EuCAIFCon 2025

Monday, 16 June 2025 - Friday, 20 June 2025

THotel, Cagliari, Sardinia, Italy

Book of Abstracts

Contents

1

ARDE: Neural network-based algorithms for discrimination between electrons and γ-rays

Author: Naomi Marchini[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: marchini@fi.infn.it

The results of the ARDE project will be presented, aiming to develop innovative algorithms based on neural network architectures to discriminate between signals induced by electrons and γ-rays in semiconductor detectors, specifically in Si(Li) and HPGe. The algorithm performances for internal conversion electron spectroscopy measurements in an energy range from ~300 keV to ~1-2 MeV will be investigated. Using techniques based on artificial intelligence and machine learning enables the simultaneous analysis of all the information of the signal shape, rather than relying on a correlation between two parameters as in traditional PSA techniques. Thanks to ARDE, the instrumentation used for internal conversion electron spectroscopy measurements will be simplified, moving away from the current reliance on magnetic γ-ray filters. These filters cause significant technical issues during measurements, such as making detection efficiency highly dependent on the energy of the electrons. Furthermore, the techniques developed in this project will provide the foundation for other applications, such as those related to the search for rare events (e.g., 0νββ decay) and medical applications, where measuring β-radiation doses and energy in the presence of γ-radiation background is crucial.

AI keywords:

PSA, Analysis

2

AI for cosmic ray direct detection in space with the DAMPE mission

Author: Andrii Tykhonov[None]

Corresponding Author: andrii.tykhonov@cern.ch

DArk Matter Particle Explorer (DAMPE) is a pioneering instrument launched in space in 2015, designed for precise cosmic ray measurements reaching unprecedented hundreds of TeV in energy. One of the key challenges with DAMPE lies in cosmic ray data analysis at such high energies. It has been shown recently that deep learning can boost the experiment precision in regression (particle reconstruction) and classification (particle identification) tasks, in some cases replacing conventional techniques such as Kalman-based track finding. The new deep learning pipeline of DAMPE allowed to extend the energy reach of its measurements and enabled a non-trivial enhancement in accuracy. In this talk, we will present the AI methods used in DAMPE emphasizing their impact for the science performed with the mission.

AI keywords:

CNNs; offline data reconstruction; pattern recognition; regression; classification

3

Adversarial Machine Learning for Robust Event Classification in Particle Physics

Author: Davide Brundu[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: davide.brundu@ca.infn.it

Adversarial machine learning is a collection of techniques used to study attacks on machine learning algorithms. It is commonly used in cybersecurity and still has few applications in fundamental physics. Since systematic effects in detector data, often absent in Monte Carlo simulations, challenge the performance of machine learning models in particle physics, in this work we model the detector as an "adversary", either as an "enemy" introducing realistic perturbations to expose model vulnerabilities or as a "friend" simulating systematic effects during training. This dual approach forces classifiers to learn invariant features, improving robustness and accuracy on real data. Applied to simulated detector events, our method demonstrates superior generalization compared to traditional approaches, addressing systematic uncertainties in experimental science.

AI keywords:

simulation-based inference; adversarial machine learning

4

A(i)DAPT Program

Author: Tommaso Vittorini[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: tommaso.vittorini@ge.infn.it

A(i)DAPT is a program which aims to utilize AI techniques, in particular generative modeling, to support Nuclear and High Energy Physics experiments. Its purpose is to extract physics directly from data in the most complete manner possible. Generative models such GANs are employed to capture the full correlations between particles in the final state of nuclear reactions. This many-fold program will allow us to to achieve various goals including accurately fitting data in a multidimensional space and unfolding detector effects to minimize their impact on the relevant physics. Moreover, it will enable us to store a large amount of realistic-like data in an extremely compact format and to extract reaction amplitudes in an alternative way. We aim at incorporating universality of scattering amplitudes, training networks with different kinematics of the same final state or different final states to recover the underlying physics. As of today, we've conducted a positive closure test on inclusive electron scattering, demonstrating that generative models are able to reproduce $2-\pi$ photoproduction data. We also showed that GANs are a viable tool to unfold detector smearing and acceptance, ensuring the preservation of initial correlations.

AI keywords:

Generative Adversarial Networks; simulation-based inference; Multi-dimensional correlations;

5

Advanced Tracking Analysis in Space Experiments with Graph Neural Networks

Author: Federica Cuna[1]

Co-authors: Fabio Gargano [2]; Maria Bossa [3]; Mario Nicola Mazziotta [3]

[1] *Istituto Nazionale di Fisica Nucleare- Sezione di Bari*

[2] *INFN-Bari*

[3] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: federica.cuna@ba.infn.it

The integration of advanced artificial intelligence techniques into astroparticle experiments marks a transformative step in data analysis and experimental design. As space missions grow increasingly complex, the adoption of AI technologies becomes critical to optimizing performance and achieving robust scientific outcomes.
This study focuses on the development of innovative AI-driven algorithms for tracking purposes, leveraging the power of Graph Neural Networks (GNNs). GNNs, a subset of geometric deep learning, are well-suited for exploiting the inherent graph structure of tracking systems, where nodes correspond to energy deposits (hits) and edges represent their interconnections. These networks enable a range of tasks, including node classification, link prediction, and graph classification, tailored to the specific challenges of space-based experiments.
A key obstacle in tracking systems for space experiments is the high-noise environment, characterized by backscattering tracks from calorimeter, which complicate the accurate identification of the primary particle trajectory. To overcome this, we propose a novel GNN-based approach for node-level classification, designed to distinguish noise hits, which include backscattering hits and electronic noise, from signal hits and accurately reconstruct particle tracks
The algorithm recognizes the primary hits among the noises one and allows to easily retrieve the track parameters.
By addressing these challenges, our work aims to improve the accuracy and reliability of data interpretation in astroparticle physics, paving the way for more precise and insightful discoveries through the application of cutting-edge AI methodologies.

AI keywords:

Graph Neural Network; SageConv; Supervised Learning;Classification

6

EMBER-2: towards implicit emulators to constrain astrophysics in FIRE-based simulations

Author: Mauro Bernardini[None]

Corresponding Author: mauro.bernardini@uzh.ch

Galaxy formation is a complex problem that links large-scale cosmology with small-scale astrophysics over cosmic timescales. The most principled method, full hydrodynamical simulations, come with high computational costs and thus, the development of faster models is essential. Modern field level emulation techniques leverage Convolutional Neural Networks (CNNs) to "paint" baryonic channels directly on top of dark matter simulations. These emulators are fast to train and reproduce relevant correlations at a fraction of the computational cost compared to traditional simulations.

I will introduce EMBER-2, an enhanced version of the EMBER (Emulating Baryonic EnRichment) framework, designed to emulate baryon channels including gas and HI density, velocity, and temperature over a broad redshift range, from z=6-0. EMBER-2 features a style-based network combined with implicit convolution kernels for fast and accurate emulations. EMBER-2 is capable of interpolating across the entire redshift range with a single CNN and small memory footprint.

EMBER-2 can accurately reconstruct HI-related metrics, such as the cosmic HI column densities, cross-correlations between dark matter, gas, and HI, as well as the correct HI fraction in gaseous

halos.
In a different application, I will also present how EMBER-2 can be used to reconstruct the underlying dark matter distribution from HI surface densities.
I will outline how these advancements lay the groundwork for future analysis pipelines, such as those for the SKA, to constrain dark matter and galaxy formation models.

AI keywords:

Generative Adversarial Networks, Distribution Learning, Physics-based Emulators

7

Advanced Particle Classification in Space Missions Using Transformers

Author: Maria Bossa[1]

Co-authors: Fabio Gargano [2]; Federica Cuna [3]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *INFN-Bari*

[3] *Istituto Nazionale di Fisica Nucleare- Sezione di Bari*

Corresponding Author: mbossa@na.infn.it

The application of advanced artificial intelligence (AI) techniques in astroparticle experiments represents a groundbreaking advancement in data analysis and experimental design. As space missions become increasingly complex, integrating AI technologies is essential for optimizing their performance and enhancing their scientific outcomes. In this study, we propose a fully custom-designed Transformer-based model tailored for calorimeters in space-based experiments. One of the goal for space calorimeter experiment is to distinguish between particle types, such as electrons and protons. By capturing the dependencies within these features, Transformers can achieve robust classifications, even when the data spans thousands of channels or dimensions. By addressing these challenges, we aim to enhance the accuracy and reliability of data interpretation in astroparticle physics through the application of advanced artificial intelligence techniques. Furthermore, this approach has the potential to extend the classification capability across a very broad energy range, spanning from 1 GeV to 100 TeV.

