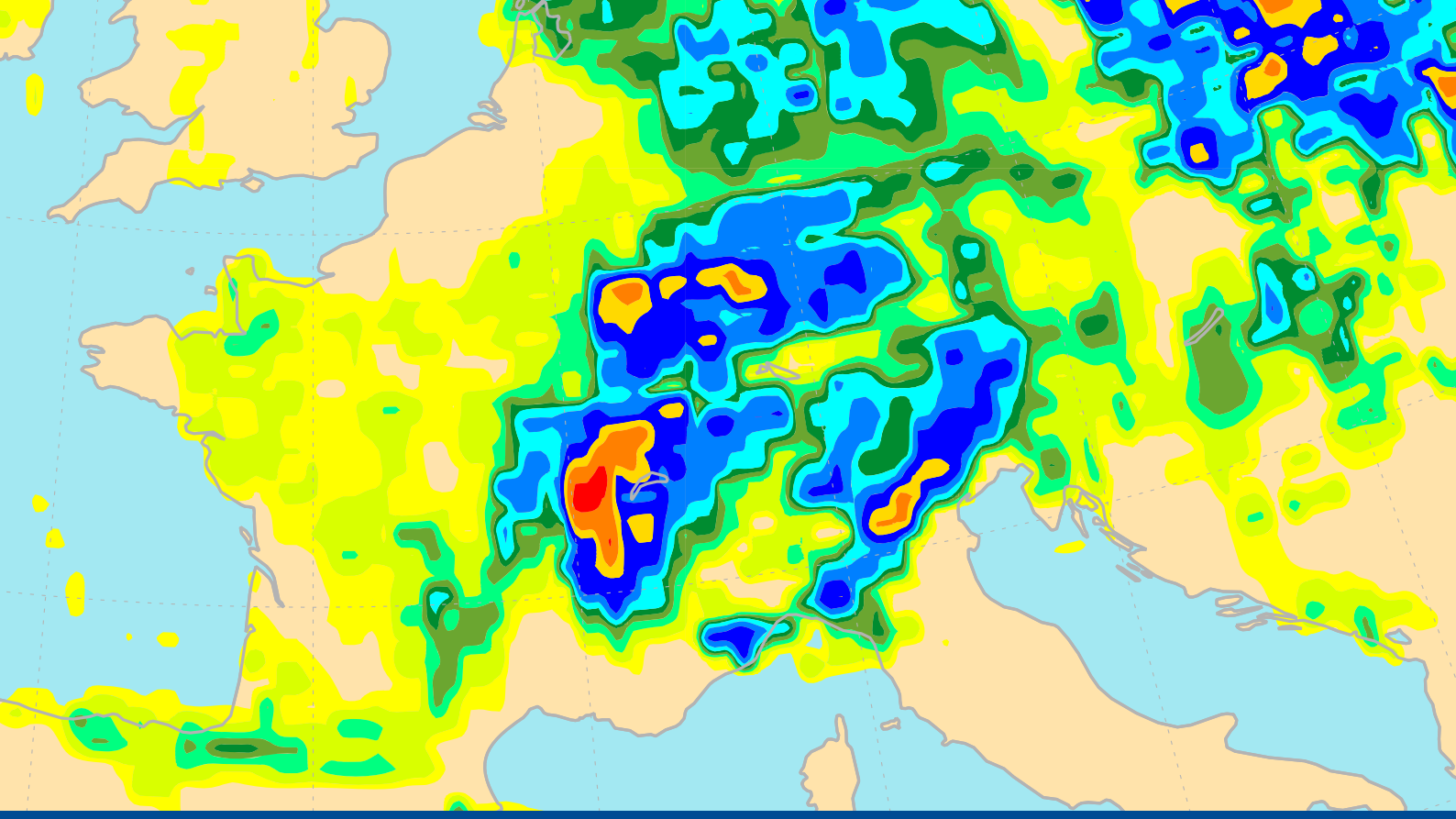


HITS 2014

Annual Report
Jahresbericht

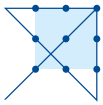
Heidelberg Institute for
Theoretical Studies

HITS

THINK BEYOND THE LIMITS!

Annual Report | Jahresbericht

Heidelberg Institute for
Theoretical Studies



HITS

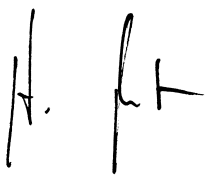
Klaus Tschira

Das Vorwort, das traditionell den Jahresbericht unter der Überschrift „Think beyond the limits“ einleitet, haben Klaus Tschira und ich immer zusammen verfasst, um bei dieser Gelegenheit darüber zu reflektieren, was im Berichtsjahr wichtig war und wo besondere Herausforderungen für die Zukunft zu finden waren. Das Vorwort zu diesem Jahresbericht ist der letzte Text, den wir gemeinsam verfasst haben, was ihm – zumindest im Hinblick auf das HITS – eine besondere Bedeutung gibt. Wenn man ihn jetzt, im Lichte der neuen Situation, liest, so enthält er zwei wichtige Botschaften.

Die erste lautet: Der Aufbau des HITS im Sinne der Ideen, die Klaus Tschira im Jahr 2007 entwickelt hatte, ist weitgehend abgeschlossen. Das ist natürlich nicht zu lesen als „Jetzt sind wir fertig.“; für ihn bedeutete die Feststellung, der Aufbau sei abgeschlossen, dass es jetzt richtig losgehen kann, dass das Konzept eines multidisziplinär angelegten Grundlagenforschungs-Institutes jetzt seinen Wert beweisen, seine Möglichkeiten erproben kann.

Die zweite Botschaft lautet: Es sind alle organisatorischen Vorkehrungen getroffen, um das HITS auf Dauer zu etablieren. Dies war eines von Klaus Tschiras besonderen Anliegen: Er wollte mit dem HITS eine neue (und neuartige) Einrichtung schaffen, die jetzt und in Zukunft eine maßgebliche Rolle in der naturwissenschaftlichen Grundlagenforschung (inklusive Mathematik und Informatik) spielt. Dazu musste eine stabile Grundfinanzierung sichergestellt werden, und es brauchte einen Gesellschaftsvertrag, der den besonderen Erfordernissen eines privaten Forschungsinstituts mit engen Verbindungen zu öffentlichen finanzierten Forschungseinrichtungen Rechnung trägt. Beides ist mit der Einrichtung der HITS Stiftung und mit der neuen Gesellschaftssatzung des HITS, die im Oktober 2014 in Kraft getreten ist, gelungen, und es war Klaus Tschira ein besonderes Anliegen, dies im Vorwort zum Jahresbericht zu betonen.

Er hat alles dafür getan, „sein“ Institut so aufzustellen, dass es seinen Ideen gemäß arbeiten und den Nutzen interdisziplinärer Forschung an vielen Beispielen (be-)greifbar machen kann. Sein überraschender und viel zu früher Tod hat nun verhindert, dass die Realisierung seiner Vision miterleben und begleiten kann. Wir merken jeden Tag in vielen großen und kleinen Dingen, wie sehr Klaus Tschira, der Gründer, der Ideengeber, der Anreger und der Ermutiger, fehlt. Was vor allem fehlt, ist der unglaublich weite Horizont seiner Interessen und Aktivitäten, die die Arbeit des HITS in vielfältiger Weise ergänzt und befruchtet haben, auch wenn sie ursprünglich gar nichts mit den laufenden Forschungsprojekten zu tun hatten. Beispiele hierfür sind Explore Science (siehe Kapitel 5.3) und die Zusammenarbeit mit der Europäischen Südsternwarte (ESO) für den Bau des „ESO Supernova“ Planetariums und Besucherzentrums. Auch das Heidelberg Laureate Forum, das seinen Ursprung in einem Vorschlag des HITS hat, ist ein Beleg dafür, wie ein weitblickender Förderer und passionierter Wissenschaftsförderer aus einem kleinen Projekt eine Veranstaltung mit globalem Format machen kann. Es ließen sich noch viele Beispiele finden, die allesamt illustrieren, mit wieviel Engagement, mit wieviel Sachkenntnis und mit wieviel Herzblut sich Klaus Tschira seiner Sache verschrieben hat: der Förderung der wissenschaftlichen Forschung und der Vermittlung ihrer Methoden und Ergebnisse an eine breite Öffentlichkeit. Wir trauern um unseren Gründer, einen Mann, der für den Wissenschaftsstandort Deutschland so viel getan hat wie kaum ein Einzelner sonst, und wir sind uns der Verantwortung bewusst, die er uns – zusammen mit nachgerade idealen Arbeitsbedingungen – hinterlassen hat: Das HITS zu einer weltweit anerkannten Forschungseinrichtung zu machen, in der hervorragende Wissenschaftlerinnen und Wissenschaftler ihre eigene Forschungsagenda verfolgen können. Ich hoffe (und glaube zuversichtlich), dass wir dieser Verantwortung gerecht werden und zu jedem Zeitpunkt sagen können: Klaus wäre stolz auf sein Institut.



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

07.12.1940 - 31.03.2015

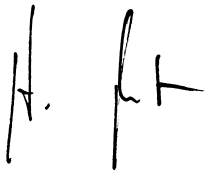
Klaus Tschira

The foreword that traditionally introduced the annual report under the heading “Think Beyond the Limits” was always written by Klaus Tschira and me together in order to reflect upon what had been important in the previous year and where particular challenges for the future might lie. The foreword to this annual report is the last text that we wrote together, which makes it particularly significant, at least in terms of HITS. If read in light of the current circumstances, the text contains two important messages.

The first message is that the development of HITS in terms of the ideas that Klaus Tschira created in 2007 has largely been completed. This of course does not mean that we are now finished with everything. For Klaus, the meaning of the phrase “the development is complete” is that everything can now finally begin, that the concept of a multi-disciplinary basic research institute can now test the waters and show its true colors.

The second message is that all necessary organizational measures to permanently establish HITS have been achieved. This was a particular concern of Klaus Tschira’s. With HITS, he wanted to create a new (and novel) institution that would play a significant role in basic natural science research (including mathematics and computer science) both now and in the future. In order to do this, it was necessary to ensure stable basic funding and to create a social contract that takes into account the special needs of a private research institute with close ties to publicly funded research organizations. Both of these requirements were achieved with the establishment of the HITS Foundation and with the new HITS charter that took effect in October 2014, and it was very important to Klaus Tschira to emphasize this fact in the foreword of the Annual Report.

Klaus did everything to set up “his” institute so that it would work in accordance with his ideas and could make the benefits of interdisciplinary research both tangible and understandable in many areas. His surprising and untimely death prevented him from being able to experience and bask in the realization of his vision. Every day, we notice in many large and small things how much we miss Klaus Tschira, the founder, the creator of ideas, the inspirer, and the encourager. Above all, we miss the incredibly broad spectrum of his interests and activities that supported and stimulated the work at HITS in many ways, even if they originally had nothing to do with the ongoing research projects. Examples include Explore Science (see Chapter 5.3) and the collaboration with the European Southern Observatory (ESO) to build the „ESO Supernova Planetarium & Visitor Centre“. Even the Heidelberg Laureate Forum, whose origin lies in a proposal from HITS, is an example of how a far-sighted supporter and passionate science promoter can create a global event out of a small project.
It would be easy to list many other examples that illustrate the tremendous commitment, knowledge, and passion with which Klaus Tschira dedicated himself to the promotion of scientific research and the imparting of its methodologies and results to a broad public. We deeply mourn the loss of our founder, a man who did more for the position of science in Germany than almost anyone else, and we are aware of the responsibility – along with the virtually ideal working conditions – that he left behind for us: making HITS an internationally recognized research facility in which outstanding scientific researchers can pursue their own research agenda. I hope (and confidently believe) that we can meet this goal and will be able to say that Klaus would be proud of his institute.



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

December 7, 1940 - March 31, 2015

1 Think Beyond the Limits!

Die erste Aufbauphase des HITS ist in allen wesentlichen Punkten abgeschlossen: Prof. Dr. Friedrich Röpke hat den Ruf der Universität Heidelberg angenommen; damit verbunden ist die Übernahme der Leitung einer neuen Forschungsgruppe am HITS (der elften des Instituts) mit dem Titel „Physics of Stellar Objects“. Diese Gruppe erforscht die physikalischen Vorgänge in Sternen, insbesondere das Phänomen der Supernovae, und versucht, die entsprechenden theoretischen Modelle mit detaillierten Computer-Simulationen zu validieren. Mit den beiden schon vorhandenen Forschungsgruppen „Theoretische Astrophysik“ und “Astro-Informatik“ verfügt das HITS nun über eine – nicht nur gemessen an der Größe des Instituts – sehr starke Forschungskompetenz in den Bereichen Astronomie und Astrophysik.

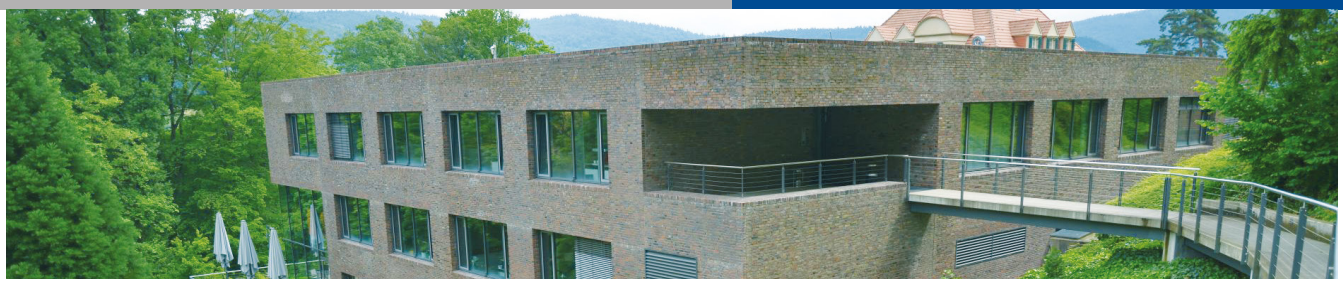
Das Berufungsverfahren für die letzte im Aufbauplan vorgesehene Forschungsgruppe („Wissenschaftliche Visualisierung“) ist auf den Weg gebracht worden, d.h. alle für die gemeinsame Berufung mit der Universität Heidelberg erforderlichen Gremienbeschlüsse liegen vor. Wir hoffen, dass das Verfahren bis Ende 2015 abgeschlossen werden kann.

Da das Institut nun seine „Reiseflughöhe“ erreicht hat, war es erforderlich, ihm eine Struktur zu geben, die einen langfristigen erfolgreichen Betrieb ermöglicht und dabei insbesondere den spezifischen Erfordernissen eines Forschungsinstituts Rechnung trägt. Zu diesen Erfordernissen zählen u.a. die Beteiligung der Wissenschaftler an allen Aspekten der Institutsleitung und die Etablierung von klaren Prozeduren zur Entscheidung darüber, ob neue Gruppen eingerichtet werden sollen, oder ob existierende Gruppen evtl. inhaltlich neu ausgerichtet werden müssen. Zu diesem Zweck wurde eine umfangreiche Änderung der Gesellschaftssatzung durchgeführt, deren wesentliche Punkte die folgenden sind:

- Die Universität Heidelberg und das KIT wurden als Gesellschafter in die gGmbH aufgenommen.
- Hauptgesellschafter ist die neu gegründete HITS Stiftung. Der primäre Zweck der HITS Stiftung besteht darin, dem HITS die für seinen Betrieb erforderliche Grundfinanzierung bereitzustellen. Die Mittel dafür erhält sie zunächst ausschließlich von der Klaus Tschira Stiftung (KTS).
- Das HITS hat einen Institutssprecher; diese Rolle wird reihum von den Leitern der regulären Forschungsgruppen wahrgenommen.
- Der Institutssprecher bildet zusammen mit dem Geschäftsführer eine Doppelspitze; alle Entscheidungen müssen gemeinsam getroffen werden.
- Die gezielte Förderung der wissenschaftlichen Mitarbeiter wurde explizit im Gesellschaftsvertrag verankert.

Wir sind sicher, dass mit diesem Regelwerk der Institutsbetrieb einerseits weiter so effektiv wie bisher durchgeführt werden kann, dass aber andererseits alle nötigen Vorkehrungen getroffen wurden, um künftig auftretenden neuen Situationen gerecht zu werden.

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



The first start-up phase of HITS has been completed: Prof. Dr. Friedrich Röpke has accepted a professorship at Heidelberg University and has thus become the leader of the (eleventh) HITS research group, “Physics of Stellar Objects”. This group examines the physical processes in stars (with particular focus on supernovas) and tries to provide evidence for the respective theoretical models with detailed computer simulations. Together with the PSO-group and the other groups “Theoretical Astrophysics” and “Astro Informatics”, HITS now has a very broad basis for astronomic and astrophysical research, especially for an institute of its size.

The appointment procedure for the last research group planned for the start-up phase (“Scientific Visualization”) has been launched, which means all administrative prerequisites for the joint appointment with Heidelberg University have been met. We hope to conclude the process by the end of 2015.

Now that the institute has reached its “cruising altitude”, we have implemented a new structure to ensure a long-term run with regard to the specific requirements of a research institute. Among these requirements is the participation of our scientists in all aspects of the management of the institute. Another goal was the establishment of clear procedures for the set-up of new groups and the reorganization of existing groups.

In order to implement these changes, we have amended our statutes. The main changes are the following:

- Heidelberg University and KIT are now shareholders of HITS gGmbH.
- The main shareholder is the newly established “HITS Stiftung” (HITS Foundation), whose main purpose is to secure the base funding of the institute. Financial support for the HITS Stiftung is initially provided by the Klaus Tschira Stiftung only.
- HITS has a Scientific Director. This function is assumed by the research group leaders in turns.
- All decisions must be taken by both the Scientific Director and the Managing Director.
- Means for specific support of our scientific employees are fixed in the statutes.

We are convinced that this reorganization guarantees that the institute will continue running as effectively as before while also setting the framework for any future changes and challenges.

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

2 Research

2.1 Junior Group Astroinformatics (AIN)

The Astroinformatics group develops new methods and tools to deal with the currently available complex, heterogeneous, and large datasets in astronomy.

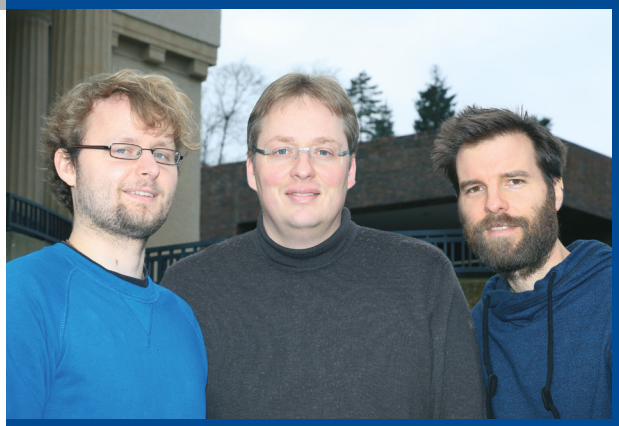
Over the last two decades, computers have revolutionized astronomy. Due to advances in technology, new detectors, complex instruments, and innovative telescope designs have been able to be realized. These advances enable today's astronomers to observe objects to an unprecedented extent and with high spatial/spectral/temporal resolution. In addition, there are new untapped wavelength-regimes still to be investigated. Dedicated survey telescopes map the sky and constantly collect data. We enable scientists to analyze this increasing amount of information.

The group is interested in the development of improved photometric redshift regression models. This is a key tool for the analysis of the data of upcoming large survey projects like the Square Kilometer Array (SKA), Gaia, and Euclid. Another scientific interest is methods and tools for the extraction and filtering of rare objects for detailed follow-up analysis with 8-m class telescopes. With estimated occurrences of only a few objects per million, a manual inspection of the existing catalogs is not possible. The group's other interests include the morphological classification of galaxies based on imaging data as well as measuring similarity in high dimensional data spaces. Both processes aim at providing astronomers with tools that will allow them to explore the data archives.

Die Astroinformatik Gruppe entwickelt neue Methoden und Werkzeuge, um eine Analyse, der heutzutage verfügbaren komplexen, heterogenen und großen Daten im Bereich der Astronomie, zu ermöglichen.

In den letzten zwanzig Jahren hat der Einsatz von Computern die Astronomie stark beeinflusst. Durch technologische Fortschritte wurde es möglich neue Detektoren sowie innovative Instrumente und Teleskopdesigns zu realisieren. Dadurch können Astronomen nun Objekte mit bisher unerreichtem Detailreichtum und in neuen Wellenlängenbereichen beobachten. Mit speziell dafür vorgesehenen Teleskopen wird der Himmel jede Nacht beobachtet und die so gewonnen Daten werden frei zur Verfügung gestellt. Durch unsere Forschung ermöglichen wir es Wissenschaftlern, diese riesigen Datenmengen durch neue Analysemethoden effizienter zu nutzen.

Unsere Gruppe beschäftigt sich mit der Entwicklung photometrischer Rotverschiebungsmodelle. Diese werden für die neuen Generationen von Himmelsdurchmusterungen, benötigt. Des Weiteren beschäftigen wir uns mit der Suche nach astronomischen Objekten, die mit einer Häufigkeit von ein paar wenigen pro Million vorkommen. Um solch seltene Objekte für detaillierte Untersuchungen zu finden, scheidet die manuelle Selektion aus. Die morphologische Klassifikation von Galaxien sowie hochdimensionale Ähnlichkeitsmaße sind weitere Forschungsbereiche. Beide Bereiche werden benötigt, um einen explorativeren Datenzugang für die Astronomen zu schaffen.



The AIN group in 2014 (f.l.t.r.): Dennis Kügler, Kai Polsterer, Nikos Gianniotis

Group Leader
Dr. Kai Polsterer

Staff Member
Dr. Nikos Gianniotis (since July 2014)

Scholarship Holder
Sven Dennis Kügler (HITS Scholarship)

MORPHOLOGY OF GALAXIES

In the last two decades, more and more all-sky surveys have created an enormous amount of data, which is publicly available on the Internet. Today, it is impossible for an expert to manually inspect all objects in the available large-scale astronomical databases. Due to the exponential growth in size and complexity of the data-sets in astronomy, new methods for explorative analysis are required. Our goal is to enable astronomers to efficiently perform a morphological analysis on huge amounts of pre-processed data (e.g., images or radio-synthesis data). Crowd-sourcing projects such as GalaxyZoo and Radio-GalaxyZoo have encouraged users from all over the world to manually conduct various classification tasks. The GalaxyZoo project is a good example of how to make use of more than 100,000 volunteers to derive a morphological analysis for about 900,000 galaxies (Lintott et al. 2011). The combination of the pattern-recognition capabilities of thousands of volunteers enabled scientists to finish the data analysis within acceptable time. For upcoming surveys with billions of sources, however, this approach is no longer feasible. Therefore, new methods need to be developed that combine semi-automatic data analysis schemes with the visual recognition capabilities, creativity, and keen perception of the human brain. By using computers to pre-process and pre-analyze the data, we wish to assist astronomers to conduct such tasks in a semi-automatic manner instead of using a fully manual analysis via crowd-sourcing projects. Similar and frequent objects can be combined/sorted by machine learning models, which yield only a single representative that needs manual inspection by the scientist. In the past, dimensionality reduction techniques that are able to compute topological maps, i.e., latent embeddings, showed good results (Kramer et al. 2013) based on images of galaxies from the Sloan Digital Sky Survey (SDSS) (Ahn et al. 2014). These dimension reduction techniques aim at projecting complex, high-dimensional data to low-dimensional feature spaces while preserving similarities and neighborhood relations between the original data points.

We have developed an unsupervised method that can automatically process large amounts of galaxy images and that generates a set of prototypes. This resulting model can be used both to visualize the given galaxy data as well as to classify currently unseen images. We employ a modified version of self-organizing maps (Kohonen 1989), which is a type of artificial neural network. Compared with several other non-linear dimensionality reduction techniques, we have found that our method can produce embeddings of higher discriminating quality. Our Parallelized rotation/flipping INvariant Kohonen map (PINK)

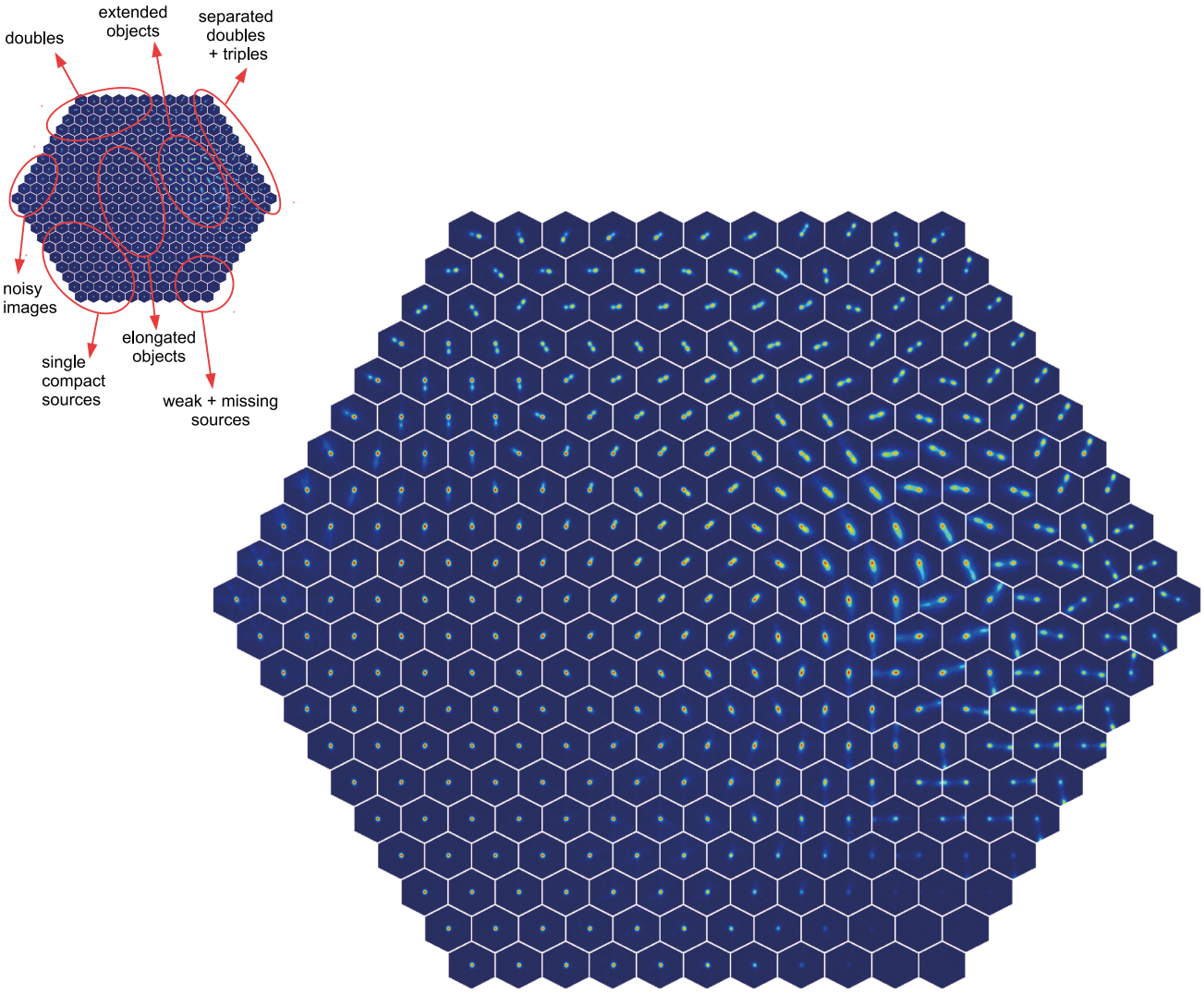


Fig.1: Map of morphological prototypes of galaxies as derived with PINK based on radio-synthesis data.

framework makes use of multi-core CPU/GPU environments to speed up calculations. The PINK framework allows for training maps either with a rectangular or hexagonal grid, both in a continuous repeating or edge-limited version. A hexagonal grid defines denser neighborhoods than a rectangular grid and hence allows for smoother

maps. We could not distinguish significant differences when comparing the two types of grids other than the fact that corner effects on the edge-limited version are less pronounced on the hexagonal grid. After the training phase is finished, it is possible to match an image/pattern to the derived prototypes and thereby retrieve a coordinate on the map. By inspecting and annotating the derived prototypes, a scientist inspects all matching objects at once. Therefore, the amount of objects to be inspected is reduced to the number of prototypes on the map.

When inspecting images by eye, the brain automatically scales, aligns, distorts, and interpolates the information such that objects are perceived to be similar or not. Pre-processing the images to align them to the principal axis of their main component and using a simple pixel-wise Euclidean distance was one of the first approaches to deal with rotation. In the past, we carried out multiple tests with rotation invariant similarity measures (Polsterer et al. 2012). Up to now, we have achieved the best results with Fourier transformed circular slices of the images. This method has the limitation of losing the information of complex and weak structures and therefore only allows for a sorting based on dominant morphological features. For the imaging data at hand, a rotation and flipping invariant similarity measure is essential to achieve satisfying results. To calculate similarity, our approach basically calculates Euclidean distances for all possible rotations/flipped/un-flipped objects on the map to determine the best match. This operation can be shown to give rise to a valid distance metric. Since the considered brute-force comparisons between an image and all the neurons are computationally very demanding, this task is an ideal candidate for massively parallel implementations.

To test the performance and usability of our approach, we performed experiments on synthetic data as well as on real astronomical images. As the synthetic data was simply

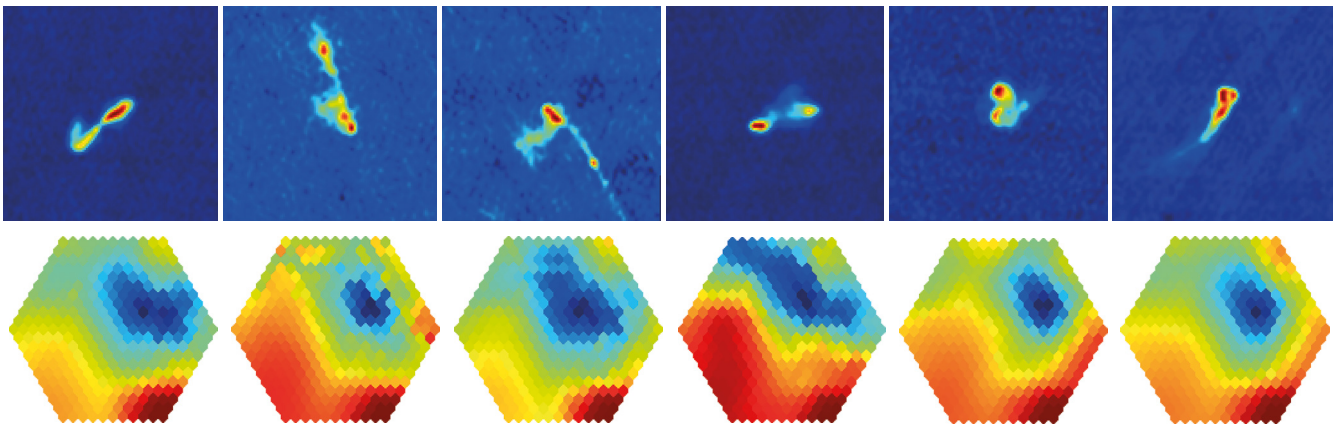


Fig. 2: Outliers selected based on their absolute similarity to the prototypes on the map. The heatmap visualizes the distance to the individual prototypes on the map, where blue denotes high similarity and red reflects dissimilarity.

used to corroborate the correctness of our algorithm, we only show results on the real-world data. After having performed the training on the 200k images from Radio Galaxy-Zoo, we retrieved the map presented in Fig. 1 (see page 12). This resulting map shows the derived prototypes, which allow for a clear separation into different morphological classes. Based on their mapping to the prototypes, it is possible to transfer the annotations created for the map directly to every individual image. Objects that are not well represented by the prototypes can be directly extracted by using the absolute similarity value of the best match. In our experiment, just a fraction of a percent turned out to be such outliers based on the analysis of the distribution of the absolute similarity values. They constitute the interesting objects that an expert would have to manually mine for (Fig.2, see page 13).

AUTOENCODING TIME SERIES FOR VISUALIZATION

Time series are often considered a challenging data type to handle in machine learning tasks. Their variable-length nature has forced the derivation of feature vectors that capture various characteristics, e.g. Richards et al. 2011. However, it is unclear how well such (often handcrafted) features express the potentially complex latent dynamics of time series. Time series exhibit long-term dependencies, which must be taken into account when comparing two time series for similarity. This temporal nature makes the use of common designs, e.g. RBF kernels, problematic. Hence, more attentive algorithmic designs are needed, and there have indeed been works in classification scenarios (Jaakkola and Haussler 1998, Jebara et al. 2004, Chen et al. 2013) that successfully account for the particular nature of time series.

We are interested in visualizing time series by dimensionality reduction. Therefore, we propose a fixed-length vector representation that is based on a special type of neural network called the Echo State Network (ESN) (Jaeger 2001). ESNs are powerful recurrent discrete-time neural networks with a large, fixed hidden layer. Typically, the hidden layer is constructed in a randomized manner; however, we take advantage of simple, cyclic, deterministically generated reservoirs that perform up to par with the standard ESN (Rodan and Tino 2011). The great advantage of ESNs is the fact that the hidden part, the reservoir of nodes, is fixed, and only the readout weights need to be trained. Given a fixed ESN reservoir, for each time series in the dataset, we determine its best readout weight vector and take this vector to be the series' new representation with respect to this reservoir.

In a second stage, we employ an autoencoder (Kramer 1991) that reduces the dimensionality of the readout weight vectors. The autoencoder is a type of neural network that compresses high-dimensional data into low-dimensional representations. It defines a fan-in fan-out architecture, with the middle layer composed of a small number of neurons referred to as the 'bottleneck'. When data are propagated through the network, the bottleneck forces the autoencoder to reduce the dimensionality of the data. By setting the number of neurons in the bottleneck to three, the bottleneck activations can be interpreted as three-dimensional projection coordinates and used for visualization.

The autoencoder learns an identity mapping by training on targets identical to the inputs. However, the presence of the bottleneck forces the autoencoder to compress the inputs, and the output is hence only an approximate reconstruction of the input. Typically, the L2 norm is employed as an objective function for measuring how well the reconstructed inputs approximate the original data inputs.

In our case, the L2 is inappropriate; what we are really interested in is not how well the readout weight vectors are reconstructed in the L2 sense, but how well each reconstructed readout weight vector can still reproduce its respective time series when plugged back to the same, fixed ESN reservoir. To that end, we introduce a more suitable objective function for measuring reconstruction quality according to this new view.

The proposed algorithm can handle out-of-sample data and hence, apart from projecting training data only, we also project unseen test data. Moreover, we additionally constructed visualizations using the popular t-SNE algorithm (van der Maaten and G. Hinton 2008). In Fig. 3 and Fig. 4 we used data from Harikrishnan et al. 2011 concerning a black hole binary system that expresses various types of temporal regimes that vary over a wide range of time scales. We extracted subsequences of length 1000 from regimes β,γ,κ,χ that were chosen on account of their similarity. Unlike t-SNE, which operates directly on the raw data, the proposed algorithm can capture the differences between the time series in lower-dimensional space. This is because our method explicitly accounts for the sequential nature of time-series; learning is performed in the space of readout weight representations and is guided by an objective function that quantifies the reconstruction error in a principled manner. Moreover, our method, by its very nature, is also capable of projecting

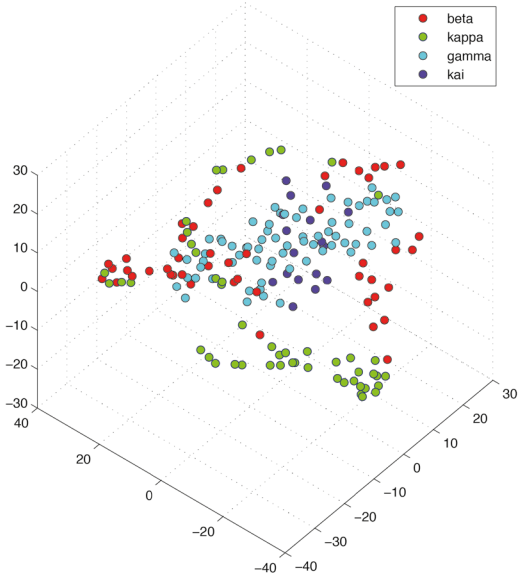


Fig. 3: Projection generated with the t-SNE algorithm of X-ray data from black hole binaries. Colors represent different classes of time-series.

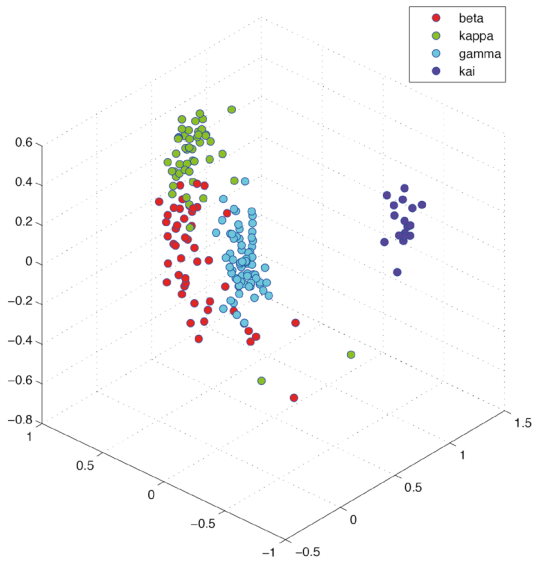


Fig. 4: Same black hole data as in Fig. 3, projected with our approach using autoencoders and echo-state networks.

unseen hold-out data. Finally, as demonstrated for the time-series data with astronomical application, we see that the visualization captures important characteristics of the time series that are otherwise ignored when treating the time series merely as vectorial data.

HIGH-DIMENSIONAL FEATURE SPACES

Analyzing datasets with a large number (>10^6) of high-dimensional (>10^3) but well-structured objects is an important challenge in astronomy. For the huge available catalogs (e.g. spectra of SDSS DR10, Ahn et al. 2014), most machine learning techniques are struck by the "curse of dimensionality" (Bellman & Bellman 1961). As a consequence, the selection of straightforward features becomes very complex if the full information content is used. Together with the continuously increasing number of entities in the databases, the number of rare but interesting objects increases accordingly. Thus, techniques that are more flexible than the classical model-based approaches are required, which are able to deal with the variety of observed behaviors. Together with the DMQ-group (see Section 2.4), we investigate a more general approach that allows for clustering, classification, and outlier detection in such complex and rich databases.

PHOTOMETRIC REDSHIFT REGRESSION MODELS

Photometric redshift estimation models are an important tool in astronomy. Large-scale all-sky surveys are typically based on broadband imaging. For this reason, only a coarse analysis of specific object parameters is possible for most of the detected objects. To verify the true nature of a particular object, spectroscopic follow-up observations (with a higher resolution) are usually required. Since obtaining such spectra is much more time-consuming than broadband imaging, detailed information is only available for a relatively small subset of detected objects. Therefore, regression tasks are common in astronomy, e.g., to estimate the redshift or the metallicity of a galaxy. From a data mining perspective, the photometric objects that are spectroscopically observed form the basis for generating new models, which in turn can then be applied

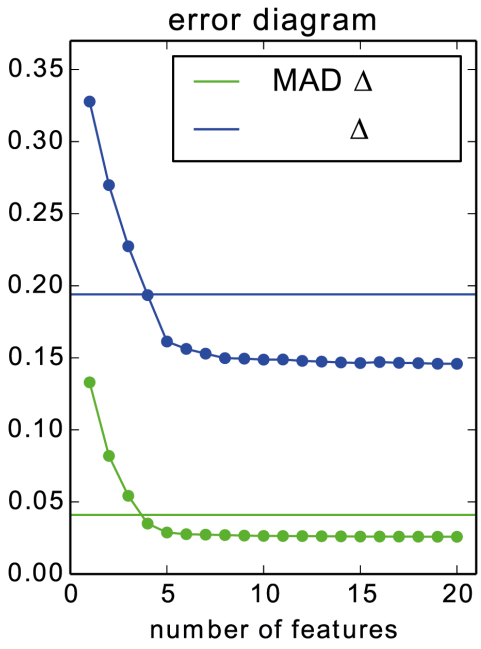


Fig. 5: Improvement in redshift estimation performance, with up to 20 selected features. A performance boost can already be observed with 4 selected features compared with the performance with just 8 standard features (horizontal lines).

to all remaining objects in the catalog. Classification and regression models based on photometric features are typical examples for such models (Bolzonella et al. 2000, Laurino et al. 2011). Besides the lack of labeled data for specific learning tasks, the available features might differ in explanatory power. This directly leads to the question of selecting good features, i.e., a representative subset of all possible features. The features that are extracted for each band range from plain magnitudes up to more complex, composed features.

Usually, the colors derived from adjacent filter bands are fed to appropriate regression techniques to solve the redshift estimation task (O'Mill et al. 2011, Wu et al. 2010). However, we considered an alternative approach: Instead of resorting to these standard features, we considered a large set of combinations of features and ignored any particular domain knowledge of the input parameters' physical properties. Instead of trying to improve the regression technique itself, we concentrated on selecting the best performing subset of features. Selecting a feature subset is, however, of combinatorial nature, and quickly becomes infeasible, even for moderate subset sizes. To accelerate the inevitable search, we therefore resorted to the massive computational resources that are provided by today's graphics processing units (GPUs) (Gieseke et al. 2014, Polsterer et al 2013). When an exhaustive test is impractical, a greedy forward selection approach, which adds the best performing features consecutively, is applied. The features selected by our framework lead to a significantly better prediction performance of e.g., photometric redshifts (Fig. 5) compared with the standard features given a nearest neighbor regression model used for this task. The sets of features found in this manner are then used for more complex and even better-performing approaches.

We are currently involved in determining appropriate features based on different catalogs for various scientific projects. The discovered features often result in long discussions with scientific partners, as improvement to the models cannot be related to simple relations/physical effects. So far, the detailed analysis of the results have often revealed relations in the data that have been neglected up to now but that help to improve the results of the regression tasks considerably.

SPECTROSCOPIC REDSHIFT VALIDATION

The determination of spectroscopic redshift has so far been treated in a classical model-based approach. As a basis for the creation of the models, 20,000 of the spectra were manually investigated by domain experts and were then assigned a class and a redshift. A set of templates was created from these hand-vetted spectra to describe the continuum behavior of the reference spectra. Additionally, a principal component analysis (PCA) was performed to extract the spectral features. Finally, the classification was performed by minimizing the deviation between model and data by aligning the continuum, the redshift, and the power of the individual principal components. This approach inevitably yields truncated results since the selection of the potential classes as well as the order of the PCA have a major impact on the classification and the redshift results. Additionally, the reference sample itself might not be representative (e.g., be missing potential classes or have a too low number of high-redshifted objects). Another limitation of the model-based approach is that the constrained models can only describe behaviors that are covered by one of the models. As a consequence, potentially interesting objects (e.g. super massive black hole binaries, Smith et al. 2010 & Tsalmantza et al. 2011, or gravitational lenses, Inada et al. 2012) are assigned to incorrect classes and cannot be detected based on their erroneous classification.

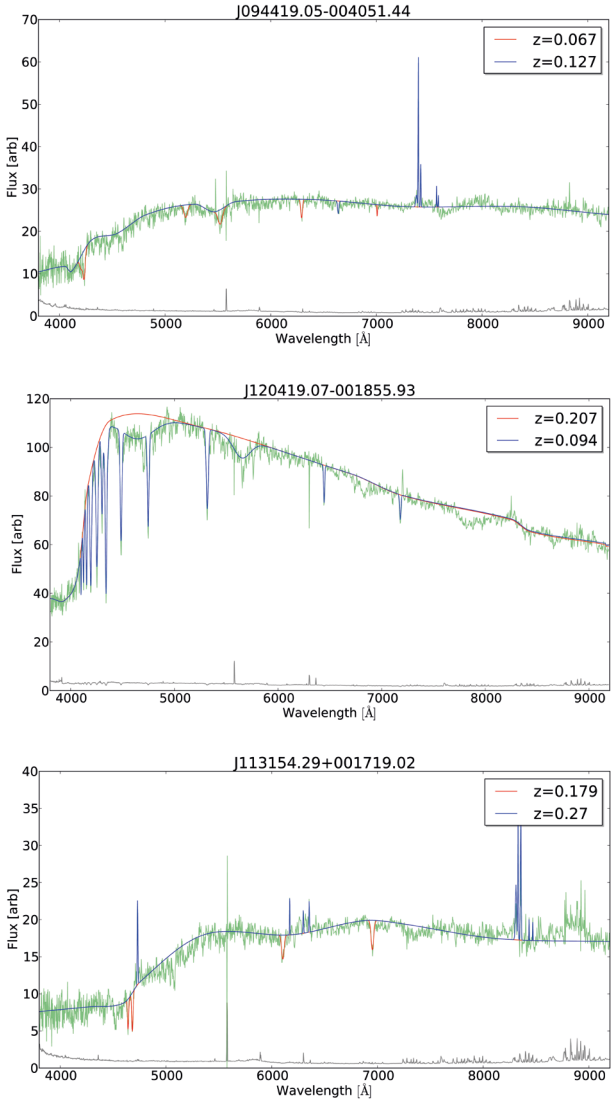


Fig. 6: SDSS spectra of objects with multiple redshift components (SDSS fit in red; our result in blue) that have been discovered with our approach.

In our work, we focus on a regression task of the well-described redshift value without including any prior knowledge about the structure of the complex data itself. As this approach is so general, it could easily be used to determine other properties, such as metallicity or star-burst ratio. The idea of our approach is to validate the result of the redshift calculation pipeline of SDSS, which performed well for the majority of the objects. For a few but still significant number of objects (≈1‰ of the spectra, which corresponds to a total number of a few thousand in total), wrong values were assigned. We applied a k-nearest-neighbors approach to a subset (~16,000) of the database, which was preprocessed to ensure scaling invariance and corrected for continuum effects. To provide independent redshift measures, we divided the spectra into an emission and an absorption component and compared them to the corresponding part of all the other available spectra. Based on this comparison, a redshift was determined statistically. Even though our values are based on the redshifts obtained by SDSS, we were able to reach a higher precision than the model-based method of SDSS (scales with an inverse square root of k). Additionally, we are easily able to detect objects with differences between emission and absorption (Fig. 6). We can even further increase the information gained with our method by dividing the spectra into specific regions of individual spectral features instead of merely dividing them into emission and absorption components.

As a result, we discovered in our subset 14 spectra with two redshift components, 10 of which had not been properly processed since the data reduction pipeline did a bad job at accounting for the night sky. Although the benefits of applying such an expensive method to the whole dataset seem not to pay off if they only find a few objects with incorrect values, the detected objects are so rare and scientifically important that validation techniques like our approach are mandatory for future surveys.

LUCI

The Astroinformatics group is involved in the LUCI project. LUCI is a pair of near-infrared imagers and spectrographs at the Large Binocular Telescope (LBT), the world largest optical telescope. The innovative new approach of the LBT is to combine two 8.4-m telescopes on a single mount to enable the interferometric combination of both mirrors. We have been responsible for the management of the control software development as well as the development of the observation preparation tool (Fig. 7). In 2014, we contributed to the binocular scheduling framework, the binocular user interaction, as well as the binocular observation preparation to provide access to the full binocular capability of the LBT.

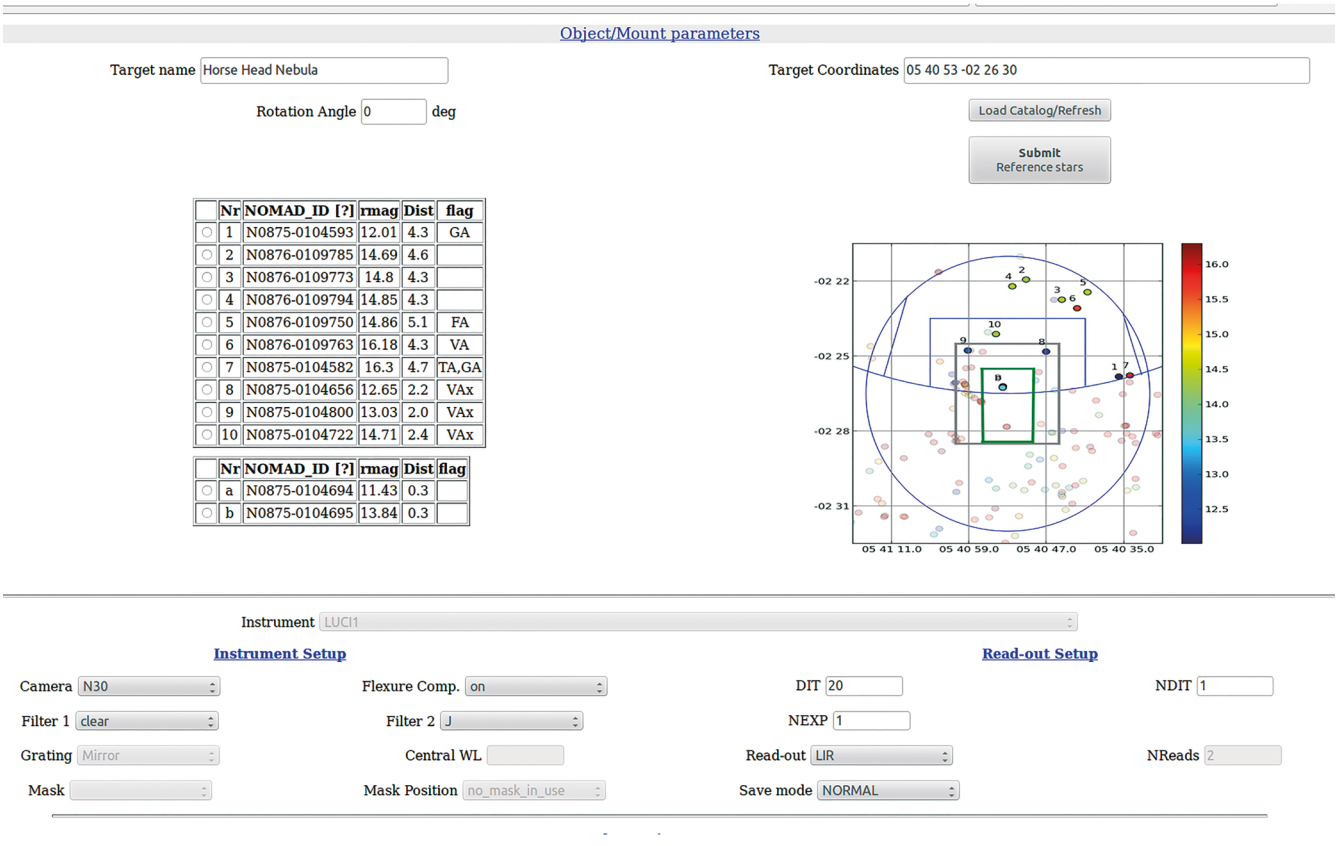


Fig. 7: New binocular observation preparation tool.

