Working Group on Physics, Modern IT and Artificial Intelligence
Arbeitskreis Physik, moderne Informationstechnologie und Künstliche Intelligenz (AKPIK)

Tim Ruhe
TU Dortmund
Otto Hahn-Straße 4a
44227 Dortmund
tim.ruhe@tu-dortmund.de

Overview of Invited Talks and Sessions

(Lecture halls KS 00.003 and KS H C)

Invited Talks

AKPIK 3.1	Tue	11:00–11:30	KS H C	**Large Language Models for Research Data Access** — •Jutta Schnabel
AKPIK 3.2	Tue	11:30–12:00	KS H C	**Deep Learning-Based Imaging of MeerKAT Observations** — •Kevin Schmitz
AKPIK 3.3	Tue	12:00–12:30	KS H C	**Towards FAIR fundamental physics: PUNCH4NFDI approach** — •Victoria Tokareva, Ivan Knežević, Harry Enke, Andreas Haungs

Sessions

AKPIK 1.1–1.1	Sun	17:30–19:00	AM 00.014	**jDPG/AKPIK Programmierworkshop (joint session AKjDPG/AKPIK)**
AKPIK 2.1–2.8	Mon	16:00–18:00	KS 00.003	**Parallel Talks**
AKPIK 3.1–3.3	Tue	11:00–12:30	KS H C	**Invited Talks**
AKPIK 4.1–4.4	Wed	13:45–15:45	KH 00.011	**Artificial Intelligence in Scientific Publishing (joint session AGI/AKjDPG/AKPIK)**

AKPIK 1: jDPG/AKPIK Programmierworkshop (joint session AKjDPG/AKPIK)

Time: Sunday 17:30–19:00 Location: AM 00.014

Tutorial AKPIK 1.1 Sun 17:30 AM 00.014

Resource-Aware Deep Learning: Tracking Energy Consumption in Scientific AI Applications — •Kevin Schmitz[1], Anno Knierim[1], and Raphael Fischer[2] — [1]TU Dortmund University, Dortumd, Germany — [2]Lamarr Institute for ML & AI, Dortmund, Germany

Deep learning has become an indispensable tool across physics and astronomy, yet its growing computational demands increasingly raise questions about energy efficiency and sustainability. This tutorial introduces physicists to the principles of resource-aware deep learning, focusing on practically quantifying, understanding, and optimizing the energy consumption of deep learning models. We begin by outlining different approaches for tracking model power usage, ranging from static estimation methods to dynamic profiling tools validated against ground-truth measurements, and demonstrate how these concepts are implemented in practice using the Lamarr Energy Tracker, developed at the Lamarr Institute for Machine Learning and Artificial Intelligence, for straightforward monitoring of GPU and CPU utilization during training or inference. Finally, we show how resource metrics can be visualized together with reconstruction accuracy using an example from radio interferometric imaging, where super-resolution neural networks reconstruct astrophysical sources from sparse visibility data. The session provides conceptual foundations and practical guidance for integrating sustainability into scientific machine learning workflows, empowering researchers to balance predictive performance with environmental responsibility.

AKPIK 2: Parallel Talks

Time: Monday 16:00–18:00 Location: KS 00.003

AKPIK 2.1 Mon 16:00 KS 00.003

Metadata Analysis for the Square Kilometre Array Max Planck Institute Telescope — •Felix Wersig and Luca Davide Di Bella — TU Dortmund University

The Square Kilometre Array Max Planck Institute (SKAMPI) telescope has been operating since 2019 at the Karoo Radio Astronomy Reserve, collecting metadata from the receiver, the antenna control unit and the MeerKAT weather sensors for many observations under diverse conditions. Searching for unusual patterns in the metadata could be a key method for assessing data quality and identifying maintenance needs.

The SKAMPI telescope is designed as a 15 m prototype dish for MeerKAT+ and the SKA-Mid component of the Square Kilometre Array Observatory. The SKA-Mid component is going to consist of 197 parabolic dishes incorporating the MeerKAT+ radio telescope. Its layout will consist of a $\sim$1 m central core and antennas radiating out from it up to a diameter of 150 km.