AI keywords:

Transformer, supervised algorithm, classification

8

Full-stack quantum machine learning on hybrid quantum-classical platforms

Author: Stefano Carrazza[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: stefano.carrazza@mi.infn.it

We discuss about applications of hybrid quantum-classical computing and present Qiboml, an open-source software library for Quantum Machine Learning (QML) integrated with the Qibo quantum computing framework. Qiboml interfaces most used classical Machine Learning frameworks such as TensorFlow, PyTorch and Jax with Qibo. This combination enables users to construct quantum or hybrid classical-quantum models that can be executed on any type of hardware accelerators: multi-threading CPU, GPU and multi-GPU for quantum simulation on classical hardware (using state-vector and tensor network approaches) and Quantum Processing Units (QPU) for execution on self-hosted quantum devices. We present a High-Energy Physics application executed on a superconducting single-qubit device through Qiboml.

AI keywords:

quantum-machine-learning, hybrid-computing

9

Field-Level Emulation with Neural Networks

Author: Drew Jamieson[1]

[1] *Max-Planck Institute for Astrophysics*

Corresponding Author: jamieson@mpa-garching.mpg.de

Upcoming galaxy surveys promise to greatly inform our models of the Universe's composition and history. Leveraging this wealth of data requires simulations that are accurate and computationally efficient. While N-body simulations set the standard for precision, their computational cost makes them impractical for large-scale data analysis. In this talk, I will present a neural network-based emulator for modelling nonlinear cosmic structure formation at the field level. Starting from the linear initial conditions of the early Universe, this model predicts the nonlinear evolution from an N-body simulation. We include the underlying cosmological parameters and time evolution, ensuring physical consistency by enforcing a key constraint: velocities correspond to time derivatives of displacements. This constraint markedly enhances the model's accuracy. Trained on an extensive suite of N-body simulations, the network achieves remarkable precision, particularly on small, nonlinear scales where traditional approximations often struggle. The model effectively captures highly nonlinear phenomena, including dark matter halo mergers, and is computationally efficient enough for tasks such as generating mock catalogs and reconstructing the Universe's initial conditions, as I will demonstrate with various applications. This method paves the way for robust, large-scale cosmological analyses using nonlinear scales at the field level.

AI keywords:

physics-informed neural network; surrogate model; simulation-based inference

10

GINGERINO signal reconstruction and classification through neural networks implementation

Author: Giuseppe Di Somma[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: giuseppe.disomma@pi.infn.it

GINGER data analysis is based on the experience gained with GINGERINO data analysis, the general analysis scheme will be described.
The reconstruction of the beat frequency of a laser gyroscope signal in the shortest possible time is

a non-trivial challenge. Advancements in artificial intelligence are used to develop a DAQ system capable of determining the beat signal frequency with higher precision than the FFT-based algorithm, achieving a delay time of just one-hundredth of a second. This neural network achieves double the precision compared to the FFT algorithm. The reconstructed signal is then classified by exploiting the relationship between laser physics phenomena and fringe contrast monitoring. Furthermore, a neural network created for seismic event recognition has, trained on real events present in GINGERINO's data, achieved an accuracy on the available test data ranging between 99% and 100%.

AI keywords:

Anomaly detection; fast frequency reconstruction; seismic event recognition; minimal time delay

11

Towards more precise data analysis with Machine-Learning-based particle identification with missing data

Author: Lukasz Graczykowski[1]

Co-authors: Kamil Deja [1]; Maja Karwowska [1]; Malgorzata Janik [1]; Milosz Kasak [1]; Monika Jakubowska [1]

[1] *Warsaw University of Technology (PL)*

Corresponding Author: lgraczyk@cern.ch

Identifying products of ultrarelativistic collisions delivered by the LHC and RHIC colliders is one of the crucial objectives of experiments such as ALICE and STAR, which are specifically designed for this task. They allow for a precise Particle Identification (PID) over a broad momentum range.

Traditionally, PID methods rely on hand-crafted selections, which compare the recorded signal of a given particle to the expected value for a given particle species (i.e., for the Time Projection Chamber detector, the number of standard deviations in the dE/dx distribution, so-called "nσ" method). To improve the performance, novel approaches use Machine Learning models that learn the proper assignment in a classification task.

However, because of the various detection techniques used by different subdetectors (energy loss, time-of-flight, Cherenkov radiation, etc.), as well as the limited detector efficiency and acceptance, particles do not always yield signals in all subdetectors. This results in experimental data which include "missing values". Out-of-the-box ML solutions cannot be trained with such examples without either modifying the training dataset or re-designing the model architecture. Standard approaches to this problem used, i.e., in image processing involve value imputation or deletion, which may alter the experimental data sample.

In the presented work, we propose a novel and advanced method for PID that addresses the problem of missing data and can be trained with all of the available data examples, including incomplete ones, without any assumptions about their values [1,2]. The solution is based on components used in Natural Language Processing Tools and is inspired by AMI-Net, an ML approach proposed for medical diagnosis with missing data in patient records.

The ALICE experiment was used as an R&D and testing environment; however, the proposed solution is general enough for other experiments with good PID capabilities (such as STAR at RHIC and others). Our approach improves the F1 score, a balanced measure of the PID purity and efficiency of the selected sample, for all investigated particle species (pions, kaons, protons).

[1] M. Kasak, K. Deja, M. Karwowska, M. Jakubowska, Ł. Graczykowski & M. Janik, "Machine-learning-based particle identification with missing data", Eur.Phys.J.C 84 (2024) 7, 691

[2] M. Karwowska, Ł. Graczykowski, K. Deja, M. Kasak, and M. Janik, "Particle identification with machine learning from incomplete data in the ALICE experiment", JINST 19 (2024) 07, C07013

AI keywords:

transformer encoder; attention; classification; incomplete data; embedding

12

Simulation-Based Inference of the Double White Dwarf Population in Gravitational Wave Data

Author: Rahul Srinivasan[1]

Co-authors: Chantal Pitte [2]; Enrico Barausse ; Roberto Trotta [2]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *SISSA, Italy*

Corresponding Author: rsriniva@sissa.it

Understanding the population properties of double white dwarfs (DWDs) in the Milky Way is a key science goal for the upcoming gravitational wave detector, LISA. However, the vast number of galactic binaries (~30×10^6) and the large data size (~6×10^6) pose significant challenges for traditional Bayesian samplers. In this talk, I present a simulation-based inference framework to infer the population characteristics of these binaries directly from LISA data. Our approach leverages a GPU-accelerated forward simulator, enabling the efficient generation of synthetic populations under various astrophysical models. We design a robust summary statistic to compress the high-dimensional input data while preserving the essential population information. We demonstrate the potential for scalable and accurate population inference for deeper insights into the evolution and distribution of galactic white dwarfs.

AI keywords:

Simulation-Based Inference; Bayesian Inference; Data-Summaries; Parameter Estimation

13

Variational inference for pile-up removal at hadron colliders with diffusion models

Authors: Chris Pollard[1]; John Raine[None]; Malte Algren[2]; Tobias Golling[3]

[1] *University of Warwick*

[2] *UNIGE*

[3] *University of Geneva*

Corresponding Author: malte.algren@unige.ch

We present a novel method for pile-up removal of pp interactions using variational inference with diffusion models, called Vipr. Instead of using classification methods to identify which particles are from the primary collision, a generative model is trained to predict the constituents of the hard-scatter particle jets with pile-up removed. This results in an estimate of the full posterior over hard-scatter jet constituents, which has not yet been explored in the context of pile-up removal. We evaluate the performance of Vipr in a sample of jets from simulated $t\bar{t}$ events overlain with pile-up contamination. Vipr outperforms SoftDrop in predicting the substructure of the hard-scatter jets over a wide range of pile-up scenarios.

AI keywords:

Diffusion; Transformers; variational inference

14

pop-cosmos: scaleable Bayesian inference of galaxy properties under a diffusion model prior

Author: Stephen Thorp[1]

Co-authors: Justin Alsing [1]; Hiranya Peiris [2]; Sinan Deger [2]; Daniel Mortlock [3]; Boris Leistedt [4]; Joel Leja [5]; Arthur Loureiro [1]; Anik Halder [2]

[1] *Oskar Klein Centre, Stockholm University*

[2] *Institute of Astronomy, University of Cambridge*

[3] *Astrophysics Group & Department of Mathematics, Imperial College London*

[4] *Astrophysics Group, Imperial College London*

[5] *Department of Astronomy & Astrophysics, The Pennsylvania State University*

Corresponding Author: stephen.thorp@fysik.su.se

Projects such as the imminent Vera C. Rubin Observatory are critical tools for understanding cosmological questions like the nature of dark energy. By observing huge numbers of galaxies, they enable us to map the large scale structure of the Universe. However, this is only possible if we are able to accurately model our photometric observations of the galaxies, and thus infer their redshifts and other properties. I will present a new approach to this problem, which uses a neural emulator to speed up a complex physical model for galaxy spectra (stellar population synthesis; SPS), and a GPU-enabled batched ensemble sampler for posterior sampling. We perform this inference under a flexible diffusion model prior on the 16 physical parameters. This prior is a population distribution that was trained to reproduce the multi-band photometry of a deep 26-band dataset taken from the Cosmic Evolution Survey (COSMOS). I will present the different stages of our pipeline, including the emulation of SPS, the initial training of our population model, and our use of this model as a prior in subsequent inference for individual galaxies. The use of neural emulation for the SPS calculations has enabled us to perform full Bayesian inference for ~300,000 individual galaxies from COSMOS with a sophisticated SPS model - with ongoing work scaling this to tens of millions of galaxies from the Kilo-Degree Survey (KiDS). I will also demonstrate that our population model prior enables more precise and less biased redshift inference than competing methods, with a significantly reduce rate of catastrophic failures.