2.2 Computational Biology (CBI)

The Computational Biology Junior Group (CBI) began its work at HITS in 2013 and grew over the course of the year to its current size of four. Philipp Kämpfer and Philipp Bongartz, two PhD scholarship holders, joined in June and August, respectively, and Martin Pippel joined in July as a postdoc. Furthermore, the group receives mentorship from Gene Myers, one of the pioneers in the field of genome assembly. Gene is a director at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden and holds the Klaus Tschira Chair of Systems Biology.

The CBI group works at the interface(s) between computer science, mathematics, and the biological sciences. Our research focuses on the computational and algorithmic foundations of genome biology. Of the multitude of issues encountered in that field, we are especially interested in whole-genome assembly, i.e., the reconstruction of a genome's sequence from the data produced by a DNA sequencer. The basic principle applied for assembly is to randomly (over-)sample overlapping fragments from the genome, sequence them, and computationally reconstruct the full sequence from these fragments.

The complexity of this task is largely dependent on two characteristics of the fragments, i.e., average length and accuracy. The current generation of sequencers produces very long fragments, but with high error rates, so new approaches to the problem of efficient assembly under such conditions are needed. The development of such algorithms and their efficient implementation and application in genome sequencing projects are the main goals of the group.

Die Computational Biology Junior Group (CBI) begann ihre Arbeit am HITS Anfang 2013 und wuchs im Laufe des Jahres zu der aktuellen Größe von vier Mitgliedern. Zwei Promotionsstipendiaten, Philipp Kämpfer und Philipp Bongartz, starten Juni bzw. August und Martin Pippel nahm seine Tätigkeit als PostDoc im Juli auf. Des weiteren hält Gene Myers, einer der Pioniere im Bereich der Genomassemblierung und Direktor am Max Planck Institut für Zellbiologie und Genetik in Dresden, die Rolle des Mentors der Gruppe inne.

Die CBI Gruppe arbeitet an der Schnittstelle von Information, Mathematik und Biologie, mit Fokus auf die informatischen und algorithmischen Grundlagen der Genombiologie. Von der Vielzahl an Problemen in diesem Feld, sind wir besonders an der Assemblierung von Genomsequenzen interessiert. Darunter ist die Rekonstruktion der Sequenz (Folge der Nukleotide) eines Genoms, basierend auf Daten die durch einen DNA-Sequenzierer produziert wurden, zu verstehen.

Das Prinzip hinter Assemblierung ist, aus dem Genom zufällig (überlappende) Fragmente auszulesen, diese zu sequenzieren und anschließend aus der Sequenz dieser Fragmente die komplette Genomsequenz mit computergestützten Verfahren zu rekonstruieren.

Die Komplexität dieses Ansatzes wird primär von der Länge der Fragmente und der Fehlerrate des DNA-Sequenzierers bestimmt. Die aktuelle Generation an Sequenzierern, welche sehr lange Fragment aber mit einer hohen Fehlerrate produzieren, erfordert neue algorithmische Ansätze, um Genome effizient unter solchen Bedingungen rekonstruieren zu können. Die Entwicklung solcher Verfahren und deren Anwendung in Genomsequenzierungsprojekten stellen die Hauptaufgaben der Gruppe dar.

A DE NOVO WHOLE-GENOME SHOTGUN ASSEMBLER FOR NOISY LONG-READ DATA

On the face of it, 10Kbp reads from single-molecule sequencers are impressive, but an error rate of 15% often makes them difficult to handle. However, truly random error positioning and near-Poisson single-molecule sampling imply that reference-quality reconstructions of gigabase genomes are, in fact, possible with as little as 30X coverage. Such a capability would resurrect the production of true reference genomes and enhance comparative genomics, diversity studies, and our understanding of structural variations within a population.

We have built a prototype assembler we call the Dazzler that can assemble 1-10Gb genomes directly from a shotgun, long-read dataset currently producible only with the PacBio RS II sequencer. It is based on the string graph paradigm, its three most important attributes being:

1) It scales reasonably to gigabase genomes, being roughly 30 times faster than current assemblers for this kind of data.
2) A “scrubbing phase” detects and corrects read artifacts, including an untrimmed adapter, polymerase strand jumps, and ligation chimers that are the primary impediments to long contiguous assemblies.
3) A read correction phase that reduces the error rate to <1% on average.

2014 marked the completion of prototypes for all modules in our assembly pipeline. With a full front-to-end assembler at hand, we shifted our focus towards improving algorithmic aspects and the scalability of the separate modules.

One of the impediments to scaling the overlapper to large genomes lies in their repetitiveness. If every fragment of the genome were unique, then only true overlaps would be calculated. However, this is only partly the case, even for the simplest bacterial genomes. The repetitiveness that is ever-present at varying degrees induces partial matches (i.e., local alignment) of the reads that are responsible in large for the overlapper’s runtime and storage requirements. In order to alleviate this problem, we coupled the overlapper with an incremental repeat finder, which begins tagging regions of a read as repeats as soon as enough alignments have been calculated. This happens parallel to overlapping and results in the continuous exclusion of parts of a read from the overlapping process. This has resulted in tremendous savings in disk space and runtime.

The CBI group in 2014 (f.l.t.r.):
Siegfried Schloissnig, Martin Bongartz, Philipp Kämpfer, Martin Pippel

Group Leader
Dr. Siegfried Schloissnig

Staff Members
Martin Pippel

Scholarship Holders
Philipp Bongartz
Philipp Kämpfer

Visiting scientists
Jason Chin (Sept.-Oct. 2014)

Developments in this work have been rapid. The Dazzler on the PacBio E. coli dataset produces a perfect result in 10 minutes on a laptop. Together with our collaborators, we have sequenced various species and use those datasets to perfect assemblies on the mega-to-gigabase base pair scale. Additional details on our genome sequencing projects can be found in the Genome Projects section below.

A HYBRID GRAPH APPROACH FOR SHORT-READ ASSEMBLY

Though many consider the assembly issue a solved problem, unordered and fragmented genome assemblies with false joins are widespread, significantly hampering any downstream analysis in which they are involved. NGS sequencers produce gigabases very quickly and cheaply, but read-lengths are mostly very short. Short read-lengths and short paired-read insert lengths are the primary reasons that most assembly tools have difficulty producing accurate assemblies with long-range contiguity despite ~100X sequencing coverage. Another issue is that most assembly projects choose insert lengths and mixes that are not optimized for the target genome.

We are currently developing approaches in this sector for systematically optimizing paired-end libraries to the repeat structure and composition of the target genome before sequencing. This entails the creation of heuristics that gauge the coverage and paired-end library insert sizes needed before any sequencing actually takes place. Given close to optimal insert sizes and coverage, the next challenge lies in the reconstruction of the genomic sequence based on the short-read data, a problem for which currently two distinct graph-based approaches are employed. The string graph concept is superior to the deBruijn graph in that the unit of assembly is a read as opposed to a small k-mer, so that the graph and its path structure are simpler.

However, most NGS assemblers rely on the deBruijn graph approach simply because it is more time- and space-efficient. In this project, we take advantage of the best of both approaches. We quickly build a deBruijn graph, noting where each read starts and ends within the graph, and employ novel graph algorithms to efficiently find the transitively invariant read overlaps. These are the edges of the string graph, with traversals to the left and right along paths from each read location. We are currently working toward prototype implementation and refining the algorithmic aspects in order to compute the string graph in linear expected time without computing all the pairwise overlaps, the Achilles heel of the approach in computation terms.

GENOME PROJECTS

HUMAN

Most eukaryotes harbor two copies of each chromosome. This is referred to as the genome being diploid. Both copies are usually very similar in their content, thereby causing problems in the assembly processIn order to simplify sequence assembly, the common strategy in the laboratory is to create so-called inbred strains by consistently mating the animals with close relatives. Over multiple generations, this leads to a homogenization of the content of the chromosome's copies but also inadvertently to significant developmental and behavioral defects.

In collaboration with Human Longevity, Inc., we are working on the diploid assembly of multiple human genomes. This project aims at establishing a new set of human reference genomes, which improve on the existing one in two main aspects. First, the resulting reference sequences will give a true view of the underlying diploid nature of the genome, and multiple reference genomes will be created from samples of individuals of varying ethnicities in order to provide references that capture human intra-species variation.

AXOLOTL

The Axolotl, a member of the order Caudata comprising the salamanders, is the vertebrate closest to humans that is capable of regenerating an entire limb and parts of its nervous system. A mouse can regenerate a fingertip, yet humans, unfortunately, can only regenerate liver tissue. So why and how has the human genome evolved progressively to lose an ability that would seem to confer a huge survival advantage?

In order to perform a comparative genomics study, one needs the entire genomes of the species to be studied. Alas, all studies are currently hampered by the fact that most of the genomes involved are poorly reconstructed or that no reconstruction is available at all. In order to gain detailed insights into the mechanism of regeneration, it is necessary to understand the rearrangement history and regulatory changes, not just the transcriptome and proteome ones. The current limitation is the lack of a complete genome that would permit access to the entire potential set of molecular actors involved in regeneration and to further allow genome-editing technologies, such as Crispr, to be deployed. Moreover, a complete genome would allow a scientist to perform whole genome profiling of epigenetic modifications that do occur in regenerating cells and that are therefore implicated in the mechanism of regeneration.

We are collaborating with the laboratory of Elly Tanaka at the Center for Regenerative Therapies in Dresden on the sequencing and assembly of the Axolotl genome, which poses unique challenges. Due to its haploid size of an estimated 32Gbp, ten times the size of the human genome, almost 1.5 machine-years of sequencing are needed to acquire data worth a 30-fold coverage. Overlapping this initial dataset would take many hundreds of thousands of CPU/hours and result in overlaps requiring something in the order of 1PB of disk space. Prompted by these CPU time- and disk-space requirements, when performing a naïve all-against-all overlapping, we developed a dynamic masking approach that incrementally masks the regions of the reads that induce local-alignments (in other words, that are responsible for a large fraction of the quadratic behaviour), thereby drastically lowering the resource requirements of the overlapper. This significant step forward and other technical enhancements have put us on track for initial assembly in 2015.

Fig. 8: The Ambystoma mexicanum, commonly known as the axolotl, possesses extraordinary regenerative abilities and is capable of reconstituting limbs, retina, liver, and even minor regions of the brain. (Picture © kazakovmaksim / Fotolia)

PLANARIANS

The ability to regenerate lost body parts is widespread in the animal kingdom. Humans, by contrast, are unable to

regenerate even minor extremities. If the “survival of the fittest” principle really holds, regeneration should be the exception rather than the rule and remains a fascinating conundrum in biology. Even amongst planarian flatworms, celebrated for their ability to regenerate complete specimens from random tissue fragments, species exist that have completely lost the ancestral ability to regenerate.

Owing to the lack of physical maps, high AT-content, and high repeat density, not one single high-quality planarian genome is currently available. We are working on a draft assembly of Schmidtea mediterranea (S.med), which has defied previous assembly attempts for many years now.

The assembly is based on DNA fragments (also called reads) that are produced with the current single-molecule real-time (SMRT) sequencing technology. This method of sequencing has two major advantages:

1) The resulting DNA fragments are very long (up to 60k base pairs).
2) The set of reads produced is nearly a Poisson sampling of the underlying genome.

The first point is helpful in resolving the numerous repeat structures within the S.med genome, whereas the second point offers a big advantage compared with amplification-based sequencing technologies that are problematic for genomes with extreme base compositions.

A putative disadvantage of SMRT sequencing is the high error rate of up to 12-15% (mostly insertions and deletions). However, these errors are compensated for by the Poisson sampling of the genome and by the fact that they are randomly distributed across the reads.

The full genome of S.med is estimated to be 800Mb-1Gb. Our primary dataset consists of a 50-fold oversampling of the genome. Based on this initial dataset, a draft genome was assembled using our custom front-to-end assembly pipeline, which encompasses the following steps:

1) overlapping of the noisy DNA fragments
2) various filter steps to remove repeat-induced overlaps and machine-specific artifacts
3) a correction phase to reduce the error to <1% on average
4) overlapping of the corrected DNA fragments
5) additional overlap filtration steps and the production of contigs (long continuous stretches of sequences)
6) a scaffolding (i.e., connecting) of contigs based on the original fragments and overlaps.

With our assembly pipeline, we were able to produce a draft genome of S.med., which improves more than tenfold over existing assemblies in the most important quality metrics. However, there is room for improvement. The repeat density and AT-richness result in huge demands on compute time and storage capacity – more than twice the amount needed for processing a human genome, which is three times larger. We also find that the reads, even though luxuriously long in comparison to previous technologies, are still not long enough to span all repeat structures of S.med. Therefore, we are developing novel approaches for detecting and resolving repetitive elements in genomes, which should lead to improved assemblies of not only S.med., but also many other genomes.

This work is performed in close collaboration with the Max Planck Institute for Molecular Cell Biology and Genetics and the Systems Biology Center in Dresden. These two institutions provide species samples and will subsequently conduct further experiments on the basis of the aforementioned genome sequence in an attempt to understand regeneration at a systems level.

TIGER FLATWORM (MARITIGRELLA CROZIERI)

Maritigrella crozieri is a member of Policladida, a highly diverse clade within the phylum Platyhelminthes. These marine turbellarian flatworms can be found on the eastern coasts of North America and the Caribbean Sea. Their main advantages are ease of collection, spiral cleavage, a biphasic life cycle, and large size (up to 55mm) with many eggs, which can be obtained and raised without eggshells. Accordingly, they represent an interesting system for evolutionary and developmental studies within their phylum.

The diploid genome of M. crozieri is estimated to be highly repetitive and about 2.5 gigabases in length, distributed across three chromosomes. The aim of our work is to generate an initial draft genome of this species using the de novo genome assembly strategies described above.

The assembly is mostly based on multiple short-read, paired-end, and mate-pair libraries originating from the Illumina sequencing platform. First, quality assessment and filtering of the reads was performed. Then, a set of assemblers based on both the deBruijn and the string-graph concept was used to produce longer contiguous sequences called contigs. To partly overcome the aforementioned repetitiveness of the genome, additional long-read sequencing will be performed on the PacBio RS II sequencer.

This work is being done in conjunction with Prof. Max Telford of University College London and the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden.

2.3 Computational Statistics (CST)

The Computational Statistics group at HITS was established in November 2013, when Tilmann Gneiting was appointed group leader and Professor of Computational Statistics at the Karlsruhe Institute of Technology (KIT). The group conducts research in two main areas: (1) the theory and practice of forecasting and (2) spatial and spatio-temporal statistics.

The group's current focus is on the theory and practice of forecasting. As the future is uncertain, forecasts should be probabilistic in nature, i.e., take the form of probability distributions over future quantities or events. Accordingly, we are currently witnessing a trans-disciplinary change of paradigms from deterministic or point forecasts to probabilistic forecasts. The CST group seeks to provide guidance and leadership in this transition by developing both the theoretical foundations of the science of forecasting, as well as cutting-edge statistical methodology, notably in connection with applications.

Weather forecasting is a key example. In this context, the group maintains research contacts and collaborative relations with national and international hydrologic and meteorological organizations, including the German Weather Service, the German Federal Institute of Hydrology, and the European Centre for Medium-Range Weather Forecasts, where group leader Tilmann Gneiting was appointed a fellow.

In 2014, the CST group grew substantially, with staff members Werner Ehm, Kira Feldmann, Stephan Hemri, Alexander Jordan, Fabian Krüger, and Evgeni Ovcharov having joined our team.

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. Sie beschäftigt sich mit zwei wesentlichen Arbeitsgebieten, der Theorie und Praxis der Vorhersage sowie der räumlichen und Raum-Zeit-Statistik.

Der Forschungsschwerpunkt der Gruppe liegt derzeit in der Theorie und Praxis von Prognosen. Im Angesicht unvermeidbarer Unsicherheiten sollten Vorhersagen probabilistisch sein, d.h., Prognosen sollten die Form von Wahrscheinlichkeitsverteilungen über zukünftige Ereignisse und Größen annehmen. Dementsprechend erleben wir einen transdisziplinären Paradigmenwechsel von deterministischen oder Punktvorhersagen hin zu probabilistischen Vorhersagen. Der CST Gruppe ist es ein Anliegen, diese Entwicklungen nachhaltig zu unterstützen, indem sie theoretische Grundlagen für wissenschaftlich fundierte Vorhersagen entwickelt, eine Vorreiterrolle in der Entwicklung entsprechender statistischer Methoden einnimmt und diese in wichtigen Anwendungsproblemen, wie etwa in der Wettervorhersage, zum Einsatz bringt.

In diesem Zusammenhang bestehen Kooperationen mit nationalen und internationalen hydrologischen und meteorologischen Organisationen, wie etwa dem Deutschen Wetterdienst, der Bundesanstalt für Gewässerkunde und dem Europäischen Zentrum für mittelfristige Wetterprognosen, wo Gruppenleiter Tilmann Gneiting zum Fellow ernannt worden ist.

Im vergangenen Jahr 2014 haben Werner Ehm, Kira Feldmann, Stephan Hemri, Alexander Jordan, Fabian Krüger und Evgeni Ovcharov unsere Gruppe verstärkt.

THE SCIENCE OF FORECASTING

A major human desire is to make forecasts for the future. As the future is inherently uncertain, forecasts should strive to be probabilistic in nature, taking the form of probability distributions over future quantities or events. Accordingly, we have been witnessing a transdisciplinary transition from deterministic or point forecasts to probabilistic forecasts. For example, the Bank of England has issued probabilistic forecasts of inflation rates and gross domestic product for nearly two decades, using fan charts to visualize the predictive distributions (http://www.bankofengland.co.uk/publications/Pages/inflationreport/), and central banks worldwide have followed its lead. Figure 9 shows the November 2014 projection of inflation in the United Kingdom, as measured by the consumer price index (CPI). The most central band depicts a pointwise 30% prediction interval. The pairs of the lighter red areas extend by 30% each, so the entire fan corresponds to a 90% interval, which is expected to cover the actual inflation rate 9 out of 10 times on average.

The CST group's work in the area of forecasting has been supported by an Advanced Grant from the European Research Council (ERC) on „The Science of Forecasting: Probabilistic Foundations, Statistical Methodology, and Applications" (ScienceFore). In addition to group leader Tilmann Gneiting, CST staff members and students Werner Ehm, Kira Feldmann, Alexander Jordan, Fabian Krüger, Evgeni Ovcharov, and Patrick Schmidt have been working on the ScienceFore project. Our overarching goal is to provide guidance and leadership in the transition to probabilistic forecasting by developing the theoretical foundations of the science of forecasting, as well as cutting-edge statistical methodology, along with applications in meteorology and economics. Here we give a broad, non-technical overview. For a more quantitatively oriented survey of recent progress in this area, we refer to the review article by Gneiting and Katzfuss (2014).

The CST group in 2014 (f.l.t.r.): Stephan Hemri, Tilmann Gneiting, Kira Feldmann, Roman Schefzik, Fabian Krüger, Alexander Jordan

Group Leader
Prof. Dr. Tilmann Gneiting

Staff Members
Dr. Werner Ehm (from Aug. 2014)
Kira Feldmann (from Aug. 2014)
Stephan Hemri (from Jan. 2014)
Alexander Jordan (from July 2014)
Dr. Fabian Krüger (from Jan. 2014)
Dr. Evgeni Ovcharov (from Sept. 2014)

Visiting Scientists
Prof. Dr. Sandor Baran (July 2014)

Students
Patrick Schmidt

WHAT IS A GOOD PROBABILISTIC FORECAST?

It is now generally recognized that the goal of probabilistic forecasting is to maximize the sharpness of the predictive distributions subject to calibration. Calibration refers to the statistical consistency between the probabilistic forecasts and the observations. This is a joint property of the forecasts and the events that materialize. Intuitively, a forecast is calibrated if the realizing observations are statistically indistinguishable from random draws from the

respective predictive distributions – a notion that can be made mathematically rigorous by using tools of measure theory and probability. Sharpness refers to the concentration of the predictive distributions and is a property of the forecasts only. In the context of prediction intervals, this goal can be rephrased simply as follows: The shorter the intervals, the sharper they are; the sharper they are, the better – subject to the nominal coverage being attained in practice.

PROPER SCORING RULES

The issues of the generation and the evaluation of probabilistic forecasts are intimately related in that improvements in forecast methodologies depend on our ability to adequately assess their quality. Scoring rules are omnibus performance measures for probabilistic forecasts that address calibration and sharpness simultaneously. In a nutshell, a scoring rule assigns a numerical reward based on the predictive distribution and on the event or value that materializes.

Under a proper scoring rule, a forecaster maximizes the expected reward by issuing a forecast that agrees with his or her best judgment. Therefore, proper scoring rules have also been referred to as providing a "truth serum". To fix the idea, suppose that Alice asks Bob to supply critical forecasts for her company. Bob's reward depends both on his forecasts and on the respective realizing quantities. If the reward corresponds to a proper scoring rule, it is designed such that Bob's best strategy is to provide the most careful and honest forecasts he can generate.

While the class of the proper scoring rules can be characterized using tools of convex analysis, important questions as to their structure remain open, some of which we have been addressing. For example, if the predictive distribution admits a probability density, it can be argued that the score ought to depend on the predictive density only via its behavior in an infinitesimal neighborhood of the realizing observation. Any such scoring rule is said to be local, with the ubiquitous logarithmic scoring rule being the most prominent example. As it turns out, local proper scoring rules other than the logarithmic score have a very appealing property in that the score can be computed without knowledge of the normalization constant of the underlying predictive density.

CONSISTENT SCORING FUNCTIONS

A proper scoring rule that depends on the predictive probability distribution only via a certain facet or feature, such as its mean or expectation, or a certain quantile, is said to be a consistent scoring function (for the given feature). We have been investigating the mathematical structure of the classes of the consistent scoring functions in important special cases and have been studying the practical implications.

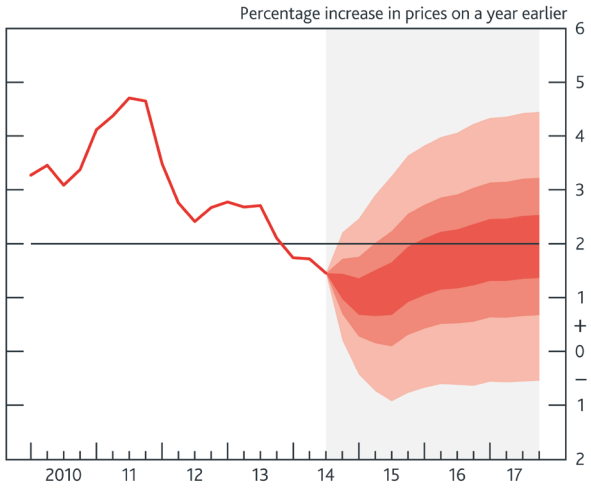


Fig. 9: Bank of England projection of United Kingdom consumer price index (CPI) inflation in percent, as issued in November 2014. The shaded bands frame prediction intervals in increments of 30%. Reproduced with kind permission from the Bank of England's November 2014 Inflation Report.

Consistent soring functions play critically important roles in the evaluation of predictive performance across all realms of science and society. For instance, the Basel protocol prescribes the ways in which regulatory bodies request financial risk assessments from banks. A natural question, then, is whether the protocol can be designed such that a bank's best strategy is to provide careful and honest assessments of its financial risks to regulators. Currently, there is a good deal of debate about a potential revision of this protocol in which banks would need to report a certain feature of their in-house generated predictive distributions, called the "expected shortfall" or "conditional value-at-risk". However, this feature is not elicitable in the technical sense that it does not allow for a respective consistent scoring function. While this is a serious drawback from a statistical theory perspective, there is currently an active debate about practical consequences and remedies. One suggestion is that regulation be based on elicitable features of predictive distributions and that banks' in-house assessments be required to perform no worse than simple reference techniques.

On June 18, 2014, the CST group hosted an interdisciplinary workshop on the topic of "Propriety and Elicitability" on the HITS premises. Together with mathematicians, statisticians, computer scientists, and economists from academia and business, including guests from Australia, Denmark, the United Kingdom, the United States, and Switzerland, we discussed the state of the art of the theory and practice in this very active strand of research. Evgeni Ovcharov, from our group, contributed a talk on the analytic and geometric structure of proper scoring rules on finite sample spaces, and Alexander Jordan and Patrick Schmidt presented posters.

STATISTICAL POST-PROCESSING OF ENSEMBLE WEATHER FORECASTS

Weather forecasting has traditionally been viewed as a deterministic exercise, drawing on highly sophisticated numerical models of the physics and the chemistry of the atmosphere. The advent of ensemble prediction systems in the 1990s marks a change of paradigms. An ensemble forecast comprises multiple runs of numerical weather prediction models that differ in their initial conditions, the lateral boundary conditions, and/or the parameterized representation of the physical and chemical processes in the atmosphere that are being used.

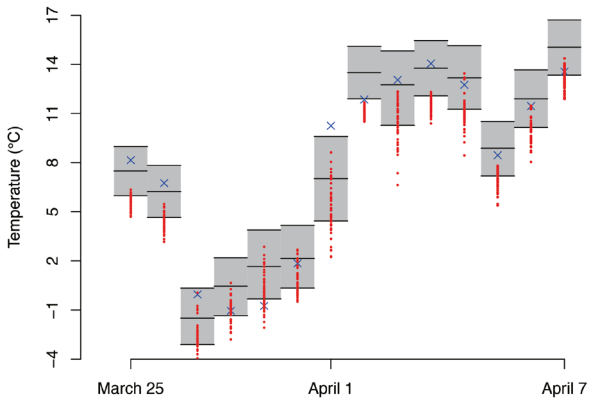


Fig. 10: 48-hour ahead predictive distributions for nighttime temperature in Berlin based on the 50-member European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble, valid March 25 – April 7, 2011. The 50 members of the forecast ensemble are shown as red dots, and the realizing observations as blue crosses. The boxes show the 10th, 50th, and 90th percentiles of the respective statistically post-processed predictive distributions by using Bayesian model averaging (BMA).

Realizing the full potential of an ensemble forecast requires statistical post-processing of the numerical model output in order to address systematic inadequacies, such as model biases and dispersion errors. Techniques for statistical post-processing include ensemble model output statistics (EMOS), a method that is based on heterogeneous regression, and Bayesian model averaging (BMA), which uses mixture models in which each mixture component is associated with an individual ensemble member. To give an example, Figure 10 displays post-processed predictive distributions of nighttime temperature in Berlin based on the 50-member European Centre for Medium-Range Weather Forecasting (ECMWF) ensemble and BMA. The underlying mixture components are Gaussian, and the statistical post-processing corrects for both a low bias and an occasional lack of dispersion. The BMA model is fitted on a rolling training set consisting of ensemble forecasts and the respective observations for the most recent 30 days that are available when the forecast is being issued. In schemes of this type, the training set is updated continually, thereby allowing the statistical model to adapt to changes in seasons and weather regimes. Tilmann Gneiting was appointed one of three inaugural Fellows in July 2014 at the ECMWF in Reading, United Kingdom, which is considered the world leader in global medium-range numerical weather prediction.

While BMA and EMOS are state-of-the-art post-processing techniques, they treat distinct weather variables at distinct geographic locations and distinct look-ahead times independently of each other. However, in key applications, such as air traffic control, flood management, or winter road maintenance, it is critically important that the post-processed forecast fields show physically realistic and coherent joint dependence structures across meteorological variables, geographic space, and look-ahead times.

The ScienceFore project addresses this challenge for the THORPEX (http://www.wmo.int/thorpex) Interactive Grand Global Ensemble (TIGGE; http://tigge.ecmwf.int/), which merges a range of global-scale ensembles into a single multi-model ensemble. Specifically, TIGGE collects and unifies ensemble forecasts from the leading global numerical weather prediction centers, including the ECMWF, and the national weather services of Australia, Brazil, Canada, China, France, Japan, Korea, the United Kingdom, and the United States. Figure 11 illustrates 2-day ahead forecasts of 48-hour accumulated precipitation from four members of the TIGGE Ensemble. While the

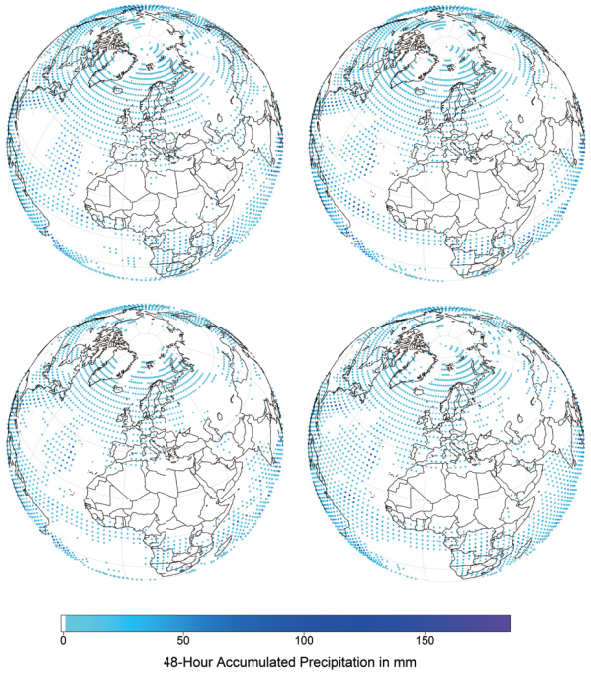


Fig. 11: 2-day ahead forecasts of 48-hour accumulated precipitation from four members of the TIGGE Multi-Model Ensemble, valid December 1, 2008. The top left, top right, bottom left, and bottom right members were generated by the national weather services of Australia, China, the United States, and Canada, respectively.

member forecasts resemble each other, there are distinctions in detail, serving to quantify the predictive uncertainty. For example, the top left member predicts rain over the Libyan desert, whereas the other three members don't.

Issues and challenges in post-processing that we are facing here include an unprecedented number of ensemble members, the multi-model character and complex structure of the ensemble, a considerable extent of missing data, the global geographic scope, and, last but not least, the aforementioned need for physically realistic dependence structures in the post-processed fields. Kira Feldmann presented initial work toward these goals at the World Weather Open Science Conference 2014 in Montreal, Canada, and at a Mini Symposium on "Spatial Statistics", which we hosted in October 2014, when colleagues from Saudi Arabia and the United States visited HITS. At the International Symposium "Extremes 2014" in Hannover, Germany, her work received a poster award.

ECONOMIC FORECASTS: FROM SURVEYS TO PREDICTIVE DISTRIBUTIONS

The availability of reliable and accurate forecasts of future economic variables, such as inflation, employment, and gross domestic product, is crucial for policymakers, investors, and management and labor unions, among others. To give an example, Figure 12 shows current quarter forecasts of the consumer price index (CPI) from the Survey of Professional Forecasters (SPF; http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/) in the United States.

There is a striking analogy between ensemble forecasts in meteorology and climatology on the one hand and surveys of experts in economics on the other hand in that both provide a collection of complementary point forecasts. Perhaps surprisingly, methods for ensemble post-processing, which have led to major improvements in weather forecasts, remain to be applied to economic surveys, and we are investigating this issue. However, there are major differences that need to be addressed. For example, the composition and size of the SPF panel varies from quarter to quarter, and training data are much scarcer.

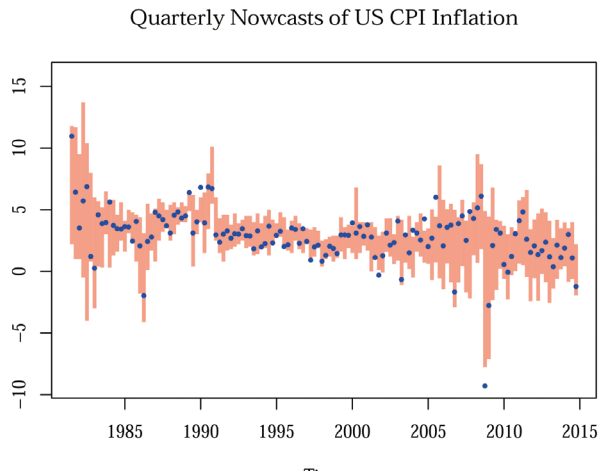


Fig. 12: Current quarter forecasts of annualized quarterly consumer price index (CPI) inflation in the United States from the Survey of Professional Forecasters (SPF), in percent. The red boxes show the range of the individual SPF experts' point forecasts, and the blue dots represent the realizing inflation rates. Note the extreme deflation in the fourth quarter of 2008, and the recent increase in the spread of the experts' forecasts as compared to the 1990s and early 2000s.

2.4 Data Mining and Uncertainty Quantification (DMQ)

Advances in sensor technology and high-performance computing enable scientists to collect and generate extremely large datasets, usually measured in terabytes and petabytes. These datasets, obtained by means of observation, experiment, or numerical simulation, are not only very large, but also highly complex in their structure. Exploring these datasets and discovering patterns and significant structures in them is a critical and highly challenging task that can only be addressed in an interdisciplinary framework combining mathematical modeling, numerical simulation, optimization, statistics, high-performance computing, and scientific visualization.

In addition to the size and complexity of these data, quality is another crucial issue in guaranteeing reliable insights into the physical processes under consideration. The associated demands on the quality and reliability of experiments and numerical simulations necessitate the development of models and methods from mathematics and computer science that are able to quantify uncertainties for large amounts of data. Such uncertainties may derive, for example, from measurement errors, lack of knowledge about model parameters, or inaccuracy in data processing.
The Data Mining and Uncertainty Quantification group is headed by Prof. Dr. Vincent Heuveline. In this group, we make use of stochastic mathematical models, high-performance computing, and hardware-aware computing to quantify the impact of uncertainties in large datasets and/or associated mathematical models and thus help to establish reliable insights in data mining. Currently, the fields of application are medical engineering, fluid dynamics, energy, astronomy, and meteorology.

Fortschritte in Sensortechnik und High Performance Computing ermöglichen Wissenschaftlern das Erfassen und Erzeugen extrem großer Datenmengen, die vorwiegend in Tera- und Petabyte gemessen werden. Sie werden durch Beobachtung, Experimente und numerische Simulation generiert und sind nicht nur umfangreich, sondern weisen auch eine hochkomplexe Struktur auf. Die Erforschung dieser Datenmengen und das Entdecken von Mustern und signifikanten Strukturen ist eine entscheidende und in hohem Maße herausfordernde Aufgabe, der man nur in einem interdisziplinären Rahmen gerecht werden kann, d.h. durch Verbindung von mathematischer Modellierung, numerischer Simulation, Optimierung, Statistik, High Performance Computing und wissenschaftlicher Visualisierung.

Neben der Größe und Komplexität der Daten spielt deren Qualität eine entscheidende Rolle, um zuverlässige Einblicke in die betrachteten physikalischen Prozesse zu garantieren. Der damit verbundene Anspruch an Qualität und Zuverlässigkeit der Experimente und numerischen Simulationen erfordert die Entwicklung von Modellen und Methoden aus Mathematik und Informatik, die Unsicherheiten für große Datenmengen quantifizieren können. Solche Unsicherheiten können z.B. durch Messfehler oder mangelnde Kenntnisse über Modellparameter entstehen. In ihrer statistischen Aussagekraft spielen diese eine zentrale Rolle z.B. im Rahmen von Predictive Analytics für große Datenmengen.

Die DMQ-Gruppe wird von Prof. Dr. Vincent Heuveline geleitet. Unsere Forschungsgruppe nutzt stochastische mathematische Modelle, High Performance Computing und Hardware-Aware Computing, um die Auswirkungen von Unsicherheiten bei großen Datensätzen und/oder zugehörigen mathematischen Modellen im Hinblick auf zuverlässige Einblicke in Data Mining zu quantifizieren. Derzeit sind Medizintechnik, Strömungsmechanik, Energieforschung, Astronomie und Meteorologie typische Anwendungsfelder.

The DMQ group in 2014 (f.l.t.r.): Philipp Gerstner, Chen Song, Vincent Heuveline, Michael Schick, Maximilian Hoecker

Group Leader
Prof. Dr. Vincent Heuveline

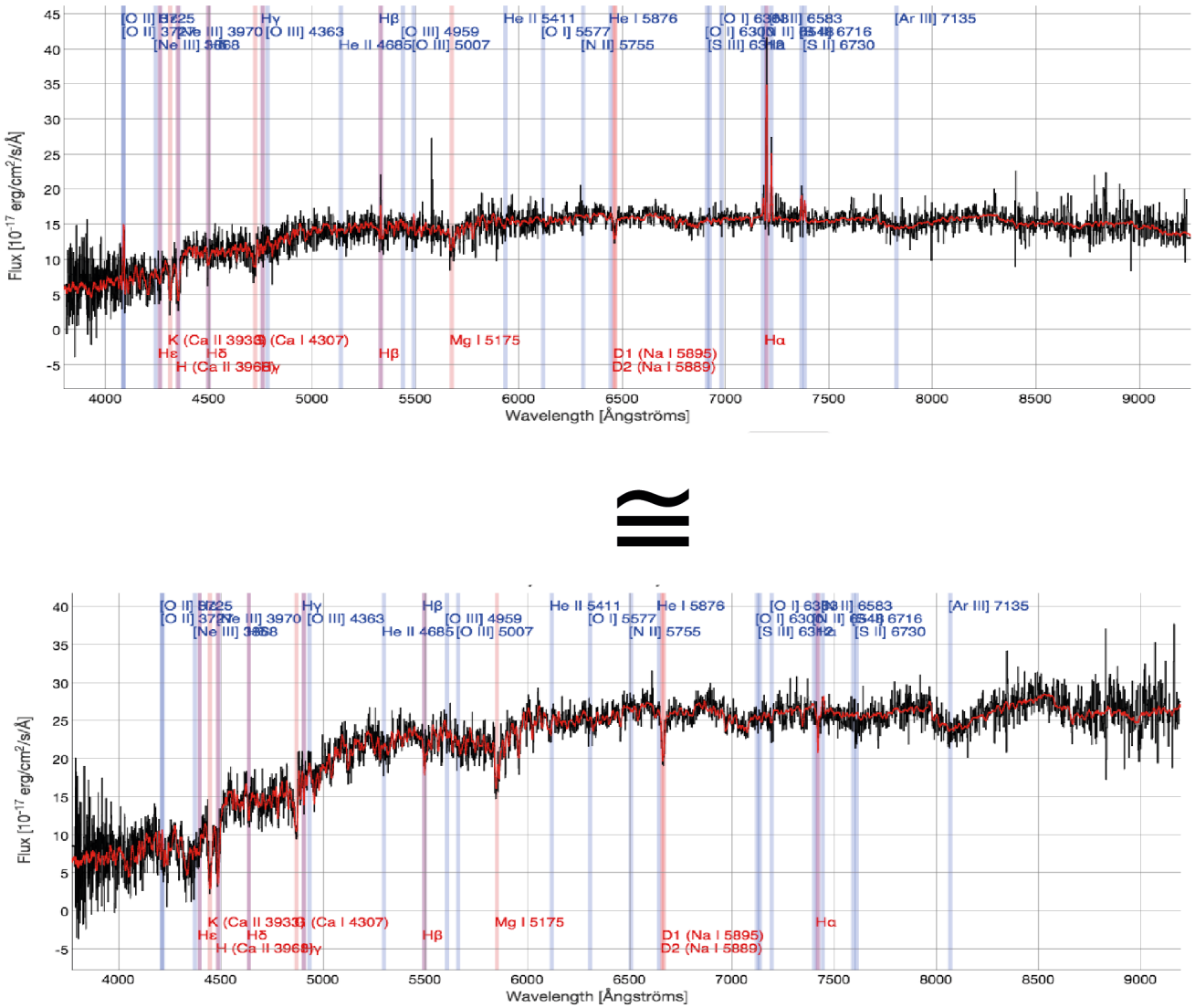
Staff Members
Maximilian Hoecker
Dr. Michael Schick
Philipp Gerstner (since April 2014)

Scholarship Holders
Michael Bromberger (HITS Scholarship)
Chen Song

Visiting Scientist
Dr. Stefanie Speidel (KIT) (since June 2014)

LARGE HIGH-DIMENSIONAL CLUSTERING

Today's scientific and industrial datasets often cover all aspects of the well-known big data characteristics (value, velocity, variety, veracity, volume). Data analysis methods with big data compatibility are the key to the problem statements of many fields, especially in astronomy. Dedicated features are extracted based on pre-processed data. In most cases, a model-driven approach is chosen to generate these features. Both the extracted features as well as the uncertainties of the model-fitting are stored in relational databases with the original data aside. Therefore, scientists have to define selection criteria explicitly in order to retrieve the objects of interest. Instead of working on the original data, the analysis is limited to the pre-extracted features only. This requires having according features in the database and an a-priori knowledge of the nature of the requested objects. Rare and odd objects are hard to be detected or filtered for follow-up analysis. To allow for a more explorative access to the scientific data, unsupervised methods like clustering and outlier-detection are helpful. Clustering in scientific environments is a challenging task as a result of the complexity and size of the data. Current datasets can no longer be analyzed efficiently with the existing scheme. New upcoming projects will increase exponentially in size and complexity, e.g., the Square Kilometre Array (SKA) archive will be limited to 1 Exabyte as a result of by the costs projected in 2011. The aim of this project is to provide a powerful method for analyzing large datasets based on similarities of items in high dimensions. In this research project, the science case is analyzing unlabeled datasets from the Sloan Digital Sky Survey 3, Data Release 10 (SDSS3 DR10), with a focus on the similarity/dissimilarity relationship of two objects. Each object is represented by a feature vector of approx. 5,000 dimensions (as depicted in Fig. 13, page 34). These vectors display a numeric value of a captured spectrum with uncorrelated noise for all specific wavelengths. The whole dataset consists of 3 million objects with 60 GBytes of raw data in total. As all objects have to be compared with each other, the resulting complexity is $O(n^2)$ in computation and storage. Reflecting the science-case mentioned, a naïve full analysis with distance-density clustering algorithms would end in 542 days of processing time using a single similarity measure on a 128-cores computing-cluster. This assumes a time of 2 ms per comparison, including loading and saving data. The 9 × 1012 comparisons would effectively produce between 24 TByte and 120 PByte of resulting data depending on the level of detail. This project aims at developing a method analyzing this data in an acceptable timeframe.



≅

Fig. 13: The comparison of two Optical Spectra is a non-trivial task. Each feature has to be detected and compared creating a single value for similarity. This pictures shows two random, exemplary selected spectra that could be compared. (Images-Source: dr10.sds)

HIGH-PERFORMANCE COMPUTING AND UNCERTAINTY QUANTIFICATION

The increasing demand on the quality and reliability of numerical simulations of physical problems results in an increasing complexity of mathematical models. In particular, the knowledge for the description and definition of model-relevant parameters often cannot be assumed to be available in a deterministic way. There are often uncertainties involved, which can arise, for example, by inexact measurements or modelling assumptions. This makes the development of appropriate and efficient numerical solution methods a crucial task.

The focus of this project is the development of parallel numerical methods for uncertainty propagation in partial differential equations (PDEs) by using Polynomial Chaos expansions and stochastic Galerkin projections. Orthogonal Polynomials are used to express the dependence of all stochastic quantities on the randomness given in the parameter of the overall physical model. The spatial components of the considered PDEs are discretized by the Finite-Element-Method, thereby giving rise to the so-called Spectral-Stochastic-Finite-Element-Method (SSFEM). PDEs ranging from the linear elliptic type to the incompressible Navier-Stokes equations in steady and unsteady formulation serve as applications based on varying probability models for the corresponding uncertain parameters. The most attention is devoted to the development of parallel numerical methods for the efficient solution of the associated discretized systems. Parallelization is carried out in the stochastic, the spatial, and the temporal domain by using distributed and shared memory approaches.

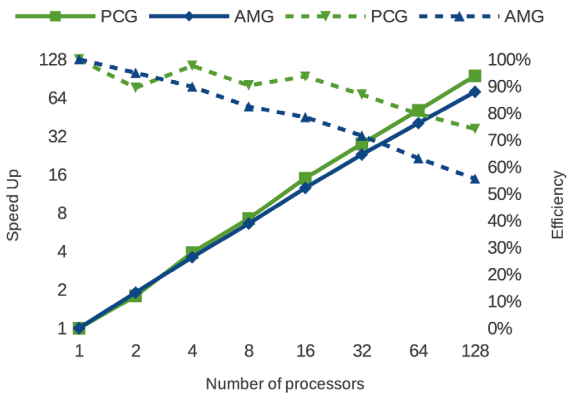


Fig. 14: Speed Up and parallel efficiency plot for a multilevel (AMG) and preconditioned CG method (PCG) for stochastic Galerkin projected systems.

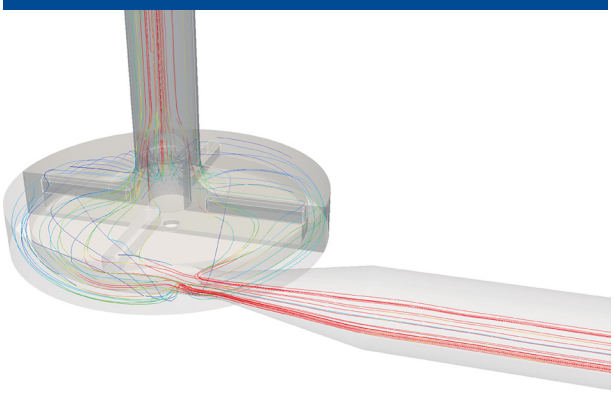
The first promising results are based on a multilevel discretization method inspired by deterministic multigrid algorithms. For these algorithms, the stochastic Problem is reformulated as a hierarchy of differently resolved problem formulations with respect to the polynomial degree of the Polynomial Chaos expansion. This allows for the use of block-smoothing algorithms, which are required for smoothing out high-frequency error fluctuations. Their block structure corresponds to uncoupled stochastic “sub problems”, which paves the way for a parallelism (compare Fig. 14).

Current work is focused on the extension of parallelism to many-core computing platforms and its hybrid use with domain decomposition approaches for memory distribution.

UNCERTAINTY QUANTIFICATION IN A BLOOD PUMP SIMULATION FOR HUMAN HEART SURGERY

The ventricular assist device (VAD) – also known as a blood pump – has been one of the most important cardiac therapeutic instruments in the last two decades since

Fig. 15: Blood pump geometry with streamline visualization of blood flow. Red colors correspond to high velocities, blue colors to low velocities.



more than 23 million people are currently suffering from heart failure (HF) worldwide. Meanwhile, the centrifugal pump is the most common concept for transcutaneous and implantable VADs, which support patients by augmenting or replacing the function of a failing heart.

Nowadays, computational fluid dynamics (CFD) plays an important role in modeling blood flow in complex geometries. For such medical devices in particular, a good deal of crucial information could be obtained via deterministic numerical simulations, such as destructive shear stresses, stagnation points, vortices, etc., in order to quantify blood damage level.

Although the deterministic numerical models have been largely and successfully introduced, there are still many unpredictable factors that we could not have the ability to anticipate, and this uncertainty could lead to unreliable situations that cannot be solved via classical simulations. Therefore, uncertainty quantification (UQ) is crucial in numerical simulations.

In this project, we develop appropriate numerical solution methods to quantify the effect of uncertainty propagation in the technical device modeling of the blood pump (compare Fig. 15). To this end, the numerical methods based on Polynomial Chaos expansions (as in the High-Performance Computing project) are used to tackle the extremely large number of degrees of freedom (many billion) associated with this simulation. However, a pure brute force parallelization approach can only resolve a limited amount of resolution due to the strong inherit couplings in the model. Therefore, we develop appropriate model reduction techniques, which can then be combined with high-performance computing.

OPTIMAL ECONOMIC POWER FLOW (DFG PROJECT WITH KIT)

Against the background of the liberalization of energy markets, increasing fuel costs, and decentralized power generation by renewable energy sources, running an electrical power grid in an efficient way nowadays is becoming more and more important. The problem of determining an optimal operation state and an optimal power grid extension is known as “Optimal Power Flow (OPF)”. Mathematically speaking, OPF is a non-linear and non-convex optimization problem in up to millions of variables. Due to its high complexity, power system operators still have to use simplified physical models to describe electrical grids.

The project Optimal Economic Power Flow (OEPF) is in cooperation with the Karlsruhe Institute of Technology and supported by German Research Foundation (DFG). It aims at developing new numerical approaches to handle the high problem size arising from an accurate alternating current (AC) physical model. In contrast to linear direct current (DC) simplifications of the electrical networks, AC models allow the consideration of reactive powers, which are often neglected due to the associated nonlinearity. In contrast to standard OPF, the economic part in OEPF is important. This part couples simulations in electrical grids with power plant predictions for power generations and

demands; however, it also significantly exacerbates the complexity of the problem.

Our point of interest lies in an efficient solution to linear systems arising from optimization methods applied to OEPF. Though much larger in dimension, these linear systems have the same sparsity structure as the underlying physical network. By exploiting this fact, it is possible to solve the full linear system by decomposing the electrical grid into smaller sub grids, solving the corresponding subsystems separately, and recomposing the overall solution (compare Fig. 16). The resulting algorithm can be run in parallel on multicore systems and allows for considering the optimization of large power grids, e.g., the complete transmission grid in Germany.

HARDWARE-AWARE COMPUTING FOR UNCERTAINTY QUANTIFICATION

In recent decades, speeding up the computation of scientific applications has mostly been achieved by increasing the frequency of single-core processors. Due to technical limitations, it is not possible to further increase this frequency nowadays. Therefore, one possibility is to increase the amount of homogenous cores on the same chip. Another option is to integrate different or heterogeneous compute units into the same system. Such systems are called hybrid or heterogeneous systems and integrate, e.g., general purpose graphic processing units (GPUs) or field-programmable gate arrays (FPGAs). The latter offers the possibility of having special hardware accelerators for a given problem.

The common approach for using accelerators in a heterogeneous system starts with a profiling of the application. Then, the most time-consuming part of the algorithm is ported, e.g., to an FPGA. We applied this for HHblits (see Fig. 19, page 39), which is a tool for iteratively finding homologous protein sequences in a large database. It still uses a two-level prefilter to accelerate the search for such sequences. Porting the first step of this prefilter to four FPGAs enables a significant speed up against the execution of the prefilter on the vector extension unit of an x86 processor.

