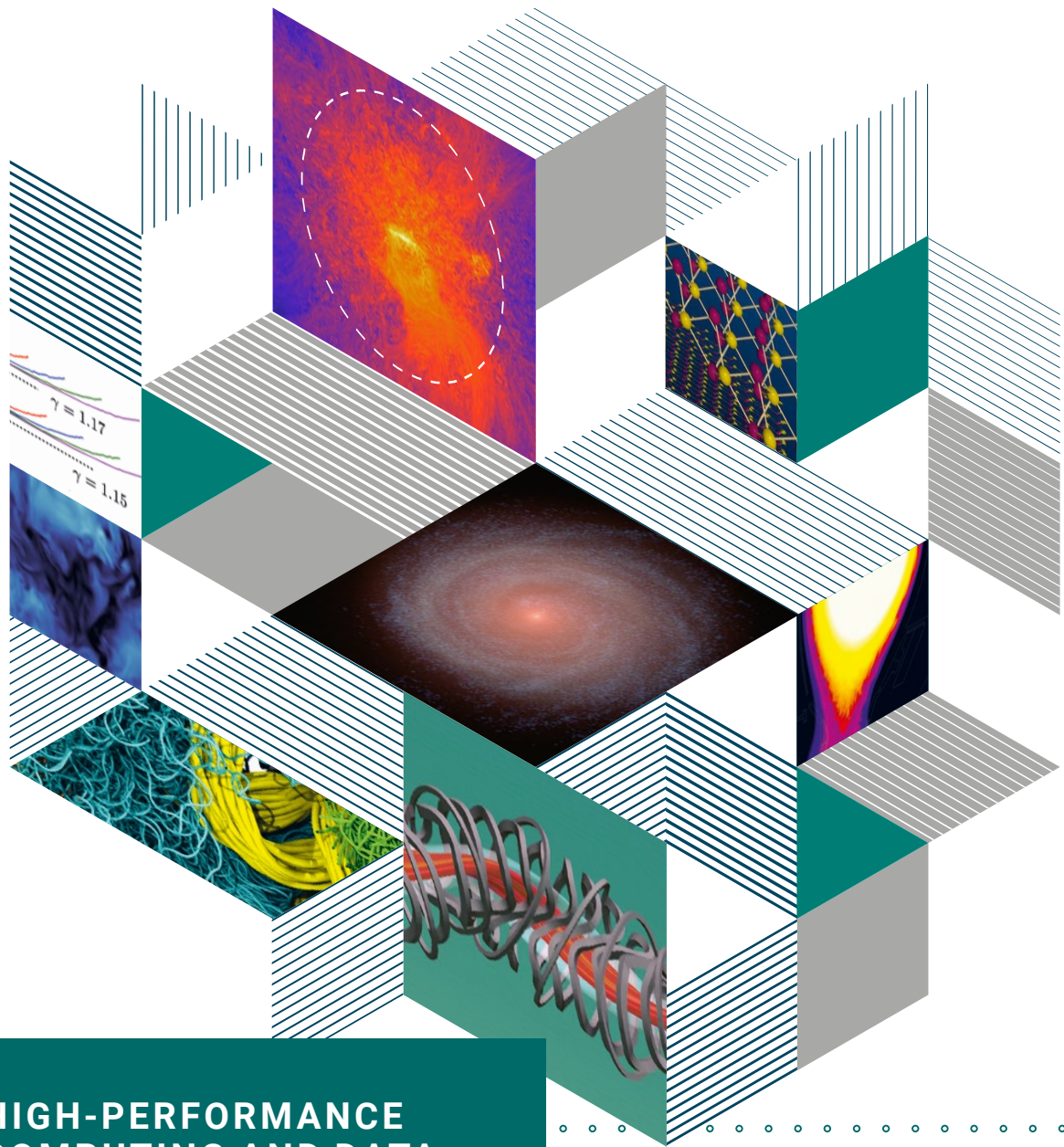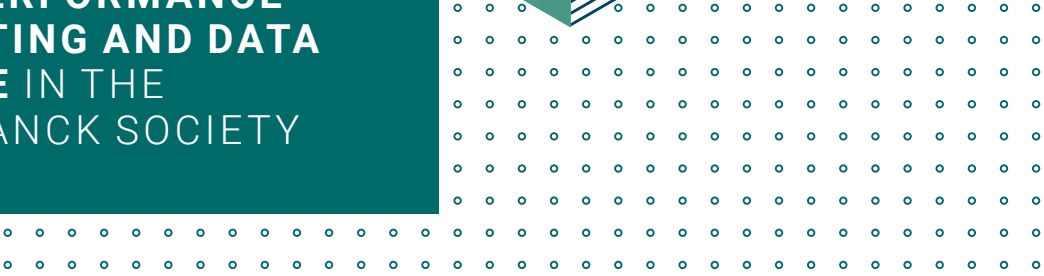MPCDF

MAX PLANCK
COMPUTING & DATA FACILITY

**HIGH-PERFORMANCE
COMPUTING AND DATA
SCIENCE** IN THE
MAX PLANCK SOCIETY

Editorial

Professor Erwin Laure

High Performance Computing (HPC) and data analysis and management are today key components of scientific workflows in many disciplines. New capabilities of supercomputers soon reaching performance levels over 1 Exaflops (10^{18} floating point operations per second) enable ever more demanding simulations and data analysis. For more and more scientific domains HPC has become an indispensable tool and thanks to the growing importance of data analysis, including machine-learning (ML) and artificial intelligence (AI) methods, also domains that have traditionally not used HPC are becoming important adopters today.

The Max Planck Society (MPG) has very early on recognized the importance of this technology for its scientists and starting in 1961 an IBM 7090 system was installed at the Institute for Plasma Physics (IPP) in Garching, which became operational in 1962. We are thus celebrating 60 years of High Performance Computing at the MPG this year!

During these 60 years, the Rechenzentrum Garching (RZG) has always provided MPG scientists with best-of-a-breed computing resources – in total 15 different systems over time, and the 16th system, Raven, became operational just this summer. Over the years, an increasing number of institutes and research groups have been using these systems, today over 40, and thus in 2015 the MPG decided to transform the RZG, which until then was a part of the IPP, into a central MPG facility under the new name Max Planck Computing and Data Facility (MPCDF).

This transformation not only reflected the increased importance of the MPCDF for the MPG at whole, but also the increasing importance of data management and storage next to computational power. Today, the MPCDF is storing over 200 PetaBytes ($2*10^{17}$ Bytes) in its tape archive, making it one of the largest data sites world-wide.

The computing and storage services of the MPCDF are complemented with high-level support activities. MPCDF experts are engaged in many projects where, together with various Max Planck Institutes, applications are being developed and improved, complex data pipelines and data management solutions are being devised. Recently, a new group dedicated to AI and High Performance Data Analytics has been created, reflecting the increasing importance of these methodologies and technologies.

The state-of-the-art MPCDF data center in Garching not only hosts the large, central MPG supercomputers and data-management solutions, but also an increasing number of medium-sized, institute-owned computing systems. Currently, over 25 Max-Planck-Institutes profit from this service and host their systems at the MPCDF facilities. With the ever increasing demand for hosting and growing infrastructure requirements of supercomputers, the current facilities in Garching are however nearing their capacity limits. To provide a long term perspective, new facilities as part of the Life Science Campus in Martinsried are being discussed.

In this brochure, we want to present selected examples of science that is currently being supported by the MPCDF services. The by no means comprehensive selection includes articles from astrophysics, brain research, materials and bio science, high-energy physics, plasma physics and fusion research, turbulence research, demographics.
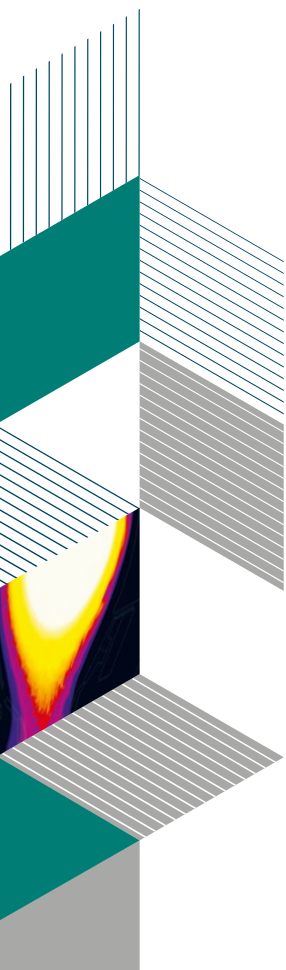
We hope, the examples collected in this brochure provide an interesting and inspiring read and if you are interested in using the MPCDF services to accelerate your science – contact us! Our experts are here to help!

Professor Erwin Laure, Director MPCDF

Contents

Formation of the Milky Way galaxy

Modern numerical techniques have opened up the possibility to evolve the initial conditions left behind by the Big Bang forward in time, accounting self-consistently for most of the relevant physics. This includes the dynamics of dark matter and of diffuse hydrogen and helium gases, magnetic fields, as well as star formation, supermassive black hole growth, and associated energy feedback. The simulations can predict the formation and evolution of galaxies in stunning detail.

AUTHOR *Volker Springel*

Fig. 1: Face-on (top) and edge-on (bottom) stellar light projection of the K-, B- and U-band luminosity of stars in our highest-resolution simulated Auriga galaxy at the present day.
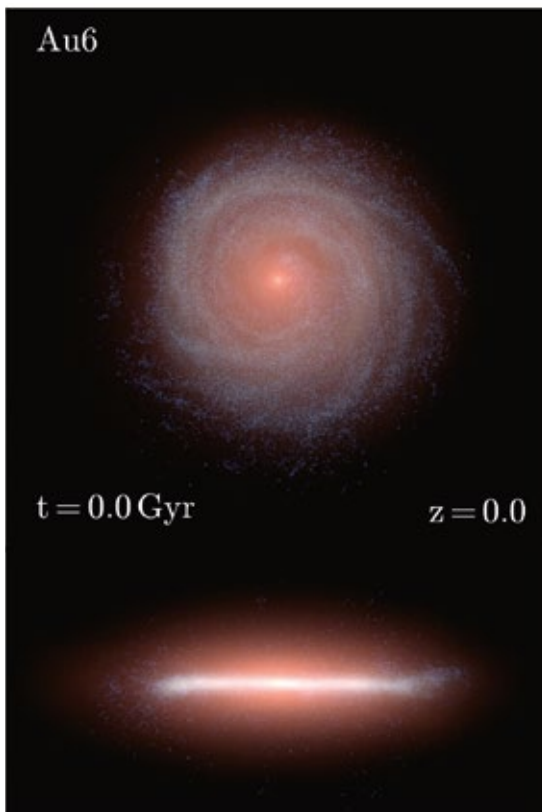
Thanks to the recent progress in observational cosmology, we now have a firmly established standard cosmological model in which most of the matter (~85%) is composed of yet unidentified collisionless dark matter particles, while ordinary baryons produced in the Big Bang make up only a subdominant part (~15%). The initial state of the universe is a hot, nearly featureless soup of dark matter and plasma. But as the cosmos expands and cools, gravity amplifies tiny density perturbations, until they collapse and virialize as non-linear dark matter halos. Baryons radiatively cool and settle in these small potential wells, forming rotationally supported disks. The relentless pull of gravity causes further collapse and fragmentation in the cold gas, leading to star formation and thus the formation of stellar disks, i.e., the birth of galaxies.

However, many of the details of this broad-brush sketch of galaxy formation are still poorly understood. When stars evolve, they release enormous amounts of energy into their environment, in the form of hard radiation, stellar winds, and supernova explosions. This drives supersonic turbulence in the interstellar medium, and leads to a critical regulation of star formation to the low rates observed, which are in fact far smaller than suggested by naïve theoretical estimates. Furthermore, large galaxies are observed to shut down their star formation eventually, for reasons apparently

Fig. 2: Gas flow patterns of the circum–galactic medium for different physical models of cosmic ray transport. Streamlines indicate the direction of gas flow, and arrow colors indicate the sign of the radial velocity, with the stellar disc being oriented edge–on.

unrelated to the energy output of the stars themselves. This is widely suspected to be related to the activity of supermassive black holes in galactic nuclei. While they grow through gas accretion, these black holes are capable of turning a sizable fraction of the swallowed rest mass into pure energy. How the energy released by these powerful engines couples to the surrounding galaxies is an unsolved problem.

Direct hydrodynamical cosmological simulations have become an extremely powerful tool to study how these highly complex processes unfold over more than 13 billion years of evolution, thanks to their ability to solve the underlying systems of partial differential equations. A particular numerical challenge arises from the nonlinear coupling of a range of different physics, and the vast range of length and time scales that need to be followed. Furthermore, not all processes can be treated *ab initio*, forcing the use of "subgrid" models to account for processes below the resolution limit of the simulations.

THE AURIGA PROJECT The Auriga simulations are a set of so-called cosmological "zoom" simulations performed with our magneto-hydrodynamics code AREPO and represent a state-of-the-art effort to study the formation of galaxies with a mass comparable to the Milky Way. They concentrate the resolution and thus most of the computational expense onto individual dark matter halos selected within a large cosmological volume that is otherwise followed at coarser resolution. The galaxy formation model implemented by the code includes primordial and metal-line cooling with

self-shielding corrections, star formation, gas recycling and chemical evolution, stellar feedback through winds and supernova explosions, black hole growth through merging and gas accretion, black hole feedback in a quasar- and radio-mode channel, and last but not least magnetic fields. In some of our newest simulations, cosmic rays are included as well. To investigate the variety of possible formation histories of Milky Way-sized galaxies, a large sample of dark matter halos were selected for study, and in each case, the same object was repeatedly simulated with different numerical resolution to precisely assess numerical convergence and the robustness of our results.

COMPUTATIONAL APPROACH We have developed the massively parallel N-body/magnetohydrodynamic code AREPO for cosmological simulations in an expanding background spacetime. It treats the self-gravity with hierarchical multipole expansion on small scales, and a particle-mesh method based on Fourier techniques on large scales. While collisionless systems (dark matter, stars) are discretized as an N-body system, gas dynamics is solved on an unstructured, fully dynamic mesh of Voronoi cells with a second-order accurate finite-volume Godunov method. The mesh geometry is determined based on a set of mesh-generating points that move along with the flow, thereby creating a manifestly Galilean-invariant numerical discretization and a continuously adaptive resolution of the flow. The resulting quasi-Lagrangian scheme is particularly well matched to the high density contrasts and bulk velocities occurring in galaxy formation. The code is optimized to

$B\,[\mu\mathrm{G}]$

Au 6 CGM 1kpc
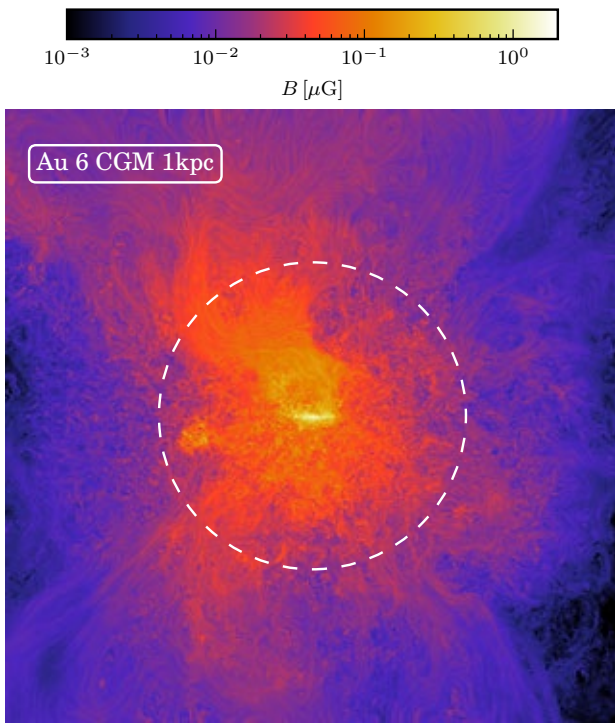
Fig. 3: Projection of the predicted volume–averaged root mean square magnetic field strength in the circum–galactic medium of one of our simulated Auriga galaxies. The galaxy's virial radius of 210 kpc is shown by the dashed circle. The faint streaks in the outskirts reflect the orientation of the predicted magnetic field.

parallelize well both for homogeneously sampled cosmological volumes and extreme zoom simulations. The currently largest calculations done with AREPO employ more than 100 billion resolution elements and are executed on more than 10^5 cores.

RESULTS FOR THE MILKY WAY: The Auriga simulations have already produced a rich body of results, ranging from detailed studies of the structure and kinematics of the formed stellar disks to properties of the circum-galactic medium. Auriga has also been used to construct mock catalogues that can be compared to the Gaia satellite mission, and for reconstructing the properties of the

putative Gaia-Enceladus merger in the Milky Way's past. In one of our most recent simulations from the project carried out at the MPCDF, we have pushed the resolution to an unprecedented level (baryonic mass resolution 800 $M_\odot$), allowing converged predictions for the full luminous satellite population of a galaxy of Milky Way mass. The exquisite resolution of this new calculation demonstrated that the predicted radial distribution of the satellites is actually in good agreement with observations of satellites around Milky Way-mass halos.

One omission in almost all cosmological simulations of galaxy formation to date is the physics of cosmic rays, even though they are known to contribute about a third of the pressure in the interstellar gas of the Milky Way, the rest being made up in equal proportions by thermal and magnetic pressure. In recent extensions of the Auriga simulations, we began to include simplified treatments of the complex Plasma Physics governing cosmic rays. Interestingly, we found a profound influence of cosmic ray physics on the gas motions in the halo of the galaxies, thereby changing how gas flows in and out of galaxies. This in turn impacts the morphology of the forming galaxies, causing quite small and more compact stellar disks in poorer agreement with observations than for simulations without cosmic rays. But it could well be that refined treatments of the complicated cosmic ray physics will rectify this problem in future work.

PUBLICATIONS

1. *Grand R.J.J., et al., "Determining the full satellite population of a Milky Way–mass halo in a highly resolved cosmological hydrody–namic simulation", arXiv:2105.04560 (2021)*

2. *Buck T., et al., "The effects of cosmic rays on the formation of Milky Way–mass galaxies in a cosmological context", MNRAS, 497, 1712 (2020)*

3. *Grand R.J.J., et al., "Gas accretion and galactic fountain flows in the Auriga cosmological simulations: angular momentum and metal redistribution", MNRAS, 490, 4786 (2019)*

4. *Grand R.J.J., et al., "The Auriga Project: the properties and formation mechanisms of disc galaxies across cosmic time", MNRAS, 467, 179 (2017)*

5. *Springel V., "E pur si muove: Galilean–invariant cosmological hydrodynamical simulations on a moving mesh", MNRAS, 401, 791 (2010)*

MAX PLANCK INSTITUTE FOR ASTROPHYSICS

Max Planck Institute for Astrophysics / Department of Computational Astrophysics Prof. Volker Springel and his group are interested in a number of topics in numerical astrophysics, including cosmic structure formation, dark matter and dark energy, galaxy formation and evolution, magneto-hydrodynamics, cosmic rays, and feedback processes from stars and supermassive black holes. They also work on developing new numerical methodologies and high-performance computing techniques.

Magnetohydrodynamic coupling of solar and stellar atmospheres

Magnetic fields couple the different layers of solar and stellar atmospheres. Numerical experiments can now be conducted with the necessary physics to allow direct comparison with observations. These simulations are critical for interpreting the observations of the Sun and stars and understanding their physics.

AUTHORS   *Cosima Breu, Robert Cameron, Mayukh Panja, Hardi Peter, Damien Przybylski, and Sami K. Solanki*

The Sun varies on timescales from seconds to millennia and displays dynamic phenomena at all observable spatial scales, from tiny bright points to coronal waves covering almost the whole visible surface. Almost all of this variability and related phenomena are associated with the Sun's highly complex and dynamic magnetic field. The magnetic field is generated beneath the solar surface by convective motions, which are affected at larger scales by the rotation of the Sun. Once the field reaches the surface, it is continuously restructured due to the highly turbulent, partly supersonic convective motions. Higher up in the atmosphere, the magnetic field governs the spatial distribution of plasma. There, the constantly evolving field drives flows and jets and makes the atmosphere extremely dynamic and spatially finely structured. Because the magnetic field is space filling, it also couples the different layers of the atmosphere. For example, magnetic and magneto-acoustic waves transport energy through the different atmospheric layers.

Understanding the physics of the solar atmosphere is difficult. In the lowest layer there are transitions from the plasma motions dominating the magnetic field to the field dominating the motions, and there is a transition from the energy being carried by convection to the energy being carried by radiation. As we move upwards into the chromosphere, the temperature increases from about 6,000 to 10,000 K, and the density drops by many orders of magnitude. Due to the drop in density, the collision rate between the particles decreases and plasma can no longer be treated in the local thermal equilibrium (LTE) or statistical equilibrium approximations. Above the chromosphere the temperature rises steeply to above a million
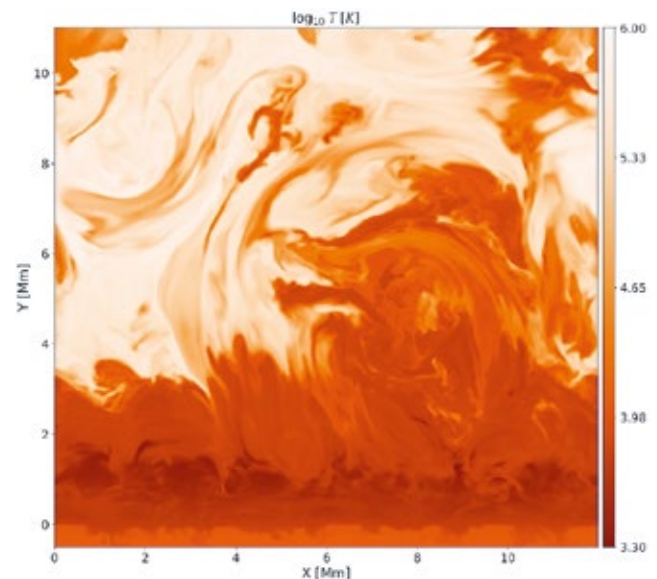
Fig. 1: Temperature through a vertical slice of a simulation of the solar atmosphere. The plasma becomes optically thick below the solar surface at Y=0. The temperature decreases outwards for the first 500 km above Y=0, after which it rises first to about 10^4 degrees, with a very sharp increase to almost a million Kelvin in the corona. The atmosphere is finely structured and extremely dynamic. This makes this type of simulation necessary for both understanding the physics and interpreting the observations. Image by Damien Przybylski.

Kelvin in the corona. There, thermal conduction along magnetic field lines by high-speed electrons becomes important.

Interpreting the observations is also difficult, especially observations of the chromosphere. In addition to the problems associated with not being able to assume LTE or statistical equilibrium, there is also the problem that only the light integrated along the line of sight is available, and in a number of interesting spectral lines the photons are scattered multiple times at atoms and ions within the atmosphere.

The main goal of the simulations performed by the members of the Solar Department at the Max Planck Institute for Solar System Research is to understand the physics of the solar atmosphere, including the coupling between the different layers. As a secondary goal, we test our understanding and the realism of our models by producing synthetic data that are then compared with real observations.

TOOLS There are several different computational tools that have been developed at, or in conjunction with, the department. Here we highlight the Max Planck Institute for Solar System Research, University of Chicago, Radiative Magnetohydrodynamic (MURaM) code. This code has recently been updated to perform *ab initio* calculations of the interaction between the radiation, magnetic fields, and plasma throughout the upper convection zone, through the photosphere, chromosphere and into the corona.

THE CHROMOSPHERE SEPARATING THE COOL SURFACE FROM THE HOT CORONA A recent major upgrade to the MURaM code now enables us to perform simulations of all layers of the solar atmosphere with the correct physics. The code had been optimized over many years for treating the magnetohydrodynamics of the solar atmosphere with as low numerical diffusivity as possible. This strength of the code means that we can now simulate the chromosphere and corona at higher resolution and with lower diffusivities than has ever been possible. In particular, the treatment of the chromosphere is now possible with the MURaM code because non-LTE and non-statistical-equilibrium effects have been included. Our numerical experiments have revealed the highly dynamic nature of the solar chromosphere.

Even in the quiet Sun, away from active regions and flares, we see shocks, weak magnetic fields and a constantly evolving, finely structured chromosphere. Figure 1 shows the temperature on a vertical cut through a snapshot of such a simulation. In a series of papers, we have shown that small-scale vortices, aligned to lie along magnetic field lines, transport enough energy to heat the chromosphere above decaying active regions.

THE HOT UPPER ATMOSPHERE AND ITS MAGNETIC COUPLING TO THE SURFACE The high temperature of the corona of the Sun has remained enigmatic since its discovery in the 1940s. Of particular interest are loops seen in extreme UV and X-ray emission that are the basic building blocks of the corona. Our numerical experiments show that motions within small-scale concentrations of the magnetic field in the intergranular lanes between convection cells at the surface generate the required energy flux. This energy flux is then channeled upwards, e.g., in small vortices, before the induced

Fig. 2: Internal structure of a coronal loop. While curved on the Sun, the numerical model solves the equations for the loop model in a rectangular grid. At both ends the loop is rooted at the surface, illustrated here by the two planes at the ends of the box that show the vertical magnetic field on a gray scale (with white and black for opposite magnetic polarities). The volume rendering shows the density in the loop (in false color) with red color showing the million Kelvin hot coronal part. The coronal part shows individual elongated strands that evolve quickly in time. Image by Cosima Breu.

G2V Temperature [K]   **M0V Temperature [K]**
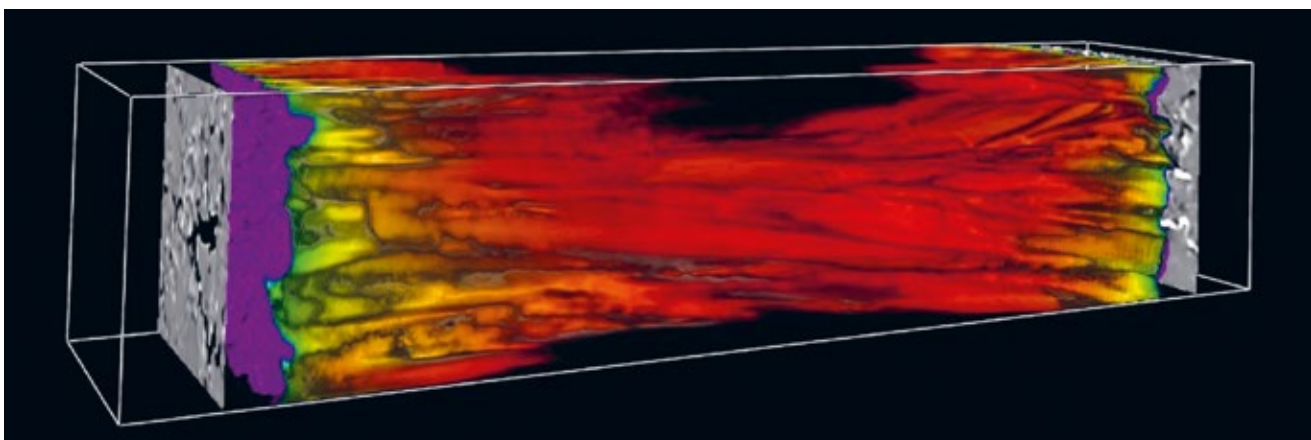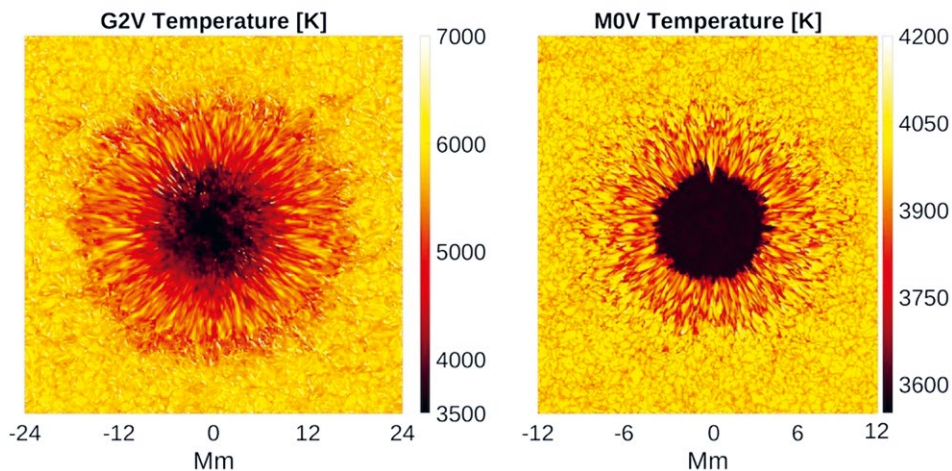
Fig. 3: Comparison between a sunspot and its equivalent on a cooler star. The left panel shows the effective temperature of the region around a spot on a star of type G2V (like the Sun) obtained from a computer simulation. The right panel shows the same, but for an M0V type star. M0V stars are colder, smaller and denser than the Sun. Image by Mayukh Panja.

currents are dissipated and heat the plasma to more than a million Kelvin (see Fig. 2). This consistent picture of the energization of the upper atmosphere has been possible only because of the low diffusivity of the code and the exquisite spatial resolution of the numerical experiments. Models with MURaM also show an abundance of short-lived small-scale brightenings originating from transiently heated pockets of plasma reaching coronal temperatures. These are caused by slow surface motions that drive bundles of magnetic field lines to interact in the upper atmosphere and to undergo component reconnection. With this model, we provide a first explanation of the so-called campfires, ubiquitous small coronal brightenings recently discovered in observations by ESA's latest science mission, Solar Orbiter.

MAGNETIC ACTIVITY ON OTHER STARS Sunspots are regions of high magnetic field strength. Here the magnetic field suppresses the convection of the ionized plasma and hence the energy transport. This leads to a lower temperature and a lower emergent radiation, letting the spots appear dark. On other stars, starspots have been inferred from changes in the radiance, as well as in studies of the spectrum of stellar light and spectropolarimetry. However, because stellar surfaces are not resolved, observations provide only limited information on the properties of starspots. Correctly inferring the properties of the spots requires good models of their structure. We have produced the first ever radiative-magnetohydrodynamic models of starpots. The structure of the starspots is determined using all the relevant physics, which we have validated in detail using the Sun, and will play a crucial role in understanding observations of magnetic fields on stars other than the Sun. For example, on stars significantly cooler (in terms of surface temperature) and more compact than the Sun, the modeled temperature difference between starspot and surrounding region is much smaller than for the Sun

(see Fig. 3). This translates to a very different intensity contrast for the two stars, which is just one example underlining the key role of models for properly interpreting stellar observations.

———

PUBLICATIONS
1. Breu, C., Peter, H., Cameron, R., Solanki, S.K., Przybylski, D., 2021, A&A submitted
2. Chen, Y., Przybylski, D., Peter, H., et al. 2021, A&A in press (DOI: https://doi.org/10.1051/0004-6361/202140638)
3. Panja, M., Cameron, R. H., and Solanki, S. K., 2020, ApJ 893, 113
4. Przybylski, D., Cameron, R., Solanki, S. K., et al., 2021, A&A submitted
5. Shapiro, A. I., Solanki, S.K., et al., 2017 Nature Astronomy, 1, 612
6. Yadav, N., Cameron, R. H., and Solanki, S. K., 2021, A&A 644, 44

Max-Planck-Institut für Sonnensystemforschung

Max Planck Institute for Solar System Research / Department Solar and Heliosphere The overarching science theme of the department is the magnetically coupled solar atmosphere. The research topics focus on the Sun's atmosphere with its rich structure and dynamics, the solar magnetic field as the main driver of solar activity and dynamics, the influence of the Sun on Earth, and the solar-stellar connection. We address these topics through a close interaction between numerical experiments and observations of the Sun. For the observations, we develop instrumentation for leading solar space missions, as well as for facilities on the ground. Most notably, these include the Sunrise solar balloon-borne telescope, and instruments on the recent Solar Orbiter mission of ESA and NASA.

Explosive instabilities in fusion plasmas and progress towards taming them

For realizing a fusion power plant, plasma confinement in a magnetic "cage" is promising. Similar to coronal mass ejections, plasma instabilities that expel heat and particles from the sun, also laboratory plasmas can develop plasma instabilities, which need control. We describe simulations on supercomputing systems to advance the research.

AUTHORS *Matthias Hoelzl, Vinodh Bandaru, Andres Cathey, Ihor Holod, Verena Mitterauer, Nina Schwarz, Fabian Wieschollek, and Sibylle Günter; in collaboration with the JOREK community, EUROfusion and the MPCDF*

About 99% of the matter in the universe is in the plasma state, and violent instabilities are ubiquitous in astrophysical plasmas. A prominent example are coronal mass ejections, which are driven by magneto-hydrodynamic (MHD) instabilities and expel large amounts of material and energy from the sun, constituting a risk for technical systems on earth.

Following the example of the stars, energy can be harvested from the fusion of light atomic nuclei. Confining a plasma of several hundred million degrees Celsius in a strong magnetic "cage" to hinder it from getting into direct contact with material surfaces is a promising path towards a power plant, which is pursued with the ITER tokamak, presently under construction in Cadarache, South France.

When optimizing a magnetically confined fusion plasma for best energy output at lowest "costs" for the confining magnetic field, the plasma inevitably hits instabilities, comparable to the sun. We work on revealing the origin of such instabilities and developing reliable control strategies that pave the way towards a robust fusion power plant.

COMPUTATIONAL APPROACH Massive spatial and temporal scale separations require efficient algorithms and powerful computing systems. Furthermore, the plasma strongly interacts with its surrounding material, e.g., with the so-called divertor components, where most of the heat loads are deposited, and

with electromagnetic currents in conducting structures around the plasma. Extended MHD models allow studying the explosive instabilities in realistic geometries. Due to long time scales of interest, the system of equations is evolved in a fully implicit way, which implies solving very large matrix systems in a parallelized manner at every time step. The challenging physics problem furthermore does not permit to apply iterative solvers with standard algebraic pre-conditioners such that dedicated methods adapted to the physics problem are required. We develop the non-linear extended MHD code JOREK (https://www.jorek.eu) within a strong international community. Both, the support by high-performance computing (HPC) experts and the availability of large computing systems at MPCDF contribute significantly to our success.