AI keywords:

diffusion models; Bayesian inference; generative models; emulators

15

Tests for model misspecification in simulation-based inference: from local distortions to global model checks

Authors: Noemi Anau Montel[1]; James Alvey[2]; Christoph Weniger[3]

[1] *Max Planck Institute for Astrophysics*

[2] *University of Cambridge*

[3] *GRAPPA, University of Amsterdam*

Corresponding Author: noemiam@mpa-garching.mpg.de

Model misspecification analysis strategies, such as anomaly detection, model validation, and model comparison are a key component of scientific model development. Over the last few years, there has been a rapid rise in the use of simulation-based inference (SBI) techniques for Bayesian parameter estimation, applied to increasingly complex forward models. To move towards fully simulation-based analysis pipelines, however, there is an urgent need for a comprehensive simulation-based framework for model misspecification analysis.

In this talk, I will describe a solid and flexible foundation for a wide range of model discrepancy analysis tasks, using *distortion-driven model misspecification tests*. From a theoretical perspective, I will introduce the statistical framework built around performing many hypothesis tests for distortions of the simulation model. I will also make explicit analytic connections to classical techniques: anomaly detection, model validation, and goodness-of-fit residual analysis. Furthermore, I will introduce an efficient self-calibrating training algorithm that is useful for practitioners. I will demonstrate the performance of the framework in multiple scenarios, making the connection to classical results where they are valid. Finally, I will show how to conduct such a distortion-driven model misspecification test for real gravitational wave data, specifically on the event GW150914.

Related work at https://arxiv.org/abs/2412.15100 .

AI keywords:

simulation-based inference, misspecification tests, out-of-distribution, anomaly detection

16

A Deep Learning approach to event reconstruction in Super-Kamiokande

Author: Nicola Fulvio Calabria[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: nicola.calabria@ba.infn.it

Super-Kamiokande is a 50-kton Water Cherenkov detector, operating since 1996 in the Kamioka mine, Japan, whose broad scientific program spans from neutrino physics to baryon number violating processes, such as proton decay. In this preliminary study I show the development of a Deep Learning model, based on Convolutional Neural Networks (CNN) and Residual Neural Networks (ResNet), for event reconstruction in Super-Kamiokande. To do so, simulated event samples have been used. This study aims to the development of a software tool to be employed alongside the official reconstruction software (fiTQun) to improve particle detection and reconstruction for proton decay analysis.

AI keywords:

CNN;ResNet;pattern recognition;image analysis

17

Rapid Identification and Classification of Eccentric Binary Black-hole mergers using Machine Learning

Author: Yuvraj Sharma[None]

Co-author: Prayush Kumar [1]

[1] *International Centre for Theoretical Sciences, Bangalore*

Corresponding Author: yuvrajasharma.acads@gmail.com

The future of Gravitational Wave (GW) detectors [LVK] have made remarkable progress, with an expanding sensitivity band and the promise of exponential increase in detection rates for upcoming observing runs [O4 and beyond]. Among the diverse sources of GW signals, eccentric Binary mergers present an intriguing and computationally challenging aspect. We address the imperative need for efficient detection and classification of eccentric Binary mergers using Machine Learning (ML) techniques. Traditional Bayesian Parameter estimation methods, while accurate, can be prohibitively time-consuming and computationally expensive. To overcome this challenge, we leverage the capabilities of ML to expedite the identification and classification of eccentric GW events. I will present our approach that employs Separable Convolutional Neural Networks (SCNN) to discriminate between non-eccentric and eccentric Binary mergers and further classifying the latter into categories of low, moderate, and high eccentricity mergers.

AI keywords:

CNN, Separable CNN, Neural Networks

18

Quantum Dynamics with Time-dependent NQS

Author: Alejandro Romero-Ros[1]

[1] *Universitat de Barcelona*

Corresponding Author: alejandro.romero.ros@fqa.ub.edu

One of the main challenges in solving quantum many-body (MB) problems is the exponential growth of the Hilbert space with system size.

In this regard, a new promising alternative are neural-network quantum states (NQS).
This approach leverages the parameterization of the wave function with neural-network architectures.
Compared to other variational methods, NQS are highly scalable with systemsize and can naturally capture complex behaviours. #nonlinearities and nonlocal correlations within the system.

Here, we present proof-of-principle time-dependent NQS simulations, involving coherent states of single-particle models, to illustrate the ability of this approach to effectively capture key aspects of quantum dynamics in the continuum.
These results pave the way to more complex MB systems with promising applications in nuclear physics, ultracold atoms, and quantum simulations.

AI keywords:

Neural Quantum States; reinforcement learning; real-time dynamics;

19

DINAMO: Dynamic and INterpretable Anomaly MOnitoring for Large-Scale Particle Physics Experiments

Authors: Arsenii Gavrikov[1]; Julián García Pardiñas[2]

Co-author: Alberto Garfagnini [1]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *University of Milano-Bicocca*

Corresponding Author: arsenii.gavrikov@pd.infn.it

Ensuring reliable data collection in large-scale particle physics experiments demands Data Quality Monitoring (DQM) procedures to detect possible detector malfunctions and preserve data integrity. Traditionally, this resource-intensive task has been handled by human shifters who may struggle with frequent changes in operational conditions. Instead, to simplify and automate the shifters' work, we present DINAMO: a dynamic and interpretable anomaly detection framework for large-scale particle physics experiments in time-varying settings [1]. Our approach constructs evolving histogram templates with built-in uncertainties, featuring both a statistical variant - extending the classical Exponentially Weighted Moving Average (EWMA) - and a machine learning (ML)-enhanced version that leverages a transformer encoder for improved adaptability and accuracy.
Both approaches are studied using comprehensive synthetic datasets that emulate key features of real particle physics detectors. Validations on a large number of such datasets demonstrate the high accuracy, adaptability, and interpretability of these methods, with the statistical variant being commissioned in the LHCb experiment at the Large Hadron Collider, underscoring its real-world impact.

[1] A. Gavrikov, J. García Pardiñas, and A. Garfagnini, DINAMO: Dynamic and INterpretable Anomaly MOnitoring for Large-Scale Particle Physics Experiments (2025). Link: https://arxiv.org/abs/2501.19237

AI keywords:

Anomaly Detection; Interpretability; Online Learning; Transformer

20

Unsupervised Particle Tracking with Neuromorphic Computing

Author: Fabio Cufino[1]

Co-authors: Eleonora Porcu [1]; Emanuele Coradin [2]; Fredrik Sandin [3]; Jinu Raj [4]; Tommaso Dorigo [5]

[1] *University of Bologna*

[2] *University of Padova*

[3] *Lulea Techniska Universitet*

[4] *Central University of Tamilnadu*

[5] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: fabio.cufino5@gmail.com

We study the application of a neural network architecture for identifying charged particle trajectories via unsupervised learning of delays and synaptic weights using a spike-time-dependent plasticity rule. In the considered model the neurons receive time-encoded information on the position of particle hits in a tracking detector for a particle collider, modeled according to the geometry of the Compact Muon Solenoid Phase II detector. We show how a spiking neural network is capable of successfully identifying in a completely unsupervised way the signal left by charged particles in the presence of conspicuous noise from accidental or combinatorial hits, opening the way to applications

of neuromorphic computing to particle tracking. The presented results motivate further studies investigating neuromorphic computing as a potential solution for real-time, low-power particle tracking in future high-energy physics experiments.

AI keywords:

neuromorphic computing, detector design, pattern recognition

21

Gaussian Processes: Machine Learning for Observable Interpolation & Data Analysis

Author: Ryan Ferguson[1]

Co-author: Bryan McKinnon [1]

[1] *University of Glasgow*

Corresponding Author: r.ferguson.3@research.gla.ac.uk

Current studies of the hadron spectrum are limited by the accuracy and consistency of datasets. Information derived from theory models often requires fits to measurements taken at specific values of kinematic variables, which needs interpolation between such points. In sparse datasets the quantification of uncertainties is problematic. Machine Learning is a powerful tool that can be used to build an interpolated dataset, with quantification of uncertainties. The primary focus here is one type of machine learning called a Gaussian Process (GP).

