. IEEE Access 12:106884-106894

Sokolov M, Hoffmann DS, Dohmen PM, Krämer M, Höfener S, Kleinekathöfer U, Elstner M (2024). Non-adiabatic molecular dynamics simulations provide new insights into the exciton transfer in the Fenna–Matthews–Olson complex. Phys. Chem. Chem. Phys., ,26, 19469-19496

Song M, Yang D, Lerch S, Xia X, Yagli GM, Bright JM, Shen Y, Liu B, Liu X, Mayer MJ (2024). Non-crossing quantile regression neural network as a calibration tool for ensemble weather forecasts. Advances in Atmospheric Sciences, 41(7):1417–1437

Stiller J, Feng S, Chowdhury A-A, Rivas-González I, Duchêne DA, Fang Q, Deng Y, Kozlov A, Stamatakis A, Claramunt S, Nguyen JMT, Ho SYW, Faircloth BC, Haag J, Houde P, Cracraft J, Balaban M, Mai U, Chen G, Gao R, Zhou C, Xie Y, Huang Z, Cao Z, Yan Z, Ogilvie HA, Nakhleh L, Lindow B, Morel B, Fjeldså J, Hosner PA, da Fonseca RR, Petersen B, Tobias JA, Székely T, Kennedy JD, Reeve AH, Liker A, Stervander M, Antunes A, Tietze DT, Bertelsen MF, Lei F, Rahbek C, Graves GR, Schierup MH, Warnow T, Braun EL, Gilbert MTP, Jarvis ED, Mirarab S, Zhang G (2024). Complexity of avian evolution revealed by family-level genomes. Nature, 629, 851–860 (2024)

Taffoni G, Mignone A, Tornatore L, Sciacca E, Guarrasi M, Lapenta G, Riha L, Vavrik R, Vysocky O, Kadlubiak K, Strakos P, Jaros M, Dolag K, Commercon B, Rezzolla L, Pierre K, Doulis G, Shen S, Marazakis M, Gregori D, Boella E, Perna G, Zanotti M, Raffin E, Polsterer K, Trujillo Gomez S, Marin G (2024). Redesign of astrophysical codes for exascale computing: the SPACE experience. Software and Cyberinfrastructure for Astronomy VIII, p.32,SPIE

Taha D, Zhao W, Riestenberg JM, Strube M (2024). Normed spaces for graph embedding. Transactions of Machine Learning Research. arXiv:2312.01502

Temaj D, Schneider FRN, Laplace E, Wei D, Podsiadlowski Ph (2024). Convective-core overshooting and the final fate of massive stars. A&A 682:A123

Togkousidis A, Stamatakis A, Gascuel O (2024). Much Ado About Nothing: Accelerating Maximum Likelihood Phylogenetic Inference via Early Stopping to evade (Over-)optimization. Biorxiv. https://doi.org/10.1101/2024.07.04.602058

Tuhtan V, Al-Belmpeisi R, Christensen MB, Kuruwita R, Haugbølle T (2024). Simulated analogues I: apparent and physical evolution of young binary protostellar systems. Monthly Notices of the Royal Astronomical Society 534(4):3176-3193

Uhl TN, Schimek M, Hübner L, Hespe D, Kurpicz F, Seemaier D, Stelz C, Sanders P (2024). KaMPIng: Flexible and (Near) Zero-Overhead C++ Bindings for MPI. SC24: International Conference for High Performance Computing, Networking, Storage and Analysis,pp.1-21, IEEE

Vantyghem AN, Galvin TJ, Sebastian B, O'Dea CP, Gordon YA, Boyce M, Rudnick L, Polsterer K, Andernach H, Dionyssiou M, Venkataraman P, Norris R, Baum SA, Wang XR, Huynh M (2024). Rotation and flipping invariant self-organizing maps with astronomical images: A cookbook and application to the VLA Sky Survey QuickLook images. Astronomy and Computing 47:100824

Vetter M, Röpke FK, Schneider FRN, Pakmor R, Ohlmann ST, Lau MYM, Andrassy R (2024). From spherical stars to disk-like structures: 3D common-envelope evolution of massive binaries beyond inspiral. A&A 691:A244

Wade RC (2024). Editorial: The Journal of Molecular Recognition: Changing of the guard. J of Molecular Recognition 37(3),e3081

Walz E-M, Henzi A, Ziegel J, Gneiting T (2024a). Easy uncertainty quantification (EasyUQ): Generating predictive distributions from single-valued model output. SIAM Review, 66(1):91–122

Walz E-M, Knippertz P, Fink AH, Köhler G, Gneiting T (2024b). Physics-based vs. data-driven 24-hour probabilistic forecasts of precipitation for northern tropical Africa. Monthly Weather Review,152(9):2011–2031

Wang C, Bodensteiner J, Xu X-T, de Mink SE, Langer N, Laplace E, Vigna-Gómez A, Justham S, Klencki J, Olejak A, Valli R, Schootemeijer A (2024). Stripped Helium Star and Compact Object Binaries in Coeval Populations: Predictions Based on Detailed Binary Evolution Models. ApJL 975(1):L20

Wei D, Schneider FRN, Podsiadlowski P, Laplace E, Röpke FK, Vetter M (2024). Evolution and final fate of massive post-common-envelope binaries. A&A 688:A87

Weidner P, Saar D, Söhn M, Schroeder T, Yu Y, Zöllner FG, Ponelies N, Zhou X, Zwicky A, Rohrbacher FN, Pattabiraman VR, Tanriver M, Bauer A, Ahmed H, Ametamey SM, Riffel P, Seger R, Bode JW, Wade RC, Ebert MPA, Kragelund BB, Burgermeister E (2024). Myotubularin-related-protein-7 inhibits mutant (G12V) K-RAS by direct interaction. Cancer Letters 588:216783

Wiegert J, Höhler D, Haag J, Stamatakis A (2024). Predicting Phylogenetic Bootstrap Values via Machine Learning. Biorxiv. https://doi.org/10.1101/2024.03.04.583288

Zaunseder E, Mütze U, Okun JG, Hoffmann GF, Kölker S, Heuveline V, Thiele I (2024). Personalized metabolic whole-body models for newborns and infants predict growth and biomarkers of inherited metabolic diseases. Cell Metabolism 36(8):1882-1897.e7

Zaunseder E, Mohammad FK, Mütze U, Kölker S, Heuveline V, Thiele I (2024). Dynamic whole-body models for infant metabolism. Biorxiv. https://doi.org/10.1101/2024.11.25.625291

Zaunseder E, Teinert J, Boy N, Garbade SF, Haupt S, Feyh P, Hoffmann GF, Kölker S, Mütze U, Heuveline V (2024). Digital-Tier Strategy Improves Newborn Screening for Glutaric Aciduria Type 1. IJNS 10(4):83

9 Teaching

Degrees

Benjamin Barth:
"Optimizing for Hyperparameters: Constraint-Adaptive Sharpness-Aware Minimization and Its Variants", Master's thesis, Karlsruhe Institute of Technology and HITS: Jan Stühmer (2024).

Sophia Ber:
"Investigation of the impact of mutations in Cytochrome P450 reductase on the rate of electron transfer to Cytochrome P450 by molecular dynamics simulations", Master's thesis, Chemistry, Faculty of Chemistry and Geosciences, Heidelberg University and HITS: Jonathan Teuffel and Rebecca C. Wade (2024).

Maximilian Bing:
"Continuous Modeling of Extracellular Matrix Invasion by Tumor Growth", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2024).

Erik Borker:
"Effects of Systematic Errors on Phylogenetic Inference", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2024).

Johannes Hengstler:
"Re-Engineering Genomic Tree Sequence Inference Algorithms", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2024).

Saber Boushehri:
"Effect of O-glycans on the Structure and Viscosity of Intrinsically Disordered and Glycosylated Protein Lubricin", PhD thesis, Department of Engineering Sciences, Heidelberg University and HITS: Frauke Gräter (2024).

Johanna Buck:
"Variously Cross-linked Martini3-Coarse-Grained Collagen Fibrils Under Force", Master's thesis,, Department of Physics and Astronomy, Heidelberg University and HITS: Frauke Gräter (2024).

Svenja de Buhr:
"Functional Dynamics and Mechano-Activation of c-Abl Kinase", PhD Thesis, Department of Biological Sciences, Heidelberg University and HITS: Frauke Gräter (2024).

Max Heller:
"One dimensional merger models for massive stars", Master's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Fabian Schneider (2024).

Anastasiya Kapinskaya:
"Asteroseismologie von dem Roten Riesen KIC 9145955: Untersuchung von potenziellen Methoden für die detaillierte Modelierung", Master's thesis, Fakultät für Physik und Astronomie, Ruprecht-Karls-Universität Heidelberg and HITS: Saskia Hekker (2024).

Jonathan Kunz:
"Structural bioinformatics approach to exploring transient binding pockets in proteins". Bachelor's thesis. Biochemistry, Faculty of Bioscience and Faculty of Chemistry and Geosciences, Heidelberg University and HITS: Stefan Richter and Rebecca Wade (2024).

Eric Laudemann:
"Validation and Cross-Species Comparison of scRNA-seq Trajectory Inference Algorithms", Master's thesis, Karlsruhe Institute of Technology and HITS and EMBL: Alexandros Stamatakis, (2024).

David Lehmann:
"Prediction of ventricular pressure and diagnosis from cardiac magnetic resonance imaging using artificial neural networks", Ph.D. thesis, Heidelberg University and HITS: Vincent Heuveline (2024).

Giovanni Leidi:
"Simulating magnetohydrodynamic phenomena in the interior of stars," Ph.D. thesis, Combined Faculty of Natural Sciences, Engineering and Mathematics, Heidelberg University and HITS: Friedrich Röpke (2024).

Maximilian Ludwig:
"Evaluation, benchmarking, and ensembling of Intrusion Detection Systems using a comparative platform", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2024).

Kiril Maltsev
"Statistical modeling of the progenitor evolution and formation of neutron stars and stellar-mass black holes," Ph.D. thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Friedrich Röpke (2024).

Luc Mercatoris:
"Quantification and Prediction of Multiple Sequence Alignment Difficulty", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2024).