EDGE-LOCALIZED MODES (ELMS) expel material from the surface of the plasma causing large transient heat loads to so-called divertor structures, which would reduce their life time in ITER or a power plant in an unacceptable way. With our simulations, we recently
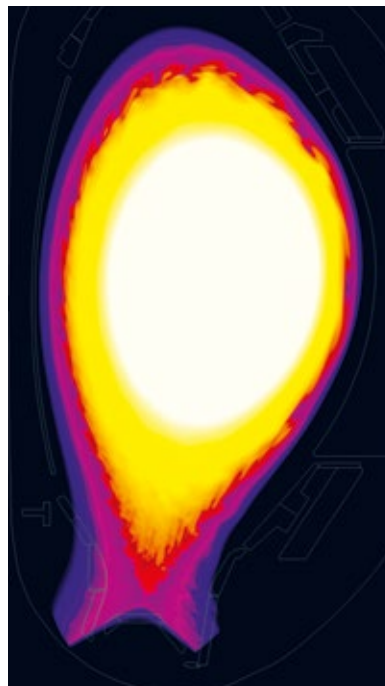
Fig. 1: Periodic large ELM crashes in ASDEX Upgrade were reproduced in simulations for the first time. The pressure distribution during an ELM crash is shown. Material is expelled from the plasma and leads to large heat loads onto divertor targets (at the bottom).
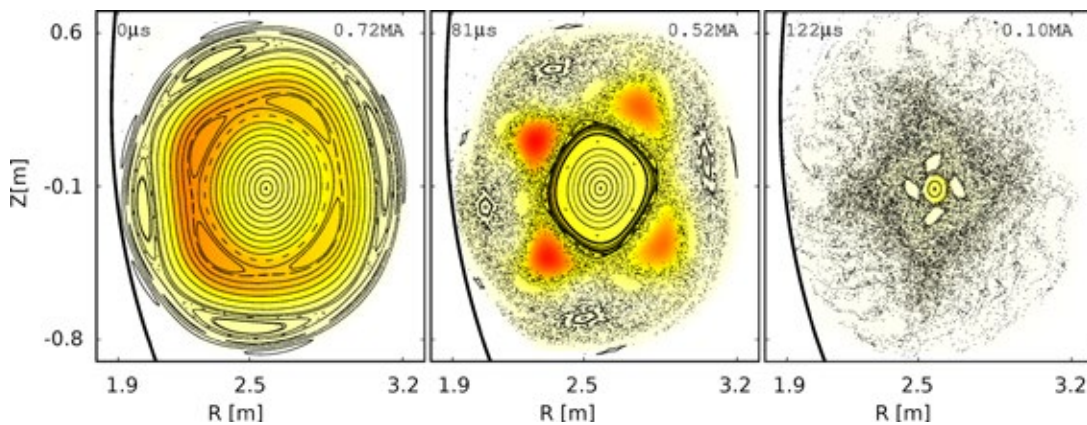
Fig. 2: Simulation results are shown for the termination of a RE beam by the formation of magnetic islands leading to an ergodization of the magnetic topology (Poincare plot in black). Within less than 100 micro–seconds almost the entire RE content (shown in color) is lost just like experimentally observed in the JET tokamak. Due to the stochastic region forming from the plasma edge, the deposition to plasma facing com–ponents does not exhibit localized hotspots in line with the experiment.

were able to describe such large periodic ELM crashes for the first time, reproducing many experimental observations. The simulations allow studying the processes in detail and reveal the role of the complex interplay between stabilizing (plasma flows) and destabilizing (pressure gradient and current density) effects. We can thus reveal the origin of the explosive onset of the instability (Fig. 1). We also study ways to mitigate ELMs, e.g., by the injection of frozen deuterium pellets, an experimentally established technique to control the ELM frequency. Further work is underway to study ELM free regimes and ELM control by magnetic perturbation fields in direct comparison to experiments, increasing the confidence in predictive modeling for ITER.

MAJOR DISRUPTIONS lead to a sudden loss of the complete plasma confinement and would have the potential of causing considerable damage in a future tokamak power plant. While the goal is to operate the plasma in regimes that render disruptions virtually impossible, reliable mitigation methods will be obligatory for large devices to protect them when a disruption cannot be avoided, e.g., following a system malfunction. A possible mitigation technique is shattered pellet injection (SPI). We simulated how SPI can mitigate disruptions at the example of the ASDEX Upgrade tokamak in Garching where an SPI system is presently being installed, allowing for one-to-one comparisons in the future. Vertical plasma instabilities, which are associated with disruptions, are being simulated including predictions of heat and electromagnetic loads.

During disruptions, electrons are often accelerated to velocities close to the velocity of light. Such relativistic runaway electrons (REs) can pose major problems when they interact with the surrounding walls. We study the effect of MHD instabilities on RE beams avoiding already its formation or a benign termination of

such a beam when already established. A termination of such a RE beam was simulated for the first time taking into account the self-consistent MHD-RE interaction showing very good agreement with key experimental observations (Fig. 2).

———

PUBLICATIONS
1. Orain F., Hoelzl M. et al. Physics of Plasmas 26, 042503 (2019) doi:10.1063/1.5091843
2. Hoelzl M., Hu D. et al. Physics of Plasmas 27, 022510 (2020) doi:10.1063/1.5133099
3. Cathey Cevallos A., Hoelzl M. et al. Nuclear Fusion 60, 124007 (2020) doi:10.1088/1741–4326/abbc87
4. Bandaru V., Hoelzl M. et al. Plasma Physics and Controlled Fusion 63, 035024 (2021) doi:10.1088/1361–6587/abdbcf
5. Cathey A., Hoelzl M., Futatani S., et al. Plasma Physics and Controlled Fusion (submitted) preprint at arXiv:2102.05850
6. Hoelzl M., Huijsmans G.T.A. et al. Nuclear Fusion (submitted) preprint at arXiv:2011.09120

IPP Max-Planck-Institut für Plasmaphysik

Max Planck Institute for Plasma Physics (IPP) / Division MHD and Fast Particles The IPP is concerned with investigating the principles underlying a power plant, which – like the sun – will produce energy from the fusion of light atomic nuclei. The MHD and fast particles division is concerned with large-scale plasma instabilities and their interactions with supra-thermal particles.

Full 6D kinetic simulations
of magnetically confined plasmas

With the increase in computational capabilities over the past years it is possible to simulate more and more complex and accurate physical models. Gyrokinetic theory was introduced in the 1960s and 1970s to describe a plasma more accurately than with fluid equations, but still eliminate the complexity of the fast gyration about the magnetic field lines. Although current gyrokinetic computer simulations are in fair agreement with experimental results in core physics, the assumptions in the derivation make them unreliable in regimes of higher fluctuation amplitudes and stronger gradients, such as the tokamak edge. With our novel optimized and scalable semi-Lagrangian solver we are able to simulate ion-temperature gradient modes with the full 6D kinetic equations, including the turbulent saturation, and we have demonstrated the presence of Bernstein waves.

AUTHORS *Klaus Hallatschek, Katharina Kormann, and Mario Räth*

The dynamics in a plasma can occur on a wide range of temporal and spatial scales. For this reason, various models exist to describe the plasma behaviour. They generally can be split into two main groups, fluid and kinetic models. Whereas in fluid simulations, the plasma is described using hydrodynamic equations (Navier-Stokes, Euler), in the kinetic models, the plasma is described by a phase-space density whose dynamics is governed by the Vlasov-Fokker-Planck equation. The plasma particles generate and interact with electromagnetic fields described by Maxwell's equations. Together, the Vlasov-Fokker-Planck-Maxwell system can describe physical phenomena in a tokamak plasma on all scales from the size of the device (~1 m) down to micro-scales of the order of the electron Larmor radius (~10^{-4} m). However, until recent years, such simulations were too expensive for the computational capabilities. Gyrokinetic models reduce the dimensionality of the problem from 6D to 5D by averaging over the fast circular motion of charged particles in a strong magnetic field. Moreover, by this method it is not necessary to resolve the period of the gyration, which is typically much smaller than the time scales of interest.

It is well known that gyrokinetic simulations, such as by the codes GENE, GS2 and GYRO, are in good agreement with experiments in the center of fusion devices, where only small perturbation amplitudes and gradients (in density and temperature) are present. However, at least the δf gyrokinetic models, which assume that the perturbations are much smaller than the background quantities, break down in
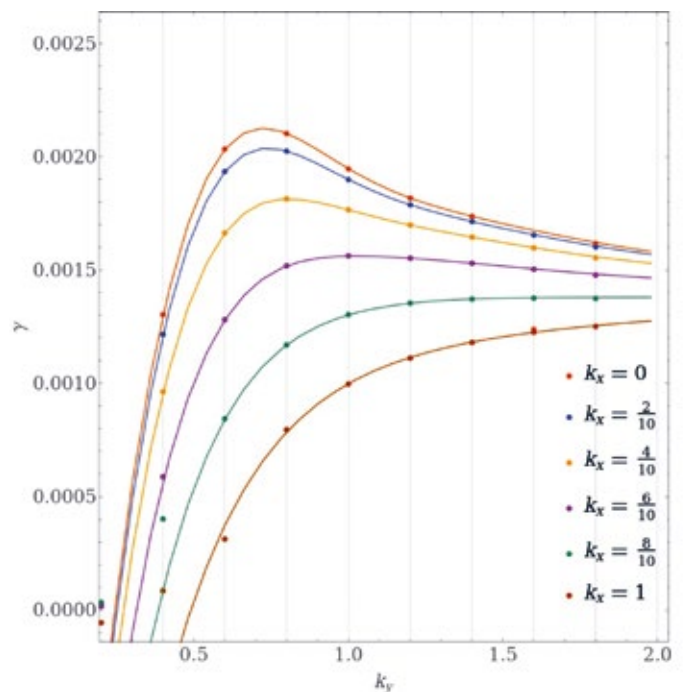
Fig. 1: Growth rate of slab–ITG instability determined with the new Lagrangian solver (bullets) in comparison to analytical calculation (lines).
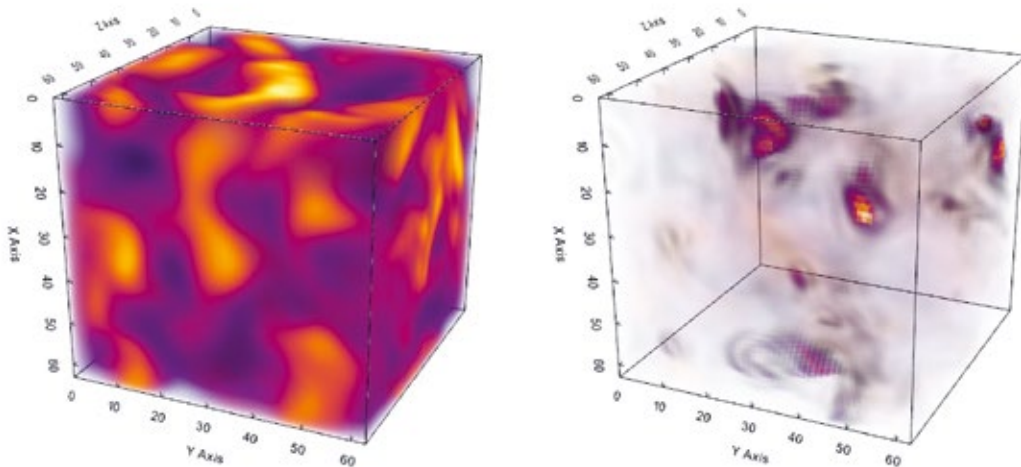
Fig. 2: 3D visualisation of the density after nonlinear saturation (left) and locally excited high–frequency IBWs (right).

regimes of higher turbulence levels, such as the plasma edge of a tokamak. Even more, modern gyrokinetic theories are only valid to a limited order in the fluctuation levels.

COMPUTATIONAL ASPECTS Due to the high dimensionality of the phase-space, the full 6D kinetic problem is computationally very intensive. For this project, a new code has been developed that is based on a semi-Lagrangian solver. The backward semi-Lagrangian method traces the grid points backwards in time during one time step which has the advantage that larger steps are allowed compared to standard Eulerian schemes. In order to cover the fast time scale induced by the background magnetic field, we use a rotating velocity grid that follows the fast gyro-motion in a Lagrangian way and thus enables larger time steps compared to the gyro-period. The major building block of the code are one-dimensional interpolation routines applied to slices of a six-dimensional array. This has been implemented in collaboration with the Max Planck Computing and Data Facility with a domain decomposition approach. Since preferably high-order interpolation is applied for accuracy reasons, the communication overhead of the code is substantial but could partly be hidden by a thread-based communication-computation overlap. Moreover, cache blocking is used to avoid discontinuous memory access. The details of the algorithm and its high-performance implementation can be found in [1].

COMPARISON WITH GYROKINETIC SIMULATIONS Predictions of the semi-Lagrangian solver have been compared to analytical and gyrokinetic calculations for the growth and the saturation of a slab ion temperature gradient instability (ITG) in a quasi-neutral plasma with adiabatic electrons. In the linear phase various diagnostic quantities have been investigated, e.g., growth rate and frequency of various modes, energy fluxes in the parallel and perpendicular direction. Furthermore, it was possible to perform simulations well into the non-linear phase. The saturation levels of the energy fluctuations and the turbulent transport agree well with results obtained from the gyrokinetic code GYRO. We showed that in the transition to the non-linear phase, waves with high frequencies close to the Larmor frequency and its harmonics are excited.

These so-called ion Bernstein waves (IBW) are only present in our code, which is a substantial difference to gyrokinetic simulations, where waves with such high frequencies are excluded by definition.

HIGH-FREQUENCY WAVES BEYOND GYROKINETIC THEORY We have shown agreement of the analytically derived ion Bernstein wave (IBW) frequencies and growth rates with the ones determined by linear code runs. This shows that the code simulates waves correctly which, due to their high frequency, are significantly beyond the reach of gyrokinetic theory.

We are currently investigating the mechanism that leads to the excitation of such high frequency waves. Simulations are performed far into the non-linear regime, and the excited IBWs are located. By studying the local conditions, such as gradients and flows, we will learn more about the excitation of such high frequency waves and their relevance for the overall plasma behavior. Understanding these high-frequency waves could lead to the missing link in the transition to high confinement operational modes (such as H-mode) in fusion devices.

──────

PUBLICATIONS
1. *K. Kormann, K. Reuter, M. Rampp: A massively parallel semi–Lagrangian solver for the six–dimensional Vlasov–Poisson equation, International Journal of High Performance Computing Applications 33, 2019*

Accelerating fusion research via high‑performance computing

Fusion energy, mimicking the physical processes that fuel the stars, is an extremely promising pathway to matching the constantly growing global electricity needs. With the help of emerging exascale supercomputers, we will be able to make reliable predictions regarding key aspects of fusion devices, guiding the design of future fusion power plants and accelerating fusion research.

AUTHOR *Frank Jenko*

FROM HIGHLY IDEALIZED TO REMARKABLY REALISTIC MODELS High-performance computing (HPC) has always been extremely important for fusion research, given that the fundamental equations as well as many physical phenomena (like turbulent transport) are inherently nonlinear, limiting the capabilities of analytical theory. Initially, the focus was on highly idealized models, since they were more tractable with the HPC resources of the day. With time, the models became more and more comprehensive and realistic, however, allowing for quantitative comparisons with experimental measurements. This happened first in the context of areas like collision-induced transport and linear magnetohydrodynamics (MHD). Later, the focus shifted to more complex issues like turbulent transport and nonlinear MHD. More recently, we have come to realize that many physical processes in fusion plasmas that have been investigated in isolation over many years are actually deeply connected. This implies that we are entering the multi-physics, multi-scale era of fusion simulation.

DEVELOPMENT AND APPLICATION OF STATE-OF-THE-ART HPC CODES In terms of code development, our flagship effort is the continuous evolution of the GENE family of gyrokinetic turbulence codes (see also: http://genecode.org). The versions of GENE targeted at describing turbulent transport in the core region of tokamaks are widely used (at IPP and by a large, world-wide user base) to explore a very wide range of physics questions. Meanwhile, two new members joined the family. After several years of code development, the first publications based on GENE-3D, a stellarator version of GENE, appeared recently [1, 2]. Importantly, GENE-3D describes the full dynamics of the system, i.e., the turbulent motion of the ions and electrons over the entire inner volume of the plasma, including the resulting fluctuations of the magnetic field. Another effort is focused on the development of a version of GENE, which is also applicable to the very important edge region. This code, GENE-X [3], leverages the mathematical and computational infrastructure originally developed for the fluid edge turbulence code GRILLIX. First simulations in an ASDEX-Upgrade-like geometry have been carried out recently. At the same time, GRILLIX

Fig. 1: Snapshot from a GENE–X simulation of plasma turbulence in the edge region of the ASDEX Upgrade tokamak at IPP Garching. Turbulent transport induces energy losses and limits the energy confinement time, which is a key figure of merit of any fusion device.
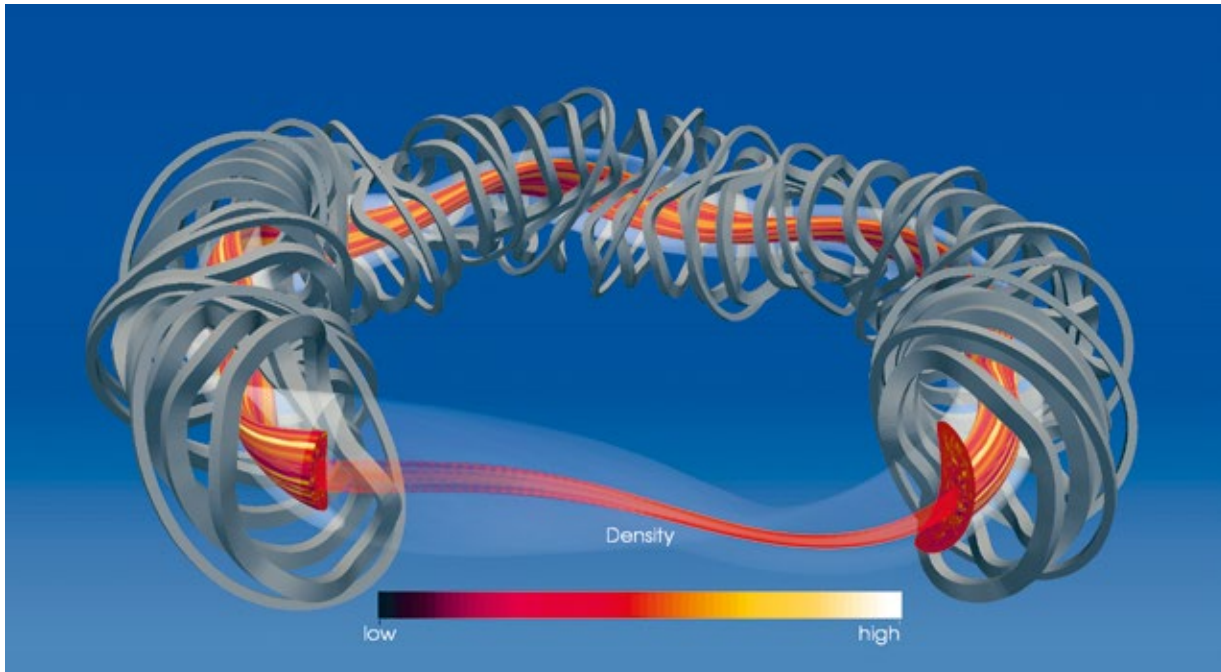
Fig. 2: Turbulence in the Wendelstein 7–X stellarator at IPP Greifswald, as computed by GENE–3D. HPC offers the opportunity to predict turbulent transport in future stellarators and optimize it according to certain criteria, exploiting the relatively large number of degrees of freedom.

has seen very significant extensions, leading to a first wave of interesting physics results. Several other codes are further developed and used in the Tokamak Theory Division, including the particle-in-cell code ORB5.

TOWARDS A VIRTUAL FUSION PLASMA VIA EXASCALE COMPUTING The prospects of exascale computing are very intriguing for fusion research, and codes like GENE are currently being prepared for upcoming exascale supercomputers [4]. For the first time, we will be able to accurately capture essential aspects of fusion plasmas based on first-principles models, including various interactions between different physical processes. This development will allow us to become truly predictive, extrapolating from existing devices and plasma regimes to future ones with a high level of confidence. Needless to say that such a capability will greatly facilitate the development and optimization of next-step devices and future power plants. While a full-blown virtual tokamak or stellarator may be out of reach for quite some time, critical aspects of fusion physics, from heat and particle exhaust to the avoidance and control of disruptions, will greatly benefit from such efforts. Exascale computing is therefore viewed by many as a likely game changer for fusion research, offering new pathways to guiding and accelerating it. It is the right new "technology" at the right time, promising exciting new R&D opportunities.

PUBLICATIONS
1. M. Maurer et al., J. Comp. Phys. 420, 109694 (2020)
2. A. Bañón Navarro et al., Plasma Phys. Control. Fusion 62, 105005 (2020)
3. D. Michels et al., Comp. Phys. Commun. 264, 107986 (2021)
4. K. Germaschewski et al., Phys. Plasmas 28, 062501 (2021)

IPP Max-Planck-Institut für Plasmaphysik

Max Planck Institute for Plasma Physics / Tokamak Theory Department The research activities of the Tokamak Theory Department in Garching span a wide range of topics and approaches. This includes, in particular, the development of state-of-the-art simulation codes and their application to many open scientific problems, from addressing fundamental theoretical questions to interpreting specific experimental results. While most of these efforts are dedicated to tokamaks, extensions to stellarators are also pursued. Moreover, there exist significant activities in the area of plasma astrophysics, in particular in the context of the Excellence Cluster ORIGINS, which involves two large universities as well as several Max Planck Institutes in the Munich area.

Stellarator optimization

Stellarators must be carefully optimized in order to achieve their goal of effectively confining hot plasmas. Fortunately, the large number of degrees of freedom in this problem makes it indeed possible to find suitable toroidal magnetic fields and coil systems, while satisfying a number of additional physical constraints critical to a successful design.

AUTHORS *Sophia Henneberg, Jim–Felix Lobsien, Michael Drevlak, Per Helander, and Gabriel Plunk*

CONTROLLED NUCLEAR FUSION holds the promise of a safe and reliable means of future energy production. To achieve fusion reactions in a reliable fashion, so-called "magnetic confinement devices" have been constructed for decades, using magnetic fields to confine high temperature plasma. Meanwhile, an experimental test reactor is under construction in Cadarache, France. It relies on the tokamak principle: a set of toroidally arranged coils provides one part of the confining magnetic field, a toroidal plasma current, inductively driven by a central solenoid, provides the other. The resulting axially symmetric device, although quite successful, also has disadvantages, which led to the development of alternative confinement concepts. In stellarators, the confining magnetic field is chiefly provided by external coils such that there is no need for a plasma current, which may be the cause of instabilities. Stellarator fields, however, are truly three dimensional and consequently much more complicated – and they are not guaranteed to confine plasma well. Therefore, a potentially successful stellarator design requires computer optimization.

STELLARATOR OPTIMIZATION The magnetic field in stellarators is calculated using the theory of magneto-hydrodynamics (MHD), balancing magnetic and plasma pressure. The shape of this field is crucial for the performance of the projected device. It must be ensured that field lines stay on so-called flux surfaces, which are concentric around a closed curve called the magnetic axis. Within this theory, the plasma equilibrium is completely determined by the shape of the last closed flux surface and the profiles of plasma pressure and plasma current. The vast space of possible boundary shapes offers the degrees of freedom necessary to optimize the field for a number of further properties, such as stability against perturbations, a good confinement of particles, and favorable transport properties, i.e., the ability to sustain temperature and density gradients between the axis and the outer boundary of the device.

The confinement of the plasma particles is tied to certain symmetries – here called quasi-symmetries – of the magnetic fields, which show up in properties of the absolute strength of the magnetic field when expressed in suitable coordinates.

The optimization of the magnetic field, however, is only the first step of the optimization process. The coils necessary to produce the field sufficiently accurately need also to be designed. The problem of finding those coils is not well-posed, i.e., its solution is not unique, but this fact can be exploited to satisfy additional criteria like limits on coil shape, distances, inter-coil forces, and so on.
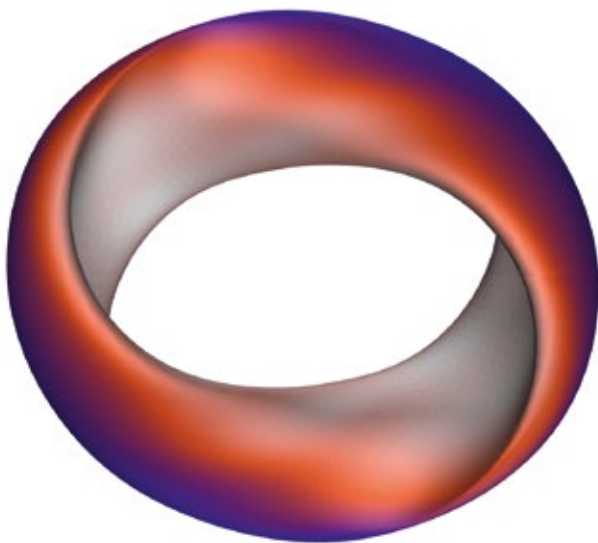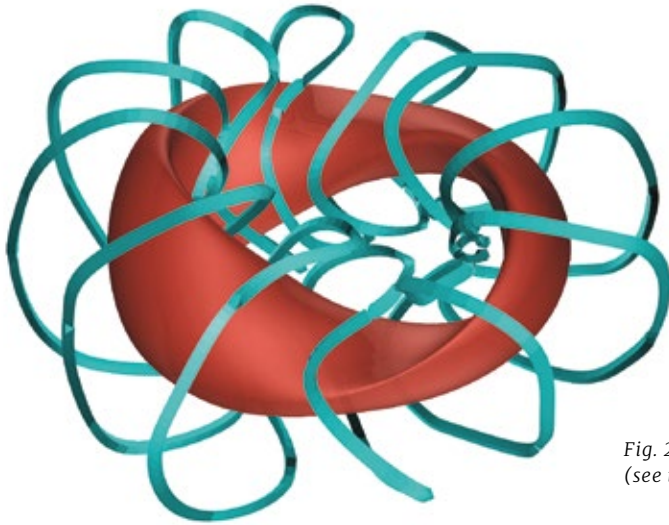

Fig. 1: Shape of the last closed magnetic surface of a quasi–axi– symmetric stellarator as described in the text. The color codes the magnetic field strength on that surface (red codes the highest, blue the lowest values).

Fig. 2: Optimized coil set for a quasi–axial stellarator (see text) with three coils per half period.

A COMPACT QUASI-AXISYMMETRIC STELLARATOR Stellarators with quasi-axial symmetry comprise a class of stellarators yet to be explored, capable of confining fast fusion-alpha particles. One reason they have drawn the attention of stellarator research in recent years is the promise of greater simplicity. Quasi-axisymmetric stellarators have naturally large intrinsic ("bootstrap") currents that could contribute to the rotational transform. A potential drawback is that this current prevents the use of a resonant divertor as currently used with great success (in the non quasi-axially symmetric W7-X device). Quasi-axisymmetric stellarators can be realized at aspect ratios much smaller than those of quasi-isodynamic or quasi-helically symmetric stellarators. Note, that the aspect ratio describes the ratio of the small radius of the torus to the large radius, which relates to the relative amount of plasma volume available. In magnetic coordinates, the field structure of a quasi-axisymmetric stellarator is equivalent to a tokamak.

In the scope of the work presented here, a quasi-axially symmetric equilibrium was optimized at an aspect ratio of A=3.4 and N_p=2 field periods. Only a small fraction of the rotational transform is provided by plasma currents and the configuration is stable at a beta of 3%.

STOCHASTIC COIL OPTIMIZATION is a technique aimed at improving the resilience of the coil set to misalignment and fabrication errors. As in regular coil optimization techniques, a coil set needs to be found that aligns as accurately as possible with the target magnetic field obtained from the preceding equilibrium optimization. In contrast to regular coil optimization, however, stochastic optimization evaluates not only the quality of a coil set with a given parameter set, but that of an entire cloud of perturbed coil sets scattered around a given reference coil set. Optimization of the resulting cloud of points in the parameter space is done with the same algorithms as in the reference case. Good optimization results have been obtained using a cloud population of 2,000 to 20,000 perturbed coil sets in each evaluation of the target function. Generally, a larger coil population yields better optimization results.

Optimizations carried out in this manner resulted in coil configurations that, in addition to showing much better resilience to perturbations, also had much better performance in the unperturbed case. Geometric properties like coil curvatures and clearances, important for the engineering feasibility of a coil system, were improved, too. In the case of an actual power plant, these improvements would result in a significant relaxation of machining and alignment tolerances, which in turn would lead to significantly reduced construction cost, as well as lower risks of schedule overruns and a lower overall construction risk.

In order to achieve computation of the point cloud fitness, the evaluation of the perturbed coil sets had to be performed in parallel. Utilizing the massively parallel computers of MPCDF was essential for this effort.

─────

PUBLICATIONS
1. *M. Drevlak et al 2019 Nucl. Fusion 59 016010*
2. *J. Lobsien et al 2018 Nucl. Fusion 58 106013*
3. *S.A. Henneberg et al 2019 Nucl. Fusion 59 026014*
4. *S.A. Henneberg et al 2020 Plasma Phys. Control. Fusion 62 014023*
5. *S.A. Henneberg et al 2020, arXiv:2012.09278, accepted by JPP*
6. *J. Lobsien et al., 2020, Journal of Plasma Physics, 86(2), 815860202*
7. *S.A. Henneberg, et al. 2021, arXiv:2105.00768, submitted to JPP*

IPP Max-Planck-Institut für Plasmaphysik

Max Planck Institute for Plasma Physics / Department Stellarator Theory The Max Planck Institute for Plasma Physics has two branches located in Garching and Greifswald. Both are dedicated to experimental and theoretical research in Plasma Physics in general and nuclear fusion in particular. The aim is to confine an ioinized gas (a plasma), made of very energetic atoms, in a magnetic field to enable collisions strong enough to overcome the Coulomb barrier between hydrogen and tritium atoms. In consequence, the atoms fuse to helium. The institute runs two large fusions experiments: the ASDEX-Upgrade tokamak in Garching and the Wendelstein 7-X stellarator in Greifswald. The latter is the world largest and newest stellarator experiment. It is also the first stellarator, which was optimized using supercomputers. The physics necessary for understanding fusion comprises electromagnetism, magnetohydrodynamcis, kinetic theory, plasma-wall interaction, turbulence and many more. For theoreticians, supercomputer codes are essential to make reliable predictions as well as to design and optimize fusion machines as illustrated in this article.

Turbulence on supercomputers

Turbulence has been an age-old subject of fascination with countless applications in nature and technology. Today, supercomputer simulations provide us with detailed insights into turbulence, helping to tackle this long-standing problem.

AUTHORS *Michael Wilczek, Cristian C. Lalescu, and Bérenger Bramas*[1]

Turbulence is virtually everywhere. It governs our atmosphere and our oceans as well as their interaction. Turbulent flows feature a broad range of dynamically active scales with a continuous transfer of energy between them. Turbulence, therefore, plays an important role for the energy budget of geophysical flows, it triggers precipitation by enhancing droplet growth in clouds and provides the main mixing mechanism for combustion processes, pollutants in the atmosphere, urban environments, and marine microorganisms in the ocean.

Since turbulent flows are chaotic and random, theories of turbulence necessarily have to be of statistical nature. Despite decades of research, developing such theories, which ultimately all modeling applications rely on, remains one of the outstanding scientific challenges.

From a physics perspective, fully developed turbulence constitutes a paradigmatic problem of a strongly driven system with an enormous number of interacting degrees of freedom. The challenge of developing a statistical theory of turbulence arises from multi-scale flow structures, which introduce long-range correlations and give rise to complex, scale-dependent statistics. Numerical simulations can provide key insights into the emergence of these structures, their dynamics, and the resulting statistics.

TURBULENCE SIMULATIONS ON LARGE-SCALE HIGH-PERFORMANCE COMPUTING (HPC) INFRASTRUCTURES

For our simulation-driven theoretical investigations, we study highly turbulent flows from two complementary perspectives: The Eulerian frame, i.e., a fixed frame of reference, is well suited to investigate spatial features of turbulence. In contrast, the Lagrangian frame, i.e., following tracer particles, is particularly useful to study spatio-temporal phenomena such as turbulent mixing.

In collaboration with the MPCDF Application Support Group, we have developed the simulation framework TurTLE (Turbulence Tools: Lagrangian and Eulerian), a flexible pseudo-spectral solver for fluid and turbulence problems implemented in C++ with a hybrid MPI/OpenMP approach. TurTLE allows for an efficient tracking of a large class of particles by means of a parallel programming pattern that is easy to adapt and implement. TurTLE scales well up to tens of thousands of computing cores for problem sizes up to 4096^3 grid points and up to 10^9 particles.

For the largest problem size (4096^3) that we considered so far, the simulation of an integral time (characteristic time for large-scale flow features) requires several million core hours on the Max Planck supercomputer Cobra. A single flow field requires 1.65TB of disk space, such that a database spanning several integral time scales
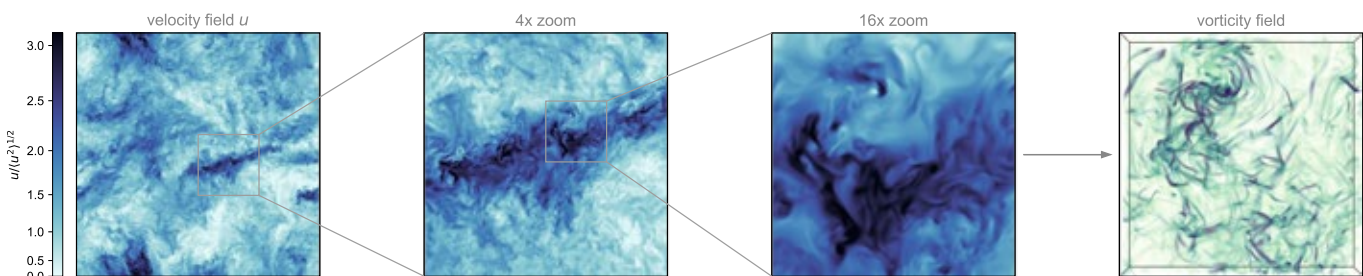
Fig. 1: Turbulence is a multi-scale phenomenon. A sequence of zooms into the velocity field (two-dimensional cuts) from one of our simulations with 4096^3 grid points reveals significant correlations on all scales. Steep gradients in the velocity field are closely related to small-scale coherent structures such as vortex filaments (volume rendering, right panel). The close connection between coherent structures and multi-scale correlations poses a major challenge in developing a statistical theory of turbulence.

[1] *Inria Nancy, France*
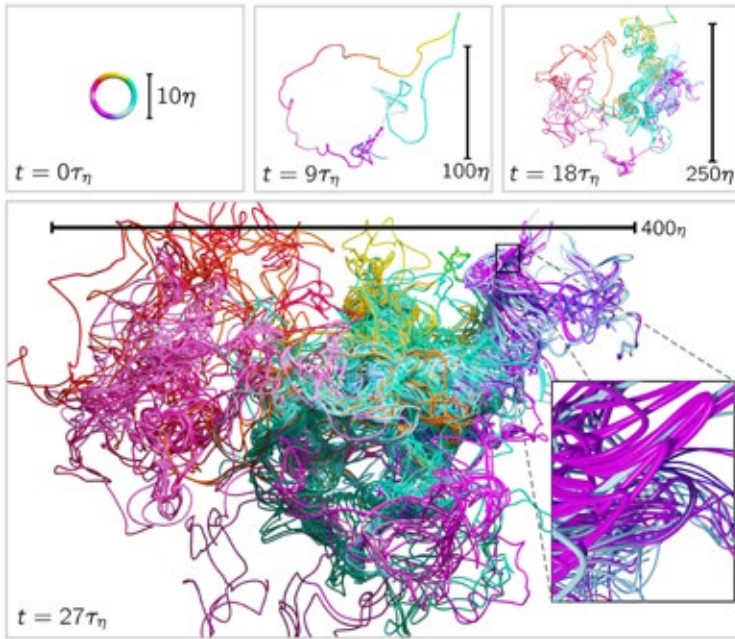
requires on the order of 100TB. Figure 1 is a visualization of the velocity field from one of our large-scale simulations. It illustrates the multi-scale nature of turbulence: while the velocity shows correlations comparable to the domain size, a considerable degree of small-scale structure in the form of coherent vortices is evident from these visualizations.

FROM FUNDAMENTAL RESEARCH TO APPLIED PROBLEMS

Over the past years, simulations on the MPCDF HPC infrastructures have been instrumental in informing new fundamental theories for turbulence and in tackling a number of applied problems.