By calculating the covariance between known datapoints, the GP can predict the mean and standard deviation of other, unknown, datapoints, with no theoretical model dependence. The GP model presented here uses a bespoke method to find the optimal hyperparameters, using Bayesian inference. Checking and testing is performed using Legendre polynomials to ensure it is unbiased and gives accurate predictions.

The GP can also be used to give a probability density prediction between conflicting, and potentially inconsistent, datasets of the same physics observable. This enables theorists to test their models using the largest likelihood available from data.

AI keywords:

Bayesian inference; uncertainty quantification; dataset creation

22

Optimal Equivariance from the Matrix-Element Method

Authors: Daniel Maitre[1]; Vishal Ngairangbam[1]; Michael Spannowsky[1]

[1] *Institute for Particle Physics Phenomenology, Durham University*

Corresponding Author: vishal.s.ngairangbam@durham.ac.uk

The Matrix-Element Method (MEM) has long been a cornerstone of data analysis in high-energy physics. It leverages theoretical knowledge of parton-level processes and symmetries to evaluate

the likelihood of observed events. We combine MEM-inspired symmetry considerations with equivariant neural network design for particle physics analysis. Even though Lorentz invariance and permutation invariance over all reconstructed objects are the largest and most natural symmetry in the input domain, we find that they are sub-optimal in most practical search scenarios. We propose a longitudinal boost-equivariant message-passing network. We present numerical studies demonstrating MEM-inspired architectures achieve new state-of-the-art performance in distinguishing di-Higgs decays to four bottom quarks from the QCD background, with enhanced sample and parameter efficiencies. This synergy between MEM and equivariant deep learning opens new directions for physics-informed architecture design, promising more powerful tools for probing physics beyond the Standard Model.

AI keywords:

Group Equivariance, Message Passing Neural Networks, Point Clouds

23

Automatizing the search for mass resonances using BumpNet

Authors: Amit Shkuri[1]; Bruna Pascual[2]; Ethan Meszaros[3]; Etienne Dreyer[1]; Eva Mayer[2]; Fannie Bilodeau[3]; Georges Azuelos[3]; Hoang Dai Nghia Nguyen[3]; Ilan Bessudo[1]; Jean-François Arguin[3]; Josephine Potdevin[4]; Julien Donini[2]; Maryna Borysova[1]; Michael Kwok Lam Chu[1]; Muhammad Usman[3]; Nilotpal Kakati[1]; Samuel Calvet[2]; Shalini Epari[3]; Shikma Bressler[1]; Émile Baril[3]

[1] *Weizmann Institute of Science (IL)*

[2] *Université Clermont Auvergne (FR)*

[3] *Université de Montréal (CA)*

[4] *EPFL - École Polytechnique Fédérale de Lausanne (CH)*

Corresponding Author: ethan.meszaros@umontreal.ca

The search for resonant mass bumps in invariant-mass histograms is a fundamental approach for uncovering Beyond the Standard Model (BSM) physics at the Large Hadron Collider (LHC). Traditional, model-dependent analyses that utilize this technique, such as those conducted using data from the ATLAS detector at CERN, often require substantial resources, which prevent many final states from being explored. Modern machine learning techniques, such as normalizing flows and autoencoders, have facilitated such analyses by providing various model-agnostic approaches; however many methods still depend on background and signal assumptions, thus decreasing their generalizability. We present *BumpNet*, a convolutional neural network (CNN) that predicts log-likelihood significance values in each bin of smoothly falling invariant-mass histograms, enhancing the search for resonant mass bumps. This technique enables a **model-independent search of many final states without the need for traditional background estimation**, making BumpNet a powerful tool for exploring the many unsearched areas of the phase space while saving analysis time. Trained on a dataset consisting of realistic smoothly-falling data and analytical functions, the network has produced encouraging results, such as predicting the correct significance of the Higgs boson discovery, agreement with a previous ATLAS dilepton resonance search, and success in realistic Beyond the SM (BSM) scenarios. We are now training and optimizing BumpNet using ATLAS Run 2 Monte Carlo data, with the ultimate goal of performing general searches on real ATLAS data. These encouraging results highlight the potential for BumpNet to accelerate the discovery of new physics.

AI keywords:

Anomaly detection; Likelihood-based inference; Pattern recognition

24

LLM-based AI assistant for HEP data analysis

Author: Ke Li[None]

Corresponding Author: ke.li@cern.ch

The data processing and analyzing is one of the main challenges at HEP experiments, normally one physics result can take more than 3 years to be conducted. To accelerate the physics analysis and drive new physics discovery, the rapidly developing Large Language Model (LLM) is the most promising approach, it have demonstrated astonishing capabilities in recognition and generation of text while most parts of physics analysis can be benefitted. In this talk we will discuss the construction of a dedicated intelligent agent, an AI assistant at BESIII based on LLM, the potential usage to boost hadron spectroscopy study, and the future plan towards a AI scientist.

AI keywords:

Automation, Large Language model, data analysis

25

ML techniques to search for antideuterons with AMS-02 on the International Space Station

Author: Francesco D'Angelo[1]

Co-author: Alberto Oliva [1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: fdangelo@bo.infn.it

Most of the antimatter in cosmic rays is produced by collisions of high energy particles with the interstellar medium while they propagate through it. The detection of an antimatter component over the collisional background can be used to investigate new sources, as the presence of dark matter annihilations in the halo. A possible smoking gun for dark matter is given by the detection of antideuterons below the GeV/n scale, where the secondary production is forbidden by kinematics and the presence of anti-deuterons can be associated only with exotic processes. The Alpha Magnetic Spectrometer (AMS) installed in 2011 on the International Space Station, is a large field of view high-energy particle detector able to measure rare antimatter components. However, the antideuteron search requires a high level of background rejection, coming from cosmic protons and antiprotons. In this talk I will discuss the experimental methods, and the ML techniques, that are used to reject the background in the AMS data.

AI keywords:

Classification, DNN, BDT

26

FAIR Universe : HiggsML Uncertainty Challenge Competition

Authors: Aishik Ghosh[1]; Benjamin Nachman[2]; Chris Harris[2]; David Rousseau[3]; Elham E Khoda[4]; Ihsan Ullah[5]; Isabelle Guyon[5]; Jordan Dudley[6]; Paolo Calafiura[2]; Peter Nugent[2]; Po-Wen Chang[2]; Ragansu Chakkappai[7]; Sascha Diefenbacher[2]; Shih-Chieh Hsu[8]; Steven Farrell[2]; Wahid Bhimji[2]; Yuan-Tang Chou[9]; Yulei Zhang[9]

[1] *University of California Irvine*

[2] *Lawrence Berkeley National Laboratory*

[3] *IJCLab-Orsay*

[4] *University of California, San Diego (UCSD)*

[5] *Chalearn*

[6] *Lawrence Berkeley National Laboratory , University of California Berkeley*

[7] *Universit\'e Paris-Saclay, CNRS/IN2P3, IJCLab*

[8] *UWashington*

[9] *University of Washington Seattle*

Corresponding Author: ragansu.chakkappai@cern.ch

The **Fair Universe** project organised the **HiggsML Uncertainty Challenge**, which took place from Sep 2024 to 14th March 2025. This groundbreaking competition in high-energy physics (HEP) and machine learning was the first to place a strong emphasis on uncertainties, focusing on mastering both the uncertainties in the input training data and providing credible confidence intervals in the results.
The challenge revolved around measuring the Higgs to tau+ tau- cross-section, similar to the **HiggsML challenge** held on Kaggle in 2014, using a dataset representing the 4-momentum signal state. Participants were tasked with developing advanced analysis techniques capable of not only measuring the signal strength but also generating confidence intervals that included both statistical and systematic uncertainties, such as those related to detector calibration and background levels. The accuracy of these intervals was automatically evaluated using pseudo-experiments to assess correct coverage.
Techniques that effectively managed the impact of systematic uncertainties were expected to perform best, contributing to the development of uncertainty-aware AI techniques for HEP and potentially other fields. The competition was hosted on **Codabench**, an evolution of the Codalab platform, and leveraged significant resources from the **NERSC infrastructure** to handle the thousands of required pseudo-experiments.
This competition was selected as a **NeurIPS competition**, and the preliminary results were presented at the **NeurIPS 2024** conference in December. As the challenge concluded in March 2025, an account of the most innovative solutions and final outcomes will be presented at this conference.