Ensuring that a large fraction of the antennas remain operational at all times is a key challenge given the large number of antennas and extensive distances present in the planned layout. Leveraging the large dataset collected by the SKAMPI telescope, a preliminary tool for quick analysis of the collected metadata from a dish is presented.

AKPIK 2.2 Mon 16:15 KS 00.003

Calibration in radio interferometer simulations using pyvisgen — •Sadiah Azeem[1], Kevin Schmitz[1,2], Anno Knierim[1,2], and Christian Arauner[1,2] — [1]TU Dortmund University, Dortmund, Germany — [2]Lamarr Institute for Machine Learning and Artificial Intelligence, Dortmund, Germany

Modern physics experiments greatly benefit from the ability to reproduce entire measurement processes through realistic simulations. In radio interferometry, such simulations enable testing of calibration and data analysis pipelines and allow estimation of the discovery potential for faint astrophysical sources. `pyvisgen` is a Python framework designed to generate radio interferometric visibilities based on the radio interferometer measurement equation (RIME) and is a part of the `radionets-project`, which uses deep learning based methods for radio interferometric imaging. Realistic simulations must include instrumental and atmospheric effects that corrupt the measured visibilities such as antenna gains and tropospheric delays. They also provide the ground truth needed to test and evaluate calibration methods. This talk will outline how to generate RIME-consistent corrupted visibilities and apply them to the development of calibration methods.

AKPIK 2.3 Mon 16:30 KS 00.003

Simulating Protoplanetary Disks as Training Data for Neural Network based Reconstruction of Radio Interferometer Measurements — •Tom Gross[1,2], Kevin Schmitz[1,2], Christian Arauner[1,2], and Anno Knierim[1,2] — [1]TU Dortmund University — [2]Lamarr Institute for Machine Learning and Artificial Intelligence

Understanding the evolution of planetary systems is an important research field in modern astronomy. Planets form through accretion processes of dust and gas in protoplanetary disks around young stars.

The thermal radiation of the disks is measurable in the radio spectrum using radio interferometers. These enable the observation of small structures but only measure samples of an incomplete Fourier space. This intrinsic limitation results in noisy images, unsuitable for physical analyses.

The `radionets-project` uses a modern deep-learning framework to reconstruct the missing information. To enhance the scientific interpretation of the reconstructions, the training data needs to represent realistic source distributions. The software `FARGO3D` simulates the distribution of dust and gas in a protoplanetary disk using a hydrodynamical model. From these distributions, the thermal radiation can be simulated using Monte Carlo methods. This presentation shows the results of the simulations and how they are integrated into the data pipeline for the deep-learning based reconstruction.

AKPIK 2.4 Mon 16:45 KS 00.003

ParaO: In situ reproducible workflows through user-centric design — Martin Erdmann, •Benjamin Fischer, and Felix Zinn — III. Physikalisches Institut A, RWTH Aachen University

In High Energy Physics (HEP) and other fields, data analyses require many different steps to produce a final result. As such, it is sensible to leverage workflow tools to codify these complex structures of tasks and their dependencies. Of course, such intricate analyses are not created ad hoc; rather, their development is a lengthy and iterative process.

Here, ParaO's design empowers users to rapidly grow and readily adapt their work while avoiding verbose boilerplate. For example, integrating a new parameter or dependency in the middle of a ten-step workflow needs just a single line of code. There, ParaO automatically propagates the resulting knock-on effects, such as changed implicit parameterization of later tasks—unlike Luigi, which requires a more manual treatment.

Thus, ParaO enables in situ usage of workflows and their benefits, such as bookkeeping and reproducibility, even in early and chaotic stages of development. We demonstrate ParaO's principles and features in the context of a comprehensive top-Higgs analysis.

AKPIK 2.5 Mon 17:00 KS 00.003

Cleaning MOJAVE observations with neural networks — •Christian Arauner[1,2], Kevin Schmitz[1,2], Anno Knierim[1,2], and Tom Gross[1,2] — [1]TU Dortmund University — [2]Lamarr Institute for Machine Learning and Artificial Intelligence

Radio interferometry enables observations of astronomical objects with high angular resolution. However, the inherent incompleteness of the (u, v)-plane sampling leads to significant noise and artifacts in the reconstructed images. Existing state-of-the-art cleaning algorithms effectively mitigate these effects but are computationally demanding, not easily reproducible and not readily scalable to the data volumes anticipated from next-generation radio telescopes. As a promising alternative, neural network-based approaches offer the potential to automate and accelerate the image reconstruction process.