For many applications that are widely used nowadays, such as machine learning, image processing, and data

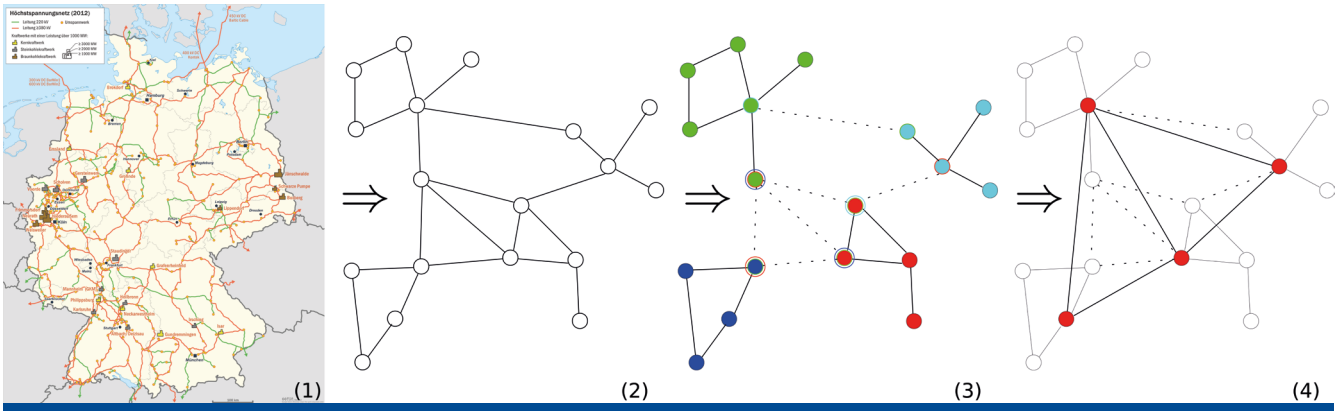


Fig. 16: Schematic diagram of developed linear solver: Physical grid (1) is represented as undirected graph (2), graph after partitioning (3), coarse grid correction (4).

mining, have a tolerance for inexact computation giving rise to “approximate computing”. We calculated the depth map from a stereo camera system and achieved massive speed up by only a slight increase of the calculated error. For a broader applicability for applications (e.g., spectral method; see Fig. 17), the amount of data that is transferred between the compute units and the main memory can be reduced by a dynamic data conversion unit.

In the future, we plan to consider numerical simulations that integrate uncertainty quantification. We want to identify different kernels that can be integrated into several applications and also be easily used by domain experts like mathematicians or engineers who are not familiar with GPUs or FPGAs. This is even more important since a new trend in the field of computer architecture is to integrate the special hardware on the same chip as the host processor. This allows the direct memory access to the host memory or even access to the cache hierarchy of the host processor.

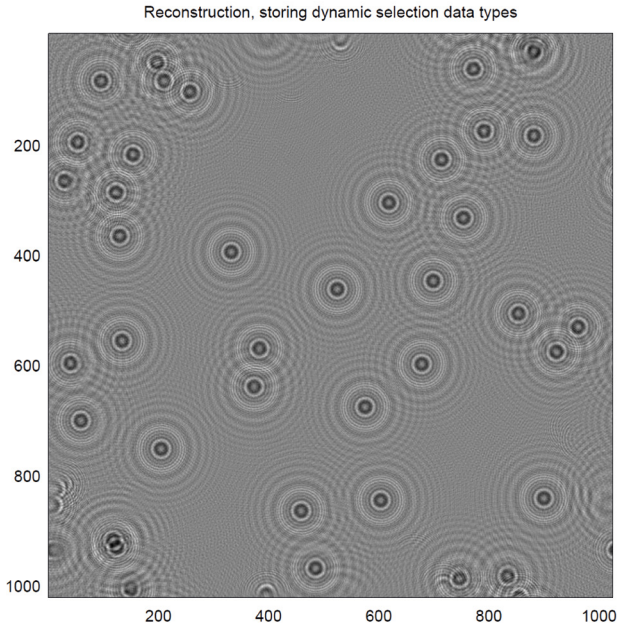


Fig. 17: Spectral method for holography reconstruction using the conversion unit.

SOFTWARE DEVELOPMENT FOR UNCERTAINTY QUANTIFICATION

HiFlow3 (www.hiflow3.org) is a multi-purpose, finite-element software providing powerful tools for the efficient and accurate solution of a wide range of problems modeled by partial differential equations (PDEs). Based on object-oriented concepts and the full capabilities of C++, the HiFlow³ project follows a modular and generic approach for building efficient parallel numerical solvers. It provides highly capable modules dealing with the mesh setup, finite element spaces, degrees of freedom, linear algebra routines, numerical solvers, and output data for visualization. Parallelism – as the basis for high performance simulations on modern computing systems – is introduced at two levels: coarse-grained parallelism by means of distributed grids and distributed data structures, and fine-grained parallelism by means of platform-optimized linear algebra back-ends.

This software is extended toward an Uncertainty Quantification module developed at HITS, thereby providing an implementation and user-friendly interface for the algorithms developed in many of our projects. The module will be accessible as open-source for the academic research community. As of now, it includes a generic framework for stochastic Galerkin projections and Polynomial Chaos expansions for PDEs with uncertain parameters. Different probability measures can thereby be used to model the uncertainties in the parameters. In addition, we integrated a parallelization possibility based on the “Message-Passing-Interface (MPI)” to distributed memory computation, which fills a significant lack of parallel UQ software in the research community.

Our current work is to extend the amount of tutorials for this module with the associated documentation. Furthermore, the amount of numerical solution methods will be steadily increased.

Fig. 18: Logo of the HiFlow3 project. Visit www.hiflow3.org for more information and downloads.

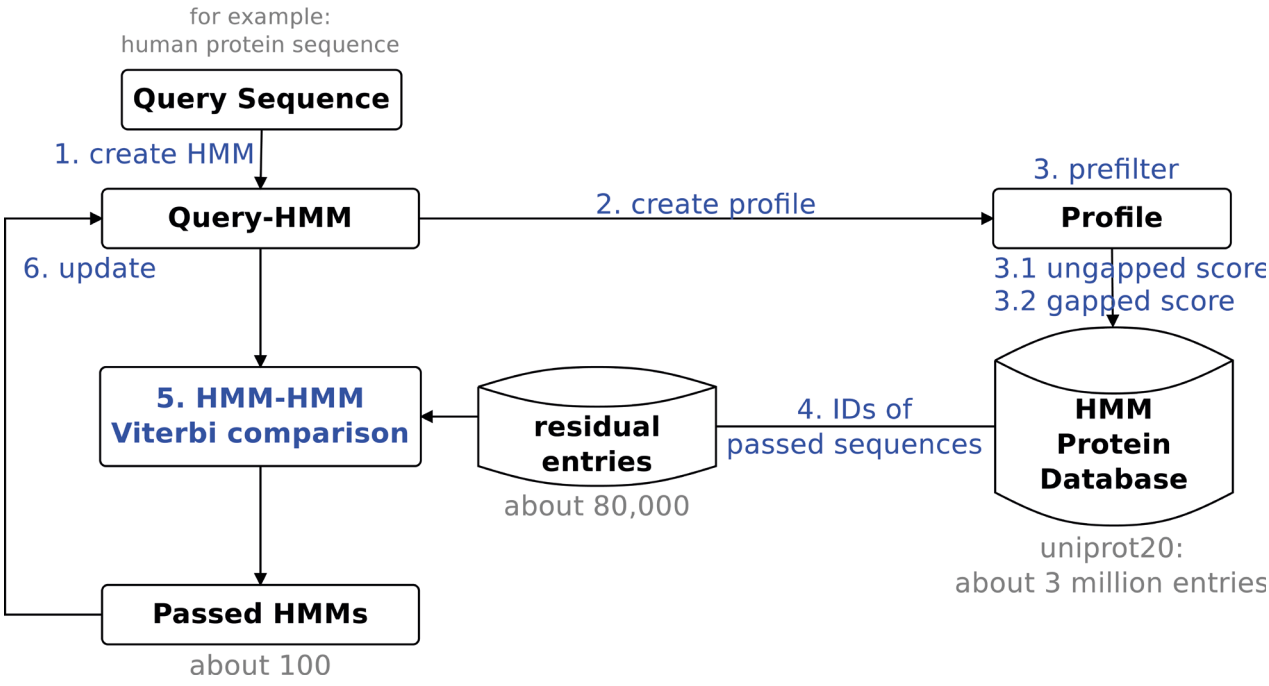


Fig. 19: HMM-HMM-based, lightning-fast iterative sequence search (HHblits).

2.5 Molecular Biomechanics (MBM)

Mechanical forces closely regulate cellular functions such as growth, motility, and signaling. Proteins play a pivotal role in such mechanically guided processes, acting as robust elements bearing cellular stress or as mechano-sensors transducing the mechanical signal into a biochemical response. One fundamental issue here is how a mechanical-function protein has been designed by evolution to withstand and transmit high levels of stress. We aim at understanding the mechanical functions of proteins by using computer simulations at multiple scales combined with bioinformatics methods. Most importantly, we developed a method to compute internal molecular forces, termed Force Distribution Analysis, to track the propagation of external mechanical stress through molecular structures. We apply these methods to various questions in mechanical biology, ranging from mechano-sensing proteins to biomaterials. Examples include the von Willebrand factor, a protein crucially involved in primary hemostasis and regulated by the shear forces of flowing blood, and spider silk fibers, nano-structured protein materials with outstanding mechanical toughness. Our aim is to identify major determinants of stress resistance and sensitivity in biological structures. These insights can serve as design principles when trying to interfere with or re-engineer the mechanical properties of biological systems.

Mechanische Kraft reguliert zentrale zelluläre Funktionen wie Wachstum, Beweglichkeit oder Signalweiterleitung. Proteine spielen dabei eine Schlüsselrolle, indem sie als robuste Bausteine mechanischen Stress tragen oder als mechanische Sensoren Kraft in ein biochemisches Signal umsetzen. Eine wichtige Frage ist, wie sich eine mechanische Funktion während der Evolution dahingehend entwickelt hat, dass das heutige Proteinrepertoire hohe Kräfte nicht nur aushalten, sondern auch weiterleiten kann.

Unser Ziel ist es, die mechanische Funktion von Proteinen mit Hilfe von Computersimulationen auf verschiedenen Skalen und mit bioinformatischen Methoden zu verstehen. Dafür haben wir unter anderem eine Methode entwickelt, die wir Kraftverteilungsanalyse nennen, um die Weiterleitung von externen Kraftsignalen durch molekulare Strukturen hindurch zu berechnen. Wir wenden diese Methoden auf verschiedene Fragen der Mechanobiologie an, die sich von kraftmessenden Enzymen bis hin zu Biomaterialien erstrecken. Anwendungsbeispiele sind der von Willebrand Faktor, ein für die Hämostase essentielles und von Scherkräften reguliertes Protein im Blut, und Spinnenseide, ein proteinbasiertes Nano-Biomaterial mit außerordentlicher Belastbarkeit. Unsere computergestützten Simulationsmethoden erlauben uns, die Hauptmerkmale biologischer Strukturen zu identifizieren, die eine spezifische Stressresistenz oder –sensitivität herbeiführen. Diese Einsichten können als Richtschnur dienen, wenn man ähnliche Materialien mit spezifischen mechanischen Eigenschaften entwerfen will.

A BOTTOM-UP COMPUTATIONAL APPROACH FOR SILK FIBER MECHANICS

Spider silks provided by the major ampullate (MA) glands are used by the spider to form the web frame and the spider's dragline. MA silk has been the most studied silk, as it has excellent mechanical properties and an unusual combination of high stiffness, toughness, strength, and extensibility, which are rarely observed in synthetic high-performance fibers. Silk fiber mechanics are ultimately defined by the nanoscale structure of the fiber. The repetitive segment of spider dragline silk is dominated by iterations of alanine- and glycine-rich regions. The alanine segments are composed of a polyalanine or polyalanylglycine, which form beta-sheets that stack together and thereby form rigid crystals. The second constituent is glycine-rich sequence motifs form the amorphous phase, which is predominantly disordered. Today, it is generally accepted that the stiff beta-sheet crystals furnish silk fibers with a high stiffness and yield strength, whereas the amorphous glycine-rich matrix provides extensibility. However, how the mechanical properties of these individual constituents and their interplay give rise to the typical, highly non-linear stress-strain behavior of silk fibers is still largely unknown. Thus, correctly assessing plastic and viscous deformations of the crystalline and amorphous phases, respectively, is required to integrate their nanoscale mechanical response into a more realistic, purely bottom-up, and therefore predictive macroscopic fiber model. The mechanical response of the crystalline phase of MA spider silk has been comparably well studied. The crystal component of silk largely behaves like an elasto-plastic material, which undergoes non-reversible rupture in response to applied forces. The second component, the amorphous phase, in contrast, is much less well-characterized. The large extensibility and viscous behavior, as evidenced by the time-dependency of silk mechanics in tensile loading experiments, is likely to originate from the amorphous phase due to the sliding of peptide chains, i.e., internal molecular friction. Here, we focused on the rate-dependent behavior of the amorphous phase of MA silk fibers. We assessed friction forces between the peptide chains of the amorphous phase by using MD simulations. This allowed us to deduce a friction coefficient and

The MBM group in 2014 (f.l.t.r.): Eduardo Cruz-Chu, Camilo Aponte-Santamaria, Maxime Louet, Johannes Wagner, Frauke Gräter, Agnieszka Bronowska, Beifei Zhou, Davide Mercadante, Sandeep Patil, Jing Zhou

Group Leader
Dr. Frauke Gräter

Staff Members
Dr. Camilo Aponte-Santamaria
Sandeep Patil (until Dec. 2014)
Dr. Maxime Louet (until Dec. 2014)
Dr. Eduardo Cruz-Chú
Dr. Katra Kolšek (from July 2014)
Ana Herrera-Rodriguez (from June 2014)

Scholarship Holders
Dr. Ulf Hensen
Dr. Davide Mercadante
Johannes Wagner
Jing Zhou

Visiting Scientists
Dr. Agnieszka Bronowska
Dr. Richard Henchman (until Dec. 2014)
Beifei Zhou (until Dec. 2014)

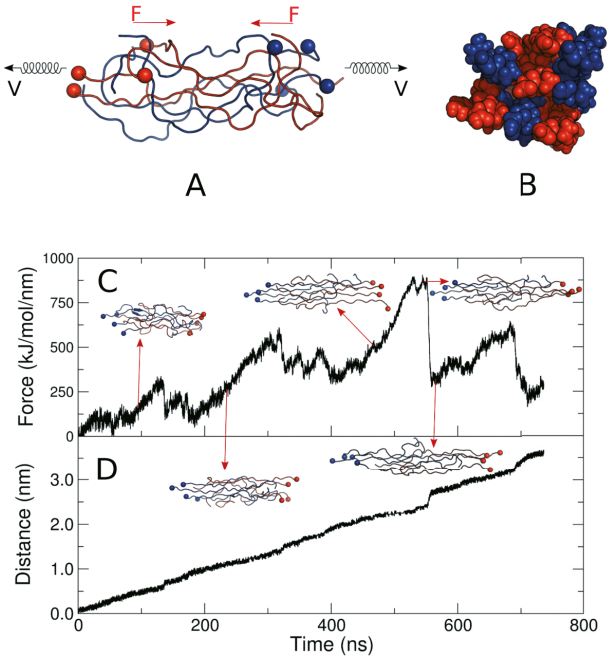


Fig. 20: Silk is a viscoelastic material, and the viscosity can be mainly attributed to its amorphous phase. (A,B) A representative fraction of the amorphous phase was pulled in MD simulation. (C,D) Forces encountered during pulling allowed for determining the internal molecular friction in silk fibers.

coefficient of viscosity at the viscous limit. We employed the coefficient of viscosity in proof-of-principle finite element models of the amorphous phase of silk. Our quantitative analysis of the viscoelasticity of the amorphous phase presents an important step towards developing a bottom-up visco-elastoplastic model for MA silk fibers. We modeled the amorphous phase of spider silk from the MA gland of Araneus diadematus. We constructed bundles of 4, 8, and 24 fully stretched peptide chains, and these bundles of the amorphous chains were subsequently solvated in TIP4P water. These simulation systems were relaxed by energy minimization, and then position-restrained simulations were performed to equilibrate the solvent. Finally, all models were fully equilibrated, allowing the silk peptides to adopt relaxed conformations and to partially entangle within the bundle. The resulting equilibrated simulation systems served as starting points for force-probe Molecular Dynamics (FPMD) simulations. In the FPMD simulations, half of the peptide chains were pulled in one direction and the other half pulled in the opposite direction, as schematically shown in the Fig. 20A,B. The pulling direction for the each peptide was chosen such that the peptide chains were maximally surrounded by peptide chains being pulled in the opposite direction, as shown in Fig. 20B. The springs were moved with constant velocities ranging between 0.01 and 100 nm/ns. Fig. 20C shows a typical force profile and related structures of the 8-chain bundle model. The resulting average displacement of the center of mass of the pulled peptide chains in one direction is shown in Fig. 20D.

There was no external force exerted on the peptides perpendicular to the pulling direction. The FPMD simulations were stopped after the amorphous peptide chains separated from each other. We obtained peak frictional forces of all the velocities and then separated the frictional forces within the silk peptide bundle from frictional forces with water. Data was obtained at different pulling velocities, and averages and standard errors over four independent FPMD simulations are given. Similarly, we calculate the shear stress and coefficient of viscosity for different velocities. The considered velocities for these simulations are in the range of intermediate and large velocities. However, our data does not include the crossover from large and intermediate velocities to the regime of linear friction at low velocities. In the experiments, e.g., in force spectroscopy experiments or when biological molecular motors are active, the applied external force causes molecular motions in the micro-m/s range. Thus, we are experimentally always in the viscous linear response regime, where friction forces are proportional to velocities. To extract the

viscous friction coefficient from the sliding of silk peptide chains in our simulations, we have to extrapolate our data to the viscous regime. To this end, we used a stochastic model, which describes the full velocity dependence of the friction force per residue. On this basis, we then extracted the primary quantity of interest, the coefficient of viscosity for the amorphous phase of spider silk, from our MD simulations. In our extrapolation at low velocities, we obtained a coefficient of viscosity of the amorphous phase of spider dragline silk of 1x10^4 Ns/m^2, which is in the range of polymer melts ~ (10^3 to 10^5 Ns/m^2). We next determined the rate-dependent behavior of the amorphous phase via the Finite Element Method (FEM), using the coefficient of viscosity determined from MD simulations as described above. Viscoelasticity is the property of materials that exhibit both viscous (dashpot-like) and elastic (spring-like) characteristics when undergoing deformation. Using FEM, we found that stress-strain curves are dependent on the rate of straining, i.e., the faster the stretching, the larger the stress required. We observed a significant hysteresis for most of the strain rates. As explained above, the hierarchical structure of spider dragline silk is composed of two major constituents, i.e., the amorphous phase and crystalline units, and its mechanical response has been attributed to these prime constituents. Silk mechanics, however, might also be influenced by the resistance to sliding of these two phases relative to each other under load. Molecular Dynamics (MD) simulations have increasingly helped us gain further insight into the force-bearing structures and interactions within the crystal and the amorphous phase under mechanical load. The crystalline component of silk fibers behaves like an elasto-plastic material, which undergoes non-reversible rupture in response to applied forces. The amorphous phase is softer and features a rate-dependent behavior, i.e., it is a viscoelastic material. However, to our knowledge, the mechanical resistance, or friction, at the crystalline-amorphous protein-protein interface within the fiber has not yet been characterized to date. Depending on the extent of interfacial friction, crystals are able to slide, and thereby redistribute within a silk fiber under mechanical load. To study the relative sliding of the amorphous phase and crystalline units, we built two crystalline units composed of the repeat units found to be present in Araneus diadematus spider silk fibers, AAAAAAAA. Next, we constructed the amorphous phase from a representative 24-residue sequence of MA silk. We constructed our friction model such that two crystalline units were 3 nm apart, and seven bundles of the amorphous, each containing eight fully stretched peptides, were positioned between and around these two units (Fig. 21A). In total, this model consisted of 50 crystalline- and 56 amorphous peptides chains. In the FPMD simulations, seven bundles of the amorphous phase were pulled in the upward direction, as schematically shown in Fig. 21A,B (see page 44). Forces were applied at the center of mass of the terminal residues of the each bundle of the amorphous phase. The goal of this study was to compute friction between the amorphous phase and crystalline units as they slide relative to each other. We applied constant velocities ranging between 0.02 and 20 m/s. The FPMD simulations were stopped after the amorphous phase detached from the crystalline units. To assess the frictional forces between the amorphous and crystalline phase of spider silk, we closely followed the protocol discussed above to assess frictional forces within the amorphous phase. We obtained a coefficient of viscosity or the viscous friction parameter between the amorphous phase and crystalline units of spider dragline silk of 2 X 10^2 Ns/m^2. The obtained coefficient of viscosity is two orders of magnitude lower than that within the amorphous phase, which is 1 X 10^4 Ns/m^2. Interestingly, the bond strength we obtain here is lower than the one we obtained within the amorphous peptide bundles, which reflects the lower hydrogen bond density at the partially hydrophobic crystal surfaces. We next determined the friction between the crystalline and amorphous blocks by FEM, using the coefficient of viscosity determined from MD simulations as described above. Our FEM simulations predict that there is no significant resistance against sliding at low to medium velo-

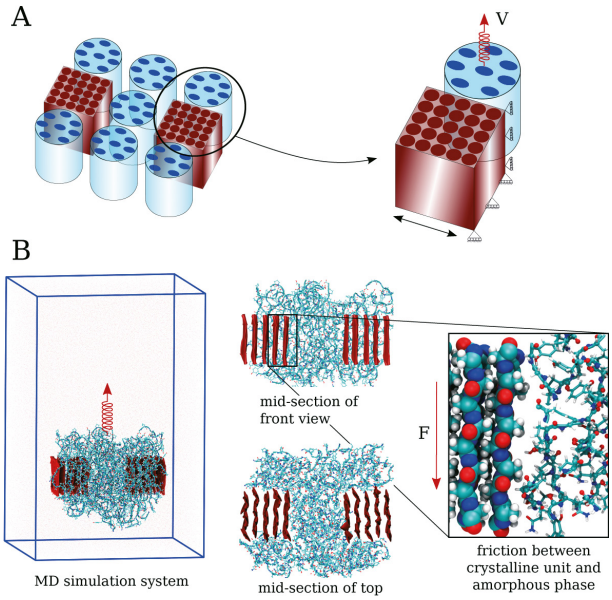


Fig. 21: Friction between crystal and amorphous phase. (A) Scheme of crystals (brown) embedded into amorphous matrix bundles (blue). (B) Lateral sliding of the two phases allowed for determining the inter-phase friction.

cities in the situation of perfect relative horizontal motion. However, slightly inclined loading may cause substantial resistance to the sliding relative to each other, and resistance increases with increasing relative velocities.

DISENTANGLING BLOOD COAGULATION WITH A COMPUTER

VWF is a giant protein that plays a key-role in blood clotting by cross-linking the extra-cellular matrix and blood platelets at sites of vascular injury. Its monomers consist of 2,050 amino acids. Functional VWF is a multimer of up to a few tens of covalently linked monomers. This size enables the protein to sense changes in the flow behavior of blood. Changes of shear rates induce a tumbling of VWF between globular and extended states. The latter can transform shear forces into a mechanical stretching along the protein chain that regulates VWF function via unfolding of domains or opening of inter-domain interfaces. By using a combined approach that makes use of computational and experimental techniques, we studied the structural changes induced by a disease-related mutation abolishing the monomer multimerization and the novel interaction of VWF with DNA, which may have potential clinical implications during inflammatory conditions.

Recent experimental studies have shown that DNA binds to VWF under shear conditions, reducing the capacity of platelets (through the glycoprotein receptor Ib GPB1b) and heparin to bind to VWF. It remains unclear, however, which regions in the VWF interact with DNA, and how DNA binding could modify the binding of GPB1b and heparin. We addressed these questions using molecular dynamics simulations and electrostatic calculations (Fig. 20). Our simulations of the isolated VWF A1 domain revealed that A1 has a positively charged region (blue), located at the surface, which remains constantly exposed over time scales of hundreds of nanoseconds. This region partially overlaps with the GPB1b binding site and fully overlaps with the heparin binding site. Our simulations therefore support the hypothesis that the positively charged region – in the A1 domain – attracts the negatively charged DNA, thereby blocking the binding of both GPB1b and heparin. These results associate a multi-binding role for the VWF A1 domain in which different biomolecules, such as DNA, GP1B and heparin, compete for similar binding regions, with consequences – in the case of DNA – during inflammatory processes.

The bleeding disorder von Willebrand disease (VWD) type 2A is caused by mutations of VWF. We provided a comprehensive picture of shear-dependent and shear-independent dysfunction of VWD type 2A mutants by performing static assays, microfluidic approaches, and molecular dynamics simulations. Two VWF chains are connected through disulfide bonds in between their C-

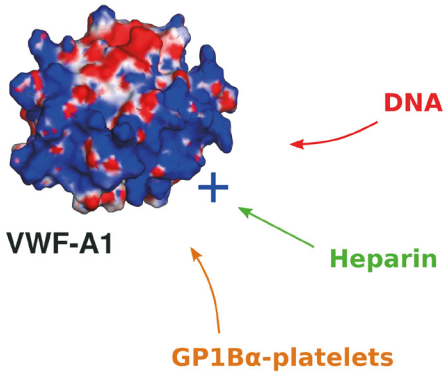


Fig. 22: VWF A1 domain is a potential binding site for DNA. The A1 domain (surface representation coloured according to its electrostatic potential) contains a positively charged region that is a potential binding site for the negatively charged DNA. This region overlaps with binding sites of Heparin and GPB1b-platelets complexes.

terminal CK domains. The cysteines C2771, C2773, and C2811 form such disulfide bonds. Moreover, the mutations Cys2771Arg and Ser2775Cys fully prevented the CK-dimer formation, whereas Cys2773Arg only partially impeded it. However, it remained unknown how these mutations influence the dimer formation. To examine the structural basis for the distinct roles of cysteines in dimer formation, we created a structural model of the CK domain based on its homology with the TGF-β2 protein and observed the disulfide-linked parallel segments of CK to be rigid, whereas the loops and the C-terminus were more flexible. The recent crystal structure of the CK domain validated the observed conformational dynamics of the CK domain. The mutations (mentioned above) induced conformational changes in the CK domain, and they were found to be highly correlated with the experimentally observed dimer abolishment (Fig. 22). Our results thus suggest that dimerization is favoured in compact confor-

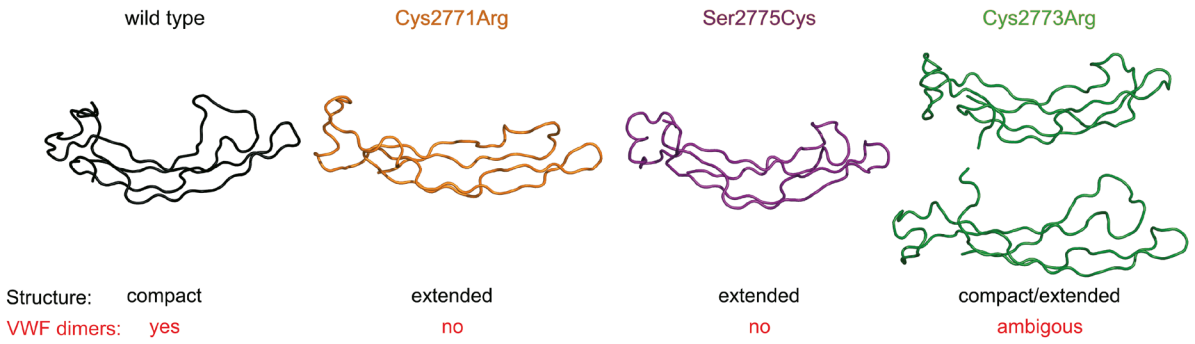


Fig. 23: Structural basis for the distinct roles of cysteines in VWF dimer formation. The structure of the CK domain was predicted by molecular dynamics simulations. In its wild type form (black), the CK domain adopted a compact conformation. The mutants Cys2771Arg (orange) and Ser2775Cys (purple) adopted extended structures, while the Cys2773Arg (green) adopted both compact and extended conformations. The mutation-induced structural changes are correlated with the experimentally observed dimer formation (indicated with red), suggesting that dimerization is favoured in compact conformations and blocked in extended conformations of the CK domain.

mations and blocked in extended conformations of the CK domain. Using this combined approach opens new possibilities for the diagnosis of new VWD phenotypes and the treatment choice for patients with VWF dysfunctions.

TOWARDS A RATIONAL DESIGN OF NEW INHIBITORS FOR THE TREATMENT OF ALZHEIMER'S DISEASE

The aim of this project is the rational development of drugs that efficiently slow down the progression of Alzheimer's Disease (AD). Small molecules have been designed as inhibitors of kynurenine pathway (KP) enzymes - namely, IDO1, IDO2, and TDO, with simultaneous antagonism of the dioxin/aryl hydrocarbon receptor (AHR). These enzymes are involved in the metabolism of tryptophan, while the activation of AHR is directly linked to the activation of the kynurenine pathway at the gene expression level. The binding pockets of AHR and KP enzymes are able to accommodate ligands of similar sizes and properties. This poses a serious challenge for the development of KP inhibitors – compounds that can also bind AHR and act as agonists should generally be avoided since the activation of AHR also activates KP enzymes, and development efforts should lead towards potent KP inhibitors, which are also antagonists or non-binders of AHR. In order to achieve this goal, the project focused on the AHR receptor in parallel with the KP enzymes. Several computational techniques, such as homology modelling, molecular dynamics (MD) simulations, and molecular docking, were combined in order to obtain an atomistic 3D model of the human AHR receptor, to deduce its binding site, and to predict its binding modes with agonists and antagonists (Fig. 24). The differences between the binding modes of antagonists and agonists were uncovered. This approach would enable us to screen high-affinity KP inhibitors to obtain a set of AHR antagonists that block the KP pathway. This would lead to the design of new drugs for the treatment of Alzheimer's Disease.

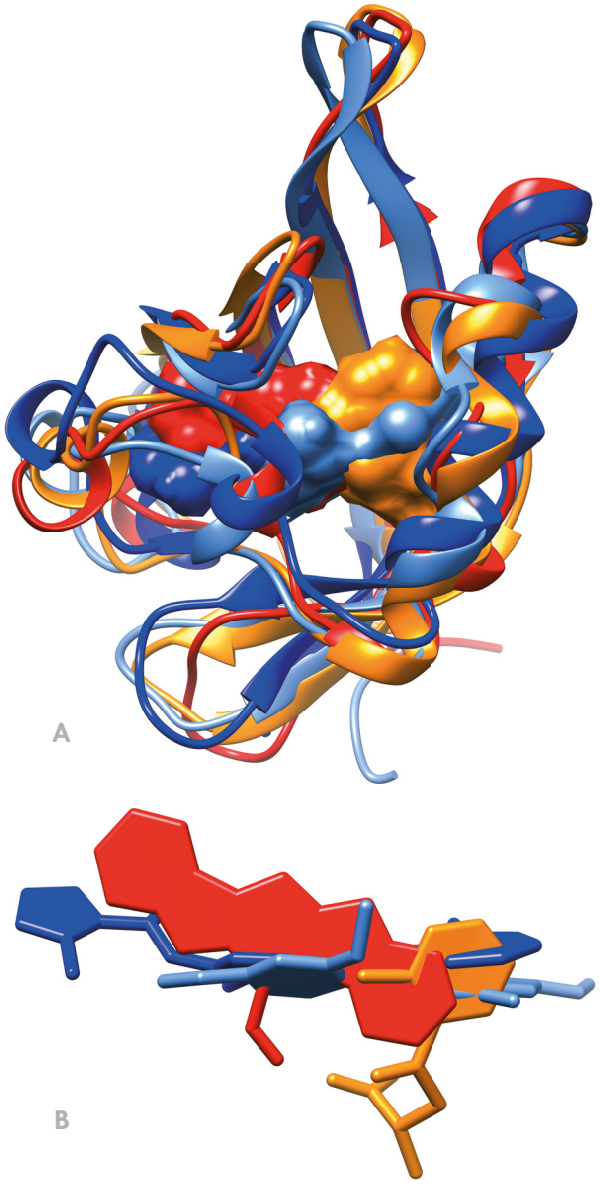


Fig. 24: (A) Atomistic model of the human AHR receptor developed with agonists (red and orange) and antagonists (blue and cyan) docked at the binding site, and (B) detailed positions of agonists (red and orange) and antagonists (blue and cyan) bound to AHR.

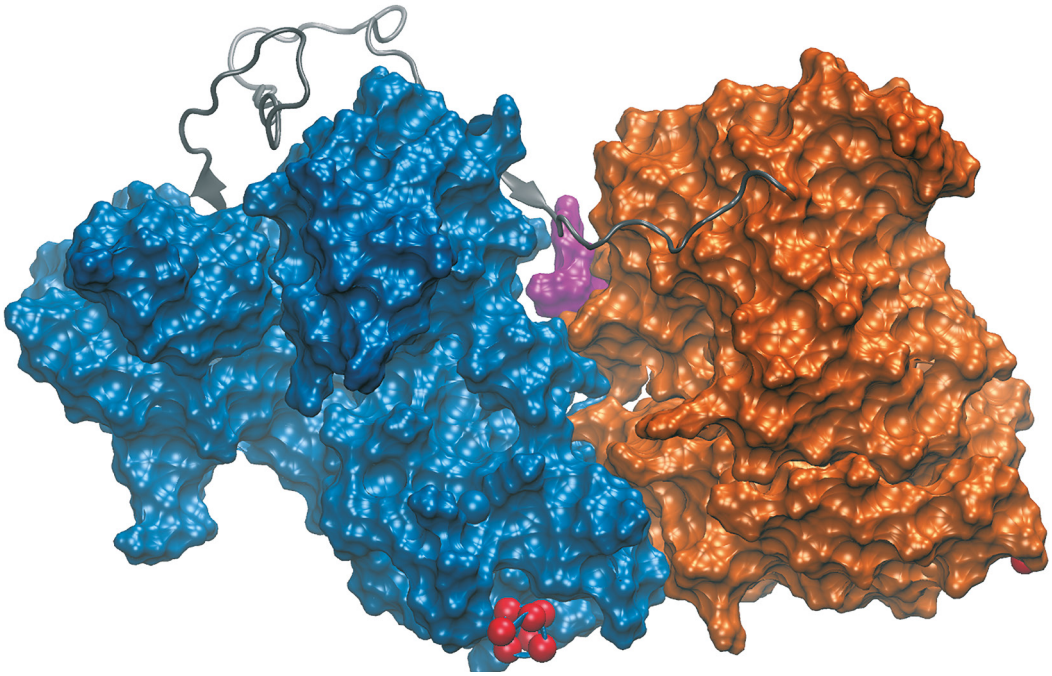


Fig. 25: Focal adhesion kinase. The kinase domain (orange) is inhibited by the FERM domain (blue), as this domain blocks the autophosphorylation site (magenta). FAK is linked to the membrane by interactions of its basic patch (red) to the lipid PIP2.

REGULATION OF FOCAL ADHESION KINASE

Focal adhesion kinase (FAK) is a component of focal adhesion sites and plays a crucial role in cell differentiation and motility. FAK exerts its activity at the crossroads of multiple-cell signaling pathways. The mechanical deformation of a cell induces the translocation of FAK at the cell periphery towards the direction of cell migration. This leads to the functional activation of FAK by tyrosine phosphorylation. After a sequence of tyrosine phosphorylation events, FAK adheres to downstream signaling proteins leading to the formation of very large multimolecular complexes called focal adhesions (FAs). Through FAs, the cytoskeleton of a cell connects to the extracellular matrix. In this way, mechanical forces can be transduced from outside to inside the cell and translated into biochemical signals. Thus, studying the mechanism of FAK activation may be very helpful for understanding cell signaling by mechano-sensing. If FAK indeed serves as a direct mechanosensor, i.e., is activated for phosphorylation by mechanical force, it would be a prime candidate for analyzing and re-engineering mesenchymal stem cell di_erentiation and bone formation in biomedicine. Despite multiple experimental studies of FAK's biological function, its activation mechanism remains elusive. Many stimuli have been reported to induce FAK activation, such as integrin signaling and direct engagement of the cytoplasmic regions of growth factor receptors [3, 14], but the molecular details of such interactions are unclear. In order to answer how FAK triggers intercellular signaling cascades and thereby guides cell differentiation, we analyzed FAK

activation mechanisms with a combination of Molecular Dynamics (MD) simulations and Force Distribution Analysis, a method developed in our group to track signal propagation through protein molecules.
We considered the two structurally known domains of FAK, namely the N-terminal FERM domain and the kinase domain Fig. 25, page 47).

The FERM domain is involved in auto-inhibition of the kinase by blocking a phosphorylation site (Tyr576/577) in the kinase domain. The exposure of this phosphorylation site induces the maximal activity of FAK. In addition, Daniel Lietha and co-workers (our collaboration partners) proposed phosphatidylinositol 4,5-bisphosphate (PIP2), which is abundant in the cell membrane, as a potential FAK activator. Their experimental data has recently shown that the negatively charged PIP2 binds to the positively charged basic patch on the FAK FERM domain, which induces conformational rearrangements. This binding promotes the exposure of the FAK autophosphorylation site (Tyr397) but cannot induce the exposure of the Tyr576/577 phosphorylation site, which is a hallmark of FAK activation. Our MD simulation data can now provide detailed molecular insights in the mode of PIP2-induced FAK phosphorylation.
Our data is consistent with the opening of FAK upon PIP2 and ATP binding as monitored by fluorescence resonance energy transfer. We demonstrated that PIP2 binds to a basic region on the FAK FERM domain, close to but not over-lapping with an autoinhibitory region on the FERM F2 lobe. ATP binding on the active site leads to partial opening between FERM F2 lobe and kinase C lobe, simultaneously closing the gap between kinase N lobe and kinase C lobe. PIP2 binding re-opened the gap, enabling the activation of the kinase domain.
Intriguingly, we found PIP2 binding to the basic patch to induce conformational changes at the kinase N-lobe, which is very distant from the binding site. This is the hallmark of an allosteric protein, which can be activated and deactivated on one side of the structure to a different side of the protein by regulatory binding events (here PIP2). For many proteins, including FAK, it has remained puzzling how allosteric signal propagation can bridge the distance between the binding and active site of the protein. Our Force Distribution Analysis was previously able to track signals in other allosteric structures with great success. A common finding was that forces can propagate through proteins very efficiently, even through regions where conformational changes appear minor. The reason is that stiff elements of a structure, by definition, only deform under high forces, or, in turn, can distribute high forces very efficiently since dissipation through conformational deformation is minor.
Moreover, we were able to obtain a signal propagation pathway from the PIP2 binding site to the ATP binding site and the site of the domain opening for FAK, which explains the observed long-range allostery (Fig. 26).
It is primarily the kinase (and to a lesser extent the FERM

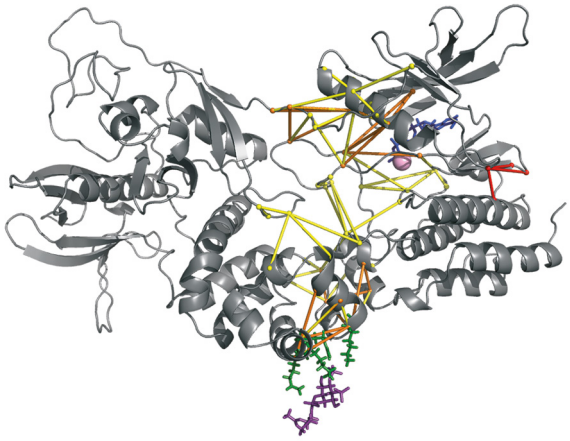


Fig. 26: Force Distribution analysis reveals the network of forces (yellow/orange sticks) propagating from PIP2 (magenta) to the basic patch (green) and all the way to the inter-domain interactions near the ATP binding site (blue).

domain) that contributes to the allosteric transition. We identified domain-domain interactions that can be expected to lead to an impairment of the allosteric mechanism if abolished through site-directed mutagenesis.

While both ATP and PIP2 very specifically changed the functional and conformational state of FAK, neither ATP nor PIP2 binding could induce full FAK activation. These results strongly support the hypothesis that an external mechanical force is required for the FAK activation, which is the subject of current investigations in our group.

2.6 Molecular and Cellular Modeling (MCM)

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, and what shape can the complex take? 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?

These questions are illustrative of the types of problem we address in our projects via the development and application of computational approaches to the study of biomolecular structure, dynamics, interactions, and reactions. We take an interdisciplinary approach, which entails collaboration with experimentalists and a concerted use of computational approaches based on physics and bio-/chemoinformatics. The broad spectrum of techniques employed ranges from interactive, web-based visualization tools to atomic-detail molecular simulations.

Die MCM-Gruppe ist primär daran interessiert, die Wechselwirkungen zwischen Biomolekülen zu verstehen. 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? Diese Fragen versuchen wir in unseren Projekten durch die Entwicklung und Anwendung rechnerischer Methoden zur Untersuchung biomolekularer Strukturen, Dynamik, Wechselwirkungen und Reaktionen zu beantworten. In enger Zusammenarbeit mit Experimentatoren verwenden wir in interdisziplinären Ansätzen rechnerische Methoden aus den Bereichen der Physik-, Bio- und Chemoinformatik. Das Spektrum unserer Methoden reicht dabei von webbasierten Visualisierungswerkzeugen bis hin zu Molekularsimulationen auf atomarer Ebene.

Die Ergebnisse unserer diesjährigen Arbeit präsentieren wir in drei ausgewählten Projekten. Sie demonstrieren einerseits die Methoden, die wir entwickeln, um makromolekulare Interaktionen zu modellieren und simulieren, und andererseits ihre Anwendungen in Biologie, Biotechnologie und Medikamentenforschung. Die Projekte beschäftigen sich mit (i) einem Webserver um Ligand-Protein Wechselwirkungen abzufragen (LigDig), (ii) Entwicklungen von unseren Methoden zur Simulation der Diffusion und Bindung von Makromolekülen und einem neuen Webserver zu diesem Zweck (WebSDA), und (iii) Simulationen der Struktur und Dynamik von parasitischen und humanen Cytochrom P450 Enzymen mit Relevanz für Wirkstoffdesign gegen Parasiten.

The MCM group in 2014 (f.l.t.r.): Rudi Tong, Mykhaylo Berynskyy, Antonia Stank, Neil Bruce, Stefan Richter, Jonathan Fuller, Daria Kokh, Ghulam Mustafa, Xiaofeng Yu, Prajwal Nandekar, Rebecca Wade, Stefan Henrich, Julia Romanowska

GENERAL NEWS:

Among this year's new group members, we welcomed Dr. Neil Bruce to work on the Human Brain Project and Ina Pöhner to start her doctoral studies contributing to the NMTrypI (New Medicines for Trypanosomatidic Infections) project. Dr. Anna Feldman-Salit returned to HITS for a short postdoc, and Renate Griffith spent her second summer here on sabbatical leave from the University of New South Wales in Australia. Several undergraduate and masters students did internships in the group, and Rudi Tong successfully completed his thesis project on a comparative analysis of adenylyl cyclase binding sites for his bachelor's degree in Molecular Biotechnology at Heidelberg University.

We completed the PPI-MARKER project supported by the EU International Research Staff Exchange Scheme (IRSES) this year. In this project, we collaborated with three other research groups on developing bioinformatics tools for analyzing protein binding sites and protein-protein interactions and identifying protein interactions of potential value as biomarkers for prostate cancer. As part of this project, Mehmet Öztürk spent some months in Prof. Biaoyang Lin's laboratory in Hangzhou in China working on the experimental characterization of the interactions of two DNA-binding proteins that he has modelled.

Along with the SDBV group, the MCM group participates in the NMTrypI (New Medicines for Trypanosomatidic Infections) project that began in February 2014. NMTrypI is

Group Leader
Prof. Dr. Rebecca Wade

Staff Members
Dr. Neil Bruce (from March 2014)
Dr. Anna Feldman-Salit (from Sept. 2014)
Dr. Jonathan Fuller
Dr. Stefan Henrich (until June 2014)
Dr. Daria Kokh
Dr. Michael Martinez (until March 2014)
Ina Pöhner (from July 2014)
Dr. Stefan Richter
Antonia Stank

Scholarship Holders
Mykhaylo Berynskyy (until Sept. 2014)
Ghulam Mustafa (DAAD Scholar)
Prajwal Nandekar (DAAD Scholar until July 2014)
Musa Özboyaci
Mehmet Öztürk
Xiaofeng Yu

Visiting Scientists
Dr. Jan Brezovsky (June – Sept. 2014)
Dr. Renate Griffith (April – July 2014)
Jie Liu (Jan. – June 2014)
Dr. Julia Romanowska (EMBO postdoctoral scholar, until June 2014)

Students
Eduard Bopp (until June 2014)
Annika Gable (Jan. - Febr. 2014)
Gaurav Ganotra (from Aug. 2014)
Anke Heit (Oct. – Nov. 2014)
Egle Maximowitsch (May – June 2014)
Rudi Tong (March - July 2014)
Talia Zeppelin (Erasmus student, from September 2014)

an interdisciplinary collaboration between eight academic research groups and five SMEs supported by the EU's 7th Framework Programme. The aim of the project is the development of drug candidates and biomarkers for the treatment of three neglected parasitic diseases, i.e., sleeping sickness, leishmaniasis, and Chagas' disease. The MCM group is contributing by applying protein modeling and computer-aided ligand design techniques.
The 7th EMBO Course on Biomolecular Simulation, which was organized by Michael Nilges (Pasteur Institute) and Rebecca Wade, took place in Paris in July (see Section 5.1.4).

This report describes the results achieved this year in three selected projects. They demonstrate the types of methods we develop to study macromolecular interactions and their application to problems in biology, biotechnology, and drug design. The projects are on (i) a web server for querying ligand-protein interactions (LigDig), (ii) improvements to our methodology for simulating the diffusional association of macromolecules and a new web server for this purpose (WebSDA), and (iii) simulations of the structure and dynamics of parasitic and human cytochrome P450 enzymes with implications for anti-parasitic drug design.

LIGDIG: A WEB SERVER FOR QUERYING LIGAND-PROTEIN INTERACTIONS

There has been growing discussion of the 'data-deluge' and the perceived problems surrounding it, namely how to store, process, and visualize data and how to combine multiple heterogeneous data sources. In the cross-disciplinary environment of systems biology research, one of the hurdles is the challenge of combining diverse types of data from a wide variety of sources. LigDig is a webserver designed to answer questions about ligand-protein interactions that previously required several independent queries to diverse data-sources [Fuller, 2014]. It thus uses the strategy of accessing multiple data sources and assisting the user to interpret results that might not be accessible by visiting a single data source alone. Additionally, LigDig assists users in performing basic manipulations of protein structures and analyses of the structures of protein-ligand complexes that can help researchers to gain new insights into their system of interest (see Fig. 27). The LigDig webserver is modular in design and consists of seven tools, which can be used separately or via linking the output from one tool to the next in order to answer more complex questions. Currently, the tools allow a user to (i) perform a free-text compound search, (ii) search for suitable ligands (particularly inhibitors) of a protein and query their interaction network, (iii) search for the likely function of a ligand, (iv) perform a batch search for compound identifiers, (v) find structures of protein-ligand complexes, (vi) compare three-dimensional structures of ligand binding sites, and (vii) prepare coordinate files of protein-ligand complexes for further calculations. LigDig is designed to be usable by non-experts in bio- and chemo-informatics. LigDig is available at http://mcm.h-its.org/ligdig.

SDA AND WEBSDA: SOFTWARE FOR THE SIMULATION OF THE DIFFUSIONAL ASSOCIATION OF MACROMOLECULES

Two important advances in our Brownian dynamics (BD) methodology and software for the Simulation of Diffusional Association (SDA) this year have been (1) the introduction of a long range Debye-Hückel correction for computing grid-based electrostatic forces between biomacromolecules [Mereghetti, 2014] and (2) a new web server, webSDA, to facilitate setting up, running, and analyzing SDA calculations.

BD is a simulation method that employs a mesoscopic model in which the solvent is treated as a continuum and the solutes are modelled as discrete entities at a level of detail appropriate for the problem being studied. BD

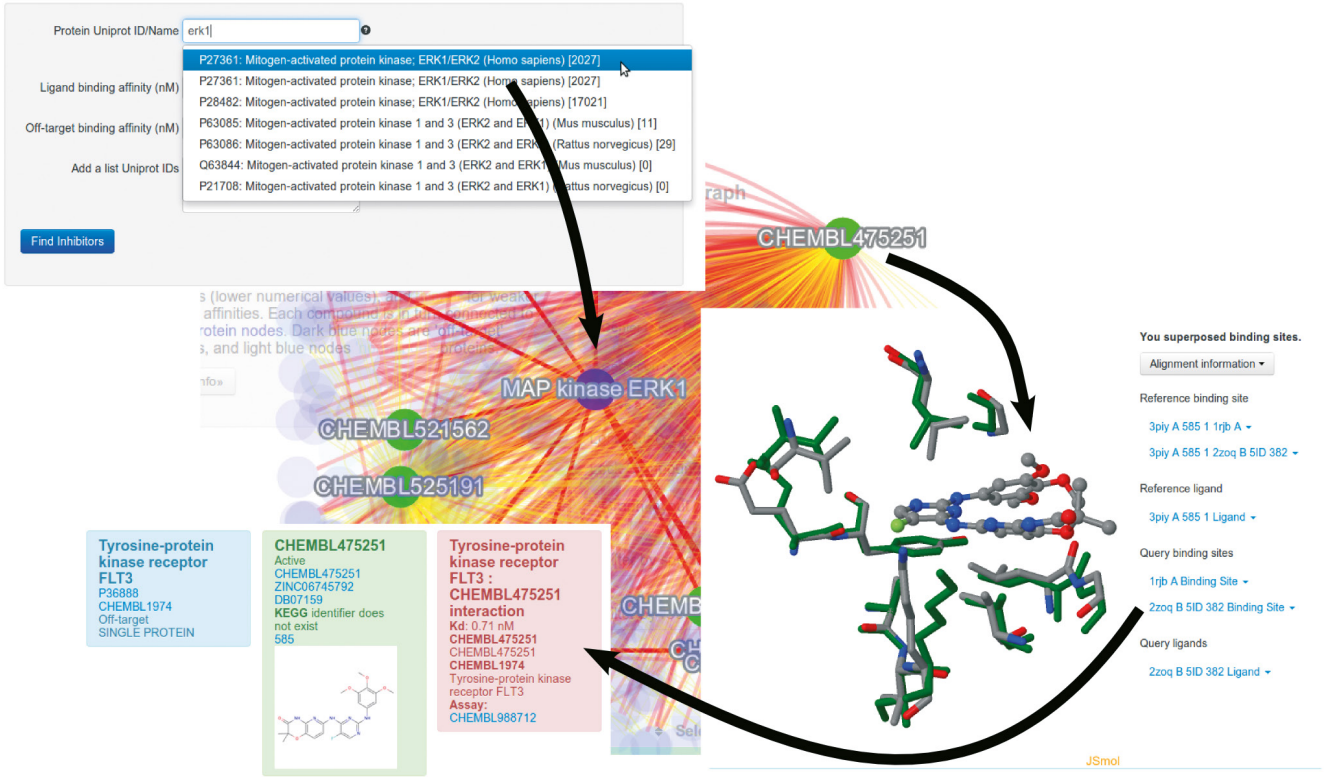


Fig. 27: Example showing a session with the LigDig webserver to investigate ligands binding to kinase proteins. [Fuller, 2014a].

can be used to study larger biomacromolecular systems on longer time-scales than is possible with classical atomic-detail molecular dynamics simulations. However, the simulation of large systems with many macromolecules using atomic-detail structures requires strategies to contain the computational costs, especially for the computation of interaction forces and energies. A common approach is to compute interaction forces between macromolecules by pre-computing their interaction potentials on three-dimensional discretized grids. For long-range interactions, such as electrostatics, grid-based methods are subject to finite size errors. We therefore introduced a Debye-Hückel correction to the grid-based electrostatic potential used in the SDA BD simulation software. We found that the inclusion of the long-range electrostatic correction increased the accuracy both of the protein-protein interaction profiles and of the protein diffusion coefficients at low ionic strength (see Fig. 28, page 54). An advantage of this method is the low additional computational cost required to treat long-range electrostatic interactions in

large biomacromolecular systems. Moreover, our implementation for BD simulations of protein solutions could also be applied in implicit solvent molecular dynamics simulations that make use of gridded interaction potentials.

webSDA is a webserver that allows users to run Brownian dynamics simulations with SDA from a web browser (see Fig. 29). webSDA generates input parameters and files automatically, runs short SDA calculations, and offers user-friendly visualization of output results. webSDA is envisaged to be useful for beginners learning how to perform Brownian dynamics simulations as well as for experts preparing input files for running simulations. webSDA currently has three modules: "SDA docking" to generate structures of the diffusional encounter complexes of two macromolecules, "SDA association" to calculate bimolecular diffusional association rate constants, and "SDA multiple molecules" to simulate the diffusive motion of hundreds of macromolecules. webSDA is available at http://mcm.h-its.org/webSDA/.