AI keywords:

Benchmark; Uncertainty Quantification; Simulation- Based Inference;

27

Robustly Dissecting the Gamma-Ray Sky at High Latitudes with Simulation-Based Inference

Authors: Christoph Weniger[1]; Christopher Eckner[2]; Florian List[3]; Francesca Calore[4]; Noemi Anau Montel[5]

[1] *GRAPPA, University of Amsterdam*

[2] *University of Nova Gorica*

[3] *University of Vienna*

[4] *LAPTh, CNRS*

[5] *Max Planck Institute for Astrophysics*

Corresponding Author: ceckner@ung.si

Over the past 16 years, the *Fermi* Large Area Telescope (LAT) has significantly advanced our view of the GeV gamma-ray sky, yet several key questions remain - such as the nature of the isotropic diffuse background, the properties of the Galactic pulsar population, and the origin of the GeV excess towards the Galactic Centre. Addressing these challenges requires sophisticated astrophysical

modelling and robust statistical methods capable of handling high-dimensional parameter spaces.
In this work, we analyse 14 years of high-latitude ($|b| > 30°$) *Fermi*-LAT data in the 1–10 GeV range using simulation-based inference (SBI) via neural ratio estimation. This approach allows us to detect individual gamma-ray sources and derive a source catalogue with estimated positions and fluxes that are consistent with the bright portion of the *Fermi*-LAT collaboration's 4FGL catalogue. Additionally, we reconstruct the source-count distribution $\mathrm{d}N/\mathrm{d}S$ in both parametric and non-parametric forms, achieving good agreement with previous literature results and detected sources. We also quantitatively validate our gamma-ray emission simulator via an anomaly detection technique demonstrating that the synthetic data closely reproduces the complexity of the real observations.
This study highlights the practical utility of SBI for complex, high-dimensional problems in gamma-ray astronomy and lays the groundwork for its application to more challenging sky regions or data from next-generation facilities such as the Cherenkov Telescope Array Observatory.

AI keywords:

simulation-based inference; neural ratio estimation; uncertainty quantification; anomaly detection

28

An implementation of neural simulation-based inference for parameter estimation in ATLAS

Authors: Aishik Ghosh[1]; Arthur Schaffer[2]; David Rousseau[3]; Jay Sandesara[4]

[1] *University of California Irvine*

[2] *IJCLab*

[3] *IJCLab-Orsay*

[4] *University of Wisconsin-Madison*

Corresponding Author: jay.ajitbhai.sandesara@cern.ch

Neural simulation-based inference is a powerful class of machine-learning-based methods for statistical inference that naturally handles high-dimensional parameter estimation without the need to bin data into low-dimensional summary histograms. Such methods are promising for a range of measurements, including at the Large Hadron Collider, where no single observable may be optimal to scan over the entire theoretical phase space under consideration, or where binning data into histograms could result in a loss of sensitivity. This work develops a neural simulation-based inference framework for statistical inference, using neural networks to estimate probability density ratios, which enables the application to a full-scale analysis. It incorporates a large number of systematic uncertainties, quantifies the uncertainty due to the finite number of events in training samples, develops a method to construct confidence intervals, and demonstrates a series of intermediate diagnostic checks that can be performed to validate the robustness of the method. A first full application of the novel techniques to an ATLAS measurement of the off-shell Higgs boson in the $H \to ZZ \to 4\ell$ final state is also presented. This approach represents an extension to the standard statistical methodology used by the experiments at the Large Hadron Collider, and can benefit many other physics analyses as well.

AI keywords:

Simulation-Based Inference, deep neural networks, uncertainty quantification

29

Synthetic data generation for the training of trigger systems and embodied models.

Author: Giuseppe Levi[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: giuseppe.levi@bo.infn.it

The sheer volume and complexity of data from high-energy physics experiments makes neural networks particularly attractive for the implementation of trigger systems. On the other hand, a large amount of classified and labelled data is required to train a network and this can be a complex task, especially if the experimental data were in the form of images. In this contribution we discuss the possibility of constructing a synthetic dataset using other neural networks that can be trained with a small number of real-world data examples. This synthetic dataset can be used to train pattern recognition systems as well as to train embodied models to make the measuring instrument an independent robotic system. An example of this technique applied to images of the plastic detectors of the MoEDAL experiment at CERN will be shown.

AI keywords:

pattern recognition; training dataset; generative networks; embodied models; multimodal models

30

Accelerating TimeSPOT Detector Simulation with Deep Learning

Authors: Zhihua Liang[1]; Angelo Loi[1]; Stefano Cabiddu[1]; Rudolf Oldeman[1]; Andrea Lampis[1]; Andrea Contu[1]; Adriano Lai[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: zhihua.liang@ca.infn.it

The TimeSPOT project has developed innovative sensors optimized for precise space and time measurements of minimum-ionizing particles in high-radiation environments. These sensors demonstrate exceptional spatial resolution (around 10 μm) and time resolution (around 10 ps), while withstanding high fluences (> 10^{17} 1 MeV n_eq/cm²). Tests on small-scale structures confirm their potential for deployment in next-generation inner trackers, including potential applications in experiments such as the new VELO detector of LHCb Upgrade II (Hi-Lumi LHC) and neutrino-tagging techniques. However, large-scale simulations within a full detector apparatus are essential to validate performance. Detailed simulations using Geant4/TCAD/TCoDe are computationally expensive, limiting design studies. To address this, we have developed a fast simulation based on deep learning, initially focusing on a Multi-Layer Perceptron (MLP) architecture. This work establishes a crucial proof-of-concept for applying machine learning to accelerate detector simulation.

Our approach involves training the MLP on a dataset of simulated TimeSPOT events (single-pixel geometry) generated with Geant4/TCAD/TCoDe. The trained model is converted to highly optimized C++ code via ONNX and ROOT TMVA SOFIE, and has been successfully integrated into the Gauss simulation framework. This integration demonstrates the feasibility of embedding machine learning models directly within complex simulation environments.

Results demonstrate a substantial speedup (10^4-10^5 times) compared to traditional simulations, while maintaining good accuracy. Subsequent work has explored advanced architectures, including Edge-Activated Adaptive Function Networks (EAAFNs), achieving even higher accuracy (R^2 > 0.99 for charge, R^2 > 0.93 for CoG). These EAAFN models are candidates for future integration.

This fast simulation, currently for a single TimeSPOT pixel, paves the way for significant advancements: multi-pixel geometries, entire detector modules, and other detector types. The proven methodology –training, conversion to optimized C++, and integration –can be extended to parameterizable simulations.

AI keywords:

fast simulation; deep learning; MLP; EAAFN; SOFIE; ONNX; TimeSPOT; Gauss

31

SKATR: A Self-Supervised Summary Transformer for the Square Kilometer Array

Authors: Ayodele Ore[1]; Caroline Heneka[1]; Tilman Plehn[1]

[1] *Heidelberg University, ITP*

Corresponding Author: ore@thphys.uni-heidelberg.de

The Square Kilometer Array (SKA) will bring about a new era of radio astronomy by allowing 3D imaging of the Universe during Cosmic Dawn and Reionization. Machine learning promises to be a powerful tool to analyze the highly structured and complex signal, however accurate training datasets are expensive to simulate and supervised learning may not generalize. We introduce SKATR, a self-supervised vision transformer whose learned encoding can be cheaply adapted for downstream tasks on 21cm maps. Focusing on regression and posterior inference of simulation parameters, we demonstrate that SKATR representations are near lossless. We also study how SKATR generalizes to differently-simulated datasets and compare to fully-supervised baselines.

AI keywords:

Self-Supervision; Simulation-Based Inference; Transformers

32

On the accuracy of posterior recovery with neural network emulators

Authors: Harry Bevins[1]; Thomas Gessey-Jones[1]; Will Handley[1]

[1] *University of Cambridge*

Corresponding Author: htjb2@cam.ac.uk

Neural network emulators or surrogates are widely used in astrophysics and cosmology to approximate expensive simulations, accelerating both likelihood-based inference and training for simulation-based inference. However, emulator accuracy requirements are often justified heuristically rather than with rigorous theoretical bounds. We derive a principled upper limit on the information loss introduced by an emulator with a given accuracy. This is quantified via the Kullback-Leibler divergence between the true posterior, which would be recovered using full simulations if computationally feasible, and the inferred posterior obtained with the emulator. Under assumptions of model linearity, uncorrelated noise, and a Gaussian likelihood, we show that accurate posterior recovery remains possible even when emulator errors reach 20% of the data noise level. We demonstrate the utility of this bound with an example from 21-cm cosmology, where neural networks are extensively used to constrain the astrophysics of the early universe with current observational limits.

AI keywords:

Emulators, simulation-based inference, inference, information theory, Bayesian analysis

33

Bridging the Gap: Unfolding and Quantification Learning for Physics Research

Author: Mirko Bunse[1]

[1] *Lamarr Institute for Machine Learning and Artificial Intelligence, Dortmund, Germany*

Corresponding Author: mirko.bunse@cs.tu-dortmund.de

The resolution of any detector is finite, leading to distortions in the measured distributions. Within physics research, the indispensable correction of these distortions is know as *Unfolding*. Machine learning research uses a different term for this very task: *Quantification Learning*. For the past two decades, this difference in terminology (and some differences in notation) have prevented physicists and computer scientists from acknowledging the fact that Unfolding and Quantification Learning cover indeed the same mathematical problem.