The Python package `radionets` implements a deep-learning framework for the reconstruction of calibrated data from radio observations. In this approach, ResNets are used to reconstruct the missing values directly in the (u, v)-plane. The validity of the neural network-based approach has been shown with simulated data, now it is possible to apply it to real measurments from the MOJAVE program. The observations of the MOJAVE archive are particular suitable for the development and testing of these neural networks as it comprises a large data set of high-quality data, measured over a long period of time.

In this talk, I will present first observations of the MOJAVE program, reconstructed with `radionets`.

AKPIK 2.6 Mon 17:15 KS 00.003

Methodological Preparation for Measurements with the DSA-2000: Simulation, Analysis, and Reconstruction Using Neural Networks — •Jonas Grossmann[1], Kevin Schmitz[1,2], Anno Knierim[1,2], Christian Arauner[1,2], and Tom Gross[1,2] — [1]TU Dortmund University, Dortmund, Germany — [2]Lamarr Institute for Machine Learning and Artificial Intelligence, Dortmund, Germany

A new generation of radio telescopes is set to begin operations later this decade. Among these is the Deep Synoptic Array 2000 (DSA-2000), a radio survey interferometer. It comprises 2000 small antennas, each 5 m in diameter, distributed over an area of 160 square kilometres in the Nevada Desert, USA. This configuration is optimised for rapid, high-resolution surveys of the entire northern sky. The DSA-2000 will achieve at least an order of magnitude improvement in angular resolution, enabling the discovery of billions of previously unknown radio sources. The resulting data volumes are unprecedented, surpassing the capabilities of conventional analysis methods and necessitating new approaches for translating interferometric measurements into cleaned images. The `radionets-project` group at TU Dortmund University aims to develop neural networks that, could automatically reconstruct dirty radio images. Training these networks requires large, realistic datasets. Given that all 2000 antennas cannot operate flawlessly at all times, the goal is to train the networks to deliver stable and reliable reconstructions even when subsets of antennas are inactive. This talk outlines the framework in `radionets-project` required to create the datasets necessary for training these neural networks.

AKPIK 2.7 Mon 17:30 KS 00.003

Resource Awareness in Deep Learning-based Imaging in Radio Interferometry — •Anno Knierim[1,2], Kevin Schmitz[1,2], Christian Arauner[1,2], Raphael Fischer[1,2], Dominik Elsässer[1,2], and Wolfgang Rhode[1,2] — [1]TU Dortmund University, Dortmund, Germany — [2]Lamarr Institute for Machine Learning and Artificial Intelligence, Dortmund, Germany

Recent approaches in radio interferometry aim to improve image cleaning of measurements using machine learning techniques. Reconstructing sources using these novel techniques has the advantage of being agnostic to the initial parameters used in traditional cleaning algorithms.

The `radionets-project` is a multi-software environment developed at TU Dortmund University. The main deep-learning framework, `radionets`, reconstructs calibrated data from radio observations using convolutional neural networks (CNNs). The framework aims to achieve a high dynamic range and produce high-resolution source images.

Due to their high complexity, applying deep learning models sustainably requires balancing predictive performance with resource consumption. The environmental impacts extend well beyond training, as model complexity also affects the efficiency during data simulation and inference. This talk presents ongoing efforts to track and estimate the environmental impact of our models using `CodeCarbon`, `PyTorch Lightning`, and `MLFlow`.