For example, analyzing millions of particle tracks from large-scale simulations, we observed that the ensemble of tracer trajectories, based on their acceleration signature, can be categorized into classes of fluid particles with much simpler statistics. This led us to develop a comprehensive theoretical framework for single-particle statistics. For the first time, this work pointed toward the exciting possibility to finally unravel the complex statistics of turbulence in a data-driven theoretical approach, which inspires novel statistical field theories of turbulence. We also showed that the transport of tracer particles can be predicted from spatial features of the flow. Essentially, this means that a snapshot of a turbulent flow is already informative about how well turbulence mixes over time.

Recently we started focusing on geometric features of turbulence. By tracking the evolution of material loops, i.e., closed material lines, we showed how fluid elements are stretched and folded as they go with the flow, leading to very complex shapes, see Fig. 2. Remarkably, we were able to compute robust statistical features of such material lines. This work has implications, for example, for our understanding of magnetic field lines in astrophysical problems or the role of polymer stretching in turbulent flows.

Using the insights of our fundamental turbulence research, we also address a range of application-focused problems. For example, we investigate how turbulence enhances the encounter rates of microorganisms in the oceans. It turns out that small-scale turbulence has a significant impact on their swimming and sedimentation dynamics. This may significantly alter their encounter rates, therefore playing a role for mating, grazing, and the formation of planktonic colonies.

A quite complementary, yet surprisingly similar problem occurs in the formation of rain in clouds. Until today, the microphysical processes that let micro-droplets grow to rain drops are not fully understood. In the turbulent environment of clouds, the small-scale vortices act as tiny centrifuges, which accelerate particles into flow regions with less swirl. As a consequence, particles tend to cluster in those regions, significantly enhancing droplet collisions and coalescence. In an ongoing project, we are investigating this process using simulations with the aim of developing better predictions.

These are a few examples from our recent and ongoing work, which illustrate that the high-performance-computing infrastructures provided by the MPCDF are an indispensable tool for pursuing simulation-driven theoretical research. Extensive large-scale simulations will continue to provide the computational foundation for such investigations.

PUBLICATIONS

1. C. C. Lalescu and M. Wilczek. How tracer particles sample the complexity of turbulence. New J. Phys., 20:013001, 2018

2. L. Bentkamp, C. C. Lalescu, and M. Wilczek. Persistent accelerations disentangle Lagrangian turbulence. Nat. Commun., 10:3550, 2019

3. L. A. Leppin and M. Wilczek. Capturing velocity gradients and particle rotation rates in turbulence, Phys. Rev. Lett., 125:224501, 2020

4. L. Bentkamp, T. D. Drivas, C. C. Lalescu, and M. Wilczek. The statistical geometry of material loops in turbulence. arXiv:2106.11622, 2021

MAX PLANCK INSTITUTE
FOR DYNAMICS AND
SELF-ORGANIZATION

Max Planck Institute for Dynamics and Self-Organization / MPRG Wilczek From the fascinating self-organization of bacterial flows to large-scale flow patterns in the atmosphere – the dynamics of fluids and soft matter exhibit a host of intriguing phenomena. The aim of our research program is to develop a rigorous understanding of and predictive theories for a range of fluid and soft matter systems by combining theory and state-of-the-art computational methods.

ATLAS at the Large Hadron Collider: particle physics at the highest energy scales

ATLAS is a particle physics experiment built and operated by a large international collaboration at the Large Hadron Collider (LHC) at CERN. It is designed to explore the smallest constituents of matter and the fundamental forces of nature that govern the structure and the dynamics of the universe, by measuring and recording collisions of protons at the highest particle energies and collision rates ever being achieved. The handling of O(50 PB) of new scientific data each year, over a period of 30 years, requires significant resources and the participation of up to 100 computer centers distributed around the globe, with MPCDF as one of the LHC Tier-2 centers within the Worldwide LHC Computing Grid (WLCG).

AUTHORS *Siegfried Bethke and Stefan Kluth*

In the Large Hadron Collider (LHC) [1], dense bunches of protons are accelerated to particle energies of up to 6.5 TeV, corresponding to 6600 times the proton's rest mass. They collide at a rate of 40 MHz at four distinct collision points, which are surrounded by large particle detector systems. These detectors record the debris of particles that emerge from the proton-proton collisions at energy densities that resemble the conditions of the ultra-hot and early universe only 10^{-15} seconds after the Big Bang. Just a small fraction of these collisions are of actual interest for current research and a sophisticated filtering system (trigger) selects from an initial data stream of ca. 200 TB/s a few hundred events per second corresponding to about 1 GB/s of data for further analysis. These data are stored at CERN and distributed to 13 Tier-1 centers worldwide. Data reduction is performed at CERN and at the WLCG centers to yield a pre-selected compact format ready for analysis by the ATLAS scientists. These compact data sets are distributed to the Tier-2 centers according to the demand by the analysis groups.

SCIENTIFIC METHODS OF HIGH-ENERGY PARTICLE PHYSICS (HEP) Quantitative scientific investigation proceeds by comparing predictions of theoretical models with measurements extracted from the data. This allows to determine
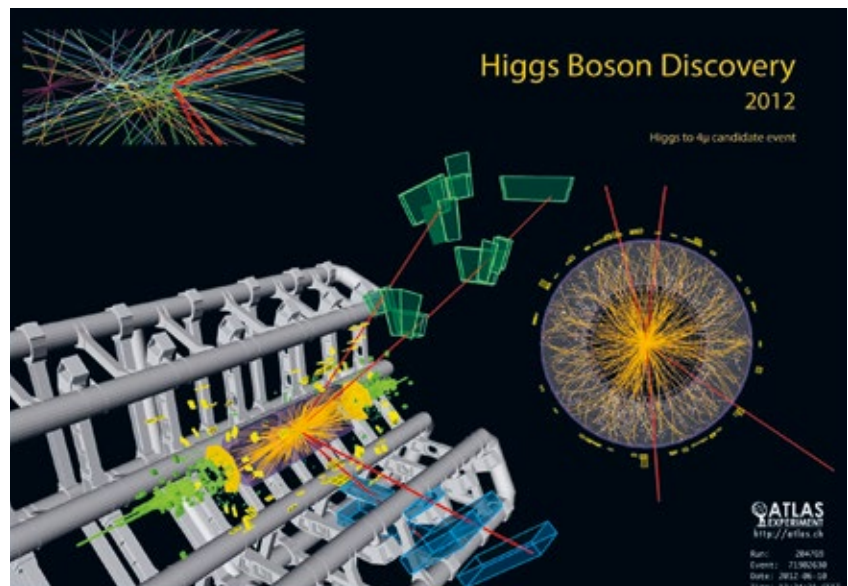
Fig. 1: Candidate event for a Higgs boson decaying to four muons, recorded with the ATLAS detector in 2012. The red lines mark the trajectories of the reconstructed muons, the yellow lines indicate all other reconstructed charged particles. Green dots and bars show energy deposits in the calorimeters. The blue and green boxes indicate the position of muon detectors that were hit by the muons. The upper-left insert shows a magnification of the proton-proton collision area inside the LHC beam pipe, together with extrapolations of all measured charged particle tracks, demonstrating that the four muons (in red) originate from one common vertex point, i.e., from one and the same high-energy collision of two protons.
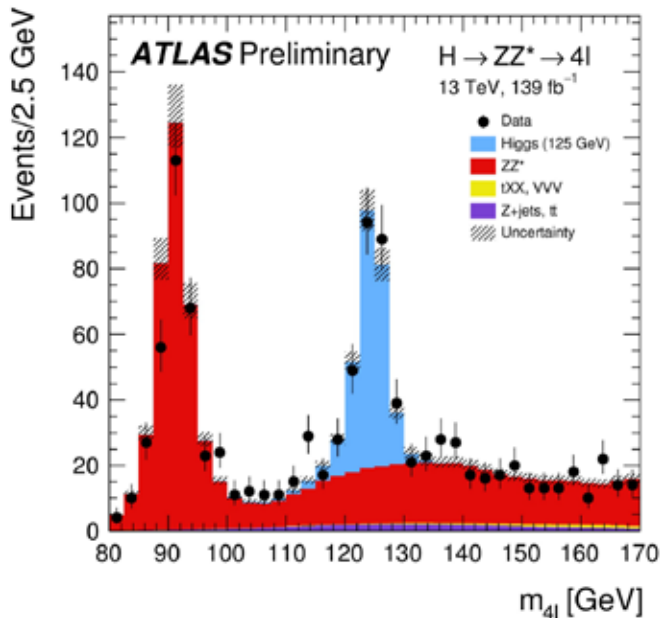
Fig. 2: *Distribution of invariant masses of four leptons measured with the ATLAS detector in proton–proton collisions at 13 TeV center–of–mass energy [5]. The black markers are the data, while the colored histograms show the predictions based on the SM without (red) and with (blue) a Higgs boson at m_H = 125 GeV.*

free parameters of the models and to perform precise tests of the theory and its validity to describe nature. Theories and models developed to address open fundamental questions are scrutinized and then discarded if they do not agree with the measurements.

Precision particle physics theory is complex and with few exceptions predictions can only be obtained numerically. Theoretical calculations and model predictions are computationally expensive and large runs use thousands of CPU cores and possibly GPUs for several weeks. A unique step in HEP is the generation of huge numbers of simulated collision events based on random processes predicted by theory models. However, before predictions extracted from simulated event samples can be compared to the data, the effects of event triggering and selection as well as of resolution and efficiency of the particle detectors must be taken into account. The ATLAS experiment [2] is a large apparatus, 45 m long and 25 m in diameter, with about 100 million electronic readout channels using semiconductor, gaseous and liquid detector technologies for measurements of the particle's momenta and energies. Simulating the response of ATLAS [3] can take about 1,000 seconds on a single CPU core for one complex collision event, thus the production of large event samples occupies more than 100,000 cores on the distributed resources of the WLCG including the MPP Tier-2 center at MPCDF.

DISCOVERY OF THE HIGGS BOSON [4] The Higgs boson is a fundamental component of the Standard Model of particle physics (SM), a complex quantum gauge-field theory of the electroweak and the strong interactions between elementary particles. However, up until 2012, its existence was not confirmed by experiment. It is predicted to be produced in high-energy proton-proton collisions at the LHC with a rate large enough for discovery with sufficiently large data samples. The Higgs boson is expected to instantly decay, among many other possibilities, into four leptons, grouped into two pairs of oppositely charged electrons or muons. These leptons can be efficiently identified in the ATLAS detector. Figure 1 shows a can-

didate for a Higgs boson decaying into four muons, together with many other particles produced in the same collision event. The physical properties of the Higgs boson can be derived from the kinematics of the four leptons into which it decayed. Figure 2 shows the distribution of invariant masses m_{4l} obtained from a combination of the four leptons in selected events in the data [5]. The structure at values around m_{4l} = 125 GeV is an indication of production of states decaying into four leptons over a background of combinations of unrelated leptons. The quantification of the significance of this structure heavily relies on the availability of large samples of simulated events including Higgs bosons decaying into four leptons, for different Higgs boson invariant masses, and of samples of simulated events without production of Higgs bosons. An indispensable prerequisite for its discovery in 2012 was the availability and functionality of the global Worldwide LHC Computing Grid with essential contributions of the MPP Tier-2 center at the MPCDF in Garching.

———

PUBLICATIONS
1. *The Large Hadron Collider, by Lynn Evans, New J. Phys. 9, 335 (2007)*
2. *The ATLAS Experiment at the CERN Large Hadron Collider, by ATLAS Collaboration (G. Aad et al.). 2008. JINST 3:S08003,2008*
3. *Simulation strategies for the ATLAS experiment at LHC, by ATLAS Collaboration (A. Rimoldi et al.), J.Phys.Conf.Ser. 331 (2011) 032026*
4. *Observation of a new particle in the search for the Standard Model Higgs boson with the AT–LAS detector at the LHC, by ATLAS collab. (G. Aad et al.), Phys.Lett. B716 (2012) 1–29*
5. *https://atlas.cern/updates/briefing/exploring–higgs–discovery–channels*

MAX PLANCK INSTITUTE FOR PHYSICS

Max Planck Institute for Physics (Werner Heisenberg Institute)
Research at the Max Planck Institute for Physics (MPP) addresses particle and astroparticle physics from both experimental and theoretical perspectives, with the mission to uncover and to investigate the basic building blocks of matter, the forces that govern them and their interactions with space and time. One main research activity in particle physics at accelerators is the participation in the ATLAS experiment at the LHC at CERN. The scientific foci of the ATLAS collaboration are the search for new physics and precision measurements and tests of the Standard Model of particle physics (SM). One major breakthrough was the discovery of the Higgs Boson, the last missing cornerstone of the SM in 2012, with significant contributions from scientists and technological developments from MPP.
.

Direct neutrino mass measurement with KATRIN

The Karlsruhe Tritium Neutrino (KATRIN) experiment is designed to measure the neutrino mass with a high sensitivity. This is achieved with a precise measurement of the tritium beta decay spectrum close to the endpoint, where the impact of the neutrino mass is maximal. The MPP group has developed one of the main analysis strategies based on the so-called Monte Carlo (MC) propagation technique. The analysis of the first KATRIN data sets were performed on the MPCDF and led to world-leading limits on the neutrino mass.

AUTHORS *Christian Karl, Susanne Mertens, Alessandro Schwemmer, and Martin Slezak*

The absolute value of the neutrino mass is one of the most pressing open questions in astroparticle physics. With a mass at least five orders of magnitude smaller than the mass of an electron, neutrinos are a clear misfit in the Standard Model (SM) of particle physics, which may be the key to physics beyond the SM.

The least model-dependent approach to assess the neutrino mass is based on the kinematics of a single beta decay. Here, the

Fig. 1: Imprint of a neutrino with non–zero mass on the theo–retical differential beta spectrum in the endpoint region.

neutrino mass manifests itself as a small distortion of the spectral shape in the close vicinity of the kinematic endpoint at 18.6 keV, see Figure 1. The leading experiment based on this approach is the Karlsruhe Tritium Neutrino (KATRIN) experiment.

KATRIN is designed to measure the neutrino mass with a sensitivity of 0.2 eV/c^2 at 90% confidence level. The experiment is situated at the Karlsruhe Institute of Technology and is operated by an international collaboration with about 150 members. The KATRIN setup consists of a 70m long beamline, combining a high-activity molecular tritium source (10^{11} Bq) with a high-resolution (eV-scale) spectrometer. This combination allows to obtain a high-statistics, high-precision measurement of the integral beta decay spectrum.

KATRIN celebrated its official inauguration in spring 2018 [1]. In 2019, the first neutrino mass campaign, still at reduced tritium activity, was performed, leading to an improved limit on the neutrino mass of m$_\nu$ < 1.1 eV/c^2 (90% CL) [2, 3] and first results on eV-scale sterile neutrinos [4]. Recently, in May 2021, the results of the second measurement campaign, which was performed at nominal tritium activity, were released. With this campaign we reach, for the first time in the history of direct neutrino mass measurements, a sub-eV sensitivity to the absolute neutrino mass of m$_\nu$ < 0.7 eV/c^2 (90% CL) [5].

One of the main analysis strategies for both neutrino mass measurements was developed at the MPP. The technique, described in detail below, relies on massive computing power, which was provided by the MPCDF.
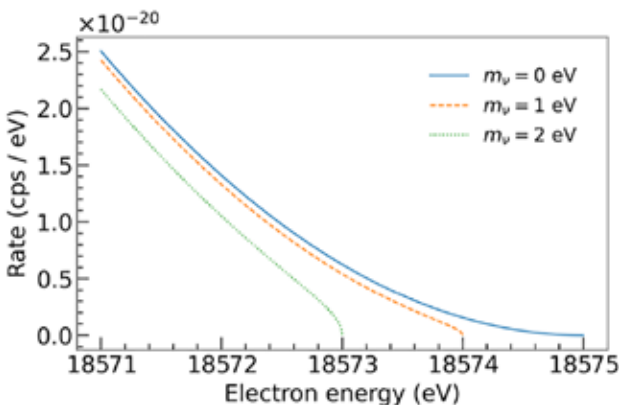
MONTE CARLO PROPAGATION TECHNIQUE To infer the neutrino mass from the measured integral spectrum, we fit a theo-
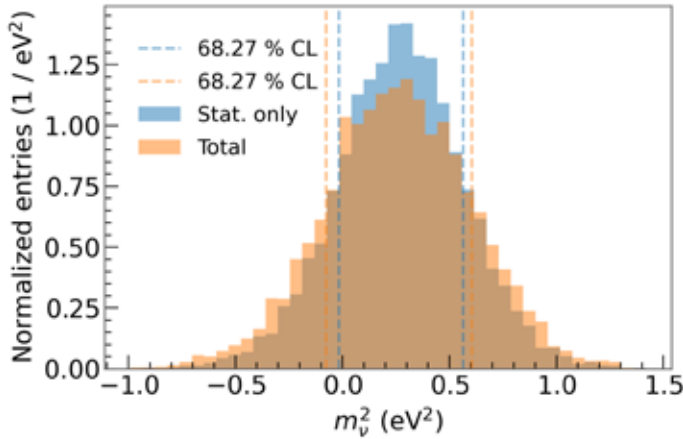
Fig. 2: Distribution of neutrino mass squared obtained by the Monte Carlo propagation technique. The blue histogram shows the distribution including only the statistical uncertainty, while the orange distribution includes both statistical and systematic uncertainty. The best fit is defined as the weighted median of the distribution and the 1–sigma uncertainty is obtained by integrating the distribution up to 16% from both sides.

retical prediction to the data. The theoretical prediction includes the theoretical differential tritium spectrum and the experimental response function. Its calculation involves various integrals and is computationally expensive. Moreover, the model contains various systematic parameters (such as tritium gas density, magnetic fields, etc.), which are known with a certain accuracy and their uncertainty needs to be propagated to the final neutrino mass result. The combination of a computationally expensive model calculation with a large number of systematic parameters motivates the development of a novel analysis technique, called Monte-Carlo propagation [6], and the extensive usage of supercomputers at the MPCDF.

The basic idea of the Monte Carlo propagation is to repeatedly fit the spectrum with random but fixed values for the systematic parameters. This allows pre-calculating the expensive response function as well as reducing the number of free parameters in the fit significantly at the cost of performing thousands of fits. As this process is embarrassingly parallel, the usage of multiple independent nodes at a computing cluster such as MPCDF's Draco makes the computation time feasible. As a result of this procedure, we obtain distributions of the fit parameters, from which we infer the final best fit (weighted median of the distribution) and the statistical and systematic uncertainties (integration of the distribution up to 16 % from both sides).

RESULT AND CONCLUSION The distributions of the best-fit parameters of interest (here: neutrino mass) obtained with MC propagation are illustrated in Figure 2, for the case of the second neutrino mass campaign. For the best fit we find $m^2_\nu = 0.26$ +/- 0.34 eV^2/c^4, translating to an upper limit of $m_\nu < 0.9$ eV/c^2 at 90% confidence level. This result is slightly higher than the sensitivity of the campaign, as the best fit result is positive. Combined with the first neutrino mass campaign we obtain an upper limit of $m_\nu < 0.8$ eV/c^2. This result is the first sub-eV limit on the neutrino mass from a direct measurement, and is highly complementary to other probes based on cosmology and the neutrino-less double beta decay.

PUBLICATIONS

1. *First operation of the KATRIN experiment with tritium, EPC–C 80, 264 (2020)*
2. *Aker, M. et al. Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN. PRL.123, 221802 (2019)*
3. *Aker, M. et al. Analysis methods for the first KATRIN neutrino-mass measurement. PRD 2101.05253. (2021)*
4. *Aker, M. et al. Bound on 3 + 1 Active-Sterile Neutrino Mixing from the First Four-Week Science Run of KATRIN. PRL126, 091803 (2021).*
5. *Aker, M. et al. First direct neutrino-mass measurement with sub-eV sensitivity, arXiv:2105.08533 [hep-ex]*
6. *Slezák, M. et al. Analysis strategies and treatment of systematic effects in the KATRIN experiment. Journal of Physics: Conference Series 1468, 012181, DOI: 10.1088/1742-6596/1468/1/012181 (2020)*
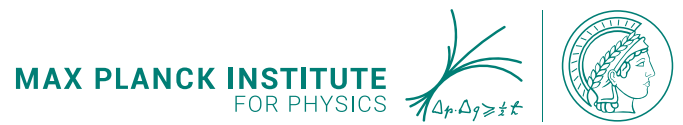
MAX PLANCK INSTITUTE FOR PHYSICS

Max Planck Institute for Physics (Werner Heisenberg Institute)
Research at the Max Planck Institute for Physics (MPP) addresses some of the most fundamental questions of Astroparticle physics. What is our universe made of? Why is there more matter than anti-matter in the universe? What is the origin of particle masses? The most abundant known matter particle in our universe, the neutrino, may help to unlock some of these mysteries. With the international Karlsruhe Tritium Neutrino (KATRIN) experiment, also sometimes called the „most precise scale of the world", the Max Planck Research Group of Prof. Mertens investigates the absolute mass of neutrinos. A team of about five students and postdocs in her group works on the KATRIN data analysis, developing new strategies to infer information on the elusive neutrino particle from the highly-complex KATRIN data. Since the start of data taking in 2019, the KATRIN collaboration succeeded in setting a new world-best limit on the neutrino mass.

Computer simulations for understanding strong electron correlation in biological, bio-inspired and solid state materials at the atomistic level

The main goal of computational quantum chemical simulations is to complement or anticipate experiments in understanding and predicting ground- and excited-state electronic structures of complex biological, bio-inspired and solid-state materials, at their equilibrium geometry, or during chemical reactions. The electronic structure theory department develops theory, algorithms and code necessary for achieving this goal and applies them to systems of increasingly large chemical, physical and computational complexity.

AUTHORS *Giovanni Li Manni and Ali Alavi*

Many biological catalysts (enzymes), bio-inspired catalysts, and attractive solid-state materials with potential application in superconductivity, sensing and batteries are based on transition metal (TM) clusters containing multiple metal centers, known as poly-nuclear transition metal (PNTM) clusters, or repeating transition metal units (crystals). In most of their common oxidation states, the TM centers exhibit a number of unpaired electrons, thus behaving as microscopic (at the atomic level) magnets. The interaction of these atomic magnets, mediated by the bridging units connecting them, determines at the macroscopic level the magnetic properties of these materials, and their catalytic activity. Exploring the local magnetic interactions in these materials is often experimentally challenging for a number of reasons, including radioactivity, toxicity, weak spectral signals or cumbersome synthetic procedures. Quantum chemical computer simulations, starting from the electronic Schrödinger equation, enable scientists to create and solve models that contain the main features of the real systems under investigation.

Broadly speaking, quantum chemical methods can be classified in single- and multi-configurational. In the former category, that contains the well-known Density Functional Theory (DFT) and the Hartee-Fock method, a single electronic configuration (distribution of electrons in orbitals) is considered and orbitals are variationally optimized in the mean field generated by this single-configurational wave function. The single-configurational nature of these methods represents their main advantage (computationally inexpensive and simple interpretation); however, it also represents their weakness, especially when studying strongly correlated materials that require a much higher degree of freedom in the representation of their wave functions.

In cases where more than one electronic configuration can be chosen as the reference (resonating structures, near-degenerate states, noncollinear spin-states in PNTM clusters) multi-configurational methodologies are to be chosen in order to obtain a qualitatively and quantitatively correct description of the modeled systems. Unfortunately, multi-configurational electronic structure simulations are computationally demanding and often require large-scale high-performance computing (HPC) resources. And yet, their applicability is generally limited to small and finite systems where electron correlation is very local, involving a handful of strongly correlated electrons, up to 18 electrons to be precise. Beyond 18 electrons exact optimization is not possible even on large HPC clusters, due to the exponential scaling of the computational requirements (prohibitively large memory and floating point operations demand) with respect to the number of correlated electrons.

Over the years, members of our department have been developing methods that accurately approximate the exact solution of the electronic Schrödinger equation, while retaining the flexibility that only multi-configurational procedures can guarantee. Of particular interest is the development of stochastic optimizations of large multi-configurational wave functions, started by Professor Ali Alavi, head of the department, through the development of the Full-Configuration Interaction Quantum Monte-Carlo (FCIQMC) algorithm [1].
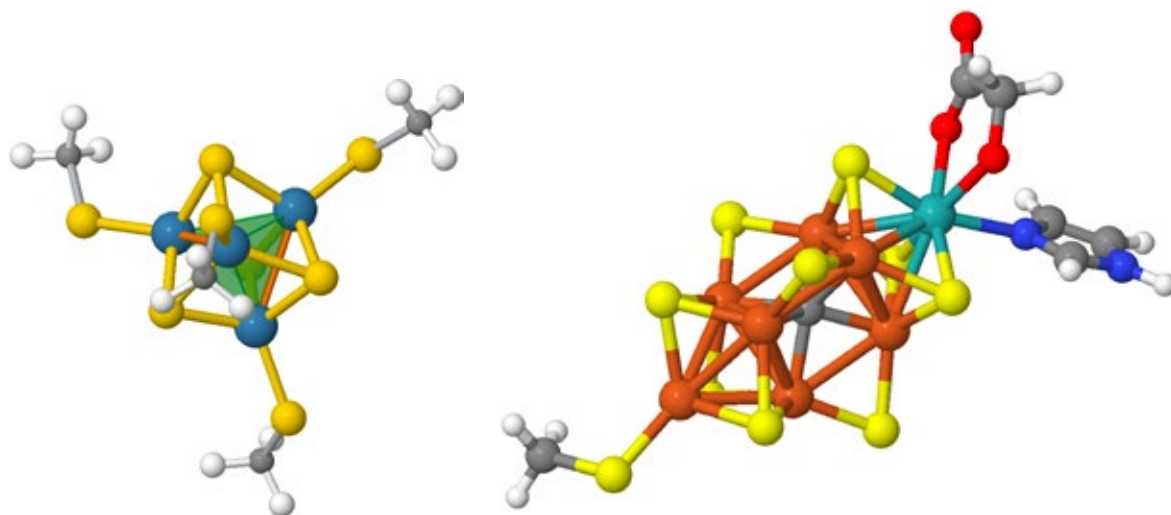
Two examples of polynuclear transition metal clusters. Left: the Fe_4S_4 cubane, with the Fe centers forming a distorted tetrahedron. Magnetic couplings (ferro/antiferro) within the two pairs (marked in orange) depend on the type of distortion. Right: The FeMo–cofactor consists of 8 metal centers... our next challenge!

A number of important contributions have made FCIQMC an accurate and practical tool to study PNTM clusters and other strongly correlated model systems.

The Stochastic-CASSCF [2] (complete active space self-consistent field) developed at the electronic structure theory department, allows to build a multi-configurational wave function by selecting an orbital window around the Fermi level, where the frontier orbitals are located, known as the "active space". All electronic configurations generated by combinatorially exciting the active electrons among the active orbitals form the basis of the CAS multi-configurational wave function. The optimization parameters (CI coefficients) are stochastically determined by FCIQMC, while the molecular orbitals are variationally optimized under the field generated by the CAS wave function, thus reducing the initial condition bias and maximizing the strong correlation effect within the active space. An FCIQMC implementation has been made available that is especially designed to target specific spin states, called spin-adapted FCIQMC [3]. This approach is key to study strongly correlated systems where spin is of utmost importance, such as exchange-coupled anti-ferromagnets, systems featuring spin frustrations, and systems that undergo spin-driven Jahn-Teller distortions [4].

Recently, the FCIQMC algorithm has also been embedded in a more general framework, where the effect of specific forms of electron correlation in solids and molecules can be selectively probed, such as super-exchange, metal-to-ligand and ligand-to-metal charge-transfer. This is of particular relevance in understanding and predicting the role of the environment (linkers, peripheral ligands, local defects) for the interaction among the magnetic centers [5].

With these tools in hand, we have been able to study Fe_4S_4 cubane systems of potential interest in the bio-mimetic nitrogen fixation process [4]. Our study has shown, for the first time, that there is a special interplay between geometrical distortions and the coupling of local spins in the energetically low-lying states, which is a core property for understanding their catalytic activity. These clusters are intrinsically frustrated anti-ferromagnets, and the

obtained spin structures together with the geometrical distortions represent two possible ways to release spin frustration (spin-driven Jahn-Teller distortion).

The fact that Fe(III)-based ferredoxins are dominated by local S = 5/2 spins has been known for decades in the inorganic chemistry community, starting from the works by Anderson and Hasegawa in the 1950s [6]. Through our methodologies we show for the first time a unique *locked-in* spin coupling within pairs of metal centers.

———

PUBLICATIONS
1. *Kai Guther et al., J. Chem. Phys. 153 (2020) 034107*
2. *G. Li Manni, S. D. Smart, A. Alavi, J. Chem. Theory Comput. 12 (2016) 1245*
3. *W. Dobrautz, S. D. Smart, A. Alavi, J. Chem. Phys. 151 (2019) 094104*
4. *(a) G. Li Manni, W. Dobrautz, A. Alavi, J. Chem. Theory Comput. 16 (2020) 2202. (b) G. Li Manni, W. Dobrautz, NA Bogdanov, K. Guther, A. Alavi, The journal of physical chemistry. A 125 (22), 4727*
5. *O. Weser, L. Freitag, K. Guther, A. Alavi, G. Li Manni, Int. J. Quantum Chem. 121 (2021) e26454*
6. *Anderson, P. W.; Hasegawa, H. Considerations on Double Exchange. Phys. Rev. 1955, 100, 675−681*

STUTTGART
MAX PLANCK INSTITUTE FOR SOLID STATE RESEARCH

Max Planck Institute for Solid State Research / Department of Electronic Structure Theory The basic research at the Electronic Structure Theory Department, headed by Prof. Ali Alavi, is aimed at developing electronic structure theories and methods for strongly correlated molecular and periodic systems. These are used to broaden the understanding of complex electronic structures in order to provide complementary knowledge for the prediction of chemical reactions, magnetic behavior and catalysis and to provide support for the design of novel materials.

Electronic correlations in superconducting nickelates

Strong electron-electron interactions play a pivotal role in the manifestation of superconductivity at high temperatures in doped cuprates. While finding superconductivity in iso-structural and iso-electronic nickel oxides is exciting, although not realized until after three decades of its discovery in cuprates, it is fascinating how different the electronic correlations are in the two seemingly identical systems.

AUTHORS *Vamshi M Katukuri and Ali Alavi*

(a)

(b)

	NdNiO$_2$	CaCuO$_2$
$d_{x^2-y^2}$	0.00	0.00
d_{xy}	−1.44	−1.70
$d_{xz,yz}$	−2.35	−2.21
d_{z^2}	−3.08	−2.73
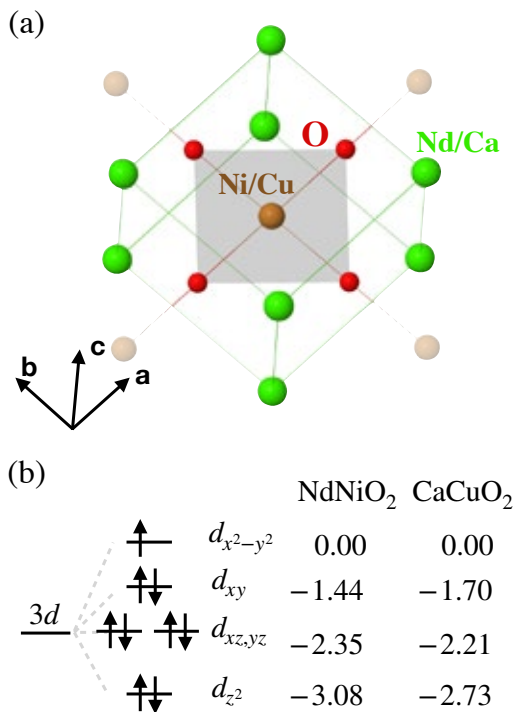
Fig. 1: (a) Crystal structure of infinite layered cuprate/nickelate superconductor parent compounds. (b) The d–level electronic multiplet structure of the d⁹ configuration in NdNiO$_2$ and CaCuO$_2$. The excitation energies shown are in units of eV.

The discovery of superconductivity in hole-doped infinite layer nickelates LaNiO$_2$ La = {Nd, Pr} [1, 2] is remarkable as these materials are isostructural and electronically similar to strongly correlated copper oxide compounds (cuprates) where high critical temperature (T_C) superconductivity was first observed. However, while T_C in cuprate superconductors is as high as 134 K under atmospheric pressures, T_C in nickelates is much smaller and in the range 10-15 K. The low T_C in nickelates is intriguing, raising questions on the microscopic understanding of the fundamental physical phenomena driving superconductivity in strongly correlated materials. Although there is no theory that can fully explain the origin of superconductivity in these systems, there is a large consensus that the antiferromagnetic (AF) correlations in the ground state (GS) are vital. The strong quantum spin fluctuations originating from the AF correlations take the role of "vibrations" analogous to phonons in conventional superconductors. The AF Néel order and corresponding fluctuations manifest themselves in cuprates from the localized moments of the Mott insulating GS that is a result of strong electron-electron interactions arising from the localized nature of the valence orbitals. In terms of the famous one-band Hubbard model, the AF exchange interaction is given by $J = -4\frac{t^2}{U}$, where t is the electron hopping amplitude from one orbital to the other and U is the one-site Coulomb repulsion when two electrons sit on the same orbital. In cuprates, the nearest-neighbor AF exchange is as high as 250 meV (in Sr$_2$CuO$_3$). In this work we investigated how similar or different the superconducting nickel parent compound NdNiO$_2$ is to the cuprate analogue CaCuO$_2$ [3]. We use advanced wavefunction quantum chemistry

NdNiO$_2$ CaCuO$_2$

— 0.04 < I < 0.08   — 0.008 < I < 0.04
— 0.004 < I < 0.008
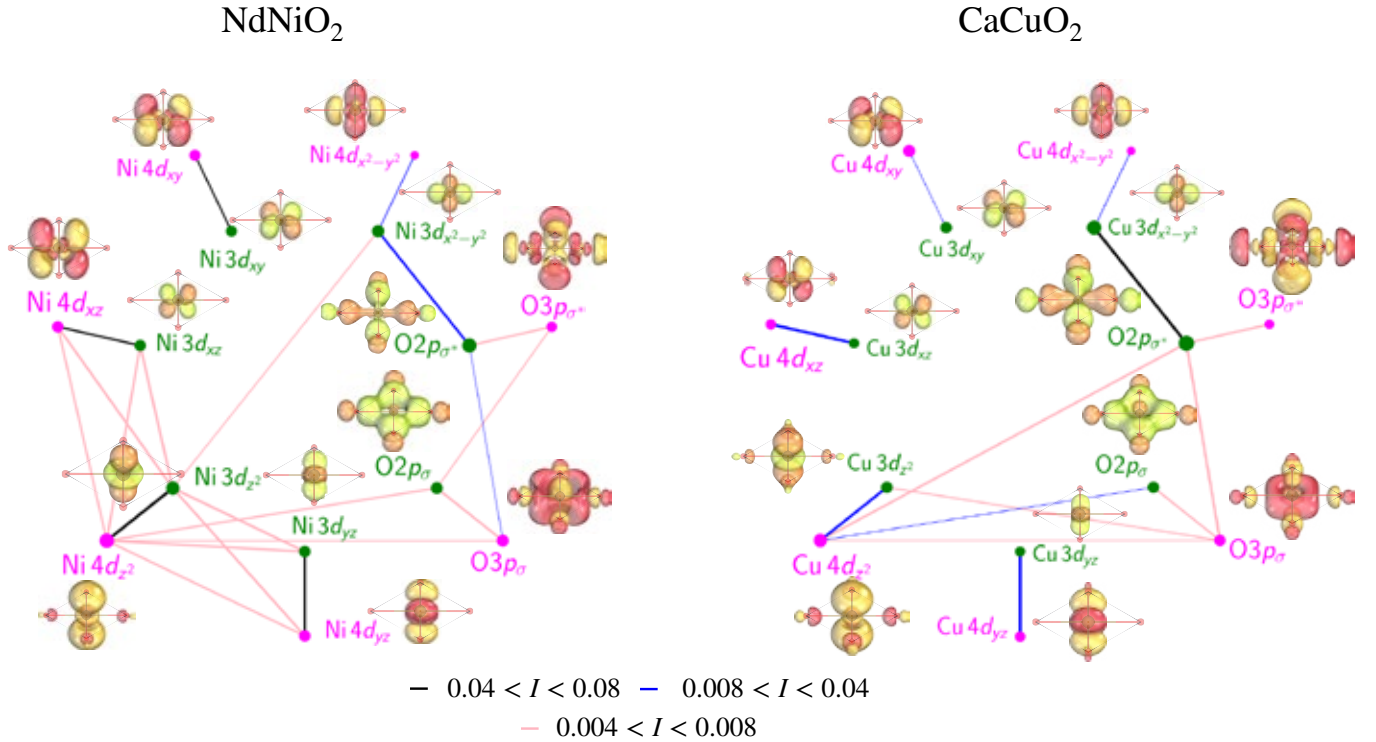
methods such as complete active space self-consistent field (CASSCF) theory using density matrix renormalization group theory and full configuration interaction quantum Monte Carlo as eigen solvers for elucidating the way electronic correlations manifest in these two material systems.