In this talk, I will bridge the gap between these two branches of research and I will provide an overview of the many key results that Quantification Learning has produced over the past two decades, covering statistical consistency, the anatomy of reconstruction errors, improved optimization techniques, more informative data representations, and arbitrary numbers of observable quantities. Each of these results has immediate and compelling implications on the practice of Unfolding, tackling questions like: Which algorithms are trustworthy? How can we increase their performance and how should we implement them? How much data do we need? Which of the current limits are inherent and which can be lifted? I will discuss these questions from an interdisciplinary perspective, taking into account recent developments from both physics and machine learning research.

AI keywords:

quantification learning; learning theory; label shift; classification; constrained optimization

34

Detecting gravitational waves using convolutional neural networks.

Authors: Gianluca Guidi[1]; Lorenzo Mobilia[1]; Tito Dal Canton[2]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *IJCLab*

Corresponding Author: l.mobilia@campus.uniurb.it

The first detection of the gravitational wave event GW150914 in 2015 opened the doors to the gravitational astronomy. Since then, hundreds of such events have been detected. Some of these have been particularly significant, such as GW170817, the first binary neutron star merger. This detection enabled a measurement of electromagnetic counterpart marking the beginning of the multi-messenger astrophysics with gravitational waves. Detecting such events is extremely challenging. The challenges go from noise isolation in interferometers to the data analysis techniques used to identify and characterize the signals in the interferometric data. In this context, machine learning techniques have been deeply explored as potential solutions to the gravitational wave detection. In this work, we explore a possible solution in data analysis for discovering binary neutron star events, which so far have represented the biggest challenge for machine learning method. Exploiting simulated data that includes Gaussian noise, transient noise and gravitational wave signals, a Residual Network is trained and tested to recognize the presence of a gravitational wave signal in the data. We compare the performance of this approach with well-established methods that are used presently for the detection of gravitational waves.

AI keywords:

Deep Learning; Residual Network; anomaly detection;

35

Simulation-Based Inference for Antenna Gain Calibration in 21 cm Cosmology

Author: Joy Sanghavi[1]

Co-authors: Sambit Giri [2]; Daan Meerburg [2]; Christoph Weniger [3]

[1] *University of Amsterdam*

[2] *University of Groningen*

[3] *GRAPPA, University of Amsterdam*

Corresponding Author: j.k.sanghavi@uva.nl

The 21 cm signal from neutral hydrogen is a key probe of the Epoch of Reionization (EoR), marking the universe's transition from a cold, neutral state to a predominantly hot, ionized one, driven by the formation of the first stars and galaxies. Extracting this faint 21 cm signal from radio interferometric data requires precise gain calibration. However, traditional calibration methods are computationally expensive and time-intensive. More efficient calibration techniques are urgently needed with next-generation radio telescopes like the Square Kilometer Array (SKA) set to host hundreds of antennas.

To address this challenge, we present a sequential simulation-based inference (SBI) approach for direction-independent gain calibration, designed to automate and accelerate the process while improving scalability and accuracy. Once a forward model is established to generate simulations—transformations of the true sky image due to antenna gain variations—neural posterior estimation (NPE) with embedding networks is employed to infer the correct gain values for multiple antennas from the joint parameter-data distribution. We leverage GPU-accelerated parallelization to efficiently estimate the large number of gain parameters involved, within a feasible time frame.

The Bayesian framework enables robust uncertainty estimation, which traditional methods often overlook while facilitating faster and more reliable analysis of real SKA data. Future work could extend this approach to direction-dependent gains and other systematic effects or involve validating existing radio data. By integrating these techniques into the analysis pipeline, we can fully exploit SKA's unprecedented sensitivity, significantly improving our ability to extract fundamental cosmological insights from large-scale observations.

AI keywords:

simulation-based inference; Neural Posterior Estimation; GPU Parallelization

36

OmniJet-alpha and beyond: foundation model expansions

Authors: Anna Hallin[1]; Gregor Kasieczka[1]; Henning Rose[1]; Joschka Birk[1]; Martina Mozzanica[1]

[1] *Universität Hamburg*

Corresponding Author: anna.hallin@uni-hamburg.de

OmniJet-alpha, released in 2024, is the first cross-task foundation model for particle physics, demonstrating transfer learning between an unsupervised problem (jet generation) and a classic supervised task (jet tagging). This talk will present current developments and expansions of the model. We will for example show how we are able to utilize real, unlabeled CMS data to pretrain the model. We will also cover how OmniJet-alpha can be used to generate calorimeter showers, showcasing its capabilities to work with very different data types.

AI keywords:

foundation models; transformers; generation; classification

37

Towards a Pixel-Based Imaging of Quantum-Correlation Functions

Author: Marco Zaccheddu[1]

Co-author: Nobuo Sato [1]

[1] *Jefferson Lab*

Corresponding Author: zacch@jlab.org

Understanding hadron structure requires the extraction of Quantum Correlation Functions (QCFs), such as parton distribution functions and fragmentation functions, from experimental data. The extraction of QCFs involves solving an inversion problem, which is ill-posed due to errors and limitations in the experimental data.

To address this challenge, we propose a novel method for extracting QCFs by conceptualizing them as images or multidimensional tensors. This approach allows us to leverage image processing techniques, including Generative Adversarial Networks (GANs), to not only extract the QCFs but also quantify the associated uncertainties.

We will present results showcasing the application of this novel framework to the extraction of Generalized Parton Distribution Functions (GPDs) and Transverse Momentum Dependent Distribution Functions (TMDs).

AI keywords:

Generative Adversarial Networks : Inversion Problem : Image Processing : Image Processing

38

BART-Lagrangian : When Transformers Write Particle Physics Lagrangians

Authors: Eliel Camargo-Molina[1]; Yong Sheng Koay[1]

Co-authors: Rikard Enberg [1]; Stefano Moretti [2]

[1] *Uppsala University*

[2] *NExT Institute (Southampton & RAL) and Uppsala University*

Corresponding Author: yongsheng.koay@physics.uu.se

In physics, Lagrangians provide a systematic way to describe laws governing physical systems. In the context of particle physics, they encode the interactions and behavior of the fundamental building blocks of our universe. By treating Lagrangians as complex, rule-based constructs similar to linguistic expressions, we trained a transformer model– proven to be effective in natural language tasks – to predict the Lagrangian corresponding to a given list of particles. We report on the transformer's performance in constructing Lagrangians respecting the SU(3)×SU(2)×U(1) gauge symmetries. The resulting model is shown to achieve high accuracies with Lagrangians up to six matter fields, with the capacity to generalize beyond the training distribution, albeit within architectural constraints. We show through embedding analysis that the model has internalized concepts such as group representations and conjugation operations eventhough it is trained to generate Lagrangians. It is also capable of generating parts of known Lagrangians such as the Standard Model and other BSM models. We make the model and training datasets available to the community. An interactive demonstration can be found at: https://huggingface.co/spaces/JoseEliel/generate-lagrangians

AI keywords:

Transformers ; Symbolic AI ; LLM ; Out-of-Distribution Generalization ; Embedding Analysis

39

Estimation of Temporal Muon Signals in Water-Cherenkov Detectors of the Surface Detector of the Pierre Auger Observatory

Author: Margita Kubátová[1]

[1] *FZU - Institute of Physics of the Czech Academy of Sciences*

Corresponding Author: majercakova@fzu.cz

The Pierre Auger Observatory is a cosmic-ray detector that uses multiple systems to simultaneously observe extensive air showers (EAS). EAS are particle cascades initiated by ultra-high-energy cosmic rays (UHECRs) interacting with the atmosphere of the Earth. Determining the sources of UHECRs requires precise knowledge of their mass composition. One key observable for estimating the mass of an impinging cosmic ray is the number of muons created in the shower cascade. The Surface Detector (SD) of the Observatory is a 3,000 km^2 array of independent detector stations. The main component of the SD stations is a water-Cherenkov detector, which records the signals of air shower particles reaching the ground. Since the particle cascade consists of many different elementary particles, such as muons, electrons, and photons, filtering the detected signals to isolate the muonic component is non-trivial. To estimate the contribution of muons in these signals, a recurrent neural network approach based on long short-term memory layers was developed. The model performs well on simulations, achieving a small muon signal bias on an unseen dataset and generalizing across different hadronic interaction models. In addition, the estimator can be calibrated using measurements from the Underground Muon Detector, a set of buried scintillator detectors that directly measure high-energy muons. Such a calibration is a key advantage due to limitations in shower and detector simulations.