Johanna Riedel:
"Probabilistic Reconstruction of Stellar Spectra from Photometry", Bachelor's thesis, Institute for Computer Science, Heidelberg University (Fred Hamprecht, 2024), co-supervised by Kai Lars Polsterer and Nikos Gianniotis (2025).

Kai Riedmiller:
"Predicting Hydrogen Atom Transfer in Collagen", PhD thesis, Department of Engineering Sciences, Heidelberg University and HITS: Frauke Gräter (2024).

Boris Schüpp:
"Advancing the Study of Ultrasonic Breaking of Nicked Double-Stranded and Single-Stranded DNA: Modeling and Experimental Investigations", Master's thesis, Faculty of Engineering Sciences, Heidelberg University and HITS: Frauke Gräter (2024).

Steven Schürstedt:
"Zuordnung der Autorenschaft bei generativen Modellen am Beispiel klassischer Klaviermusik", Master's thesis, Karlsruhe Institute of Technology and HITS: Jan Stühmer (2024).

Benedict Schuster:
"Four-Species Tumor Growth Model", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2024).

Dominik Siebelt:
"Faster Sorting of Aligned DNA-Read Files", Bachelor's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2024).

Alexander Suhrkamp:
"Improving DNA-Barcoding Databases: A Framework for Error Detection- and Correction in the Barcode of Life Data System (BOLD)", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2024).

David Trajkovski:
"Passwortschemata zur Erstellung sicherer Passwörter allein durch Menschenhand", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2024).

Elefterios Tselegkidis:
"DNS-Angriffe: Erkennung und Umgehung von Anomalieerkennungssystemen", Master's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2024).

Viliuga Vsevolod:
"Introducing flexibility conditioning to protein design", Master's thesis, Department of Biological Sciences, Heidelberg University and HITS: Frauke Gräter (2024).

Eva-Maria Walz:
"Statistical methods for probabilistic forecasts of real-valued outcomes", Ph.D. thesis, Department of Mathematics, Karlsruhe Institute of Technology and HITS: Tilmann Gneiting (2024).

Julius Wiegert:
"Use Cases of Predictive Modeling for Phylogenetic Inference and Placements", Master's thesis, Karlsruhe Institute of Technology and HITS: Alexandros Stamatakis (2024).

Nicolas Wolf:
"Learning to Generate Protein Conformations and Application in Crosslink Design", Master's thesis, Department of Physics and Astronomy, Heidelberg University and HITS: Frauke Gräter (2024).

Elaine Zaunseder:
"New approaches in mathematical and data-based modeling for newborn screening", Ph.D. thesis, Heidelberg University and HITS: Vincent Heuveline (2024). DOI: doi:10.11588/heidok.00035789.

Sven Zelch:
"Reproducible Virtual Private Network Benchmarking using Declarative Environments", Bachelor's thesis, Faculty for Mathematics and Computer Science, Heidelberg University and HITS: Vincent Heuveline (2024).

Lectures, courses and seminars

Camilo Aponte-Santamaría:
Seminar: *"Molecular dynamics and machine learning for the study of biomolecules"*, SSF+UN Seminar, UNAL Bogotá, 8 August (2024).

Camilo Aponte-Santamaría and Frauke Gräter:
Course *"Data Analysis and Simulations"*, Heidelberg University, summer semester 2024.

Camilo Aponte-Santamaría, Frauke Gräter and Rebecca Wade:
Course: *"Computational Molecular Biophysics"*, Heidelberg University, winter semester 2024/2025.

Tommaso Bartoloni, Riccardo Beccaria, Rebecca C. Wade (MCM):
M.Sc. seminar course on *"Machine Learning for the Biomolecular World"*, with Tristan Bereau (Heidelberg University), Heidelberg University, summer semester, 2024.

Beatriz Bordadagua:
Teaching assistant *"Cosmology"*, Heidelberg University, winter semester 2023/24. Teaching assistant *"Astronomical Techniques"*, Heidelberg University, summer semester 2024.

Iliana Isabel Cortés Pérez:
Tutor for the course *"Astro-lab"*, lecturer Prof. Dr. Jochen Heidt, Heidelberg University, winter semester 2024/25.

Francisca Espinoza:
Teaching assistant *"Einführung in die Astronomie I"*, Heidelberg University, winter semester 2023/24. Teaching assistant *"FP30: CCD Photometry in modern Astronomy"*, Heidelberg University, summer semester 2024.

Christina Fakiola:
Tutorial of *"Einführung in die Astronomie und Astrophysik I,"* Heidelberg University, winter semester 2024/25.

Robert Fisher:
Summer School Lectures, *"Hydrodynamics and Ideal Magnetohydrodynamics for Stellar Transients,"* ICESUN Summer School, Kunming, China, August 20, 2024.

Tilmann Gneiting:
Lecture on *"Forecasting: Theory and Practice I"*, Karlsruhe Institute of Technology, winter semester 2024/2025.

Martin Golebiewski and Gerhard Mayer:
Software demonstration of FAIRDOM SEEK, EDITH technical consortium meeting, Athens, Greece, 15 - 17 April 2024.

Frauke Gräter, Camilo Aponte-Santamaria (MBM), Rebecca C. Wade, Stefan Richter, Riccardo Beccaria (MCM):
M.Sc. lecture and practical course on *"Computational Molecular Biophysics"*, Heidelberg University, winter semester, 2024/25.

Fabian Grünewald:
Lectures on *"Thermodynamics for Life Sciences"* within Molecular Biotechnology Bachelor, Heidelberg University, summer semester 2024. Lecture on *"Digital Twins for Nanomedicine"* within Molecular Biotechnology Bachelor, Heidelberg University winter semester 2024/2025.

Saskia Hekker:
Lecture *"Asteroseismology"*, Heidelberg University, winter semester 2023/24. Seminar *"Applications of Asteroseismology"*, Heidelberg University, summer semester 2024. Lecture *"Asteroseismology"*, winter semester 2024/25. Course *"Scientific Writing Basics for IMPRS PhD Students"*, winter semester 2024/25.

Vincent Heuveline:
Lecture on *"IT-Sicherheit 1"*, Heidelberg University, winter semester 2023/2024. Seminar on *"IT-Security"*, Heidelberg University, winter semester 2023/2024. Seminar on *"Special Topics in FEM"*, Heidelberg University, winter semester 2023/2024. Lecture on *"IT-Sicherheit 2"*, Heidelberg University, summer semester 2024. Seminar on *"IT-Security"*, Heidelberg University, summer semester 2024. Lecture on *"IT-Sicherheit 1"*, Heidelberg University, winter semester 2024/2025. Lecture on *"Einführung in die numerische Mathematik"*, Heidelberg University, winter semester 2024/2025.

Xiaoming Hu:
Aligning the FAIR data principles and HL7 FHIR profiling. Semantic Web Applications and Tools for Healthcare and Life Sciences (SWAT4HCLS) 2024, Leiden, Netherlands, 26-29 February 2024. Software demonstration of FAIRDOM SEEK, COMBINE 2024, Stuttgart, Germany, 1 - 4 September 2024.

Xiaoming Hu, Maja Rey, Andreas Weidemann, Ulrike Wittig:
de.NBI Course *"Virtual 2-Day Workshop SABIO-RK and FAIRDOMHub/FAIRDOM-SEEK"*, online, 25-26 November 2024.

Alexander Jordan and Kristof Kraus:
Lecture on *"Time Series Analysis"*, Karlsruhe Institute of Technology, summer semester 2024.

Alexey Kozlov:
"RAxML-NG: Introduction and Tutorial", Phylogenomics 2024 Workshop, Český Krumlov, Czech Republic, January 2024. *"Green Computing"*, Phylogenomics 2024 Workshop, Český Krumlov, Czech Republic, January 2024. Talk slides.

Giovanni Leidi:
Practicals on *"Fundamentals of simulation techniques,"* Heidelberg University, winter semester 2023/24.

Tobias van Lier:
Teaching assistant *"Physik A (Physik für Nebenfächler)"*, Heidelberg University, winter semester 2024/25.

Jonas Müller:
Teaching assistant *"Einführung in die Astronomie und Astrophysik I"*, Heidelberg University, winter semester 2023/24. Teaching assistant *"Einführung in die Astronomie und Astrophysik II"*, Heidelberg University, summer semester 2024.

Wolfgang Müller:
"Rightfield: Semantische Daten-Annotation", NFDI4Biodiversity/SMon cross-community workshop, 22 - 23 January 2024. *"Three useful tools to get data in shape"*, NFDI4Biodiversity iDiv Seasonal School 2024, 2 - 6 December 2024, DOI: 10.5281/zenodo.14269511.

Giulia Paiardi:
"Meet-EU Heidelberg-team", European Bioinformatics Master Network (https://cu-bioinformatics.github.io/meet-eu-2023/), with Carl Herrmann (Heidelberg University), winter semester, 2023/24.

Friedrich Röpke:
Lecture course *"Computational Astrophysics,"* Heidelberg University, summer semester 2024.

Fabian Schneider, Friedrich Röpke:
Lecture course *"The Stellar Cookbook: A practical guide to the theory of stars"*, Heidelberg University, winter semester 2023/2024. Lecture course *"Stars Squared: Evolution of Binary Stars"*, summer semester 2024. Lecture course *"The Stellar Cookbook: A practical guide to the theory of stars"*, Heidelberg University, winter semester 2024/2025.

Alexandros Stamatakis:
Guest Lecture *"Phylogenetic Inference"*, University of Thrace MSc program *"Applied Bioinformatics & Data Analysis"*, Alexandroupoli, Greece, on-line, March 2024. Seminar *"Reproducibility in Bioinformatics"*, computer science Master's program at University of Crete, spring semester, 2024.

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, Dimitri Höhler:
Summer School *"Computational Molecular Evolution"*, Welcome Trust Genome Campus, Hinxton, UK, May 2024.

Alexandros Stamatakis, Alexey Kozlov, Lukas Hübner:
Lecture *"Introduction to Bioinformatics for Computer Scientists"*, computer science Master's programs at Karlsruhe Institute of Technology and the University of Crete, winter semester, 2023/2024.

Alexandros Stamatakis, Anastasis Togkousidis, Alexey Kozlov:
Seminar *"Hot Topics in Bioinformatics"*, computer science Master's program at Karlsruhe Institute of Technology, summer semester, 2024.