NICKELATES IN d⁹ CONFIGURATION In most inorganic solids Ni is stable in the +2 oxidation state with $3d^8$ valence configuration. In octahedrally coordinated compounds, the d^8 configuration results in a high spin configuration where two unpaired electrons are parallelly aligned in the crystal field split e_g orbitals. However, in the infinite layered crystal structure of NdNiO$_2$, the Ni^{+1} is stabilized albeit with a square-planar coordination [1] as shown in Figure 1a. The $3d$ multiplet structure of the Ni^{1+} ion in NdNiO$_2$ is strikingly similar to that of the Cu^{2+} ion in CaCuO$_2$, where a single hole is localized on the $3d_{x^2-y^2}$ orbital, see Figure 1b. Despite this, one noticeable difference is the excitations into the d_{z^2} orbital are almost 0.4 eV higher in the nickel compound. However, the GS wavefunction is quite different in the two compounds: the first three terms of the GS wavefunction for NdNiO$_2$ (NNO) and CaCuO$_2$ (CCO) expressed in hole excitations from the reference (first) configuration state function (CSF) are

$\Psi_{NNO} = 0.890|\downarrow\rangle_1 + 0.072|\downarrow\rangle_2|\uparrow\downarrow\rangle_1 + 0.068|\square\rangle_3|\downarrow\rangle_1|\uparrow\downarrow\rangle_4 + ...$

$\Psi_{CCO} = 0.896|\downarrow\rangle_1 + 0.131|\downarrow\rangle_2|\uparrow\downarrow\rangle_1 + 0.054|\uparrow\rangle_3|\downarrow\rangle_1|\downarrow\rangle_4 + ...$

where the subscripts 1,2,3 and 4 corresponds to $3d_{x^2-y^2}$, $2p_{\sigma^*}$, $3d_{z^2}$ and $4d_{z^2}$ -like orbitals, respectively, as shown in Figure 2, and $\uparrow$/$\downarrow$, $\uparrow\downarrow$ and $\square$ represent singly occupied, doubly occupied, and empty orbital states, respectively.

Fig. 2: Orbital entanglement diagram showing how correlations are different in NdNiO$_2$ and CaCuO$_2$. The green and pink colors signify occupied and unoccupied orbitals, respectively, in the reference CSF. The size of the dots corresponding to each of the orbitals is proportional to s(1) (0.047 ≤ s(1) ≤ 0.22) and the thickness of the lines connecting different orbitals is proportional to I$_{i,j}$.

While about 80% of the wavefunction is dominated by the CSF with a hole in the $3d_{x^2-y^2}$-like orbital in both the compounds, the rest of the composition of the wavefunction is quite different. The contribution from the O $2p_{\sigma^*}$ charge transfer (CT) CSF (second term) is considerably smaller in NdNiO$_2$, aproximately 0.5%, whereas in CaCuO$_2$ it is approximately 1.7%, over three times larger. The third term indicates a strong dynamic correlation (double excitation) between the $3d_{z^2}$- and $4d_{z^2}$ -like orbitals in the Ni compound, whereas a single excitation (orbital relaxation) with a smaller weight in the Cu compound.

ORBITAL ENTANGLEMENT ENTROPY When representing a wavefunction in a Slater determinant/CSF basis, the electron correlation effects can be quantified by the von Neumann entropy of a particular orbital (single orbital entropy, $s(1)_i$) or between any pair of

orbitals (mutual orbital information, $I_{i,j}$) in which the electrons reside. $s(1)_i$ quantifies the correlation between the ith orbital and the remaining set of orbitals, contained in the wavefunction expansion and $I_{i,j}$ illustrates the correlation of an orbital with another, in the embedded environment comprising all other orbitals. $s(1)_i$ takes a maximum of $\ln(4) \sim 1.39$ when all four possible occupations of an orbital are equally probable. Thus, an orbital with a large $s(1)$ experiences strong charge fluctuations, implying a strongly correlated orbital. In Figure 2, $s(1)_i$ (dots) and $I_{i,j}$ (lines) computed from CAS(35,36)SCF (see [3] for details) wavefunctions are shown. While the O $2p_{\sigma*}$ charge transfer is much smaller in NdNiO$_2$, $I_{i,j}$ between the $3d$- and $4d$- like orbitals is significantly larger compared to CaCuO$_2$, implying the presence of strong dynamic correlation within the d orbital manifold in NdNiO$_2$.

EXCHANGE INTERACTIONS The exchange (J) in these systems is primarily of the superexchange type that depends on the virtual hopping of electrons (or holes), t, through the bridging oxygen and the effective on-site Coulomb repulsion U_{eff} on the 3d orbitals of the Ni (Cu) ions. The J values computed from CAS(24,26)SCF plus perturbative correction for dynamical correlation calculations are shown in Table 1. There are two competing effects that result in a smaller value of J in NdNiO$_2$. While a smaller O $2p_{\sigma*}$ charge transfer CSF contribution results in a smaller t and hence a smaller J, on the other hand, due to the increased dynamical correlation U_{eff} is reduced in NdNiO$_2$, enhancing J. Interestingly, the latter many-body effect is crucial for the unusually large J in a nickel oxide, e.g., a mean-field type calculation of J results is a much smaller value of 10 meV [4].

Compound	NdNiO$_2$	CaCuO$_2$
J (meV) (AF)	77.6	208.1

Table 1: Computed exchange

In conclusion, while the nickelates superconductor parent compounds are very similar to the cuprate analogues, the interplay of microscopic interactions is very different. The effective onsite Coulomb interactions are smaller in d^9 nickelates, resulting in a larger than expected exchange even though the Ni $3d$ – O $2p$ hybridization is small.

PUBLICATIONS

1. *D. Li, K. Lee, B. Y. Wang, M. Osada, S. Crossley, H. R. Lee, Y. Cui, Y. Hikita, and H. Y. Hwang, Superconductivity in an Infinite-Layer Nickelate, Nat. Lond. 572, 624 (2019)*
2. *M. Osada, B. Y. Wang, B. H. Goodge, K. Lee, H. Yoon, K. Sakuma, D. Li, M. Miura, L. F. Kourkoutis, and H. Y. Hwang, A Superconducting Praseodymium Nickelate with Infinite Layer Structure, Nano Lett. acs.nanolett.0c01392 (2020)*
3. *V. M. Katukuri, N. A. Bogdanov, O. Weser, J. van den Brink, and A. Alavi, Electronic Correlations and Magnetic Interactions in Infinite-Layer NdNiO₂, Phys. Rev. B 102, 241112 (2020)*
4. *M. Jiang, M. Berciu, and G. A. Sawatzky, Critical Nature of the Ni Spin State in Doped NdNiO₂, Phys. Rev. Lett. 124, 207004 (2020)*

STUTTGART
MAX PLANCK INSTITUTE FOR
SOLID STATE RESEARCH

Multi-physics simulations with Octopus

In our recent research, we have developed an extension of density-functional theory to quantum electrodynamics. This approach allows for a self-consistent forward-backward coupling of the electromagnetic field to a density functional description of matter. Here, we briefly summarize the main features and our implementation of this approach in the real-space real-time code Octopus.

AUTHORS *Heiko Appel and Angel Rubio*

Traditionally, the description of light-matter interactions has been treated in different research fields with different emphasis. Disciplines like solid-state physics or quantum chemistry focus on the matter description and often either neglect the electromagnetic field, or treat it perturbatively, or only as external and unmodified input. The back-reaction of the matter on the radiation field is in most cases neglected. In quantum optics on the other hand, a strong focus is placed on the quantum nature of the electromagnetic field and the matter description is often simplified with few-level systems, preventing a full microscopic picture of molecules and solids [1].

COUPLING MAXWELL'S EQUATIONS TO FIRST-PRINCIPLES TIME-DEPENDENT DENSITY-FUNCTIONAL THEORY The goal of the present research project is to bridge the gap between different disciplines and to develop a coherent self-consistent coupling of
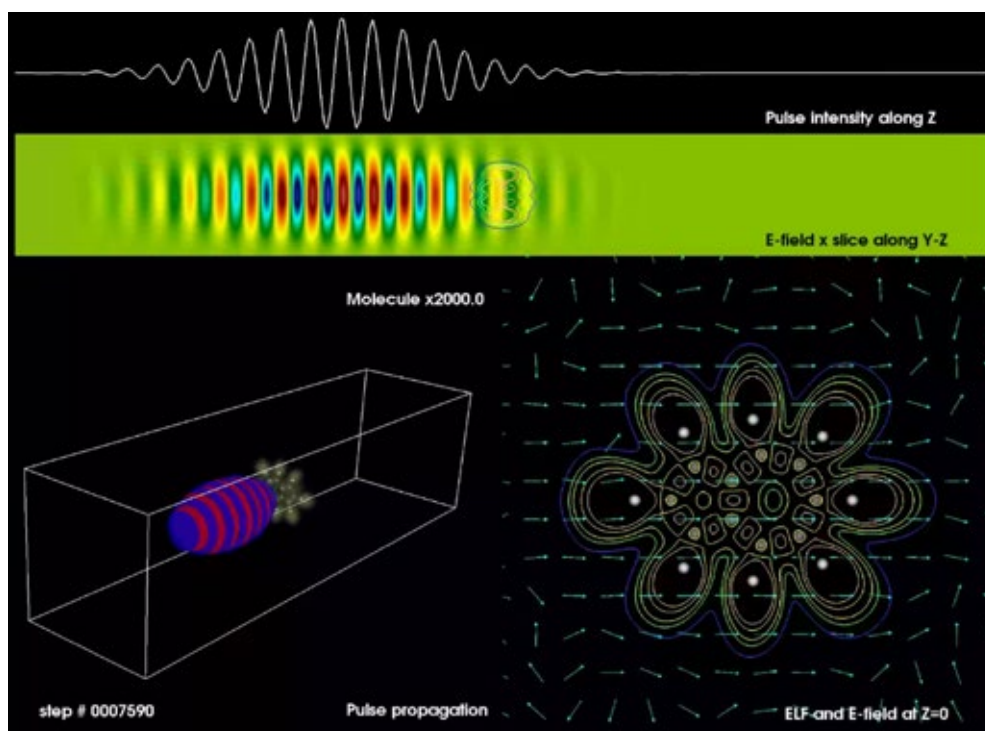
Fig. 1: The image shows a snapshot in time of the interaction of an Azulene molecule with a laser pulse. The panel on the lower left of the image shows the electric field amplitude of the laser pulse in real-space, and the top two panels illustrate respective cuts of this pulse in two and one dimension along the propagation axis. The panel in the lower right shows the electron localization function of the Azulene molecule and the positions of the atomic nuclei. As overlay, we display vector arrows, which indicate the local electric field surrounding the molecule.
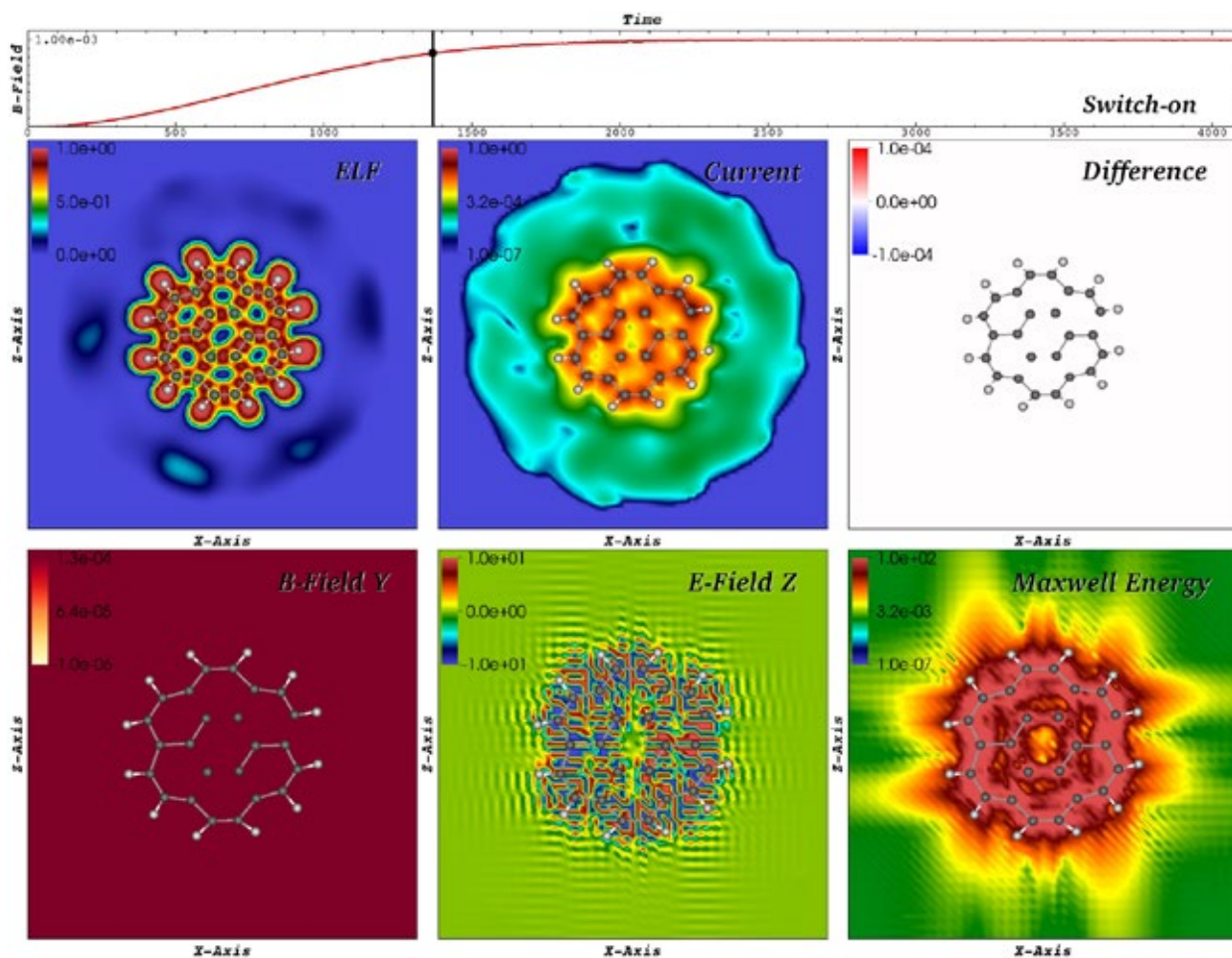
Fig. 2: These images show a snapshot in time of a Coronene molecule in a strong magnetic field perpendicular to the plane of the molecule. The top panel shows the ramping function, which is used to increase the magnitude of the magnetic field over time. The three panels in the center row show from left to right the electron localization function, the electronic current, and the positions of the atomic nuclei. In the bottom row, the three panels show the (uniform) external magnetic field, the electric field and the Maxwell energy density. While the current is ramped up, strong ring currents are introduced in the aromatic ring structure of the molecule. At a critical value of the external magnetic field, the electrons are ionized and the resulting lack of charge results in a rearrangement of the chemical bonds.

light with matter at a full *ab initio* level. To this end, we have developed an extension of time-dependent density-functional theory (TDDFT) to quantum electrodynamics (QED) [2, 3]. This approach is a formally exact reformulation of non-relativistic QED and allows treating the quantum nature of matter and the quantum nature of the electromagnetic field on an equal footing. Applications of this approach range from polaritonic chemistry, nano-plasmonics, and tip-enhanced spectroscopies to the description and tailored design of novel states of matter in optical cavities. Moreover, the present setup allows us to model on the same footing both the spectroscopic properties of matter and the effect on the photon sector. This allows us to study scattering of linear and nonlinear light phenomena, which so far has not been accessible in first-principles calculations.

We have implemented the real-time version as well as the perturbative limits [4] of this density-functional approach for QED in the real-time real-space density-functional code Octopus [5]. In the classical limit for the electromagnetic field, our approach simplifies to the self-consistent forward-backward coupling of Maxwell's equations to the time-dependent Kohn-Sham equations [3]. The forward-backward coupling of Maxwell's equations to a density-functional description of matter gives rise to a multi-scale problem in space and time and at the same time two entirely different subsystems have to be self-consistently coupled and propagated in real-time. In recent experimental collaborations with the groups of Martin Wolf at the Fritz Haber Institute in Berlin and Rupert Huber at the University of Regensburg, this approach has already provided useful theoretical insight [6].

NOVEL MULTI-PHYSICS IMPLEMENTATION IN OCTOPUS
The aspects arising in a multi-scale self-consistent coupling of different subsystems, as occurring in the Maxwell-TDDFT propagation, has triggered us to generalize the concept of subsystems and their mutual coupling, and we have started a new multi-physics implementation in Octopus. This implementation allows the creation of a

multitude of new subsystems for which explicit forward and backward couplings can be defined. Our new framework then allows propagating all these subsystems and their couplings in real-time. So far, examples of possible subsystems that are already included in Octopus are electrons on the TDDFT level, electrons on the density-functional tight-binding level, classical particles, Maxwell subsystems, and subsystems using the auxiliary differential equation approach for dispersive media. The creation of new subsystems has been made very flexible using the object oriented features of Fortran 2003. In addition, subsystems can be grouped, arranged in a graph topology, and multiple replicas of a given subsystem with different initial conditions or type-bound data can be generated and propagated simultaneously. Based on these features, methods such as the recently proposed multi-trajectory Maxwell Ehrenfest [6], or schemes like replica-exchange molecular dynamics, ring-polymer dynamics, or subsystem density-functional theory are almost trivial to implement. In addition, any external code which can be linked as a library can be included as an own subsystem. One example for such a case is the LAMMPS code opening the possibility to couple to molecular mechanics simulations.

Combining these subsystems and their mutual couplings at various levels now allows performing novel multi-physics *ab initio* simulations with Octopus that have not been possible before. In addition, these developments are ideally suited for a parallelization over (sub-)systems and hence for large-scale simulations towards exascale computing.

COLLABORATION WITH THE MPCDF All aspects of the new implementations mentioned above have involved a frequent and continued collaboration with experts from the MPCDF over the past years. The fruitful collaboration with the MPCDF application group has made it possible to implement new GPU kernels for frequently occurring computations, created many regression test tools for our continuous integration environment, and led to fundamental code improvements for scalability and performance. In addition, an extensive visualization toolkit has been developed over time, which was also used to generate the images in Figures 1 and 2 from the corresponding Octopus simulations.

PUBLICATIONS

1. *From a quantum-electrodynamical light-matter description to novel spectroscopies, M. Ruggenthaler, N. Tancogne-Dejean, J. Flick, H. Appel, A. Rubio Nature Reviews Chemistry 2 (3), UNSP 0118 (2018)*

2. *Quantum-electrodynamical density-functional theory: bridging quantum optics and electronic-structure theory, Michael Ruggenthaler, Johannes Flick, Camilla Pellegrini, Heiko Appel, Ilya V. Tokatly, and Angel Rubio Phys. Rev. A 90, 012508 (2014)*

3. *Light-matter interactions within the Ehrenfest-Maxwell-Pauli-Kohn-Sham framework: fundamentals, implementation, and nano-optical applications, R. Jestädt, M. Ruggenthaler, M. J. T. Oliveira, A. Rubio, H. Appel Advances in Physics 68 (4), 225-333 (2019)*

4. *Light-Matter Response in Nonrelativistic Quantum Electrodynamics J. Flick, D. Welakuh, M. Ruggenthaler, H. Appel, A. Rubio ACS Photonics 6 (11), 2757-2778 (2019)*

5. *Octopus, a computational framework for exploring light-driven phenomena and quantum dynamics in extended and finite systems N. Tancogne-Dejean, M. J. T. Oliveira, X. Andrade, H. Appel, C. H. Borca, G. Le Breton, F. Buchholz, A. Castro, S. Corni, A. A. Correa, U. de Giovannini, A. Delgado, F. G. Eich, J. Flick, G. Gil, A. Gomez, N. Helbig, H. Hübener, R. Jestädt, J. Jornet-Somoza, A. H. Larsen, I. V. Lebedeva, M. Lüders, M. A. L. Marques, S. T. Ohlmann, S. Pipolo, M. Rampp, C. A. Rozzi, D. A. Strubbe, S. Sato, C. Schäfer, I. Theophilou, A. Welden, A. Rubio The Journal of Chemical Physics 152 (12), 124119 (2020)*

6. *Quantitative sampling of atomic-scale electromagnetic waveforms D. Peller, C. Roelcke, L. Z. Kastner, T. Buchner, A. Neef, J. Hayes, F. Bonafé, D. Sidler, M. Ruggenthaler, A. Rubio, R. Huber, J. Repp Nature Photonics 15 (2), 143-147 (2021)*

7. *Capturing vacuum fluctuations and photon correlations in cavity quantum electrodynamics with multitrajectory Ehrenfest dynamics N. Hoffmann, C. Schäfer, A. Rubio, A. Kelly, H. Appel Physical Review A 99 (6), 063819 (2019)*

mpsd

Max Planck Institute for the Structure and Dynamics of Matter, Theory department The research of the Max Planck Institute for the Structure and Dynamics of Matter in Hamburg is focussed on dynamical phenomena within matter down to femtosecond or attosecond time scales to resolve atomic and electronic motions. To reach insight in these regimes, short wavelength ultrafast probes, such as X-rays or electron pulses, are employed, which are capable of measuring atomic and electronic structures in matter of all kinds.

The research of the Theory department is embedded in this general focus of the institute and is in particular concerned with the electronic and structural properties of advanced materials, nanostructures and molecular complexes. Besides this, the department is developing novel theoretical tools and computational codes to investigate and control the electronic response of such systems to arbitrary time-dependent external electromagnetic (quantum) fields. The research in the department aims to provide a detailed, efficient, and at the same time accurate microscopic approach for the *ab initio* description and control of the dynamics of decoherence and dissipation in quantum many-body systems and seeks to characterize new non-equilibrium states of matter.

Twisted bilayer systems

Atomically-thin layers of maters stacked with a small twist angle has recently emerged as an exciting path to study correlations between electrons. Here we present how we investigate these challenging systems by means of first-principle simulations, using the Octopus code.

AUTHORS *Nicolas Tancogne–Dejean and Angel Rubio*

Two-dimensional materials are fascinating objects in the field of condensed matter research, and these materials exhibit a variety of physical effects. One example is the recent breakthrough of twisted bilayer graphene, where two layers of graphene are stacked together. Depending on the angle formed by the two layers, the system can, for instance, become superconducting. However, most existing phenomena, like superconductivity, only occur at small twist angles, making the theoretical modeling of these systems challenging. Due to the large size of the simulation cell (with many thousands of atoms) and the fact that electron correlations beyond mean field need to be accounted for, the theoretical modeling of these systems is very demanding. In the following, we briefly describe how we treat the correlation effects and how we used this framework to investigate twisted bilayer boron nitride (TBBN) using the real-space real-time code Octopus (https://octopus-code.org).

BRIDGING THE HUBBARD MODEL AND DENSITY-FUNCTIONAL THEORY The Hubbard model and derived many-body interacting model Hamiltonians are the cornerstone of numerical methods targeting the description of strongly correlated electrons. While these methods, such as frozen renormalization group (FRG), can accurately solve these models and precisely describe correlation-induced effects such as superconductivity, these methods are limited by the fact that effective electronic parameters entering the low-energy Hamiltonian need to be obtained from other theories or fitted to experimental results. In order to allow for studying twisted bilayer systems from an *ab initio* perspective, we recently developed numerical tools allowing to bridge density-functional theory (DFT) and a strongly correlated method based on model Hamiltonians. These tools are based on the DFT+U method, in which DFT is complemented by a Hubbard U describing on-site interactions. However, unlike most approaches in which the Hubbard U is specified as an input of the simulation, we implemented in the Octopus code [1], a framework allowing us to evaluate the effective electronic parameters *ab initio* and self-consistently [2, 3]. Using this framework, it is possible to evaluate the effective electronic parameters *ab initio*, and to then use methods like FRG to solve accurately an effective Hamiltonian and to explore the phase diagram of strong-correlated materials.
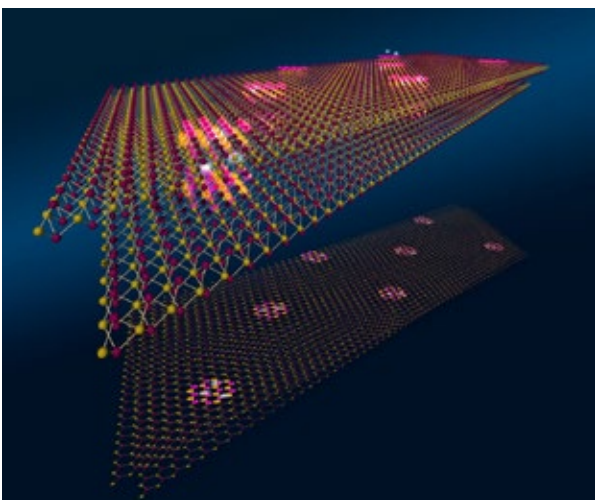

Fig. 1: Schematic of twisted bilayer WSe_2, where a correlated insulator phase and a possible superconducting transition emerge at a relatively wide range of small twist angles.
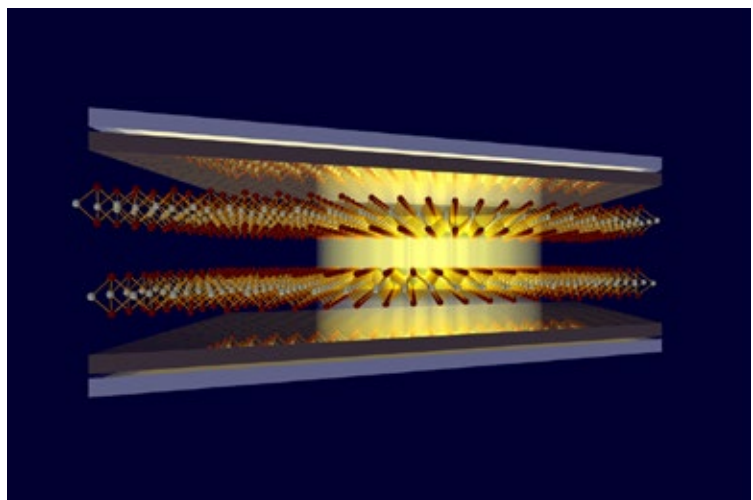

Fig. 2: Twisted bilayer system inside a cavity, as a new platform to investigate strong coupling between light and electrons.
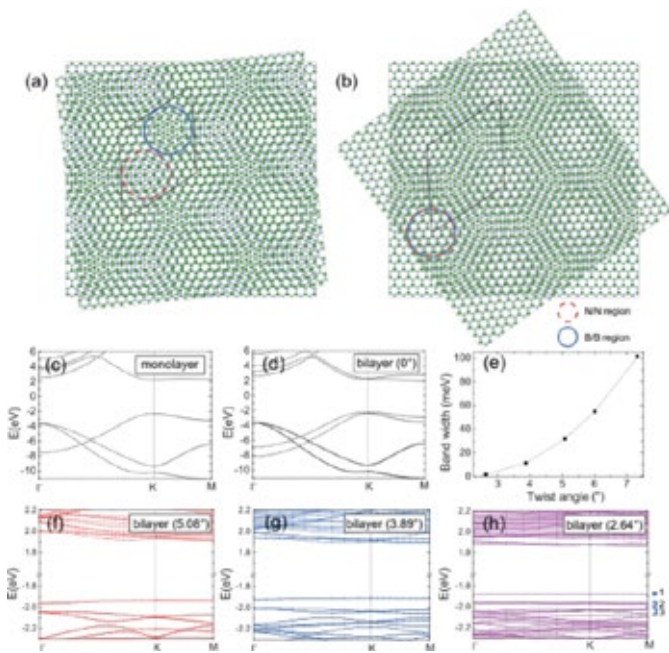
Fig. 3: Atomic and electronic structure of twisted bilayer boron nitride. a), b) Schematic illustration of the two possible configurations in twisted bilayer hBN that have the same Moiré pattern. c)–d), f)–h) Band structure of twisted bilayer hBN for different twist angles. e) Evolution of the bandwidth of the so-called flat band versus the twist angle.

PUBLICATIONS

1. Octopus, a computational framework for exploring light-driven phenomena and quantum dynamics in extended and finite systems N. Tancogne-Dejean, M. J. T. Oliveira, X. Andrade, H. Appel, C. H. Borca, G. Le Breton, F. Buchholz, A. Castro, S. Corni, A. A. Correa, U. de Giovannini, A. Delgado, F. G. Eich, J. Flick, G. Gil, A. Gomez, N. Helbig, H. Hübener, R. Jestädt, J. Jornet-Somoza, A. H. Larsen, I. V. Lebedeva, M. Lüders, M. A. L. Marques, S. T. Ohlmann, S. Pipolo, M. Rampp, C. A. Rozzi, D. A. Strubbe, S. Sato, C. Schäfer, I. Theophilou, A. Welden, A. Rubio The Journal of Chemical Physics 152 (12), 124119 (2020)
2. N. Tancogne-Dejean, M. J. T. Oliveira and A. Rubio, Self-consistent DFT+U method for real-space time-dependent density functional theory calculations Phys. Rev. B 96, 245133 (2017)
3. N. Tancogne-Dejean, A. Rubio, Octopus, a computational framework Parameter-free hybridlike functional based on an extended Hubbard model: DFT+U+V Physical Review B 102, 155117 (2020)
4. Moiré heterostructures as a condensed-matter quantum simulator D. M. Kennes, M. Claassen, L. Xian, A. Georges, A. J. Millis, J. Hone, C. R. Dean, D. N. Basov, A. N. Pasupathy, and Angel Rubio Nature Physics 17, 155–163 (2021)
5. Multiflat Bands and Strong Correlations in Twisted Bilayer Boron Nitride: Doping-Induced Correlated Insulator and Superconductor L. Xian, D. M. Kennes , N. Tancogne-Dejean, M. Altarelli, and A. Rubio Nano Letters (2019)
6. Enhanced tunable second harmonic generation from twistable interfaces and vertical superlattices in boron nitride homostructures. K. Yao, N. R. Finney, J. Zhang, S. L. Moore, L. Xian, N. Tancogne-Dejean, et al. Science Advances 7, eabe8691 (2021)

TWISTED BILAYER BORON NITRIDE Among the possible systems of interest, the twisted bilayer systems represent a particularly exciting class of materials. In fact, these heterostructures can be seen as a condensed-matter quantum simulator [4], in which a large number of fundamental many-body quantum models of condensed-matter physics can be realized. Based on this, we propose that these systems can be used as a robust quantum simulation platform that enables the study of strongly correlated physics and topology in quantum materials (making it possible to realize new phases of matter broadening the landscape of accessible physics and holding a promise for future technological applications).

Among all possible materials of interest, we decided to investigate twisted bilayer hexagonal boron nitride (hBN) [5]. The atomic and electronic structure of this system is shown in Figure 1. As one can see, for small twist angles, relevant for the correlated physics, the unit cell of the system quickly becomes very large, making it challenging to study using real-space DFT(+U). Thanks to the collaboration with the MPCDF over the past few years, we have been able to perform simulations of this system up to a twist angle of 2.64°. This required improving the algorithms used to iteratively determine the ground state of the system, as well as optimizing the calculation. For each twisted angle, we evaluated the Hubbard U, and we then used FRG to compute the phase diagram of twisted bilayer hBN [5].

Those theoretical and algorithm developments have enabled the study of their nonlinear optical properties [6], that should be extended to explore the possibility of exciton localization and condensation in these systems (see Fig. 1), as well as the description and control of out-of-equilibrium correlation phenomena triggered by light pulses. This plethora of exciting new physics can be extended to add the quantum nature of the photon field (quantum vacuum fluctuations) by embedding them in optical cavities (see Fig. 2 as example of the new field of cavity-twistronics that the group is leading). Many interesting phenomena are waiting to be unraveled!

mpsd

Max Planck Institute for the Structure and Dynamics of Matter, Theory department The research of the Max Planck Institute for the Structure and Dynamics of Matter in Hamburg is focussed on dynamical phenomena within matter down to femtosecond or attosecond time scales to resolve atomic and electronic motions. To reach insight in these regimes, short wavelength ultrafast probes, such as X-rays or electron pulses, are employed, which are capable of measuring atomic and electronic structures in matter of all kinds.

The research of the Theory department is embedded in this general focus of the institute and is in particular concerned with the electronic and structural properties of advanced materials, nanostructures and molecular complexes. Besides this, the department is developing novel theoretical tools and computational codes to investigate and control the electronic response of such systems to arbitrary time-dependent external electromagnetic (quantum) fields. The research in the department aims to provide a detailed, efficient, and at the same time accurate microscopic approach for the *ab initio* description and control of the dynamics of decoherence and dissipation in quantum many-body systems and seeks to characterize new non-equilibrium states of matter.

Overcoming barriers in nuclear dynamics with first principles electronic structure methods

We develop new computational techniques that, coupled with increased computational power and new machine learning methods, allow the determination of finite-temperature structures of weakly bonded systems and interfaces, accounting for nuclear quantum effects and based on the first principles of quantum mechanics.

AUTHORS *Mariana Rossi, Alaa Akkoush, Karen Fidanyan, Marcin Krynski, Alan Lewis, Yair Litman, Dmitrii Maksimov, Eszter Pos, Nathaniel Raimbault, and Haiyuan Wang*

The research projects in the Lise Meitner group "Simulations from *Ab Initio* Approaches (SAbIA)" led by Dr. Mariana Rossi, focus on understanding the impact of temperature and nuclear quantum effects in electronic-structure calculations. The nuclear fluctuations induced by temperature can completely change the structure, the thermodynamic equilibrium, and also the electronic structure of materials, especially in weakly bonded systems. In addition, given the ubiquitous presence of light nuclei in organic materials, not only the electrons but also the nuclei must be treated

within the first principles of quantum mechanics. We achieve an all-atom, all-electron quantum description of these systems by developing methodologies and tools that join density-functional-theory calculations to path-integral methods. We belong to the core developers of the FHI-aims electronic-structure package (https://fhi-aims.org/) and the i-PI software (http://ipi-code.org/), that performs nuclear dynamics. This combination has allowed us to answer questions related to structural properties and vibrational fingerprints of complex materials, level broadening in semiconductor interfaces, isotope effects in electron-phonon coupling and apparently contradictory observations in hydrogen-transfer reactions. None of these investigations would have been possible if the atomic structure of the systems was unknown or incorrect. Hence, we also conduct *ab initio* structure searches of single molecules and of self-assembled monolayers within the frameworks that allow for conformational flexibility.