STRUCTURE AND DYNAMICS OF CYTOCHROME P450 ENZYMES

Cytochrome P450 (CYP or P450) comprises a large superfamily of heme-thiolate monooxygenases found in all domains of life. We are studying the structural and dynamic properties of several human CYPs important in drug metabolism and as targets for drug design [Ghulam, 2014]. We have also recently focused on human and parasitic CYP51 enzymes. CYP51 enzymes are important in sterol biosynthesis, and although all CYPs share the same overall three-dimensional fold, the CYP51 enzy-

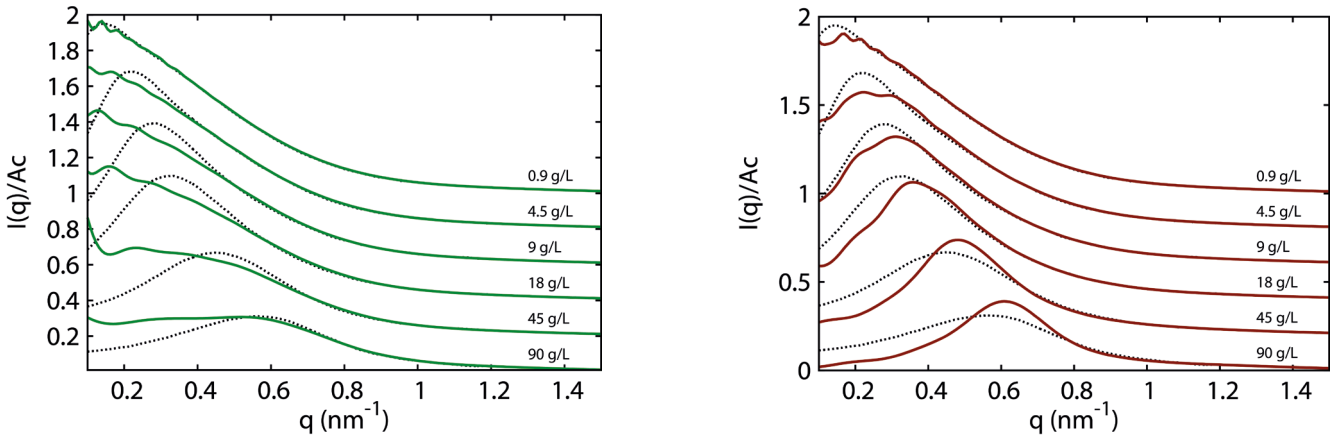


Fig. 28: Comparison with experimental data shows improvement in the modeling of protein interactions by using the Debye-Hückel long-range correction. Experimental (dashed lines) and computed (continuous lines) normalized small angle x-ray scattering intensities at different concentrations (indicated on the plots) of bovine serum albumin at low ionic strength. Computed curves (shifted by 0.2 on the vertical axis for visibility) from simulations without (A) and with (B) the Debye-Hückel long-range approximation. [Mereghetti, 2014]

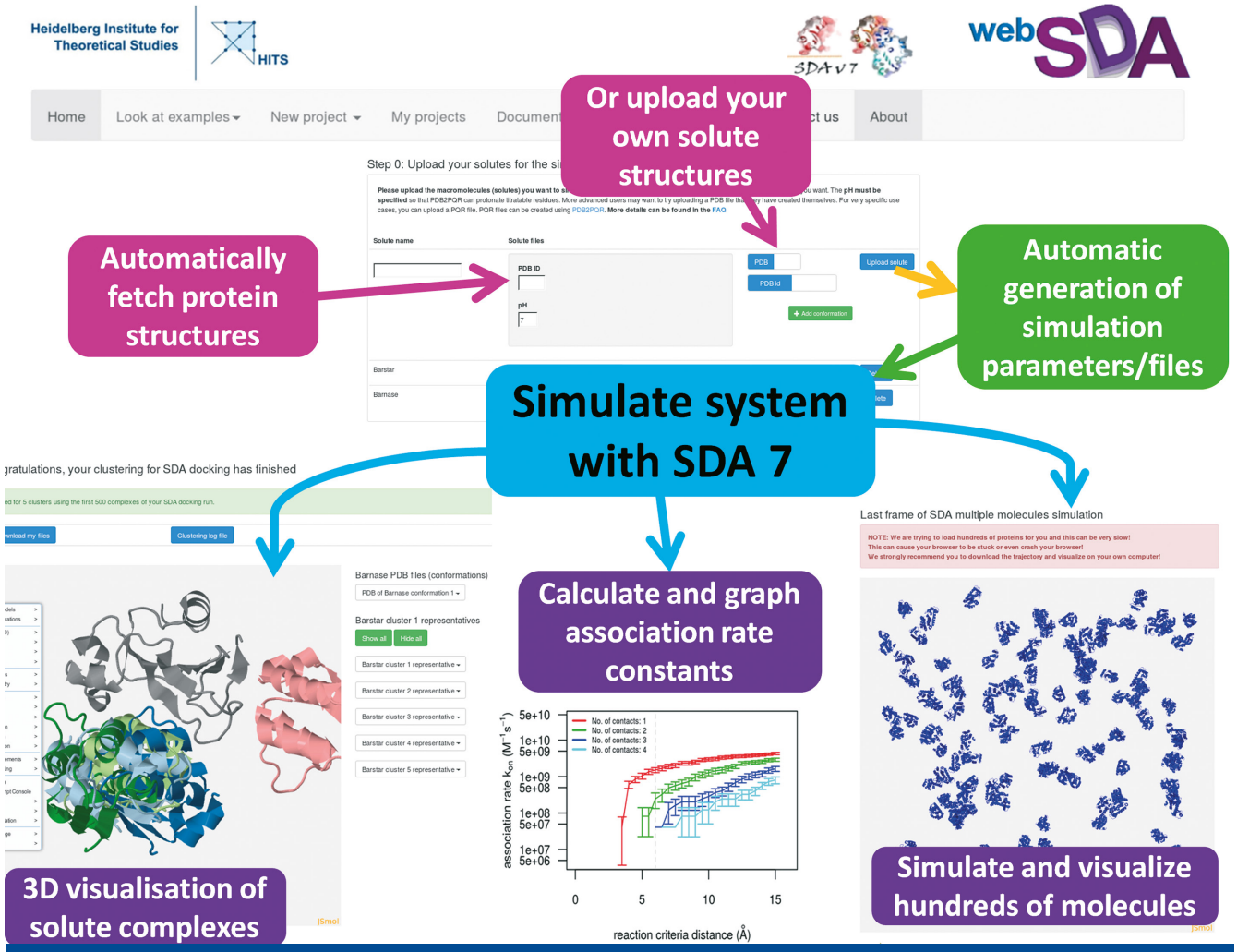


Fig. 29: Screenshots of webSDA. [Yu et al, 2015].

mes display a number of distinguishing features.

Our collaborator Galina Lepesheva (Vanderbilt University, Nashville, USA) and her colleagues have discovered selective inhibitors of parasitic trypanosomal CYP51s that do not affect the human CYP51 and that cure acute and chronic Chagas disease in mouse models without severe side effects. Crystal structures indicate that CYP51 may be more rigid than most CYPs and it has been proposed that this property may facilitate anti-parasitic drug design. Therefore, to investigate the dynamics of trypanosomal CYP51, we built a model of membrane-bound T. brucei CYP51 (Fig. 30, left) and then performed molecular dynamics simulations of T. brucei CYP51 in membrane-bound and soluble forms. We compared the dynamics of T. bru-

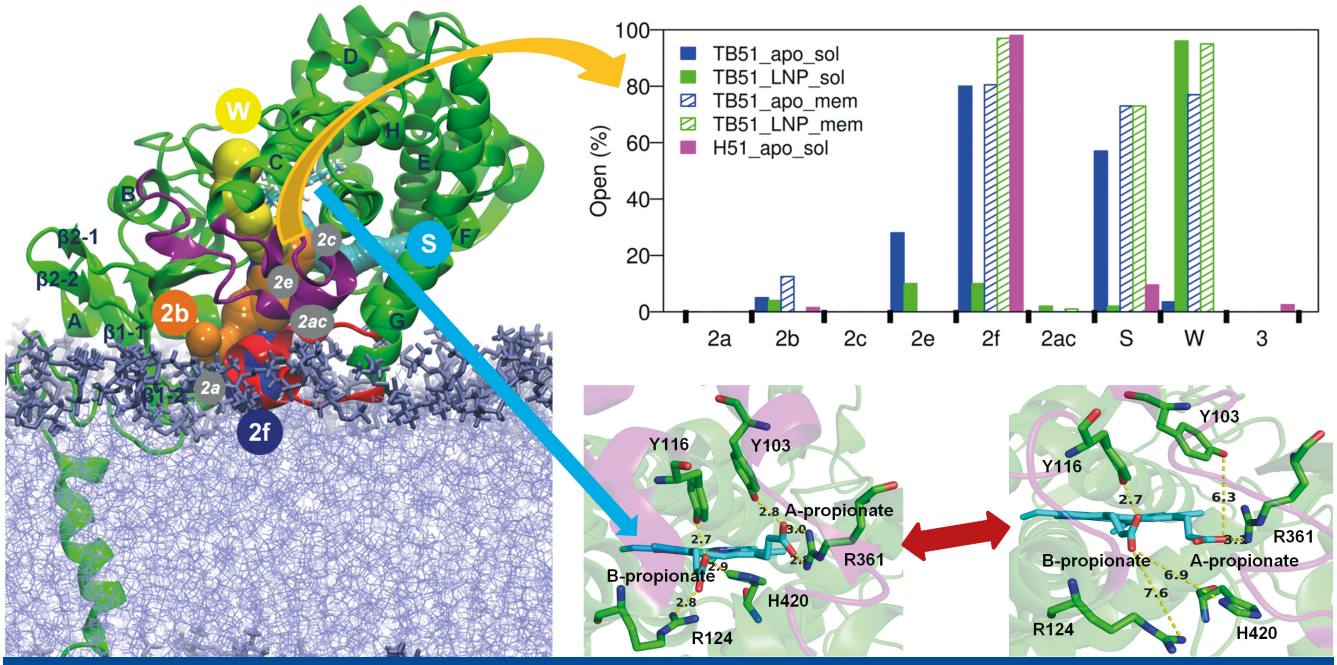


Fig. 30: Left: A model of the parasitic T. brucei CYP51 enzyme (green cartoon representation) in a membrane lipid bilayer (blue) indicating ligand egress routes from the buried active site. Right: Opening of the ligand egress tunnels in T. brucei CYP51 is influenced by the presence of the membrane or by a ligand in the active site, and a hydrogen-bonding network around the heme affects its conformation and the opening of tunnels in the protein. Yu et al. J. Molec. Recogn. (2015), 28, 59-73.

cei CYP51 with those of human CYP51, and two human CYPs. In the simulations, the CYP51s display low mobility in the buried active site, although overall mobility is similar in all the CYPs studied. The rigidity of the active site is consistent with the high substrate selectivity of CYP51. The simulations suggest that in CYP51, a different tunnel between the active site and the membrane is used compared with the human drug-metabolizing enzymes studied. Furthermore, the simulations reveal that the tunnel entrance residues have higher mobility, possibly because of the relatively large size of the CYP51 substrates. Further features peculiar to CYP51 are the heme conformations sampled and the opening behavior of a tunnel for water molecules (Fig. 30, right). Our simulations give insights into the dynamics of CYP51 that complement the available experimental data and have implications for drug design against CYP51 enzymes.

2.7 Natural Language Processing (NLP)

The Natural Language Processing (NLP) group develops methods, algorithms, and tools for the automatic analysis of natural language. The group focuses on discourse processing and related applications, like automatic summarization.

The NLP achieved great success in 2014, with papers accepted at all important NLP conferences. We were also successful again at our participation in the shared task of the Text Analysis Conference. Despite strong international and industrial competition, our team, headed by Alex Judea, ranked among the top 25% for mono- and cross-lingual entity linking.

In October 2014, Angela Fahrni submitted her PhD thesis, entitled “Joint Discourse-aware Concept Disambiguation and Clustering”. She will defend it in 2015. In November 2014, she began working as a Postdoc at IBM Research in Zurich.

In the fall of 2014, we received the good news that two new projects are going to be funded soon: In April 2015, we will establish a DFG-funded research training group focusing on the “Adaptive Preparation of Information from Heterogeneous Sources” (AIPHES) together with colleagues from the TU Darmstadt and the University of Heidelberg. The second project that we will initiate deals with “Scalable Author Disambiguation for Bibliographic Databases”. The project partners are the online computer science bibliography DBLP and Zentralblatt MATH, which has a similar service in mathematics.

Finally, Michael Strube was appointed co-chair of the largest and most important conference in Computational Linguistics, the 53rd Annual Meeting of the Association for Computational Linguistics to be held in Beijing in July 2015. He is looking forward to a busy year.

Die Natural Language Processing-Gruppe (NLP) widmet sich der automatischen Verarbeitung natürlicher Sprache. Der Arbeitsschwerpunkt der Gruppe liegt auf dem Text- oder Diskursverstehen und darauf aufbauenden Anwendungen wie etwa der automatischen Zusammenfassung.

Das Jahr 2014 war wissenschaftlich sehr erfolgreich für die NLP-Gruppe. Wir schafften es, auf allen wichtigen NLP-Konferenzen zu publizieren. Darüber hinaus nahmen wir wieder erfolgreich an einem Wettbewerb im Rahmen der Text Analysis Conference teil. Das von Alex Judea geleitete Team erreichte eine Platzierung unter den Top 25% im Bereich mono- und cross-linguales Entity Linking.

Angela Fahrni reichte im Oktober 2014 ihre Doktorarbeit mit dem Titel “Joint Discourse-aware Concept Disambiguation and Clustering” ein. Ihre Verteidigung wird im Laufe des nächsten Jahres stattfinden. Schon im November 2014 trat sie eine neue Stelle als Postdoktorandin bei IBM Research in Zürich an.

Im Herbst 2014 erhielten wir die guten Nachrichten, daß unsere zwei laufenden Projektanträge angenommen wurden: Im April 2015 werden wir zusammen mit Kollegen von der TU Darmstadt und der Universität Heidelberg ein Graduiertenkolleg zum Thema “Adaptive Preparation of Information from Heterogenous Sources” (AIPHES) beginnen. Das zweite neue Projekt wird sich “Scalable Author Disambiguation in Bibliographic Databases” widmen. Das Projekt wird zusammen mit der Online-Informatik-Bibliographie DBLP und dem Zentralblatt MATH, das einen ähnlichen Dienst für Mathematik anbietet, durchgeführt.

Schließlich wurde Michael Strube zum Co-Chair der größten und wichtigsten Konferenz in der Computerlinguistik, dem 53rd Annual Meeting of the Association for Computational Linguistics, das im Juli 2015 in Peking stattfinden wird, ernannt. Er erwartet ein mit viel Arbeit erfülltes nächstes Jahr.

CONCEPT DISAMBIGUATION

Concept disambiguation is the task of linking common nouns and proper names in a text (henceforth called "mentions") to their corresponding concepts in a predefined inventory, which, in our case, is Wikipedia. For instance, in the sentence „Watson was not connected to the internet for the match," „Watson" denotes a supercomputer from IBM and not the tennis player Heather Watson, and „match" is not a tool used to create fire, but is used in the sense of game. Resolving such ambiguities in an automatic way is a difficult task. Although it has been studied for years, it has not been entirely resolved.

One of the reasons that concept disambiguation is difficult is that the decisions for the different mentions are dependent on each other. For instance, the disambiguation decision taken for „match" depends on the disambiguation decision that is taken for „Watson", and vice versa. Modelling these dependencies in a time-efficient way is challenging. What makes concept disambiguation even more difficult is that only some but not all mentions in a text are dependent on each other with respect to their concepts. For instance, if the sentence "After a short break, the game continued" appears in the same text as our Watson example sentence, then the concepts for "Watson" and "break" do not depend on each other, while the concepts for "match" and "game" depend on each other. We argue that the other mentions to which a specific mention is interrelated depends on its embedding in discourse. However, how a mention is embedded in discourse depends, in turn, on its concepts. In the last three years, we have investigated how we can model these interdependencies and proposed a discourse-aware approach in the framework of Markov Logic that approximates these interdependencies. We have achieved good results for various datasets.
We have re-implemented the concept of a disambiguation system based on UIMA (Unstructured Information

The NLP group in 2014 (f.l.t.r.): Alexander Judea, Daraksha Parveen, Michael Strube, Angela Fahrni, Mohsen Mesgar, Nafise Moosavi, Benjamin Heinzerling, Yufang Hou, Sebastian Martschat

Group Leader
Prof. Dr. Michael Strube

Staff Members
Angela Fahrni (until Sept. 2014)

Scholarship Holders
Yufang Hou (HITS Scholarship, June – Aug. 2014, from Dec. 2014)
Alex Judea (HITS Scholarship)
Sebastian Martschat (HITS Scholarship)
Mohsen Mesgar (HITS Scholarship)
Nafise Moosavi
(HITS Scholarship, from Feb. 2014)
Daraksha Parveen (HITS Scholarship)

Visiting Scientists
Gordana Ilic Holen (March 2014)
Yufang Hou (Promotionskolleg Scholarship, until June 2014, Sept. – Nov. 2014)
Minsu Ko (April 2014)
Nafise Moosavi (until Jan. 2014)
Caecilia Zirn

Students
Nicolas Bellm
Thierry Göckel (until June 2014)
Benjamin Heinzerling (until Nov. 2014)
Hans-Martin Ramsl (until Aug. 2014)

Management Architecture), an up-to-date framework for systems dealing with unstructured information, like texts. The new system is easier to design, components can be easily exchanged or dropped, and new components can be plugged in. It is easier to extend because low-level functionalities are encapsulated in components and components can form groups in order to achieve high-level functionalities; it is faster because we have parallelized critical parts; it is embedded in a web service, which makes it easier to be used; and it is better suited for experiments because of its modular architecture.

The NLP group has participated in two shared tasks with the new version of the disambiguation system, namely the Entity Recognition and Disambiguation Challenge (ERD2014) organized by Microsoft and the Entity Discovery and Linking Task (EDL, TAC2014) organized by NIST. In both cases, the system achieved competitive results. Our presentation proposal for TAC2014 was accepted by the organizers because of our scientific contributions mentioned above.

ERROR ANALYSIS FOR COREFERENCE RESOLUTION

This year, we have extended and deepened our work on error analysis for coreference resolution. In particular, we have put our analysis framework on sound linguistic and representational foundations. We have furthermore applied our framework to analyze the errors made by four state-of-the-art coreference resolution systems on a large English benchmark dataset.

Consider the short text: “After the discussion, Obama confirmed he will return. Then the president and his bodyguards left.” Most current systems not equipped with world knowledge will output two entities {“Obama”, “he”} and {“president”, “his”}. Obviously, the system failed to recover a link. However, due to the complex nature of the coreference resolution task, it is not clear how to represent the error made by the system: Is it missing the link between “the president” and “Obama”? Can the error be attributed to deficiencies in pronoun resolution?

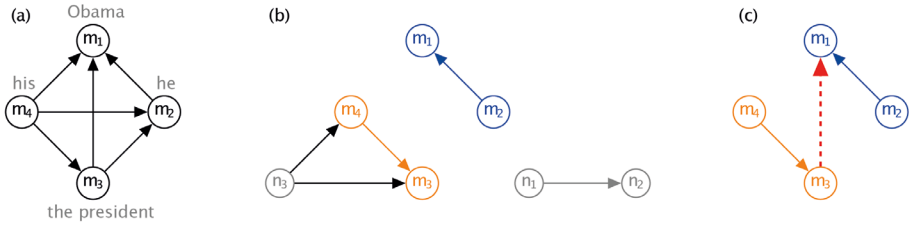


Fig. 31: Linguistic information can be injected into the framework during the spanning tree construction. Our current instantiation of the framework makes use of Ariel’s accessibility theory: Names and nouns refer to less-accessible entities than pronouns do. For such mentions, we prefer descriptive (name/nominal) preceding mentions. For pronouns, we choose the preceding mentions by distance.

Our error analysis framework is rooted in a graph-based entity representation. Given an entity we want to recover (the gold entity), we represent it as a complete one-directional graph in which the nodes are the entity's mentions (Fig. 31a). We employ the same graph-based representation for the system output (Fig. 31b). Since all of the entity information can be restored from a spanning tree of the gold entity, we extract each edge in the spanning tree that is not in the system output as an error (Fig. 31c).

We applied the above method to analyze errors of four state-of-the-art systems, including our in-house system, on a large and diverse benchmark dataset. In our analysis, we were especially interested in errors common to all systems since these errors constitute main challenges in coreference resolution. We focus on errors involving only proper names and common nouns, as preliminary experiments showed that these constitute a large fraction of all errors. We find that most errors in this category are common to all systems. While many errors are due to deficiencies in mention extraction, most errors are not resolved due to missing world knowledge, e.g., in "Florida" and "the Sunshine State" or "the Prime Minister" and "Mr. Papandreou". Our work highlights and quantifies the usefulness of world knowledge for coreference resolution.

PATTERN MINING FOR COREFERENCE RESOLUTION

Based on the results of our last year's research on coreference resolution, we have come to the conclusion that features play a major role in coreference resolution, and the impact of features on coreference results outweighs the impact of resolution methods. It is also the case in

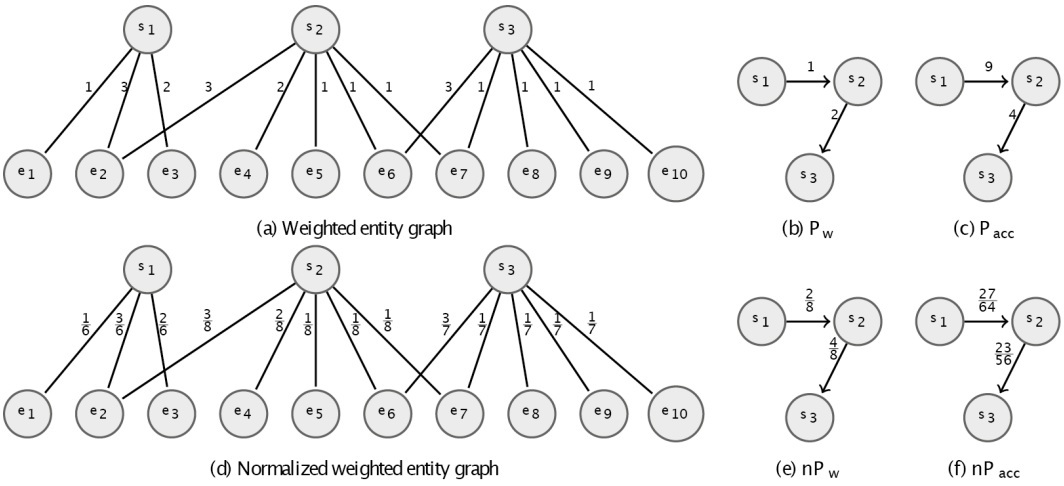


Fig. 32: Two weighted and directed projection graphs (P_w and P_acc) are obtained from the entity graph. P_w creates an edge between two sentence nodes if they share at least one entity. The weight of each edge in P_w is the number of shared entities of the two corresponding sentences (Fig. 32b). P_acc integrates grammatical information of shared entities in the corresponding sentences (Fig. 32c).

other areas of natural language processing (like word-sense disambiguation) that different learning methods result in similar performance when they use the same set of features.

Based on this conclusion, we began working on mining precise features from raw data so that we can have an informative representation of the data to be provided to learning methods.

Frequent patterns that occurred in the input data are promising candidates to be used as features so that the features can be defined as the statistically significant structures that are hidden in raw data. Mining frequent patterns is a well-known and essential problem in data mining, data analysis, and knowledge discovery, and there is a vast number of work on pattern and association rule mining. However, frequent pattern enumeration is proven to be an NP-hard problem, and it also produces a huge number of redundant features. Therefore, efficient discovery of discriminative patterns is an important and interesting challenge in data mining.

In this regard, we have developed a new discriminative pattern mining approach that can efficiently mine discriminative features from large-scale data. Our approach is based on the well-known frequent pattern tree (FP-Tree) structure, which only needs to scan data in two passes to represent data in a compact structure and is therefore appropriate for mining large amounts of data. Each node of the tree is labeled with an attribute-value of the data. The information of each node represents the information of a path of the tree that runs from the root to the current node. Therefore, each node represents the information of its corresponding attribute-value when combined with other attribute-values (attribute-values of its ancient nodes).

In our approach, the nodes of the FP-Tree are ordered based on the discriminative power of their corresponding attribute-values so that each intermediate node represents information of its corresponding attribute-value when it is combined with more-discriminating attribute-values. We propose an efficient approach for bottom-up processing of this tree that generates a relatively complete set of informative rules for classifying the data.

COMPUTING LOCAL COHERENCE

It is essential for NLP applications like multi-document summarization and text generating systems to produce a coherent text. Coherence shows that sentences in a text are semantically connected. One kind of connection between sentences in a text is entity transitions. The entity transitions in a text can be modeled by the entity graph model. This model represents which entity appears in which sentences and how these entities are distributed among sentences. The entity graph model uses a weighted bipartite graph, which consists of two disjoint sets of nodes: sentences and entities. Each edge connects a sentence node to an entity node if and only if the corresponding entity appears in the corresponding sentence. The weight of each edge is a numerical value that depends on the grammatical role of the entity in the corresponding sentence. Fig. 32a, page 61) illustrates an example of a weighted entity graph for three sentences.

These projection graphs ignore the effect of unshared entities and just take into account the information of shared entities. Inspired by linguistic theories, the importance of each entity should be defined relative to the other entities in a specific sentence. To this end, we propose a normalization method for the entity graph model.

Our graph normalization method has two phases. First, the importance of each entity in each sentence is affected by other entities in the sentence (Fig. 32d). Second, we compute normalized projections by integrating the relative importance of each entity with respect to other entities in a specific sentence (Imp(s_i,e)) and the relative importance of each sentence with respect to another sentence (Iw_i^j). The relative importance of each sentence in comparison with another sentence in a text is defined as its degree in the entity graph divided by sum of the degrees of these two sentences (Fig. 32e and 32f).

The average outdegree of a projection graph measures

the local coherence of a text. This value is equal to the sum of the weights in the projection graph divided by the number of sentences.

We evaluated this normalization on two tasks: insertion and readability assessment. In the insertion task, we removed one sentence and asked the normalized entity graph model to find the best position for the removed sentence. Our model proposes that the original position of the removed sentence is its best position with higher accuracy than the entity graph model does. In the readability assessment, the normalized entity graph is evaluated to distinguish texts from the Encyclopedia Britannica from the Britannica Elementary. Since the Elementary version is for children, its articles are supposed to be more coherent. Results of the normalized entity graph outperform the entity graph model and state-of-the-art models on readability assessment (Accuracy=0.897).
In the future, we will mine coherence patterns from the entity graph presentation of a collection of texts.

AUTOMATIC SUMMARIZATION

In NLP, summarization is an important task that automatically produces a summary of one or more documents. The main focus of research on summarization is to find the summary by dealing with the following three tasks simultaneously:
1. Importance
2. Redundancy
3. Coherence

The final summary should contain important information without any redundancy and must be readable. These three tasks have to be accomplished by all summarization systems.
Some well-known approaches for summarization are based on a graphical representation of the document in which sentences are considered nodes and edges are created by calculating the similarity between sentences. We also represent documents graphically. This graph is a bipartite graph with two types of nodes: one set of nodes is the entities while the other set is the sentences.
Our technique is based on the ranking and optimization of sentences in a document using a bipartite graph. The HITS (Hub and Authority) algorithm is applied on this graph to rank sentences. We use Integer Linear Programming for optimization. We optimize between importance and redundancy. Importance is represented by ranks of sentences.
The domain of our application is scientific journal articles from the bio-medical domain. Summarizing scientific articles is quite different than the standard NLP task of summarizing news articles. We introduce a new dataset with scientific articles. This dataset consist of 50 different scientific articles from the fields of biology and medicine. We choose these data because these articles are very long in contrast to news articles, for instance. Additionally, in terms of application, researchers and companies in the bio-medical domain cannot keep up with the research being published. Hence, researchers are very interested in a summarization system that can condense the important information in these articles. Our dataset is prepared by picking the articles from PLOS medicine journals. This is experimentally convenient: PLOS medicine is published with a Creative Commons License. The articles are not only published in PDF and HTML format, but also in a well-structured XML-format, which makes automatic processing less error-prone. Finally, the articles are headed not only by an abstract written by the authors, but also by a summary written by an editor who follows certain guidelines when writing this summary. We can use this summary as gold standard for our summarization system.
In future work, we will add a means for producing more coherent summaries to our system.

2.8 Scientific Computing (SCO)

The Scientific Computing Group (SCO) focuses on developing algorithms, computer architectures, and high-performance computing solutions for bioinformatics.
We mainly focus on

- computational molecular phylogenetics
- large-scale evolutionary biology data analyses
- supercomputing
- quantifying biodiversity
- next-generation sequence-data analyses

Secondary research interests include, but are not limited to,

- emerging parallel architectures (FPGAs, GPUs, Xeon PHI)
- discrete algorithms on trees
- population genetics

In the following section, we outline our current research activities. Our research is situated at the interface(s) between computer science, electrical engineering, biology, and bioinformatics. The overall goal 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, our overarching goal is 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 and can nowadays be analyzed, observed, and tracked at the DNA level. A famous dictum widely quoted in this context comes from evolutionary biologist Theodosius Dobzhansky: “Nothing in biology makes sense except in the light of evolution.”

Die Gruppe wissenschaftliches Rechnen (SCO) beschäftigt sich mit Algorithmen, Hardware-Architekturen und dem Hochleistungsrechnen für die Bioinformatik.
Unsere Hauptforschungsgebiete sind:

- Rechnerbasierte molekulare Stammbaumrekonstruktion
- Analyse großer evolutionsbiologischer Datensätze
- Hochleistungsrechnen
- Quantifizierung von Biodiversität

Sekundäre Forschungsgebiete sind unter anderem:

- Neue parallele Rechnerarchitekturen (FPGAs, GPUs, Xeon PHI)
- Diskrete Algorithmen auf Bäumen
- Methoden der Populationsgenetik

Im Folgenden beschreiben wir unsere Forschungsaktivitäten. Unsere Forschung setzt an der Schnittstelle zwischen Informatik, Elektrotechnik, 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”.

WHAT HAPPENED AT THE LAB IN 2014?

The following is an account of the important events at the lab in 2014.

In the summer of 2014, Alexis taught a class entitled “Introduction to Bioinformatics for Computer Scientists” at the Karlsruhe Institute of Technology (KIT).
The course was well-received, and Alexis received a very positive teaching evaluation from the students. In the 2014/15 winter term, we also introduced the new bioinformatics practical programming course and are working with a team of students from the summer class to design useful software for biologists.

The year was also very important for our former PhD student Fernando Izquierdo- Carrsco, who successfully defended his PhD thesis at the computer science department of the Karlsruhe Institute of Technology (KIT) in June 2014. By that time, Fernando had already started working at a startup company in the UK that is developing novel DNA sequencing technology.

Our PhD student Jiajie Zhang submitted his PhD thesis toward the end of the year and is the next in line to graduate. He will defend his thesis in March 2015. He will soon join Fernando in the development of bioinformatics software and methods at the same UK company.

We were also happy to welcome our new PhD student Lucas Czech, the second PhD candidate we recruited from KIT. Lucas won the Heidelberg science slam with a presentation on speech recognition.

In 2014, we hosted a visiting PhD student via our visiting PhD student program. Paschalia Kapli, a biologist from the University of Crete who works on the evolution of lizards, completed her second visit at our lab in March 2014. In November 2014, she successfully defended her PhD thesis at the department of Biology at the University of Crete. Alexis was a member of her PhD committee. We are happy that she will come back to Heidelberg once more to work with us as a postdoc in 2015.
We were also pleased to welcome Mark Holder and Emily McTavish. Mark Holder is an associate professor at the

The SCO group in 2014 (f.l.t.r.): David Dao, Alexey Kozlov, Diego Darriba, Tomáš Flouri, Jiajie Zhang, Paschalia Kapli, Kassian Kobert, Alexandros Stamatakis, Andre Aberer

Group Leader
Prof. Dr. Alexandros Stamatakis

Staff Members
Andre Aberer
Tomáš Flouri
Diego Darriba

Scholarship Holders
Kassian Kobert (HITS Scholarship)
Alexey Kozlov (HITS Scholarship)
Jiajie Zhang (HITS Scholarship)
Lucas Czech (HITS Scholarship, from July 2014)

Visiting Scientists
Paschalia Kapli (until March 2014)
Mark Holder (from Aug. 2014)
Emily Jane McTavish (from Dec. 2014)

Students
David Dao

University of Kansas and one of the world-leading experts in Bayesian reconstruction of evolutionary trees. He will be with us until mid-2015 and is enjoying his sabbatical in Europe. Emily is Mark's postdoc and joined him in our lab via a Humboldt fellowship. There is a lot of overlap in the research interests within our group, as evidenced by Emily's use of a software pipeline that was developed by our former PhD student Fernando.

Another highlight of 2014 was the summer school on computational molecular evolution that took place for the 6th time at the Hellenic Institute of Marine Research near Heraklion, Crete. Alexis was the main organizer of this event. We received 200 applications for the 35 available places. Our former postdoc Pavlos Pavlidis, our visiting PhD student Paschalia Kapli, and Andre Aberer also participated in the summer school as teaching assistants.

The biggest highlight of 2014 was that two lab members became parents and are doing very well. There is now a total of three babies and three fathers in the group.

INTRODUCTION

The term "evolutionary bioinformatics" is used to refer to computer-based methods for reconstructing evolutionary trees from DNA or from, say, protein or morphological data. The term also refers to the design of programs for estimating statistical properties of populations, i.e., for disentangling evolutionary events within a single species.
The very first evolutionary trees were inferred manually by comparing the morphological characteristics (traits) of the species under study. Nowadays, in the age of the molecular data avalanche, manual reconstruction of trees is no longer feasible. This is why evolutionary biologists have to rely on computers for phylogenetic and population-genetic analyses. In fact, following the introduction of so-called short-read sequencing machines (machines used in the wet-lab by biologists to extract DNA data from organisms) that can generate over 10,000,000 short DNA fragments (containing between 30 and 400 DNA characters), the community as a whole is facing novel and exciting challenges. One of the key problems that need to be tackled is the fact that the amount of molecular data available in public databases is growing at a significantly faster pace than the computers capable of analyzing the data can keep up with. In addition, the cost of sequencing a genome is decreasing at a faster pace than the cost of computation.
Accordingly, as computer scientists, we are facing a scalability challenge, constantly trying to catch up with the data avalanche and make molecular data analysis tools more scalable with respect to dataset sizes. At the same time, we also want to implement more complex and hence more realistic and compute-intensive models of evolution.
Another difficulty is that next generation sequencing technology is changing rapidly. Accordingly, the output of these machines, with respect to the length and quality of the sequences they can generate, is also changing constantly. This requires the continuous development of new algorithms and tools for filtering, puzzling together, and analyzing these molecular data.
Yet another big challenge is reconstructing the tree of life based on the entire genome sequence data of each living organism on earth.
Phylogenetic trees (evolutionary histories of species) are important in many domains of biological and medical research. The programs for tree reconstruction developed in our lab can be deployed to infer evolutionary relationships among viruses, bacteria, green plants, fungi, mammals, etc. In other words, they are applicable to all types of species. In combination with geographical and climate data, evolutionary trees can be used, e.g., to disentangle the geographical origin of the H1N5 viral outbreak, determine the correlation between the frequency of speciation events (species diversity) and climatic changes in the past, or analyze microbial diversity in the human gut. For conservation projects, trees can also be deployed to determine endangered species that need to be protected based on how many non-endangered close relatives they have.

Studies of population-genetic data, i.e., genetic material from a large number of individuals of the same species (e.g., a human population) can be used to identify mutations leading to specific types of cancer or other serious diseases.

As we have seen, one key challenge for computer science is scaling existing analytical methods to the huge new datasets produced by next-generation sequencing methods. We face these challenges every day since we are involved in a number of large-scale empirical data-analysis projects.

In the one thousand insect transcriptome project (1KITE, www.1kite.org), for example, we intend to disentangle the evolution of insects by using 1,000 insect transcriptome sequences. To analyze these data, we need to use a large supercomputer, such as the SuperMUC system in Munich. The transcriptome is the fraction of the DNA that is translated into RNA in each cell and encodes important functions with respect to the processes in and development of that cell.

In fact, the results of two such large data analysis projects both made it to the cover of Science in November (analysis of the first 140 insect transcriptomes of 1KITE) and December 2014 (analysis of 45 bird genomes). Our main contribution to these projects was the development of substantially more scalable software that reduces time to solution, i.e., the time for calculating the evolutionary trees, from 24 to just one month.

There is more to come with the 1KITE project since we are in the process of preprocessing the data of approximately 1,500 transcriptomes for phylogenetic analysis.

Another major challenge is the so-called multi-core revolution in parallel computing architectures. Throughout the 1980s and 1990s, computers became faster and faster as a result of increases in clock frequencies (the clock frequency of a processor essentially represents the number of arithmetic operations that can be executed per second). We have now reached a point at which sheer physical limitations dictate that clock frequencies cannot be increased beyond approximately 4GHz (4,000,000,000 instructions per second). Accordingly, the computer industry has started producing systems with more than one processor, so-called multi-core processors, so that further speed improvements (e.g., for analyzing larger and more complex phylogenetic trees) can be achieved. This represents a significant paradigm shift because in order to exploit the available computational resources (cores) in a new-generation processor, programs now need to be executed in parallel. In other words, programs need to be re-written so that they can perform their computational steps simultaneously on several processing cores.

The transition to multi-core architectures is a genuine revolution given that transforming a serial/sequential program into a parallel program is a non-trivial task. In other words, the task of parallelizing a code so that it can be executed simultaneously requires human intuition. In addition, each new generation of processors makes more cores available to the application programmer. As a consequence, the programming environments are becoming increasingly complex. Current hardware is also becoming hybrid, i.e., a collection of processors with distinct characteristics for different types of computations are integrated on the same chip or system. Evidently, this development yields the process of developing efficient parallel codes that are even more complex and error-prone. A substantial amount of manpower needs to be invested to efficiently exploit the capabilities of modern hardware for molecular evolutionary analyses.

In the following section, we briefly outline some research highlights in 2014.

TWO LARGE DATA ANALYSIS EFFORTS

The highlight of the year was our contribution to two Science papers that both made it to the cover of the journal in November and December 2014, respectively. The projects were fairly similar in the sense that huge datasets (about 45 entire bird genomes and 140 insect transcriptomes) needed to be analyzed. In both projects, our lab was

responsible for the last stage in the analysis pipeline, i.e., the inference of the phylogenetic trees. This included applying for CPU time at the Munich Supercomputer system SuperMUC, further optimizing our ExaML code for large-scale phylogenetic inference, and providing phylogenetic expertise to our project partners.
In addition, for the 1KITE project, we also conducted so-called divergence-time analyses. The task consists of transforming a phylogenetic tree that has relative branch lengths into a dated tree that has absolute branch lengths that represent real times. This is done by annotating the tree with fossils that can actually be dated and by subsequently calibrating the branch lengths based on this information under rather complex and compute-intensive statistical models.

Such a timed tree can then be used to correlate certain events in earth history (e.g., climate changes, tectonic events) with accelerated diversification rates of the species under study. Unfortunately, while we now have scalable methods at our disposal to infer trees with ExaML and ExaBayes (see below), the scalability of dating methods lags behind. In a heroic effort, our postdoc Tomas used several clusters and complex job-submission- and control scripts to carry out the divergence time analyses for the insect transcriptome project under substantial time pressure. Clearly, we need to develop novel and more scalable methods for dating because the ad hoc solution we deployed for dating the tree of 145 insects will never scale to the tree of 1,400 insects that we will infer in 2015. To this end, we are currently working in collaboration with our visiting professor Mark Holder toward developing novel and more scalable methods for this purpose.

We were also happy that a large number of our lab members were on the respective author lists: major contributions from Andre (bird paper) and Tomas (insect paper) as well as regular contributions from Paschalia, Alexey, and former PhD students Simon Berger and Fernando-Izquierdo Carrasco.

A rather sad consequence of these projects is that our moments of glory in Science are pretty much over. As a computer scientist, one rarely gets the chance to publish in Science. This was mainly feasible in the above cases because the software for analyzing these data had to be actively developed and adapted in the course of the projects. Now that it has become available as production-level open-source code, biologists will, sadly enough, be able to publish 'fancy' Science papers without us. Nonetheless, this is in line with the major goal of our lab: to enable research in evolutionary biology.

At the technical level, there has been substantial progress with our ExaML (Exasacle Maximum Likelihood) code that was deployed for both projects. Alexey Kozlov conducted some initial exploratory work with the greatly appreciated help of Christian Goll from the IT group to explore if the main computational kernel of ExaML can be efficiently mapped onto the Intel Xeon PHI multi-core processor. Note that the PHI processor seems to be a promising new accelerator architecture. It is being increasingly purchased by supercomputing centers. In fact, the accelerator performance share of the PHI for the top 500 supercomputers increased from 5% in the June 2014 list of top systems to 30% in November 2014. Alexey's initial work was a pure proof-of-concept approach. Based on the rather encouraging results and because of the availability of a new PHI-based cluster at the Munich supercomputer facility, we decided to fully integrate the PHI version into ExaML at production level. This required a substantial amount of re-engineering. Now, ExaML is fully hybrid, i.e., it can simultaneously use the PHI and 'normal' x86 multi-core processors. This also required the use of a nested load-balance algorithm (see below) that distributes data among 'normal' cores and then, in a second step, within the PHI processor.
We worked with Andre Aberer on improving the I/O performance of ExaML. At program start, we needed to read in a large file that contains the genomic sequences. Since each processor only processes a part of each genome,

the file reading can become complex and time-consuming. Imagine the following scenario: We are using 4,000 cores to infer the tree, and just reading the alignment at program start-up requires 15 minutes. This means that we are wasting 1,000 hours of accumulate CPU time just by waiting to read the data. By re-designing the input file format and file indexing, we managed to reduce this I/O time from 15 minutes to just 1 minute, i.e., in our example, we now just waste 67 hours of accumulated CPU time for file I/O.

Another major improvement was the development of a new load balance algorithm that distributes computational load more evenly among processors. This new algorithm is presented in a separate section below. Finally, we have recently extended ExaML by additional substitution models for binary as well as protein sequence data, completely re-written the documentation, and included more error checks so that it becomes more difficult for users to make 'silly' mistakes with the code.

On a slightly different note, some philosophical issues also exist in conjunction with the above-mentioned two papers. Hence, one of the next steps is to conduct a critical assessment of what has been achieved. The two main issues are reproducibility and software quality.

With respect to reproducibility, it is in fact not really feasible to reproduce the results. One of the main reasons for this is that no supercomputer center in the world will grant millions of CPU hours to researchers that simply want to reproduce results obtained by others. Unfortunately, computational resources are too scarce and competition for obtaining CPU time on these systems is too intense to allow for this.

Furthermore, there is the issue of parallel reproducibility. Because most scientific applications rely on floating point arithmetics that can yield different results depending on the order of additions (for instance, i.e., a +(b + c) might

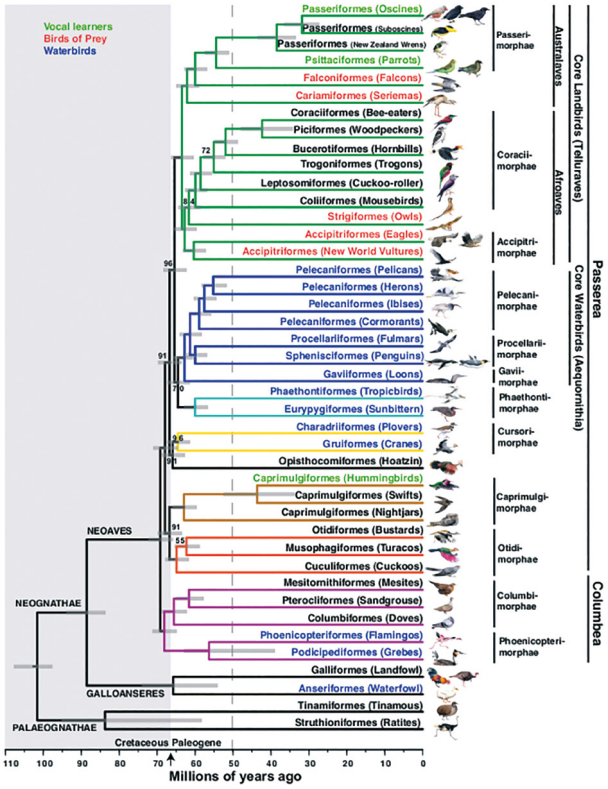


Fig. 33: The phylogeny of birds.

not be equal to (a+b)+c), results may be different if we run ExaML with 2,000 or, say, 4,000 cores in parallel. These slight numerical variations can and actually do produce different final tree topologies. In particular, this phenomenon can have a pronounced impact with Maximum Likelihood models since frequentist methods strive to return a point estimate, i.e., the parameter configuration that maximizes the likelihood. This numerical sensitivity of point estimates might be the best argument in favor of using Bayesian methods that merely sample the posterior probability distribution.

Another major area of concern is software quality and analysis pipeline complexity per se. Note that here we focus on implementation quality and errors and not on conceptual or mathematical errors that occurred before the actual coding step. With the advent of Next Generation Sequence (NGS) data, bioinformatics analysis pipelines have become increasingly complex.

In the good old 'Sanger sequencing days', the analysis pipeline was rather straight-forward, once the sequences were available. For a phylogenetic study, it consisted of the following steps: align → infer tree → visualize tree.

For NGS data and huge phylogenomic datasets, such as the insect transcriptome or bird genome projects mentioned above, pipelines have become substantially longer and more complex. They also require user expertise in an increasing number of bioinformatics areas (e.g., orthology assignment, read assembly, dataset assembly, partitioning of datasets, divergence times inference, etc.). In addition, these pipelines typically require a plethora of helper scripts, typically written in languages such as perl (a language that is highly susceptible to coding errors due to the lack of typing) or python to transform formats, partially automate the workflow, and connect the components. Our main concern is that if each code or script component i used in such a pipeline has a probability of being 'buggy' Pi(bug), the probability that there is a bug in the pipeline increases dramatically with the number of components. If detected too late, errors in the early stages of pipelines (e.g., NGS assembly) for large-scale data analysis projects can have a dramatic impact on all downstream analyses (e.g., phylogenetic inferences, dating): They will all have to be repeated. In fact, this has happened in every large-scale data analysis project we have been involved in thus far.

Since we are concerned about software quality with Postdocs Diego and Tomas, we initiated the so-called 'crappy software project''. In this project, we intend to assess the quality of 15 often-cited bioinformatics codes by using a set of simple criteria and to draft a list of the best practices for improving software quality. We also intend to more closely collaborate with colleagues from the computer science department at KIT who work on software engineering. The overall goal of this project is to raise the awareness of the community with respect to software quality and application of 'better' software development practices.

INTRODUCING EXABAYES

In 2014, Andre worked very hard on developing, publishing, and releasing a novel code for Bayesian phylo-

Fig. 34: The cover of the Science issue containing the insect phylogeny (Cover: Science).

genetic inference: ExaBayes. While ExaML was already available for computing phylogenies on supercomputers under the Maximum Likelihood criterion, a comparable scalable code for Bayesian phylogenetic inference was not yet available. Using a great many insights, the parallelization scheme, and the likelihood kernel of ExaML, Andre designed ExaBayes. Using the SuperMUC supercomputer, Andre was able to show that his code scales up to 32,000 cores for a dataset of 200 sequences that each had a length of 100,000,000 DNA characters. This means that one can now calculate whole-genome phylogenies on datasets comprising 200 species and more. Thus, with ExaML and ExaBayes, our lab has now developed scalable tools for phylogenetic inference on supercomputers for the two most widely used statistical approaches to building phylogenies. It is worth noting that numerous concepts we had developed for achieving this type of scalability, e.g., the parallelization approach, the load-balancing algorithm, and the parallel I/O method, are generally applicable to all codes that conduct phylogenetic likelihood calculations. To this end, we believe that we have substantially contributed toward transforming evolutionary biology into a true computational science.