AI keywords:

recurrent neural networks, long short-term memory, simulation-based inference

40

Emulating CO Line Radiative Transfer with Deep Learning

Author: Shiqi Su[1]

Co-authors: Frederik Ceuster [2]; Jaehoon Cha [3]; Jeyan Thiyagalingam [3]; Jeremy Yates [4]; Yihang Zhu [5]; Jan Bolte [6]; Mark Wilkinson [1]

[1] *Department of Physics and Astronomy, University of Leicester*

[2] *Department of Physics and Astronomy, Institute of Astronomy, KU Leuven*

[3] *Scientific Computing, Rutherford Appleton Laboratory, Science and Technology Facilities Council*

[4] *Department of Computer Science, University College London*

[5] *School of Computing and Mathematical Sciences, University of Leicester*

[6] *Department of Mathematics, Kiel University*

Corresponding Author: ss1421@leicester.ac.uk

The adoption of AI-based techniques in theoretical research is often slower than in other fields due to the perception that AI-based methods lack rigorous validation against theoretical counterparts. In this talk, we introduce COEmuNet, a surrogate model designed to emulate carbon monoxide (CO) line radiation transport in stellar atmospheres.

COEmuNet is based on a three-dimensional residual neural network and is specifically trained to generate synthetic observations of evolved star atmospheres. The model is trained on data from hydrodynamic simulations of Asymptotic Giant Branch (AGB) stars perturbed by a binary companion. Given a set of input parameters, including velocity fields, temperature distributions, and CO molecular number densities, the COEmuNet model emulates spectral line observations with a mean relative error of ~7% compared to a classical numerical solver of the radiative transfer equation, while being 1,000 times faster.

This presentation will also include some of our preliminary results, demonstrating the improved performance achieved through Physics-Informed Machine Learning (PIML) applied to the same problem, highlighting its potential for accelerating radiative transfer modelling in AGB stars.

AI keywords:

surrogate models, physics informed machine learning, convolutional neural networks

41

Neural Network-Based Particle Identification: Towards Physics-Informed Loss Functions

Authors: Marvin Kohls[1]; Simon Spies[2]

[1] *GSI Helmholtzzentrum für Schwerionenforschung GmbH*

[2] *Goethe University Frankfurt*

Corresponding Authors: s.spies@gsi.de, m.kohls@gsi.de

In this conference contribution, we present our findings on applying Artificial Neural Networks (ANNs) to enhance off-vertex topology recognition using data from the HADES experiment at GSI, Darmstadt. Our focus is on decays of Λ and K^0_S particles produced in heavy ion as well as elementary reactions. We demonstrate how ANNs can enhance the separation of weak decays from combinatorial background by performing a MultiVariate Analysis (MVA), taking into account strong nonlinear correlations between topology parameters.

Furthermore, we introduce Physics Informed Loss Functions to test whether it improves the results significantly. This contribution also discusses the potential of improving the identification of single charged particle tracks coming from the primary vertex.

AI keywords:

Artificial Neural Networks (ANN); Physics Informed Neural Networks (PINN); MultiVariate Analysis (MVA)

42

Rings of Light, Speed of AI: YOLO for Cherenkov Reconstruction

Authors: Giovanni Laganà[1]; Martino Borsato[2]; Maurizio Martinelli[3]

[1] *Università degli Studi di Milano Bicocca*

[2] *Milano Bicocca University and INFN*

[3] *Università degli Studi di Milano Bicocca e INFN*

Corresponding Authors: g.lagana2@campus.unimib.it, maurizio.martinelli@unimib.it, martino.borsato@mib.infn.it

Cherenkov rings play a crucial role in identifying charged particles in high-energy physics (HEP) experiments. The size of the light cone depends directly on the mass and momentum of the particle that produced it. Most Cherenkov ring pattern reconstruction algorithms currently used in HEP experiments rely on a likelihood fit to the photo-detector response, which often consumes a significant portion of the computing budget for event reconstruction. As the field moves toward real-time event reconstruction, faster and more efficient techniques are needed.
We present a novel approach to Cherenkov ring reconstruction using YOLO, a computer vision algorithm capable of real-time object identification with a single pass through a neural network. The pipeline is trained on a simulated dataset containing approximately 60 Cherenkov rings per event, with significant overlaps on the detector plane. The performance meets the requirements of modern HEP experiments, achieving a reconstruction efficiency above 95% and a pion misidentification rate below 5% across a wide momentum range for all particle species.

AI keywords:

object detection; attention; computer vision; YOLO; edgeML

43

Signals in the Noise: Learning Invariant Representations for Ultra-Fast Edge Applications in High-Energy Physics

Authors: Brais Fernandez Rodino[None]; Chang Sun[None]; Marius Köppel[None]; Mattia Cerrato[None]; Patrick Odagiu[None]; Thea Aarrestad[1]

[1] *ETH Zurich*

Corresponding Authors: brais.fernandez.rodino@rai.usc.es, chsun@cern.ch, podagiu@ethz.ch, mkoepp@phys.ethz.ch

At CERN's Large Hadron Collider (LHC), hardware trigger systems are crucial in the first stages of data processing: they select a tiny fraction of the 40 million collision events per second for further analysis, within a few microseconds.
Machine Learning (ML) techniques being used more and more frequently to enable the efficient selection of extremely rare events.
These ML algorithms are deployed on custom computing platforms equipped with Field-Programmable Gate Arrays (FPGAs) to satisfy the extreme throughput and latency constraints.

Moreover, the loss of valuable data can be further minimised by decorrelating these algorithms from certain features in this data, thus ensuring that their performance remains robust across varying conditions.
For example, rare event searches at the LHC require methods that can effectively leverage both simulated and real-world data to train robust and accurate models.
Additionally, anomaly detection methods are highly susceptible to biases in the data, such as pile-up.
% Standard invariant representation techniques face significant challenges due to the stringent throughput and latency constraints.

In this work, we propose novel methods for learning invariant representations designed for ultra-fast inference and high-throughput edge applications.
Our contributions are fourfold:
i) we introduce mutual information-based measures to learn invariant representations for both supervised and unsupervised domains;
ii) we develop a new dataset to benchmark and evaluate these techniques;
iii) we implement a stochastic Bernoulli-layer in hardware description language (HDL) to enable seamless integration into FPGAs; and
iv) we demonstrate these techniques on two important physics applications: precision measurements of the $\tau \to 3\mu$ process in the LHCb experiment, and real-time anomaly detection in the CMS Experiment.

AI keywords:

Real-Time ML, anomaly detection, FPGAs

44

EagleEye: A general-purpose density anomaly detection method

Author: Andre Scaffidi[1]

[1] *SISSA*

Corresponding Author: ascaffid@sissa.it

Modern high-energy physics experiments generate massive, high-dimensional datasets that demand advanced strategies for anomaly detection. This talk presents *EagleEye*, a novel density-based method designed to compare two multivariate distributions on a point-by-point basis using a distribution-free approach rooted in Bernoulli and binomial statistics. *EagleEye's* deterministic framework, which analyzes local neighborhoods to pinpoint deviations, can be shown to outperform established techniques on challenging domain searches such as the LHC Olympics R&D dataset in a completely unsupervised manner, without the need to specify signal regions and/or control regions, or any other weakly-supervised prescription. In this talk I will discuss the statistical properties of *EagleEye*, detailing how the algorithm remains computationally efficient and parallelizable, whilst explaining how the method can locate anomolous regions of over/underdensity in feature space, and even estimate the total and local signal-to-background ratio. I will also show how *EagleEye* can be readily adapted to a diverse range of science tasks—from new particle searches to climate data.

AI keywords:

Anomaly detection; unsupervised learning; theory

45

Developing Artificial Intelligence in the Cloud: the AI_INFN Platform

Authors: Daniele Spiga[1]; Diego Ciangottini[2]; Diego Michelotto[3]; Giulio Bianchini[1]; Lucio Anderlini[1]; Rosa Petrini[1]; Stefano Dal Pra[1]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *INFN Perugia*

[3] *CNAF*

Corresponding Author: rosa.petrini@fi.infn.it

Machine Learning (ML) is driving a revolution in the way scientists design, develop, and deploy data-intensive software. However, the adoption of ML presents new challenges for the computing infrastructure, particularly in terms of provisioning and orchestrating access to hardware accelerators for development, testing, and production. The INFN-funded project AI_INFN ("Artificial Intelligence at INFN") aims at fostering the adoption of ML techniques within INFN use cases by providing support on multiple aspects, including the provisioning of AI-tailored computing resources. It leverages cloud-native solutions in the context of INFN Cloud, to share hardware accelerators as effectively as possible, ensuring the diversity of the Institute's research activities is not compromised. In this contribution, we provide an update on the commissioning of a Kubernetes platform designed to ease the development of GPU-powered data analysis workflows and their scalability on heterogeneous distributed computing resources, also using the offloading mechanism with Virtual Kubelet and InterLink API, in synergy with InterTwin. This setup can manage workflows across different resource providers, such as Leonardo CINECA, and hardware types, which are crucial for scientific use cases that require dedicated infrastructures for different parts of the workload. Initial test results, emerging case studies and integration scenarios will be presented with functional tests and benchmarks.