The research performed in the group relies on a large amount of computational resources, due to the system sizes involved and the complexity of solving quantum-mechanical problems for a high number of relevant degrees of freedom. Therefore, we also invest in optimizing code packages for massively parallel architectures and the facilities of MPCDF are fundamental for realizing the development and applied work in the group.

QUANTUM STRUCTURE AND DYNAMICS OF NUCLEI *With: Yair Litman, Karen Fidanyan and Eszter Pos /* The inclusion of nuclear zero-point-energy and tunneling effects in high-dimensional anharmonic systems represents a considerable challenge in atomistic simulations. We develop and optimize methodology in the realm of path-integral-based methods (path-integral molecular dynamics, ring polymer instanton), that can give access to reaction rates and vibrational spectra, including nuclear quantum effects, and is fully applicable to high-dimensional anharmonic systems. With these methods,
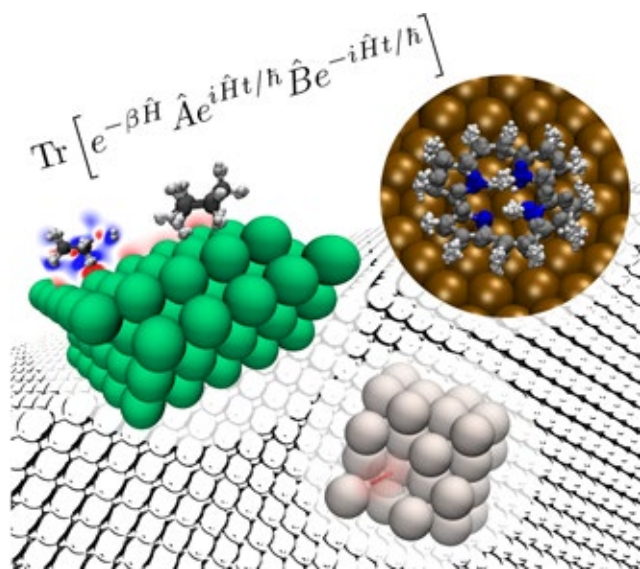

Fig. 1: Representative systems in which we investigate nuclear quantum effects (static and time dependent) – organic–metal interfaces, and hydrogen–based reactions.
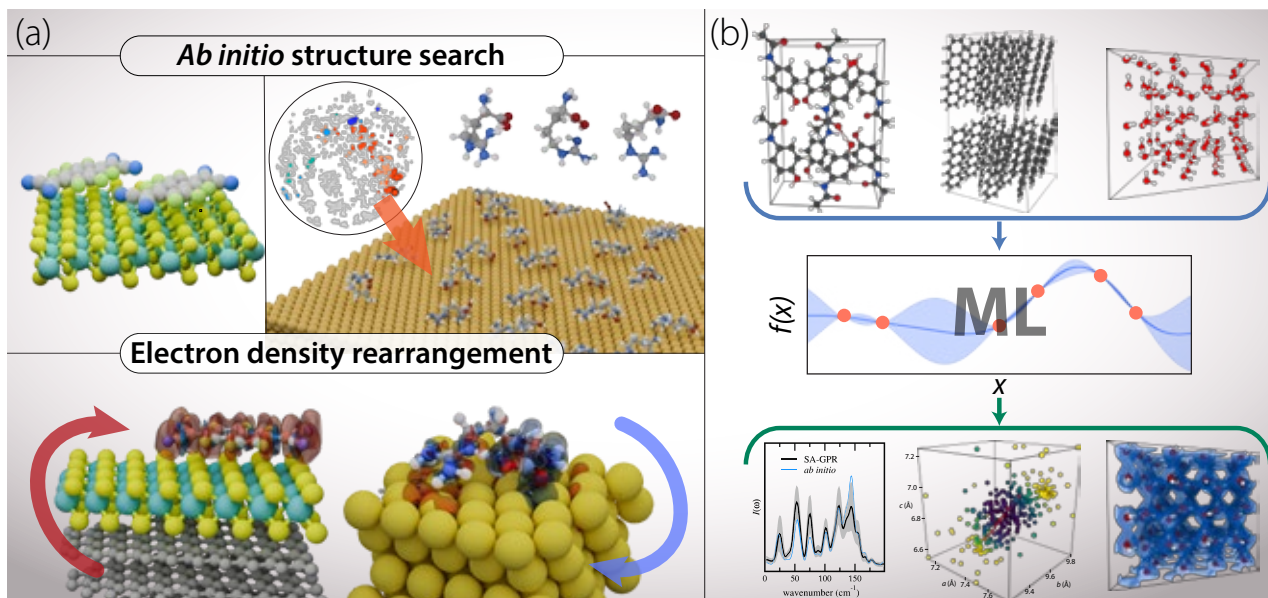
Fig. 2: (a) Structure–property relationship of flexible and rigid molecules at diverse interfaces (metals and semiconductors): a good sampling of the conformational space leads to accurate prediction of electronic properties; (b) Different flavors of super–vised machine learning methods used to speed up the simulation of crystals, allowing the prediction of vibrational spectra, free energies, and electron densities.

we were able to resolve isotope effects in work-function changes in metal-organic interfaces [1] and discover that surface fluctuations can enhance the intramolecular hydrogen-tunneling rate in molecules adsorbed on metallic surfaces by orders of magnitude [2].

ADVANCING THE SIMULATION OF VIBRATIONAL SPEC-TROSCOPY TECHNIQUES *With: Nathaniel Raimbault, Yair Litman and Alaa Akkoush /* Vibrational spectroscopic signals give access to detailed structural information of molecules and solids, and can probe time scales on the order of femto- to picoseconds. We develop simulation techniques that are able to account for anharmonic effects, nuclear quantum effects and spatial localization in IR, Raman and sum frequency generation signals. This has required developments on both the electron dynamics side (response properties to electric field perturbations from density-functional perturbation theory) and the nuclear dynamics side. With this methodology, we could account for anharmonic effects on the polymorph-sensitive low-frequency Raman spectra of molecular crystals [3] and we are now developing spatially-localized techniques, as well as interface-sensitive techniques.

HYBRID ORGANIC-INORGANIC SYSTEMS AT FINITE TEM-PERATURE *With: Haiyuan Wang and Dmitrii Maksimov /* A major challenge in describing organic-inorganic interfaces is predicting and controlling the structure-function relationship. The weakly-bonded organic components show flexibility and polymorphism, that are exacerbated by the typical temperatures at which these systems operate. With efficient structure-search techniques that are developed in the group and the possibility to perform large-scale electronic structure calculations with accurate methodology, we have navigated the structure space and created a database for adsorbed amino-acids [4]. In addition, we have studied the temperature dependence of charge transfer in hybrid interfaces of organic and inorganic semiconductors [5].

MACHINE LEARNING ELECTRONIC STRUCTURE PROPERTIES
With: Marcin Krynski, Nathaniel Raimbault and Alan Lewis / The increased cost of accounting for nuclear fluctuations and dynamics with *ab initio* potentials calls for the development of techniques that can decrease this cost without sacrificing the accuracy of the simulations. We devised a machine-learning model based on a symmetry-adapted Gaussian process regression algorithm (SA-GPR) that utilizes models trained for isolated molecules to learn static dielectric responses and thus predict anharmonic Raman spectra of organic crystals [6]. We are employing the same technique to predict the electronic density itself for a wide variety of solids, as well as different flavors of GPR to predict finite-temperature free energies and lattice expansion of organic crystals [7].

PUBLICATIONS
1. *K. Fidanyan, I. Hamada, M. Rossi, Adv. Theory Simul. 4, 2000241 (2021)*
2. *Y. Litman, M. Rossi, Phys. Rev. Lett. 125, 216001 (2020)*
3. *N. Raimbault, V. Athavale, M. Rossi, Phys. Rev. Mater. 3, 053605 (2019)*
4. *D. Maksimov, C. Baldauf, M. Rossi, Int. J. Quantum Chem. 121, e26369 (2020)*
5. *S. Park, H. Wang, et al, Adv. Mater. 2008677 (2021)*
6. *N. Raimbault, A. Grisafi, M. Ceriotti, M. Rossi, New J. Phys. 21, 105001 (2019)*
7. *M. Krynski, M. Rossi, arXiv:2106.08612 (2021)*

mpsd

Max Planck Institute for the Structure and Dynamics of Matter / Independent Lise Meitner Group The Lise Meitner group of Dr. Mariana Rossi started in Jan. 2020 at the MPI for the Structure and Dynamics of Matter in Hamburg. Previously, from 2016 to 2019, Dr. Rossi led an Otto Hahn group at the Fritz Haber Institute of the Max Planck Society in Berlin. Throughout these years, much of the research in the group was only possible due to the existence of the facilities at the MPCDF, which allowed the performance of all the research showcased in this article.

Computational materials design: complexity and real operating environments

Modern engineering materials have reached a level of structural and chemical complexity that render conventional brute-force trial-and-error approaches inefficient. Combining quantum mechanical approaches with thermodynamic and engineering concepts to run calculations at high-performance computing facilities provides powerful tools to explore and discover novel materials that sustain extreme technical environments.

AUTHORS *Mira Todorova, Christoph Freysoldt, Tilmann Hickel, and Jörg Neugebauer*

To address the grand environmental challenges faced by our society requires inexpensive, ecologically green materials with a long lifetime that can also survive harsh operational conditions. Historically, materials with targeted properties were discovered and improved by trial and error. Modern engineering materials, however, have evolved from simple materials dominated by a single chemical component to nano-composites and compositionally complex alloys consisting of multiple principal components. To design

this new generation of materials requires activating and controlling dynamic deformation mechanisms down to the atomistic scale. As a consequence, the increased structural and thermodynamic complexity renders conventional material design strategies less efficient.

Fully parameter-free quantum mechanical approaches provide perfect tools to replace the conventional ones, but face serious challenges to sample high-dimensional chemical and structural configuration spaces systematically. Linking advanced statistical sampling approaches with our python based workflow framework pyiron [1] allows us to combine first principles calculations with big data analytics in a highly automated way and to obtain accurate *ab initio* descriptors. The flexibility and power of these approaches will be highlighted in the following, discussing two prototype examples: First, the discovery of a new alloy that combines lightweight with excellent mechanical properties by a high-throughput study scanning over thousands of possible materials compositions. Second, understanding and predicting the behavior of materials under the extreme environments realized in an electrochemical cell.

MATERIALS DISCOVERY BY HPC-BASED HIGH-PERFOR-MANCE SCREENING A recurrent task in materials design is to explore whether existing limitations can be overcome by adding or substituting alloying elements [2]. For example, magnesium alloys
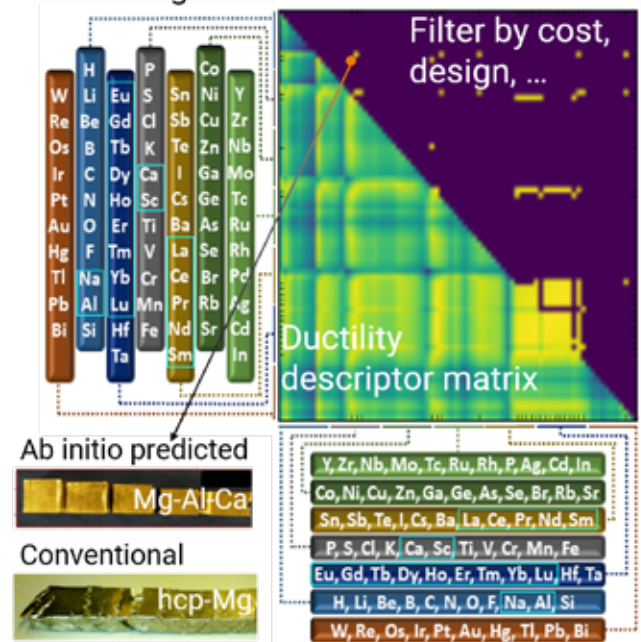
Fig. 1: Computed values of the Yttrium–similarity index (YSI, see text) for solute pairs in a Mg matrix visualized in the form of a symmetric matrix with yellow (blue) indicating high (low) similarity. Only solute pairs with YSI > 0.95 (high ductility) are shown in the upper triangle. The predicted optimum Mg–Al–Ca alloy composition has been synthesized experimentally and clearly overcomes the brittle behavior of conventional Mg alloys (bottom left picture).
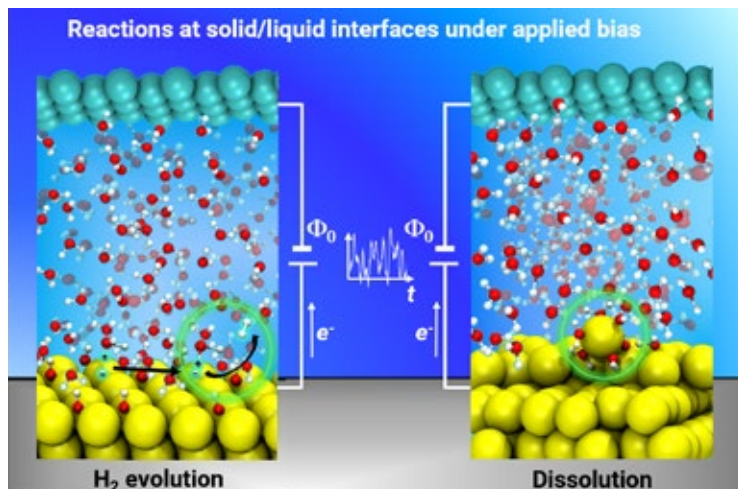
Fig. 2: Reactions at solid/liquid interfaces under realistic conditions of applied bias. The voltage is applied between two electrodes (yellow and cyan balls), which sandwich the water electrolyte (oxygen: red, hydrogen: white). Calculated reactions and reaction paths occurring at the interface are highlighted.

mechanism but also to open the way to search for alloy compositions that reduce or even eliminate this mechanism. Furthermore, they helped to establish a realistic picture of the dissolution of a metal surface (Fig. 2, right) and were successfully used to highlight the role of water co-adsorption on the formation of the electrode potential [6] – a key quantity in electrochemistry. The two examples given here highlight the opportunities in computational materials design that become available by recent progress in physical concepts, algorithms, and compute power.

———

PUBLICATIONS

1. *J. Janssen, S. Surendralal, Y. Lysogorskiy, M. Todorova, T. Hickel, R. Drautz, J. Neugebauer, Pyiron: An integrated development environment for computational materials science, Comput. Mat. Sci. 163, 24 (2019)*
2. *X. Zhang, H. Wang, T. Hickel, J. Rogal, Y. Li, J. Neugebauer, Mechanism of collective interstitial ordering in Fe–C alloys, Nature Materials 19, 849–854 (2020)*
3. *S. Sandloebes, M. Friak, S. Korte-Kerzel, Z. Pei, J. Neugebauer, D. Raabe, A rare-earth free magnesium alloy with improved intrinsic ductility, Sci. Rep. 7, 10458 (2017)*
4. *S. Surendralal, M. Todorova, M.W. Finnis, J. Neugebauer, First-principles approach to model electrochemical reactions: Understanding the fundamental mechanisms behind Mg corrosion, Phys. Rev. Lett. 120, 246801 (2018)*
5. *F. Deißenbeck, C. Freysoldt, M. Todorova, J. Neugebauer, S. Wippermann, Dielectric Properties of Nanoconfined Water: A Canonical Thermopotentiostat Approach, Phys. Rev. Lett. 126, 136803 (2021)*
6. *S. Surendralal, M. Todorova, J. Neugebauer, Impact of water co-adsorption on the electrode potential of H-Pt(111)-liquid water interfaces, Phys. Rev. Lett. 126, 166802 (2021)*

are technologically highly attractive due to their light weight. However, their high brittleness presently prevents broad application. Alloying Yttrium had been known to eliminate brittleness and to make Mg-alloys ductile, but the high price of this rare earth element made this approach economically unfeasible. After developing a suitable descriptor to predict the mechanical properties from *ab initio* calculations (the "Yttrium similarity index" [3]), a high-throughput study scanning over thousands of possible ternary Mg alloys could be performed (Fig. 1). Having this information for each alloy composition allowed us to employ various filters to identify compositions that combine best mechanical performance, economic costs, absence of toxic elements etc. This combinatorial approach revealed a new single-phase composition where Mg is alloyed with low concentrations of Al and Ca [3]. The new alloy contains only inexpensive and non-toxic elements making it very appealing for industrial applications. It also turned out to be scientifically of high interest and formed the basis to successfully establish a new DFG-funded collaborative research centre (CRC) with the RWTH Aachen (SFB 1394).

MATERIALS UNDER EXTREME CONDITIONS Materials have to show the targeted properties not only under idealized lab conditions but also under harsh operating conditions. An example are electrochemical conditions, which are of high interest in electrocatalysis, but also the origin of materials degradation, such as wet corrosion. The relevant mechanisms and chemical reactions take place at the interface between material and electrolytes such as water, and this makes their identification and understanding challenging and often impossible for experiment. New approaches developed at the Max-Planck-Institut für Eisenforschung provide realistic and computationally efficient algorithms to realize an electrochemical cell [4] and a thermodynamically consistent potentiostat [5]. Figure 2 shows examples of how these new approaches in connection with high-performance computing (HPC) enabled a new level of understanding of corrosion mechanisms. For instance, it became possible to solve the 150-year old mystery why magnesium evolves H_2 gas even under anodic conditions (Fig. 2, left), whereas all other metals would rather evolve oxygen, get oxidized, or consume H_2. Applying the new techniques allowed us to not only to identify and explain the underlying

MAX-PLANCK-INSTITUT
FÜR EISENFORSCHUNG GmbH

Accelerating molecular materials discovery with machine–learning

Computational molecular design is posed to accelerate the discovery of new organic materials. A brute-force screening approach is ineffective to this end, however, due to the sheer size of chemical space and the non-trivial relation between device performance and molecular properties. We tackle this challenge by combining physics and machine learning.

AUTHORS *Simon Wengert, Christian Kunkel, Johannes T. Margraf, and Karsten Reuter*

RESEARCH OBJECTIVES Organic semiconductors (OSC) enable important future technologies such as portable solar cells or rollable displays. For such applications, improved functional molecules – which make up these materials – need to be discovered. Computational screening has played a pivotal role in accelerating this process, as it avoids the laborious experimental synthesis and characterization for all but the most promising candidates. Even for computational screening, the number of potentially interesting small organic molecules is daunting, however (estimated to be on the order of 10^{33}). This overwhelming number of possibilities means that a brute-force screening approach cannot hope to truly explore the large diversity of chemical space. Indeed, most of the possible molecules are not suitable as OSC, making a complete screening quite wasteful, even if it were possible. One is essentially looking for the proverbial needle in a haystack.

Beyond this combinatorial challenge (which plagues all molecular design tasks) OSC discovery is additionally hampered by the fact that the performance of a molecule in a device cannot reliably be predicted from studying isolated molecules. Instead, charge transport crucially also depends on the arrangement of the molecules in the condensed phase (e.g., a molecular crystal). Predicting these condensed phase arrangements requires large-scale calculations and extensive sampling. This further emphasizes the need to judiciously select promising candidates, for which these costly simulations are carried out.

The goal of this project is therefore two-fold: On the one hand, we aim to develop methods for the active discovery of organic semiconductors, which avoid brute-force screening and reliably select promising candidates. On the other hand, we aim to accelerate the accurate prediction of molecular crystal structures. In both cases, this is achieved by combining physics-based simulations with machine learning (ML).
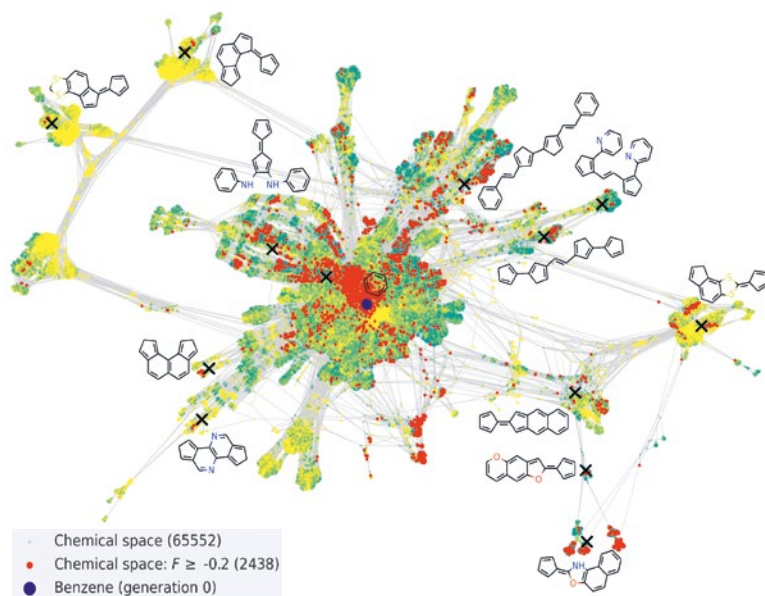
Fig. 1: Visualization of the chemical space explored with active discovery.

COMPUTATIONAL APPROACH Our approach is characterized by the combination of different levels of theory. These include density-functional theory (DFT) calculations using the full-potential FHI-aims code, wavefunction theory calculations with ORCA and semi-empirical density-functional tight-binding (DFTB) calculations using DFTB+. Machine learning models are developed using the QUIP code and custom in-house software.
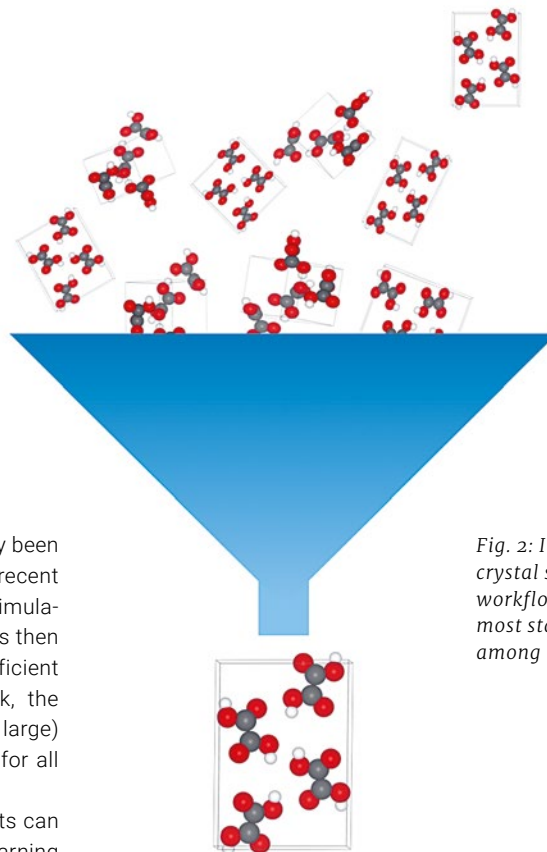
PROJECT DESCRIPTION Machine learning has increasingly been used to accelerate computational screening approaches in recent years. The basic idea is to run costly quantum mechanical simulations only for a fraction of the database of interest. This data is then used to train a ML model, which provides computationally efficient predictions for the remaining candidates. For this to work, the search space of interest must be somehow limited (even if large) and the resulting ML model must have reasonable accuracy for all remaining systems in the database.

In our recent work, we have shown that these requirements can be circumvented by so-called active learning [1]. Instead of learning from a predefined database, the ML algorithm iteratively decides for which candidates quantum mechanical calculations are performed. In this manner, a database of diverse and promising candidates is assembled and the ML algorithm is continually improved. Importantly, the computational resources are spent efficiently, so that most quantum mechanical calculations are performed for high-performing molecules. This also means that the algorithm can be applied to virtually unlimited search spaces, since the complete enumeration of candidates is not necessary.

In a parallel project, we aimed to accelerate the prediction of molecular crystal structures [2]. Indeed, the capability to reliably predict the structure of molecular crystals is considered one of the holy grails of molecular modeling, beyond the OSC context discussed herein. Applications for such crystal structure prediction (CSP) methods also include, for example, the design of pharmaceuticals with improved dissolution properties (and thus bioavailability). The goal of a typical CSP algorithm is in principle straightforward, namely to find the most stable periodic arrangement of molecules. In practice, the competition of different intermolecular forces (e.g., dispersion, dipolar interactions and hydrogen bonding) often leads to the coexistence of multiple similarly stable crystal structures – so-called polymorphs – each exhibiting different physical properties, however. Finding all relevant polymorphs typically requires sampling an enormous configurational space. Meanwhile, distinguishing between energetically similar structures can be challenging even for high-level quantum mechanical methods. A methodology with a large accuracy/cost ratio is therefore paramount for effective CSP.

To achieve this, we combined state-of-the-art ML potentials (trained on high-quality quantum mechanical reference data) with a low-cost tight-binding baseline model. The use of this baseline leads to a highly data-efficient method, since not all physical interactions are learned from scratch. A further beneficial characteristic of our approach is that no reference data from molecular crystals is used. Instead, the ML model is trained on data from small molecular clusters. This even allows the use of high-level target methods beyond semi-local DFT. The approach is broadly applicable to different molecular materials and, importantly, also allows for reliable structure optimizations. All this is achieved with a computational effort that is orders of magnitude smaller than the high-level target methods, even taking training costs into account.

Taken together, these developments open the door towards highly accurate molecular materials discovery workflows, which actively select promising candidates and are able to incorporate condensed phase information.

PUBLICATIONS

1. C. Kunkel et al. "Active discovery of organic semiconductors", *Nature Commun. 12, 2422 (2021)*
2. S. Wengert et al. "Data–efficient machine learning for molecular crystal structure prediction", *Chem. Sci. 12, 4536 (2021)*

FRITZ-HABER-INSTITUT
MAX-PLANCK-GESELLSCHAFT

Fritz Haber Institute of the Max Planck Society / Department Theory
The research in the Theory Department focuses on a quantitative modeling of materials properties and functions, and in particular on processes in working catalysts and energy conversion devices. For this we advance and employ predictive-quality multiscale models, advanced data science techniques and machine learning, thereby straddling the frontiers of physics, chemistry, computing sciences, as well as materials science and engineering.

Identifying novel thermal insulators in material space

Heat transport plays a pivotal role in a multitude of scientific and industrial applications, e.g., energy conversion, catalysis, and combustion. However, surprisingly little is known about the trends governing thermal transport across material space and the underlying microscopic mechanisms. At the NOMAD Laboratory, novel first-principles methods are developed to shed light on these questions.

AUTHORS *Christian Carbogno and Matthias Scheffler*

Macroscopic heat transport, as measured by the temperature- and pressure-dependent thermal conductivity $\kappa(T,p)$, is an ubiquitous phenomenon in condensed matter that plays a crucial role in a multitude of scientific and industrial applications, e.g., energy conversion, catalysis, and combustion. In particular, materials with better insulating properties are urgently needed for enabling emerging technologies with important environmental impact. For instance, this includes so-called *thermal barrier coatings*, which are used in gas and airplane turbines to thermally insulate the metal alloys from the extreme temperatures generated during combustion. These strongly insulating coatings do not only increase the longevity of turbines, but also allow for an increase in operational temperature, thus resulting in a higher fuel efficiency. Furthermore, better thermal insulators are also needed for *thermoelectric converters*, which allow to recycle otherwise unused, wasted heat – which, at present, makes up over 40% of the world's raw energy consumption –

into useful electric power. Currently existing thermoelectric materials are not sufficiently efficient to enable an economically viable deployment; designing materials with higher efficiency, which scales inversely with $\kappa(T,p)$, is one of the most promising strategies to overcome this hurdle.

Despite these important challenges, only very little is known to date about thermal transport: For instance, the database *Springer Materials* (https://materials.springer.com) lists less than 100 entries for the thermal conductivity of single crystals. At the NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society, we develop, optimize, and apply computational methodologies to shed light on the mechanisms that drive thermal transport. This allows us to identify novel and better thermal insulators to enable the aforementioned applications. For this purpose, we employ a hierarchical approach: First, materials with the potential for ultra-low thermal conductivities are identified among the billions of possible compounds. Second, high-precision first-principles techniques are used to accurately predict the thermal conductivity of these materials and to understand the underlying microscopic mechanisms.

ACCURATE THERMAL CONDUCTIVITIES FROM FIRST PRINCIPLES An accurate computation of thermal conductivities in semiconductors and insulators involves solving several fundamental challenges that arise from the fact that heat transport is determined from subtle details of the nuclear motion. Since the occurrence of these effects spans several orders of magnitude in time and length scales, obtaining accurate thermal conductivities entails the development of first-principles methods that both accurately capture the interactions at the microscopic level and allow bridging the necessary time and length scales, so that it becomes possible to extract information on the relevant macroscopic material properties. The *ab initio* Green-Kubo method developed at the NOMAD laboratory is the first approach that succeeds in this regard. In particular, no approximation is employed for describing the nuclear motion and the associated heat flux. This is of particular importance for the
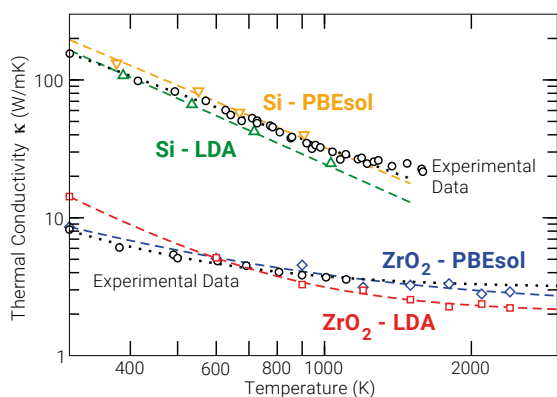

Fig. 1: Thermal conductivity $\kappa(T,p)$ as a function of temperature computed using density–functional theory and different exchange–correlation functionals (LDA, PBEsol) with the ab initio *Green–Kubo approach. Black symbols denote experimental results compiled from literature.*

accurate description of thermal insulators, the dynamics of which deviates considerably from the parabolic harmonic models typically used for this purpose in solid-state physics. Fully accounting for these anharmonic effects, which limit heat transport in solids, is thus key for an accurate assessment of materials with low and ultra-low $\kappa(T,p)$. As an example, Fig. 1 shows the thermal conductivities computed with the *ab initio* Green-Kubo method for Silicon and Zirconia, two materials featuring particularly weak and strong anharmonic effects, respectively. In both cases, the developed methodology achieves excellent agreement with experiment – despite the fact that the time and length scales of heat transport differ considerably in Silicon and Zirconia, as evident by the fact that the values of $\kappa(T,p)$ span two to orders of magnitude. Besides reproducing and predicting thermal conductivities, the *ab initio* Green-Kubo approach gives full access to the microscopic details of the nuclear motion and thus yields important insights on the mechanisms that drive the nuclear dynamics and heat transport in different materials. In this context, it is important to note that the precision and the wealth of information of the *ab initio* Green-Kubo method come at a considerable computational cost: Each computation of $\kappa(T,p)$ involves performing several ten-thousand electronic-structure calculations. Accordingly, such simulations are currently only possible on massively parallel high-performance computing (HPC) installations.

HIGH-THROUGHPUT MATERIAL SPACE EXPLORATION Although the described *ab initio* Green-Kubo approach allows to compute heat transport in thermal insulators with unprecedented accuracy, it does not lend itself to an extended exploration of material space, since the computational cost to investigate all possible candidates would be prohibitive. To overcome this hurdle, the NOMAD Laboratory has developed metrics and workflows for rapidly and reliably discerning good thermal conductors from potential thermal insulators. In a nutshell, each material is classified by a single number, the σ^A metric, which quantifies how influential anharmonic effects are on its nuclear motion. As shown in Figure 2, this σ^A metric then allows estimating the thermal conductivity of a material. Compared to full *ab initio* Green-Kubo calculations, these estimates come at a negligible computational cost, since they only require a dozen electronic-structure calculations per material. Accordingly, this enables a *high-throughput* exploration of material space, covering thousands of different compounds. Being able to rapidly scan over so many materials is key for identifying those compounds with optimal properties – like finding a needle in a haystack – and thus again requires massively parallel HPC resources. From this huge pool of investigated compounds, the employed hierarchical approach allows to single out the few, most promising candidate materials, which are then further investigated with the *ab initio* Green-Kubo method, as sketched in Figure 2.

PUBLICATIONS

1. *C. Carbogno, R. Ramprasad, and M. Scheffler, Ab Initio Green–Kubo Approach for the Thermal Conductivity of Solids, Phys. Rev. Lett. 118, 175901 (2017)*
2. *F. Knoop, T. A. R. Purcell, M. Scheffler, and C. Carbogno, Anharmonicity measure for materials, Phys. Rev. Mater. 4, 083809 (2020)*
3. *F. Knoop, T. A. R. Purcell, M. Scheffler, and C. Carbogno, FHI–vibes: Ab Initio Vibrational Simulations, J. Open Source Softw. 5, 2671 (2020)*

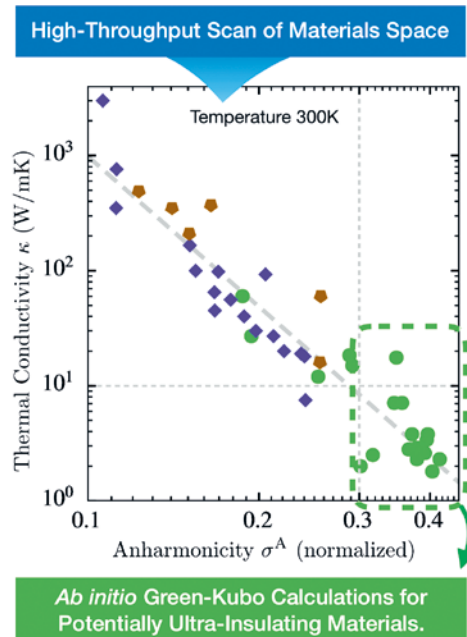*Fig. 2: Sketch of the employed hierarchical workflow. After determining the anharmonicity metric σ^A, good, but uninteresting thermal conductors can be discarded. As an example, we here show a representative subset of experimentally measured thermal conductivities κ at 300K as a function of the anharmonicity metric σ^A. Those potentially ultra–insulating compounds with high values of σ^A are then investigated using the ab initio Green–Kubo approach, to accurately determine $\kappa(T,p)$.*

Fritz Haber Institute of the Max Planck Society / The NOMAD (Novel Materials Discovery) Laboratory The NOMAD Laboratory at the FHI addresses fundamental aspects of the chemical and physical properties of surfaces, interfaces, clusters, and nanostructures. Specifically, the research concerns heterogeneous catalysis, thermal and electrical transport, thermoelectric materials, crystal growth, and defects in semiconductors. These studies combine quantum mechanical, *ab initio* calculations of the electronic structure and molecular dynamics with methods from thermodynamics and statistical mechanics. To address these questions, novel, more accurate, and more efficient first-principles methodologies are developed and implemented in massively parallel scientific software. Along these lines, the group is also a leader in developments and applications of artificial-intelligence methods for materials science.

The group initiated and coordinates a web-based platform for FAIR sharing of materials science data and a Center of Excellence for exascale software developments in the field of computational materials science. Together with the MPCDF it maintains the world's largest *Repository* for computational materials science, as well as a *Material Encyclopedia* and a *Big-Data Artificial Intelligence Toolkit*. It is also co-leading the consortium FAIRmat (FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids) which extends the NOMAD concepts to materials synthesis, experiments, and theory.