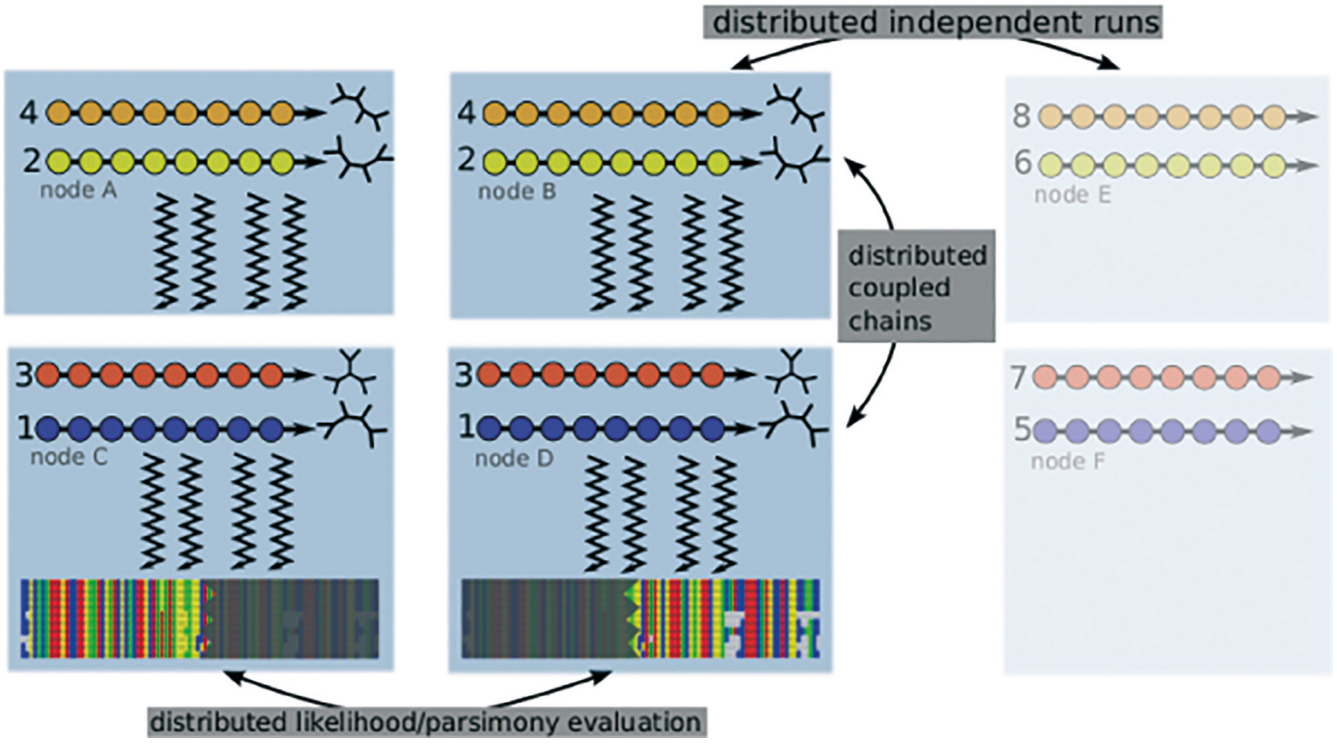


Fig. 35: Schematic parallelization of ExaBayes.

THE BEAUTY OF THEORY AND PRACTICE

In a paper with Andre, Kassian, and Tomas that I am quite fond of, we combined the beauty of theory with practice. As mentioned in the previous section within the framework of ExaML and ExaBayes development, we had to work on improving the data distribution to the computing cores to attain better parallel efficiency. The problem could be formulated as a 'classic' optimization problem in theoretical computer science. We showed that the problem is NP-hard, i.e., all exponentially possible data distributions need to be evaluated to find the optimal one. In addition, we developed a so-called approximation algorithm, i.e., an algorithm that does not solve the problem exactly, but is guaranteed to find a solution that is provably only worse than the optimal solution by a certain factor. This factor or bound for our approximation algorithm was very low and we are hence guaranteed to find near-optimal solutions. This was nice theoretical work. However, the beauty of it was that this was not a mere theoretical exercise, but that we implemented the algorithm in ExaML and ExaBayes and were able to show that it could improve parallel efficiency by a factor of up to 8 in the best possible case.

THE PEAR SOFTWARE

In early 2014, we published and released the PEAR software that can be used for overlapping (puzzling together) corresponding sequence pairs generated by Illumina NGS machines. As we showed in the paper, the software is both fast and accurate. Fortunately, Jiajie was clever enough to predict that PEAR might become of interest for commercial companies that provide NGS services. While the software is freely available to the academic community, it is not for companies. Therefore, we have already sold a couple of licenses to commercial companies. The income generated by these licenses will be used for what we have called the 'PEAR scholarship'. The scholarship will be granted to master students at KIT who want to do their master's thesis with us so that they can work abroad in labs that we collaborate with during a part of their thesis.

A NEW RAXML VERSION

Another highlight in 2014 was the release and publication of a new version 8 of our flagship code RAxML for phylogenetic inference. Alexis invested a lot of time into coding and adding new features. More importantly, the documentation was finally completely re-written, resulting in a 60-page manual. The paper that was published in January 2014 has already been cited 200 times according to Google scholar.

FUNDING

Our Postdoc Tomas Flouri is funded by the DFG (German Research Foundation).
Andre Aberer, Diego Darriba, David Dao, Jiajie Zhang, Kassian Kobert, Lucas Czech, Mark Holder, and Alexey Kozlov are funded directly by HITS. Emily McTavish is funded by the Humboldt Foundation. Our visiting PhD student Paschalia Kapli was funded partially by HITS and partially via the EU as well as by Greek national funds. The summer school in Crete was funded by EMBO. Alexis' visits at the University of Arizona at Tucson will be funded via an NSF grant for work on terraces in tree space.

OUTLOOK

We have assembled a strong team of bioinformaticians, computer scientists, biologists, and mathematicians to address the challenges that lie ahead. We intend to make use of this broad and diverse reservoir of knowledge to improve statistical models of evolution, scale up evolutionary bioinformatics algorithms, analyze and exploit emerging parallel computer architectures, and infer even larger phylogenies of insects in the framework of the 1KITE project. One of the key challenges will be to design new

methods and codes for dating large phylogenies. In 2015, we also intend to work more intensively on code quality issues and to generally begin to re-assess and call into question the available methods and tools.

2.9 Scientific Databases and Visualization (SDBV)

The main focus of our interdisciplinary group is data management for scientists (mainly systems biologists). SABIO-RK is a reaction kinetics database that is populated and curated (structured and enriched) by professionals. SEEK is a system that enables users to collect, curate, share, and disseminate their own data, models, and standard operation procedures (SOP) with research partners. Excemplify is a system that reduces work in curating immunoblot experiments.

While these systems all cater to researchers in their respective application domains, the Operations Explorer and the discover the liver portal are geared towards other parts of the public. The Operations Explorer is a tool for journalists that visualizes both medical and social data on the German population from Destatis (Statistisches Bundesamt). The discover the liver portal is designed for the science-interested but non-specialist public.

2014 was an exciting year for science infrastructure builders like us. With our contribution, ISBE, the Infrastructure for Systems Biology in Europe, we have advanced to the point of presenting a business case document. In Germany, the German Network for Bioinformatics Infrastructure (de.NBI) is forming. HITS will contribute data management via a SEEK instance, as well as SABIO-RK curation. de.NBI and ISBE blend well with FAIRDOM. Together with partners from Manchester and Zurich (Uni ZH & ETH), the new project performs data and model management for the European ERASysAPP funding initiative and beyond.

Data management as an infrastructure needs standardising. For this reason, SDBV has been active in standardisation communities for a long time. The new NormSys project granted this year by the BMWi is linking our community standards work to international standardization organisations.

Finally, we continue to provide the infrastructure for the data management of the Virtual Liver Network (VLN) and the SBEpo. In the new EU project NMTrypI (New Medicines for Trypasoma Infections) we are also the data management provider, taking again the opportunity to collaborate with Rebecca Wade‘s MCM group.

Im Zentrum unserer Aktivitäten steht das Datenmanagement für Wissenschaftler, hauptsächlich Systembiologen. SABIO-RK ist eine Datenbank für biochemische Reaktionskinetiken, die von Biologen und Biochemikern gepflegt wird. SEEK ist ein System, das seinen Nutzern ermöglicht, eigene Daten, Modelle und Prozeduren zu speichern und zu teilen. Excemplify ist ein Software Werkzeug, welches es erlaubt, Immunoblot-Experimente sowohl semi-automatisch zu planen als auch die Daten zu verwalten und mit anderen Experimentatoren zu teilen. Neben diesen Systemen, die Wissenschaftlern bei ihrer Arbeit helfen, sind der Operations Explorer und das Portal “Entdecke die Leber” auf ein anderes Publikum ausgerichtet. Das erstere bereitet Daten des statistischen Bundesamtes für Journalisten auf, das letztere ist für die wissenschaftlich interessierte breitere Öffentlichkeit gedacht.

2014 war ein sehr positives Jahr für Gruppen, die sich wie wir mit Daten-Infrastruktur befassen. Mit unserem Beitrag hat ISBE, die Infrastruktur für Systembiologie in Europa, das Business-Case-Dokument präsentiert. In Deutschland formiert sich das Deutsche Netzwerk für Bioinformatik Infrastruktur (de.NBI). HITS wird hier SEEK-basiertes Datenmanagement und Kuratierung von SABIO-RK-Daten beitragen. ISBE und de.NBI passen gut zu FAIRDOM, unserem neuen Projekt, in dem wir mit Partnern aus Manchester (Koordination) und Zürich das Datenmanagement für ERASysAPP und weitere Projekte leisten.

Standardisierung ist essentiell für ein effektives Datenmanagement. Aus diesem Grund ist die SDBV seit Jahren in Standardisierungsgremien aktiv. Das neue NormSys-Projekt, das in diesem Jahr durch das BMWi bewilligt wurde, stellt eine Verbindung zwischen unserer Arbeit in der Wissenschaft und internationalen Standardisierungsorganisationen wie der ISO her.

Zudem betreuen wir weiterhin die Datenmanagement-Infrastruktur für das Virtual Liver Network und SBEpo. Neu hinzu kam „New Medicines for Trypanosoma Infections“. Hier arbeiten wir mit Rebecca Wades MCM-Gruppe zusammen.

INFRASTRUCTURES FOR SYSTEMS BIOLOGY

Most of the projects in the group aim at constructing and maintaining data management infrastructure and databases for the up-to-date research field of systems biology, in which experimentalists and theoreticians as well as clinicians work together to construct and simulate complex computer models that help to better unravel the biological processes going on in cells, tissues, organs, and the entire organism. Such mathematical models of systems can be created by integrating heterogeneous biological knowledge about parameters changing in their respective contexts as a result of different temporal and environmental conditions. Thus, the modeling process is highly dependent on the ability to access and integrate heterogeneous data from databases and other published sources as well as directly from ongoing experimental investigations. The data in systems biology projects are obtained from a wide variety of sources, and the methods employed are just as variegated. Data and corresponding metadata (data describing the data) have to be structured, combined, and integrated in order to make them comparable and to construct simulatable computer models based on these data. Likewise, resulting models have to provide coherent and compatible interfacing options for their eventual integration in more complex and modularized computer models. Together with its collaboration partners, the SDBV group develops and maintains complex data management platforms that help to structure, exchange, integrate, and publish experimental data, models, workflows, and additional information pertaining to them. Beyond that, the group is involved in large national and European consortia that aim at providing the infrastructure backbone for systems biology research in other projects (like ISBE, FAIRDOM, and de.NBI) or do their own research based on the infrastructure provided by our group (such as the Virtual Liver Network, SBEpo, and NMTrypI).

SDBV

The SDBV group in 2014 (f.l.t.r.):Renate Kania, Wolfgang Müller, Lihua An, Martin Golebiewski, Jill Zander, Ivan Savora, Meik Bittkowski, Olga Krebs, Andreas Weidemann, Quyen Nguyen, Iryna Ilkavets

Group Leader
Priv.-Doz. Dr. Wolfgang Müller

Staff Members
Lihua An
Meik Bittkowski (until Dec. 2014)
Martin Golebiewski
Dr. Iryna Ilkavets
Renate Kania
Dr. Olga Krebs
Quyen Nguyen
Dr. Maja Rey
Ivan Savora
Lei Shi (until Feb. 2014)
Dr. Andreas Weidemann
Dr. Ulrike Wittig

Student
Jill Zander

DE.NBI (GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE)

de.NBI contains a number of local, well-equipped, and specialized service units with outstanding expertise in many areas of life sciences research and also has local nodes managing high-quality data resources. In collaboration with the University of Rostock, the SDBV represents the Data Management Node of the de.NBI network with a focus on data management (SEEK), data curation and exchange (SABIO-RK), and experimental data workflow management (Excemplify).
The central mission of this network is to provide and continuously further develop bioinformatics services and software solutions for basic and applied life-sciences research and to establish standards for data storage, analysis, management, and exchange. The network is complemented by a coordinating body for the development of long-term strategies for the sustainability of the offered services and available data resources. In addition to providing comprehensive bioinformatics services to users, de.NBI coordinates bioinformatics training and education in Germany and the cooperation of the German bioinformatics community with international bioinformatics network structures.

ISBE (INFRASTRUCTURE FOR SYSTEMS BIOLOGY EUROPE)

The SDBV group is part of the ISBE preparation project. The mission of the ISBE is to establish and enable access to an integrated, distributed infrastructure of state-of-the-art facilities for systems biology across Europe with a view to transforming our understanding of the life sciences, human health, and the environment. It will be composed of national Systems Biology Centres (nSBCs) across Europe that offer overlapping and complementary services at the national and European levels and thereby enhance the life sciences across Europe for society, industry, and research. Our group is part of the model and data management work package, which is responsible for surveying the state of the art and best practices for model and data management for Systems Biology, promoting a framework for it, and to collaborating with standardisation activities.

ISBE is in its preparatory phase with the following objectives:

- A proposal outlining the recommendations for the technical specifications of the physical infrastructure, required technologies and access policies.
- National workshops and roadshows involving users, providers, national funding bodies, and commercial stakeholders.
- The development and negotiation of a business plan and a sustainable funding model for interim and the 'steady state' phase of ISBE.

At the current time, funding for the establishment phase is still open. However, ISBE gives us the opportunity to shape the European systems biology data management architecture for the near future.

VLN (VIRTUAL LIVER NETWORK)

The Virtual Liver Network is a major national research initiative in systems biology funded by the German Federal Ministry for Education and Research. According to its mission statement, it aims at developing a dynamic mathematical model that represents (rather than fully replicates) human liver physiology, morphology, and function, integrating qualitative and quantitative data from all levels of organization, from sub-cellular levels to the whole organ. The main focus is on delivering a true multiscale representation of liver physiology that helps in understanding the dynamics of liver functions in normal and diseased states. The network is made up of 70 research groups distributed across Germany and involves about

250 researchers from both experimental and theoretical science. In 2010, the SDBV group implemented the VLN data management system (http://seek.virtual-liver.de/) as project and data hub that has been constantly improved upon and extended to include new features that provide better support for users (see section 'Software update SEEK' for recent changes and improvements in 2014).

FAIRDOM (FINDABLE, ACCESSIBLE, INTEROPERABLE, RE-USABLE DATA, OBJECTS, & MODELS)

FAIRDOM (http://fair-dom.org/) builds on the outcomes of the successful SysMO-DB and SyBIT data-management projects, uniting their tool and database development as well as their experience serving large systems biology projects. The main goals of the FAIRDOM project are (i) the development of the necessary toolset and the setup of a data and model management platform for EraSysApp projects (https://www.erasysapp.eu), (ii) merging the successful platforms SEEK and openBIS into openSEEK, and (iii) establishing a support and service network for European Systems Biology in standardising, managing, and disseminating data and models in a FAIR manner (Findable, Accessible, Interoperable, and Reusable).
Furthermore, FAIRDOM will document and disseminate the outcomes and activities to funding agencies, projects, and centres with the goal of establishing a sustainable business model for this infrastructure to enable its progressive exchange and use across the borders of projects and nations.
FAIRDOM is comprised of five partner institutions from four countries: the University of Manchester (UK), HITS (Germany), the University of Zurich (Switzerland), the ETH Zürich (Switzerland), and the University of Leiden (NL). It is intended to work closely together with the ISBE.

NMTRYPI (NEW MEDICINES FOR TRYPANOSOMATIDIC INFECTIONS)

The NMTrypI project (http://www.nmtrypi.eu) aims at obtaining new candidate drugs against Trypanosomatidic infections, i.e., sleeping sickness, leishmaniasis, and Chagas disease. The NMTrypI concept is based on the development of innovative drug leads, including a mechanism-based combination of a known and investigational drug and dual targets inhibition by using a common drug-discovery platform. NMTrypI is an EU-funded, three-year research project with experts from 14 partners in Europe and in disease-endemic countries (Italy, Greece, Portugal, Sudan, and Brazil). SDBV leads the data management within the project. A SEEK instance for NMTrypI was established and is used for the uploading, exchanging, and structuring of the heterogeneous research datasets from the project partners. Our special focus within NMTrypI SEEK is on the management of diverse chemical compound libraries. To facilitate this work, we developed new functionalities to search for compound details within data files, e.g., Excel files, which give a better overview with fewer clicks, as shown in the Fig. 2.9.1.

SBEPO (SYSTEMS BIOLOGY OF ERYTHROPOIETIN)

The SBEpo project (http://www.sbepo.de/) aims at the multi-level mathematical modeling of erythropoiesis for optimized expansion of erythroid progenitor cells and improved treatment regimes. The hormone Erythropoietin (Epo) is the key regulator of definitive erythropoiesis that is a unidirectional proliferation and maturation process, ensuring the renewal of red blood cells. Erythropoiesis represents one of the best-understood differentiation processes in the human body and thus serves as an ideal test case for data-based mathematical modeling to provide insights into mechanisms regulating cell fate decisions and to quantitatively predict targeted perturbations.
SDBV hosts the data management for the SBEpo project

using the SEEK data management platform (http://seek.sbepo.de/) and supports data workflow during the experimental phase in the laboratory using Excemplify.

NORMSYS - MODELLING STANDARDS IN SYSTEMS BIOLOGY

NormSys – 'Normalization and standardization for the exchange of models and data in systems biology research' is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi). The collaborative project with the University of Potsdam and LifeGlimmer GmbH in Berlin aims at enhancing and promoting the formal normalization of community standards for computational modelling in systems biology (http://www.normsys.org). NormSys brings different stakeholders together who all have an interest in further harmonizing and unifying the standardization of systems biology, including researchers from academia and industries as well as recognized official standardization institutions like the national German DIN (German Institute for Standardization: http://www.din.de), the European CEN (European Committee for Standardization: http://www.cen.eu), and the international ISO (International Organization for Standardization: http://

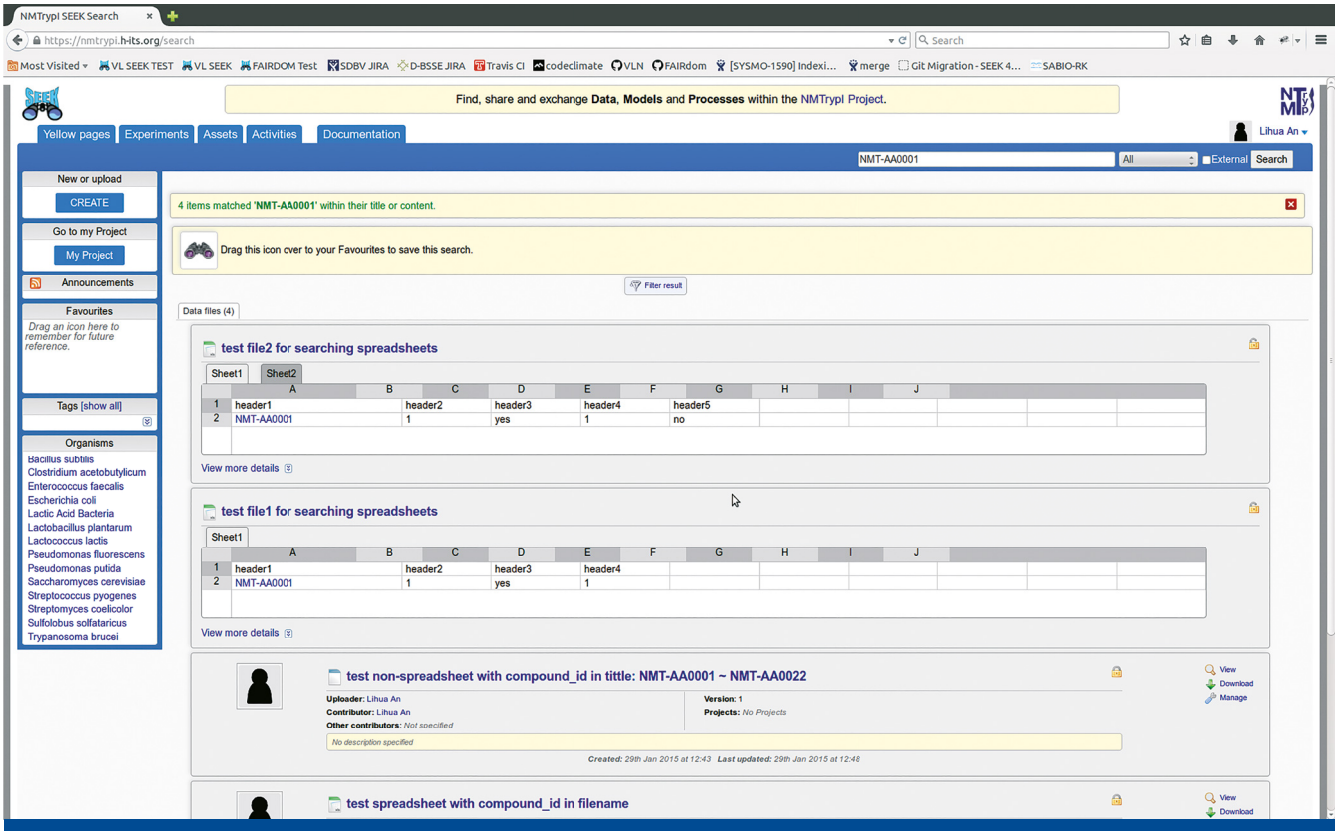


Fig. 36: NMTrypI SEEK screenshot for searched matched spreadsheets and non-spreadsheets.

www.iso.org). The aim is to build a bridge between the official national and international standardization bodies on the one hand, and existing community standardization efforts on the other hand, such as scientific grass-root standardization communities like the COMBINE network (Computational Modeling in Biology Network: http://www.co.mbine.org) and research infrastructure consortia like ISBE.

We are building a registry that will conjointly provide details about existing community standards (e.g., SBML - Systems Biology Markup Language, BioPAX - Biological Pathway Exchange, SBGN - Systems Biology Graphical Notation, SED-ML - Simulation Experiment Description Markup Language) with their key features and classify them according to their scope of application. This will help modellers to find and apply the community standard tailored to their requirements and intended use.

SEEK

In the past year, SEEK has seen an impressive list of new features. For example, faceted browsing and search is the core change in the method of exploring and navigating in SEEK. It provides an easy way to explore and navigate through a large amount of assets, both in asset indices and search results. It does this by applying multiple filters, called facets, enabling the assets to be accessed and ordered in multiple ways rather than in a single, pre-determined order.

Asset pipelining is an infrastructure change that boosts performance and improves caching. It means that all images, including avatars, javascript, and stylesheets are optimised and cached by the browser, thereby reducing the amount of information retrieved and how often it needs to be fetched. This reduces both the load on the server running SEEK and the network bandwidth required to interact with SEEK - particularly important in places with poor internet, such as conferences.

Spreadsheets are well supported by SEEK. Moreover, it has been continuously improved to meet new demands. Paging and better caching of spreadsheet exploring has been added to increase usability and support for large spreadsheets.

In order to provide more flexible parsing of templates as well as the transformation of spreadsheets, we have developed a Domain Specific Language to specify and execute parsers (DSL Parser). There are spreadsheet parsing and transformation operations that need to run over and over again, e.g., loading XLS, extracting columns, building a column model, applying transformation to cells, etc. The Parser DSL was developed to specify such parsers easily in the context of our work. DSL Parser provides flexibility in defining the input format as well as in mapping and output format. It allows for the creation of a flexible header format for source Excel files, transposition, entities to parse, and other parameters. Similarly, export functionality is flexible in content format and format, providing XLS and JSON files. As an added benefit, the DSL enables the tracking of provenance information, i.e., leaving a trace to indicate which input data lead to which output data. This enables verification of transformation outcomes, an important aspect of reproducibility. This very useful DSL has been stress tested in transforming data from several research projects into one common Excel-based format.

SABIO-RK

SABIO-RK (http://sabio.h-its.org/) is a manually curated database containing data about biochemical reactions and their kinetic properties. These data are mainly based on information reported in the scientific literature that are manually extracted and stored in a structured format. In addition, SABIO-RK also offers a direct automatic data upload from lab experiments.

As of December 2014, the database contains more than 50,300 different entries related to 4,861 publications from about 240 different journals. Kinetic data are available for more than 6,450 biochemical reactions catalysed by 1,471 enzymes (represented as EC numbers) in 823 dif-

ferent organisms and 303 tissues or cell types. The list of kinetic parameters comprises more than 35,500 velocity constants (e.g. Vmax, kcat, rate constants), about 40,900 Km or S_half values, and more than 10,700 inhibition constants (Ki and IC50).

SABIO-RK is cross-referenced from the external pathway database KEGG (Kyoto Encyclopedia of Genes and Genomes), the chemical compound database ChEBI (Chemical Entities of Biological Interest), and the UniProtKB (Universal Protein knowledgebase). Currently, 1,467 reactions in KEGG, 2,817 chemical compounds in ChEBI, and 3,507 database entries in UniProtKB refer to corresponding SABIO-RK database entries.

EXCEMPLIFY

Excemplify is a web-based application that was developed to support the exchange and long-time storage of experimental data, especially spreadsheets in general and immunoblot-related experiments in particular. It is able to parse these data from the initial experimental setup stage and to automatically generate the following spreadsheet stages in the experimental workflow. Apart from Excemplify's data storage capabilities, experimentalists are relieved of their burden of the time-consuming data-handling procedures and error-prone manual impositions. Excemplify was developed in close collaboration with the group for Systems Biology of Signal Transduction, which finally resulted in the release of a production version of Excemplify at the German Cancer Research Institute (DKFZ).

In 2014, an improved version of Excemplify was developed. This new version addressed the following major issues: I) speed acceleration, which was required due to the large amounts of data currently being stored in Excemplify, which caused speed problems for database queries that were not time critical in the initial stages of the project; II) the extraction of information from spreadsheets (parsing) for direct display and review on webpages; III) the extraction of information from spreadsheets (parsing) was also used to allow more fine-grained searches for documents; IV) search functionality in general was extended and made more user-friendly; V) the usability of the front end was highly improved in regard to navigation and survey; for that, new pages were added, allowing a more detailed view of the data; as an added benefit, this version allows the direct editing of spreadsheets in the browser window instead of consecutive document download-change-upload cycles, as before; VI) the security layer was extended; the users are now able to tag individual experimental results as visible for other users (read only), which allows for the sharing of documents and information within Excemplify; VII) the new version supports the direct upload of data to different configurable SEEK instances, currently including VLN-, SBEpo-SEEK, and LungSys-SEEK.

DISCOVER THE LIVER

With the help of VLN researchers and the HITS press department, we further collected, edited, and published the text and illustration content in discover the liver. Usability testing using the eye tracker helped us to improve the text writing for this site, whose purpose is to inform the interested public about the liver and the reasons that motivate scientific interest in it. Internal brainstorming meetings, regular user feedback rounds, and thorough analysis of the text and image collection process allowed us to refine the workflow of this process, to include additional editing steps, and to adapt the writing style to reader's different knowledge levels.

In addition to this distributed, collaborative editing effort, we applied changes to the visual design of the portal itself in order to adjust its visual aspect to the look and feel of the main site about the virtual liver (www.virtual-liver.de). This work was done in collaboration with Klug Newmedia, a media-centered SME from Berlin. This iterative process took some time, as the visual concept of the portal had to be changed quite drastically in order to adapt to the VLN

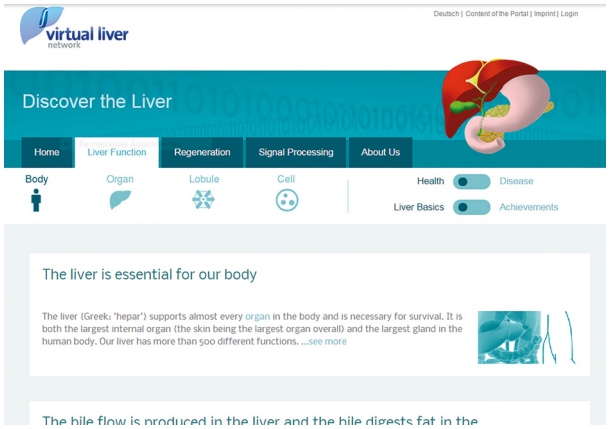


Fig. 37: Screenshot of the current discover the liver webpage.

visual identity, yet we wanted to keep the interaction style. Preliminary tests support that this iterative adaption was successful. Discover the liver is available in English- and German-language versions.

OPERATIONS EXPLORER

The Operations Explorer (a collaboration with Volker Stollorz, funded by the Robert Bosch foundation) enables the browsing of data (e.g., from the Statistisches Bundesamt) that describe the statistics of diagnoses and operations sorted by their ICD-10 and OPS-5 codes. The goal and motivation was to provide a tool that enables data journalists to easily use these data to find and investigate ideas for stories about the German health system. In 2014, we looked into how the Operations Explorer is used with the

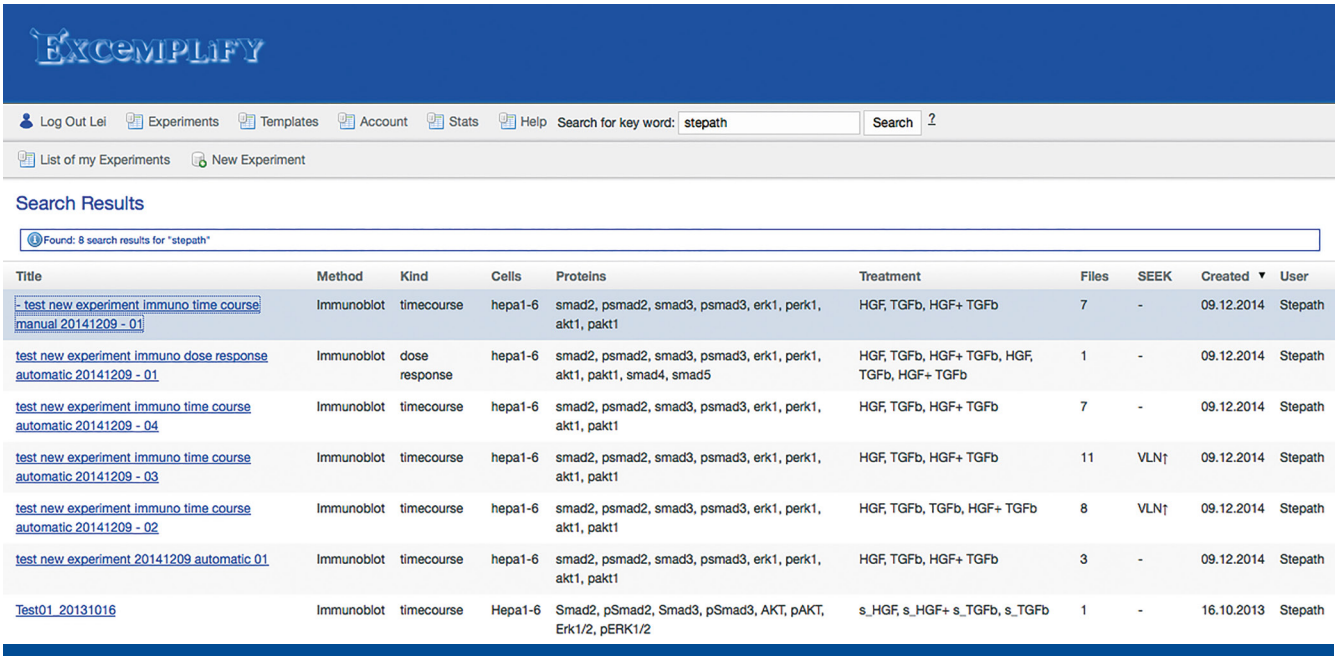


Fig. 38: Screenshot of the newly developed Excemplify version showing the main page of the graphical user interface.

goal of improving the interface. To this end, we formulated a set of questions that had to be answered by test users and tracked the eye movement of participants. Eye trackers provide insights in the way users use the system and recognizes which points the eyes concentrate on and how long they look at individual points. The duration is represented by a colour-code from green (short) to red (long).

Fig. 39 represents a user search for a specific button at the old interface, which shows that the user was searching almost everywhere for it to find the right place.
In contrast, Fig. 40 shows the gaze heatmap of a user who was asked to look for the button needed to share the map. It was found almost at once. Most areas of the user interfaces are only very shortly touched (green area) by the user's search for the correct button (red area).

FUTURE

The previous sections have demonstrated the infrastructure activities of the SDBV group. 2015 promises to be exciting, for it will bear witness to the start of both de.NBI and FAIRDOM as well as to the end of the preparative phase of ISBE. Furthermore, NormSys will intensify its activities in the standardisation domain, which is active in DIN, ISO, and VDI.
We are planning exciting new features for the existing tools within the existing projects, and we are currently setting up new research collaborations, some of which have already led to research proposals that are awaiting review. Furthermore, we are investigating joint doctoral theses with other groups inside and outside HITS.

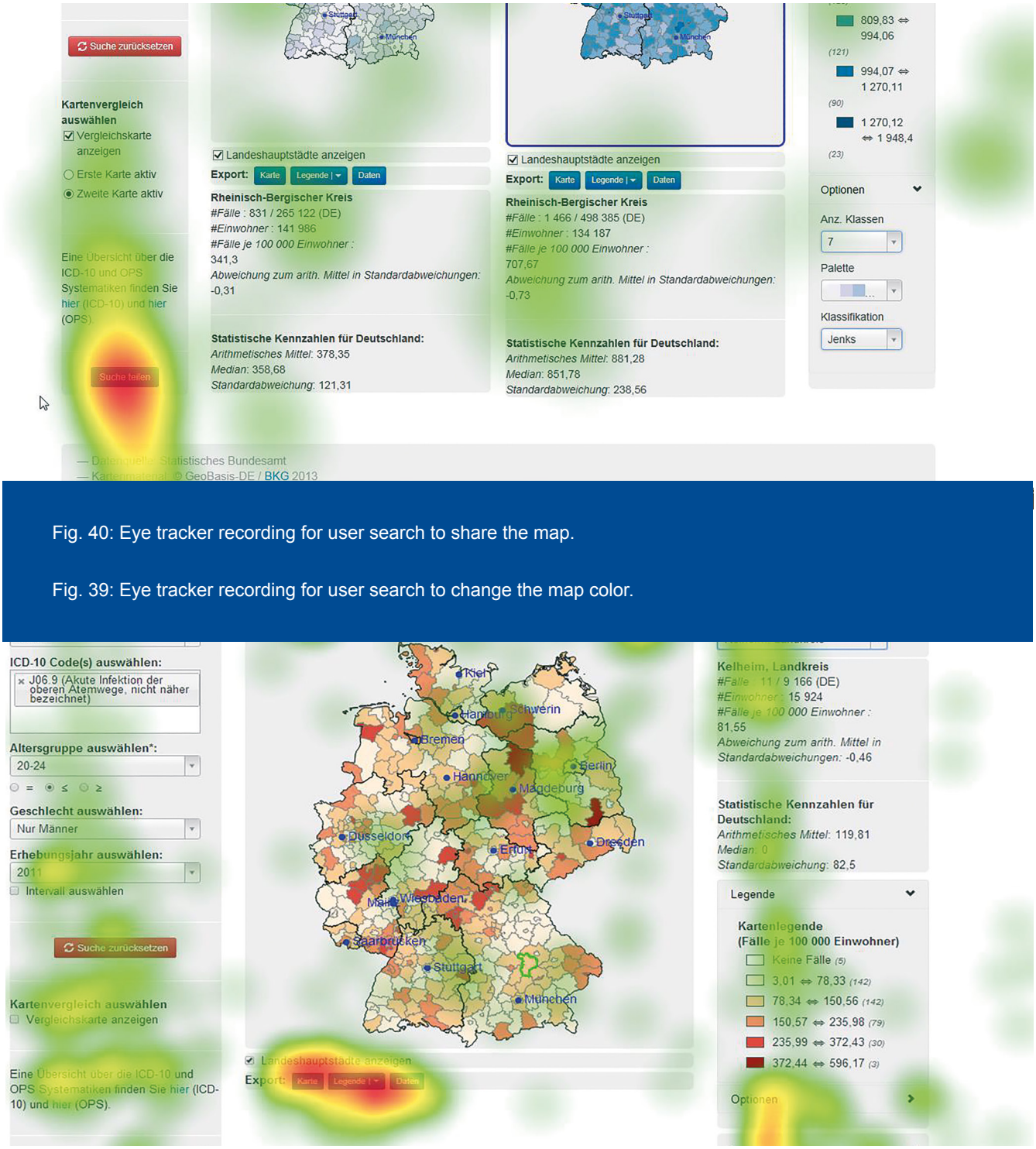


Fig. 40: Eye tracker recording for user search to share the map.

Fig. 39: Eye tracker recording for user search to change the map color.

2.10 Theoretical Astrophysics (TAP)

The Theoretical Astrophysics group at HITS seeks to understand the physics of cosmic structure formation over the last 13.5 billion years, from briefly after the Big Bang until today. We are especially interested in how galaxies form, ultimately producing magnificent systems like our own Galaxy, a busy metropolis of more than a hundred billion stars. We also aim to more accurately identify the properties of dark matter and dark energy, the two enigmatic matter and energy components that dominate today's universe and that are at the root of some of the most fundamental problems in modern physics.

A prominent role in our work is played by numerical simulations on a variety of scales, both of collisionless and hydrodynamic type. To this end, we develop novel numerical schemes that can be used efficiently on very large supercomputers, with the goal of exploiting them at their full capacity for linking the initial conditions of the universe with its complex evolved state today. The simulation models are indispensable for the interpretation of observational data and their comparison to theoretical models.

Using simulations, we are in particular able to study how diverse physical processes relevant in structure formation interact in a complex and highly non-linear fashion. A current priority in our group is the incorporation of physics into our models that has thus far been considered important but often been neglected, such as supermassive black hole formation, cosmic rays, and radiative transfer. In this report, we highlight a few results from our work of the past year in an exemplary fashion.

Die Theoretische Astrophysik Gruppe am HITS versucht die Physik der kosmischen Strukturentstehung während der letzten 13.5 Milliarden Jahre, vom Urknall bis heute, zu verstehen. Unser besonderes Interesse gilt der Entstehung von Galaxien, welche schließlich zur Bildung von spektakulären Systemen wie unserer Milchstraße führt, einer geschäftigen Metropole mit mehr als einhundert Milliarden Sternen. Wir arbeiten auch an einer Bestimmung der Eigenschaften der Dunklen Materie und der Dunklen Energie, jenen rätselhaften Komponenten, die den heutigen Kosmos dominieren und die zu den fundamentalsten Problemen der modernen Physik gehören.

Eine besonders wichtige Rolle in unserer Arbeit spielen numerische Simulationen auf verschiedenen Skalen. Zu diesem Zweck entwickeln wir neue numerische Verfahren, die effizient auf sehr großen Supercomputern eingesetzt werden können, mit dem Ziel, deren volle Kapazität für eine Verknüpfung der Anfangsbedingungen des Universums mit seinem heutigen komplexen Zustand auszunutzen. Die Simulationen sind für die Interpretation von Beobachtungen und deren Vergleich mit theoretischen Modellen unverzichtbar.

Mit der Hilfe von Simulationen sind wir insbesondere in der Lage, das komplexe und nichtlineare Zusammenspiel verschiedener physikalischer Prozesse zu studieren. Eine aktuelle Priorität in unsere Gruppe besteht darin, Physik in unsere Modelle einzubauen, die zwar als wichtig erachtet wird, die aber bisher vernachlässigt wurde, etwa superschwere Schwarze Löcher, kosmische Strahlen oder Strahlungstransport. In diesem Bericht stellen wir beispielhaft Ergebnisse unserer Arbeit im vergangenen Jahr vor.

THE ILLUSTRIS SIMULATION

Galaxies are comprised of up to several hundred billion stars and display a variety of shapes and sizes. Their formation involves a complicated blend of astrophysics, including gravitational, hydrodynamical, and radiative processes, as well as dynamics in the enigmatic “dark sector” of the universe, which is composed of dark matter and dark energy. Dark matter is thought to consist of a yet-unidentified elementary particle, making up about 85% of all matter, whereas dark energy opposes gravity and has induced an accelerated expansion of the universe in the recent past. Because the governing equations are too complicated to be solved analytically, numerical simulations have become a primary tool in theoretical studies of cosmic structure formation. The TAP group uses large parallel supercomputers to model representative pieces of the universe, aiming to predict how galaxies formed and how they cluster in space. Such calculations connect the comparatively simple initial state left behind by the Big Bang some 13.5 billion years ago with the complex, evolved state of the universe today.

A recent highlight of our group‘s work is the “Illustris Simulation” reported in Nature [Vogelsberger et al. 2014], currently the largest and most advanced hydrodynamical simulation of galaxy formation. Carried out together with colleagues at Harvard University, MIT, and the University of Cambridge, we have employed a refined treatment of galaxy formation physics implemented in our AREPO code. This method for cosmological hydrodynamics uses a finite-volume approach on a three-dimensional, fully dynamic Voronoi tessellation. The moving mesh is particularly well-suited to the high dynamic range in space and time posed by the galaxy formation problem. In addition to gravity in the dark matter and cosmic baryons, the processes modeled by the code also include radiative cooling, star formation and stellar evolution, energy feedback by supernova explosions and growing black holes,

The TAP group in 2014 (f.l.t.r.): Federico Marinacci, Christopher Hayward, Volker Springel, Christoph Pfrommer, Rüdiger Pakmor, Kevin Schaal, Denis Yurin, Andreas Bauer, Christine Simpson, Christian Arnold, Dandan Xu, Dominik Steinhauser, Juan Carlos Basto Pineda, Jolanta Krzyszkowska

Group Leader
Prof. Dr. Volker Springel

Postdocs
Dr. Robert Grand (from Oct. 2014)
Dr. Christopher Hayward (until Sept. 2014)
Dr. Federico Marinacci (until Oct. 2014)
Dr. Rüdiger Pakmor
PD Dr. Christoph Pfrommer
Dr. Christine Simpson
Dr. Dandan Xu

Graduate Students
Andreas Bauer
Kevin Schaal
Denis Yurin
Rainer Weinberger (from Dec. 2014)
Christian Arnold (from May 2014)

Visiting Scientists
Juan Carlos Basto Pineda (until March 2014)

Undergraduate Students
Jolanta Krzyszkowska
Svenja Jacob (from Oct. 2014)

as well as metal enrichment by stellar and galactic winds. The Illustris calculation was carried out in a periodic box that co-moves with the cosmic expansion, covering a region of about 350 million light years across. Fig. 41 shows an overview of the cosmic structures resolved by the simulation, such as the cosmic web of dark matter and diffuse gas, as well as stars in galaxies. The particular strength of the simulation is that it makes rich predictions for a diverse set of physical properties, including the X-ray emission of the intracluster gas, metal lines in the intergalactic medium, and the Sunyaev-Zeldovich effect, a kind of shadow cast by foreground galaxy clusters on the cosmic microwave background. Additionally, more exotic processes, such as the faint gamma-ray glow expected from dark matter annihilation, can be calculated with the simulation.

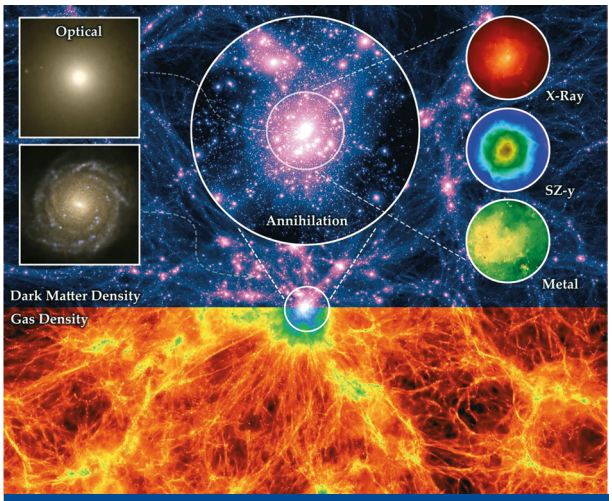


Fig. 41: Overview of different physical probes and spatial scales of the Illustris simulation, including the cosmic web detailed by the dark matter and gas density fields, the structure of individual galaxies, and the properties of the diffuse gas in the intracluster and intergroup medium [Vogelsberger et al. 2014b].

Most importantly, Illustris also makes very detailed predictions for galaxy formation. In fact, it was the first simulation that yielded a realistic mix of elliptical and spiral galaxies, thereby overcoming a decade-old impasse in the field (Fig. 42).

Moreover, we found that the simulation can explain the enrichment of heavy elements (collectively called metals in astronomy) in neutral hydrogen gas found in so-called damped Lyman-alpha absorbers. Furthermore, carefully prepared mock observations of the simulated universe show that the calculated galaxies are distributed in space as observed with telescopes such that deep mock images of galaxies created by Illustris show a stunning similarity to real observations, like the Hubble Ultradeep

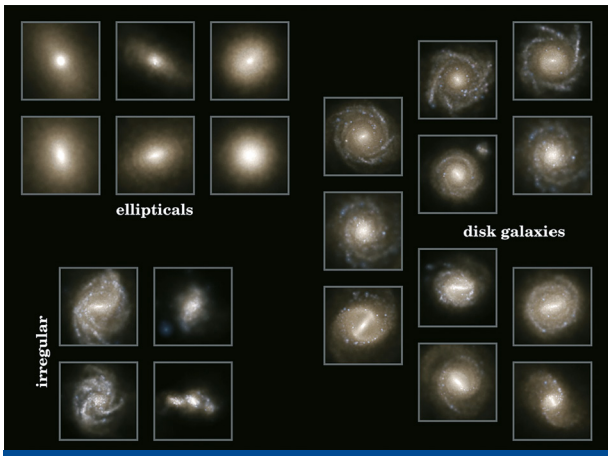


Fig. 42: Images of the simulated population of galaxies, which are arranged along the classical Hubble sequence (“tuning fork” diagram) for morphological classification. The Illustris simulation produces a variety of galaxy types, from elliptical and disk galaxies to irregular systems, which are mainly the result of mergers and interactions in galaxy clusters and mergers [Vogelsberger et al. 2014a].

Field. Quantitatively, this is reflected in a good match of the simulation's predictions for the abundance of galaxies at different epochs as a function of their stellar mass, as shown in Fig. 43. In order to achieve this success, energetic feedback processes from supernova and supermassive black holes are invoked in the physics model calculated by the simulation. These two powerful actors suppress star formation in small and large dark matter halos, respectively, producing a characteristic mass scale at which galaxy formation is most efficient (Fig. 44, page 88). This reconciles observations of the galaxy abundance with theoretical expectations for the abundance of dark matter halos. It thus appears that the standard model of cosmology is in principle capable of explaining galaxy formation despite our present ignorance of the true physical nature of dark matter and dark energy.

The Illustris simulation produced more than 200 terabytes of data and required the combined power of 8,192 processors for several months, using the supercomputers CURIE in France and SuperMUC in Germany. It employed more than 18 billion particles and cells and bridged a dynamic range of more than one million per space dimension. The total CPU cost of the main simulation was about 19 million hours, with a total memory requirement of over 25 TB RAM. The rich data set of Illustris now allows a variety of novel predictions as well as comprehensive tests of cosmological theories of galaxy formation. In particular,

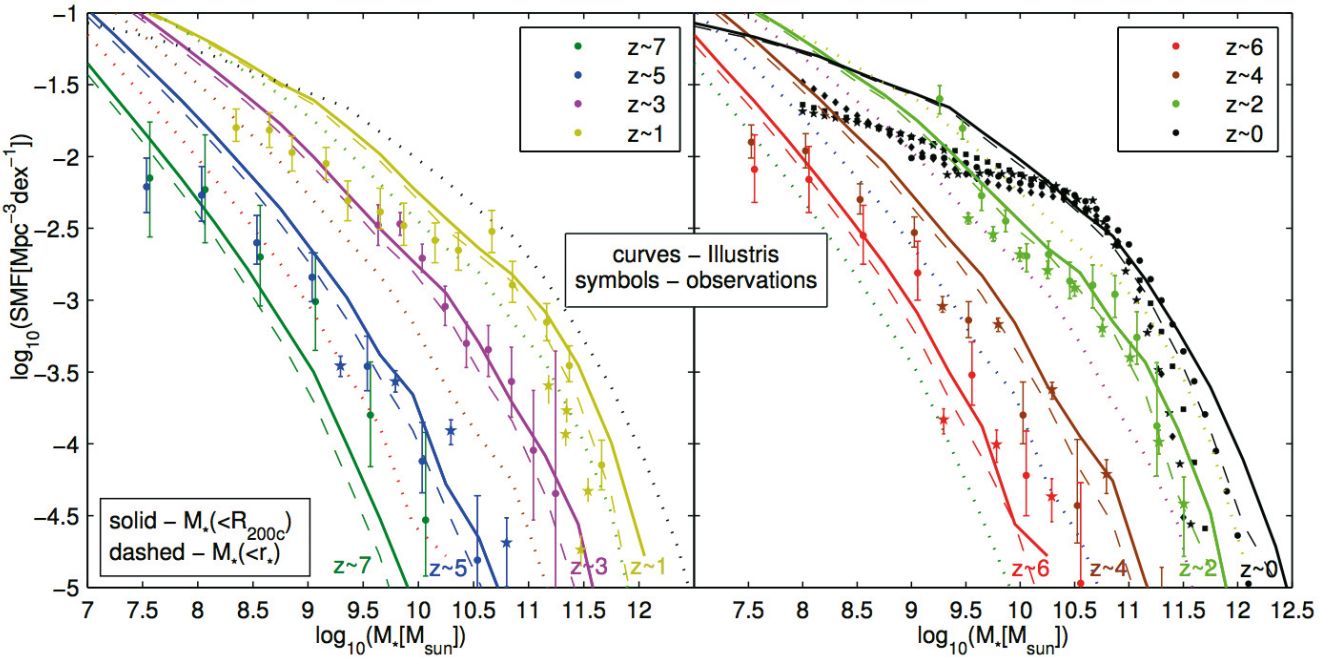


Fig. 43: Stellar mass function of galaxies in the Illustris simulation at different epochs. The comparison to observational data (symbols) shows that the galaxy population grows consistently with observations and matches the expected number density over a wide dynamic range [Genel et al. 2014].

it also describes how processes related to baryonic physics impact the cosmic web of dark matter that permeates the universe with a complicated tangle of nodes and filaments. Understanding this reliably is very important for the success of future observational missions that target the nature of dark energy, such as the European EUCLID satellite.

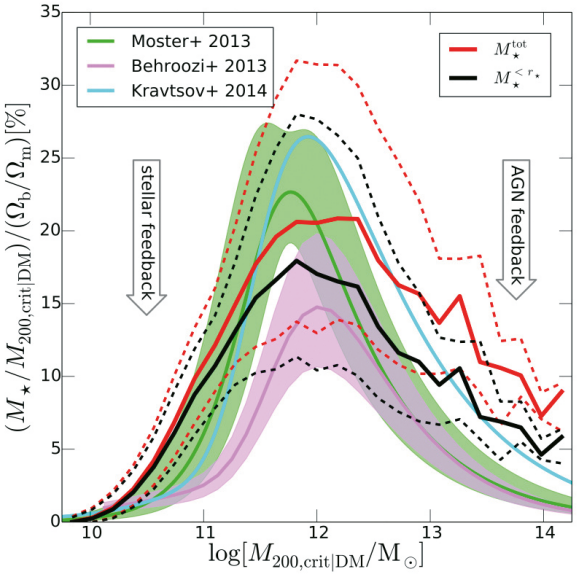


Fig. 44: Efficiency of star formation as a function of stellar mass. The physics of supernova and black holes jointly establish a halo mass scale in which star formation is most efficient. The star formation suppression at smaller and larger masses expected from empirical abundance matching models is roughly reproduced by Illustris [Vogelsberger et al. 2014b].