Michael Strube:
PhD Colloquium, Department of Computational Linguistics, Heidelberg University (winter semester 2023/2024). PhD Colloquium, Department of Computational Linguistics, Heidelberg University (summer semester 2024).

Jan Stühmer:
Lecture *"Geometric Deep Learning"*, computer science Master's program at Karlsruhe Institute of Technology, winter semester, 2024/2025. Compact course *"Geometric Deep Learning"*, Heidelberg Physics Graduate Days, Heidelberg University, summer semester, 2024. Seminar *"Interpretability and Causality in Machine Learning"*, computer science Master's and Bachelor's program at Karlsruhe Institute of Technology, summer semester, 2024. Lecture *"Geometric Deep Learning"*, computer science Master's program at Karlsruhe Institute of Technology, winter semester, 2023/2024.

Sebastian Trujillo Gomez:
CASTIEL2 Online Training Event: Machine Learning prototype tools for SPACE applications, 7 May 2024.

Marco Vetter:
"Computational Astrophysics" tutorial, Department of Physics and Astronomy, Heidelberg University, summer semester 2024.

Kristián Vitovský:
Exercise group for *"Theoretical Astrophysics (MKTP2),"* taught by Matthias Bartelmann, Heidelberg University, winter semester 2024/25.

Rebecca Wade:
Lecture contributions to the M. Sc. Biochemistry *"Computational Biochemistry"* course (winter semester, 2023/24), M.Sc. Molecular & Cellular Biology Module 4 ("Special Topics") (winter semester, 2023/24), and Module 3 on *"Protein Modelling"* (summer semester, 2024).

Rebecca Wade, Stefan Richter, and Manuel Glaser:
BSc. lecture and practical course on Bioinformatics, Heidelberg University, winter semester 2023/24.

Ulrike Wittig:
"Datenbanken + Datenmanagement" in SETAC/GDCh-Postgradualkurs *"Entwicklung von Alternativmethoden"*, Heidelberg, Germany, 8-12 April 2024. Software demonstration of FAIRDOM SEEK, COMBINE 2024, Stuttgart, Germany, 1 - 4 September 2024.

Alexander Zeilmann:
Advanced software practical *"IT-Sicherheit"*, Heidelberg University, winter semester 2024/2025.

Wei Zhao:
Seminar *"Diachronic Language Models"*, Department of Computational Linguistics, Heidelberg University (winter semester 2023/2024).

10 Miscellaneous

10.1 Guest Speaker Activities (invited talks):

Aksel Alpay:
"The Community-Driven AdaptiveCpp SYCL Compiler Project: From High-Level C++ Programming to the Automatic Synthesis of Specialized Kernels", Minisymposium Application Perspective on SYCL, a Modern Programming Model for Performance and Portability, PASC '24, Zurich, 4 June 2024.

Róbert Andrássy:
"Convective boundary mixing: Theory and numerical simulations," conference *"Stellar convection: Modelling, Theory and Observations,"* Nordita, Stockholm, Sweden, 30 August 2024. *"Magnetohydrodynamics simulations of stellar interiors: challenges and methods,"* Université de Bordeaux, Bordeaux, France, 2 October 2024.

Camilo Aponte:
"About my career path in academia", Third Infinity 2024, Göttingen, Germany, 17 October 2024.

Beatriz Bordadagua:
"The efficiency of mixed modes for angular momentum transport", Stars Day, Porto, Portugal, 2024.

Mislav Brajkovic:
"Investigation of effect of loop motion on the residence time of histamine-1-receptor (H1R) antagonists by τRAMD.", 3rd German workshop on structural predictions of membrane proteins: From ion channels to G protein-coupled receptors, Jülich, Germany. 26-27 February 2024. Tunneling group symposium, Katowice, Poland, 30 June – 1 July 2024.

Lynn Buchele:
"Probing the internal structure of solar-like oscillators", Asterics Group, ISTA, Austria, 2024.

Jeong Yun Choi:
"Decoding the puzzle of overlapping oscillations in asteroseismic binary star systems", Europe-Korea Conference on Science and technology (EKC) 2024, Warwick University, UK, 2024.

Robert Fisher:
"Turbulent Detonation in Type Ia Supernovae," Max Planck Institute for Astrophysics (MPA) Seminar on Stellar Astrophysics (SESTAS) Seminar, 31 January 2024; *"An Emerging Consensus on White Dwarf Supernovae,"* Heidelberg Joint Astronomy Colloquium, 18 June 2024. *"Recent Progress on Modeling Helium-Ignited Sub-Chandrasekhar Mass and near-Chandrasekhar Mass SNe Ia,"* The Progenitors of Supernovae and Their Explosions Meeting, Dali, China, 26-30 August 2024.

Javier Morán Fraile:
"3DMHD simulations of white dwarf mergers and their GW signals," LGWA workshop, Castel Gandolfo, Rome, Italy, 8 October 2024. *"3DMHD simulations of compact object mergers,"* Seminar at Institute for Advanced Study, Princeton, USA, 23 October 2024; seminar at the Center for Computational Astrophysics, Flatiron Institute, New York, US, 24 October 2024.

Nikos Gianniotis:
"Probabilistic Flux Variation Gradient". Star Formation Across Cosmic Scales: Machine Learning Insights and Applications, Budapest, Hungary, 14 May 2024.

Martin Golebiewski
"The role of standards in defining an ecosystem for Virtual Human Twins (VHTs)" (invited talk), 1st International Symposium Digital Twins for Healthcare, Ayia Napa, Cyprus, 16 - 17 May 2024.

Frauke Gräter:
"Learning force fields and reactivity to enhance classical biomolecular simulations", ISQBP2024, Athens, Greece, 19-23 May 2024; CECAM 55 conference, Lausanne, Switzerland, 2-4 September 2024; BioExcel Conference, Brno, Czech Republic, 20-23 October 2024. *"Mechano-sensing biomolecular systems"*, Seminar, IBI, EPFL, Lausanne, Switzerland, January 2024; Colloquium NCCR Bio-inspired materials, University of Fribourg, Fribourg, Switzerland, 19 February 2024. *"Converting physical forces into chemical cues - directly: mechanoradicals in tissue"*, Conference Mechanics of Life, EMBL, Heidelberg, Germany, 15-19 April 2024; Seminar, Skintegrity, ETH Zuerich, Switzerland, 23 April 2024. *"Reactive Molecular Dynamics simulations aided by Machine Learning"*, CECAM workshop Rare events, Mainz, Germany, 17-19 September 2024.

Fabian Grünewald:
"Strings to Systems Describing Molecules Using CGsmiles", CECAM Workshop on *"Frontiers of Coarse-Grained Models: From New Developments to Modeling Dynamics, Assemblies, and Macromolecular Machines"* ENS Lyon France, 5-3 April 2024. *"Predicting interactions of polymers with cellular matter"*, Giersch International Conference on *"From Multiscale Models to Digital Twins"* FIAS Frankfurt Germany, 25-27 April 2024.

Jan Henneco:
"Crash course binary mass transfer: what do we know, what do we want to learn", Lorentz Center Workshop - From discovery to a population: benchmarking stripped stars and companions, Lorentz Center, Leiden, the Netherlands, 23 July 2024.

Saskia Hekker:
"Internal structures of low-mass stars", The Crafoord Prize Symposium in Astronomy: New avenues in solar and stellar physics, Stockholm, Sweden, 2024. *"Asteroseismology and binarity: two of a kind"*, European Astronomical Society (EAS) Annual Meeting, Padova, Italy, 2024.

Vincent Heuveline:
„Gesellschaftliche Perspektiven der KI. Chancen und Herausforderungen" in der Reihe *"Unsupervised Thinking"*, Providenzkirche, Heidelberg, 21 June 2024. *"Do management and administration need AI, or vice versa? Opportunities and challenges"*, 19th Israeli-German Administrators Conference (IGAC), Heidelberg, 24 June 2024. *"AI in Medicine: Opportunities and Challenges, with Possible Side Effects"*, Joint DFH/UFA workshop on AI in Medicine: Optimised Trials with Machine Learning, Sorbonne University, Paris, 11 September 2024. *"AI in medicine: potential and challenges with possible side effects"*, Farewell Symposium The Future Vision of Cancer Prevention in honor of Prof. Dr. Magnus von Knebel Doeberitz, Heidelberg, 20 September 2024. *„Brauchen Verwaltung und Management KI oder andersherum? Ein Plädoyer für Innovation"*, Bundestagung Studierendenverwaltungen, Heidelberg, 7 November 2024.

Lukas Hübner:
"Future-Proofing Bioinformatic Applications: Handling CPU-failures, abstracting MPI & reproducible experiments", Universidade da Coruña, La Coruña, Spain, October 2024.

Rajika Kuruwita:
"Were all fast-rotating stars born in a binary?", Invited talk, Stockholm University, Sweden, 16 May 2024. Invited Review Talk, European Astronomical Society ASM, Padua, Italy, 1-5 July 2024.

Eva Laplace:
"Compact objects from massive single and binary stars and their signatures in gravitational-wave sources" - seminar, University of Amsterdam, Netherlands, 5 May 2024.

Mike Lau:
"Common envelopes and planetary engulfment in SPH," Joint Franco-Australian 5th Phantom and MCFOST Users Workshop, online, 15 February 2024; online group meeting presentation, Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam, Netherlands, 1 August 2024. *"Expansion of accreting main-sequence stars during rapid mass transfer,"* SESTAS meeting, Max Planck Institute for Astrophysics, Garching bei München, Germany, 18 June 2024. *"Radiation hydrodynamics in common-envelope evolution & mass ablation of engulfed planets,"* presentation for the Stars & Compact Objects Group, Center for Computational Astrophysics, Flatiron Institute, New York, United States of America, 26 September 2024. *"Hydrodynamical simulations of planet ablation in stellar envelopes,"* The Chinese University of Hong Kong, Hong Kong, 27 December 2024.

Kiril Maltsev:
"Stellar evolution emulators," Astrophysics group seminar, University of Surrey, Surrey, United Kingdom, 22 February 2024 (online). *"Explodability criteria for rapid binary population synthesis,"* Gravitational Wave Astrophysics group seminar, University of Heidelberg, Heidelberg, Germany, 3 May 2024; SESTAS seminar, Max Planck Institute for Astrophysics, Garching, Germany, 13 November 2024. *"Progenitor evolution, formation and collapse of neutron stars,"* Astro AI Lab seminar, University of Heidelberg, Heidelberg, Germany, 8 July 2024.