AKPIK 2.8 Mon 17:45 KS 00.003

FAIR-Driven Retrospective Metadata Enrichment for KCDC Digital Resources — •Victoria Tokareva — Karlsruhe Institute of Technology, Institute for Astroparticle Physics, 76021 Karlsruhe, Germany

The KASCADE Cosmic-Ray Data Centre (KCDC) is an established open data archive in astroparticle physics with more than a decade of experience in long-term data preservation. It aims to apply modern technologies to maximise the accessibility, usability, and scientific value of research data for a broad audience, from professional researchers to students, citizen scientists, and science enthusiasts, while advancing the use of FAIR (Findable, Accessible, Interoperable, Reusable) practices. Operating within a data-intensive high-energy physics environment, we address community-wide challenges such as the large data volumes produced in astroparticle physics and the complexity of formats like ROOT and HDF5, which require domain-specific software and specialised approaches for machine-actionable metadata extraction. To support interdisciplinary discovery and integration into infrastructures such as PUNCH4NFDI, we extend established metadata with additional discovery-oriented elements and references, including domain knowledge historically dispersed in scientific publications. In this contribution, we present an enhanced metadata model for KCDC digital resources featuring expanded use of persistent identifiers and increased semantic connectivity. We also demonstrate how natural language processing and large language models support retrospective metadata enrichment for archival KASCADE datasets.

AKPIK 3: Invited Talks

Time: Tuesday 11:00–12:30 Location: KS H C

Invited Talk AKPIK 3.1 Tue 11:00 KS H C

Large Language Models for Research Data Access — •Jutta Schnabel for the KM3NET-ERLANGEN-Collaboration — ECAP, FAU Erlangen-Nürnberg

In a strive to provide ever more interoperable, complex scientific data to a wide range of users across scientific domains, new approaches to lower the boundaries for the access to research data are called for. To address the collection, reduction and analysis of data, as well as its storage, sharing and finding, Large Language Models (LLMs) offer an intriguing opportunity. They can facilitate the transition in research data management from Big Data to Smart Data by increasing the effectiveness of research workflows and the efficiency of data discovery, curation, and reuse especially in the physics domain. This contribution will introduce the uptake of LLMs for data management as pursued in the Physics-LLM project in ErUMData, starting from the application of LLMs in the KM3NeT collaboration.

Invited Talk AKPIK 3.2 Tue 11:30 KS H C

Deep Learning-Based Imaging of MeerKAT Observations — •Kevin Schmitz — TU Dortmund University, Dortumd, Germany — Lamarr Institute for ML & AI, Dortmund, Germany

Modern radio interferometers such as MeerKAT produce vast amounts of data that enable high-resolution imaging of astrophysical sources but also pose significant challenges for image reconstruction. Classical deconvolution algorithms like CLEAN struggle with noise artifacts and limited scalability. The deep-learning framework radionets addresses these issues by combining simulated observations with neural-network-based image reconstruction.

Using the Radio Interferometer Measurement Equation (RIME), realistic simulations model signal propagation and systematic effects in the measured data. These synthetic datasets allow convolutional neural networks to be trained under controlled conditions, enabling quantitative studies of image quality, uncertainty estimates, and computational efficiency.

The framework reconstructs sparse data directly in Fourier space before transforming them into clean sky images, achieving higher positional accuracy and flux recovery than standard tools, especially for wide-field observations containing both compact sources and diffuse emission. I will present its application to real MeerKAT observations, compare the results with classical methods, and illustrate how physics-guided machine learning improves data quality and interpretation in radio astronomy.

Invited Talk AKPIK 3.3 Tue 12:00 KS H C

Towards FAIR fundamental physics: PUNCH4NFDI approach — •Victoria Tokareva[1], Ivan Knežević[2], Harry Enke[3], and Andreas Haungs[1] — [1]Karlsruhe Institute of Technology, Karlsruhe, Germany — [2]GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt, Germany — [3]Leibniz-Institute for Astrophysics,

Potsdam, Germany

Research data management is becoming increasingly important in data-intensive fundamental physics, including astroparticle, particle, nuclear and astrophysics. Growing data volumes and complexity, increasing interdisciplinarity, high costs of data acquisition and processing, and rising expectations for open science and AI-ready digital research outputs are driving these communities toward FAIR (Findable, Accessible, Interoperable, Reusable) data management practices. Key efforts in this direction include reuse and practical harmonisation of semantic artefacts such as metadata schemas, thesauri and controlled vocabularies; making explicit expert knowledge embedded in collaboration-internal data transformations, specialised software tools, and internal data formats (e.g., FITS, ROOT, HDF5); and defining minimal requirements for reusable digital research objects. This talk presents the multi-layer metadata model and the concept and the use-case-based implementations of Digital Research Product, developed within the PUNCH4NFDI (Particles, Universe, NuClei and Hadrons for the German National Research Data Infrastructure, punch4nfdi.de) consortium in the context of a modern FAIR Data landscape and as core elements of the PUNCH4NFDI Science Data Platform.