THE ORIGIN OF MAGNETIC FIELDS IN GALAXIES

Our universe is permeated with magnetic fields – they are found on Earth, in and around the Sun, as well as in our home galaxy, the Milky Way. Often the magnetic fields are quite weak: for example, the magnetic field on Earth is not strong enough to decisively influence the weather on our planet. In galaxies like the Milky Way, however, the field is so strong that its pressure on the interstellar gas in the galactic disc is of about the same size as the gas thermal pressure. Astrophysicists hence conclude that magnetic fields could play an important role there, although their origin still remains mysterious. Current hypotheses argue that these black holes either already existed directly after the Big Bang and were then greatly amplified with time, or that the field was produced by the first stars and was then dispersed into the Galaxy.

In principle, computer simulations that follow the formation and evolution of galaxies starting at the Big Bang ought to be able to answer these questions. However, until recently, they have mostly failed because the predicted galaxies did not agree with astronomical observations. In fact, the formation of disk galaxies has long been a puzzling conundrum for cosmologists: Computer simulations typically produced far too massive and too small disks, a problem that persisted for decades, and only quite recently could simulations such as Illustris make progress on it. Nevertheless, even Illustris has not been able to follow the dynamics of magnetic fields within the full cosmological context. Treating the latter is mathematically and numerically considerably more challenging than the plain gas dynamics, which has typically been employed for computing galaxies so far.

Within a project granted by the Gauss Centre for Supercomputing (GCS) on SuperMUC, we used our massively parallel simulation code AREPO to study galaxy formati-

on within a fully cosmological setting, employing a comprehensive treatment of the physics and much higher numerical resolution than had been used before, including the one available in Illustris. Thanks to AREPO's very low advection errors, it is particularly well-suited to the highly supersonic flows occurring in cosmology and to treating subsonic turbulence within the gas of virialized halos. These properties make it superior to smoothed particle hydrodynamics and adaptive mesh refinement codes that use a stationary Cartesian mesh. AREPO also follows the dynamics of dark matter with high accuracy, as is required to compute cosmic structure growth far into the nonlinear regime.

In our recent work, we have succeeded in including additional physical processes, such as magnetic fields, in the simulations, thereby improving them in a decisive way [Pakmor et al. 2014]. In fact, with the help of our novel numerical methods and the achieved progress in parallel scalability, we have been able to leverage the power of SuperMUC to form a virtual galaxy that closely resembles our own Milky Way (see Fig. 46, page 90). It has the right stellar mass for its dark matter halo mass of about 10^{12} solar masses, forms a disk of the right scale length, and the relation between the mean age of stars and the total stellar mass of the disk is also consistent with observations. Moreover, the predicted content of metals in the gas synthesized in stars and stellar explosions is consistent with observational data.

Importantly, we were for the first time able to predict the expected structure of the magnetic field in a spiral galaxy directly from the initial conditions left behind after the hot Big Bang (see Fig. 46). It turns out that an extremely tiny magnetic field left behind by the Big Bang is sufficient to explain today's observed field strengths, which are orders of magnitude larger. We were able to show that the magnetic field first grows exponentially for about 1 billion years due to gas motions in the early universe before the field reaches a stationary average value that is independent of its initial strength at the beginning. Once the first disk galaxies have formed (ca. 2.5 billion years after the Big Bang), the rotational motion of the disk further amplifies the field linearly with time, yielding field strengths at the micro-Gauss level. The revolving flow of the gas

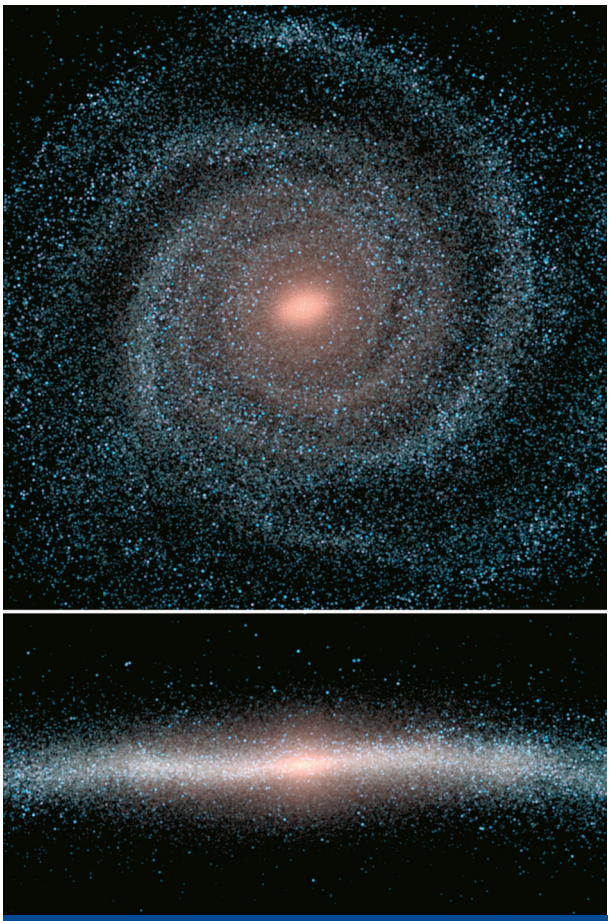


Fig. 45: Stellar structure of a Milky Way-like galaxy formed in one of our magneto-hydrodynamical cosmological simulations after 13 billion years of evolution. The face-on projection on top nicely reveals spiral structure in the disk [Pakmor et al. 2014].

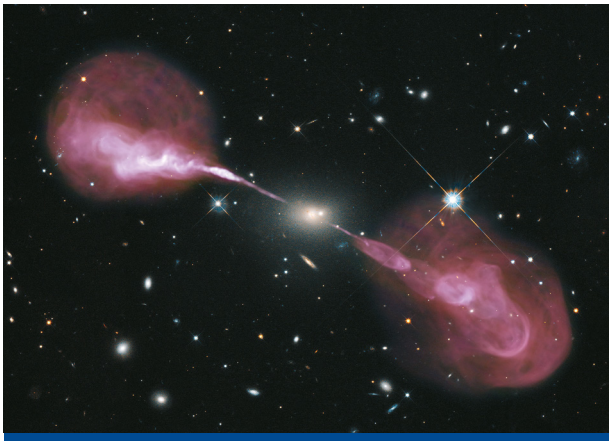
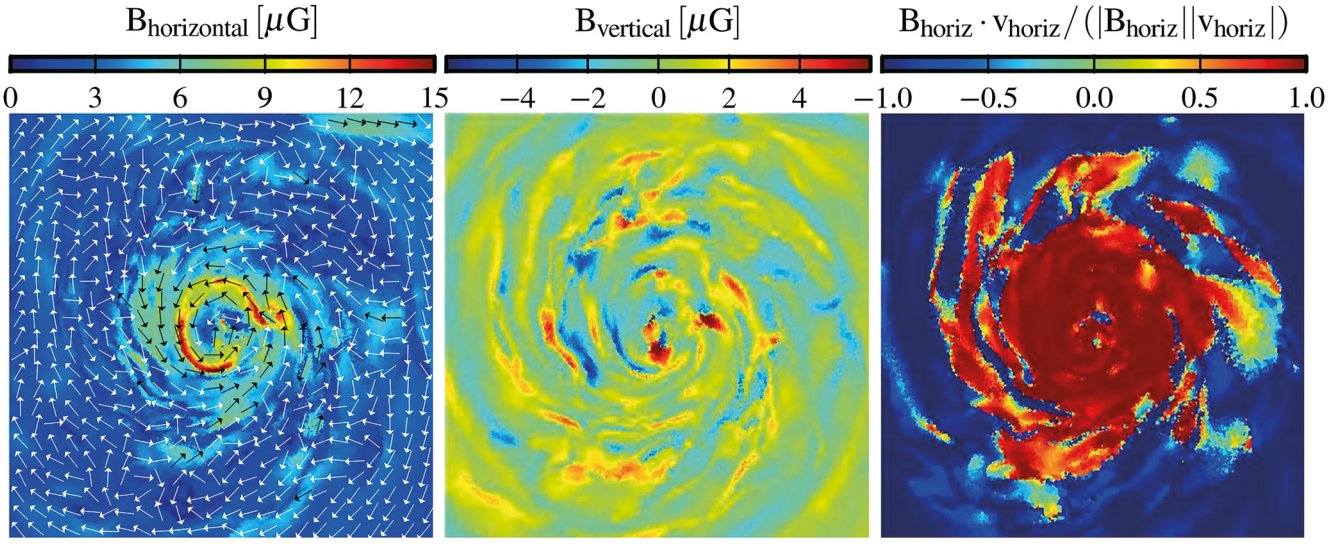


Fig. 47: Radio jets around the supermassive black hole in the center of the elliptical galaxy Hercules A, as observed with the Hubble Space Telescope (credit: NASA/ESA).

in the disk also pulls the magnetic field lines and directs them tangentially along this motion. Interestingly, the magnetic field strength found in the simulation does not only agree very well with the values measured for the Milky Way and neighboring galaxies, but it also reproduces the observed vertical and horizontal profiles. This is remarkable given that there are no free parameters that could be tuned to influence the final field strength. The successful formation of disk galaxies with a small bulge-to-disk ratio constitutes a long-sought advance in the intricate problem of the formation of galaxies in hydrodynamic cosmological simulations. It is fascinating that this can at the same time explain the formation of typical magnetic fields found in galaxies like the Milky Way. These findings also promise to help in understanding the deflection of cosmic ray particles in the magnetic field of the Milky Way and in providing clues for tracking down the sources of these particles, which is still an unsolved problem in observational astronomy.

Fig. 46: Structure of the magnetic field in one of our simulated Milky Way-sized galaxies. The left and center panels show the strength of different components of the magnetic field. The right panel shows the pitch angle, illustrating that the field develops a large-scale orientation aligned with the azimuthal rotation [Pakmor et al. 2014].



THE PHYSICS AND COSMOLOGY OF BLAZARS

Black holes are among the most fascinating objects in the universe. They span a large range of masses, from about a solar mass up to 10 billion solar masses, which constitute the class of super-massive black holes. Those "active galactic nuclei" are situated at the center of every galaxy and are able to drive powerful relativistic jets and electromagnetic radiation out to cosmological distances.

The jets are ejected back-to-back and leave the accretion region of the black hole along the rotation axis of the spinning black hole (see Fig. 47). In the unified model of active galactic nuclei, the geometrical orientation of these jets with respect to our line of sight determines their observational appearance: If the jets are lying in the plane of the sky, they constitute the population of radio galaxies, whereas systems in which the jets are pointing directly at us have been dubbed blazars. These blazars dominate the source population of the extragalactic gamma-ray sky by far. They are visible as gamma-ray point sources due to the limited resolution of current gamma-ray telescopes. If one masks out the brightest gamma-ray point sources (as well as the Galactic foreground), there is a diffusely glowing background remaining, the so-called extragalactic gamma-ray background, which is characterized by a unique spectral shape (see data points of Fig. 48).

There have been many papers written about the possible physical origin of this gamma-ray background, ranging from active galactic nuclei, starburst galaxies, and large-scale structure formation shocks, up to more exotic suggestions, such as the decay or annihilation of dark matter. The simplest explanation would be that the most abundant resolved gamma-ray sources (namely blazars) also give rise to the gamma-ray background through the contribution of many faint and unresolved blazars. However, there was a major obstacle to this solution: If blazars

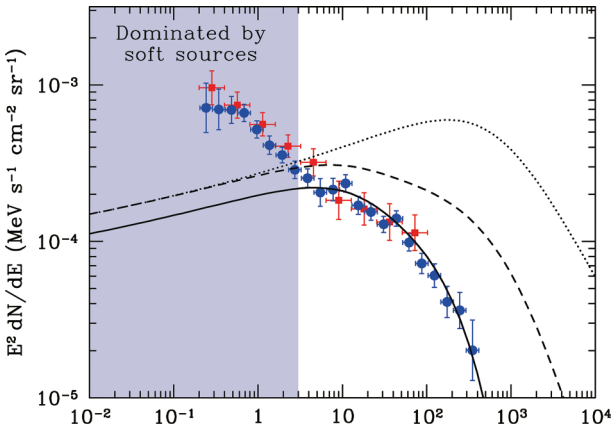


Fig. 48: Isotropic extragalactic gamma-ray background anticipated by the hard gamma-ray blazars. The dotted, dashed, and solid lines correspond to the unabsorbed spectrum, the spectrum corrected for absorption on the extragalactic background light, and the spectrum corrected for resolved point sources (assuming all nearby hard gamma-ray blazars are resolved). These are compared with the measured extragalactic gamma-ray background reported by the Fermi Collaboration (red squares: 2010; blue circles: 2014). Note that the EGRB is dominated by soft sources below 3 GeV [Broderick et al. 2014].

follow the strong time evolution found for other active galactic nuclei (of which they are supposed to be a subclass), then the total energy emitted in the gamma-ray band would vastly overproduce the detected background. Thus, either the unified model for active galactic nuclei is wrong, or there must be additional physics at play in the gamma-ray band – two very disturbing conclusions that hold the promise for exciting discoveries!

In the TAP group, we have been looking at the second possibility and have dissected the physics associated

with the propagation of tera-electron-volt (TeV) gamma rays emitted by hard-spectrum blazars. The universe is opaque to the emitted TeV gamma rays because they annihilate and pair produce on the extragalactic background light, which is provided by all galaxies and quasars that have ever existed in the universe. The resulting ultra-relativistic pairs of electrons and positrons are commonly assumed to lose energy primarily through scattering processes with photons of the cosmic microwave background (the relic radiation from the Big Bang). This would allow the original emission to cascade down by a factor of 1,000 to giga-electron-volt (GeV) energies. The previously mentioned problem of overproduction of the gamma-ray background from an evolving blazar population could be immediately solved if the kinetic energy of the ultra-relativistic pairs were not reprocessed into the GeV gamma-ray band, but rather channeled into some other form of energy. This is demonstrated by the solid line in Fig. 48, which shows the contribution of unresolved hard blazars that exhibit the same time evolution as other active galactic nuclei but for which the gamma-ray cascading process is switched off. Moreover, in the same paper, we demonstrated that the same blazar model provides an astonishing match to the luminosity distribution and time evolution of all resolved sources by the Fermi gamma-ray space telescope. This provides strong empirical motivation to search for the detailed physics of such a mechanism.

The difficulty was not only to identify a process that does this, but one that does this faster than the competing scattering process with photons of the cosmic microwave background. Two years ago, we identified such a process: Powerful plasma instabilities driven by the highly anisotropic nature of the ultra-relativistic pair distribution provide a plausible way to dissipate the kinetic energy of the TeV-generated pairs locally, heating the intergalactic medium. At that time, we had shown that the linear growth rate of the instability is indeed fast enough be the dominant mechanism. Since then, a number of works that have been inspired by our novel approach have further developed and complemented it with interesting insights.

However, one aspect has remained controversial: whether the nonlinear interaction of two coherent (high-frequency) waves driven by the instability would produce a beat wave that drains energy at a faster rate than the rate at which the unstable wave grows. If true, this would have rendered the proposed mechanism physically irrelevant by reducing the effective damping rate to a low level. Using numerical calculations, we could instead show that the effective damping rate of our plasma instability is fast enough to be the dominant cooling mechanism of these pair beams [Broderick et al. 2014]. In particular, we realized that previous estimates of this rate assumed that the damping of scattered waves entirely depends on collisions, ignoring faster collisionless processes that are ubiquitous. We found that the total wave energy eventually grows to approximate equipartition with the beam by increasingly depositing energy into long wavelength modes. While this discovery provides a major step forward for the understanding of fundamental plasma physics in a very dilute regime (which is inaccessible to laboratory experiments), there is nevertheless a lot of work ahead to fully understand the physics of the non-linear regime of the instability. In any case, our work provides some tantalizing physical understanding as to how accreting super-massive black holes impact not only upon their immediate environment, but also upon the universe on very large scales.

The activities of the institute's administration in 2014 were primarily determined by two requirements: the consolidation of processes after the institute had expanded significantly by four research groups in the previous year and the preparation of organizational changes resulting from the institute's new constitution.

The first activity focused on transferring the externally funded projects, which the newly appointed group leaders had acquired at their present facilities, to our institute. In some cases – for example, with DFG projects – this was possible without problems. For other funding streams, such as ERC grants, the transfer process took several months, but all projects and staff could ultimately be transferred to HITS.

The revised charter of the institute provides for a dual leadership consisting of a managing director and a scientific director. This has implications for the majority of the administrative processes, and the ultimate goal was to not affect the quick execution of all operations that have been so positively commented on by the staff. It remains to be seen in 2015 whether the specific provisions will satisfy this requirement.

A side effect of this revision of all processes is re-employment with the question of how the various processes can be handled more effectively with IT-based tools – if they have not already been. We investigated this problem in previous years from the perspective of project planning and controlling, and while we did not find a practical solution, we came to a very detailed understanding of the difficulties involved. These difficulties have to do with the very different accounting rules of the funding providers (which also continue to change with every new framework program) on the one hand and with the need for the coupling of such a tool with existing IT systems, such as the SAP R/3 that we use, on the other hand.

Since a comprehensive project accounting system will probably not be available soon and in-house development will not be possible due to capacity issues, we will first focus on other aspects of the administrative work and hope to reach useful results in this area more quickly.

The Administration group in 2014 (f.l.t.r.): Ingrid Kräling, Rebekka Riehl, Andreas Reuter, Christina Bölk-Krosta, Christina Blach, Stefanie Szymorek

Group Leader

Prof. Dr. Andreas Reuter (acting)

Group Members

Christina Blach I office

Christina Bölk-Krosta I controlling

Benedicta Frech I office

Ingrid Kräling I controlling

Kerstin Nicolai I controlling

Rebekka Riehl I human resources and assistant to managing director

Stefanie Szymorek I human resources

3.2 IT Infrastructure and Network

Although the IT Services group was only formed in the summer of 2013, it underwent a big change in 2014 with the introduction of scientific and web software development services. The reason for this change is a well-known problem in academic institutions: PhD students or postdocs-level researchers develop scientific software or web applications but go on to the next step in their careers after a short time (usually one to three years), taking with them the knowledge and leaving behind applications with little or no documentation, obscure code, or missing features. Further development, fixing bugs, or simply maintaining functionality become difficult or even impossible tasks, such that an application is often rewritten from scratch by the next PhD student or researcher... and the cycle repeats. With the addition of two professional software developers to the ITS group, we hope to minimize such effects for the scientific groups at HITS. Not only is the continued support guaranteed for the whole lifetime of the applications, but the professional developers can also share their knowledge with the less-experienced programmers at HITS, either directly while working together or through the organization of internal workshops. As the developers become gradually involved in more and more software projects, they can also act as a connecting bridge between the scientific groups, revealing common patterns and functionality. This could lead to further interdisciplinary projects and less code duplication.

In connection with the newly introduced software development activities, we have started offering platforms for centralized software project management (based on Redmine, Subversion and git) and for continuous integration (based on Jenkins). These platforms have replaced individual solutions installed over the years by the scientific groups, thereby reducing learning time and maintenance overhead. They also allow for a much easier collaboration both between the HITS groups as part of interdisciplinary projects as well as with outside researchers.

To keep up with the increase in demand from the scientific groups, the HPC resources have been further expanded. The cluster has grown with the addition of 1024 cores based on the Intel Ivy Bridge architecture, with Infiniband used for inter-node communication. Two special cluster nodes have also been added for applications that do not fit well with the distributed computation and memory model imposed by MPI: one with 1TB RAM and 2 nVidia Kepler GPUs, and one with 4TB RAM and 120 CPU cores. These nodes will be used primarily for machine learning, data mining, and bioinformatics. The high performance storage has also seen a significant capacity increase and now reaches more than 1.6PB in BeeGFS (formerly known as FhGHS) volumes. For automatic node deployment and cluster management, we are exploring a new solution called Hive (http://www.gemmantics.com/software/hive/), which promises a simplification and a higher scalability of the usual maintenance tasks, leading to an increased availability of the cluster resources.

Towards the end of 2014, HITS also became part of the "eduroam" community (https://www.eduroam.org/), providing free roaming network access for students, researchers, and staff of the member institutions. This allows HITSters visiting other academic institutions to readily get network access, eliminating the need for registration procedures or password management. In turn, HITS also provides the guest scientists with easy access to WiFi - at the moment, primarily in the seminar rooms, with plans underway for expansion to the whole building. To further support the mobile access trend and allow HITSters remote connections to internal computation and storage resources, we have installed a new Virtual Private Network (VPN) solution with high performance and high availability features.

Although less visible but critically important for our continued services, the IT infrastructure has also witnessed some significant additions. We have introduced a monitoring system that provides a valuable overview of the wide landscape of IT devices and services at HITS, as well as notifications for the prediction and indication of failures. For the installation and maintenance of Linux servers, we have introduced a central configuration management

system and are working toward automatic provisioning of servers and services. The Active Directory setup has begun to take a more central role in all IT services offered at HITS and was supplemented with additional servers in order to achieve a very fast and highly redundant user identification and authentication environment. We have also migrated the e-mail accounts to new servers running the latest version of Microsoft Exchange in order to benefit from an increase in storage capacity and better compatibility with client software. The backup system has also undergone a partial reorganization, an upgrade of the central backup server, and a further capacity increase. Moreover, we have been heavily involved in the extensive preparations for the re-launch of the HITS website, planned for the very beginning of 2015. One of the major goals was switching to a new Content Management System (CMS) to support a dynamic website that perfectly reflects the high pace of research happening at HITS. Although WordPress is a CMS renowned for its easy deployment and configuration, the integration of a large number of plugins, an external theme, and security- and performance-related settings has been a complex task requiring countless hours of prototyping, configuration changes, and testing. All this effort should have yielded a highly functional, modern-looking, secure, and fast website by the time you are reading these sentences.

Several events in 2014 have had a noticeable impact on the whole IT industry and have also required special preparations or a swift response at HITS. The end of support for Microsoft Windows XP and Office 2003, very popular products among scientists, has demanded a careful planning of upgrades to ensure that all HITS computers, particularly the mobile ones, are running software versions which continue to receive security-relevant patches. The discovery of a serious vulnerability in the popular OpenSSL cryptographic software library (the so-called Heartbleed bug (http://heartbleed.com/)) has called for an immediate installation of updates and reissued certificates for all servers and the institute firewall. A very quick response was also required upon the further discovery of weaknesses in the SSLv3 protocol, often used for securing connections between computers (the so-called POODLE bug (https://www.openssl.org/~bodo/ssl-poodle.pdf)).

In addition to all these activities, we have tried to remain responsive to users' questions and requests for help. Although we haven't always succeeded, we hope to do an even better job in 2015!

The ITS group in 2014 (f.l.t.r.): Norbert Rabes, Christian Goll, Bernd Doser, Andreas Ulrich, Bogdan Costescu, Nils Wötzel

Group Leader
Dr. Ion Bogdan Costescu

Staff Members
Dr. Bernd Doser I Software Developer
(from August 2014)
Dr. Christian Goll I System Administrator
Norbert Rabes I System Administrator
Andreas Ulrich I System Administrator
Dr. Nils Wötzel I Software Developer
(from April 2014)

Student
Philipp Grüner

In the year 2014, the HITS communications team continued its manifold activities to enhance the institute's outreach activities and to help shape the name "HITS" as a brand for a small but excellent research science institute with an international, interdisciplinary, and inspiring atmosphere.

The first and basic prerequisite for our work is scientific excellence. Fortunately, in 2014, the HITSters were able to demonstrate that this prerequisite is being met. For instance, Volker Springel (TAP) co-authored the "Illustris" simulation, the most detailed numerical simulation of the Universe, published in Nature in May 2014. The worldwide media response was almost overwhelming. Alexandros Stamatakis (SCO) and several of his team members were the only German authors in two international phylogenetic studies on insects and birds. Both articles hit the front page of the "Science" magazine in November and December, causing widespread media resonance.

Finally, according to the "Highly Cited Researchers" report by the Thomson Reuters Group, Volker Springel and Tilmann Gneiting (CST) belong to the group of highly cited researchers worldwide. The two HITSters rank among the scientists most cited for their subject field and year of publication, which is an important indicator for the scientific impact of a publication. Of the 3,125 most-cited scientists worldwide, only 167 have a primary or secondary affiliation with a German institution.

Last but not least, in the 2014 ranking of the Humboldt Foundation, HITS was listed among the German research institutions and cities that are leading in the international contest of the best minds. In the last five years, we have welcomed a total of five scholars – a remarkable figure given the size of the institute. As a result, in 2014, it was a true pleasure to be able to report these excellent findings and HITSter rankings to the public.

Another prerequisite for successful communication is the development of reliable and sustainable journalistic contacts. An important project for HITS is the "Journalist in Residence" program. It is addressed to science journalists and offers them a paid sojourn at HITS. During their stay, they can learn more about data-driven science and get to know researchers and research topics in more detail and without pressure from the "daily grind".

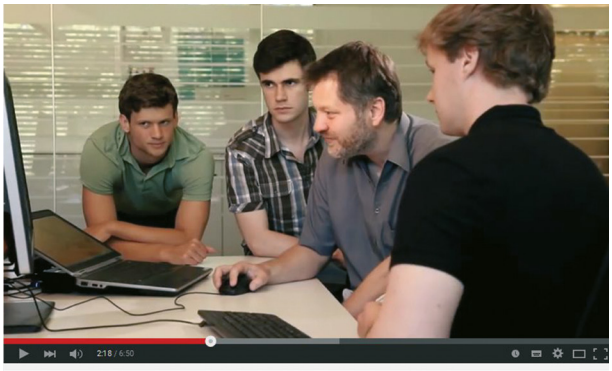


Fig. 49: In August 2014 our Math@HITS Video went online on the HITS Youtube-Channel TheHITSters.

On 15 August 2014, Barcelona-based science journalist Michele Catanzaro began his stay as third and first international "Journalist in Residence". As a physicist and a network specialist by training, he was keenly interested in the methods and tools deployed at the institute. During his stay, Michele managed to interview all group leaders and spent some time deepening his knowledge of the increasing amount of data in science. Moreover, he held a public lecture on research in the media from an international, non-naïve point of view ("Put more journalism into science journalism"), and he organized an internal seminar with the HITSters on science journalism. Michele Catanzaro participated in several conferences, such as the Heidelberg Laureate Forum (see Chapter 5.5), where he was one of the bloggers reporting on the event.

In November, he went to the "Wissenswerte" conference of science journalists in Magdeburg. He was only supposed to meet his German colleagues, but suddenly found himself as a session panelist on the topic of international

journalism as an ad hoc stand-in for a colleague who had missed the conference on short notice. Overall, Michele's stay at HITS proved to be valuable for the credibility of the "Journalist in Residence" program in the community of journalists, both nationally and internationally.

An indispensable prerequisite for public relations work is up-to-date information. In 2014, we produced a new HITS flyer in a modular design, taking into account that the amount of research groups was still growing. To demonstrate the mathematical focus at the institute, a movie called "Math@HITS" was produced, depicting data-driven research from a mathematical view. A short version of the movie was shown at the International Congress of Mathematicians (ICM) in Seoul, Korea (see CD).
In addition to these overseas outreach activities, HITSters were eagerly active in the Heidelberg region and in Germany, beginning at the end of June with a booth and two talks at the "Informatiktag" ("Compute Science Day") at Heidelberg University, followed by "Explore Science" in July with three hands-on stations and a talk by Alexandros Stamatakis on "living diversity" (see Chapter 5.3). In August, HITS was "on board" of the "MS Wissenschaft", a floating science center that lay at anchor in Mannheim. We presented the opportunities of

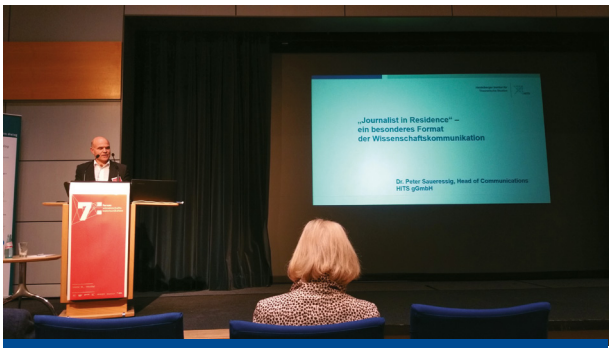


Fig. 50: Peter Saueressig presenting the Journalist in Residence Program at the "Forum Wissenschaftskommunikation" 2014 in Potsdam.

The HITS Communications team in 2014 (f.l.t.r.): Peter Saueressig, Isabel Hartmann, Elisa Behr

Head of Communications
Dr. Peter Saueressig

Members
Isabel Hartmann I Public Relations
Juliane Repp BA I student (until August 2014)
Elisa Behr BA I student (from October 2014)

Communications

data-driven research and organized a panel discussion with Theresia Bauer, the Minister of Science for the state of Baden-Württemberg (see Chapter 5.4). In September, HITSters hosted a group of young researchers who participated at the Heidelberg Laureate Forum (see Chapter 5.5). Finally, in December, Peter Saueressig gave a short talk at the "Forum Wissenschaftskommunikation" ("Science Communication Forum") in Potsdam, setting the "Journalist in Residence" program in the context of science communication.

5 Events

5.1 Conferences, Workshops & Courses

5.1.1 EMBO PRACTICAL COURSE COMPUTATIONAL MOLECULAR EVOLUTION

5 – 14 May 2014, Heraklion/Greece

The need for an effective and informed analysis of biological sequence data is increasing with the explosive growth of biological sequence databases. A molecular evolutionary framework is central to many Bioinformatics approaches used in these analyses, e.g., de novo gene finding from genomic sequences.

Additionally, the explicit use of molecular evolutionary and phylogenetic analyses provides important insights into its own right, such as the analysis of adaptive evolution in viruses, which provides clues to the viruses' interaction with host immune systems.

The EMBO Practical Course took place for the 6th time at the Hellenic Institute of Marine Research near Heraklion Crete and provided graduate and postgraduate researchers with the theoretical knowledge and practical skills to carry out molecular evolutionary analyses on sequence data. It entailed data retrieval and assembly, alignment techniques, phylogeny reconstruction, hypothesis testing, and population genetic approaches. The course covered sessions on analysis of both protein and nucleotide sequences, including NGS data.

The course offered the opportunity for direct interaction with some of the world-leading scientists and authors of famous analysis tools in evolutionary bioinformatics, such as John Huelsebeck, Olivier Gascuel, Nick Goldman, Bruce Rannala, Alexandros Stamantakis, and Ziheng Yang.

Alexandros Stamatakis (SCO group leader) was the principal organizer of this event. The committee received 200 applications for the 35 available places. Former SCO postdoc Pavlos Pavlidis, visiting PhD student Paschalia Kapli, as well as Andre Aberer also participated in this early "summer school" as teaching assistants. HITS was among the sponsors of this course.

5.1.2 HARVARD-HEIDELBERG WORKSHOP, HEIDELBERG

23 – 26 June 2014, Heidelberg/Germany

Fig. 51: Group picture of the Harvard-Heidelberg Workshop. (Picture: Haus der Astronomie)

The 2014 summer meeting "HHSF14: Star Formation: Data, Models, and Visualization. A Harvard-Heidelberg Workshop" at the "Haus der Astronomie" ("House of Astronomy") in Heidelberg from 23 – 26 June 2014 was the first joint science conference between Heidelberg and Harvard. The conference gathered experts in star formation and visualization from Harvard University (including the Harvard-Smithsonian Center for Astrophysics), the Max Planck Institute for Astronomy (MPIA), Heidelberg University, the Heidelberg Institute for Theoretical Studies (HITS), and the Max Planck Institute for Extraterrestrial Physics (MPE).

The focus of the workshop was on star formation within our Galaxy, as well as visualization tools and techniques that have or can be used in studies of star formation. The scientific scope of the workshop covered studies of star

formation from the smallest scales to Milky-Way-scale star formation, and from the initial conditions for star formation to the end of the main accretion phase.
In the framework of the program, Theoretical Astrophysics (TAP) group leader Volker Springel gave a talk about his work at HITS, where he and is group focus on simulating galaxies. Kai Polsterer, group leader of the Astroinformatics (AIN) group who is working on processing the large data in astronomy with his group, gave an insight into "Machine Learning in Astronomy: Why the Data Deluge is not Just Pain".

5.1.3 PDE SOFT, HEIDELBERG

July 14-18, 2014, Heidelberg/Germany

Simulation software for complex phenomena based on models involving partial differential equations (PDE) has become an important topic in modern research from mathematics and scientific computing and provides methods and algorithms for application fields that utilize codes and define requirements on their technical abilities. PDESoft conferences provide a discussion venue for developers and users of multi-purpose libraries for PDE simulation as well as for researchers who study the implementation of computer algorithms for PDE. The conferences serve as a focus point where the state-of-the-art implementation of state-of-the-art algorithms are discussed. They give developers a chance to share ideas with other developers and users. Additionally, users of such software have an opportunity to learn about the development principles of software and to contribute their knowledge from the point of view of application.
After the first conference in Münster in 2012, the 2nd conference took place in summer 2014 in Heidelberg. The conference was jointly hosted by HITS and Heidelberg University's Interdisciplinary Center for Scientific Computing (IWR). HITS Managing Director Prof. Andreas Reuter and DMQ group leader Prof. Vincent Heuveline were part of the local conference committee. The presentations and discussions took place at the Studio Villa Bosch, followed by coding days (July 17/18) at the IWR – two days of joint coding.

Fig. 52: Group picture of the PDE soft conference in Heidelberg 2014.

5.1.1 7TH EMBO PRACTICAL COURSE ON BIOMOLECULAR SIMULATION

20 – 27 July 2014, Pasteur Institute, Paris/France

Organizers: Michael Nilges (Pasteur Institute, Paris) and Rebecca Wade (HITS, Heidelberg)

Course faculty: Arnaud Blondel (Pasteur Institute, Paris, France), Monika Fuxreiter (Univ. Debrecen, Hungary), Francesco Gervasio (UCL, London, UK), Konrad Hinsen (CEA, Saclay, France), Tru Huynh (Pasteur Institute, Paris, France), Richard Lavery (CNRS, Lyon, France), Katrina Lexa (UCSF, USA), Therese Malliavin (Pasteur Institute, Paris, France), Michael Nilges (Pasteur Institute, Paris,

France), Tom Simonson (Ecole Polytechnique, Palaiseau, Paris, France), Anna Tramontano (Univ. Rome, Italy), Rebecca Wade (HITS, Heidelberg, Germany), Christopher Woods (Bristol University, UK), Willy Wriggers (DE Shaw Research, USA)
Website: http://events.embo.org/14-simulation/

Molecular simulation techniques constitute an important component of the biologist's toolbox. The function of biological macromolecules is determined by their three-dimensional structure and dynamics. Even with structural genomics projects, the structure-sequence gap is widening at increasing speed. Modelling techniques are thus a major source of structural information, and simulation techniques allow dynamic features to be explored at levels of detail rarely possible experimentally. The EMBO Practical Course addressed three types of simulation appropriate for studying biomolecules at different temporal and spatial scales: quantum mechanics, molecular dynamics, and Brownian dynamics. These topics were supplemented by practical introductions to force fields, macromolecular electrostatics, coarse-graining of molecular models, protein modeling, free energy calculations, structure-based drug design, data fitting to obtain macromolecular structures, as well as programming techniques. The aim of the course was to provide the basic theory and practical hints for using these methods so that the students would know how to begin to put them into practice when they returned to their laboratories. Each topic was addressed by 1-2 lectures, followed by practical sessions. Further talks on the lecturers' own research gave the students insights into cutting-edge applications of these methods.
As in previous years, the course was oversubscribed. The accepted students had a range of scientific backgrounds, both experimentally and theoretically oriented, mostly at the doctoral and post-doctoral level. Apart from Rebecca Wade, Antonia Stank and Mustafa Ghulam also participated from HITS, assisting with the practical session on electrostatics and Brownian dynamics simulation.

Sponsor: European Molecular Biology Organisation (EMBO)

5.1.5 SDBV WORKSHOPS, MELBOURNE / AUSTRALIA

In September 2014, Martin Golebiewski (SDBV) organized two workshops on data standards and computer modelling at the 15th International Conference on Systems Biology (ICSB2014) in Melbourne, Australia (http://www.icsb14.com):
At the NORMSYS & ISBE workshop "Standards for data and model exchange in systems biology", about 30 renowned participants from academic research and industries, as well as editors of scientific journals and representatives of research funding agencies, discussed the implementation and distribution of standards for data and models in system biology. Interoperability and interfacing between the standards and also options for transferring grass-root community standards into common norms were identified during the meeting. It helped to better coordinate the standardization efforts and provided a foundation for further harmonization and alignment of systems biology standards.
The COMBINE & ERASysAPP tutorial "Modelling and Simulation of Biological Models" (http://co.mbine.org/events/tutorial2014) showed young scientists how to set up quantitative computer models of biological networks using experimental kinetic data and how to simulate them in different systems biology platforms. Tutors from all over

the world taught the more than 60 attendees to use the modelling and simulation tools and databases they provide and showed them how to use standardized formats to exchange models. The SDBV group also presented their database SABIO-RK and the SEEK system.

5.1.6 SYMPOSIUM "EXTREMES", HANNOVER

6 – 7 October 2014, Hannover/Germany

Extreme events occurring in natural and man-made systems are neither fully excludable nor reliably predictable. However, precautions and early warnings can diminish the loss of lives and damages to health, ecosystems, infrastructure, and property. Among other things, improvements depend on a better understanding of extreme events in complex systems. To this end, computer simulations of extremes have gained considerable interest across disciplines.

The Symposium "Extreme Events: Modeling, Analyses, and Prediction" by the Volkswagen Foundation took place in October 2014 in Hannover. The key areas were the geo-, environmental-, climate-, and life sciences, as well as mathematical statistics. A specific focus was set on improved methodologies and their application to real-world problems. Prof. Tilmann Gneiting, head of the CST group at HITS, was one of the organizers.

Kira Feldmann, a PhD Student of the CST group, was awarded a poster prize at the Symposium. Out of 21 participants, she and three other young researchers from Potsdam, Hamburg, and Munich were chosen as winners. Kira's poster was entitled "Spatial postprocessing for forecasts of temperature minima and maxima".

5.1.7 SYSTEM BIOLOGY DATA MANAGEMENT FOUNDRY, HEIDELBERG

Data management is generally recognized as a crucially important but challenging part of Systems Biology. Data management systems that are useful, usable, and used are difficult to get right. It is not easy to find out about already-existing software or the state of standards that could or should be adopted.

The first systems biology data management foundry workshop, a meeting of data management practitioners organized by Wolfgang Müller from HITS, Bernd Rinn from ETH Zurich, and Carole Goble from University of Manchester, was held in 2012. Its goal was precisely meeting the communication needs of practicing data managers.

From 6 to 7 October 2014, ERASysAPP gave us the possibility to organize a follow-up. The organizing team was extended by Dagmar Waltemath (University of Rostock). We chose the SVB as the venue.

Fig. 53: The Motto of the Workshop: "Better Software, Better Research".

35 scientists from five countries participated in the meeting. As an opener, Neil Chue Hong, director of the Software Sustainability Institute (UK), gave a talk on the topic "Scientific Software: Sustainability, Skills & Sociology" that covered the wide range of challenges scientific software has to meet and the many open problems that are to be solved.

The participants then showed each other their software in a sequence of show-and-tell sessions, which generated discussions. A combined dinner provided the opportunity to continue discussions as well as to socialize in a more relaxed setting.

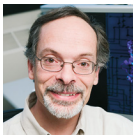
On the second day, we discussed the impact of ESFRI infrastructures on the work we had seen on the first day, as well as ways towards increasing the likelihood of making infrastructures in systems biology successful.

5.1.8 WORKSHOP ON HIGH DIMENSIONAL, HIGH FREQUENCY, AND SPATIAL DATA

29 – 31 October 2014, Karlsruhe/Germany

Analyzing large and complex (indeed, "big") data sets poses intriguing challenges to statisticians. These challenges were in the focus of the workshop on High Dimensional, High Frequency, and Spatial Data that was organized by KIT professors Claudia Kirch and Vicky Fasen along with CST group leader Tilmann Gneiting. The workshop participants presented statistical perspectives on topics as diverse as precipitation fields, financial regulation, solar power production, and brain signals. This range of topics illustrates how statistical methods can generate new insight and lead to better decision making in many areas of modern society. In practice, data analysis nowadays relies heavily on the use of computers. Accordingly, several workshop contributions tackled computational aspects, ranging from parallel computing to simulation algorithms. Furthermore, new data structures call for theoretical advances in order to analyze the properties of statistical methods in complex situations. A number of theoretical contributions responded to this need, relating to the fields of forecast evaluation, space-time modelling, causality analysis, and high frequency data analysis.

Overall, the workshop gave an impressive overview of statistics in data-rich scenarios. A particularly positive aspect was the smooth interaction between theory and applications, with each area inspiring and challenging the other. Finally, the workshop also delivered unexpected insights into the field of genealogy: In addition to presenting his scientific results, participant Will Kleiber of the University of Colorado at Boulder noted that his forefathers once migrated to the United States from the city of Durlach, just a few kilometers away from the workshop venue on the KIT campus.

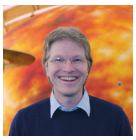


Prof. Dr. Gustavo Caetano-Anollés
Institute for Genomic Biology, and Illinois Informatics Institute,
University of Illinois at Urbana-Champaign
February 17, 2014: Computing the origin and history of life

Prof. Dominik Marx
Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum
March 17, 2014: Molecular Nano (Newton) Mechanics

Prof. Anna Wienhard
Mathematical Institute, University of Heidelberg
April 28, 2014: Geometry through symmetry

Prof. Michael K. Gilson
Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego
May 19, 2014: Plumbing the depths of entropy and enthalpy in molecular recognition



Dr. Bruno Leibundgut
ESO Headquarters, Garching
June 16, 2014: Exploring our universe with supernovae

Prof. Dr. Daniel A. Keim
University of Konstanz, Department of Computer and Information Science
July 17, 2014: The Power of Visual Analytics: Unlocking the Value of Big Data

Prof. Dr. Max Mühlhäuser
Technical University of Darmstadt, Telecooperation Lab
September 15, 2014: User Centric Smart Spaces

Dr. Gerhard Hummer
Max Planck Institute of Biophysics, Department of Theoretical Biophysics, Frankfurt
October 20, 2014: Molecular simulation of protein dynamics and function

Dr. Michele Catanzaro
Freelance Journalist / HITS “Journalist in Residence” 2014
November 11, 2014: Put more Journalism into Science-Journalism.
An international, non-naive view of research in the media

Dr. Stuart Parkin
Max Planck Institute for Microstructure Physics, Halle
November 17, 2014: The Spin on Electronics! Science and Technology
of spin currents in nano-materials and nano-devices

5.3 Explore Science

Like every year, in 2014, HITS participated in Explore Science, the science event for children, students, and their families organized by the Klaus Tschira Foundation. The event took place on 9 – 13 July and used the motto "Living Diversity", attracting more than 43,000 visitors to Mannheim's Luisenpark.

The HITSters created three hands-on stations for the event that revealed that there is diversity everywhere in nature – from molecules to genealogical trees. Using the slogan "What do we have in common with a banana?" the Molecular Biomechanics (MBM) group extracted DNA from a banana and revealed in it all the necessary building blocks of life. The scientists conducted the experiment together with the children, showing them how to use daily kitchen tools to extract DNA from bananas. In another step, the scientists explained what DNA is and how it defines different organisms – revealing that humans share approximately fifty percent of their DNA with a banana – a fun scientific fact that surprised young and old alike.

Making relations between different species was topic of the shared hands-on station of the Scientific Database and Visualization (SDBV) and Scientific Computing (SCO) group. The fact that humans and primates are related was one of the easier parts of this quiz – but what about birds, lizards, and insects? Here, the visitors could test their knowledge about the tree of life and find out how diversity developed throughout evolution.

Showing the diversity on a molecular scale was the idea of the Molecular and Cellular Modeling (MCM) group. Using Origami (the Japanese art of paper folding), the scientists created small protein models made of paper that displayed how the same "building blocks" can form different molecules depending on their assembly. "Dancing" molecule simulations on the computer rounded off the hands-on station.

As an expert of algorithms to analyze genealogical trees, Alexandros Stamatakis, leader of the SCO group, held a presentation in the "Festhalle Baumhain" as part of the framework program of Explore Science and explained how software and supercomputers help to calculate diversity and family trees.

Fig. 54: The MBM group shows how to use kitchen tools to extract DNA from bananas. Fascinating!

Fig. 55: Ulrike Wittig (SDBV) and some young visitors deciphering the tree of life.

Fig. 56: Diversity on small scale: Stefan Richter (MCM) builds proteins using origami.

5.4 HITS on Board - Digital Research on the Science Boat MS Wissenschaft

On 6 August 2014, HITS went “on board” the “MS Wissenschaft”, a floating science center that lay at anchor in Mannheim. The boat cruises through Germany’s rivers annually, bringing with it a unique exhibition on a different topic every year. In 2014, on the occasion of the Year of Science, the exhibition featured the topic “Digital Society”, and HITS seized the chance to contribute to this matching theme.

Using the motto “Digital Research – From Molecules to the Universe”, HITS presented the opportunities of data-driven research. The event was open to the public and attracted regular visitors of the exhibition as well as visitors who had come especially for the HITS event. HITSters from three different groups presented their work in interactive hands-on stations, showing opportunities of computer-aided, data-driven research methods.

Kai Polsterer’s Astroinformatics group sent the visitors up into space using the Oculus Rift, a virtual reality display. At the Molecular and Cellular Modeling group’s hands-on station, the visitors could see moving proteins through 3D glasses and learn how these simulations of proteins can help in matters of drug design. The third station, from the Theoretical Astrophysics group, showed the Illustris Simulation – the biggest computer simulation of the universe. The scientists explained how such simulations are made and why they’re so important for our basic understanding of astrophysics.

The highlight of the event was the panel discussion in the evening, in which Theresia Bauer (the Minister of Science of the state of Baden-Württemberg), HITS researchers Volker Springel (group leader Theoretical Astrophysics) and Tilmann Gneiting (group leader Computational Statistics), and science journalist Volker Stollorz (freelance journalist and HITS “Journalist in Residence” 2012) discussed the opportunities and risks of “e-science”. After a vivid and stimulating discussion, the participants (moderator: Hans-Georg Bock, Heidelberg University) agreed that the digitization of research helps break new ground in natural science. “But knowledge alone is not enough,” Bauer reminded the audience, “it must also be evaluated.”

As a “rounding-off” of the evening, the guests and HITS researchers enjoyed various finger food and refreshments on the deck of the boat and continued their conversations to while away the evening.

Fig. 57: (f.l.t.r.) Hans Georg-Bock (Heidelberg University), Volker Springel (HITS), Theresia Bauer (Minister of Science, Baden-Württemberg), Tilmann Gneiting (HITS) and Volker Stollorz (Journalist).

Fig. 58: Exploring the physiology of proteins with 3D glasses.

5.5 Heidelberg Laureate Forum 2014

The 2nd Heidelberg Laureate Forum took place from 21 – 26 September. Laureates of the most prestigious awards in Mathematics and Computer Science (the Alan Turing Awards, the Fields Medal, and the Nevanlinna Prize) came together in Heidelberg to meet young researchers from over sixty different countries. As in the year before, there were several talks given by the laureates. Young researchers and laureates interacted in various framework activities and the young researchers had the opportunity to get to know the Heidelberg research landscape. Part of the framework program involved a dinner at the Schwetzingen and Heidelberg castles, a river cruise on the Neckar, as well as an own Oktoberfest. There were plenty of opportunities for the young researchers and laureates to exchange ideas and to connect with another generation.

The five-day program also offered a new format: the “Hot Topic” discussion on 23 September. The focus of this first round was “The Role and the Potential of Mathematics and Computer Science in Developing Nations/Emerging Economies”. Scientists from Niger, Bangladesh, Ecuador, Cambodia, and India held presentations on the status of science in their countries, followed by a vivid discussion on the challenges and chances for researchers in these countries.

HITS has co-initiated this networking event and supports the HLF in its scientific expertise. Further, HITS and its scientists also hosted an event in the framework of the “young researcher’s day”. This special program allows the young researchers to visit and connect to several scientific institutes in Heidelberg. On 24 September, HITS welcomed 25 young scientists from all over the world to the institute. The HITSters presented their work and eagerly engaged in discussions with the international guests.

The next Heidelberg Laureate Forum will take place 23 – 28 August 2015.

Fig. 59: HITS Managing Director Andreas Reuter speaks at the 2nd HLF. (Picture: HLFF / Flemming)

Fig. 60: Volker Springel (TAP) debates with the Young Researchers.

Fig. 61: Group Picture: The Young Researchers at the end of their visit at HITS.

Fig. 62: The HITS Scientific Advisory Board: Prof. Dieter Kranzlmüller (LMU Munich, 2nd from left) was elected chair at the first Board meeting. The HITS Managing Directors are Klaus Tschira (3rd from left) and Prof. Andreas Reuter (far right). Prof. Rebecca Wade (6th from left) is the first Scientific Director of HITS, and her deputy is Prof. Michael Strube (4th from left).

On November 14th, 2014, the HITS Scientific Advisory Board (SAB) held its constitutive meeting. Nine external experts from different research fields are members of this board that advises the institute and orchestrates a regular scientific evaluation process. At the board meeting, Prof. Dieter Kranzlmüller (LMU Munich) was elected the first chairman.

With the “HITS Stiftung” and the Scientific Advisory Board, the institute has been given a structure that is supposed to foster the scientific goals of HITS sustainably by following the motto “Think Beyond the Limits!”

More about the Scientific Advisory Board in Chapter 6.

6 Scientific Advisory Board

The Scientific Advisory Board (SAB) of HITS is a group of internationally renowned scientists that supports the management of HITS in various aspects of running, planning, and directing the institute. In particular, the SAB orchestrates the periodic evaluation of all the research groups of HITS. The evaluation is a continuous process involving three to four research groups per year. The SAB presents the results to the HITS management and makes recommendations of how to further improve the institute's research performance.