Evans Owusu:
"At Intermediate Ages, [Na/Fe] Effectively Separates the Young and Old Sequences of Disc Stars in the Galaxy," Bergemann Independent Research Group meeting, MPIA, Heidelberg, Germany, 17 October 2024. *"Exploring the elemental abundance of sodium for separating the young and old sequences of the disc stars in the Galaxy,"* Centre for Astrophysical Research, University of Hertfordshire, UK, 29 October 2024.

Marc-Oliver Pohle:
"Generalised Covariances and Correlations", Kolloquium der Fächergruppe Mathematik und Statistik, Helmut Schmidt University, Hamburg, Germany, 22 May 2024. *"Forecasting under Long Memory"*, Kolloquium Ökonometrie & Statistik, WHU - Otto Beisheim School of Management, Vallendar, Germany, 13 November 2024.

Kai Lars Polsterer:
"From Supervised to Unsupervised ML: lessons learned from learning machines". Colloquium at the University of Perth, Australia, 15 May 2024. *"Spherinator & HiPSter & Jasmine: Using HiPS tiling to allow for explorative access to simulations"*: IVOA Sydney, Australia, 22 May 2024.

Francisco Pozo Nunez:
"Measuring quasars accretion disk sizes with the LSST". Galactic Nuclei in the Cosmological Context 2024, Szczecin, Poland, 3-6 June 2024. *"Accretion disk sizes in distant quasars"*, Araucaria ERC Meeting, Paris Observatory, Paris, France, 2-6 September 2024.

Friedrich Röpke:
"Simulations of common-envelope interaction in binary stellar systems," seminar talk at the University of New South Wales Canberra, Australia, 25 January 2024. *"Simulating thermonuclear astrophysical transients,"* conference *"Transients Down Under"*, Melbourne, Australia, 30 January 2024. *"Jet formation in stellar interactions,"* Annual Meeting of the European Astronomical Society, Padua, Italy, 5 July 2024. *"Surprises in 3D stellar hydro: common-envelope interactions and related processes,"* Wolfgang Hillebrandt's 80th Birthday Symposium, Garching, Germany, 5 August 2024. *"Simulations of thermonuclear astrophysical transients,"* conference *"The Progenitors of Supernovae and their Explosions"*, Dali, China, 26 August 2024. *"Physics of Stellar Objects research at Heidelberg,"* Université de Bordeaux, Bordeaux, France, 2 October 2024.

Ashley Ruiter:
"Type Ia supernovae from double white dwarf merger populations," Wolfgang Hillebrandt's 80th Birthday Symposium, Garching, Germany, 5 August 2024. *"The first supernova remnant observations that reveal a clear signature of a double-detonation SN Ia,"* conference *"The Progenitors of Supernovae and their Explosions,"* Dali, China, 26–30 August 2024.

Fabian Schneider:
"Binary star evolution", Workshop *"360º Approach to Common Envelope Evolution"*, University of Barcelona, Spain, 10 June 2024.

Ivo Seitenzahl:
"Type Ia supernova nucleosynthesis," Wolfgang Hillebrandt's 80th Birthday Symposium, Garching, Germany, 5 August 2024.

Alexandros Stamatakis:
"Eliminating Subjectivity, Quantifying Uncertainty, and using Machine Learning for Phylogenetic Inference", Barcelona Supercomputing Center, Barcelona, Spain, February 2024: keynote talk, 2nd Panhellenic Conference of Bioscientists, Alexandroupoli, Greece, March 2024. *"Recent Advances in General Phylogenetic Inference and Energy-efficient Computing"*, Conference on Cross-Linguistic Explorations in Space, Time and Evolution, Tübingen, Germany, June 2024. *"Machine Learning in Phylogenomics"*, EMBO Lecture Course on Evolutionary and Comparative Genomics, Nafplio, Greece, November 2024. *"The National Female Researchers and Professors Program"*, joint event organized by FORTH and the regional government of Crete on *"Female Researchers in Positions of Responsibility"*, FORTH, 9 December 2024.

Jan Stühmer
"Generating Highly Designable Proteins with Geometric Algebra Flow Matching", Chan Zuckerberg Initiative, Redwood City, CA, USA, 19 September 2024.

Jonathan Teuffel:
"DYNATHOR: Dynamics of the Complex of Cytochrome P450 and Cytochrome P450 Reductase in a Phospholipid bilayer", 27th Results and Review Workshop of the HLRS, Karlsruhe (Germany), 10-11 October 2024.

Anastasis Togkousidis:
"RAxML-NG v2.0: Enhancing Phylogenetic Inference with MSA Difficulty Prediction, Adaptive Heuristics, and Accelerated Bootstrapping", Natural History Museum London, December 2024.

Sebastian Trujillo Gomez:
"Simulations as powerful but challenging tools for understanding dwarf galaxies".Astronomische Gesellschaft, Meeting 2024, Cologne, Germany, 9-13 September 2024. *"Spherinator + HiPSter: from the known unknowns to the unknown unknowns"*. IVOA November 2024 Interoperability Meeting, La Valletta, Malta, 14-17 November 2024.

Rebecca C. Wade:
"Bridging timescales to investigate protein binding kinetics", MD@60 Conference, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore, India, 26-29 Feb 2024. *"Brownian dynamics simulation to investigate molecular diffusion and protein complex formation"*, Symposium on Simulations of Physical and Biological Systems, Jawaharlal Nehru University (JNU), New Delhi, India, 1 March 2024. *"Bridging timescales to investigate protein binding kinetics"*, American Chemical Society Spring Meeting, New Orleans, USA, 20 March 2024. *"New directions for structure-based drug design"*, Retreat, Digestive Health and Life Science Alliance Heidelberg-Mannheim, Asselheim, Germany, 11-12 April 2024. *"Towards predicting the effects of protein mutations on binding kinetics"*, NIC Days on Biomolecular Evolution, Function and Assembly: Theory meets Experiment, Forschungszentrum Jülich, Germany, 24-26 April, 2024. *"Protein dynamics and binding kinetics"* European Workshop on Drug Design (EWDD), Siena, Italy, 19-23 May 2024. *"Multi-resolution molecular dynamics simulations to investigate how cytochrome P450 conformational variability affects its interactions and function"*, 25th International Symposium on Microsomes and Drug Oxidations, Prague (Czech Republic), 7-10 July 2024. *"Bridging timescales to predict protein binding kinetics"*, CECAM workshop: Leveraging Machine Learning for Sampling Rare Events in Biomolecular Systems, Max Planck Institute for Polymer Research, Mainz, Germany, 17-19 September 2024. *"Bridging timescales to predict protein-ligand binding kinetics: Applications to G protein-coupled receptors"*, Nxera Pharma UK, Granta Park, Cambridge, UK, 6 November 2024. *"Bridging timescales to predict protein-ligand binding kinetics"*, AstraZeneca, The Discovery Centre, Cambridge, UK, 19 November 2024. *"In silico exploration of the dynamics of biomolecular interactions"*, MRC Laboratory of Molecular Biology, Cambridge, UK, 20 November 2024. *"From protein 3D-structure to function: Learning from computer simulation"*, Department of Chemistry, University of Bristol, Bristol, UK, 5 December 2024.

Ulrike Wittig:
"ELIXIR Data Platform", ELIXIR Toxicology Community Webinar, online, 2 May 2024. *"Using EnzymeMLto exchange and publish reaction kinetics data in SABIO-RK"*, 5th EnzymeML Workshop, Rüdesheim, Germany, 24-26 September 2024. *"Research Data Management Activities in FAIRDOM and ELIXIR"*, STRENDA/ESAB Symposium *"The Need for Reproducibility of Enzyme Reaction Data"*, Lodz, Poland, 14 November 2024.

Alexander Zeilmann:
"Petabyte-scale image analysis and visualization", in Science Cafe, Natural History Museum Stuttgart, 13 March 2024. *"KI-Morph — A Platform for petabyte-scale image analysis"*, in Community Forum Data-Intensive Computing (DIC), Heidelberg University, online, 5 June 2024.

10.2 Presentations

Talks (Contributed talks)

Felix Ahlborn:
"Accurate asteroseismic envelope rotation rates of evolved red giants", KASC15/TASC8 Conference 2024, Porto, Portugal, 15-19 July 2024. *"Hydrodynamic simulations as a test bed for turbulent convection models"*, Nordita program: Stellar convection: Modelling, Theory and Observations, Stockholm, Sweden, 28 August 2024.

Aksel Alpay:
"AdaptiveCpp: Portable Heterogeneous Computing in C++", Indo-German Workshop on Hardware-aware Scientific Computing, Heidelberg, 28 October 2024.

Michaël Bazot:
"The SONG network for PLATO benchmark stars", SONG 2024 Science Meeting, Tenerife, Spain, 18-20 September 2024. *"Saturated stars PLATO WG"*, PLATO Week #15, Göttingen, Germany, 23-25 October 2024.

Beatriz Bordadagua:
"The efficiency of mixed modes for angular momentum transport", KASC15/TASC8 Conference 2024, Porto, Portugal, 15-19 July 2024.

Mislav Brajkovic:
"Investigation of effect of loop motion on the residence time of histamine-1-receptor (H1R) antagonists by τRAMD.", 3rd German workshop on structural predictions of membrane proteins: From ion channels to G protein-coupled receptors, Computer Simulation and Theory of Macromolecules, Hünfeld, Germany. 19-20 April 2024. HBIGS retreat 2024, Speyer, Germany. 30 September – 2 October 2024.

Vincent Bronner:
"Type II-L and II-P supernovae from pulsating red-supergiants", Liège Symposium LIAC41, Liège, Belgium, 16 July 2024. *"Going from 3D common-envelope simulations to fast 1D simulations"*, Liège Symposium LIAC41, Liège, Belgium, 18 July 2024. *"Explosions of pulsating red supergiants - A natural pathway to both Type II-P and II-L supernovae from the very same progenitor stars"*, The Progenitors of Supernovae and their Explosions, Dali, China, 30 August 2024. *"The next generation of Bonnsai"*, VFTS collaboration meeting, Madrid, Spain, 17 September 2024.