Materials for turning greenhouse gases into useful chemicals and fuels: a route identified by high–throughput calculations and artificial intelligence

Routes already exist to convert CO_2 into useful chemicals or fuels, instead of dumping it into the atmosphere. The key is catalysis, but the known catalyst materials are too inefficient for making this conversion practicable. To find better catalysts, an important step is the identification of materials that "activate" CO_2, i.e., make the molecule ready for doing a chemical reaction. Using high-throughput calculations and subgroup discovery, scientists at the NOMAD Laboratory identified the materials genes and rules that classify potentially good catalyst materials for the desired purpose.

AUTHORS *Aliaksei Mazheika[2], Sergey Levchenko[3], Luca Ghiringhelli, and Matthias Scheffler*

Because of its chemical stability, CO_2 is presently one of the critical mankind-created greenhouses gases. However, at some point CO_2 may well become a raw material for creating fuels and valuable chemicals. Figure 1 displays a reaction network that starts from CO_2 showing the catalytic, chemical conversion towards methane (viable for combustion engines and heating) and other useful chemicals. This is all possible, already today, but the process is inefficient. We need better catalysts.

The NOMAD Laboratory at the Fritz Haber Institute developed and advanced artificial-intelligence (AI) methods [2] that enable the identification of *basic materials parameters* that correlate with materials properties and functions of interest (here the activation of CO_2).

These parameters are also called *materials genes* as they correlate with different mechanisms that trigger, but maybe just facilitate or hinder the different processes, which play together – very much as genes do in biology. In the coordinate system of these genes, regions are identified where good catalytic materials can be found. Specifically, this CO_2 study employed the AI method called "subgroup discovery" [3], and focused on the wide class of metal oxides. Catalysis happens at surfaces. Thus, the study considered various surfaces of these materials. Altogether, 141 different surfaces of 71 different materials, typically with different adsorption sites, were calculated with state-of-the-art density-functional theory and high-throughput methodology. The results were then used for training the AI.

[2] UniCat BASF Joint Lab, Berlin, Germany
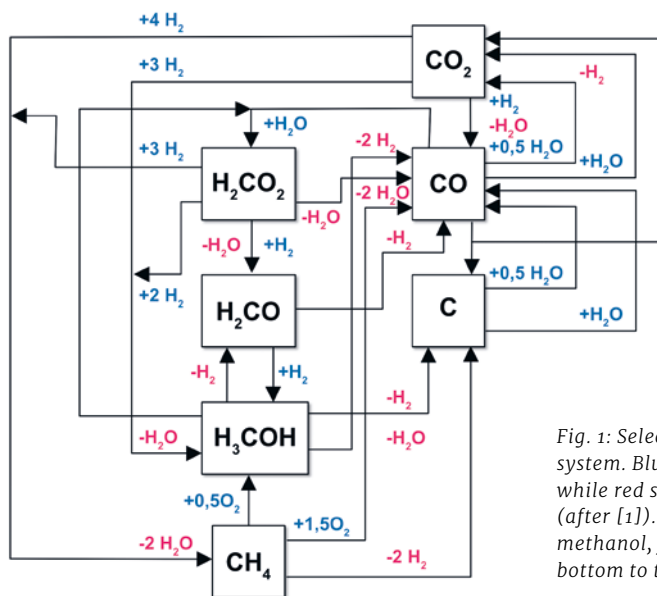*[3] now at Skolkovo Institute of Science and Technology, Moscow, Russia*

Fig. 1: Selected reactions in the C/H/O ternary system. Blue species are added to the reactants, while red species are liberated during the reactions (after [1]). The reaction products are methane, methanol, formaldehyde, and formic acid (from bottom to top).

In general terms, this study also represents a conceptual change of modeling heterogeneous catalysis (and other materials functions). In the past it was attempted to calculate the full catalytic process [1]. However, it became clear that the many aspects that rule heterogeneous catalysis, e.g., the dynamical restructuring of the surface under reaction conditions, are too intricate, and that a full theoretical description is not an expedient approach. Thus, a combined approach linking high-throughput calculations, AI, and experimental results appears more appropriate and was suggested in this study.

The idea of CO_2 activation is well established. In the gas phase, it is achieved by charging the molecule. This weakens the C-O bonds, and is reflected in a change of geometry, from linear to bent (see top of Fig. 2). This bent molecule will then happily do a chemical reaction. Analogously, for heterogeneous catalysis it was assumed that surfaces that bind CO_2 with a bent geometry may be good catalysts [4].

The new study took a more general perspective and considered five different scenarios that may be indicators of weakening the chemical bond between C and O in adsorbed CO_2. The O-C-O bond angle was one of the five, and we only mention one other indicator of bond weakening, namely the C-O bond length. More can be found in the respective publication [5].

Our AI method of choice, subgroup discovery, can find the coordinate systems where these subgroups of materials that promote a bond weakening live, i.e., it identifies the important descriptive parameters (or materials genes). For this study, we offered 46 different features, i.e., potential genes. These are basic physical parameters of the free atoms composing the material and the pristine surfaces. They don't carry explicit information about the adsorption or the catalytic processes. In the end, subgroup discovery will select the relevant genes. In simple words, essentially everything was offered that might be relevant.

Fig. 2: CO_2 activation in the gas phase (top) and the present model that is assumed relevant for heterogeneous catalysis (see text and [4]).
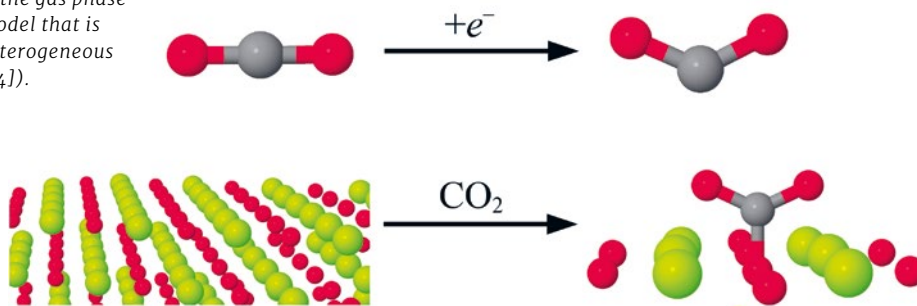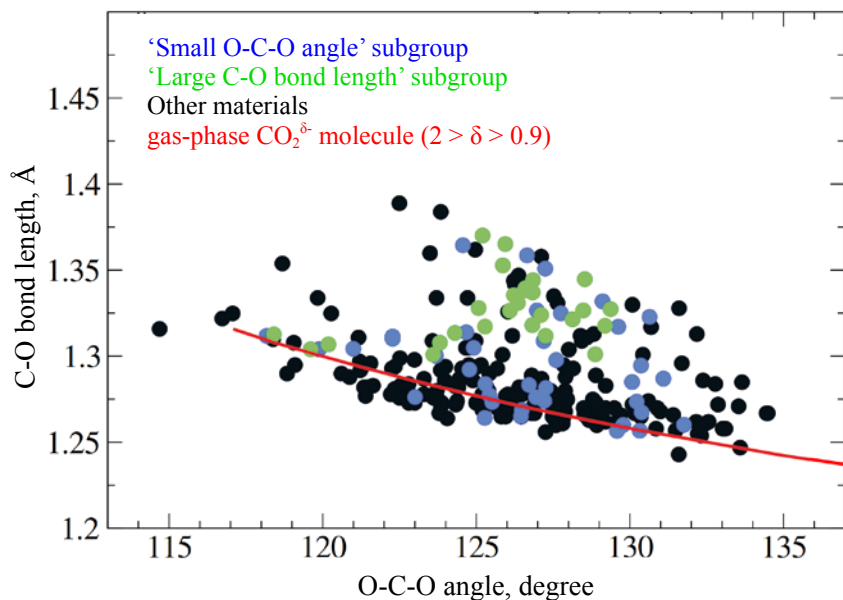
Fig. 3: Two subgroups (blue and green) identified by subgroup discovery. The relevant materials genes are quantities of the clean material, i.e., of the free atoms composing the materials and of the pristine surfaces. The red line shows the well-defined correlation between the C–O bond length and the O–C–O angle of a charged CO_2 molecule in the gas phase. Clearly, for adsorbed CO_2, this correlation is gone.

Figure 3 shows the results for all 255 calculated data points, displayed in terms of the O-C-O angle and the C-O bond length. The blue data points belong the "small O-C-O angle" subgroup and the green data points belong to the "large C-O-bond length" subgroup. In the coordinate system of this plot, the blue or the green dots are scattered; they do not look like groups. Obviously, the group character is not defined by the coordinates, but by the identified materials genes (see [5] for details).

As said above, evaluating the catalytic efficiency under realistic conditions numerically and reliably is hardly possible. Too many different processes are playing together. Thus, after identifying the two subgroups (blue and green), the study considered experimental information. It turns out that most known materials with good catalytic performance belong to these two groups. Thus, when searching for a good catalyst, one should look at the region in materials spaces that is defined by these two subgroups. This reduces the search space significantly, and improves the search process.

PUBLICATIONS

1. H.-J. Freund, G. Meijer, M. Scheffler, R. Schlögl, and M. Wolf, CO oxidation as a prototypical reaction for heterogeneous processes. Angew. Chem. Int. 50, 10064 (2011)
2. C. Draxl and M. Scheffler, Big-Data-Driven Materials Science and its FAIR Data Infrastructure. Plenary Chapter in Handbook of Materials Modeling (eds. S. Yip and W. Andreoni), Springer (2020)
3. B.R. Goldsmith, M. Boley, J. Vreeken, M. Scheffler, and L.M. Ghiringhelli, Uncovering structure-property relationships of materials by subgroup discovery. New J. Phys. 19, 013031 (2017)
4. H.-J. Freund, M.W. Roberts, Surface chemistry of carbon dioxide. Surface Science Reports, Volume 25, 225 (1996)
5. A. Mazheika, Y. Wang, R. Valero, F. Viñes, F. Illas, L. M. Ghiringhelli, S. V. Levchenko, M. Scheffler

NOMAD NOVEL-MATERIALS DISCOVERY LABORATORY
at the Fritz Haber Institute of the Max Planck Society

Fritz Haber Institute of the Max Planck Society / The NOMAD (Novel Materials Discovery) Laboratory The NOMAD Laboratory at the FHI addresses fundamental aspects of the chemical and physical properties of surfaces, interfaces, clusters, and nanostructures. Specifically, the research concerns heterogeneous catalysis, thermal and electrical transport, thermoelectric materials, crystal growth, and defects in semiconductors. These studies combine quantum mechanical, *ab initio* calculations of the electronic structure and molecular dynamics with methods from thermodynamics and statistical mechanics. To address these questions, novel, more accurate, and more efficient first-principles methodologies are developed and implemented in massively parallel scientific software. Along these lines, the group is also a leader in developments and applications of artificial-intelligence methods for materials science.

The group initiated and coordinates a web-based platform for FAIR sharing of materials science data and a Center of Excellence for exa-scale software developments in the field of computational materials science. Together with the MPCDF it maintains the world's largest *Repository* for computational materials science, as well as a *Material Encyclopedia* and a *Big-Data Artificial Intelligence Toolkit*. It is also co-leading the consortium FAIRmat (FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids) which extends the NOMAD concepts to materials synthesis, experiments, and theory.

Preparing electronic–structure theory for the exascale

With the upcoming deployment of (pre-)exascale high-performance computers, the computational materials science community is facing several challenges. Nonetheless, harvesting this huge amount of computational power enables urgently needed opportunities. Together with the MPCDF and several European partners, the NOMAD Laboratory of the FHI has initiated, installed, and is leading the NOMAD Center of Excellence (CoE) to enable the next generation of electronic-structure calculations. Key items of the NOMAD CoE are extreme-scale data and exascale computing.

AUTHORS *Christian Carbogno, Volker Blum[4], Sebastian Kokott, Hermann Lederer[5], Andreas Marek[5], Florian Merz[6], Markus Rampp[5], Xinguo Ren[7], and Matthias Scheffler*

Over the past decades, electronic-structure theory has evolved into an indispensable pillar in materials science. From its early days as a tool to reproduce and explain experimental results, it is now increasingly employed for predictions of yet-unmeasured properties of compounds that might not even have been synthesized so far. Studies are performed in *high-throughput* workflows calculating several hundred or even thousand mostly yet-unknown materials. These developments were driven by both the exponential increase in computational power and by the increased availability of mature, stable electronic-structure theory software.

Despite the impressive success and scientific impact, it is important to realize that the utmost majority of electronic-structure calculations focus on idealized and static crystals, but do not describe materials under realistic conditions. Accounting for these effects in a reliable fashion is possible, e.g., by performing extensive *ab initio* molecular dynamics simulations for extended systems, but obviously involves much more computationally costly calculations. Similarly, today's *high-throughput* studies are almost exclusively based on semi-local density-functional theory, a numerically very efficient, but also 30 years old methodology with distinct flaws. Accordingly, such calculations involve several approximations that critically limit their predictive power, for instance in the description of localized electronic states, e.g., in transition metal oxides; or in the case of defects, impurities or polarons; in the assessment of chemical reaction barriers; or in the prediction of (opto-)electronic properties. To the uttermost extent, these shortcomings are overcome in electronic-structure methods beyond semi-local density-functional theory. However, such advanced techniques come with a considerable increase in computational cost, which, so far, has limited a widespread usage. At the NOMAD Laboratory of the Fritz Haber Institute of the Max Planck Society, we develop, implement, and advance these techniques to make them available for the whole materials science community. In close collaboration with programming and application experts at the MPCDF and the NOMAD CoE we design and optimize the software to be exascale-ready, to be able to harvest the computational power in next-generation supercomputers, enabling studies that are urgent for important application in the environment, energy, mobility, IT, or health sectors but unfeasible today. This implies achieving higher single-node performances, e.g., by designing novel, more granular algorithms that allow to offload computationally intensive tasks to accelerators such as GPUs. In turn, this requires novel approaches for load balancing to minimize computational and communication overhead, hence achieving better performance and scalability on massively parallel architectures featuring 100,000 nodes and more, each equipped with hundreds of CPU and GPU cores. Together, these efforts aim not only at reducing the time-to-solution, but also the cost-to-solution, and thus lay the foundation for performing more accurate and precise and realistic materials science calculations in the next decades.

MASSIVELY PARALLEL EIGENVALUE SOLVERS In large-scale, semi-local density-functional theory calculations, the solution of (generalized) complex eigenproblems is the task that dominates the

---

[4] *Duke University, Durham, USA*
[5] *Max Planck Computing and Data Facility, Garching, Germany*
[6] *Lenovo HPC Innovation Center, Stuttgart, Germany*
[7] *Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Science*
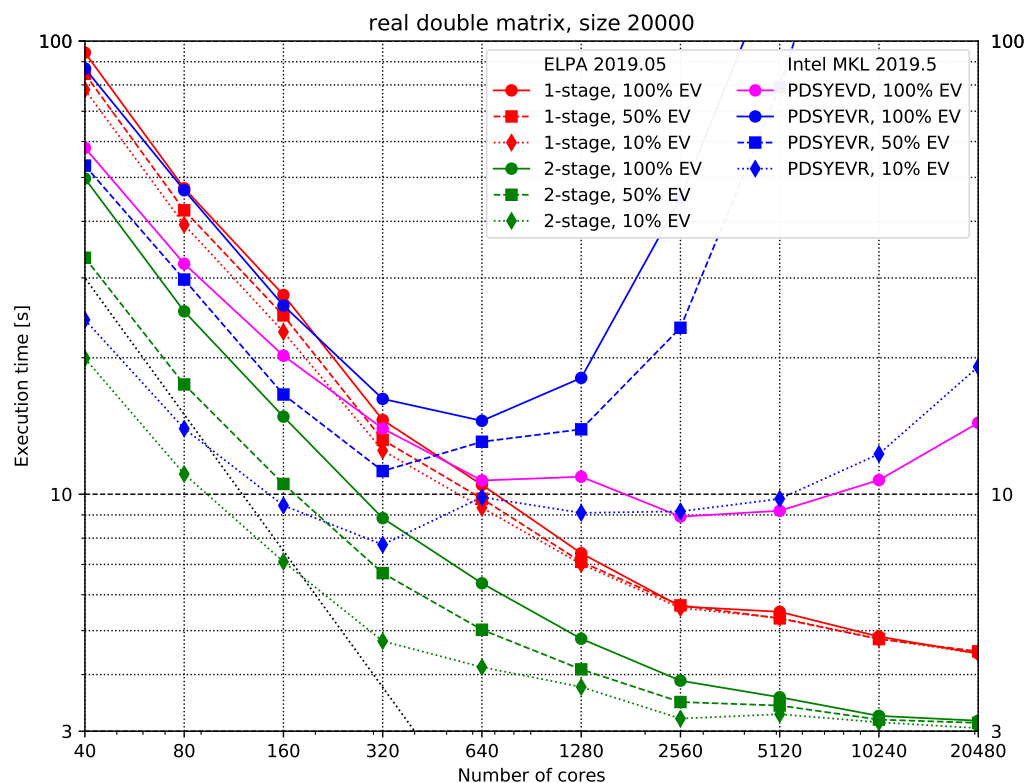
real double matrix, size 20000

Fig. 1: Performance and scaling comparison between ELPA and the Intel Math Kernel Library (MKL) for the diagonalization of a double–precision, real matrix of size n = 20,000. Percentage of computed eigenvectors and the respective employed algorithms are given in the legend.

computational cost. Since its cost scales as N^3 with the system size N, this step severely limits the complexity of the questions that can be investigated at all. To alleviate this problem, the MPCDF and the NOMAD Laboratory were involved in the collaborative development of the open-source, efficient, and highly scalable eigensolver library ELPA. To date, ELPA is superior to all competing commercial and open-source alternatives and "*outperforms ScaLAPACK on all architectures tested and all matrix sizes and concurrencies*", as recently demonstrated by NERSC and Cray. This is demonstrated in Figure 1, in which we compare the performance of ELPA with the corresponding ScaLAPACK routines in the Intel Math Kernel Library (MKL). As the plot shows, ELPA is not only faster under all circumstances, but it also exhibits improved scalability to a much higher number of cores. These advantages directly translate into immense computational savings in actual electronic-structure calculations and are thus the key to enable the most ambitious density-functional theory studies on current and future architectures. For this exact reason, ELPA is now the recommended eigensolver library in the majority of density-functional theory codes. This is possible since ELPA is freely available to the whole community as an open-source package (LGPLv3 licence) at https://elpa.mpcdf.mpg.de.

HYBRID FUNCTIONALS INCLUDING EXACT EXCHANGE One of the main sources of error in semi-local density-functional theory is the so-called *self-interaction*, i.e., the fact that an electron is not only subject to the electrostatic field generated by all other electrons, but that it also feels the field generated by itself. In response to this unphysical interaction, electrons tend to spread out in semi-local density-functional theory, which leads, in turn, to erroneous delocalized electronic states and to an underestimation of energetic barriers. One way to cure these artifacts is to go beyond the semi-local level of theory by describing non-local exchange interactions exactly. This, however, requires evaluating non-local, high-dimensional integrals. The cost of these operations competes with or even prevails over the cost of the eigensolvers, since it scales with N^4. For this reason, hybrid functionals are, so far, almost exclusively used in simulations of simple, small, and idealized model systems, but not for assessing complex and extended systems under realistic thermodynamic conditions. In this regard, the MPCDF and the NOMAD Laboratory have co-operated in recent years to improve the performance and scalability of the exact-exchange evaluation, whereby the so-called "resolution-of-identity" is employed to achieve a granular subdivision of computational tasks on different threads and is thus

key to enable scalability and parallelizability. By this means, systems with up to 2,000 atoms in the unit cell can now be routinely treated at this higher-level of theory. For the first time, this allows for a systematic investigation of complex, functional materials in a realistic fashion. For instance, this allows to address the role of atomic defects, impurities and vacancies — which are well known for their influence on electronic properties — with unprecedented accuracy.

Let us mention in passing that an alternative approach would be the coupled-cluster theory of quantum chemistry. The team is also working on implementing and advancing this "gold standard for molecules" for extended systems of materials science.

GREEN-FUNCTION-BASED MANY-BODY PERTURBATION THEORY *(GW)* Another shortcoming of semi-local density-functional theory is its inability to correctly handle electronic excitations. With that, many electronic properties are not accurately accessible, e.g., band structures, band gaps, electron addition or removal energies, and more. This severely hinders progress in the development of novel, better, and improved materials for, e.g., opto-electronic applications, for which these exact properties play a pivotal role. In this context, we have recently developed a *GW* implementation based on numeric-atomic orbitals. As in the case of hybrid functionals, a variant of the "resolution-of-identity" approximation is employed to reduce both the computational cost of evaluating and the memory overhead required to store the non-local interaction integrals. As an example, we compare experimentally measured electronic band gaps with band gaps computed using the non-selfconsistent G_0W_0 approximation on top of a semi-local (PBE) density-functional theory calculation for a set of benchmark materials in Figure 2. In contrast to standard density-functional theory calculations, which, on average, underestimate band gaps by a factor of two or more, the G_0W_0 yields band gaps that are in quantitative agreement with experiment. Accordingly, these developments thus lay the foundation for more accurate and predictive first-principles calculations of opto-electronic properties.

In the studies mentioned, worldwide leading success was achieved. We are well on the route to exascale performance, but more ideas and advancements are needed. And, unfortunately, the architecture(s) of these upcoming computers are still not clearly known.

PUBLICATIONS

1. *S. Levchenko, X. Ren, J. Wieferink, R. Johanni, P. Rinke, V. Blum, M. Scheffler, Comp. Phys. Comm. 192, 60 (2015)*
2. *P. Kus, A. Marek, S Köcher, H.H Kowalski, C. Carbogno, C. Scheurer, K. Reuter, M. Scheffler, H. Lederer, Parallel Computing 85, 167 (2019)*
3. *V. Wen–zhe Yu, J. Moussa, P. Kůs, A. Marek, P. Messmer, M. Yoon, H. Lederer, V. Blum, Comp. Phys. Comm. 262, 107808 (2021)*
4. *X. Ren, F. Merz, H. Jiang, Y. Yao, M. Rampp, H. Lederer, V. Blum, and M. Scheffler, Phys. Rev. Materials 5, 013807 (2021)*
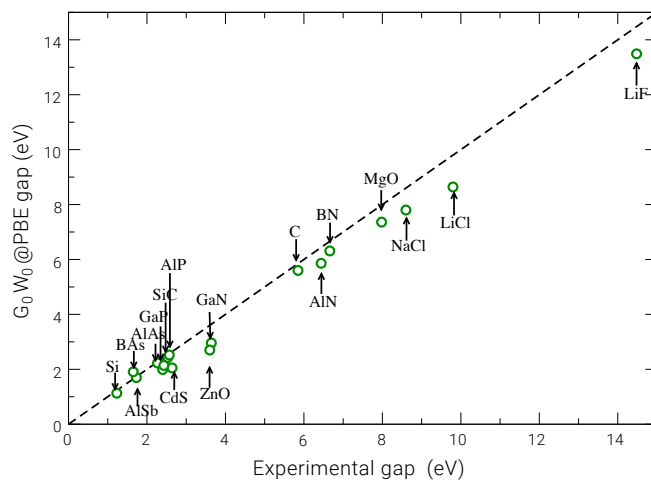

Fig. 2: Experimental band gaps versus band gaps computed with the G_0W_0 approximation on top of a semi–local density–functional theory calculation using the PBE exchange correlation functional for a set of benchmark materials.

Fritz Haber Institute of the Max Planck Society / The NOMAD (Novel Materials Discovery) Laboratory The NOMAD Laboratory at the FHI addresses fundamental aspects of the chemical and physical properties of surfaces, interfaces, clusters, and nanostructures. Specifically, the research concerns heterogeneous catalysis, thermal and electrical transport, thermoelectric materials, crystal growth, and defects in semiconductors. These studies combine quantum mechanical, *ab initio* calculations of the electronic structure and molecular dynamics with methods from thermodynamics and statistical mechanics. To address these questions, novel, more accurate, and more efficient first-principles methodologies are developed and implemented in massively parallel scientific software. Along these lines, the group is also a leader in developments and applications of artificial-intelligence methods for materials science.

The group initiated and coordinates a web-based platform for FAIR sharing of materials science data and a Center of Excellence for exascale software developments in the field of computational materials science. Together with the MPCDF it maintains the world's largest *Repository* for computational materials science, as well as a *Material Encyclopedia* and a *Big-Data Artificial Intelligence Toolkit*. It is also co-leading the consortium FAIRmat (FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids) which extends the NOMAD concepts to materials synthesis, experiments, and theory.

The NOMAD Laboratory

The enormous amounts of research data produced in condensed-matter physics, chemistry, and materials science every day represent a gold mine for science, engineering, and society. This gold mine is, however, of little value, if these data are not comprehensively characterized and made available. How can we refine this feedstock, i.e., turn data into knowledge and value? A FAIR data infrastructure is a must, and in this spirit, the Novel Materials Discovery (NOMAD) Laboratory has established a user-driven platform for sharing computational materials science data to enable their re-use and exploitation.

AUTHORS *Claudia Draxl, Markus Scheidgen, Thomas Zastrow[8], Raphael Ritz[8], Hermann Lederer[8], Stefan Heinzel[8], and Matthias Scheffler*

Recognizing that the sciences can be enhanced by data-centric[9] research has triggered both high-throughput studies (HTS) as well as artificial intelligence (AI) approaches in various fields. Obviously, this is also a vital topic in materials research. The big picture is to enable researchers in basic science and engineering to utilize materials data in order to identify improved, new, or even novel materials and, in turn, pave the way to better scientific understanding as well as novel products and technologies [1].

WHERE IS THE COMMUNITY ON THIS ROAD?

Since more than a decade, HTS computations and experiments create huge data collections all over the world. However, though such studies provide much information, they will never cover the practically infinite chemical and structural space of possible materials. Moreover, each of them aims at a specific effect or application, typically considering only a specific material class. Likewise, standard machine-learning techniques can typically interpolate such HTS data quite well but they do not have the predictive power to suggest something spectacularly new. Considering both limitations, two aspects are crucial for a real breakthrough: First, data from different sources must be brought together to widen the scope of data-centric approaches. Second, novel artificial-intelligence (AI) methods need to be developed that allow for thinking out of the box. For both these reasons, the NOMAD Laboratory [2] was created. Soon, it has turned into the biggest store of computational materials data worldwide (https://nomad-lab.eu).

In its present form, the NOMAD Laboratory processes, curates, and hosts computational materials science data, as obtained by all important materials science codes available today. Publish – Explore – Analyze: This is the essence of the NOMAD Laboratory and visualized in Figure 1, and the NOMAD services are briefly summarized below.

Fig. 1:
Landing page of the NOMAD Laboratory, which enables publishing, exploring, and analyzing computational materials data: https://nomad–lab.eu.

[8] *Max Planck Computing and Data Facility, Garching, Germany*

[9] *The term data-centric indicates a radical shift in the way information is handled and research is performed. It refers to extensive data collections, digital repositories, and new concepts and methods of data analytics. It also implies that the traditional purpose-oriented research is complemented by concepts for using data from other studies.*

NOMAD SERVICES The NOMAD *Repository* forms the basis of all NOMAD services. Users can upload their raw data as created by the employed computer code. No adjustments are necessary. Uploaded data are tagged by a persistent identifier (PID), and users can also request a digital object identifier (DOI) to make data citable. Parsers for currently 40 different software packages take care of the data conversion, thus the individual researchers are not concerned with data formats or units. The NOMAD *Archive*, containing the corresponding normalized data (unified formats and units) – at present more than 100 million calculations – represents the worldwide largest data collection of this field. To explore the wealth of all these data, the NOMAD *Encyclopedia* serves the purpose of in-depth presentation, characterization, and understanding of the materials of interest by providing knowledge of their various properties, like structural features, mechanical and thermal behavior, electronic and magnetic properties, the response to light, and more. All this information is easily accessible through a user-friendly graphical user interface (GUI). The NOMAD *AI Toolkit* provides a collection of examples and tools to demonstrate how materials data can be turned into knowledge. This platform hosts "interactive codes", i.e., data-science notebooks that run data-analytics methods. They concern topics like crystal-structure prediction, property prediction, error estimates, classification of materials, etc. They often serve the purpose of supplementary information to a peer-reviewed publication.

COLLABORATIONS AND OUTREACH NOMAD collaborates with many researchers and all other big computational materials databases, specifically AFLOW (aflow.org), Materials Project (materialsproject.org), and OQMD (oqmd.org) and it provides their (raw) data to the community. A further activity initiated by NOMAD concerns the foundation of the international non-profit association FAIR-DI e.V. (https://fair-di.eu) where the FHI, the Max Planck Institute for Polymer Research and the MPCDF are founding members. NOMAD has been and is being represented in numerous plenary, keynote, and invited talks at international conferences and institutions all over the world. Regular online tutorials are given by the developers team, and all videos can be found on the NOMAD YouTube channel [3].

HISTORY AND TIMELINE The NOMAD Repository as established in its first form as a collaborative project between Berlin (FHI and HU Berlin) and the MPCDF (going online already 2014), was largely extended within the EU Center of Excellence called NOMAD Laboratory CoE (2015-2018), adding the Archive, the Encyclopedia, and the Analytics Toolkit together with advanced visualization tools. With the help of the Max Planck network BiGmax (https://www.bigmax.mpg.de/), the underlying digital infrastructure was homogenized and modernized to allow for substantial extensions. The hardware is hosted by the MPCDF.

CURRENT ACTIVITIES AND OUTLOOK The next huge steps will be extensions in terms of materials synthesis, various experimental techniques, as well as more theoretical and computational methods. All this shall be realized in the realm of *FAIRmat* (FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids, https://fair-di.eu/fairmat), a consortium within the German NFDI (Nationale Forschungsdateninfrastruktur). The consortium consists of 60 top scientists from various fields all over Germany, and is embedded in the international landscape with collaborations and memoranda of understanding with the key players from Europe, the

US, and Asia. The challenges ahead are enormous, considering the breadth of the field, the wealth of physical methods and techniques, and the heterogeneity of data. FAIRmat sets out to tackle these challenges by an inclusive, user-driven, and non-bureaucratic approach to develop easy-to-use tools and an infrastructure towards FAIR data processing, storage, curation, sharing, and future use of materials data. Obviously, this includes data quality aspects as much as electronic lab books and laboratory information systems, materials synthesis recipes, and AI tools.

In summary, NOMAD has initiated data sharing in the field of computational materials science, changing the scientific culture towards openness and re-purposing of results. In this spirit, it is a successful forerunner and the nucleus for FAIRmat that aims at making all materials science data from synthesis, experiment, theory and computation FAIR, i.e., *Findable, Accessible, Interoperable*, and *Reuseable* – or, according to a forward looking interpretation of the acronym, *Findable and AI Ready*. Achieving this will advance (and change) the way how science is done today.

PUBLICATIONS
1. *C. Draxl and M. Scheffler, Big–Data–Driven Materials Science and its FAIR Data Infrastructure. Plenary Chapter in Handbook of Materials Modeling (eds. S. Yip and W. Andreoni), Springer, p. 49 (2020)*
2. *C. Draxl and M. Scheffler, The NOMAD Laboratory: From Data Sharing to Artificial Intelligence, J. Phys. Mater. 2, 036001 (2019)*
3. *NOMAD tutorials on YouTube: https://www.youtube.com/channel/ UCG1B2E82zuJ–MChCuofOBaA*

Fritz Haber Institute of the Max Planck Society / The NOMAD (Novel Materials Discovery) Laboratory The NOMAD Laboratory at the FHI addresses fundamental aspects of the chemical and physical properties of surfaces, interfaces, clusters, and nanostructures. Specifically, the research concerns heterogeneous catalysis, thermal and electrical transport, thermoelectric materials, crystal growth, and defects in semiconductors. These studies combine quantum mechanical, *ab initio* calculations of the electronic structure and molecular dynamics with methods from thermodynamics and statistical mechanics. To address these questions, novel, more accurate, and more efficient first-principles methodologies are developed and implemented in massively parallel scientific software. Along these lines, the group is also a leader in developments and applications of artificial-intelligence methods for materials science.

The group initiated and coordinates a web-based platform for FAIR sharing of materials science data and a Center of Excellence for exascale software developments in the field of computational materials science. Together with the MPCDF it maintains the world's largest *Repository* for computational materials science, as well as a *Material Encyclopedia* and a *Big-Data Artificial Intelligence Toolkit*. It is also co-leading the consortium FAIRmat (FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids) which extends the NOMAD concepts to materials synthesis, experiments, and theory.

Force generation by polymerizing and depolymerizing microtubules

Eukaryotic microtubules, mesoscopic cellular filaments, grow by addition of tubulin proteins at their dynamic ends and operate as chemical-to-mechanical energy transducers with stochastic transitions to an astounding shortening motion. We used large-scale atomistic simulations to reveal the mechano-chemical origin of microtubule-based force generation, with the ultimate goal to understand the functioning of these machines at the atomistic scale.

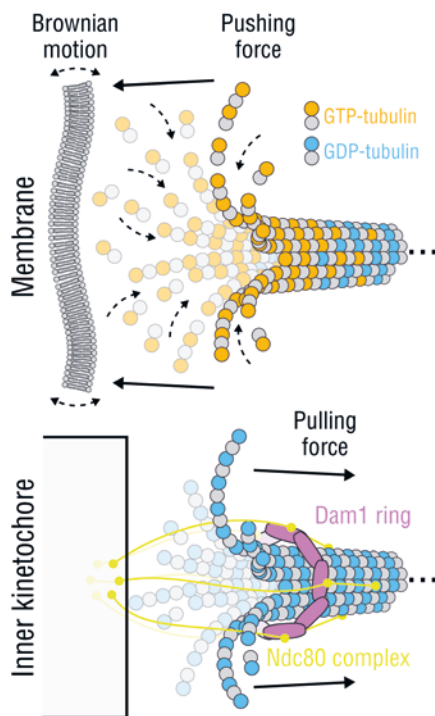AUTHORS *Maxim Igaev and Helmut Grubmüller*

Fig. 1: Biochemical basis of force generation by polymerizing and depolymerizing microtubules (adapted from [2, 6]). Top: A MT grows against an intracellular 'obstacle', e.g, a lipid membrane. Stochastic fluctuations allow for incorporation of tubulin dimers into the lattice, leading to a displacement and/or deformation of the membrane. Bottom: Schematic showing the attachment of the MT tip to the kinetochore Ndc80–Dam1 complex that is thought to slide towards the minus–end using the mechanical force transmitted from pealing protofilaments during depolymerization. α–tubulins are shown as gray circles, whereas β–tubulins are depicted as orange (GTP–bound) or blue (GDP–bound) circles.

Together with other filaments, microtubules (MTs) constitute the skeleton of eukaryotic cells (called the *cytoskeleton*). In contrast to its literal meaning, the MT cytoskeleton is generally very dynamic and covers a broad range of functions. For example, during cell division, the mitotic spindle composed of MTs, chromosomes and a variety of accessory proteins, separates the genetic cell material into two parts [1]. Yet, dynamic MTs are able to produce forces by simply colliding with membranes, organelles or other protein complexes and to explore the intracellular space, filling it with arrays of cargo tracks [2]. MTs elongate by the addition of tubulin proteins with bound guanosine triphosphate (GTP). The bound GTP is then hydrolysed inside MTs to guanosine diphosphate (GDP) and phosphate, driving conformational changes that render MTs more unstable. When a critical density of GDP-tubulins is reached, MTs initiate 'explosive' shortening events (called catastrophes), releasing GDP-tubulins from their tips. Subsequently, these released tubulins exchange GDP for GTP and repeat the cycle.

