# T 25: Data Analysis, Information Technology and Artificial Intelligence

Time: Monday 16:15–18:25 Location: T-H38

**Group Report** T 25.1 Mon 16:15 T-H38

**CROWN - A Software framework for fast analysis ntuple production** — •Sebastian Brommer[1], Markus Klute[1], Nikita Shadskiy[1], Guenter Quast[1], Roger Wolf[1], and Sebastian Wozniewski[2] — [1]Karlsruhe Institute of Technology — [2]Universität Göttingen

With the ever-increasing data recorded by the LHC experiments, new software solutions are required to handle the rising demand in computational power and to ensure fast and efficient processing of the data. The CROWN framework is designed to provide such a fast and robust solution for the conversion of stuctured event data into flat ntuples for histogramming and further analysis. Within the framework, code generation is used to create compiled C++ executables based on ROOT data frames, ensuring fast data processing with minimal dependencies.

In this talk, the core concepts of the framework as well as performance comparisons to existing solutions are presented.

**Group Report** T 25.2 Mon 16:35 T-H38

**Verringerung systematischer Unsicherheiten durch systematics-aware training** — Markus Klute, Günter Quast, •Lars Sowa, Roger Wolf und Stefan Wunsch — Karlsruhe Institute of Technology (KIT)

Eine Aufgabe für Analysen in der Hochenergiephysik besteht in der Trennung von Signal und Untergrundereignissen. Durch statistische Anpassung an Daten werden mit Hilfe dieser Trennung Fitparameter und deren Unsicherheiten, die die Genauigkeit der Fitparameter quantifizieren, bestimmt. Um statistische Unsicherheiten dieser Fitparameter zu minimieren, nehmen moderne Teilchenbeschleuniger enorme Datenmengen auf. Infolgedessen treten systematische Unsicherheiten verstärkt in den Vordergrund und Methoden zu deren Unterdrückung gewinnen für Analysen zunehmend an Wichtigkeit.

Dieser Vortrag präsentiert Studien zur Verringerung systematischer Unsicherheiten. Dabei wird eine diagnostische, auf Taylorkoeffizienten basierende Methode verwendet, um den Einfluss systematischer Variationen der Eingangsparameter auf die Ausgabefunktion eines Neuronalen Netzes zu untersuchen. Darauf aufbauend werden erprobte Methoden für systematics-aware training erläutert und vielversprechende, zur Umsetzung geplante Methoden vorgestellt.

T 25.3 Mon 16:55 T-H38

**Understanding Event-Generation Networks via Uncertainties** — Marco Bellagente[1], Manuel Haussmann[2], •Michel Luchmann[3], and Michel Luchmann[4] — [1]Institut für Theoretische Physik, Universität Heidelberg, Germany — [2]Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany — [3]Institut für Theoretische Physik, Universität Heidelberg, Germany — [4]Institut für Theoretische Physik, Universität Heidelberg, Germany

Following the growing success of generative neural networks in LHC simulations, the crucial question is how to control the networks and assign uncertainties to their event output. We show how Bayesian normalizing flows or invertible networks capture uncertainties from the training and turn them into an uncertainty on the event weight. Fundamentally, the interplay between density and uncertainty estimates indicates that these networks learn functions in analogy to parameter fits rather than binned event counts.

T 25.4 Mon 17:10 T-H38

**Evaluating Uncertainties in Measurements of the Production of a Single Top-Quark in Association with a Photon with Bayesian Neural Networks** — Johannes Erdmann[1], Burim Ramosaj[2], and •Daniel Wall[1] — [1]TU Dortmund University, Department of Physics — [2]TU Dortmund University, Department of Statistics

Multivariate approaches including neural networks constitute powerful and established methods in experimental particle physics. However, using these methods, it is difficult to account for uncertainties from statistical and systematic sources in a consistent and efficient way. By employing weight distributions instead of fixed weights and by utilising the process of Bayesian inference, Bayesian Neural Networks not only suffer significantly less from overfitting, but also allow to obtain an uncertainty estimate on the output.