The SAB had its constitutive meeting in November 2014. In December 2014, Nobel Prize winner Stefan Hell joined the committee that currently consists of the following members:

- Dr. Adele Goldberg, past President of the Association for Computing Machinery (ACM)
- Prof. Gert-Martin Greuel, University of Kaiserslautern, past Director of the "Mathematical Research Institute of Oberwolfach"
- Prof. Dr. Stefan Hell, Director at the Max-Planck-Institute for Bio-Physical Chemistry, Göttingen
- Prof. Dr. Tony Hey, University of Southampton
- Prof. Dr. Masaru Kitsuregawa, University of Tokyo, Director General of the National Institute of Informatics, Japan
- Dr. Heribert Knorr, Head of Department at Ministry of Science, Research and the Arts Baden-Württemberg (retired)
- Prof. Dr. Dieter Kranzlmüller, Ludwig Maximilians University, Munich, Director of the Leibniz Super Computing Center (Chair)
- Prof. Dr. Thomas Lengauer, Max-Planck-Institute for Computer Science, Saarbrücken
- Prof. Dr. Alex Szalay, Johns Hopkins University
- Prof. Dr. Jeannette Wing, Carnegie-Mellon University, Corporate VP of Microsoft Research

[Aberer et al., 2014] A.J. Aberer, K. Kobert, A. Stamatakis: "ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era". In Molecular Biology and Evolution, 2014.

[Ackermann et al., 2014] M. Ackermann, M. Ajello, A. Albert, A. Allafort, W. B. Atwood, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, E. D. Bloom, E. Bonamente, E. Bottacini, T. J. Brandt, J. Bregeon, M. Brigida, P. Bruel, R. Buehler, S. Buson, G. A. Caliandro, R. A. Cameron, P. A. Caraveo, E. Cavazzuti, R. C. G. Chaves, J. Chiang, G. Chiaro, S. Ciprini, R. Claus, J. Cohen-Tanugi, J. Conrad, F. D'Ammando, A. de Angelis, F. de Palma, C. D. Dermer, S. W. Digel, P. S. Drell, A. Drlica-Wagner, C. Favuzzi, A. Franckowiak, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, S. Germani, N. Giglietto, F. Giordano, M. Giroletti, G. Godfrey, G. A. Gomez-Vargas, I. A. Grenier, S. Guiriec, M. Gustafsson, D. Hadasch, M. Hayashida, J. Hewitt, R. E. Hughes, T. E. Jeltema, G. Jóhannesson, A. S. Johnson, T. Kamae, J. Kataoka, J. Knödlseder, M. Kuss, J. Lande, S. Larsson, L. Latronico, M. Llena Garde, F. Longo, F. Loparco, M. N. Lovellette, P. Lubrano, M. Mayer, M. N. Mazziotta, J. E. McEnery, P. F. Michelson, W. Mitthumsiri, T. Mizuno, M. E. Monzani, A. Morselli, I. V. Moskalenko, S. Murgia, R. Nemmen, E. Nuss, T. Ohsugi, M. Orienti, E. Orlando, J. F. Ormes, J. S. Perkins, M. Pesce-Rollins, F. Piron, G. Pivato, S. Rainò, R. Rando, M. Razzano, S. Razzaque, A. Reimer, O. Reimer, J. Ruan, M. Sánchez-Conde, A. Schulz, C. Sgrò, E. J. Siskind, G. Spandre, P. Spinelli, E. Storm, A. W. Strong, D. J. Suson, H. Takahashi, J. G. Thayer, J. B. Thayer, D. J. Thompson, L. Tibaldo, M. Tinivella, D. F. Torres, E. Troja, Y. Uchiyama, T. L. Usher, J. Vandenbroucke, G. Vianello, V. Vitale, B. L. Winer, K. S. Wood, S. Zimmer, Fermi-LAT Collaboration, A. Pinzke & C. Pfrommer. Search for Cosmic-Ray-induced Gamma-Ray Emission in Galaxy Clusters. The Astrophysical Journal. 787:18, 2014.

[Angulo et al., 2014] R. E. Angulo, S. D. M. White, V. Springel & B. Henriques. Galaxy formation on the largest scales:the impact of astrophysics on the baryonic acoustic oscillation peak. Monthly Notices of the Royal Astronomical Society. 442:2131-2144, 2014.

[Arnold et al., 2014] C. Arnold, E. Puchwein & V. Springel. The Lyman-alpha forest in f(R) modified gravity. ArXiv e-prints. arXiv:1411.2600, 2014.

[Arnold et al., 2014] C. Arnold, E. Puchwein & V. Springel. Scaling relations and mass bias in hydrodynamical f (R) gravity simulations of galaxy clusters. Monthly Notices of the Royal Astronomical Society. 440:833-842, 2014.

[Baldi et al., 2014] M. Baldi, F. Villaescusa-Navarro, M. Viel, E. Puchwein, V. Springel and L. Moscardini. Cosmic degeneracies - I. Joint N-body simulations of modified gravity and massive neutrinos. Monthly Notices of the Royal Astronomical Society. 440:75-88, 2014.

[Battaglia et al., 2014] N. Battaglia, J. R. Bond, C. Pfrommer and J. L. Sievers. On the Cluster Physics of Sunyaev-Zel'dovich and X-ray Surveys IV:Characterizing Density and Pressure Clumping due to Infalling Substructures. ArXiv e-prints. arXiv:1405.3346, 2014.

[Bergmann et al., 2014] Frank T. Bergmann, Richard Adams, Stuart Moodie, Jonathan Cooper, Mihai Glont, Martin Golebiewski, Michael Hucka, Camille Laibe, Andrew K. Miller, David P. Nickerson, Brett G. Olivier, Nicolas Rodriguez, Herbert M. Sauro, Martin Scharm, Stian Soiland-Reyes, Dagmar Waltemath, Florent Yvon, Nicolas Le Novère (2014). COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project. BMC Bioinformatics, 2014, 15:369.

[Bird et al., 2014] S. Bird, M. Vogelsberger, M. Haehnelt, D. Sijacki, S. Genel, P. Torrey, V. Springel and L. Hernquist. Damped Lyman α absorbers as a probe of stellar feedback. Monthly Notices of the Royal Astronomical Society. 445:2313-2324, 2014.

[Brehm et al., 2014] M. A. Brehm, V. Huck, C. Aponte-Santamaría, T. Obser, S. Grässle, F. Oyen, U. Budde, S. Schneppenheim, C. Baldauf, F. Gräter, S. W. Schneider, R. Schneppenheim. von Willebrand disease type 2A phenotypes IIC, IID and IIE: A day in the life of shear-stressed mutant von Willebrand factor. Thromb Haemost., 112(1):96-108.

[Broderick et al., 2014] A. E. Broderick, C. Pfrommer, E. Puchwein & P. Chang. Implications of Plasma Beam Instabilities for the Statistics of the Fermi Hard Gamma-Ray Blazars and the Origin of the Extragalactic Gamma-Ray Background. The Astrophysical Journal. 790:137, 2014.

[Broderick et al., 2014] A. E. Broderick, C. Pfrommer, E. Puchwein, P. Chang & K. M. Smith. Lower Limits on the Anisotropy of the Extragalactic Gamma-Ray Background Implied by the 2FGL and 1FHL Catalogs. The Astrophysical Journal. 796:12, 2014.

[Bryan et al., 2014] G. L. Bryan, M. L. Norman, B. W. O'Shea, T. Abel, J. H. Wise, M. J. Turk, D. R. Reynolds, D. C. Collins, P. Wang, S. W. Skillman, B. Smith, R. P. Harkness, J. Bordner, J.-h. Kim, M. Kuhlen, H. Xu, N. Goldbaum, C. Hummels, A. G. Kritsuk, E. Tasker, S. Skory, C. M. Simpson, O. Hahn, J. S. Oishi, G. C. So, F. Zhao, R. Cen, Y. Li und Enzo Collaboration. ENZO: An Adaptive Mesh Refinement Code for Astrophysics. The Astrophysical Journal Supplement Series. 211:19, 2014.

[Buschkamp et al., 2014] P. Buschkamp, W. Seifert, K. Polsterer, J. Heidt, S. Rabien, H. Gemperlein, R. K. Gredel, M. Lehmitz, G. Orban de Xivry, A. Pramskiy, W. Raab, D. Thompson, M. D. De La Peña, and J. Ziegleder. "LUCI: binocular and LGS/NGS AO modes of LUCI at the LBT" Ground-based and Airborne Instrumentation for Astronomy V. Proceedings of SPIE Vol. 9147.

[Chang et al., 2014] P. Chang, A. E. Broderick, C. Pfrommer, E. Puchwein, A. Lamberts and M. Shalaby. The Effect of Nonlinear Landau Damping on Ultrarelativistic Beam Plasma Instabilities. The Astrophysical Journal. 797:110, 2014.

[Cooper et al., 2014] A. P. Cooper, L. Gao, Q. Guo, C. S. Frenk, A. Jenkins, V. Springel and S. D. M. White. Surface photometry of BCGs and intracluster stars in Lambda-CDM. ArXiv e-prints. arXiv:1407.5627, 2014.

[Costescu and Gräter, 2014] B. I. Costescu, F. Gräter. Graphene mechanics: II. Atomic stress distribution during indentation until rupture. Phys Chem Chem Phys., 16(24):12582-90.

[Costescu at al., 2014] B. I. Costescu, I. B. Baldus, F. Gräter. Graphene mechanics: I. Efficient first principles based Morse potential. Phys Chem Chem Phys., 16(24):12591-8.

[Costi, 2014] Costi MP, Marverti G, Cardinale D, Venturelli A, Ferrari S, Ponterini G (Univ. Modena), Henrich S, Salo-Ahen O, Wade R. (HITS) "Peptides binding to the dimer interface of thymidylate synthase for the treatment of cancer" United States Patent No.: US 8,916,679 B2, Date of Patent: 23.12.2014. Filed: 01.12.2009.

[Dao et al., 2014] D. Dao, T. Flouri, A. Stamatakis: "Automated Plausibility Analysis of Large Phyolgenies". Book chapter, to appear 2014.

[Dunthom et al., 2014] M. Dunthorn, J. Otto, S.A. Berger, A. Stamatakis, F. Mahé, S. Romac, C. de Vargas, S. Audic, A. Stock, F. Kauff and T. Stoeck, BioMarKs Consortium: "Placing environmental next generation sequencing

amplicons from microbial eukaryotes into a phylogenetic context". In Molecular Biology and Evolution, 2014.

[Dueck et al., 2014] Dueck, J., Edelmann, D., Gneiting, T. and Richards, D. (2014).The affinely invariant distance correlation. Bernoulli, 20, 2305-2330.

[Fahrni et al., 2014a] Angela Fahrni and Michael Strube. A latent variable model for discourse-aware concept and entity disambiguation. In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, 26-30 April 2014, pages 491-500, 2014.

[Fahrni et al., 2014b] Angela Fahrni, Benjamin Heinzerling, Thierry Göckel, and Michael Strube. HITS' monolingual and cross-lingual entity linking system at TAC 2013. In Proceedings of the Text Analysis Conference, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 18-19 November 2013, 2014.

[Fink et al., 2014] M. Fink, M. Kromer, I. R. Seitenzahl, F. Ciaraldi-Schoolmann, F. K. Röpke, S. A. Sim, R. Pakmor, A. J. Ruiter and W. Hillebrandt. Three-dimensional pure deflagration models with nucleosynthesis and synthetic observables for Type Ia supernovae. Monthly Notices of the Royal Astronomical Society. 438:1762-1783, 2014.

[Flouri et al., 2014a] T. Flouri, F. Izquierdo-Carrasco, D. Darriba, A.J. Aberer, L.-T. Nguyen, B.Q. Minh, A. von Haeseler, A. Stamatakis: "The Phylogenetic Likelihood Library". In Systematic Biology, 2014.

[Flouri et al., 2014b] T. Flouri, K. Kobert, S.P. Pissis, A. Stamatakis: "An optimal algorithm for computing all subtree repeats in trees". In Phil. Trans. R. Soc. A 372:2016, 2014.

[Flouri et al., 2014c] T. Flouri, A. Stamatakis, K. Kobert, A.J. Aberer: "The divisible load balance problem and its application to phylogenetic inference". In Proceedings of WABI 2014, Wroclaw, Poland, September 2014, accepted for publication

[Foley et al. 2014] R. J. Foley, O. D. Fox, C. McCully, M. M. Phillips, D. J. Sand, W. Zheng, P. Challis, A. V. Filippenko, G. Folatelli, W. Hillebrandt, E. Y. Hsiao, S. W. Jha, R. P. Kirshner, M. Kromer, G. H. Marion, M. Nelson, R. Pakmor, G. Pignata, F. K. Röpke, I. R. Seitenzahl, J. M. Silverman, M. Skrutskie and M. D. Stritzinger. Extensive HST ultraviolet spectra and multiwavelength observations of SN 2014J in M82 indicate reddening and circumstellar scattering by typical dust. Monthly Notices of the Royal Astronomical Society. 443:2887-2906, 2014.

[Fontanot et al., 2014] F. Fontanot, S. Cristiani, C. Pfrommer, G. Cupani and E. Vanzella. On the evolution of the cosmic ionizing background. Monthly Notices of the Royal Astronomical Society. 438:2097-2104, 2014.

[Fuller et al., 2014a] Fuller, J.C., Martinez, M., Henrich, S., Stank, A., Richter, S. and Wade, R.C.
LigDig: a web server for querying ligand-protein interactions, Bioinformatics, (2015) 31(7):1147-1149.

[Fuller, 2014b] Fuller, J.C., Martinez, M. and Wade, R.C. On Calculation of the Electrostatic Potential of a Phosphatidylinositol Phosphate-Containing Phosphatidylcholine Lipid Membrane Accounting for Membrane Dynamics. PLoS ONE, (2014), 9(8): e104778, doi:10.1371/journal.pone.0104778.

[Gao et al., 2014] L. Gao, T. Theuns & V. Springel. Star forming filaments in warm dark models. ArXiv e-prints. arXiv:1403.2475, 2014.

[Genel et al., 2014] S. Genel, M. Vogelsberger, V. Springel, D. Sijacki, D. Nelson, G. Snyder, V. Rodriguez-Gomez, P. Torrey & L. Hernquist. Introducing the Illustris project:the evolution of galaxy populations across cosmic

time. Monthly Notices of the Royal Astronomical Society. 445:175-200, 2014.

[Gieseke, 2014] Fabian Gieseke, Kai Polsterer, Cosmin Eugen Oancea, and Christian Igel. “Speedy Greedy Feature Selection: Better Redshift Estimation via Massive Parallelism”.
ESANN 2014: 22nd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 23-25 April

[Ghulam, 2014] Ghulam Mustafa, Xiaofeng Yu and Rebecca Wade. Structure and Dynamics of Cytochrome P450 enzymes. Drug Metabolism Prediction, Ed. Kirchmair, J., Wiley-VCH Verlag, (2014), Ch. 4, pp 77-101.

[Gneiting, 2014] T. Gneiting. Calibration of medium-range weather forecasts. European Centre for Medium-Range Weather Forecasts Technical Memorandum 719, 2014.

[Gneiting and Katzfuss, 2014] T. Gneiting and M. Katzfuss. Probabilistic forecasting. Annual Review of Statistics and its Application, 1, 125-151, 2014.

[Goñi et al, 2014] G. M. Goñi, C. Epifano, J. Boskovic, M. Camacho-Artacho M. Camacho-Artacho, J. Zhou, A. Bronowska, M. T. Martín, M. J. Eck, L. Kremer, F. Gräter, F. L. Gervasio, M. Perez-Moreno, D. Lietha. Phosphatidylinositol 4,5-bisphosphate triggers activation of focal adhesion kinase by inducing clustering and conformational changes. Proc Natl Acad Sci U S A., 111(31):E3177-86 , 2014.

[Grässle et al., 2014] S. Grässle, V. Huck, K. I. Pappelbaum, C. Gorzelanny, C. Aponte-Santamaría, C. Baldauf, F. Gräter, R. Schneppenheim, T. Obser, S. W. Schneider. von Willebrand factor directly interacts with DNA from neutrophil extracellular traps. Arterioscler Thromb Vasc Biol., 34(7):1382-9, 2014.

[Grimm et al., 2014] G. Grimm, P. Kapli, B. Bomfleur, S. McLoughlin, S.S. Renner. “Using more than the oldest fossils: Dating Osmundaceae by three Bayesian clock approaches.” In Systematic Biology, 2014.

[Hayward et al., 2014] C. C. Hayward, L. Lanz, M. L. N. Ashby, G. Fazio, L. Hernquist, J. R. Martínez-Galarza, K. Noeske, H. A. Smith, S. Wuyts & A. Zezas. The total infrared luminosity may significantly overestimate the star formation rate of quenching and recently quenched galaxies. Monthly Notices of the Royal Astronomical Society. 445:1598-1604, 2014.

[Hayward et al., 2014] C. C. Hayward, P. Torrey, V. Springel, L. Hernquist & M. Vogelsberger. Galaxy mergers on a moving mesh:a comparison with smoothed particle hydrodynamics. Monthly Notices of the Royal Astronomical Society. 442:1992-2016, 2014.

[Hemri et. al, 2014a] Hemri, S., Lisniak, D., and Klein, B. (2014). Ascertainment of probabilistic runoff forecasts considering censored data (in German). Hydrologie und Wasserbewirtschaftung, 58, 84-94.

[Hemri et al., 2014b] S. Hemri, M. Scheuerer, F. Pappenberger, K. Bogner and T. Haiden. Trends in the predictive performance of raw ensemble weather forecasts. Geophysical Research Letters, 41, 9197–9205, 2014.

[Henriques et al., 2014] B. Henriques, S. White, P. Thomas, R. Angulo, Q. Guo, G. Lemson, V. Springel & R. Overzier. Galaxy formation in the Planck Cosmology I - Matching the observed evolution of star-formation rates, colours and stellar masses. ArXiv e-prints. arXiv:1410.0365, 2014.

[Hensen et al., 2014] U. Hensen, F. Gräter, R. H. Henchman. Macromolecular entropy can be accurately computed from force. J. Chem. Theory Comput., 10:4777-4781 , 2014.

[Heuveline and M. Schick, 2014] V. Heuveline and M. Schick. A hybrid generalized Polynomial Chaos method for stochastic dynamical systems. International Journal for Uncertainty Quantification 4(1): 37-61, 2014.

[Hou, 2014] Yufang Hou, Katja Markert, and Michael Strube. A rule-based system for end-to-end bridging resolution. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25-29 October 2014, pages 2082-2093, 2014.

[Izquierdo-Carrasco et al., 2014] F. Izquierdo-Carrasco, J. Cazes, S.A. Smith, A. Stamatakis. "PUmPER: Phylogenies Updated Perpetually". In Bioinformatics, 2014.

[Jarvis et al., 2014] E.D. Jarvis, S. Mirarab, A.J. Aberer, B. Li, P. Houde, C. Li, S.Y.W. Ho, B.C. Faircloth, B. Nabholz, J.T. Howard, A. Suh, C.C. Weber, R.R. da Fonseca, J. Li, F. Zhang, H. Li, L. Zhou, N. Narula, L. Liu, G. Ganapathy, B. Boussau, Md.S. Bayzid, V. Zavidovych, S. Subramanian, T. Gabaldon, S. Capella-Gutiérrez, J. Huerta-Cepas, B. Rekepalli, K. Munch, M. Schierup, B. Lindow, W.C. Warren, D. Ray, R.E. Green, M.W. Bruford, X. Zhan, A. Dixon, S. Li, N. Li, Y. Huang, E.P. Derryberry, M. Frost Bertelsen, F.H. Sheldon, R.T. Brumfield, C.V. Mello, P.V. Lovell, M. Wirthlin, M. Paula Cruz Schneider, F. Prosdocimi, J. Alfredo Samaniego, A. Missael Vargas Velazquez, A. Alfaro-Nunez, P.F. Campos, B. Petersen, T. Sicheritz-Ponten, A. Pas, T. Bailey, P. Scofield, M. Bunce, D.M. Lambert, Q. Zhou, P. Perelman, A.C. Driskell, B. Shapiro, Z. Xiong, Y. Zeng, S. Liu, Z. Li, B. Liu, K. Wu, J. Xiao, X. Yinqi, Q. Zheng, Y. Zhang, H. Yang, J. Wang, L. Smeds, F.E. Rheindt, M. Braun, J. Fjeldsa, L. Orlando, F. Keith Barker, K. Andreas Jonsson, W. Johnson, K.-P. Koepfli, S. O'Brien, D. Haussler, O.A. Ryder, C. Rahbek, E. Willerslev, G.R. Graves, T.C. Glenn, J. McCormack, D. Burt, H. Ellegren, P. Alström, S.V. Edwards, A. Stamatakis, D.P. Mindell, J. Cracraft, E.L. Braun, T. Warnow, W. Jun, M. Thomas P. Gilbert, G. Zhang: "Whole-genome analyses resolve early branches in the tree of life of modern birds". In Science, 46(6215):1320-1331, 2014.

[Judea, 2014] Alex Judea, Hinrich Schütze, and Sören Brügmann. Unsupervised training set generation for automatic acquisition of technical terminology in patents. In Proceedings of Coling 2014: Poster Volume, Dublin, Ireland, 23-29 August 2014, pages 290-300, 2014.

[Kobert et al., 2014] K. Kobert, J. Hauser, A. Stamatakis: "Is the Protein Model Assignment Problem NP-hard?". In Theoretical Computer Science, 2014.

[Kosenko et al., 2014] D. Kosenko, W. Hillebrandt, M. Kromer, S. I. Blinnikov, R. Pakmor & J. S. Kaastra. Oxygen emission in remnants of thermonuclear supernovae as a probe for their progenitor system. ArXiv e-prints. arXiv:1411.4126, 2014.

[Kozlov et al., 2014] A.M. Kozlov, C. Goll, A. Stamatakis: "Efficient Computation of the Phylogenetic Likelihood Function on the Intel MIC Architecture". In Proceedings of HICOMB workshop, held in conjunction with IPDPS 2014, Phoenix, Arizona, May 2014, accepted for publication

[Lanfear et al., 2014] R. Lanfear, B. Calcott, D. Kainer, C. Mayer, A. Stamatakis: "Selecting optimal partitioning schemes for phylogenomic datasets". In BMC Evolutionary Biology 14:1, 82, 2014.

[Lanz et al., 2014] L. Lanz, C. C. Hayward, A. Zezas, H. A. Smith, M. L. N. Ashby, N. Brassington, G. G. Fazio & L. Hernquist. Simulated Galaxy Interactions as Probes of Merger Spectral Energy Distributions. The Astrophysical Journal. 785:39, 2014.

[Ludlow et al., 2014] A. D. Ludlow, J. F. Navarro, R. E. Angulo, M. Boylan-Kolchin, V. Springel, C. Frenk and S. D. M. White. The mass-concentration-redshift relation of cold dark matter haloes. Monthly Notices of the Royal Astronomical Society. 441:378-388, 2014.

[Marinacci et al., 2014] F. Marinacci, R. Pakmor and V. Springel. The formation of disc galaxies in high-resolution

moving-mesh cosmological simulations. Monthly Notices of the Royal Astronomical Society. 437:1750-1775, 2014.

[Marinacci et al., 2014] F. Marinacci, R. Pakmor, V. Springel and C. M. Simpson. Diffuse gas properties and stellar metallicities in cosmological simulations of disc galaxy formation. Monthly Notices of the Royal Astronomical Society. 442:3745-3760, 2014.

[Martschat et al., 2014] Sebastian Martschat and Michael Strube. Recall error analysis for coreference resolution. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25-29 October 2014, pages 2070-2081, 2014.

[Mereghetti et al., 2014] Mereghetti, P., Martinez, M. and Wade, R.C. Long range Debye-Hückel correction for computation of grid-based electrostatic forces between biomacromolecules. BMC Biophys., (2014), 7:4, http://www.biomedcentral.com/2046-1682/7/4.

[Mesgar et al., 2014] Mohsen Mesgar and Michael Strube. Normalized entity graph for computing local coherence. In Proceedings of TextGraphs-9: Graph-based Methods for Natural Language Processing, Workshop at EMNLP 2014, Doha, Qatar, 29 October 2014, pages 1-5, 2014.

[Michałowski et al., 2014] M. J. Michałowski, C. C. Hayward, J. S. Dunlop, V. A. Bruce, M. Cirasuolo, F. Cullen and L. Hernquist. Determining the stellar masses of submillimetre galaxies:the critical importance of star formation histories. Astronomy and Astrophysics. 571:AA75, 2014.

[Misof et al., 2014] B. Misof, S. Liu, K. Meusemann, R.S. Peters, A. Donath, C. Mayer, P.B. Frandsen, J. Ware, T. Flouri, R.G. Beutel, O. Niehuis, M. Petersen, F. Izquierdo-Carrasco, T. Wappler, J. Rust, A.J. Aberer, U. Aspöck, H. Aspöck, D. Bartel, A. Blanke, S. Berger, A. Böhm, T.R. Buckley, B. Calcott, J. Chen, F. Friedrich, M. Fukui, M. Fujita, C. Greve, P. Grobe, S. Gu, Y. Huang, L.S. Jermiin, A.Y. Kawahara, L. Krogmann, M. Kubiak, R. Lanfear, H. Letsch, Y. Li, Z. Li, J. Li, H. Lu, R. Machida, Y. Mashimo, P. Kapli, D.D. McKenna, G. Meng, Y. Nakagaki, J.L. Navarrete-Heredia, M. Ott, Y. Ou, G. Pass, L. Podsiadlowski, H. Pohl, B.M. von Reumont, K. Schütte, K. Sekiya, S. Shimizu, A. Slipinski, A. Stamatakis, W. Song, X. Su, N.U. Szucsich, M. Tan, X. Tan, M. Tang, J. Tang, G. Timelthaler, S. Tomizuka, M. Trautwein, X. Tong, T. Uchifune, M.G. Walzl, B.M. Wiegmann, J. Wilbrandt, B. Wipfler, T.K.F. Wong, Q. Wu, G. Wu, Y. Xie, S. Yang, Q. Yang, D.K. Yeates, K. Yoshizawa, Q. Zhang, R. Zhang, W. Zhang, Y. Zhang, J. Zhao, C. Zhou, L. Zhou, T. Ziesmann, S. Zou, Y. Li, X. Xu, Y. Zhang, H. Yang, J. Wang, J. Wang, K.M. Kjer, X. Zhou: “Phylogenomics resolves the timing and pattern of insect evolution”. In Science, 346(6210): 763-767, 2014.

[Mocz et al., 2014] P. Mocz, M. Vogelsberger, D. Sijacki, R. Pakmor & L. Hernquist. A discontinuous Galerkin method for solving the fluid and magnetohydrodynamic equations in astrophysical simulations. Monthly Notices of the Royal Astronomical Society. 437:397-414, 2014.

[Moosavi et al.,2014] Nafise Sadat Moosavi and Michael Strube. Unsupervised coreference resolution by utilizing the most informative relations. In Proceedings of the 25th International Conference on Computational Linguistics, Dublin, Ireland, 23-29 August 2014, pages 644-655, 2014.

[Müller et al.,2014] Thomas Müller, Richárd Farkas, Alex Judea, Helmut Schmid, and Hinrich Schütze. Dependency parsing with latent refinements of part-of-speech tags. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25-29 October 2014, pages 963-967, 2014.

[Muñoz et al., 2014] D. J. Muñoz, K. Kratter, V. Springel and L. Hernquist. Planet-disc interaction on a freely moving mesh. Monthly Notices of the Royal Astronomical Society. 445:3475-3495, 2014.

[Nelson et al., 2014] D. Nelson, S. Genel, M. Vogelsberger, V. Springel, D. Sijacki, P. Torrey & L. Hernquist. The impact of feedback on cosmological gas accretion. ArXiv e-prints. arXiv:1410.5425, 2014.

[Ohlmann et al., 2014] S. T. Ohlmann, M. Kromer, M. Fink, R. Pakmor, I. R. Seitenzahl, S. A. Sim & F. K. Röpke. The white dwarf's carbon fraction as a secondary parameter of Type Ia supernovae. Astronomy and Astrophysics. 572:AA57, 2014.

[Pakmor et al., 2014] R. Pakmor, F. Marinacci & V. Springel. Magnetic Fields in Cosmological Simulations of Disk Galaxies. The Astrophysical Journal. 783:LL20, 2014.

[Palmai et al., 2014] Z. Palmai, C. Seifert, F. Gräter, E. Balog. An allosteric signaling pathway of human 3-phosphoglycerate kinase from force distribution analysis. PLoS Comput Biol., 10(1):e1003444.

[Parveen et al., 2014] Daraksha Parveen and Michael Strube. Multi-document summarization using bipartite graphs. In Proceedings of TextGraphs-9: Graph-based Methods for Natural Language Processing, Workshop at EMNLP 2014, Doha, Qatar, 29 October 2014, pages 15-24, 2014.

[Patil et al., 2014a] S. P. Patil, S. Xiao, K. Gkagkas, B. Markert, F. Gräter. Viscous friction between crystalline and amorphous phase of dragline silk.
PLoS One., 9(8):e104832.

[Patil et al., 2014b] S. P. Patil, B. Markert, F. Gräter (2014). Rate-dependent behavior of the amorphous phase of spider dragline silk. Biophys J., 106(11):2511-8.

[Peters et al., 2014] R.S. Peters, K. Meusemann, M. Petersen, C. Mayer, J. Wilbrandt, T. Ziesmann, A. Donath, K.M. Kjer, U. Aspöck, H. Aspöck, A. Aberer, A. Stamatakis, F. Friedrich, F. Hünefeld, O. Niehuis, R.G. Beutel, B. Misof: "The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data". In BMC Evolutionary Biology, 14:52, 2014.

[Pillepich et al., 2014] A. Pillepich, M. Vogelsberger, A. Deason, V. Rodriguez-Gomez, S. Genel, D. Nelson, P. Torrey, L. V. Sales, F. Marinacci, V. Springel, D. Sijacki, and L. Hernquist. Halo mass and assembly history exposed in the faint outskirts:the stellar and dark matter haloes of Illustris galaxies. Monthly Notices of the Royal Astronomical Society. 444:237-249, 2014.

[Pradhan et al.,2014] Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng, and Michael Strube. Scoring coreference partitions of predicted mentions: A reference implementation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, Md., 22-27 June 2014, pages 30-35, 2014.

[Sales et al., 2014] L. V. Sales, F. Marinacci, V. Springel and M. Petkova. Stellar feedback by radiation pressure and photoionization. Monthly Notices of the Royal Astronomical Society. 439:2990-3006, 2014.

[Salichos et al., 2014] L. Salichos, A. Stamatakis, A. Rokas: "Novel Information Theory-Based Measures for Quantifying Incongruence among Phylogenetic Trees". In Molecular Biology and Evolution, 2014.

[Sanderson et al., 2014] M.J. Sanderson, M.M. McMahon, A. Stamatakis, D.J. Zwickl, M. Steel. "Impacts of terraces on phylogenetic inference", Arxiv preprint, 2014.

[Sasaki et al., 2014] M. Sasaki, P. C. Clark, V. Springel, R. S. Klessen and S. C. O. Glover. Statistical properties of dark matter mini-haloes at $z \geq 15$. Monthly Notices of the Royal Astronomical Society. 442:1942-1955, 2014.

[Scheuerer and Gneiting, 2014] Scheuerer, M. and Gneiting, T. (2014). Evaluating predictive performance. In: Mathematics of Planet Earth. Proceedings of the 15th Annual Conference of the International Association for Mathematical Geosciences. Pardo-Igúzquiza, E., Guardiola-Albert, C., Heredia, J., Moreno-Merino, L., Durán, J. J. and Vargas-Guzmán, J. A. (eds.).Springer, Berlin, Lecture Notes in Earth System Sciences, pp. 15-18.

[Schick et. al. 2014], M. Schick, V. Heuveline and O. P. Le Maître. A Newton-Galerkin method for fluid flow exhibiting uncertain periodic dynamics. SIAM/ASA Journal on Uncertainty Quantification (2)1: 153-173, 2014.

[Schick 2014] M. Schick. A parallel multilevel spectral Galerkin solver for linear systems with uncertain parameters. Proceedings of the 22nd Euromicro Conference on Parallel, distributed and network-based Processing: 352-359, 2014.

[Sijacki et al. 2014] D. Sijacki, M. Vogelsberger, S. Genel, V. Springel, P. Torrey, G. Snyder, D. Nelson & L. Hernquist. The Illustris simulation:Evolving population of black holes across cosmic time. ArXiv e-prints. arXiv:1408.6842, 2014.

[Smith and Stamatakis, 2014] S. A. Smith, A. Stamatakis: “Inferring and Postprocessing Huge Phylogenies”. In M. Elloumi, A.Y. Zomaya, editors. Biological Knowledge Discovery Handbook: Preprocessing, Mining, and Postprocessing of Biological Data, pages 1049-1072, John Wiley & Sons, Inc., 2014.

[Smith et al. 2014] R. J. Smith, S. C. O. Glover, P. C. Clark, R. S. Klessen & V. Springel. CO-dark gas and molecular filaments in Milky Way-type galaxies. Monthly Notices of the Royal Astronomical Society. 441:1628-1645, 2014.

[Song et al. 2014] Chen Song, Kristian Stavåker, Martin Wlotzka , Peter Fritzson and Vincent Heuveline. PDE modeling with modelica via FMI import of HiFlow3 C++ components with parallel multi-core simulations. 55th SIMS Conference on Simulation and Modelling: 184-191, 2014.

[Springel 2014] V. Springel. High performance computing and numerical modelling. ArXiv e-prints. arXiv:1412.5187, 2014.

[Stamatakis, 2014] A. Stamatakis: “RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies”. In Bioinformatics, 2014.

[Suwelack et al., 2014]. S. Suwelack, M. Stoll, S. Schalck, N. Schoch, R. Dillmann, R. Bendl, V. Heuveline, S. Speidel. The Medical Simulation Markup Language (MSML) – Simplifying the Biomechanical Modeling Workflow. Journal paper, accepted for „Medicine Meets Virtual Reality, MMVR2014“, 2014.

[Torrey et al., 2014] P., M. Torrey, Vogelsberger, S. Genel, D. Sijacki, V. Springel and L. Hernquist. A model for cosmological simulations of galaxy formation physics:multi-epoch validation. Monthly Notices of the Royal Astronomical Society. 438:1985-2004, 2014.

[Valle et al., 2014] M. Valle, H. Schabauer, C. Pacher, H. Stockinger, A. Stamatakis, M. Robinson-Rechavi, N. Salamin. “Optimisation strategies for fast detection of positive selection on phylogenetic trees”. In Bioinformatics, 2014.

[Vogelsberger et al., 2014a] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, S. Bird, D. Nelson & L. Hernquist. Properties of galaxies reproduced by a hydrodynamic simulation. Nature. 509:177-182, 2014.

[Vogelsberger et al., 2014b] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson and L. Hernquist. Introducing the Illustris Project.

simulating the coevolution of dark and visible matter in the Universe. Monthly Notices of the Royal Astronomical Society. 444:1518-1547, 2014.

[Vogelsberger et al., 2014c] M. Vogelsberger, J. Zavala, C. Simpson and A. Jenkins. Dwarf galaxies in CDM and SIDM with baryons:observational probes of the nature of dark matter. Monthly Notices of the Royal Astronomical Society. 444:3684-3698, 2014.

[Waltemath et al., 2014] Dagmar Waltemath, Frank T Bergmann, Claudine Chaouiya, Tobias Czauderna, Padraig Gleeson, Carole Goble, Martin Golebiewski, Michael Hucka, Nick Juty, Olga Krebs, Nicolas Le Novère, Huaiyu Mi, Ion I Moraru, Chris J Myers, David Nickerson, Brett G Olivier, Nicolas Rodriguez, Falk Schreiber, Lucian Smith, Fengkai Zhang, and Eric Bonnet. Meeting, 9:1285-1301, 2014. Meeting report from the fourth meeting of the Computational Modeling in Biology Network (COMBINE). Standards in Genomic Sciences, 9:1285-1301, 2014.

[Wellons et al., 2014] S. Wellons, P. Torrey, C.-P. Ma, V. Rodriguez-Gomez, M. Vogelsberger, M. Kriek, P. van Dokkum, E. Nelson, S. Genel, A. Pillepich, V. Springel, D. Sijacki, G. Snyder, D. Nelson, L. Sales & L. Hernquist. The Formation of Massive, Compact Galaxies at z=2 in the Illustris Simulation. ArXiv e-prints. arXiv:1411.0667, 2014.

[Wittig et al., 2014a] Ulrike Wittig, Maja Rey, Renate Kania, Meik Bittkowski, Lei Shi, Martin Golebiewski, Andreas Weidemann, Wolfgang Müller, and Isabel Rojas 281(2):572-582, 2014. Challenges for an enzymatic reaction kinetics database.FEBS Journal, 281(2):572-582, 2014.

[Wittig et al., 2014b] Ulrike Wittig, Renate Kania, Meik Bittkowski, Elina Wetsch, Lei Shi, Lenneke Jong, Martin Golebiewski, Maja Rey, Andreas Weidemann, Isabel Rojas and Wolfgang Müller Perspectives in Science, 33-40, 2014. Data extraction for the reaction kinetics database SABIO-RK. Perspectives in Science (2014) 1, 33–40.

[Xiao and Gräter, 2014] S. Xiao and F. Gräter. Molecular basis of the mechanical hierarchy in myomesin dimers for sarcomere integrity. Biophys J., 107(4):965-73, 2014.

[Yesuf et al., 2014] H. M. Yesuf, S. M. Faber, J. R. Trump, D. C. Koo, J. J. Fang, F. S. Liu, V. Wild & C. C. Hayward. From Starburst to Quiescence:Testing Active Galactic Nucleus feedback in Rapidly Quenching Post-starburst Galaxies. The Astrophysical Journal. 792:84, 2014.

[Yurin et al. 2014] D. Yurin & V. Springel. The stability of stellar disks in Milky-Way sized dark matter halos. ArXiv e-prints. arXiv:1411.3729, 2014.

[Yurin et al., 2014a] D. Yurin & V. Springel. An iterative method for the construction of N-body galaxy models in collisionless equilibrium. Monthly Notices of the Royal Astronomical Society. 444:62-79, 2014.

[Zandanel et al., 2014b] F. Zandanel, C. Pfrommer & F. Prada. A phenomenological model for the intracluster medium that matches X-ray and Sunyaev-Zel'dovich observations. Monthly Notices of the Royal Astronomical Society. 438:116-123, 2014.

[Zandanel et al., 2014] F. Zandanel, C. Pfrommer & F. Prada. On the physics of radio haloes in galaxy clusters:scaling relations and luminosity functions. Monthly Notices of the Royal Astronomical Society. 438:124-144, 2014.

[Zhou et al., 2014] B. Zhou, I. B. Baldus, W. Li, S. A. Edwards, F. Gräter. Identification of allosteric disulfides from prestress analysis. Biophys J., 107(3):672-81.

[Ziegel and Gneiting, 2014] J. Ziegel and T. Gneiting. Copula calibration. Electronic Journal of Statistics, 8, 2619-2638, 2014.

DEGREES

Habilitation

[Pfrommer, 2014] Christoph Pfrommer: “Interfacing High-Energy Astrophysics and Cosmological Structure Formation”, Habilitation Thesis, Venia Legendi in Physics, Heidelberg University and HITS, 2014.

Ph.D.

[Feng, 2014] Feng, Vanessa Wei: RST-style Discourse Parsing and its Applications in Discourse Analysis. PhD. Thesis, Computer Science Department, University of Toronto: Graeme Hirst and HITS: Michael Strube (External Reviewer), 2014.

[Izquierdo-Carrasco, 2014] Fernando Izquierdo-Carrasco: “Inference of Many-Taxon Phylogenies”, Ph.D. Thesis, Computer Science, Karlsruhe Institiute of Technology and HITS: Alexandros Stamatakis, 2014.

[Kapli, 2014] Paschalia Kapli: “Phylogeography and species delimitation of the Saharo-Arabian genus Mesalina (Sauria: Lacertidae) based on genetic markers”, Ph.D. Thesis, Biology, University of Crete: Alexandros Stamatakis, 2014.

[Yurin, 2014] Denis Yurin: “Construction and stability of disk galaxies, and the radial migration of their stars”, PhD Thesis, Physics, Heidelberg University and HITS: Volker Springel, 2014.

Master & Bachelor

[Arnold, 2014] Christian Arnold: “Scaling relations and mass bias in hydrodynamical f(R) gravity simulations of galaxy clusters”, Master´s Thesis, Physics, Heidelberg University and HITS: Volker Springel, 2014.

[Broscheit, 2014] Broscheit, Samuel: A Vector Space Model Induced by Non-Negative Tensor Factorization as Feature for Coreference Resolution, Master’s Thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2014.

[Dao, 2014] David Dao: “Automated Plausibility Analysis of Large Phylogenies”, Bachelor´s Thesis, Computer Science, Karlsruhe Institiute of Technology and HITS: Alexandros Stamatakis, 2014.

[Flick, 2014] Patrick Flick: “Analysis of human tissue-specific protein-protein interaction network”, Master’s Thesis, Computer Science, Karlsruhe Institiute of Technology and HITS: Alexandros Stamatakis, 2014.

[Göckel, 2014] Göckel, Thierry: Incorporating World Knowledge into Coreference Resolution, Magister Thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2014.

[Heinzerling, 2014] Heinzerling, Benjamin: Incorporating World Knowledge into Sentence Compression, Magister Thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2014.

[Krzyszkowska 2014] Jolanta Krzyszkowska: “Angular Momentum Properties of Baryonic and Dark Matter in the Illustris Simulation”, Master´s Thesis, Physics, Heidelberg University and HITS: Volker Springel, 2014.

[Ramsl, 2014] Ramsl, Hans-Martin: Leveraging Topic Models for Graphical Local Coherence Representation, Master’s Thesis, Neuphilologische Fakultät, Heidelberg University and HITS: Michael Strube, 2014.

[Schmidt, 2014] Patrick Schmidt: “Parametric Estimation of Loss Functions”, Diploma Thesis, Mathematics, Heidelberg University and HITS: Tilmann Gneiting, 2014.

[Tong, 2014] Tong, Rudi: “Comparative analysis of Adenylyl Cyclase binding sites”, Bachelor’s Thesis, Molecular Biotechnology, Heidelberg University and HITS: Neil Bruce and Rebecca Wade, 2014.

LECTURES

Vincent Heuveline

“Numerische Mathematik 2 – Numerik partieller Differentialgleichungen”. Uni Heidelberg. WS 14/15.

„Numerik“. Uni Heidelberg. SoSe 14.

„Einführung in die Numerik“. Uni Heidelberg. WS 14/15.

Tilmann Gneiting

Forecasting: Theory and Practice I, Karlsruhe Institute of Technology (KIT), Winter Semester 2014/15.

Rüdiger Pakmor

Computer Physics, Department of Physics and Astronomy, Heidelberg University (April 2014 - July 2014).
Cosmic Explosions, Department of Physics and Astronomy, Heidelberg University (October 2014 - February 2015).

Christoph Pfrommer

Cosmology, Department of Physics and Astronomy, Heidelberg University (October 2013 - February 2014).
Cosmology, Department of Physics and Astronomy, Heidelberg University (October 2014 - February 2015).

Andreas Reuter

Data Driven Science, Heidelberg University (May - June 2014).

Michael Schick

„Nonlinear Optimization“. Uni Heidelberg. WS 14/15

Volker Springel

Fundamentals of Simulation Methods, Department of Physics and Astronomy, Heidelberg University (October 2013 - February 2014).

Alexandros Stamatakis, Andre Aberer, Tomas Flouri, Alexey Kozlov, Kassian Kobert

Introduction to Bioinformatics for Computer Scientists, Department of Computer Science, Karlsruhe Institute of Technology (summer term 2014).

Introduction to Bioinformatics for Computer Scientists, Department of Computer Science, Karlsruhe Institute of Technology (winter term 2014/15).

Michael Strube

PhD Colloquium, Department of Computational Linguistics, Heidelberg University (October 2013 - February 2014).

Seminar: “Opportunities and Risks of Natural Language Processing”, Department of Computational Linguistics, Heidelberg University (October 2013 - February 2014).

PhD Colloquium, Department of Computational Linguistics, Heidelberg University (April 2014 - July 2014).

Blockseminar: “From Word Sense to Concept Disambiguation”, Computer Science Department, Korea Advanced Institute for Science and Technology, Daejeon, Korea, May 2013.

Rebecca Wade

Module 4,“ Protein Dynamics and Biomolecular Recognition: Insights from Simulations”, M.Sc. Molecular Cell Biology, Heidelberg University, 20.02.2014.

Module 3,“Protein Modeling”, M.Sc. Molecular Cell Biology, Heidelberg University, 12-13.05.2014.

Ringvorlesung “Structure and Dynamics of Biological Macromolecules”, “Electrostatics, solvation and protein interactions”, B.Sc. Biosciences, Heidelberg University, 08.07.2014.

Ringvorlesung „Biophysik“,“Receptor-Ligand Interactions: Structure and Dynamics”, B.Sc. Molecular Biotechnology, Heidelberg University, 27.11.2014.

COURSES AND SEMINARS

Jonathan Fuller, Stefan Henrich, Daria Kokh, Stefan Richter, Julia Romanowska, Rebecca Wade
HBIGS Practical Course “Computational analysis of protein binding properties”, HBIGS Postgraduate program, University of Heidelberg, ZMBH, 3-4 April 2014.

Anna Feldman-Salit, Jonathan Fuller, Stefan Richter, Julia Romanowska, Antonia Stank, Rebecca Wade
B.Sc. Biosciences Practical Course “Grundkurs Bioinformatik”, University of Heidelberg, Bioquant, 27-31 January 2014.

Tilmann Gneiting
Seminar on “Statistical Forecasting”, Karlsruhe Institute of Technology (KIT), Winter Semester 2013/14.
Tutorial on “Spatial Statistics for Energy Challenges”, King Abdullah University of Science and Technology (KAUST), Saudi Arabia, March 8, 2014.
Seminar on “Spatial Statistics”, Karlsruhe Institute of Technology (KIT) (with Eva Ochsenreither), Summer Semester 2014.

Frauke Gräter
HBIGS course “Deciphering protein function by molecular simulations”, Heidelberg, Germany, Apr 2-3.
Two-day course “Biomolecular function from multi-scale simulations” within Master´s programme, Universidad Politécnica de Madrid, Spain, March 26-27, 2014.

Vincent Heuveline
“Seminar: Wissenschaftliches Arbeiten”. WS 2014/15 Uni Heidelberg. Ausgerichtet von HITS, 27. November 2014
„Ringvorlesung: Wer regiert das Internet“. Uni Heidelberg, 13 November 2014
„Modelling and Numerical Methods for Uncertainty Quantification“. French-German Summer School Ecole Thématique CNRS. Porquerolles Island, France, September 1-5, 2014.

Andreas Reuter
Course on “Wissenschaftliches Arbeiten“ (in German), Heidelberg University, Heidelberg, winter semester 2014/2015.

Volker Springel
Experimental Physics I, tutoring classes, Department of Physics and Astronomy, Heidelberg University (October 2014 - February 2015).

Alexandros Stamatakis
Main Seminar “Hot Topics in Bioinformatics”, Department of Computer Science, Karlsruhe Institute of Technology (winter term 2014/15).
Hands-on Bioinformatics Programming Practical, Department of Computer Science, Karlsruhe Institute of Technology (winter term 2014/15).

Alexandros Stamatakis, Andre Aberer
Computational Molecular Evolution Summer School, Hellenic Center for Marine Research, Heraklion, Greece, 5-14 May 2014.

Michael Strube
Tutorial on “Entity Linking”. Computer Science Department, Sogang University, Seoul, Korea, May 12th, 2014.

Andre J. Aberer

"Big Computers for Big Trees: Challenges of Petascale Phyloinformatics", Pawsey Petascale Bioinformatics Symposium, Perth, Australia, November 2014.

Agnieszka Bronowska

"Protein dynamics, halogen-bonding, and structure-based design". University of Durham, Faculty of Chemistry. November 2014.

"The devil is in detail. Towards novel treatments for Alzheimer's Disease: KP inhibitors and AHR antagonists". Evotec Ltd, November 2014.

Tomas Flouri

"Merging Illumina paired-end reads", FEW/FAIR 2014, Stellenbosch, South Africa, December 2014.

Tilmann Gneiting

"Statistische Nachbearbeitung von Ensemble-Vorhersagen", Karlsruher Meteorologisches Kolloquium, Karlsruher Institut für Technologie, Karlsruhe, Germany, February 4, 2014.

"Statistische Nachbearbeitung von Ensemble-Vorhersagen", Deutsche Meteorologische Gesellschaft, Deutscher Wetterdienst, Offenbach, Germany, April 23, 2014.

"Uncertainty quantification in complex simulation models using ensemble copula coupling", FNRB SNAPLE Closing Workshop, Politecno di Milano, Milano, Italy, May 15-16, 2014.

"Quantifying uncertainty in simulation models using ensemble copula coupling", Innsbruck University, Innsbruck, Austria, May 21, 2014.

"From risk measures to predictive distributions", International Workshop on Systemic Risk and Regulatory Market Risk Measures, Parmenides Foundation, Pullach, Germany, June 2-3, 2014.

"Combining predictive distributions", Workshop on Uncertainty and Probabilistic Forecasting During the Financial and Economic Crisis, Heidelberg, Germany, June 20-21, 2014.

"Uncertainty quantification in complex simulation models using ensemble copula coupling", 22nd Meeting of the Belgian Statistical Society, Louvain-la-Neuve, Belgium, November 5-6, 2014.

"Thinking aloud about the Basel Accord", Heidelberg-Mannheim Stochastics Colloquium, Mannheim, Germany, November 21, 2014.

Martin Golebiewski

"From Grassroots Initiatives to Approved Standards", ISO International Conference on Standardization and Innovation, CERN, Geneva, November 13 – 14, 2014.

Frauke Gräter

"Mechanochemistry: bonds under force", Seminar, Frankfurt University, Chemistry Dept, Frankfurt, Germany, January 13, 2014.

"Molecular force sensors", Linz Winter Workshop, February 2-4.

"Unfolding and folding under forces", Heraeus Seminar "Protein folding and assmebly", Bad Honnef, Germany, February 2-6, 2014.

"Entropic allostery in gene expression factors", Cecam Workshop "Entropy" Vienna, Austria, March 14-17, 2014.

"Protein allosteric regulation from force distribution analysis", CCP-BioSim Conference, "Frontiers of Biomolecular Simulation" Edinburgh, UK, March 21-23, 2014.

"Multi-scale modeling of biomaterials: silk", Institute Seminar, Technical University Eindhoven, Netherlands, June 13, 2014.

"Mechano-sesning proteins in adhesion and blood", Gordon Research Conference "Single molecules", Barga, Italy, July 13-20, 2014.

"Mechano-sensing proteins in adhesion and blood", International Titisee conference, Titsee, Germany, October 8-12, 2014.

Daria Kokh

"Study of slow conformational changes in proteins", TSR Le Houches, France, May 18-23, 2014.