AKPIK 4: Artificial Intelligence in Scientific Publishing (joint session AGI/AKjDPG/AKPIK)

Time: Wednesday 13:45–15:45 Location: KH 00.011

Invited Talk AKPIK 4.1 Wed 13:45 KH 00.011

Scientific publishing in the era of AI — •Dominik Elsässer — TU Dortmund, Department of Physics

The dissemination of results to colleagues and to the general public has been an indispensable part of the scientific process for as long as humankind has expanded our treasure of knowledge with scientific methods. The forms and the means by which this dissemination happens have however been subject to fundamental changes over time. One such change surely was that many fields of physics have moved towards working in large and international collaborations. And a more recent, yet profound change is the emergence of AI systems, which can be used to support many steps of the publication process, but also pose new challenges. In this talk, I will present key steps of the publication process typically encountered in fundamental physics and adjacent areas, and discuss options, methods, and tools available to the publishing scientist in the era of AI, with a specific focus on modern Large Language Models (LLMs), and on systems based on LLMs. While a focus will be on publication in peer-reviewed journals, there will also be a discussion of other forms of publication.

Invited Talk AKPIK 4.2 Wed 14:15 KH 00.011

Intelligence and the Art of Scientific Publishing - an Editor's Perspective — •Andreas Buchleitner — Physikalisches Institut, Albert-Ludwigs-Universität Freiburg

Artificial intelligence (AI) shatters fundamental rules of the scientific publishing process. It induces a certain level of perplexity and disorientation within the academic realm, on how to react and/or to adapt - at a speed which is dictated from the outside, by a rapidly developing technology, together with the economic traction which comes with it. Different stakeholders of the publication process suggest distinct remedies, certainly guided by their respective perspectives, levels of expertise, and interests. And, clearly, there are highly nonlinear interdependencies between the thus reorganizing publication process, standards of good scientific practice, and the - in many respects highly disputable - incentives which constrain science and, in particular, academic careers. After stating some of the immediate challenges AI poses to the inner workings of the editorial process, the talk will expand upon the above interdependencies, and contemplate the genuine role and responsibility of the scientific community in shaping them.

Invited Talk AKPIK 4.3 Wed 14:35 KH 00.011

Prompt or perish - the research life cycle in times of genAI — •Sandra Geisler — RWTH Aachen University, Aachen, Deutschland

Large Language Models (LLMs) are rapidly reshaping how research is conducted and communicated. In this talk we will explore and spark discussions about where generative AI could or already does add value along the research life cycle, as well as the limitations and risks that must be carefully considered. From brainstorming and literature discovery to FAIR-compliant research data management, science communication and the review process, LLMs offer powerful new opportunities for researchers. At the same time, researchers face significant uncertainty, ethical concerns, and policy gaps. Drawing on recent studies and practical examples from our own research projects, this presentation highlights both the promise and the perils of an AI-assisted research life cycle.

Discussion AKPIK 4.4 Wed 15:05 KH 00.011

Discussion — Andreas Buchleitner[1], Dominik Elsässer[2], Sandra Geisler[3], •Uwe Kahlert[3], Simon Neuhaus[4], and •Tim Ruhe[2] — [1]Physikalisches Institut, Albert-Ludwigs-Universität Freiburg — [2]TU Dortmund, Department of Physics — [3]RWTH Aachen University — [4]Bergische Universität Wuppertal

Scientific publishing has long been said to be in crisis. "Publish or perish" is too often the prevailing motto. AI tools can now be used at almost every stage of the process. What risks and perhaps also opportunities does this development present? We will discuss this with the speakers of this session.