Lynn Buchele:
"Probing the internal structure of low-mass main-sequence stars using structure inversions", KASC15/TASC8 Conference 2024, Porto, Portugal, 15-19 July 2024. Unveiling the interiors of the stars to grasp stellar populations, Cefalù, Italy, 23-27 September 2024.

Marcus Buchwald:
"Challenges in Learning to Defer to Experts", 1. TRN - HI4AI Workshop, Mannheim, February 19th, 2024. *"Discovering Predictive Biomarkers for Liver Staging Using Machine Learning Methods"*, Workshop on Artificial Intelligence (AI) in Medicine, Sorbonne University and Heidelberg University under the 4EU+ European University Alliance, Paris, September 11-13th, 2024.

Jeong Yun Choi:
"Tackling the complexities of overlapping oscillations in asteroseismic binaries using Kepler data", KASC15/TASC8 Conference 2024, Porto, Portugal, 15-19 July 2024.

Iliana Isabel Cortés Pérez:
"Exploring SMBH Activity in AGNs Through Photometric Variability". *"Stellar award presentation"*, Astroinformatics 2024, Llanuras de Diana, Puerto Natales, Patagonia, Chile, 9-13 December 2024.

Rostislav Fedorov:
"Exploration of redox properties in chemical space", 2nd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Studio Villa Bosch, Heidelberg, Germany, 15-17 May 2024.

Javier Morán Fraile:
"Thermonuclear explosions from low-mass CO white dwarfs," presentation at the European Astronomical Society Annual Meeting 2024, Padova, Italy, 1–5 July 2024.

Nikos Gianniotis:
"Astronomy for Machine Learning". ML4ASTRO - Machine Learning in Astrophysical Studies Workshop, Heidelberg, Germany, 25 June 2024. *"Positive and Scale Invariant Gaussian Process Latent Variable Model for Astronomical Spectra"*. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Brugge, Belgium, 10 October 2024. *"A Probabilistic Reformulation of the Cross Correlation Method for Delay Detection"*. Gravitational Lensing Seminary, ARI, Heidelberg University, Germany, 3 December 2024.

Martin Golebiewski:
"Standards for human digital twins and their implementability", 1st Ecosystem meeting on building the Virtual Human Twin, Paris, France, 18-19 January 2024. *"Standardization needs for Virtual Human Twins"*, EDITH technical consortium meeting, Athens, Greece, 15 - 17 April 2024. *"Task Area TA2 – Standards for FAIR Data"*, NFDI4Health Steering Committee Meeting, Bremen, Germany, 6 - 7 May 2024. *"Introduction into the standardization landscape for the Virtual Human Twin"* and *"FAIRDOM SEEK: share your data and models FAIR"*, 2nd EDITH ecosystem meeting, Amsterdam, Netherlands, 15 - 16 July 2024. *"The role of standards in defining an ecosystem for virtual human twins (VHTs)"*, VPH Conference 2024 - Data-driven Simulation Technologies for Clinical Decision Making, Stuttgart, Germany, 4 - 6 September 2024. *"FAIR Health data management using the SEEK platform and NFDI-4Health Local Data Hubs"*, Gesundheit Gemeinsam 2024 - 69th annual meeting of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS), Dresden, Germany, 8 - 13 September 2024. *"Consolidation of the metadata schema and interfaces (including mapping)"*, *"Task Area TA2 – Standards for FAIR Data"* and *"FAIR data sharing – Community aspects"*, Annual NFDI4Health Meeting, Potsdam, Germany, 7 - 8 October 2024. *"The NFDI4Health metadata schema as an example for collecting medical metadata from epidemiological and clinical studies"*, FORSCHUNGSDATEN@RMU 2024 - 5. Virtueller Forschungsdatentag der Rhein-Main Universitäten, online, 7 November 2024.

Fabian Grünewald:
"Advances in the Polyply software for macromolecular modeling", talk at the ACS Fall Meeting 2024, Denver, USA, 17-22 August 2024.

Julia Haag:
"Simulations of Sequence Evolution: How (Un)realistic They Are and Why", Legend2024, Heraklion, Crete, Greece, May 2024. *"Educated Bootstrap Guesser: Predicting Phylogenetic Bootstrap Values"*, Legend2024, Heraklion, Crete, Greece, May 2024.

Luise Häuser:
"Are Sounds Sound for Phylogenetic Reconstruction?", 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP (SIGTYP 2024), St. Julians, Malta, March 2024. *"Computational Approaches for Integrating out Subjectivity in Cognate Synonym Selection"*, Proceedings of the Society for Computation in Linguistics Conference, Irvine, USA, June 2024.

Saskia Hekker:
"The Milky Way Assembly Tale: the Plato era", The Milky Way Assembly Tale Conference, Bologna, Italy, 2024. What curves Δv?", KASC15/TASC8 Conference 2024, Porto, Portugal, 15-19 July 2024. *"Stellar structure variations in red giants derived from their oscillations"*, Unveiling the interior of the stars to grasp stellar populations, Cefalù, Italy, 23-27 September 2024.

Jan Henneco:
"Stable mass-transfer channels leading to stellar mergers", Workshop on stable mass transfer in binaries: from onset to remnants, CCA (Flatiron Institute), New York City, USA, 12 March 2024. *"Merger seismology - Asteroseismic Properties of Massive Merger Products"*, 8th TESS/15th Kepler Asteroseismic Science Consortium Workshop, University of Porto, Porto, Portugal, 18 July 2024.

Alexander Holas:
"Electron-capture supernovae – Thermonuclear explosion or gravitational collapse? The fate of sAGB stars on a knife's edge," Nuclear Physics in Astrophysics XI, Dresden, Germany, 15–20 September 2024.

Xiaoming Hu:
"NFDI4Health Local Data Hubs for Finding and Accessing Personal Health Data", Harmony 2024, London, United Kingdom, 8-11 April 2024.

Lukas Hübner:
"Memoization on Shared Subtrees Accelerates Computations on Genealogical Forests", WABI 2024 Conference, London, Great Britain, September 2024.

Alexander Jordan:
"Evaluating probabilistic classifiers: The triptych", Bernoulli-IMS 11th World Congress in Probability and Statistics, Bochum, Germany, 14 August 2024.

Alexey Kozlov:
"EcoFreq: Compute with Cheaper, Cleaner Energy via Carbon-Aware Power Scaling", 39th ISC High Performance conference, Hamburg, Germany, May 2024.

Rajika Kuruwita:
"Simulated analogues: a new methodology for non-parametric matching of models to observations", Lightning talk. HAMLET-PHYSICS 2024 Conference/Workshop. Copenhagen, 19-21 August 2024.

Eva Laplace:
"The role of stellar physics in the formation of black holes from single and binary stars", Liège Symposium LIAC41, Liège, Belgium, 19 July 2024.

Mike Lau:
"Accretor inflation during rapid mass transfer," 41st Liège International Astrophysical Colloquium: The eventful life of massive star multiples, Liège, Belgium, 18 July 2024.

Giovanni Leidi:
"MHD effects in core-collapse supernova progenitors," talk at the European Astronomical Society Annual meeting 2024, Padova, 1-5 July 2024. *"Magnetohydrodynamic simulations of a thermally relaxed carbon-burning shell,"* talk at the 18th Würzburg Winter Workshop, Heidelberg, Germany, 11 December 2024.

Kiril Maltsev:
"Convective core overshooting effects on compact remnant mass and Type II explosion energy landscapes from massive single star evolution," talk at the Transients Down Under conference, Melbourne, 29 January 2024. *"Prediction of stellar evolution tracks, gravitational waves and core collapse supernova outcomes with machine learning,"* talk at the 18th workshop of the Australian National Institute for Theoretical Astrophysics, Monash, Australia, 7 February 2024. *"Stellar evolution forecasting with a timescale-adapted evolutionary coordinate and machine learning,"* talk at the Spring Meeting 2024 of the German Physical Society, Berlin, Germany, 8 March 2024. *"Explodability criteria for rapid binary population synthesis,"* talk at the Annual Meeting of the European Astronomical Society, Padua, Italy, 1 July 2024.

Gerhard Mayer:
"Standards for human digital twins and their implementability - Part 2", EDITH High level partner's meeting, 1st Ecosystem meeting on building the Virtual Human Twin, Paris, France, 18-19 January 2024. *"Overview of EDITH documents and collections on standardization"*, 2nd EDITH ecosystem meeting, Amsterdam, Netherlands, 15 - 16 July 2024.

Jonas Müller:
"Magnetic suppression of mode amplitudes: A toy model", KASC15/TASC8 Conference 2024, Porto, Portugal, 15-19 July 2024.

Wolfgang Müller:
"Systems Biology Data Management in de.NBI", MaRDI (Math Research Data Infrastructure) meeting, 17 - 18 October 2024.

David Nickerson, Martin Golebiewski, Thomas E. Gorochowski, Sarah Keating, Matthias König, Chris J. Myers, Falk Schreiber, Dagmar Waltemath, Padraig Gleeson:
"The COmputational MOdelling in BIology NEtwork in 2024: Standards and services for the computational physiology community and beyond", VPH Conference 2024 - Data-driven Simulation Technologies for Clinical Decision Making, Stuttgart, Germany, 4 - 6 September 2024.

Giulia Paiardi:
"The role of heparin in spike SARS-CoV-2 infection: from a model for heparan sulfates to a starting structure for antivirals". Biophysical Society Meeting, Philadelphia, USA, 10-14 February 2024. ZAPP Seminar, Center for Molecular Biology of Heidelberg University (ZMBH), Germany, 29 February 2024. ISQBP President's meeting, Athens, Greece, 19-23 May 2024.

Marc-Oliver Pohle:
"Statistical Inference for Rank Correlations", 11th HKMetrics Workshop, Heidelberg, Germany, 30 July 2024. *"Generalised Covariances and Correlations"*, Bernoulli-IMS 11th World Congress in Probability and Statistics 2024, Bochum, Germany, 15 August 2024. 26th International Conference on Computational Statistics (COMPSTAT), Gießen, Germany, 29 August 2024. *"Uncertainty Quantification in Forecast Comparisons"*, Workshop on Postprocessing, Bielefeld, Germany, 3 September 2024.