MTs are hollow cylinders and are composed of heterodimers of α- and β-tubulin subunits assembled in a head-to-tail fashion. Linear arrays of tubulin dimers (along the MT main axis) are termed protofilaments (PFs), 13-14 of which are coupled laterally to make up the MT lattice [3]. This polarized arrangement of tubulin in the lattice extends into a supra-molecular polarity of MTs, with an α-tubulin exposed 'minus-end' and a β-tubulin exposed 'plus-end'. While the minus-end dynamics are usually suppressed by MT nucleation complexes [4], the plus-end is constantly switching between phases of growth and shrinkage, a property known as *dynamic instability* [5]. Despite the stochastic nature of this process, growing/shrinking MTs are capable of generating substantial pushing (e.g., depolymerization in the presence of load-bearing kinetochore attachment; see Fig. 1, top) and pulling (e.g., polymerization against a lipid membrane; Fig. 2, bottom) forces, resulting in the displacement or deformation of movable or flexible intracellular obstacles. The pulling forces associated with MT depolymerization are in the range of
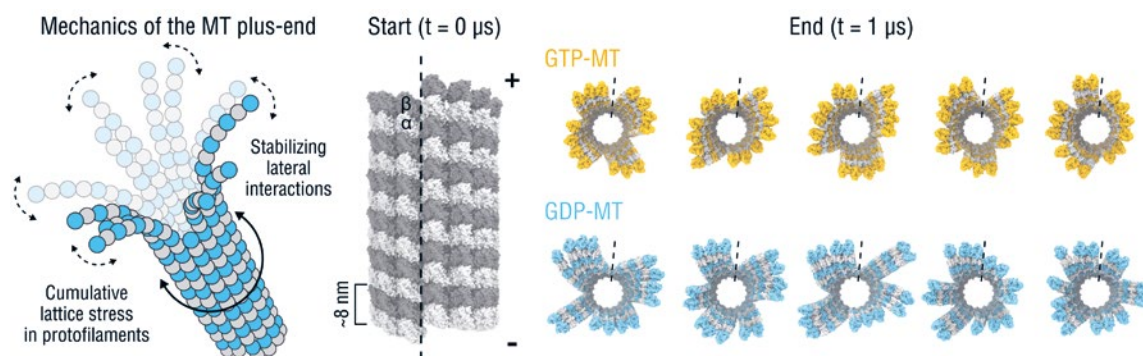
Fig. 2: Left: Conceptual framework of the MT plus–end dynamics: at every moment of its evolution, the mechanical stability of MT tips is determined by the fine balance between lateral interactions of protofilaments and intrinsic lattice stress. Middle: Side view of the MT tip model at the start. Atomistic structures are shown in surface representation, and solvent molecules are not shown for clarity. The lowest row of the α–tubulins were restrained during the simulations. Right: Top view of the GTP– (orange) and GDP–MT (blue) tips after 1 μs of MD simulation.

30-65 pN and are, therefore, almost an order of magnitude larger than the pushing forces produced by MT growth (3-4 pN) [2].

With their well-separated modes of operation – chemical energy consumption, storage, and release – MTs are prototypes for studies on fundamental nanomechanics principles. Our goal is to understand the functioning of these astounding machines at the atomistic level. The bulk kinetics of MT assembly have been fairly well characterized compared to the MT plus-end structure, the dynamics of which are still hard to sufficiently resolve in time and space with modern techniques, despite its critical role in setting up these kinetic rules. What remains unknown is, therefore, what role the plus-end morphology plays in the MT-based force generation. A clear mechanochemical picture that describes and explains all these phenomena is not yet available.

To answer these questions, we carried out intensive, atomistically resolved molecular dynamics (MD) simulations at the MPCDF. Here, the motion of millions of interacting atoms constituting the MTs and the surrounding water are calculated from the fundamental laws of physics. On the one hand, our simulations rest upon accurate structural models of tubulin and MTs obtained by high-resolution X-ray crystallography and cryo-electron microscopy. On the other hand, this allows us to analyse and explain functional motions of the MT plus-ends – information that is not easily amenable in experiment. Our MD simulations, which can take several months of computer time in order to cover sufficient simulation time scales (microseconds), hence pose significant numerical challenges, which can be tackled only by high-performance supercomputer facilities such as the MPCDF.

For example, we currently study the structure and dynamics of MT plus-ends before and after GTP hydrolysis to infer the differences in elastic properties responsible for MT-based force generation (Fig. 2; [7]). For the first time, we were able to perform MD simulations of the primary desintegration steps of *complete* MT tips (approx. 15 million atoms including water molecules; Fig. 2, middle

and right). We have also recently shown how the nucleotide state (GTP or GDP) affects the fine mechanical balance between protofilament interactions and intrinsic lattice stress induced by conformational changes in tubulins upon GTP hydrolysis, thus determining if the MT collapses or continues to grow (Fig. 2, left). Further simulations will allow us to exactly determine what functional motions of the MT plus-ends are crucial for the elastic energy transmission, as well as to explain the asymmetry in the forces generated by growing and shrinking MTs.

PUBLICATIONS
1. *Annu. Rev. Biophys, 45(1): 279–298 (2016)*
2. *Cell Adhesion & Migration, 10(5): 475–494 (2016)*
3. *Cell, 96(1): 79–88 (1999)*
4. *Nat. Commun., 11(1): 5705 (2020)*
5. *Nature, 312(5991): 237–242 (1984)*
6. *Science, 360(6388): 552–558 (2018)*
7. *Igaev, M. and H. Grubmüller (2021), in preparation*

MAX PLANCK INSTITUTE
FOR BIOPHYSICAL CHEMISTRY

(Ring) polymers and DNA Territories

Extended computer simulations support the hypothesis that generic polymer properties of topological constraints are behind many aspects of chromatin folding in eukaryotic cells. For that purpose, concentrated, topologically simple (unknotted and unlinked) chains are considered and compared to experimental findings. Recent extensions to the physics of active systems are also mentioned.

AUTHOR *Kurt Kremer*

POLYMER DYNAMICS for a long time already, is a subject of intense experimental and theoretical research due to both its scientific and technological relevance and its richness. This equally holds for synthetic polymers such as traditional commodities (polyethylene, polystyrene, …) as well as for polymers of biological origin (DNA, IDPs, …). In the simplest models, such chain molecules are modelled as random walks, meaning that their radius of gyration is proportional to $N^{1/2}$, when N is the number of repeat units or beads. While this simple structural model does not apply to chains in dilute solution, it is a very good approximation for dense solutions and melts of polymers.

As a consequence of such an open structure (the chains' fractal dimension is $d_f=2$), the overall volume of each chain in equilibrium is shared by $O(N^{1/2})$ other chains. This has dramatic consequences for the dynamics. Chains cannot cut through each other, i.e., their thermal motion is subject to severe topological constraints; they are entangled. For short chains this is not easily detectable, as their averaged motion is isotropic in space and all the chains and their beads act like a frictional surrounding and a heat bath. The diffusion constant scales as $D\sim1/N$ and the longest relaxation time is $\tau\sim N^2$. This so called Rouse behavior is well supported by simulation and experiment. When the chains are becoming very long, this changes dramatically to $D\sim1/N^2$ and the longest relaxation time $\tau\sim N^{3.4}$, with severe consequences for their rheological behavior. By now, the underlying tube (reptation) model of De Gennes and Edwards is well supported by experiment and simulations, where in this case simulations actually provided the first microscopic evidence. The characteristic chain length where the crossover occurs depends on details of the chemistry and can be predicted quite accurately. For very long chains, the longest relaxation time easily can reach days or even much longer times.

CHROMOSOME TERRITORIES In contrast to an equilibrated melt of long chain polymers, chromosomes in the interphase cell nucleus of eukaryotes, once the corresponding 30nm fibre is long enough, do not interpenetrate and are organized in territories. This is a structure far from equilibrium. However, there are good arguments that DNA in the nucleus does not form an equilibrated dense solution. The dynamics simply would be far too slow. Rosa and Everaers [1] estimated the conformational relaxation time, and thus the
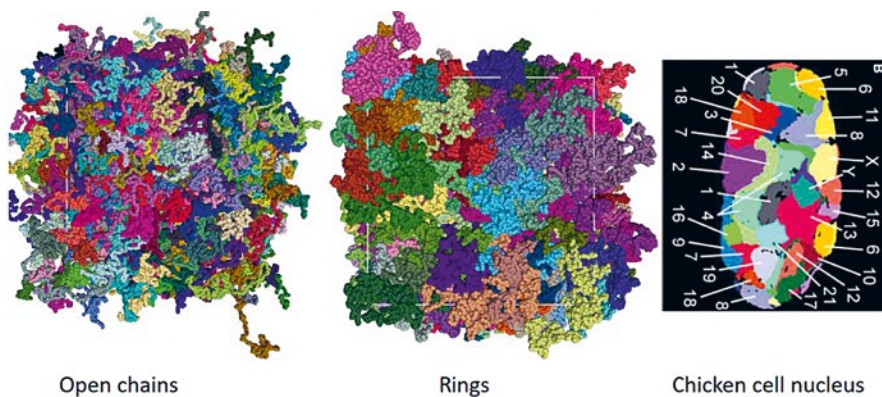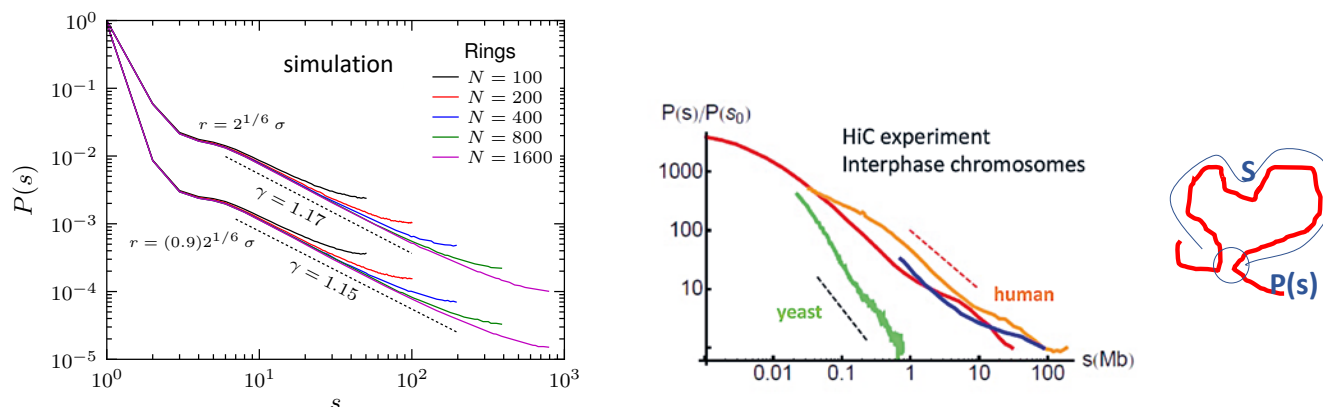

Fig. 1: Comparison of the territorial organization of chromosomes in a chicken cell nucleus (right), an equilibrated melt of non–concatenated ring polymers (middle) and an equilibrated melt of long chain polymers (left) [3].

Fig. 2: Intra–chain/intra–chromosome distance distribution for different organisms (right [3]) and ring polymers (left [4]). Simulations based on ring polymers, experiment HiC data, after [4].

time scale also related to cell division, for the human cell nucleus to be about a year or even longer, while for unentangled, segregated chains it is several orders of magnitude faster. Based on this, there is no reason for yeast to develop a territorial structure, the chromosomes are simply too short to be affected by the slow reptation dynamics [1, 3]. This raises several questions from a polymer physics point of view. How is this peculiar structure formed and – that is the question we are addressing here – what kind of measures are needed to keep that structure stable? (At least for the lifetime of the interphase nucleus.)

TOPOLOGICAL REPULSION The simplest model for a dense polymer system with conserved topological constraints is a melt of non-concatenated and un-knotted ring polymers [2, 3, 4]. This has the advantage that even for short chains we can study the role of topology conservation on the chain conformation. The fact that the missing free ends on the relevant time scales are irrelevant has been considered separately. We could show that for such ring polymers the equilibrium structure is a compact object, which to a large extent avoids chain overlap, as illustrated in Figure 1. Furthermore, the internal probabilities P(s) that two monomers located s bonds apart along the chain meet each other in space compare very well to HiC experiments on chromosomes (Fig. 2.)

Upon relaxation, which roughly follows the fast Rouse type of relaxation, the territorial organization is preserved. This does not yet explain local details of DNA function. However, these results clearly show that once the territorial state in the cell nucleus is reached, it is metastable and there is no complex biochemical apparatus needed to preserve it.

Recently, (bio-)physics is more and more turning to active systems, i.e., systems where some constituents are permanently driven out of equilibrium. Again, polymers offer excellent opportunities to study such problems. By variation of the active/inactive part of the chains, properties can be fine-tuned very efficiently. Based on extensive simulations of partially active non-concatenated ring polymers the new phase of an active topological glass has been predicted [5].

This work benefited significantly from fruitful collaboration with AY Grosberg, J. Smrek, JD Halverson and GS Grest.

PUBLICATIONS

1. *A Rosa, R. Everaers: Structure and Dynamics of Interphase Chromosomes PLoS Comp. Biol. 4, e1000153 (2008)*
2. *T Vettorel, AY Grosberg, K Kremer: Statistics of polymer rings in the melt: a numerical simulation study. Physical Biol. 6, 025013 (2009)*
3. *JD Halverson, J. Smrek, K Kremer, AY Grosberg: From a melt of rings to a chromosome territories: the role of topological constraints in genome folding. Rep. Prog. Phys. 77, 022601 (2014)*
4. *JD Halverson, WB Lee, GS Grest, AY Grosberg, K Kremer: Molecular dynamics simulation study of nonconcatenated ring polymers in a melt. I. Statics, J. Chem. Phys. 134, 204904 (2011), ...II. Dynamics. Ibid 134, 204905 (2011)*
5. *J Smrek, I Chubak, CN Likos, and K Kremer: Active topological glass; Nature communications 11, 1–11 (2020), DOI:10.1038/*

MAX PLANCK INSTITUTE FOR POLYMER RESEARCH

Max Planck Institute for Polymer Research / Theory Department
The Max Planck Institute for Polymer Research (MPI-P) in Mainz is one of the leading international research centers in the field of polymer or, more general, soft matter research. We conduct basic research on soft materials and systems, which are at the center of scientific and technological challenges related to major societal needs, such as energy, healthcare, mobility, and information technology. Science at the Institute is carried out by six departments ranging from macromolecular synthesis to theory. To achieve that, man-made as well as systems of biological origin are studied. The institute follows a comprehensive interdisciplinary approach in order to understand and control the properties of soft matter in order to create new functions. Complementing the experimental groups, the theory department uses a variety of different advanced (mostly) computational methods to understand and predict properties of and processes in soft matter. This also includes utilizing soft matter as a means to study new physics problems.

# *In silico* atomistic visualization of proteins engaging with genomic DNA

Genomic DNA is wrapped around histone proteins in nucleosomes. Arrays of nucleosomes form chromatin, a highly dynamic structure that encodes gene regulatory programs. Structural rearrangements of chromatin are hallmarks of biological development and pathology and occur with mechanisms that are still elusive. To study how nucleosome motions contribute to these mechanisms, we harvested the power of the so-called "computational nanoscope" which enables the visualization of biomolecular structures and dynamics at atomistic resolution. We strive to reveal how nucleosome flexibility is controlled by nucleosome interacting proteins such as the transcription factors that regulate and induce stem cell pluripotency.

AUTHORS *Jan Huertas, Hans R Schöler, and Vlad Cojocaru*

Every cell in the body contains two meters of DNA. In order to fit all the DNA in the cell's small nucleus, the DNA is tightly packed in a structure known as chromatin. Chromatin is an array of smaller, identical structures named nucleosomes. In a single nucleosome, the DNA is wrapped around eight proteins called histones. Chromatin is not uniformly compact across the genome. The tightness of the packaging is important in regulating which genes are expressed and therefore which proteins are produced by a cell. Transitions from tightly to loosely packed DNA – from closed to open chromatin – are essential for cells to convert to another cell type. These cell conversions are hallmarks of development and disease, but are also often used in regenerative therapies. Understanding how such transitions occur may help to understand disease mechanisms and to optimize therapeutical cell type conversions. One contribution to chromatin opening comes from the motions within single nucleosomes and the DNA that connects them. Like all molecular structures in our cell, nucleosomes are dynamic. They move, twist, breathe, unwrap and rewrap. Visualizing these motions using experimental methods is often very challenging. An alternative is to use the so-called "computational nanoscope".

COMPUTATIONAL NANOSCOPE is a term that refers to a collection of molecular modeling and simulation methods that enable the visualization of bio-molecular dynamics at atomistic and sub-atomistic resolution. The structures of biomolecules are either available from experiments or can be built using structural mod-

Fig. 1: 3 µs from the life of a nucleosome. Snapshots in time were superimposed on the core region of the histones (white). The positions of the DNA (yellow) and the flexible histone tails (blue, green, orange, red, cyan) in the snapshots are shown. The image was chosen as the cover of the June 2021 issue of PLoS Computational Biology.
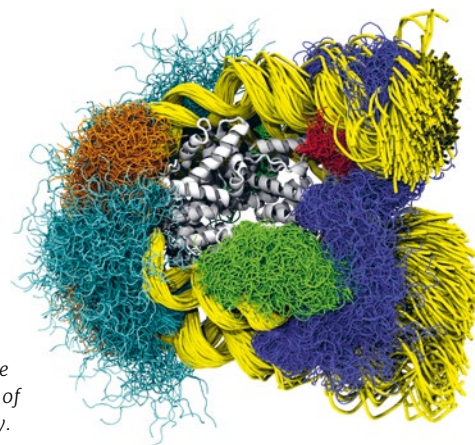
eling approaches that range from *ab inito* prediction to sequence and structure homology based modeling. The dynamics of these structures can be studied using molecular dynamics (MD) simulations, in which trajectories in time of molecular systems are generated by integrating Newton's classical equation of motion. The forces needed to integrate this equation are obtained from the potential energy of the system, also known as the force field. The quality and accuracy of the simulations depend on the parameters used to estimate this energy (the force field parameters). The integration time step depends on the resolution of the system. When all atoms are included, the time step is limited by fast chemical bond vibrations and can be set to 1 to 4 femtoseconds. If the molecular system is modeled with a reduced, coarse-grained representation, the time step can be increased and longer times can be simulated. Over the

past years, atomistic MD simulations have become so accurate that they can be viewed as a computational nanoscope; observing the molecules moving on the computer is similar to observing them under a very high-resolution nanoscope. For high accuracy, simulations are performed in an explicit solvent (e.g., water) under conditions resembling *in vitro* experiments.

Recently we reviewed the application of the computational nanoscope to study chromatin dynamics [3]. The largest system simulated to date at atomistic resolution is a gene locus with 428 nucleosomes. However, such large systems can be simulated only for a very short time (a few nanoseconds). Even for smaller systems, the major challenge remains to sample time scales long enough to observe motions that are relevant for the function of biological molecules. For single nucleosomes, using resources available at the Max Planck Computing and Data Facility (MPCDF), we are now able to generate a trajectory of one microsecond in one to two months of computing time depending on the queuing times.

NUCLEOSOME MOTIONS change the relative position of nucleosomes, regulating the compaction of chromatin. In each nucleosome, 145-147 base pairs of DNA are wrapped around 8 histones. At each end, the nucleosomal DNA extends with a linker DNA arm that connects to the next nucleosome. The histones have a structured core and one or two terminal unstructured regions known as tails. Nucleosome motions include breathing, twisting, wrapping and unwrapping and depend on the sequence of the DNA, the histone composition (canonical or variant histones) and the histone tails. In particular, the linker DNA and the histone tails are highly flexible. The opening-closing motion of the linker DNAs (breathing, Fig. 1) may influence the relative arrangement of neighboring nucleosomes. The histone tails contain sites that are chemically modified and control the degree of chromatin compaction in the genome. Specific modifications of these tails are found in active and inactive chromatin. Two histone tails (blue and red in Fig. 1) are located near the linker DNA arms.

Although it has been studied extensively how these tails influence chromatin dynamics, it is still poorly understood how these tails control nucleosome breathing. To elucidate this, we performed multiple atomistic MD simulations, over 25 microseconds in total, of three nucleosomes, two with genomic DNA sequences and one with a sequence engineered for strong positioning of the histones. From these simulations we found that nucleosome breathing is modulated by the DNA sequence [2, 4] and the motions of the two histone tails [2]. The tails restrict the motion of the nucleosomes, keeping them closed. However, large amplitude motions of the tails allowed nucleosome opening. We were able to pinpoint interactions between amino acids in the histone tails and the DNA that influence nucleosome breathing. With these findings, we revealed the correlation between nucleosome breathing and histone tail dynamics. Next, we aimed to understand how nucleosome binding proteins influence these motions.

PIONEER TRANSCRIPTION FACTORS are a special class of proteins that regulate gene expression upon binding to the DNA wrapped in nucleosomes in closed chromatin. Many of them are involved in cell fate transitions that require large chromatin structural rearrangements. For example, only three pioneer factors (Oct4, Sox2, and Klf4) are sufficient to transform an adult cell into a cell almost identical to an embryonic pluripotent stem cell. How these factors recognize their binding sites on the nucleosome and how

Fig. 2: Structure of the Oct4–nucleosome complex. The nucleosome coloring is the same as in Figure 1; only the histone tail that influences nucleosome opening is highlighted in blue. The nucleosome is in an open conformation stabilized by Oct4. The two DNA binding subdomains of Oct4 are in cyan and orange.
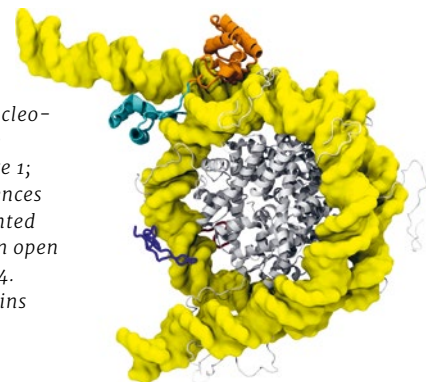
their binding affects nucleosome dynamics is still unknown. To answer these questions, we built structural models of Oct4 bound to genomic nucleosomes at different locations [4]. We then performed multiple MD simulations with a length of one microsecond for each model. From these simulations, totaling more than 25 microseconds, we found that Oct4 interprets the intrinsic nucleosome dynamics as it requires a partial opening of the nucleosome to bind to one class of binding sites. Oct4 stabilizes the partially open nucleosome conformations. On other binding sites, Oct4 enhances nucleosome flexibility and under certain conditions, it induces a large opening of the nucleosome (Fig. 2). This depends on the two histone tails mentioned above. In addition, it also depends on the motion of both DNA binding subdomains of Oct4. One subdomain binds to the sequence specific sites whereas the other explores the DNA unspecifically. These findings reveal the structural features and dynamics involved in Oct4 binding to different locations on the nucleosome and have implications for understanding chromatin dynamics and the function of pioneer transcription factors.

PUBLICATIONS
1. *MacCarthy CM, Huertas J, Ortmeier C, vom Bruch H, Reinke D, Sander A, Bergbrede T, Schöler HR, Cojocaru V (2021) OCT4 interprets and enhances nucleosome flexibility. BioRxiv, https://doi.org/10.1101/2021.04.27.441583*
2. *Huertas J, Schöler HR, Cojocaru V (2021). Histone tails cooperate to control the breathing of genomic nucleosomes. PLoS Comput. Biol., PLoS ComputBiol, 17(6):e1009013*
3. *Huertas J, Cojocaru V (2021). Breaths, twists, and turns of atomistic nucleosomes. J. Mol. Biol., 433:166744*
4. *Huertas J, MacCarthy CM, Schöler HR, Cojocaru V (2020). Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding. Biophys. J., 118(9):2280–2296 (journal cover)*

MAX PLANCK INSTITUTE
FOR MOLECULAR BIOMEDICINE

Max Planck Institute for Molecular Biomedicine / Department of Cell and Developmental Biology The Max Planck Institute for Molecular Biomedicine in Münster, Germany, is devoted to basic science investigating the molecular mechanisms underlying physiopathological processes in development and disease. The researchers use gene technology, molecular biology, electron and laser microscopy and computational approaches to reveal how cells and biomolecules shape up the endothelium, the vascular system, the stem cells, and the embryo. Research in the Department of Cell and Developmental Biology aims at understanding the mechanisms controlling the maintenance and induction of stem cell pluripotency *in vivo, in vitro,* and *in silico.*

Tracking the movements and lives of wild animals

The Movebank platform was created to support data owners in archiving animal movement data for future use, enable collaborations, help scientists address new questions by sharing data, promote open access to data, and allow the public to explore the amazing animal movements recorded by movement ecologists.

AUTHORS *Sarah Davidson and Martin Wikelski*

REMOTE SENSING FOR BIODIVERSITY Animal-borne sensors allow us to follow the lives of wild animals – to document their movements, study their behavior and evolution, and monitor and conserve their populations. Combined with other ecological information, we can address challenges of climate and land use change, biodiversity loss, invasive species, wildlife trafficking, and infectious disease. To these ends, tens of thousands of animals around the world are equipped with sensors that record data for researchers and wildlife managers (Fig. 1). Through these sensors and the expertise of the biologists that deploy them, we can detect animals starting migrations, breeding and caring for young, encountering possible threats, interacting with each other, and, ultimately, dying.

The *Movebank* platform (https://www.movebank.org), hosted by MPCDF and the Max Planck Institute of Animal Behavior, contains billions of locations and other sensor measurements for over a thousand species, and offers services and tools to help researchers and the public manage, share, discover, and analyze these data.

- We work with sensor manufacturers to let researchers *monitor wild animals in near-real time*, supporting public engagement and dynamic responses to wildlife threats.
- Through our online services, researchers *"translate" data from hundreds of formats into one*, with documentation and tools allowing these data to be understood, preserved, and combined.
- Data owners have *flexible sharing options* allowing public and controlled-access sharing, to support data management starting with initial data collection, as well as sensitive data about protected species.
- Citizen scientists around the world can see live data and engage directly with researchers through contributing photographs and messages on the *Animal Tracker App*.
- The *Env-DATA System*, Movebank's data annotation service, lets users link animal occurrences with hundreds of other environmental variables from global weather models and remote sensing products.
- Developers can use Movebank's APIs to *build websites, code and software* that are compatible with data from Movebank.
- Our archiving tools, including a public data repository, allow researchers to *publish data and preserve it for future generations* (Fig. 2).

Fig. 1: Locations of studies visible to the public at movebank.org as of 28 April 2021.
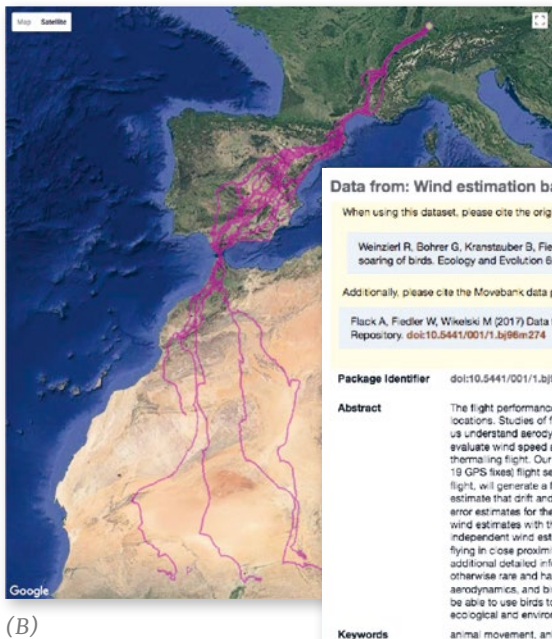
(B)

(A)

Fig. 2: Example of data re-use using Movebank, showing the published data summary (A), a map of the tracks on Movebank (B), and three peer-reviewed research papers using the dataset (C) (Flack et al., 2017).

(C)

GLOBAL PARTICIPATION Over 3,000 data owners around the world use Movebank to store data cited in nearly 700 research publications, and tens of thousands of people use the website and mobile app to learn about wildlife, where they live and beyond. By offering a global "living" platform, we provide a unique space to discover and access data and expertise, supporting large-scale collaborations such as the Covid-19 Bio-Logging Initiative, the Audubon Migratory Bird Initiative, and the Arctic Animal Movement Archive. We partner with governments and conservation organizations to support such initiatives and to meet data management and archiving requirements.

MAX PLANCK INSTITUTE
OF ANIMAL BEHAVIOR

Max Planck Institute of Animal Behavior / Department of Migration
The Department of Migration is part of the Max Planck Institute of Animal Behavior, which aims to understand and predict animal decision-making in the natural world. By developing and applying emerging technological and analytical approaches, our mission is to reveal the drivers of behavior across temporal, spatial, organizational, and taxonomic scales. To take up this challenge in all its complexity, we bring together a diverse, interdisciplinary team to produce rigorous and reproducible science that is shared openly. In this way, we contribute positively to the global research community and provide scientific training to empower the next generation. We foster a supportive and inclusive work environment that promotes intellectual exchange and productive collaboration.

PUBLICATIONS
1. *Davidson SC, Bohrer G, Gurarie E, LaPoint S, Mahoney PJ, Boelman NT, Eitel JUH, Prugh LR, Vierling LA, Jennewein J, et al. 2020. Ecological insights from three decades of animal movement tracking across a changing Arctic. Science. https://doi.org/10.1126/science.abb7080*
2. *Dodge S, Bohrer G, Weinzierl R, Davidson SC, Kays R, Douglas D, Cruz S, Han J, Brandes D, Wikelski M. 2013. The Environmental-Data Automated Track Annotation (Env-DATA) System: linking animal tracks with environmental data. Movement Ecology. 1:3. https://doi.org/10.1186/2051-3933-1-3*
3. *Douglas DC, Weinzierl R, Davidson SC, Kays R, Wikelski M, Bohrer G. 2012. Moderating Argos location errors in animal tracking data. Methods Ecol Evol. 3(6):999–1007. https://doi.org/10.1111/j.2041-210X.2012.00245.x*
4. *Flack A, Fiedler W, Wikelski M. 2017. Data from: Wind estimation based on thermal soaring of birds. Movebank Data Repository. https://doi.org/10.5441/001/1.bj96m274*
5. *Kays R, McShea WJ, Wikelski M. 2020. Born-digital biodiversity data: millions and billions. Divers Distrib. 26(5):644–648. https://doi.org/10.1111/ddi.12993*
6. *Kranstauber B, Cameron A, Weinzierl R, Fountain T, Tilak S, Wikelski M, Kays R. 2011. The Movebank data model for animal tracking. Environ Model Softw. 26(6):834–835. https://doi.org/10.1016/j.envsoft.2010.12.005*
7. *Rutz C, Loretto M-C, Bates AE, Davidson SC, Duarte CM, Jetz W, Johnson M, Kato A, Kays R, Mueller T, Primack RB, Ropert-Coudert Y, Tucker MA, Wikelski M, Cagnacci F. 2020. COVID-19 lockdown allows researchers to quantify the effects of human activity on wildlife. Nat Ecol Evol. 4:1156–1159. https://doi.org/10.1038/s41559-020-1237-z*
8. *For more, see https://www.movebank.org/cms/movebank-content/literature*

A dynamic general equilibrium approach to migration in economic history

Our research agenda relies on a theory-based empirical approach to investigate the interplay between natural and spatial population movements and economic development during Western Europe's industrialization. We pin down the role of alternative mechanisms during the transition to modernity, using advanced calibration techniques based on dynamic general equilibrium models.

AUTHORS *Thomas Baudin[10] and Robert Stelter[11]*

In their seminal paper, Galor and Weil (2000) introduced the *Unified Growth Theory* (UGT). It explains historical facts on the economic and demographic evolution in a unique and consistent framework. While natural population movements (fertility and mortality) have been widely studied in the UGT context, only very few studies have focused on urban-rural differences and internal migration (see e.g., Adams 2021). Our research agenda aims to deepen the understanding of the role of the countryside and cities – historically two different economic systems – and rural-urban population mobility in the transition from stagnation to modern growth. To do so, we build and calibrate *Dynamic General Equilibrium* models (DGE). The availability of increasing computing power on computational clusters enables us to go beyond simple numerical applications. We thus, first, collect data to quantify historical facts; second, develop economic DGE models; third, confront the theoretical framework with the collected data; and, fourth, run counterfactual experiments that allow testing hypotheses to gain new insights into the role of selected mechanisms.

DATA AND FACTS From a macro perspective, the long-run trends in natural population movement and in economic evolution are well documented. However, as soon as a grid finer than the national one is needed, data become scarcer; in particular if observation windows span over a long time. Here, well-established historical facts often arise from a mixture of few, selective observations and qualitative conclusions from historians, demographers, and economists. The adjustment of definitions hampers the comparability of available data over time. As to our research topic, i.e., cities experienced a transformation during the Industrial Revolution and corresponding statistical definitions may have changed. We thus first collect data
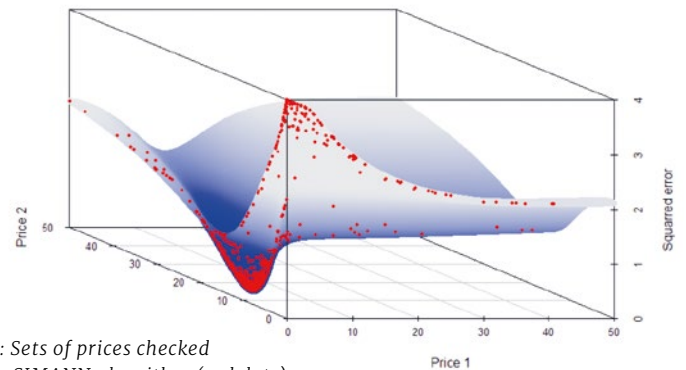

Fig. 1: Sets of prices checked by the SIMANN algorithm (red dots) and the squared distance to the general equilibrium of the economy.

on important variables such as fertility, mortality, urbanization, and sector-specific production at the local level. We then combine these data to infer historical trends. As our further analysis only relies on these trends, we reduce the data requirements and limit the impact of data issues, such as missing observations. This is a major advantage of our research strategy.

FROM ECONOMIC MODELLING TO DATA We develop a Dynamic General Equilibrium model with sector-specific production, cities and countryside, and heterogeneous agents. These rational agents are all characterized by the same well-defined preferences but equipped with varying abilities to generate income; the income is used to consume, bring up and educate children and to pay migration costs if needed. At each point in time, these activities are coordinated by markets. The sequence of these market equilibria describes the evolution of the economy over time.

[10] *IÉSEG School of Management Lille, France*
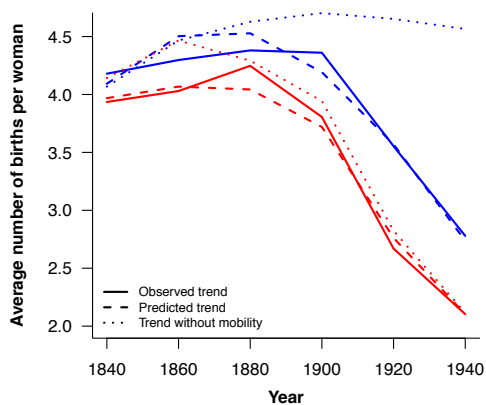[11] *University of Basel, Switzerland*

Fig. 2: Average number of births per woman in the Danish country-side (blue) and cities (red).