"Modeling of Protein Adsorption on Solid Surfaces using Brownian and Molecular Dynamics Simulations", CECAM meeting, Toulouse, France, March 23-26, 2014

Davide Mercadante

"Floppy yet fast. A sampling-independent binding mechanism for a mesh-forming intrinsically disordered protein". Joint workshop by Imperial College London and Istituto di Biostrutture e Bioimmagini "Approaching intricate biological processes by NMR and ancillary techniques", Napoli, Italy, June 20, 2014.

Wolfgang Müller

SEEK: Datenmanagement aus Software und Service, 1st RADAR Workshop, Karlsruhe, 16 September 2014.

Christoph Pfrommer

"Schwarze Löcher im Universum", Max-Planck Institute for Astronomy, Heidelberg, Germany, October 2014.

"The Physics and Cosmology of TeV Blazars", Würzburg University, Würzburg, Germany, July 2014.

"The Flea Circus", Heidelberg University, Heidelberg, Germany, May 2014.

"The Physics and Cosmology of TeV Blazars", Heidelberg University, Heidelberg, Germany, January 2014.

"The Physics of the Boomerang", Habilitation Talk, Heidelberg University, Heidelberg, Germany, January 2014.

Kai Polsterer

"Machine Learning in Astronomy: why the data deluge is not just pain" Harvard-Heidelberg Meeting on Star Formation, Heidelberg, Germany, June 2014.

"Machine Learning in Astronomy: lessons learned from learning machines" Splinter on E-Science & Virtual Observatory at the Annual Meeting of the Astronomische Gesellschaft, Bamberg, September 2014.

"Machine Learning in Astronomy: examples of data-driven science" Astrophysical Colloquium, Naples, Italy, November 2014.

Andreas Reuter

Welcome Address, 25 years IPVS at Stuttgart University, May 9, 2014.

"Computational Science - Was hat Informatik damit zu tun?" Tag der Informatik, Heidelberg University, Germany, June 27, 2014.

"Do young scientists need role models?" – Impressions of the 1st Heidelberg Laureate Forum, Schloss Dagstuhl , July 18, 2014.

"Profilbildung – Was hat das mit Bildung zu tun?", Fraunhofer ITWM, Kaiserslautern, December 16, 2014.

Christine Simpson

"Feedback and metal enrichment in cosmological models of dSph analogues" MPIA Theory Seminar, Heidelberg, Germany, April 10, 2014.

Volker Springel

"Simulationen der kosmischen Strukturbildung", Deutsches Museum Bonn, Bonn, November 2014.

“Hydrodynamical simulations of galaxy formation”, Keynote Lecture CAASTRO Workshop, Brisbane, Australia, November 2014.

“Forming the Milky Way Galaxy on a Supercomputer” , Physics Colloquium, University of Duisburg-Essen, Duisburg, January 2014.

“Simulierte Universen: Können wir Computermodellen trauen?”, Deutsch-Amerikanisches Institut Heidelberg, Heidelberg, Germany, November 2014.

“Simulating the Universe” , Keynote Human Brain Project Summit Meeting, Heidelberg, Germany, September 2014.

“Forming the Milky Way Galaxy on a Supercomputer”, Keynote International Conference of Physics Students, Heidelberg, Germany, August 2014.

“Supercomputer Simulationen der kosmischen Strukturentstehung”, Informatik Tag, Universität Heidelberg, Heidelberg, Germany, June 2014.

“Hydrodynamical simulations of galaxy formation”, Biermann Lectures, Max-Planck Institute for Astrophysics, Garching, June/July 2014.

“Promises and perils of scientific simulation”, Biermann Lectures, Max-Planck Institute for Astrophysics, Garching, June/July 2014.

“The feedback conundrum: What physics regulates galaxy and star formation?”, Biermann Lectures, Max-Planck Institute for Astrophysics, Garching, June/July 2014.

“Forming the Milky Way on a Supercomputer”, Astronomy Colloquium, Ludwig-Maximilians-Universität München, Munich, January 2014.

“Forming the Milky Way Galaxy on a Supercomputer”, Astrophysical Colloquium, École normale supérieure de Lyon, Lyon, March 2014.

“Tessellating the Universe: Cosmic Structure Formation on a Moving Mesh” , Maison de la Simulation Colloquium, Paris/Orsay, March 2014.

“Forming the Milky Way Galaxy on a Supercomputer”, Physics Colloquium, University of Würzburg, Würzburg, June 2014.

Alexandros Stamatakis

“My personal view of Bioinformatics“, Institut de Biologie Computationelle, Montpellier, France, May 2014.

“Inference and Post-Analysis of Huge Phylogenies“, Annual Society for Bioinformatics in Northern Europe (SocBiN) conference, Oslo, Norway, June 2014.

“Inference and Post-Analysis of Huge Phylogenies“, IWR seminar, University of Heidelberg, Germany, November 2014.

Michael Strube

”Issues I Don‘t Understand about Coreference and Coherence”, University of Edinburgh, UK, February 28, 2014.

”The Dark Side of NLP: An Introduction into Natural Language Processing for Spooks, Stalkers, and other Scoundrels“, University of Mannheim, Germany, March 24, 2014.

”The Dark Side of NLP: When Linguistics is Used to Monitor and Profile You“, Regionalgruppe Heidelberg der Deutschen Gesellschaft der Humboldtianer, July 9, 2014.

”The Dark Side of NLP: When Linguistics is Used to Monitor and Profile You“, Computational Linguistics Department, University of Heidelberg, Germany, November 13, 2014.

Rebecca Wade

"From Protein Diffusion to Protein Assembly: Putting Simulation and Experiment together", Winter Modeling Workshop, University of Modena and Reggio Emilia, Modena, Italy, March 13-14, 2014.

"From Protein Diffusion to Protein Assembly: Putting Simulation and Experiment together", Soft Matter Physics Group Seminar, Department of Physics, University of Leeds, Leeds, UK, March 21, 2014.

"From Protein Diffusion to Protein Assembly and Binding: Insights from Simulations", at „Binding free energy and kinetics: computation meets experiments" CECAM workshop, IIT, Genoa, Italy, June 10 -12 2014.

"Computational Approaches for Studying Drug Binding Kinetics", Satellite Meeting on "Computation for K4DD", Sanofi, Frankfurt, July 1, 2014.

"From Protein Diffusion to Protein Assembly and Binding: Insights from Simulations" Center for Multiscale Theory and Computation, Department of Chemistry, University of Münster, July 16, 2014.

"Macromolecular Brownian Dynamics Simulation", EMBO Practical Course on Biomolecular Simulation, Pasteur Institute, France, July 26, 2014.

"Organism-adapted Specificity of Allosteric Regulation of Central Metabolic Enzymes in Lactic Acid Bacteria", Biophysical Society Thematic Meeting "Modeling of Biomolecular Systems Dynamics, Allostery and Regulation: Bridging Experiments and Computations" Istanbul, Turkey, September 10-14, 2014.

"Computational exploration of allosteric sites and transient pockets on proteins", at 3rd International Drug Design Conference on „Drug Design 2014: fragment- and ligand-based drug design", St. Hilda's College, Oxford University, UK, September 23-25, 2014.

"From Protein Diffusion to Protein Assembly and Binding: Insights from Simulations", International Conference on „Virus and Cells – Computational Challenges and Approaches", IWR, Heidelberg University, Germany, October 10-11, 2014.

Denis Yurin

"A new method for the construction of N-body galaxy models in collisionless equilibrium", SFB Seminar at ARI, Heidelberg, Germany, November 5, 2014.

"A new method for the construction of N-body galaxy models in collisionless equilibrium", Blackboard Colloquium at ITA, Heidelberg, Germany, November 17, 2014.

DEMOS

Lihua An, Meik Bittkowski, Martin Golebiewski, Olga Krebs, Quyen Nguyen, Ivan Savora, Andreas Weidemann
SEEK & hands-on, 6th Virtual Liver Data Management and PALs Meeting, HITS, Heidelberg, Germany, May 8 – 9, 2014.

Jonathan Fuller
"LigDig: a web application for investigating ligand-protein interactions. LigDig can be used to query structural and functional properties", Systems Biology Data Management Foundry: A meeting for practitioners, Villa Bosch Studio, October 6-7, 2014.

Martin Golebiewski, Ursula Kummer, Jürgen Pahle, Jacky Snoep, Natalie Stanford
"JWS Online, SEEK, SABIO-RK and COPASI", COMBINE & ERASysAPP tutorial "Modelling and Simulation of Biological Models", Melbourne, Australia, September 14, 2014

Martin Golebiewski, Wolfgang Müller
"The PORTAL into Virtual Liver", Annual Meeting Virtual Liver Network WP A2, B and F, Heidelberg, Germany, April 1 – 2, 2014.

Iryna Ilkavets, Ivan Savora
The Virtual Liver Portal, 6th Virtual Liver Data Management and PALs Meeting, HITS, Heidelberg, Germany, May 8 – 9, 2014.

Wolfgang Müller, Olga Krebs
Computer practicals "Data mangement in practice", Advanced Lecture Course on Systems Biology, SysBio2014, Innsbruck, Austria, March 8 – 12, 2014.

Stefan Richter
„Ob Groß oder Klein, Biologische Vielfalt gibt es überall", Explore Science, Luisenpark Mannheim, July 9-13, 2014.

"Ligand egress from Cytochrome P450", Forschung digital – vom Molekül bis zum Universum, MS-Wissenschaft Mannheim, August 6, 2014.

POSTERS

Leandro Almeida, Neil Bruce, Paolo Carloni, Elisa Frezza, Omar Gutierrez-Arenas, Jeanette Hellgren-Kotaleski, Daniel Keller, Richard Lavery, Pierre Magistretti, Henry Markram, Anu Nair, Antoine Triller, Pietro Vidossich, Rebecca Wade
"Subcellular and Molecular Simulations", Human Brain Project Annual Summit 2014, Heidelberg, Germany, September 28 – Oct 1, 2014.

Camilo Aponte-Santamaria and Frauke Gräter
"Force-dependent auto-inhibition of the von Willebrand factor mediated by inter-domain interactions". Workshop on computer simulations and theory of macromolecules. Huenfeld, Germany, April 11-12, 2014.

Neil Bruce, Rudi Tong, Paolo Carloni, Omar Gutierrez-Arenas, Jeanette Kotaleski, Richard Lavery, Anu Nair, Rebecca Wade
"Constraining the kinetic parameters of signaling pathways through molecular-level modelling: isoform specific regulation of adenylyl cyclase", FENS Satellite Meeting: Biologically-based models of neurons and microcircuits, Pavia, Italy, July 4, 2014.

Kira Feldmann, Mikyoung Jun and Tilmann Gneiting
"Spatial postprocessing for forecasts of temperature minima and maxima", Summer School on the Modelling and Prediction of Weather Extremes, Annweiler, Germany, June 10-13, 2014.

"Spatial postprocessing for forecasts of temperature minima and maxima", International Symposium Extremes 2014, Hannover, Germany, October 6-7, 2014.

"Spatial postprocessing for forecasts of temperature minima and maxima", Workshop on High-Dimensional, High-Frequency and Spatial Data, Karlsruhe, Germany, October 29-31, 2014.

Tomas Flouri

"PLL: A software library for rapid development of phylogenetic applications", Poster at Symposium of the SMBE, San Jose, USA, June 2014.

Jonathan Fuller, Michael Martinez, Stefan Henrich, Stefan Richter, Rebecca Wade

"LigDig: a web application for investigating ligand-protein interactions. LigDig can be used to query structural and functional properties", Systems Biology of Mammalian Cells 2014, Berlin, Germany, May 12-14, 2014.

Martin Golebiewski, Olga Krebs, Stuart Owen, Katy Wolstencroft, Quyen Nguyen, Lihua An, Meik Bittkowski, Ivan Savora, Andreas Weidemann, Natalie J. Stanford, Dawie van Niekerk, Franco du Preez, Jacky L. Snoep, Wolfgang Müller and Carole Goble

"Data Needs Structure: Data and Model Management for Distributed Systems Biology Projects", 5th Conference on Systems Biology of Mammalian Cells (SBMC), Berlin, Germany, May 12 – 14, 2014.

Martin Golebiewski, Natalie J. Stanford, Lihua An, Meik Bittkowski, Olga Krebs, Stuart Owen, Quyen Nguyen, Ivan Savora, Dawie van Niekerk, Andreas Weidemann, Katy Wolstencroft, Jacky L. Snoep, Wolfgang Müller and Carole Goble

"Data Needs Structure: Data and Model Management for Distributed Systems Biology Projects", 15th International Conference on Systems Biology (ICSB), Melbourne, Australia, September 14 – 18, 2014.

Stephan Hemri, Dimitri Lisniak and Bastian Klein

"Addressing the uncertainty of extremal river discharge forecasts by multivariate calibration techniques", International Symposium Extremes 2014, Hannover, Germany, October 6-7, 2014.

Iryna Ilkavets, Martin Golebiewski, Ivan Savora, Meik Bittkowski, Elina Wetsch, Jill Zander, Wolfgang Müller

"The Virtual Liver Portal: Showcasing Scientific Data for the Non-Expert Public", 5th Conference on Systems Biology of Mammalian Cells (SBMC), Berlin, Germany, May 12 – 14, 2014.

Alexander Jordan

"Tests for equal predictive accuracy using proper scoring rules", Workshop on High-Dimensional, High-Frequency and Spatial Data, Karlsruhe, Germany, October 29-31, 2014.

Renate Kania

"SABIO-RK: New Features", Gordon Conference on Enzymes, Coenzymes and Metabolic Pathways - Enzymatic Catalysis in Health and Disease, Waterville Valley, NH, USA, July 13 – 18, 2014.

Daria Kokh, Stefan Richter Stefan Henrich, Paul Czodrowski , Friedrich Rippmann, Rebecca Wade

"Trapp - a tool for simulation of protein cavity dynamics and identification of transient binding pockets in proteins" EuroQSAR, St.Petersburg, Russia, September 1-4, 2014.

Fabian Krüger, Todd E. Clark and Francesco Ravazzolo

"Combining survey and Bayesian VAR forecasts of US macro variables: Evidence from entropic tilting", Workshop on Uncertainty and Economic Forecasting, London, UK, May 8-9, 2014.

"Using entropic tilting to combine BVAR forecasts with external nowcasts", 25th (EC)² Conference on Advances in Forecasting, Barcelona, Spain, December 12-13, 2014.

Fabian Krüger, Sebastian Lerch, Thordis Thorarinsdottir and Tilmann Gneiting

"Probabilistic forecasting based on MCMC output", International Symposium Extremes 2014, Hannover, Germany, October 6-7, 2014.

"Probabilistic forecasting based on MCMC output", Workshop on High-Dimensional, High-Frequency and Spatial Data, Karlsruhe, Germany, October 29-31, 2014.

Dennis Kügler, Kai Polsterer, Maximilian Hoecker

"Spectral Redshift Estimates using k-Nearest-Neighbors Regression." Astroinformatics 2014, Valparaiso, Chile, August 25-29, 2014.

Dennis Kügler, Kai Polsterer, Maximilian Hoecker

"Spectral Redshift Estimates using k-Nearest-Neighbors Regression." Splinter on E-Science & Virtual Observatory at the Annual Meeting of the Astronomische Gesellschaft, Bamberg, Germany, September 22-26, 2014.

Sebastian Lerch, Thordis Thorarinsdottir, Francesco Ravazzolo and Tilmann Gneiting

"Forecaster's dilemma: Extreme events and forecast evaluation", Summer School on the Modelling and Prediction of Weather Extremes, Annweiler, Germany, June 10-13, 2014.

"Forecaster's dilemma: Extreme events and forecast evaluation", Workshop on Propriety and Elictability, Heidelberg, Germany, June 18, 2014.

"Forecaster's dilemma: Extreme events and forecast evaluation", International Symposium Extremes 2014, Hannover, Germany, October 6-7, 2014.

"Forecaster's dilemma: Extreme events and forecast evaluation", Workshop on High-Dimensional, High-Frequency and Spatial Data, Karlsruhe, Germany, October 29-31, 2014.

Davide Mercadante, Sigrid Milles, Edward A. Lemke and Frauke Gräter

"A new mechanism for the binding of intrinsically disordered proteins to structured partners". Gordon Research Conference on Intrinsically disordered proteins. Boston, USA, July 6-11, 2014.

Ghulam Mustafa, Antonia Stank, Prajwal P. Nandekar, Xiaofeng Yu and Rebecca C. Wade

"Modeling and Simulation of Macromolecular Complexes". 2nd DKFZ-ZMBH Alliance Retreat, Kloster Schöntal, Germany, July 16-18, 2014.

Ghulam Mustafa, Prajwal Nandekar, Xiaofeng Yu and Rebecca C. Wade

"Modeling and Simulation of Cytochrome P450 - Membrane, Substrate and Product Interactions". EMBO Practical Course on Computational Structural Biology, Paris France, 20-27 July, 2014.

Prajwal Nandekar, Ghulam Mustafa, Xiaofeng Yu, Abhay Sangamwar and Rebecca C. Wade

"Modeling and Simulation of Cytochrome P450 - Membrane, Substrate and Product Interactions". 2nd International Symposium on Microsomes and Drug Oxidation, Stuttgart, Germany, May 18-22, 2014.

Musa Özboyaci, Antonia Stank, Daria Kokh, Rebecca C. Wade

„Molecular and Cellular Modeling", IWR and HGS MathComp, Heidelberg, Germany, November 17-18, 2014.

Ina Pöhner, Egle Maximowitsch, Talia Zeppelin, Stefan Henrich, Rebecca C. Wade

"Towards Improved Selective Pteridine Reductase Inhibi-

tors As Antiparasitic Agents". Annual meeting of the COST Action CM1307 Chemotherapy towards diseases caused by endoparasites, Calvi, France, October 27-29, 2014.

Kai Polsterer, Fabian Gieseke, Christian Igel
"Automatic Classification of Galaxies via Machine Learning Techniques: parallelized rotation/flipping invariant Kohonen map." ADASS 2014, Calgary, Canada, October 5-9, 2014.

Maja Rey, Ulrike Wittig, Lei Shi, Meik Bittkowski, Renate Kania, Wolfgang Müller
"SABIO-RK database: new features", 7th International Biocuration Conference, Toronto, Canada, April 6 – 9, 2014.

Julia Romanowska, Daria Kokh and Rebecca C. Wade
"New insights into the process of lysozyme adsorption to surfaces from Brownian dynamics simulations", ISQBP 2014 President Meeting: Challenges and celebrations of biomolecular simulation, Telluride, Colorado, USA, June 15-18, 2014.

Roman Schefzik, Thordis Thorarinsdottir and Tilmann Gneiting
"Physically coherent probabilistic weather forecasts via ensemble copula coupling", Summer School on the Modelling and Prediction of Weather Extremes, Annweiler, Germany, June 10-13, 2014.

"Modeling dependencies in probabilistic weather forecasting: Member-by-member postprocessing and ensemble copula coupling", International Symposium Extremes 2014, Hannover, Germany, October 6-7, 2014.

"Modeling dependencies in probabilistic weather forecasting: Member-by-member postprocessing and ensemble copula coupling", HGS MathComp Annual Colloquium 2014, Speyer, Germany, November 24-25, 2014.

Siegfried Schloissnig
"A genome-scale resource for in vivo tag-based protein function exploration in C. elegans.", 19th International C.elegans Meeting, University of California, Los Angeles, June 26-30, 2013.

"C. briggsae genomic fosmid library", 19th International C.elegans Meeting, University of California, Los Angeles, June 26-30, 2013.

Patrick Schmidt
"Identifying the elicitable functional behind a point forecast", Workshop on Propriety and Elicitability, Heidelberg, Germany, June 18, 2014.

Antonia Stank, Rebecca C. Wade
„Computational studies on the relation between macromolecular dynamics and protein binding and function", EMBO practical course on „Biomolecular simulation", Paris, France, July 20-27, 2014.

Antonia Stank, Daria B. Kokh, Stefan Richter, Jonathan Fuller, Rebecca C. Wade
"Computational studies on the relation between macromolecular dynamics and protein binding and function", HGS MathComp Annual Colloquium, Speyer, Germany, November 24-25, 2014.

Ulrike Wittig, Wolfgang Müller
"Data Management for Systems Biology", 2nd Data Management Workshop, Cologne, Germany, November 28 – 29, 2014.

Ulrike Wittig, Lei Shi, Meik Bittkowski, Maja Rey, Renate Kania, Martin Golebiewski, Wolfgang Müller
"Data upload into SABIO-RK via SBML", 7th International Biocuration Conference, Toronto, Canada, April 6 – 9, 2014.

Xiaofeng Yu, Ghulam Mustafa, Vlad Cojocaru and Rebecca C. Wade

"Multiscale Simulations of Cytochrome P450 Systems". Bayer IT for Life Science Workshop, Leverkusen, Germany, December 15-16, 2014.

TALKS

Andre J. Aberer

"Bayesian Tree Inference on Whole-Genome Datasets is Possible!", Symposium of the SMBE, San Jose, USA, June 2014.

Christian Arnold

"Galaxy cluster scaling relations and mass bias in f(R) gravity", Workshop on modified gravity simulations, Garching, Germany, April 9 - 11, 2014.

"Hydrodynamical simulations in f(R) gravity", Transregio Winter School, Passo del Tonale, Italy, December 7 - 12, 2014.

"Galaxy cluster scaling relations and the Lyman-alpha forest in f(R) gravity", The quest for Dark Energy II, Ringberg Castle, December 14 - 19, 2014.

"Hydrodynamical cosmological simulations in f(R) gravity", Virgo Consortium Meeting, Garching, Germany, December 17 - 19, 2014.

Andreas Bauer

"Comparing possible source models of reionization using GPUs", The 10th Sino-German Workshop on Galaxy Formation and Cosmology, Xian, China, May 18 – 23, 2014.

Meik Bittkowski

„From Spreadsheets to Standards", SeqAhead Workshop: "Managing Big Data", Berlin, Germany, July 9 – 11, 2014.

Werner Ehm

"Reproducibility from the perspective of meta-analysis", Reproducibility Conference, Munich, Germany, October 1-3, 2014.

Kira Feldmann

"Statistical postprocessing for TIGGE", World Weather Open Science Conference 2014, Montreal, Canada, August 16-21, 2014.

"Spatial post-processing for temperature forecasts", Minisymposium on Spatial Statistics, Heidelberg, Germany, October 28, 2014.

Tilmann Gneiting

"Statistical postprocessing of ensemble weather forecasts", Summer School on the Modelling and Prediction of Weather Extremes, Annweiler, Germany, June 10-13, 2014.

Martin Golebiewski

"COMBINE – The World Wide Web Consortium of Modelling in Biology", ERASysAPP/ISBE Workshop "Networking Systems Biology: Academia-Industry", Berlin, Germany, May 14 – 15, 2014.

HARMONY 2014: The Hackathon on Resources for Modeling in Biology, Manchester, UK, April 22 – 25, 2014.

"Setting the Standards for Data and Model Exchange in Systems Biology", SeqAhead Workshop: "Managing Big Data", Berlin, Germany, July 9 – 11, 2014.

"Standardisation in distributed research networks: The Virtual Liver Experience", SeqAhead Workshop: "Managing Big Data", Berlin, Germany, July 9 – 11, 2014.

"Upload of Kinetic Data into SABIO-RK via SBML", COMBINE 2014: 5th Computational Modeling in Biology Network Meeting, Los Angeles, California, USA, August 21, 2014.

"From grassroots to common standards", COMBINE 2014: 5th Computational Modeling in Biology Network Meeting, Los Angeles, California, USA, August 22, 2014.

"Bridging Experiments and Modelling: SABIO-RK - Reaction Kinetics Database", COMBINE & ERASysAPP tutorial "Modelling and Simulation of Biological Models" Melbourne, Australia, September 14, 2014.

"Setting the Standards for Data and Model Exchange in Systems Biology", NORMSYS & ISBE workshop "Standards for data and model exchange in systems biology", Melbourne, Australia, September 18, 2014.

Robert Grand

"Spiral arms in numerical simulations: observational predictions", Gaia Challenge 2, MPIA, Heidelberg, Germany, October 27-31, 2014.

"Radial migration in Spiral Galaxies", Virgo Consortium Meeting, MPA, Garching, Germany, December 18-19, 2014.

Christopher Hayward

"The total infrared luminosity may significantly overestimate the star formation rate of recently quenched galaxies" Digging Deep into the Extragalactic Infrared Sky, European Week of Astronomy and Space Sciences, Geneva, Switzerland, 30 June - 1 July 2014.

"The heterogeneity of the submillimetre galaxy population", The Formation and Growth of Galaxies in the Young Universe, Obergurgl, Austria, 26-30 April 2014.

Stephan Hemri

"Post-processing of multi-model ensemble river discharge forecasts using censored EMOS", 11th German Probability and Statistics Days (GPSD 2014), Ulm, Germany, March 4-7, 2014.

"Post-processing of multi-model river discharge forecasts using censored EMOS", European Geosciences Union General Assembly 2014, Vienna, Austria, April 27 - May 2, 2014.

Kassian Kobert

"The divisible load balance problem and its application to phylogenetic inference", 14th Workshop on Algorithms in Bioinformatics (WABI), Wrocław, Poland, September 2014.

Katra Kolšek

"An Open Source Prediction Tool for Assessing Endocrine Disruption Potential through Nuclear Receptors Binding". 28th Molecular Modeling Workshop 2014, Erlangen, Germany, March 17-19, 2014.

Fabian Krüger

"Combining survey and MCMC based forecasts of US macro variables", Workshop on Uncertainty and Probabilistic Forecasting During the Financial and Economic Crisis, Heidelberg, Germany, June 20-21, 2014.

"Combining density forecasts under various scoring rules: An analysis of UK inflation", 34th International Symposium on Forecasting, Rotterdam, The Netherlands, June 29 - July 2, 2014.

"Probabilistic forecasting based on MCMC output", 8th International Conference on Computational and Financial Econometrics, Pisa, Italy, December 6-8, 2014.

Alexey Kozlov

„Efficient Computation of the Phylogenetic Likelihood Function on the Intel MIC Architecture", IPDPS 2014, Phoenix, USA, May 2014.

Dennis Kügler

"Spectral Redshift Estimates using k-Nearest-Neighbors Regression", Splinter on E-Science & Virtual Observatory, Annual Meeting of the Astronomische Gesellschaft, Bamberg, September 2014.

Davide Mercadante

"An Unconventional Ultrafast Binding Mechanism of a Disordered Protein". Workshop on computer simulations and theory of macromolecules. Huenfeld, Germany, April 11-12, 2014.

Rüdiger Pakmor

"Magnetic fields in disks", Virgo Consortium Meeting, Leiden, Netherlands, January 22-24, 2014.

"SubCh Explosions and Uncertainties in Nucleosynthesis", 17th Workshop on Nuclear Astrophysics, Ringberg, Germany, April 7-12, 2014.

"Latest results for first principle explosion simulations of Type Ia Supernovae", The Unquiet Universe, Cefalu, Italy, June 9-14, 2014.

"Conservation of Angular Momentum", 6th Würzburg Workshop, Würzburg, Germany, July 10-11, 2014.

"MHD simulations with Arepo", Arepofest-2, Boston, USA, September 2-4, 2014.

"Type Ia Supernovae from double degenerate binary systems", Astronomische Gesellschaft Jahrestagung 2014, Bamberg, Germany, September 22-26, 2014.

"Magnetic fields in cosmological simulations of disk galaxies", Magnetisation of Interstellar and Intergalactic Media, Eitorf, Germany, September 29 - October 2, 2014

Christoph Pfrommer

"Cosmic ray feedback in galaxies and cool core clusters", Astrophysics of High-Beta Plasmas in the Universe, Jeju Island, Korea, November 10 - 13, 2014.

"Review: Gamma-ray Astronomy", Astroparticle Physics in Germany: Status and Perspectives, Karlsruhe Institute for Technology, Germany, September 30 - October 1, 2014.

"Cosmic ray heating in cool core clusters", 3rd ICM Theory and Computation Workshop, Niels Bohr Institute, Copenhagen, Denmark, August 11 - 14, 2014.

"Magnetic fields in galaxy clusters", ICM Inhomogeneities in the Intracluster Plasma, Stanford University/KIPAC, USA, July 28 - 30, 2014.

"Cosmic ray feedback from active galactic nuclei", ICM Inhomogeneities in the Intracluster Plasma, Stanford University/KIPAC, USA, July 28 - 30, 2014.

"Radio mode theory: mechanical versus cosmic-ray heating", Quenching and Quiescence, Max-Planck Institute for Astronomy, Heidelberg, Germany, July 14 - 18, 2014.

"How cosmic rays shape the faint- and bright-end of the galaxy population", Gravity's Loyal Opposition: The Physics of Star Formation Feedback, Kavli Institute for Theoretical Physics, Santa Barbara, USA, April 14 - July 3, 2014.

"Plenary talk: The Physics of Propagating TeV Gamma-rays: from Plasma Instabilities to Cosmological Structure Formation", DPG Spring Conference on Particle and Astroparticle Physics, Mainz University, Germany, March 24, 2014

Julia Romanowska

"New insights into the process of lysozyme adsorption to surfaces from Brownian dynamics simulations", ISQBP 2014 President Meeting: Challenges and celebrations of biomolecular simulation, Telluride, Colorado, USA, June 15-18, 2014.

Kevin Schaal

"A shock finder for Arepo", Virgo Consortium Meeting, Leiden, Netherlands, January 22 - 24, 2014.

"Finding and interpreting shocks in Illustris", Arepo Workshop, Cambridge, MA, USA, September 2 - 4, 2014.

"Discontinuous Galerkin Hydrodynamics with Adaptive Mesh Refinement", Virgo Consortium Meeting, Garching, Germany, December 18 - 19, 2014.

Christine Simpson

"Momentum feedback in dwarf galaxy simulations", Virgo Meeting 2014, Leiden, Netherlands, January 22 – 24, 2014.

"High resolution simulations of dwarf galaxies or High resolution modeling of ISM", Arepofest Workshop, Cambridge, MA, USA, September 2- 4, 2014.

Volker Springel

"Diffuse gas properties, stellar metallicities and magnetic fields in cosmological simulations of disc galaxy formation", Gas in and around galaxies, Ringberg Castle, May 12-16, 2014.

"Winds and their impact in simulations of Milky Way-sized galaxies", Fire Down Below: The Impact of Feedback on Star and Galaxy Formation, Santa Barbara, USA, April 14-18, 2014.

"Hydrodynamical simulations of galaxy formation", From dark matter to galaxies, Xi'an, China, May 18-23, 2014.

"Simulations of galaxy formation: The challenge to globally regulate star formation", Star formation: Data, Models and Visualization, House of Astronomy, Heidelberg, Germany, June 23-26, 2014.

"Hydrodynamical simulations of AGN and their clustering", Clustering Measurements of AGN, ESO Garching, July 14-18, 2014.

"Feedback in cosmological simulations with AREPO", Semianalytic Models and Hydrodynamic Simulations, Marseille, France, June 10-13, 2014.

Antonia Stank

„Modeling the domain interactions of Hsp40 proteins", DKFZ-ZMBH Alliance Retreat, Kloster Schöntal, Germany, July 16-18, 2014.

Dandan Xu

"CDM substructures and lensing flux ratio anomalies", Cambridge Dark Matter Substructure Workshop, Cambridge, UK, October 13-14, 2014.

"CDM substructures and lensing flux ratio anomalies", The 10th Sino-German Workshop on Galaxy Formation and Cosmology, China, May 18-23, 2014.

Xiaofeng Yu

"Using Computational Methods to Study Cytochrome P450 Systems". DKFZ-ZMBH Alliance seminar, Heidelberg, Germany, November 12, 2014

PANELS

Tilmann Gneiting, Volker Springel

"E-Science: Wissenschaft in der digitalen Gesellschaft", MS Wissenschaft, Mannheim, Germany, August 6, 2014.

Tilmann Gneiting

"Young researchers meet editors", 22nd Meeting of the Belgian Statistical Society, Louvain-la-Neuve, Belgium, November 5-6, 2014.

Martin Golebiewski

"From Grassroots Initiatives to Approved Standards", ISO International Conference on Standardization and Innovation, CERN, Geneva, November 13 – 14, 2014.

Camilo Aponte-Santamaria
Member of the Biophysical Society

Nikos Gianniotis
Member of the International Astrostatistics Association

Tilmann Gneiting
Editor: Physical Science, Computation, Engineering, and the Environment, Annals of Applied Statistics (until May 2014)

Senior Editor: Annals of Applied Statistics (since June 2014)

Guest Editor: Nonlinear Processes in Geophysics

Fellow, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading (UK)

Affiliate Professor, Department of Statistics, University of Washington, Seattle (USA)

Guest faculty member of the Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University

Associated faculty member of HGS MathComp Graduate School, Heidelberg University

Faculty member of Research Training Group 1653, Spatial/Temporal Probabilistic Graphical Models and Applications in Image Analysis, Heidelberg University
Faculty member of Research Training Group 1953, Statistical Modeling of Complex Systems and Processes: Advanced Nonparametric Methods, Heidelberg University and Mannheim University

IMS Representative, Committee of Presidents of Statistical Societies (COPSS) Awards Committee

Martin Golebiewski
Board member of the COMBINE network (Computational Modeling in Biology network)

German delegate at ISO (International Organization for Standardization) Technical committee 276 Biotechnology

Leader of the national German working group “Data Processing and Integration in Biotechnology”, German Institute for Standardization (DIN)

Member of the national German standardization committee (“nationaler Arbeitsausschuss”) NA 057–06–02 AA Biotechnology, German Institute for Standardization (DIN)

Frauke Gräter
Member of the Biophysical Society, the German Physical Society, the German Biophysical Society, the German Chemical Society

Member of BIOMS (Heidelberg Center for Modeling and Simulations in the Biosciences) Steering Committee

Faculty member, Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg

Faculty member, HGS MathComp Graduate School, University of Heidelberg

Faculty member, Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology (HBIGS), University of Heidelberg

Christopher Hayward
Member of the American Astronomical Society

Member of the American Physical Society

Member of the Association for the Advancement of Science

Memberships

Dennis Kügler
Member of Förderkreis Landessternwarte Heidelberg

Wolfgang Müller
Member of IEEE

Member of Gesellschaft für Informatik

Member of FB ILW (Informatik in den Lebenswissenschaften) of the GI

Member of the GMDS

Davide Mercadante
Member of the Biophysical Society

Christoph Pfrommer
Associate member of the LOFAR Magnetism Key Science Project

External collaboration member of the MAGIC Cherenkov Telescope Collaboration

External collaboration member of the Fermi Space Telescope Collaboration

Kai Polsterer
Member of the Astronomische Gesellschaft (A.G.)

Member of the International Astrostatistics Association

Member of the IEEE Task Force on Mining Complex Astronomical Data
Member of the Standing Committee on Science Priorities of the International Virtual Observatory Alliance

Member of the Knowledge Discovery in Databases Interest Group of the International Virtual Observatory Alliance

Andreas Reuter
Scientific Member of Max-Planck-Gesellschaft (Max Planck Institute of Computer Science, Saarbrücken)

Member of the Scientific Committee, BIOMS, Heidelberg

Member of the Advisory Board of Fraunhofer Gesellschaft Informations- und Kommunikationstechnik (IuK)

Member of the Heidelberg Club International

Member of the Board of Trustees of the Wissenschaftspressekonferenz, Bonn

Co-editor: Database Series, Vieweg-Verlag

Member of a research group on scientific computing named “WIR”

Member of Dagstuhl`s Industrial Curatory Board of „Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI)“, Dagstuhl (Leibniz Center for Computer Science)- bis Mai 2015

Member of Schloss Dagstuhl´s Scientific Advisory Board

Chairman of the Supervisory Board of SICOS GmbH, Stuttgart

Member of the Board of Directors at IWR, University of Heidelberg

Member of the search committee for a professorship on “Scientific Visualization” at University of Heidelberg

Christine Simpson
Member of the American Astronomical Society

Volker Springel
Member of the Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University

External Scientific Member of the Max-Planck-Institute for Astronomy, Heidelberg.
Member of the Interdisciplinary Astronomical Union (IAU).

Member of the Cosmological Simulation Working Group (CSWG) of the EUCLID mission of ESA.

Member of the Research Council of the Field of Focus "Structure and pattern formation in the material world" at Heidelberg University

Member of the Board of SFB 881 "The Milky Way System"

Member of the Scientific Advisory Board of the Gauss Centre for Supercomputing (GCS)

Alexandros Stamatakis
Member of the steering committee of the Munich Supercomputing System HLRB at LRZ

Member of the scientific advisory board of the Computational Biology Institute in Montpellier, France

Michael Strube
Editorial Board: Dialogue & Discourse Journal; The Journal of Data Semantics

Rebecca Wade
Associate Editor: Journal of Molecular Recognition, PLOS Computational Biology

Section Editor: BMC Biophysics

Editorial Board: BBA General Subjects; Journal of Computer-aided Molecular Design; Biopolymers; Current Chemical Biology; Protein Engineering, Design and Selection; Computational Biology and Chemistry: Advances and Applications; Open Access Bioinformatics

Member of Scientific Advisory Council of the Leibniz-Institut für Molekulare Pharmakologie (FMP), Berlin-Buch

Member of BIOMS Steering Committee, Heidelberg

Member at Heidelberg University of: CellNetworks Cluster of Excellence, HBIGS (Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology) faculty, HGS MathComp faculty, IWR, DKFZ-ZMBH Alliance

Member of CellNetworks Postdoctoral Applications Evaluation Committee

Mentor, BioMedX, Heidelberg, "Selective Kinase Inhibitors" Team

9.4 Contributions to the Scientific Community

REFEREE WORK

Andre J. Aberer
Bioinformatics

Werner Ehm
Journal of Approximation Theory

Tomas Flouri
Science

Bioinformatics

BMC Bioinformatics

Nucleic Acids Research

PLoS ONE

Theoretical Computer Science
Information Processing Letters

Algorithmica

Acta Informatica

Discrete Mathematics

Jonathan Fuller
Journal of Molecular Recognition

Nikos Gianniotis
Pattern Analysis and Application

Tilmann Gneiting
Annals of Statistics

Finance and Stochastics
Journal of Risk

Journal of Statistical Software

Meteorology and Atmospheric Physics

Quarterly Journal of the Royal Meteorological Society

Stat

Martin Golebiewski
Bioinformatics

BMC Systems Biology

CPT: Pharmacometrics & Systems Pharmacology

DATABASE - The Journal of Biological Databases and Curation

Robert Grand
Monthly Notices of the Royal Astronomical Society
Frauke Gräter
Biophysical Journal

Journal of the American Chemical Society

Journal for Physical Chemistry B

Nature Journals

Proceedings of the National Academy of Sciences

German Research Society (DFG)

PRACE

Christopher Hayward
Monthly Notices of the Royal Astronomical Society
Astrophysical Journal

Stephan Hemri
Hydrologie und Wasserbewirtschaftung

Daria Kokh
The European Physical Journal D

BMC Bioinformatics

Fabian Krüger
Annals of Applied Statistics

International Journal of Forecasting

Sebastian Martschat
Language Resources and Evaluation Journal

Wolfgang Müller
F1000 Research

Mehmet Öztürk
International Food Research Journal

Rüdiger Pakmor
Astrophysical Journal

Monthly Notices of the Royal Astronomical Society

Science

National Science Foundation, USA

Christoph Pfrommer
Astrophysical Journal

Astrophysical Journal Letters

Monthly Notices of the Royal Astronomical Society
Physical Review Letters
Physical Review D

Ina Pöhner
Journal of Molecular Recognition

Kai Polsterer
Annals of Applied Statistics

Astronomy and Computing

Siegfried Schloissnig
Biotechnology and Biological Sciences Research Council (BBSRC)

Bioinformatics

BMC Evolutionary Biology

RECOMB2014 (18th Annual International Conference on Research in Computational Molecular Biology)

Christine Simpson
Monthly Notices of the Royal Astronomical Society

Astrophysical Journal

Volker Springel
Monthly Notices of the Royal Astronomical Society

Nature

Max-Planck Society

European Research Council

National Science Center Poland
Department of Energy, USA

National Research Foundation, South Africa
Reseach Councils UK
Bayerisches Staatsministerium für Wissenschaft, Forschung und Kunst

Jülich Supercomputer Center

Leibniz Computing Centre

German-Israel Science Foundation

Alexandros Stamatakis
IEEE Transactions on Computational Biology and Bioinformatics

Bioinformatics

BMC Bioinformatics

Systematic Biology

Michael Strube
Computational Linguistics Journal

Journal of Artificial Intelligence Research

Dandan Xu
Monthly Notices of the Royal Astronomical Society

PROGRAMM COMMITTEE MEMBERSHIPS

Angela Fahrni
14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, April 26-30, 2014.

52nd Annual Meeting of the Association for Computational Linguistics (Area Chair), Baltimore, Md., USA, June 22-27, 2014.

Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, October 25-29, 2014.

Tomas Flouri
String Processing and Information Retrieval, Ouro Preto, Brasil, October 21-24, 2014.

Combinatorial Pattern Matching 2014, Moscow, Russi, June 16-18, 2014.

International Supercomputing Conference 2014, Leipzig, Germany, June 22-26, 2014.

International Workshop on Data Mining in Bioinformatics, New York, USA, August 24-27, 2014.

Jonathan Fuller
Heidelberg Unseminars in Bioinformatics (HUB), Heidelberg, Germany, 2014.

Xiaofeng Yu
Heidelberg Unseminars in Bioinformatics (HUB), Heidelberg, Germany, 2014.

Tilmann Gneiting
International Symposium Extremes 2014, Hannover, Germany, October 6-7, 2014.

Workshop on High-Dimensional, High-Frequency and Spatial Data, Karlsruhe, Germany, October 29-31, 2014.

Sebastian Martschat
Conference on Natural Language Processing (KONVENS ,14), Hildesheim, Germany, October 8-10, 2014.

Wolfgang Müller
Workshop of Ontologies and Data in Life Sciences (ODLS 2014); Freiburg, Germany, October 7-8, 2014.

Andreas Reuter
Deutsche Forschungsgemeinschaft; Fonds zur Förderung der wissenschaftlichen Forschung (Österreich).

Member of the Scientific Committee of the 2nd Heidelberg Laureate Forum, Heidelberg, Germany, September 21-26, 2014.

Alexandros Stamatakis
Pattern Recognition in Bioinformatics 2014, Stockholm, Sweden, August 21-23, 2014.

13th European Conference in Computational Biology, Strasbourg, France, September 7-10, 2014.

International Conference on Communication and Information Systems (ICICS 2014), Ibrid, Jordan, April 1-3, 2014.

International Supercomputing Conference 2014, Leipzig, Germany, June 22-26, 2014.

13th IEEE International Workshop on High Performance Computational Biology, Pohenix, Arizona, May 19, 2014.

Volker Springel
From Dark Matter to Galaxies, Xi'an, China, May 18 - 23, 2014.

Star Formation: Data, Models and Visualization, Heidelberg, Germany, June 23 - 26, 2014.

Michael Strube
25th International Conference on Computational Linguistics, Dublin, Ireland, August 23-29, 2014.

14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, April 26-30, 2014.

Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, October 25-29, 2014.
Third Joint Conference on Lexical and Computational Semantics (StarSEM ,14), Dublin, Ireland, August 23-24, 2014.

Rebecca Wade
Scientific Advisory Board, "Modeling of Biomolecular Systems Interactions, Dynamics, and Allostery: Bridging Experiments and Computations", Biophysical Society Thematic Meeting, Istanbul, Turkey, September 10-14, 2014.

Scientific Advisory Committee, „From Computational Biology to System Biology" workshop (CBSB14), Gdansk, Poland, May 25-27, 2014.

ORGANIZATION COMMITTEE MEMBERSHIP (CHAIR)

Martin Golebiewski
COMBINE 2014: 5th Computational Modeling in Biology Network Meeting, Los Angeles, California, USA, August 18-22, 2014.

Frauke Gräter
"Soft-matter, polymer, and biological physics", Co-chair of Subcommittee, IUPAP Conference for Computational Physics, Boston, USA, August 11-14, 2014.

Andreas Reuter
Scientific Chair of the 2nd Heidelberg Laureate Forum, September 21-26, 2014.

Michael Strube
Area Chair, 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Md., USA, June 22-27, 2014.

WORKSHOP ORGANIZATION

Kira Feldmann and Tilmann Gneiting
Minisymposium on Spatial Statistics, Heidelberg, Germany, October 28, 2014.

Martin Golebiewski
COMBINE & ERASysAPP tutorial “Modelling and Simulation of Biological Models“, Melbourne, Australia, September 14, 2014.

NORMSYS & ISBE workshop “Standards for data and model exchange in systems biology”, Melbourne, Australia, September 18, 2014.

SeqAhead Workshop: “Setting the standards for analysing and integrating big data”, Berlin, Germany, July 9 – 11, 2014.

6th Virtual Liver Data Management and PALs Meeting, HITS, Heidelberg, Germany, May 8 – 9, 2014.

Fabian Krüger and Tilmann Gneiting
Workshop on Propriety and Elicitability, Heidelberg, Germany, June 18, 2014.

Kai Polsterer
Co-organizing the splinter on E-Science & Virtual Observatory at the Annual Meeting of the Astronomische Gesellschaft 2014, September 22 – 26, 2014.

Volker Springel
AREPO-2 Workshop 2014, American Academy of Arts and Sciences, Cambridge, USA, September 2 - 4, 2014.

Rebecca Wade
Co-organizer of the EMBO Practical Course on Biomolecular Simulation, Pasteur Institute, Paris, July 20-27, 2014.

9.5 AWARD

Kira Feldmann
Poster Award, International Symposium Extremes 2014, Hannover, Germany, October 6-7, 2014.

Edited by

HITS gGmbH
Schloss-Wolfsbrunnenweg 35
D-69118 Heidelberg
www.h-its.org
@HITStudies
Facebook.com/HITStudies

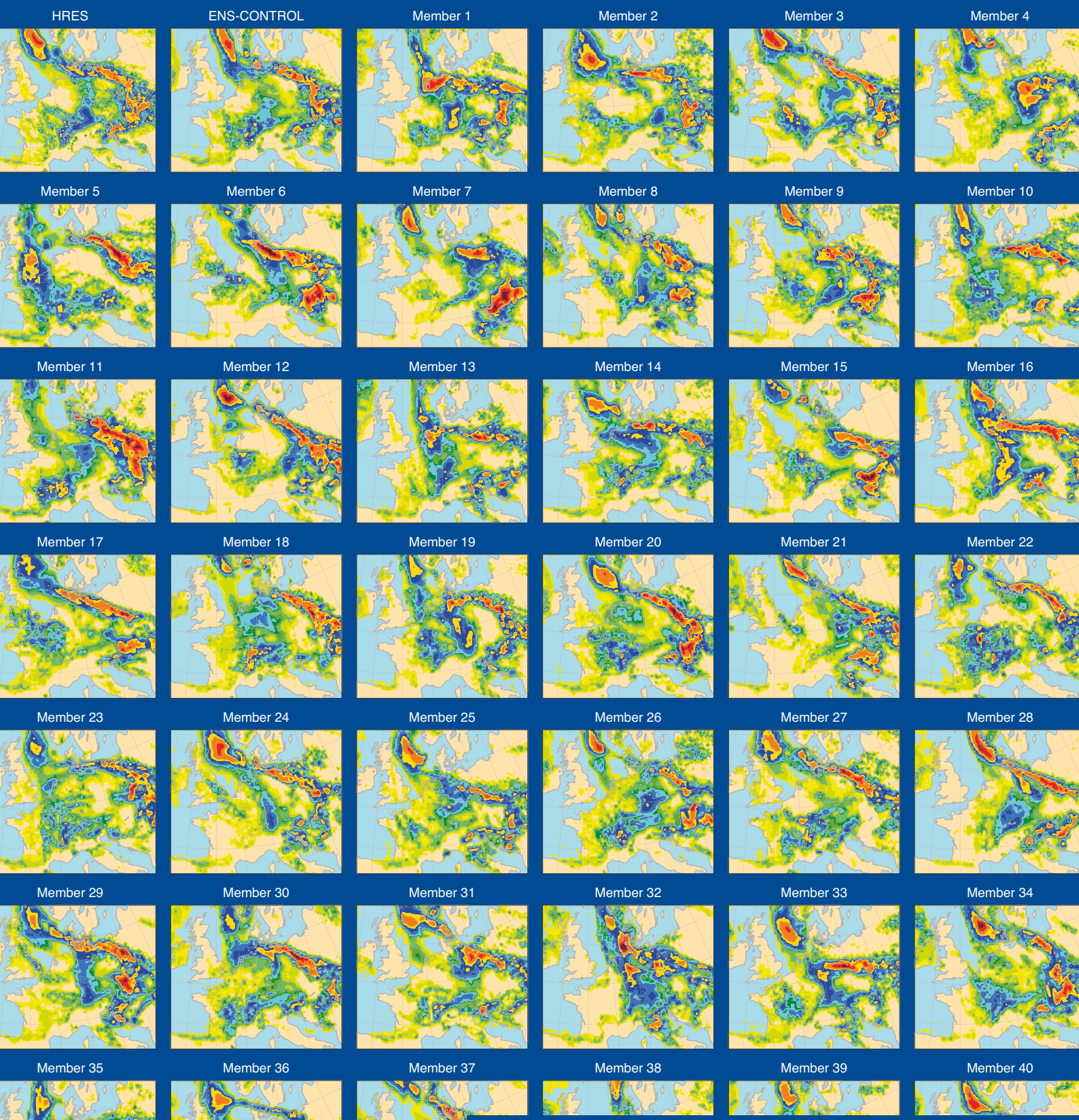
Our e-mail adresses have the following structure:
Firstname.lastname@h-its.org

Contact
Dr.Peter Saueressig
Phone: +49-6221-533 245
Fax: +49-6221-533 298
@HITStudies

Editor
Dr.Peter Saueressig
Public Relations

Pictures
HITS gGmbH,
(unless otherwise indicated)

ISSN 1438-4159



Rainfall forecast for 10 July 2014 produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). The ECMWF forecast comprises a set of 52 realisations to show the range of scenarios that may occur. Here the high resolution HRES, lower-resolution CONTROL and the 50 perturbed ensemble members are shown. The predictions were initialized on 7 July 2014 00 GMT and made for the 12-hour daytime period (0600-1800) on 10 July 2014. Colours in the images correspond to precipitation accumulation amounts. Many thanks to ECMWF for providing this stamp map.

Niederschlagsvorhersage des Europäischen Zentrums für Mittelfristige Wettervorhersage (EZMW) für den 10. Juli 2014. Die EZMW-Vorhersage umfasst 52 Einzelvorhersagen, um die Bandbreite der möglichen Szenarien darzustellen. Obenstehend sind HRES (hochauflösend), CONTROL (niedriger auflösend) und die 50 perturbierten Einzelvorhersagen des Ensembles abgebildet. Die Vorhersagen wurden am 7. Juli 2014 um 00 Uhr GMT initialisiert und betreffen den 12-stündigen Tageszeitraum (0600-1800) am 10 Juli 2014. Die Farben entsprechen der akkumulierten Niederschlagsmenge. Wir danken dem EZMW für die Bereitstellung der Abbildungen.