Erik Poppleton:
"RNA nanopores without chemical functionalization", Foundations of Nanotechnology (FNANO), April 2024, Snowbird, Utah, USA; Artificial Biology (ArtBio), August 2024, Aarhus, Denmark. *"Custom analysis with oxDNA Analysis Tools"*, oxDNA Users and Developers Meeting, September 2024, Oxford, UK. *"Biomimetic hardware for synthetic cells"*, International Conference for the Ethics of Engineering Life (ICEEL), October 2024, Rome, Italy.

Friedrich Röpke:
"Type Ia supernovae from thermonuclear explosions in white dwarf stars," Nuclear Physics in Astrophysics XI, Dresden, Germany, 17 September 2024.

Jonas Roller:
"FEM Simulation of Taylor-Couette Flows Under Dielectrophoretic Force". European Fluid Dynamics Conference (EFDC1), Aachen, 16-20 September 2024.

Ashley Ruiter:
"The first supernova remnant observations that reveal a clear signature of a double-detonation SN Ia," talk given at the SNEx seminar (Technion, Haifa, Israel, given online), 20 November 2024.

Valentin Schmid:
"Exploring Colorectal Cancer Development at the Crypt Level with Continuous Mathematical Models". Lake Como School, Como, 10-14 May 2024.

Fabian Schneider:
"Bimodal black-hole mass distribution and chirp masses of binary black-hole mergers", General Assembly of the International Astronomical Union, IAU 389 *"Gravitational Wave Astrophysics"*, Cape Town, South Africa, 6 August 2024. *"Pre-supernova evolution and final fates of binary stars"*, General Assembly of the International Astronomical Union, FM4 *"Bridging the final stages of massive stars to supernovae and transients"*, Cape Town, South Africa, 14 August 2024.

Boris Schüpp:
"DNA and Ultrasound - A New Approach in Mechanochemistry", Max Planck Schools Matter to Life Ringberg Symposium (Talk, Schloss Ringberg, Germany), 9 July 2024.

Leif Seute:
"Grappa – A Machine Learned Molecular Mechanics Force Field". Applied Machine. Learning Days at EPFL. Lausanne, Switzerland, 24 March 2024; Hünfeld Workshop on Computer Simulation and Theory of Macromolecules. Hünfeld, 27 April 2024; 2nd SIMPLAIX. Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, 17 May 2024.

Nuriza Suleimenova:
"Collagen as a redox-scavenging protein", 8TH FREIBERG COLLAGEN SYMPOSIUM, Freiberg, Germany, 19-20 September 2024.

Sebastian Trujillo Gomez:
"Representation Learning for Explorative Knowledge Discovery in Astrophysics Simulations and Beyond." HiPEAC 2024 Workshop Plasma Physics towards the Exascale Era, Garching, Germany, 1 January 2024. *"How to test gravity and cosmology using next-generation observatories"*. Lorentz Center Workshop: A new dawn of dwarf galaxy research, Leiden, Netherlands, 8-12 April 2024. *"Interpreting the largest cosmological simulations using Representation Learning"*. ML4ASTRO2 Conference, Catania, Italy, 11 July 2024. "Spherinator + HiPSter: beyond the 'known unknowns' towards the 'unknown unknowns". Astronomische Gesellschaft Meeting 2024, Cologne, Germany, 9-13 September 2024.

Kristián Vitovský:
"Hydrodynamical flows in the Interiors of Blue Supergiants," talk at the 18th Würzburg Winter Workshop, Heidelberg, Germany, 10 December 2024.

Ulrike Wittig:
"Unlocking Bio-Curation: Harnessing Text Highlighting for Neural Named Entity Models of SABIO-RK database", 17th Annual International Biocuration Conference, New-Delhi, India, 5-8 March 2024. *"Increasing data evidence by text-fragment references using nanopublications"*, 17th Annual International Biocuration Conference, New-Delhi, India, 5-8 March 2024.

Alexander Zeilmann:
"KI-Morph – User-friendly large-scale image analysis & AI on bwHPC systems", in 10th bwHPC Symposium, Freiburg, Germany, 25 September 2024.

Posters

Camilo Aponte:
"Mechanical Control of Cell Adhesion in Infection: Force-induced Multivalendy". Biophysical Society Meeting 2024, Philadelphia, USA, 10-14 February 2024; EMBO | EMBL Symposium: The mechanics of life: from development to disease, Heidelberg, 15-18 April 2024.

Tommaso Bartoloni:
"Towards peptide-based therapeutics against cardiac disease: Prediction and simulation of the S100A1ct peptide with a membrane environment". 2nd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany. 15-17 May 2024. 37th European Peptide Symposium/14th International Peptide Symposium, Florence, Italy, 25-29 August 2024. CECAM workshop: Computational Structural Biology, Florence, Italy, 30-31 August 2024. 24th EuroQSAR (European Symposium on Quantitative Structure-Activity Relationship), Barcelona, Spain, 22-26 September 2024.

Riccardo Beccaria:
"Brownian Dynamics Simulation of the Protein-Ligand Association Process in a Crowded Environment", 2nd Simplaix Workshop on *"Machine Learning for Multiscale Molecular Modeling"*, Heidelberg, Germany, 15-17 May 2024. CCP5 Summer School 2024, Newcastle, England, 15-25 July 2024. *"Simulation of Diffusional Association: a Brownian Dynamics Simulation Software for Protein-Protein, Protein-Ligand and Protein-Surface Interactions"*, CECAM workshop on Biomolecular Simulations at the Mesoscale, Trento, Italy, 26-29 August 2024.

Mislav Brajkovic:
"Investigation of effect of loop motion on the residence time of histamine-1-receptor (H1R) antagonists by τRAMD." 2nd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Heidelberg, Germany. 15-17 May 2024. 24th European Symposium on Quantitative Structure-Activity Relationship, Barcelona, Spain, 22-26 September 2024.

Marcus Buchwald:
"Intelligente Entlastung der Infektionssprechstunde – wie mit Hilfe von selbstberichteten Anamnesen Delegation erleichtert werden könnte". 58. Kongress für Allgemeinmedizin und Familienmedizin DEGAM, Würzburg, 26-28 September 2024.

Quentin Coppee:
"Various morphologies observed in the power spectra of suppressed-dipole mode red giants", KASC15/TASC8 Conference 2024, Porto, Portugal, 15-19 July 2024.

Iliana Isabel Cortés Pérez:
"Disentangling the Interplay of the Inner Regions of AGNs Via Probabilistic Photometry". Annual Meeting of the Astronomische Gesellschaft 2024. Cologne, Germany. Poster Prize. 9-13 September 2024.

Charlott Danielson, Marc Horner, Martin Golebiewski, Heike Moser, Gerhard Mayer:
"Towards international standardization of computational modeling and simulation in the field of medical devices", VPH Conference 2024 - Data-driven Simulation Technologies for Clinical Decision Making, Stuttgart, Germany, 4 - 6 September 2024.

Ayse Erozan:
"Analysis and Segmentation of Nucleus with U-Net", Machine Learning Conference for X-Ray and Neutron-Based Experiments, München, Germany, 2024.

Francisca Espinoza:
"A comprehensive catalogue of asteroseismic properties of red-giant stars", The Milky Way Assembly Tale, 2024, Bologna, Italy, 27-31 May 2024. KASC15/TASC8 Conference 2024, Porto, Portugal, 15-19 July 2024.

Rostislav Fedorov:
"Exploration of redox properties in chemical space". 2nd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling. Heidelberg, Germany, 15-17 May 2024. Chemical Compound Space Conference 2024, Heidelberg, Germany, 21-24 May 2024. ChemAI Conference, Amsterdam, Netherlands, 29 November 2024.

Fabian Grünewald:
"CGsmiles a line notation for representing molecules at different resolutions", 2rd SIMPLAIX Workshop, Heidelberg, Germany, 15–17 May 2024.

Alexander Holas:
"Ares – Simulating Type Ia Supernovae on Heterogeneous HPC Architectures," The Salishan Conference on High Speed Computing, Salishan Costal Lodge, Oregon, USA, 22–25 April 2024.

Olga Krebs, Susan Eckerle, Martin Golebiewski, Anne Elin Hegglan, Xiaoming Hu, Maja Rey, Ulrike Wittig, Wolfgang Müller:
"Standard compliant data and model management for systems medicine projects", COMBINE 2024, Stuttgart, Germany, 1 - 4 September 2024.

Mike Lau:
"Accretor inflation during rapid mass transfer: Implications for modelling binary populations and gravitational wave sources," Gravitational Wave Physics and Astronomy Workshop (GWPAW) 2024, Birmingham, United Kingdom, 28-31 May 2024.

Gerhard Mayer, Martin Golebiewski, Wolfgang Müller:
"EDITH Use case 6: Registration of SEEK models in the EDITH repository", 2nd EDITH ecosystem meeting, Amsterdam, Netherlands, 15 - 16 July 2024. *"The role of standards in defining an ecosystem for Virtual Human Twins (VHTs)"*, COMBINE 2024, Stuttgart, Germany, 1 - 4 September 2024, and VPH Conference 2024 - Data-driven Simulation Technologies for Clinical Decision Making, Stuttgart, Germany, 4 - 6 September 2024.

Lukrécia Mertová, Severin Polreich, Oleg Lewkowski, Wolfgang Müller:
"The BeeProject: Advanced Digitisation and Creation \\ of a Dataset for the Monitoring of Beehives", KIDA-KON 2024 | AI for Research in Food, Agriculture and Environment, Leipzig, Germany, 2 - 3 December 2024.

Giulia Paiardi:
"The role of heparin in spike SARS-CoV-2 infection: from a model for heparan sulfates to a starting structure for antivirals". Biophysical Society Meeting, Philadelphia, USA, 10-14 February 2024. *"The role of GAGs in spike SARS-CoV-2 infection: from a model for heparan sulfates to a starting structure for antivirals"*. 24th EuroQSAR Symposium. Barcelona, Spain, 22-26 September 2024. *"Combining AI and molecular simulation for anticancer peptide and peptidomimetic design"*. Joachim Herz Stiftung add-on fellows meeting, Hamburg, Germany, 1-2 October 2024.