These characteristics are of particular interest in measurements of processes suffering from limited statistics and challenging signal-to-background ratios. The analysis of top-quark production in association with a photon ($tq\gamma$), probing the structure of the electroweak couplings of the top quark, is one of such processes, as the corresponding cross section is considerably lower than those of relevant background processes, most importantly top-quark pair production ($t\bar{t}\gamma$).

In this talk, studies of Bayesian Neural Networks for their application in the classification of top-quark processes in association with a photon are presented.

T 25.5 Mon 17:25 T-H38

**Non-parametric background models for axion haloscopes** — •Johannes Diehl[1], Jakob Knollmüller[2], and Oliver Schulz[1] — [1]Max Planck Institute for Physics, Munich, Germany — [2]Max Planck Institute for Astrophysics, Munich, Germany

Axions have been introduced to solve the strong CP problem of the standard model of particle physics and turned out to be an excellent candidate to explain cold dark matter. "Haloscopes" are searching world wide for axions from the galactic dark matter halo, mostly by axion conversion to photons at radio frequencies in a strong B-field. Finding an axion signal in haloscope data means finding a small peak in a vast non-uniform RF background. One crucial challenge is therefore to selectively suppress larger frequency scales while inducing as little attenuation and correlation as possible at smaller frequency scales. This has so far been tackled using filter theory, e.g. through Savitzky-Golay filters for the HAYSTAC experiment, but proof that this is the optimal filter to use is still lacking. Using simulated data from the MADMAX haloscope, I present a novel machine-learning based approach to separate scales and subtract the background without attenuating the signal which lends itself well to being incorporated into a final Bayesian analysis.

T 25.6 Mon 17:40 T-H38

**Open Science in KM3NeT** — •Jutta Schnabel for the ANTARES-KM3NET-ERLANGEN-Collaboration — Erlangen Centre for Astroparticle Physics, FAU Erlangen-Nürnberg

The KM3NeT neutrino detectors are currently under construction at two locations in the Mediterranean Sea, with a first taking of data from high-energy neutrino interactions already under way. This scientific data is valuable both for the astrophysics and neutrino physics communities as well as for marine biologists. In order to facilitate FAIR data sharing of the research results, the KM3NeT collaboration is actively working towards an open science infrastructure to provide high-level scientific data, software, and analysis pipelines in an interoperable research environment suited both for research and education. This contribution introduces the open science program of KM3NeT and gives an overview of its current architecture and implementation.

T 25.7 Mon 17:55 T-H38

**Columnar data analysis with ATLAS analysis formats** — •Nikolai Hartmann — Ludwig-Maximilians-Universität München

Future analysis of ATLAS data will involve new small-sized analysis formats to cope with the increased storage needs. The smallest of these, named DAOD_PHYSLITE, has calibrations already applied to allow fast downstream analysis and avoid the need for further analysis-specific intermediate formats. This allows for application of the "columnar analysis" paradigm where operations are applied on a per-array instead of a per-event basis. This presentation shows the latest developments of tools within the scientific python ecosystem and discusses a prototype analysis for testing both on single Machines as well as Analysis Facilities or similar scale-out systems.

T 25.8 Mon 18:10 T-H38

**Information visualization platform for data quality monitoring of CMS tracker** — •Abhit Patil — Ruprecht-Karls-Universität Heidelberg, Germany

The tracker of the CMS detector consists of silicon sensors arranged in concentric cylinders and endcap disks to track muons, which requires continuous monitoring during operation and certification of the recorded data for physics analysis. The process relies on shifters who assess the data quality by comparing data distributions with references. This challenging task requires examining possible types of failures with

expert-based rule systems and manual profiling of a large number of histograms. To assess the quality of data volumes with finer granularity and to improve the quality of the data certification, this work proposes to augment the monitoring process with information visualization based methods, which aims to pre-process large amount of multidimensional data during the data taking period and provide a visual abstraction of the data quality and provide hints for potential anomalies. The visualisation methods are deployed on a platform built using Python-Django framework and Postgres database.