We implement our theoretical frameworks using the FORTRAN programming language together with a hybrid OpenMP-MPI parallelization to quantify and test our theories. Computing the equilibrium at each point in time is a typical fixed-point problem: For a given set of parameters and initial conditions, equilibrium prices (i.e., the prices at which supply and demand are equal) depend on individual activities. The individual activities – computed using OpenMP parallelization – in turn, depend on the prices. As the solutions of the equations are typically neither continuous nor monotonic, computing the market equilibriums using standard local algorithms such as Newton or Powell cannot be done without suitable starting points. Thus, we first search for the equilibrium applying the simulated annealing algorithm SIMANN (Goffe et al. 1994). If the algorithm does not approximate the market equilibrium sufficiently well, we use SIMANN's outcome as starting point for Powell's method. Figure 1 shows an example of the economy's general equilibrium as a function of the economy's prices. In this equilibrium, the squared distance between market supply and demand is zero. The dots depict price sets checked by SIMANN to find the prices that balance supply and demand in the markets.

To determine the parameter set that minimizes the distance between the model's predictions and the long-run trends, we proceed in two steps following Baudin et al. (2015): First, we search for a

suitable starting point in the high-dimensional (typically more than 15) parameter space, using the genetic optimization algorithm PIKAIA (Charbonneau, 2002). We take advantage of the fact that parameter sets (or individuals) of the same generation can be parallelized rather straightforwardly across nodes of a high-performance computing cluster, using the Message Passing Interface (MPI). Still, as the computational time to solve the model itself by the numerical algorithms potentially varies drastically – not all parameter sets lead to such a well-behaved equilibrium as plotted in Figure 1 – a static parallelization as provided by MPIKAIA (Metcalfe and Charbonneau 2003) is rather inefficient. To avoid wasting computational resources, we implemented in collaboration with MPCDF a modification of MPIKAIA that schedules the parameter sets dynamically across the MPI processes. Once we know the suitable starting point of the parameter space, we launch UOBYQA, another Powell algorithm. It improves the precision of the solution and thus further improves the fit between predicted and estimated empirical trends.

WHAT WOULD HAVE HAPPENED IF? We can run historical experiments when the models' predictions reproduce historical facts both qualitatively and quantitatively. These counterfactual experiments allow us to go beyond the pure replication of historical trends and to simulate alternative historical evolution paths that quantify the impact of certain conditions on the demographic and economic evolution. Would the absence of mortality improvements have prevented the fertility transition or delayed the take-off to modern growth? How would economies have developed if mobility conditions had been worse or better? Figure 2 illustrates Danish fertility dynamics in such a hypothetical situation with migration costs that prohibit any spatial population movement. The consequences of such an extreme case are dramatic. The fertility transition would have been stronger in cities but would have not set in on the countryside until 1940. Rural fertility simply would have been locked into the state of pre-industrialization while economic urban-rural inequalities would have exploded. Our results thus show: Mobility conditions in the past are a major driver of the contemporary state of the economy.

PUBLICATIONS

1. *Adams, J.J. (2021). Urbanization, long-run growth, and the demographic transition. Journal of Demographic Economics, 1–47. doi:10.1017/dem.2020.36*

2. *Baudin, T., de la Croix, D. and P.E. Gobbi (2015). Fertility and childlessness in the United States. The American Economic Review, 105 (6), 1852–1882. doi: 10.1257/aer.20120926*

3. *Charbonneau, P. (2002). An introduction to genetic algorithms for numerical optimization. NCAR Tech. Note TN–450+ IA, 74pp.*

4. *Galor, O. and D.N. Weil (2000). Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. The American Economic Review, 90 (4), 806–828. doi : 10.1257/aer.90.4.806*

5. *Goffe, W. L., Ferrier, G. D. and J. Rogers (1994). Global optimization of statistical functions with simulated annealing. Journal of Econometrics, 60 (1), 65–99. doi: 10.1016/0304-4076(94)90038-8*

6. *Metcalfe, T. S. and P. Charbonneau (2003). Stellar structure modeling using a parallel ge-netic algorithm for objective global optimization. Journal of Computational Physics, 185 (1), 176–193. doi: 10.1016/S0021-9991(02)00053-0*

MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

The PANDORA initiative

Valuable lessons for present-day societies can be found in the study of the human past. The PANDORA initiative is an international collaborative network that combines the use of big historical data, advanced modeling tools, and artificial intelligence to achieve a deeper understanding of the dynamics of past human societies.

AUTHORS *Ricardo Fernandes and Nicole Boivin*

The PANDORA initiative seeks to improve our understanding of the human past by developing research work along three main components (Fig. 1). It offers a data platform (https://pandoradata.earth/) where an international network of research communities involved in the study of the human past (e.g., history, archaeology, paleo-environmental sciences) can self-manage data storage. Accumulated data (e.g., isotopic proxy data) can then be modeled for spatiotemporal reconstructions of various past human and paleo-environmental phenomena such as the subsistence or spatial mobility strategies of past populations, or the paleo-climatic conditions under which they lived. The PANDORA initiative is also actively developing modeling tools, particularly involving the use of Bayesian methods, which are made available via open access to any researcher involved in the study of the human past (https://pandoraapp.com). Finally, the PANDORA initiative is also introducing the use of artificial intelligence including already developed machine learning methods employed in the study of historical causation.

A COLLABORATIVE EFFORT The PANDORA initiative is supported by the Max Planck Institute for the Science of Human History. However, it represents an international collaboration, which includes also other members of the Max Planck Society. The Max Planck Computing and Data Facility supported the development of the data platform and is currently responsible for its maintenance. The Max Planck Digital Library offers access to unique identifiers (DOIs) employed to track deposited datasets. Funding for the development of open-access modeling methods and artificial intelligence has been provided by the Max Planck Institute for the Science of Human History, the University of Warsaw, the Masaryk University, and the Eurasia3angle and Palaeo-Science and History research groups. In addition, several research institutions and groups have been involved in data research projects within the PANDORA initiative.

RESEARCH PROJECTS The modeling approaches developed within the PANDORA initiative have been employed in several big historical data projects. One such example was the use of a dataset of isotopic measurements on archaeological pottery remains to reconstruct the subsistence of the first farmers along the European Atlantic shoreline (Cubas et al. 2020). Data modeling revealed that the frequency of dairy products in pottery increased as farming was progressively introduced along a northerly latitudinal gradient suggesting a direct influence in the evolution of adult lactase persistence across Europe.

From isotopic measurements of human remains found in Mongolia it was possible to observe that the production and consumption of millet within the region intensified during the Iron Age (Wilkin et al. 2020). This shows that the Xiongnu and Mongol empires were built on a diverse economy in contrast with previous views that described these as being specialized on pastoralism.

Further illustrating the versatility of isotopic data in historical studies, a dataset of isotopic measurements on archaeological animal remains and on environmental samples was
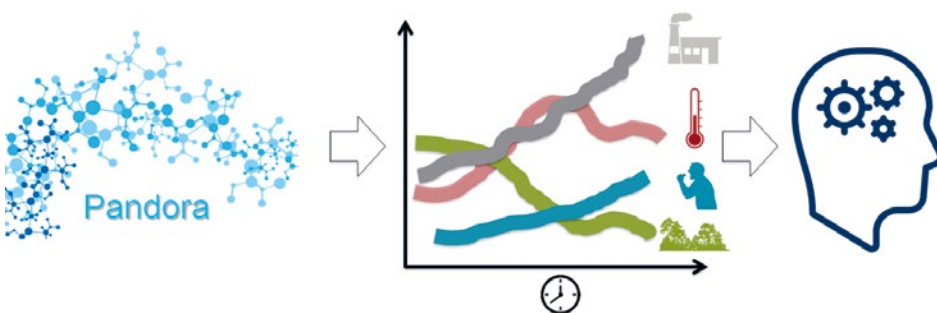

Fig. 1: The Pandora initiative has three main components: building networks of historical and archaeological data, developing and making accessible the modeling tools for spatiotemporal reconstruction of past phenomena, and using artificial intelligence to aid in our understanding of historical dynamics.
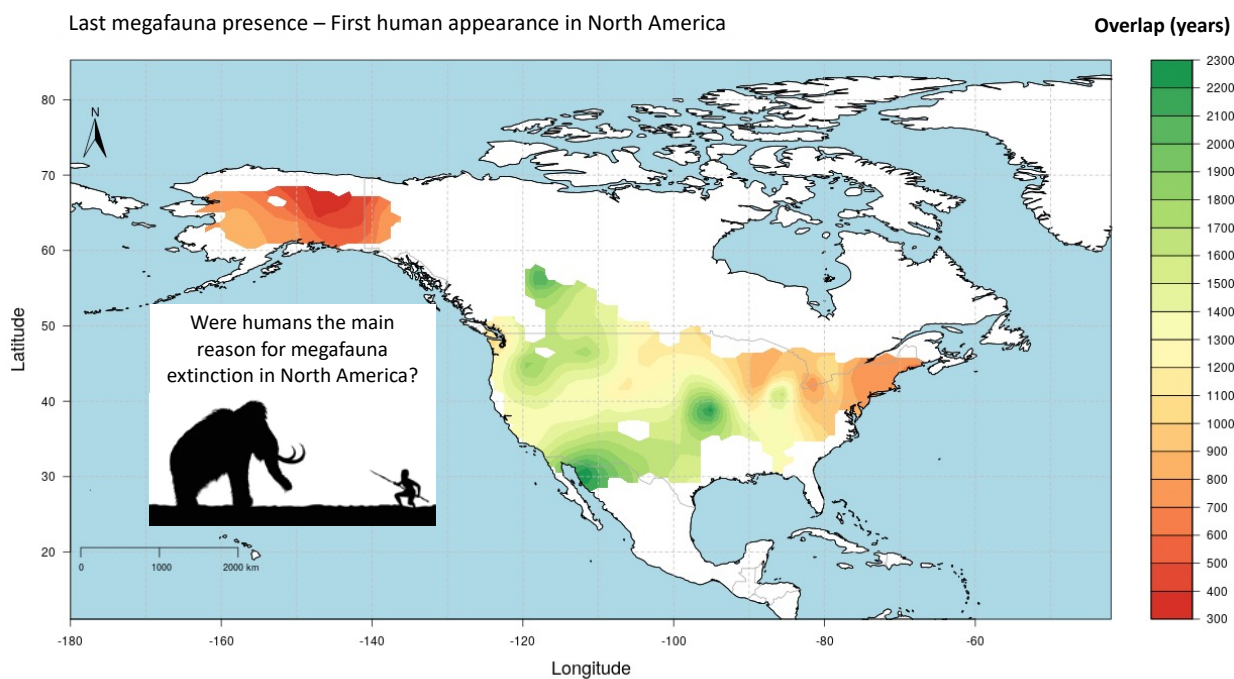
Fig. 2: Modeling results for temporal overlap between humans and megafauna across North America. These were generated using datasets of human and megafauna radiocarbon measurements, from which the last presence of megafauna and the first human appearance in North America were estimated.

used to investigate the past circulation of animals in the Yellow River Basin across a wide extent of Chinese history (Wang et al. 2021). Novel Bayesian modelling revealed that animal trade was already practiced during the Late Neolithic. In some cases, it was possible to pinpoint the locations of origin for traded animals, which were up to several hundreds of kilometers from their final destination.

MEGAFAUNA EXTINCTION IN NORTH AMERICA The largest dataset of human and megafauna radiocarbon measurements for North America is being gathered as part of a data project at the Max Planck Institute for the Science of Human History. Once data

collection is complete, this dataset will be made available via the PANDORA data platform. Preliminary modeling of the data is already offering insights into the causes of megafauna extinction in North America (Fig. 2). The difference between modeled dates for the last presence of megafauna and the first appearance of humans across North America shows that these overlapped between ca. 300 to ca. 2,000 years prior to megafauna extinction. The shortest overlap is observed for Alaska, the entry region for humans into North America. This suggests that human hunting, rather than climate change, was the driving force for megafauna extinction in North America.

PUBLICATIONS

1. *Wang, X., Roberts, P., Tang, Z., Yang, S., Storozum, M., Groß, M. and Fernandes, R., 2021. The circulation of ancient animal resources across the Yellow River Basin: a preliminary Bayesian re-evaluation of Sr Isotope data from the Early Neolithic to the Western Zhou Dynasty. Frontiers in Ecology and Evolution, 9, p.16*

2. *Wilkin, S., Miller, A.V., Miller, B.K., Spengler, R.N., Taylor, W.T., Fernandes, R., Hagan, R.W., Bleasdale, M., Zech, J., Ulziibayar, S. and Myagmar, E., 2020. Economic Diversification supported the growth of Mongolia's Nomadic empires. Scientific reports, 10(1), pp.1–12*

3. *Cubas, M., Lucquin, A., Robson, H.K., Colonese, A.C., Arias, P., Aubry, B., Billard, C., Jan, D., Diniz, M., Fernandes, R. and Valcarce, R.F., 2020. Latitudinal gradient in dairy production with the introduction of farming in Atlantic Europe. Nature communications, 11(1), pp.1–9*

MAX PLANCK INSTITUTE
FOR THE SCIENCE OF HUMAN HISTORY

Max Planck Institute for the Science of Human History / Department of Archaeology The Department of Archaeology is focused on developing and applying the latest multidisciplinary methodologies in archaeology to transform our understanding of the human past. Combining traditional field research with cutting-edge laboratory analyses, it investigates the ways in which culture, biology and ecology have intersected in the past, across the entire spatial and temporal spectrum of human history. By combining insights from both the humanities and natural sciences, the Department of Archaeology will seek to tell a more holistic story of the emergence, expansion, and diversification of our species and its social forms.

Inside the fabric of a voxel: Large–scale computational analysis of cortical microstructure

With the help of the MPCDF, we established a computational pipeline to systematically analyze the three-dimensional microstructure of the human cortex from large-scale imaging data. Using light-sheet microscopy imaging and deep neural network models for automated cell detection, we could quantify the distribution of cell types in large samples of cortical tissue. The combination of three-dimensional imaging techniques and advanced image analysis brings us closer to understanding cortical microstructure and building a bridge towards non-invasive, mesoscale MRI imaging, i.e., *in vivo* histology.

AUTHORS *Nico Scherf, Konstantin Thierbach, Kornelius Podranski, Andreas Marek[12], and Nikolaus Weiskopf*

The brain is arguably one of the most fascinating and mysterious biological systems that we know. Its intricate, connective structure both locally (between neurons) and globally (between brain regions) may hold the key to understanding how it works and why and when it fails. Thus, a critical goal for neuroscience is to map the interconnected brain architecture across these scales: (i) Mapping the brain's local microstructure is essential to understanding the basic principles of information processing, study its development, and build the next generation of noninvasive imaging biomarkers that can help us detect diseases and treat patients early enough. (ii) A statistical map of the variation between microstructure across cortical areas and their correlation to gene expression patterns, development, or neural activity will be a crucial step in understanding the relationship between the brain's structure and function.

To reach this goal, we will ultimately need systematic multi-scale imaging and quantitative analysis. However, there are limits to what we can achieve to date. We can reveal microstructural details only by microscopy, an invasive technique that can only cover small regions of interest. High-resolution magnetic resonance imaging
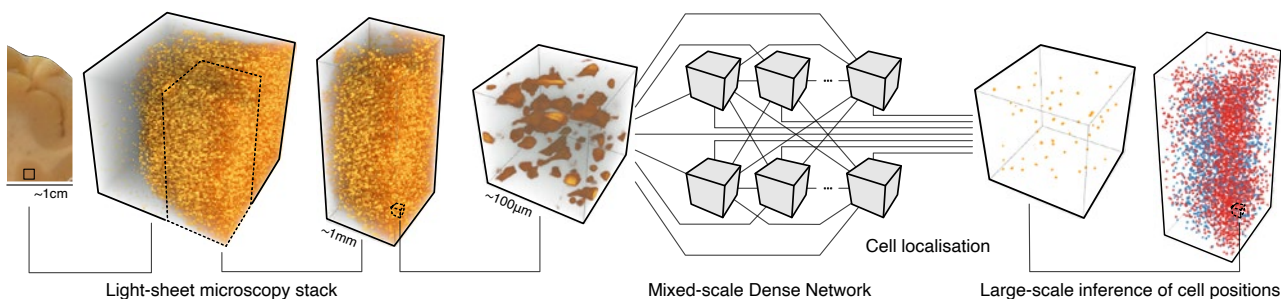

Fig. 1: Schematic of the analysis flow: We image a small piece of cortex (leftmost image) from a tissue sample in a light–sheet microscope in 3D using a fluorescent marker to visualise neurons. We train a neural network to predict cell positions on small patches of the large image stacks. This network then needs to be deployed over the entire image volume to predict the localization of cells, shown in the right–most image with a colour code indicating cell type. (Approximate scales and positions are given as a rough orientation).

(MRI) is ideal for noninvasively visualizing structure and activation across the brain [1]. However, its resolution is still severely limited to a few hundred microns per voxel. Thus, a single voxel can contain a few million cells and a few hundred kilometers of axons [2]: There is an unseen universe of structure within a single pixel on our screens that we cannot account for in this way. Ideally, we need a map between patterns of cells (called the cytoarchitecture [3]), axons (myeloarchitecture [4]) and other microstructural features and the agglomerate signals that we can measure with multi-contrast MRI. Such a bridge would allow us to dive deeper into the fabric of a single MRI voxel and make comprehensive *in vivo* histology a reality [1].

QUANTITATIVE ANALYSIS OF CORTICAL MICROSTRUCTURE
Our current scientific understanding of the brain's microstructure stems from two-dimensional histological sections. A combination of three-dimensional imaging techniques and advanced image analysis would bring us closer to understanding cortical microstructure and build a bridge towards non-invasive, mesoscale MRI imaging. As the first step, we need large-scale, three-dimensional imaging of the cortical architecture at the cellular level. Here, light-sheet microscopy is an ideal and efficient tool due to its fast, parallel imaging. With this data, we can systematically quantify the 3D microarchitecture for the first time. However, the microscope can generate datasets up to the order of Terabytes for an individual experiment. Thus, manually delineating all the structures of interest (e.g., marking the position of each neuron) is infeasible at this scale. Here, we need to rely on computational methods. In particular, we employ deep neural networks trained on sparsely annotated, small patches until they reach the desired accuracy. The final model can then be applied to process the entire dataset in a step called inference. In this part of the project, we trained a 3D multi-scale, convolutional neural network (MSD-net) to predict cellular positions.

THE COMPUTATIONAL CHALLENGE However, with data of this size comes a challenge for deep learning systems. Even deploying multi-scale deep neural networks on large-scale microscopy data hits the limits of current HPC infrastructure. The inference has to be done on a full 3D dataset that is several orders of magnitude larger than the typical examples from computer vision. For this task, our 3D neural network would require about 24TB of RAM. In our collaboration, the AI&HPDA division of the MPCDF, together with Intel, followed two approaches to tackle the computational difficulties: on the one hand, the MPCDF implemented a parallelised version of the inference pipeline, which allowed to distribute the inference step over multiple nodes and GPUs. In the end, a prediction of the neural network could be computed in less than 500 seconds on 16 nodes, each equipped with 2 Nvidia GPUs. In addition, developers at Intel extended the Open Visual Inference and Neural Network Optimization toolkit (Openvino) to 3D convolutional neural networks. By reducing the memory requirements by a factor of 15, these computa-

tions could also be done on a single compute node with 2 TB of main memory in less than one hour. Thus, for the first time, we could establish a feasible computational pipeline – either on multiple compute nodes or on special-purpose nodes with 2 TB of memory – to process the 3D cytoarchitecture of the human cortex from large-scale microscopy stacks. This is the crucial foundation for us to extend the analysis to the processes of neurons (myeloarchitecture) and to ultimately bridge the microscopic and mesoscopic (MRI) level using computational models.

PUBLICATIONS
1. *Weiskopf N, Mohammadi S, Lutti A, Callaghan MF. Advances in MRI-based computational neuroanatomy: from morphometry to in-vivo histology. Curr Opin Neurol. 2015;28: 313–322*
2. *Logothetis NK. What we can do and what we cannot do with fMRI. Nature. 2008;453: 869–878*
3. *von Economo CF, Koskinas GN. Die Cytoarchitektonik der Hirnrinde des erwachsenen Menschen. J. Springer; 1925*
4. *Vogt C. Allgemeine Ergebnisse unserer Hirnforschung I--IV. J Psychol Neurol (Lpz) 25, Erg heft. 1919;1: 279–462*
5. *Morawski M, Kirilina E, Scherf N, Jäger C, Reimann K, Trampel R, et al. Developing 3D microscopy with CLARITY on human brain tissue: Towards a tool for informing and validating MRI-based histology. Neuroimage. 2018;182: 417–428. doi:10.1016/j.neuroimage.2017.11.060*
6. *Thierbach K, Bazin P-L, Back W de, Gavriilidis F, Kirilina E, Jäger C, et al. Combining Deep Learning and Active Contours Opens The Way to Robust, Automated Analysis of Brain Cytoarchitectonics. Machine Learning in Medical Imaging. Springer International Publishing; 2018. pp. 179–187. doi:10.1007/978-3-030-00919-9_21*

MAX PLANCK INSTITUTE
FOR HUMAN COGNITIVE AND BRAIN SCIENCES

Max Planck Institute for Human Cognitive and Brain Sciences / Department of Neurophysics Research at the Max Planck Institute for Human Cognitive and Brain Sciences revolves around human cognitive abilities and cerebral processes, with a focus on the neural basis of brain functions like language, emotions and human social behavior, music and action. At the Department of Neurophysics, our vision is to develop and apply functional microstructure imaging and *in vivo* histology using magnetic resonance imaging (MRI), in order to reliably characterize the detailed functional and anatomical microstructure of the human brain including their interplay and plasticity.

Neuronal network mapping: connectomics

Our brains contain the most complex interaction networks known to date. Using novel three-dimensional electron microscopes and high-performance computing, the complete mapping of neuronal networks in brain tissue has become a reality. First major steps include the readout of *learned* network structure and the analysis of network development over individual and evolutionary time scales.

AUTHOR *Moritz Helmstaedter*

Unlike any other organ, our brains contain extremely densely packed networks of membranous cables that are used by our about 86 billion nerve cells for communication amongst each other. Each nerve cell in the main part of mammalian brains, the so-called cerebral cortex, communicates with about 1,000 other nerve cells via synapses placed along these cables over long distances. Therefore, one expects a total of about 5 million kilometers of wires packed into our skulls – more than 10 times longer than all highways on our planet, in each of our brains. The cables we find in our (and other mammalian) brains are as thin as 50 to 100 nanometers in diameter, about a 1000th of the diameter of our hairs. The resulting cable convolute is of such density and magnitude, that for more than 100 years, researchers have been able to map the connectivity between only a miniscule fraction of neurons in a given piece of brain.

CONNECTOMICS With the development of faster electron microscopic techniques ("three-dimensional EM") and more efficient image analysis routines, the dense mapping of neuronal networks at synaptic resolution has become possible. The novel field of "connectomics" has been pursuing the dense mapping of ever-larger circuits in several species and brain regions.

A FIRST DENSE CONNECTOME FROM THE CEREBRAL CORTEX
As an important first step (Motta et al., 2019), we were able to image and analyze a piece of tissue from the cerebral cortex of a 4-week old mouse, obtained via biopsy from the somatosensory cortex, a part of the cortex occupied with the representation and processing of touch. We applied optimized AI-based image processing and efficient human-machine interaction to analyze all of the about 400,000 synapses and about 2.7 meters of neuronal cable in the volume. With this, we produced a connectome between about 7,000 axons and about 3,700 postsynaptic neurites, yielding a connectome about 26 times larger than the one obtained from the
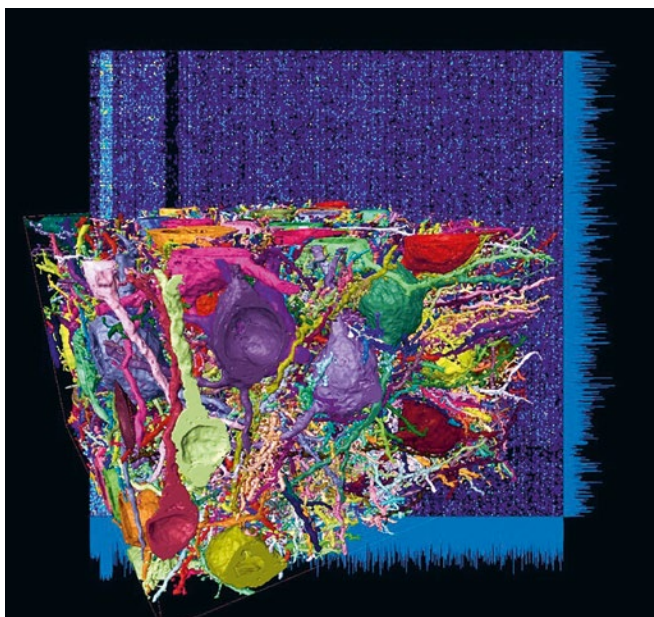

Fig. 1: First connectome from the mammalian cerebral cortex: connectivity matrix between thousands of axons and dendrites in a piece of layer from the somatosensory cortex of a mouse. The reconstruction only required about 4,000 work hours by student annotators, the rest of the work was taken over by improved artificial intelligence. (Motta et al., 2019)
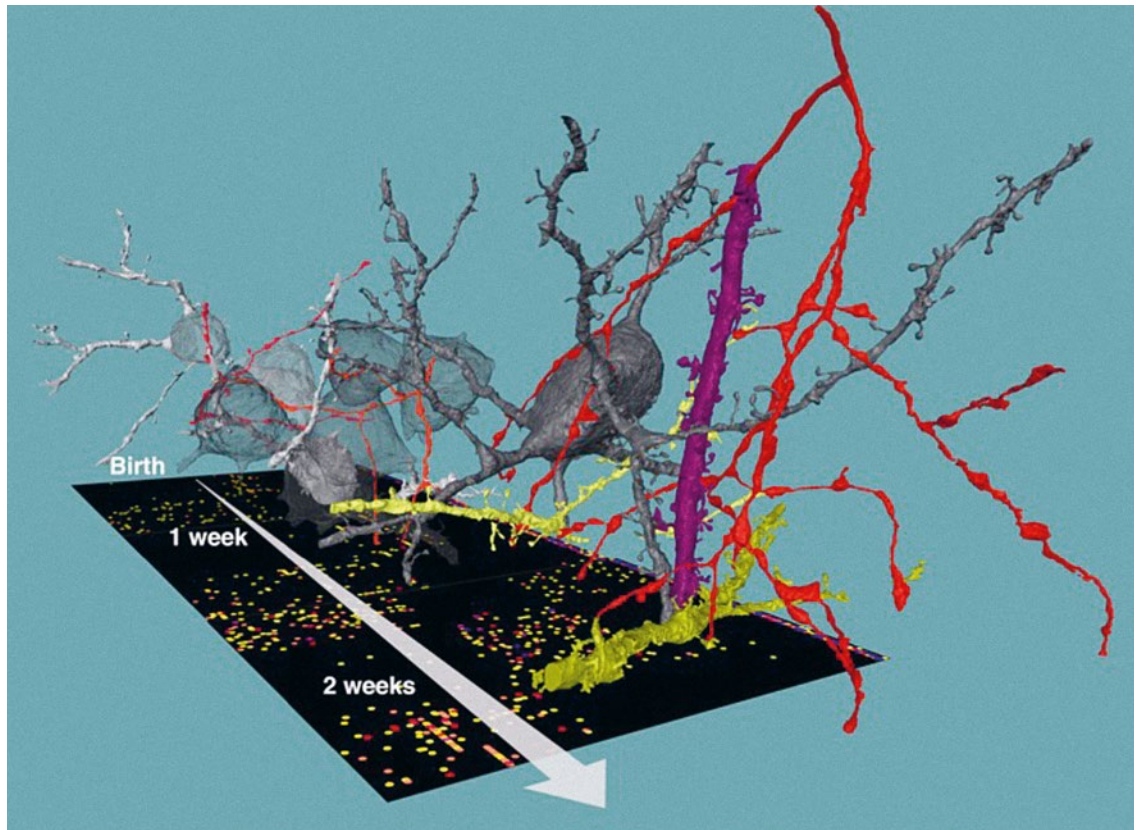
Fig. 2: Connectomes from the developing brain: monitoring the formation of inhibitory circuits from birth to adulthood. (Reprinted with permission from A. Gour et al., 2020 DOI: 10.1126/science.abb4534)

mouse retina more than half a decade ago (Helmstaedter et al., 2013). Importantly, this reconstruction was at the same time larger and about 33-times more efficient than the one applied to the retina, setting a new benchmark for dense connectomic reconstruction in the mammalian brain.

READING TRACES OF LEARNING FROM A NEURONAL CONNECTIVITY MAP Fueled by this methodological breakthrough in connectomics, we analyzed the connectome for the patterns of circuitry present. In particular, we asked what fraction of the circuit showed properties that were consistent with the growth of synapses, mechanisms known to contribute to circuit formation and "learning". We analyzed particular configurations of synapse pairs to study the degree to which they were in agreement with activity-related learning processes ("long-term potentiation"). Because some models of synaptic plasticity make concrete predictions about the increase in synaptic weight when learning, say, to identify a tree or a cat, we were able to extract the imprint of such potential processes even from a static snapshot of the circuit. Since the mouse had had a normal laboratory life until the brain biopsy at 4 weeks of age, one can argue that the degree to which circuits are shaped by learning in "normal" sensory states can be mapped using this approach.

We were surprised how much information and precision is found even in a still relatively small piece of cortex: Especially the extraction of the likely learned circuit fraction was a major eye opener for us.

HOW THE BASIC PROPERTIES OF NEURONAL NETWORKS ARE FORMED AFTER BIRTH How can the precise and convoluted neuronal circuitry found in the cerebral cortex be built in the first place? We know how neurons are born, travel to their location in the gray matter, grow and differentiate. But how and by what rules do the trillions of synapses – the sophisticated contact points via which neurons "talk to teach other" – unfold, often at highly precise locations to form our brain's networks? To address these questions (Gour et al., 2020/21), we analyzed a total of thirteen three-dimensional datasets from the cortex of mice during different stages of development: after birth, at time points comparable to baby, child, teenager and young adult. We used our connectomic methods to map out the neuronal circuitry found in the gray matter of the cerebral cortex, where most of the cerebral synapses are placed. By focusing on synapses of a type of nerve cells called interneurons, which are known to inhibit the activity of other neurons in highly specific ways, we were able to track the development of synaptic partner choice for these particular types of nerve cells.

SYNAPSE GENERATION AND ELIMINATION FOR CIRCUIT FORMATION Surprisingly, different types of interneurons followed very different time courses to establish their favorite synaptic partners. Some were able to innervate their synaptic targets with adult-like preference already in the first investigated circuit stages that correspond to baby brains. This happened immediately when the first chemical synapses were formed in the cortical gray matter. Others showed steep improvements of target choice, which were most likely caused by removal of incorrectly placed synapses. Studies had found before that in some parts of the brain, development not only involves the creation of new synapses but required the removal of synapses as well. The finding that synapse removal (or "pruning") has a precise and highly specific function for inhibitory circuit formation in the cerebral cortex was, however, a major surprise.

These insights were possible in spite of the fact that the mapping of connectomes is a "snapshot" technique: neuronal networks can be measured in biopsies of brain tissue, but cannot be further followed over time in the same piece of brain. Rather, many measurements from different brains need to be made. That we were able to still extract a clear developmental profile from this data illustrates the density of information present in connectomic data.

CONNECTOPATHIES? The developmental processes of neuronal network formation and their possible disruption are thought to be major contributors to some of the main psychiatric disorders, and a particular focus of research has identified a contribution of inhibitory circuits to these dysfunctions. Hence, a precise and detailed understanding of inhibitory circuits is a prerequisite for targeted analysis and possible interference in such disease conditions. We hope to be able to map much more precisely the normal and disrupted network formation in cortical circuits for understanding possible alterations in psychiatric disease, and possibly identify the phenotypes of connectopathies.

The approach used in this work corresponds to "connectomic screening", made possible by the much higher throughput of connectomic methods achieved recently. We expect this approach to become as widely applicable as genetic screening: studying the structure of neuronal networks under a large range of normal and diseased circumstances to understand the alterations and commonalities that are found in mammalian brains.

3D IMAGE ANALYSIS AS THE KEY CHALLENGE In all of the work described here, we are making use of the excellent high-performance computing and data storage infrastructure available at the MPCDF. Connectomic datasets are currently at the scale of petabytes per dataset, and their analysis is only possible by a combination of several machine-learning techniques, together with efficient human-machine interaction via in-browser data annotation software (webknossos.org, Boergens et al., 2017). Thanks to the massive support by MPCDF, we were able to push the benchmark of connectomic data analysis (Berning et al., 2015; Staffler et al., 2017; Motta et al., 2019). Next key goals are the connectome of a so-called cortical column, ideally mapped in species like mouse, non-human primate and human, to learn about the structure and evolutionary change of this critical neuronal circuitry. This work has the potential to also allow insights into why biological "computers" are still far outperforming artificial intelligence in the efficient usage of labeled data and energy (Helmstaedter, 2015).

PUBLICATIONS

1. Gour A, Boergens KM, Heike N, Hua Y, Laserstein P, Song K, Helmstaedter M (2020/21) Postnatal connectomic development of inhibition in mouse barrel cortex. Science 371 (6528)

2. Motta A, Berning M, Boergens KM, Staffler B, Beining M, Loomba S, Schramm C, Hennig P, Wissler H, Helmstaedter M (2019) Dense connectomic reconstruction in layer 4 of the somatosensory cortex. Science 366 (6469) doi: 10.1126/science. aay3134 (Cover article)

3. Schmidt H, Gour A, Straehle J, Boergens KM, Brecht M, Helmstaedter M (2017) Axonal synapse sorting in medial entorhinal cortex. Nature 28;549(7673):469–475

4. Boergens KM, Berning M, Bocklisch T, Bräunlein D, Drawitsch F, Frohnhofen J, Herold T, Otto P, Rzepka N, Werkmeister DW, Wiese G, Wissler H, Helmstaedter M (2017) webKnossos: Efficient Online 3D Data Annotation for Connectomics. Nature Methods 14(7): 691–694

5. Staffler B, Berning M, Boergens KM, Gour A, van der Smagt P, Helmstaedter M (2017) SynEM, automated synapse detection for connectomics. eLife, doi: 10.755/elife.26414

6. Berning M, Boergens KM, Helmstaedter M (2015) SegEM: Efficient Image Analysis for High–Resolution Connectomics. Neuron 87(6):1193–206 (Cover article)

7. Helmstaedter M, (2015) The mutual inspirations of machine learning and neuroscience. Neuron 86(1):25–28

8. Helmstaedter M, Briggman KL, Turaga S, Jain V, Seung HS, Denk W (2013) Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature 500: 168–174

Max Planck Institute for Brain Research / Department of Connectomics
Our goal is to make connectomics a high-throughput screening technique for neuroscience, applicable to the large circuits in the mammalian cerebral cortex. We are using connectomic data from the cerebral cortex to analyze brain-implemented algorithms, extract possible imprints of sensory experience, investigate the postnatal development of neuronal circuits and study connectome alterations in models of psychiatric disease.

July 2021

D2 · MPCDF

· Max Planck Computing and Data Facility

Ausbildungszentrum / Desktop Support

Arnulf-Schlüter-Hörsaal ·

Abteilung Zentrale Sicherheit

Numerische Methoden der Plasmaphysik ·