Laura López Pérez, Elena Martinelli, Marc Kirschner, Sylvia Krobitsch, Heike Moser, Giuseppe Fico, Tito Poli, Martin Golebiewski (presenting author):
"Recommendations and requirements for implementing computational models in clinical integrated decision support systems (ISO/TS 9491-2)", COMBINE 2024, Stuttgart, Germany, 1 - 4 September 2024, and VPH Conference 2024 - Data-driven Simulation Technologies for Clinical Decision Making, Stuttgart, Germany, 4 - 6 September 2024.

Maja Rey, Ulrike Wittig, Andreas Weidemann, Wolfgang Müller:
"de.NBI-SysBio: Research data management and curation of systems biology applications", de.NBI/ELIXIR-DE All Hands Meeting, Berlin, 28-29 November 2024.

Maja Rey, Frank T. Bergmann, Axel von Kamp, Pavlos Bekiaris, Ulrike Wittig, Andreas Weidemann, Steffen Klamt, Ursula Kummer, Wolfgang Müller:
"de.NBI-SysBio: Systems biology data and modelling tools and related services", 9th international conference on Systems Biology of Mammalian Cells (SBMC), Leipzig, 13-15 May 2024.

Maja Rey, Ulrike Wittig, Frank T. Bergmann, Axel von Kamp, Pavlos Bekiaris, Andreas Weidemann, Steffen Klamt, Ursula Kummer and Wolfgang Müller:
"Systems biology data and modelling support by ELIXIR Germany", ELIXIR All Hands, Uppsala, Sweden, 10-12 June 2024.

Friedrich Röpke:
"Morphology of common-envelope ejecta and shaping of planetary nebulae," conference "360° approach to common envelope evolution from binary progenitors to remnants", Barcelona, Spain, 10–14 June 2024.

Jonas Roller:
"Model Order Reduction for Unsteady Dielectrophoretic Force-Driven Flows". IUTAM Symposium on Laminar-Turbulent Transition, Nagano, 2-6th September 2024.

Valentin Schmid:
"Exploring Colorectal Cancer Development at the Crypt Level with Continuous Mathematical Models". ECMTB, Toledo, 22-26 July 2024.

Daniel Sucerquia:
"How a Stretching Force Differently Destabilizes Chemical Bonds on a Protein Backbone", GRS: Multiscale Mechanochemistry and Mechanobiology. Lewiston, Maine, USA, 20 -21 July 2024; GRC: Multiscale Mechanochemistry and Mechanobiology, Lewiston, Maine, USA, 21 -26 July, 2024.

Nuriza Suleimenova:
"A new link between tissue mechanics and oxidative stress", EMBO | EMBL Symposium: The mechanics of life: from development to disease, Heidelberg, Germany 15-18 April 2024. *"Investigating the linkage between tissue mechanics and oxidative stress"*, Thiol-Based Redox Regulation and Signaling Gordon Research Seminar, Barcelona, Spain, 13-19 July 2024.

Jonathan Teuffel:
"Multiscale simulation of cytochrome P450 electron transfer complexes: the reduction of CYP17A1 and its implications for the regulation of human sex hormone biosynthesis", Hünfeld Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, 19-20 April 2024. *"Towards predictors of electron transfer rates in the Cytochrome P450 family"*, 2nd SIMPLAIX Workshop on *"Machine Learning for Multiscale Molecular Modeling"*, Heidelberg, 15-17 May 2024. *"Martini, water and antimycotic drugs: Multiresolution molecular dynamics simulations reveal the interplay between conformational variability and functional interactions in membrane-bound cytochrome 2B4"*, 25th International Symposium on Microsomes and Drug Oxidations, Prague, Czech Republic, 7-10 July 2024.

Sebastian Trujillo Gomez:
"Representation Learning for Knowledge Discovery in Cosmological Simulations", Building Galaxies from Scratch: Advances and Challenges in Simulating Galaxy Evolution, Vienna, Austria, 19-23 February 2024. *"From data to scientific breakthroughs with tools powered by Generative Deep Learning"*. ADASS 2024, Valletta, Malta, 10-14 November 2024.

Evgeni Ulanov:
"Predicting hydrogen atom transfer energy barriers using Gaussian process regression", Workshop on Uncertainty Quantification in Molecular Simulation, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany, 29-30 August 2024.

Marco Vetter:
"Magnetically driven outflows and circumbinary disks in common-envelope events," conference "360° approach to common envelope evolution from binary progenitors to remnants", Barcelona, Spain, 10–14 June 2024.

Daniel Wolffram:
"Collaborative nowcasting of COVID-19 hospitalization incidences in Germany", Poster at 2nd National Conference on Modeling Infectious Diseases, Halle, Germany, 13 March 2024.

Alexander Zeilmann:
"KI-Morph – User-friendly large-scale image analysis & AI on bwHPC systems", in 10th bwHPC Symposium, Freiburg, Germany, 2024.

10.3 Memberships

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

Tilmann Gneiting:
Affiliate Professor at the Department of Statistics, University of Washington, Seattle, Washington, United States. Member of the Committee on Fellows at Institute for Mathematical Statistics.

Martin Golebiewski:
Convenor (chair) of the ISO/TC 276 Biotechnology working group 5 *"Data Processing and Integration"* (reelected in 2024), International Standardization Organization (ISO). Chair of the working group *"FAIR Data Infrastructures for Biomedical Informatics"* of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS). Member of the board of coordinators of COMBINE (Computational Modeling in Biology network). German delegate at the ISO technical committee 276 Biotechnology (ISO/TC 276), International Organization for Standardization (ISO). Member of the national German 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 sys-tems, and its ad-hoc group for *"Establishing the credibility of computational modelling in the field of medical devices through verification, validation, and uncertainty quantification"* (ahG 11), International Electrotechnical Commission (IEC). Member of the Virtual Physio-logical Human Institute for Integrative Biomedical Research (VPH Institute). Member of the steering committee of the German National Research Data Infrastruc-

ture for Personal Health Data (NFDI4Health). Member of the Scientific Advisory Board of the European IN-VENTS project (Innovative designs, extrapolation, simulation methods and evidence-tools for rare diseases addressing regulatory needs).

Frauke Gräter:
Scientific Member of the Max Planck Society (as of July 2024). Max Planck Fellow of the Max Planck School Matter to Life. Member of DFG Graduate school *'Tailored Scale-Bridging Approaches to Computational Nanoscience'* at KIT (GRK 2450). Member of the Biophysical Society, the German Physical Society, the German Biophysical Society, the German Chemical Society. Faculty member, Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg. Faculty member, HIDDS4HEALTH Graduate School, Heidelberg University, KIT & DKFZ. Associated faculty member, HGS MathComp Graduate School, University of Heidelberg.

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 Board of Directors (BoD) of Informatics for Life. Member of the Scientific Advisory Board (SAB) Potsdam Institute for Climate Impact Research. PI of HIDSS-4Health 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.

Eva Laplace:
PhD examination committee member, KU Leuven University, Belgium, May 2024.

Wolfgang Müller:
Leadership Team of LiSyM research network Liver Systems Medicine

Giulia Paiardi:
Member of the European Glycoscience Community (from May 2024).

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

Alexandros Stamatakis:
Member of scientific committee of the SMPGD (Statistical Methods for Post Genomic Data analysis) workshop series. Member of the advisory board of the SciLifeLab & Wallenberg National Program for Data-Driven Life Science program in Sweden.

Jan Stühmer:
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). Member of ELLIS - the European Laboratory for Learning and Intelligent Systems.

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

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 Scientific Advisory Board of the Max Planck Institute of Biophysics, Frankfurt, Germany. Member of International Advisory Board, 'Complexity in Chemistry' C2 Programme, Department of Chemical Sciences, University of Padova, Italy. 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.

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).

10.4 Contributions to the Scientific Community

Aksel Alpay:
Reviewer for the Journal of Parallel and Distributed Computing, and the conference International Workshop on OpenCL and SYCL (IWOCL) '24. Program chair for the conference International Workshop on OpenCL and SYCL (IWOCL) '24. Panelist and Co-Organizer of the Birds-of-a-Feather session C++ BoF: Welcome to C++ 23, ISC '24, Hamburg. Instructor and Co-Organizer of the Tutorial session Hands-On HPC Application Development Using C++ and SYCL, ISC'24, Hamburg. Instructor and Co-Organizer of the Tutorial session Advanced SYCL Tutorial, IWOCL '24, Chicago, US. Panelist at the panel discussion OpenCL and SYCL, IWOCL' 24, Chicago, US.

Ina Biermayer, Susan Eckerle, Elisa Brost, Wolfgang Müller:
Systems Biology of Mammalian Cells (SBMC 2024) Conference, *"Translating Systems Medicine into the clinics"*, Paulinum Leipzig, 13-15 May 2024 (organization).

Lynn Buchele:
Teaching Assistant and Organizer, MESA Down Under, Sydney, Australia.

Martin Golebiewski:
Chair of the Committee Meeting of ISO/TC 276 Biotechnology working group WG5 *"Data Processing and Integration"*, Glasgow, Scotland, UK, 17 - 22 June 2024. Member of the Organization Committee, COMBINE 2024 - Computational Modeling in Biology Network Meeting, Stuttgart, Germany, 1 - 4 September 2024. Workshop *"FAIR Data Infrastructures for Biomedical Communities"*, Gesundheit Gemeinsam 2024 - 69th annual meeting of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS), Dresden, Germany, 8 - 13 September 2024. Annual NFDI4Health Meeting, chair of the session *"Consolidation of the metadata schema and interfaces"*, Potsdam, Germany, 7 - 8 October 2024.

Martin Golebiewski and Gerhard Mayer:
Breakout session *"Standards for digital twins and their implementability"*, 1st Ecosystem meeting on building the Virtual Human Twin, Paris, France, 18-19 January 2024. Breakout session *"Standards for Virtual Human Twins"*, 2nd EDITH ecosystem meeting, Amsterdam, Netherlands, 15 - 16 July 2024.

Frauke Gräter:
Member of the Senate, Leibniz Association. Member of the Editorial Board of Biophysical Journal. Member of the coordinating committee of the excellence cluster *"3D Matter Made to Order"* (KIT and Heidelberg University).

Saskia Hekker:
AFP-Chair of European Space Science Committee, 2024. Spokesperson, IMPRS-HD, Heidelberg.

Eva Laplace:
Member of the Scientific Organizing Committee for the international workshop: *"From discovery to a population: benchmarking stripped stars and companions"*, Lorentz Center, Leiden, Netherlands 22-26 July 2024.

Wolfgang Müller:
ACM & IEEE Joint Conference on Digital Libraries, 16 - 20 December, Hong Kong, China.

Giulia Paiardi:
Co-organizer of the Career Workshop at the 24th EuroQSAR Symposium, Barcelona, Spain, 22-26 September 2024.

Kai Lars Polsterer:
Member of the Program Committee for *"ACM Conference for Reproducibility and Replicability"*, Inria, Rennes, France, 18-20 June 2024. Main organizer of *"ML4ASTRO - Machine Learning in Astrophysical Studies"*, Heidelberg, Germany, 25 June 2024. Organizer of the Splinter Meeting *"E-science and E-infrastructure"* at Astronomische Gesellschaft Meeting 2024, Cologne, Germany, 10-12 September 2024. Main organizer of Astroinformatics 2024, Llanuras de Diana, Puerto Natales, Patagonia, Chile, 9-13 December 2024.

Friedrich Röpke:
Member of the Scientific Organizing Committee for the conference *"360° approach to common envelope evolution from binary progenitors to remnants"*, Barcelona, Spain, 10–14 June 2024.

Ashley Ruiter:
Member of the Scientific Organizing Committee for the 23rd European White Dwarf Workshop, Barcelona, Spain, 8–12 July 2024.

Alexandros Stamatakis:
Organizer of 2024 Computational Molecular Evolution Summer School, Welcome Trust Genome Campus, Hinxton, UK. Main organizer of Legend2024 : Machine Learning for Evolutionary Genomics Data conference, FORTH, Crete, Greece.

Michael Strube:
General Chair of EACL 2024, The Eighteenth Conference of the European Chapter of the Asscociation for Computational Linguistics, St. Julians, Malta, 17-22 March 2024. Program Co-Chair of CODI 2024, The Fifth Workshop on Computational Approaches to Discourse at EACL 2024, St. Julians, Malta, 21 March 2024.

Rebecca Wade:
Co-organizer (as QCMS Scientific Chair) with Jordi Mestres (University of Girona, Chair), 24th EuroQSAR Symposium, Barcelona, Spain, 22-26 September 2024. Rebecca Wade (Chair), Rostislav Fedorov, Daniel Sucerquia (HITS) (with Pascal Friederich, Marcus Elstner, David Hoffmann (KIT)), 2nd SIMPLAIX Workshop on Machine Learning for Multiscale Molecular Modeling, Studio Villa Bosch, Heidelberg, 15-17 May 2024.

Ulrike Wittig:
COST Action OneHealthDrugs WG1 Workshop *"Structural and functional aspects of targets involved in vector borne diseases"*, online, 17 April 2024.

Other contributions

Jeong Yun Choi
"How do we study the stars with sound?" JuForum Sciencepub meets HITS, Cafe Leitstelle, Heidelberg, 14 May 2024. Session leader of *"AI tools in Research"*, IMPRS-Astro Hackathon, Haus der Astronomie (HdA), Heidelberg, Germany, 21-24 May 2024.

Mila Coetzee and Leif Seute:
"Machine Learning for Causal Modelling and Molecular Science", ZOrA - Zukunfts-Orientierungs-Akademie (public outreach presentation), Studio Villa Bosch & HITS, Heidelberg, 28 October 2024.

Robert Fisher:
"The Fascination of Exploding Stars," Heisenberg Gymnasium Bruchsal, Germany, February 27, 2024; Heidelberg Talks Online DAAD Seminar (Virtual), March 15, 2024; Physics Literacy for Kids Seminar (Virtual), June 19, 2024.

Saskia Hekker:
Interview with school student Stella Neumeyer from Georg-Wilhelm-Steller-Gymnasium Bad Windsheim on *"Female astronomers in History and today"*, 2024.

Debora Monego:
"Solving Biological Puzzles with Computer Simulations", HITS Open-House Event, Heidelberg, Germany, 13 July 2024.

Kai Lars Polsterer:
„Wie künstliche Intelligenz Astronomen bei der Arbeit hilft." Public Outreach talk at Planetarium Bochum, Bochum, Germany, 21 February 2024. *"A Universe in Data: Artificial Intelligence in Astronomy"*. Astronomy Open Night 2024, Macquarie University Wallumattagal Campus, Australia, 23 May 2024.

Fabian Schneider:
"The Sound of Silence: Wie klingen Schwarze Löcher?", HITS Open House Event, Heidelberg, Germany, 13 July 2024.

Leif Seute:
"Generating Proteins with Geometric Algebra Flow Matching". *"Science Insight"* Talk and Q & A at the *"Wissenswerte"* Conference, Heidelberg, 31 October 2024.

Jan Stühmer:
„Wie wird Faktenwissen in Sprachmodellen gespeichert?" ("Where is Factual Knowledge Stored in an LLM?), Science Notes (public outreach talk and Q & A), MAINS, Heidelberg, Germany, 25 January 2024; HITS Open House Event, Heidelberg, Germany, 13 July 2024.

10.5 Awards

Rostislav Fedorov:
Poster Award, 2nd Place, ChemAI Conference, Amsterdam, Netherlands, 29 No-vember 2024.

Luise Häuser:
Best paper award, SIGTYP2024 conference, Malta, 2024.

Mike Lau:
Best linguistic invention, 41st Liège International Astrophysical Colloquium: The eventful life of massive star multiples, Liège, Belgium, July 2024.

Stefan Machmeier and Vincent Heuveline:
Best Paper Award –, "Detecting DNS Tunnelling and Data Exfiltration using Dynamic Time Warping", in 2024 8th Cyber Security in Networking Conference (CSNet), Paris, France, 2024.

Ivo Seitenzahl:
Friedrich Wilhelm Bessel-Forschungspreis der Alexander von Humboldt-Stiftung.

Alexandros Stamatakis:
Highly Cited Researcher, Clarivate Analytics, 2024. Honored by president of KIT for being a highly cited researcher, November 2024.

Jonathan Teuffel:
Poster Award, Runner-up, 25th International Symposium on Microsomes and Drug Oxidations, Prague (Czech Republic), 7-10 July 2024.

Alexander Zeilmann:
Best Poster Award – , *"KI-Morph – User-friendly large-scale image analysis & AI on bwHPC systems"*, in 10th bwHPC Symposium, Freiburg, Germany, 2024.

11 Boards and Management

The HITS Scientific Advisory Board and the HITS management (November 2024). From left to right: Victoria Stodden, Kai Polsterer (HITS Deputy Scientific Director), Alex Szalay, Gesa Schönberger (HITS Managing Director), Thomas Lengauer, Barbara Ercolano, Viola Vogel, Jeffrey Brock.

Scientific Advisory Board

The HITS Scientific Advisory Board (SAB) is a group of internationally renowned scientists that supports the management of HITS in various aspects of running, planning, and directing the institute. The SAB is responsible for orchestrating the periodic evaluation of all 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 2024, 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**
  University of Southern California, Los Angeles, USA, (Vice Chair, SAB)
- **Prof. Dr. Alex Szalay**
  Johns Hopkins University, Baltimore, 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
© 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"

HITS Management

The HITS Management consists of the Managing Director and the Scientific Director. The Scientific Director is one of the group leaders appointed by the HITS shareholders for a period of two years and represents the Institute in all scientific matters vis-à-vis cooperation partners and the public.

Managing Director:
Dr. Gesa Schönberger

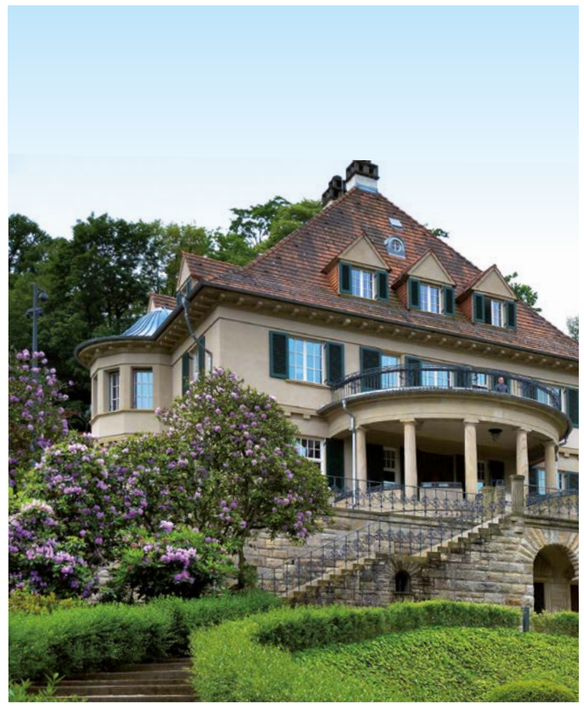
Scientific Director:
Prof. Dr. Tilmann Gneiting
(2023–2024)

Deputy Scientific Director:
Dr. Kai Polsterer
(2023–2024)

HITS

HITS, the Heidelberg Institute for Theoretical Studies, was established in 2010 by physicist and SAP co-founder Klaus Tschira (1940-2015) and the Klaus Tschira Foundation as a private, non-profit research institute. HITS conducts basic research in the natural, mathematical, and computer sciences. Major research directions include complex simulations across scales, making sense of data, and enabling science via computational research. Application areas range from molecular biology to astrophysics.

An essential characteristic of the Institute is interdisciplinarity, implemented in numerous cross-group and cross-disciplinary projects. The base funding of HITS is provided by the Klaus Tschira Foundation.



HITS gGmbH
Schloss-Wolfsbrunnenweg 35
D-69118 Heidelberg

Editor
Dr. Peter Saueressig
Head of Communications

Contact
info@h-its.org
Phone: +49 6221-533 533
www.h-its.org

Our e-mail addresses have the following structure:
firstname.lastname@h-its.org

Pictures
HITS gGmbH (unless otherwise indicated)

Layout and Design
FEUERWASSER | grafik . web . design
www.feuerwasser.de

ISSN 1438-4159 | © 2025 HITS gGmbH

Bluesky: hitsters.bsky.social
Facebook: /HITStudies
Youtube: /TheHITSters
LinkedIn: company/the-heidelberg-institute-for-theoretical-studies
Instagram: the_hitsters