AI keywords:

MLOps; distributed computing; collaborative development platform

46

DeepExtractor: Time-domain reconstruction of signals and glitches in gravitational wave data with deep learning

Author: Tom Dooney[None]

Corresponding Author: tom.dooney@ou.nl

Gravitational wave (GW) interferometers, such as LIGO, Virgo, and KAGRA, detect faint signals from distant astrophysical events. However, their high sensitivity also makes them susceptible to background noise, which can obscure these signals. This noise often includes transient artifacts called "glitches", that can mimic genuine astrophysical signals or mask their true characteristics. Fast and accurate reconstruction of both signals and glitches is crucial for reliable scientific inference. In this study, we present \textit{DeepExtractor}, a deep learning framework that is designed to reconstruct signals and glitches with power exceeding interferometer noise, regardless of their source. We design DeepExtractor to model the inherent noise distribution of GW interferometers, following conventional assumptions that the noise is Gaussian and stationary over short time scales. It operates by predicting and subtracting the noise component of the data, retaining only the clean reconstruction of signal or glitch. Our innovative approach achieves superior generalization capabilities for arbitrary signals and glitches compared to methods that directly map inputs to the clean training waveforms. We focus on applications related to glitches and validate DeepExtractor's effectiveness through three experiments: (1) reconstructing simulated glitches injected into simulated detector noise, (2) comparing its performance with the state-of-the-art BayesWave algorithm, and (3) analyzing real data from the Gravity Spy dataset to demonstrate effective glitch subtraction from LIGO strain data. Our proposed model achieves a median mismatch of only 0.9 for simulated glitches, outperforming several deep learning baselines. Additionally, DeepExtractor surpasses BayesWave

in glitch recovery, offering a dramatic computational speedup by reconstructing one glitch sample in approximately 0.1 seconds on a CPU, compared to BayesWave's processing time of approximately one hour per glitch.

AI keywords:

time-series; denoising; reconstruction; model-agnostic; u-net

47

Physics-guided Machine Learning methods in QUBIC

Author: Leonora Kardum[None]

Corresponding Author: kardum@apc.in2p3.fr

Today, many physics experiments rely on Machine Learning (ML) methods to support their data analysis pipelines. Although ML has revolutionized science, most models are still difficult to interpret and lack clarity of the process with which they calculate results and the way they utilize information from used datasets. In this work, we introduce physics-guided ML methods that keep the reliability of traditional statistical techniques (e.g. minimization, likelihood analysis), which are accurate but often slow, and use the speed and efficiency of deep learning - without loosing interpretability. We show methods that offer insight into details of the dataset by informing the models with the underlying physics and analyzing information gain, allowing interpretability while also optimizing data usage. The approach is presented in the context of QUBIC, an unconventional experiment designed to investigate the Cosmic Microwave Background using bolometric interferometry, with the goal of detecting primordial gravitational waves. Methods are applied to the process of convolved map reconstruction, and show how physics-guided methods can aid in interpretability, parameter estimation, fitting, memory optimization, and more. This approach is not limited to cosmology, and can be applied in many areas of research.

AI keywords:

physics-guided machine learning; interpretable AI; map reconstruction; deconvolution

48

Physics Instrument Design with Reinforcement Learning

Authors: Nicola Serra[1]; Patrick Owen[1]; Shah Rukh Qasim[1]

[1] *University of Zurich*

Corresponding Author: shah.rukh.qasim@cern.ch

We present a case for the use of Reinforcement Learning (RL) for the design of physics instrument as an alternative to gradient-based instrument-optimization methods (arXiv:2412.10237). It's applicability is demonstrated using two empirical studies. One is longitudinal segmentation of calorimeters and the second is both transverse segmentation as well longitudinal placement of trackers in a spectrometer. Based on these experiments, we propose an alternative approach that offers unique advantages over differentiable programming and surrogate-based differentiable design optimization methods. First, Reinforcement Learning (RL) algorithms possess inherent exploratory capabilities, which help mitigate the risk of convergence to local optima. Second, this approach eliminates the necessity of constraining the design to a predefined detector model with fixed parameters. Instead, it allows for the flexible placement of a variable number of detector components and facilitates discrete decision-making. We then discuss the road map of how this idea can be extended into designing very

complex instruments. The presented study sets the stage for a novel framework in physics instrument design, offering a scalable and efficient framework that can be pivotal for future projects such as the Future Circular Collider (FCC), where most optimized detectors are essential for exploring physics at unprecedented energy scales.

More: https://arxiv.org/abs/2412.10237

AI keywords:

Reinforcement Learning; Machine Learning; Physics Instrument Design; Detector Optimization; AI-Based Experiment Design

49

End-to-end optimization of a Muon Collider calorimeter

Authors: Federico Nardi[1]; Tommaso Dorigo[1]; Julien Donini[2]; Long Chen[3]; Xuan Tung Nguyen[4]; Alessandro Breccia[None]; Enrico Lupi[1]; Fredrik Sandin[5]; Max Aehle[3]; Muhammad Awais[1]; Nicholas R Gauger[3]; Pietro Vischia[6]; Riccardo Carroccio[None]; Fabio Cufino[7]; Kylian Schmidt[8]; Emanuele Coradin[9]; Jan Kieseler[10]; Andrea Giammanco[11]; Mia Tosi[1]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *Université Clermont Auvergne (FR)*

[3] *University of Kaiserslautern-Landau*

[4] *INFN Padova, University of Kaiserslautern-Landau*

[5] *Lulea Techniska Universitet*

[6] *Universidad de Oviedo and Instituto de Ciencias y Tecnologías Espaciales de Asturias (ICTEA)*

[7] *University of Bologna*

[8] *KIT*

[9] *University of Padova*

[10] *DESY*

[11] *UCLouvain, CP3*

Corresponding Author: federico.nardi@pd.infn.it

Setup design is a critical aspect of experiment development, particularly in high-energy physics, where decisions influence research trajectories for decades. Within the MODE Collaboration, we aim to generalize Machine Learning methodologies to construct a fully differentiable pipeline for optimizing the geometry of the Muon Collider Electromagnetic Calorimeter.

Our approach leverages Denoising Diffusion Probabilistic Models (DDPMs) for signal generation and Graph Neural Networks (GNNs) for photon reconstruction in the presence of Beam-Induced Background from muon decays. Through automatic differentiation, we integrate these components into a unified framework that enables end-to-end optimization of calorimeter configurations. We present the structure of this pipeline, discuss key generation and reconstruction techniques, and showcase the latest results on proposed geometries.

AI keywords:

generative models; graph neural networks; automatic differentiation; simulation-based optimization

50

Interaction-Aware and Domain-Invariant Representation Learning for Inclusive Flavour Tagging

Authors: Mirko Bunse[1]; Quentin Führing[1]

[1] *Lamarr Institute for Machine Learning and Artificial Intelligence, Dortmund, Germany*

Corresponding Author: mirko.bunse@cs.tu-dortmund.de

Measurements of neutral, oscillating mesons are a gateway to quantum mechanics and give access to the fundamental interactions of elementary particles. For example, precise measurements of CP violation in neutral B mesons can be taken in order to test the Standard Model of particle physics. These measurements require knowledge of the B-meson flavour at the time of its production, which cannot be inferred from its observed decay products. Therefore, multiple LHC experiments employ machine learning-based algorithms, so-called flavour taggers, to exploit particles that are produced in the proton-proton interaction and are associated with the signal B meson to predict the initial B flavour. A state-of-the-art approach to flavour tagging is the inclusive evaluation of all reconstructed tracks from the proton-proton interaction using a Deep Set neural network.

Flavour taggers are desired to achieve optimal performance for data recorded from proton-proton interactions while being trained with a labelled data sample, i.e., with Monte Carlo simulations. However, the limited knowledge of QCD processes introduces inherent differences between simulation and recorded data, especially in the quark-fragmentation processes that are relevant for flavour tagging. Existing flavour taggers neither model these differences nor do they model interactions between tracks explicitly, being at danger of overfitting to simulations, of not providing optimal performance for physics analyses, and of requiring a careful calibration on data.

We present an inclusive flavour tagger that builds on set transformers (to model particle interactions via set attention) and on domain-adversarial training (to mitigate differences between data sources). These foundations allow the tagger to learn intermediate data representations that are both interaction-aware and domain-invariant, i.e., they capture the interactions between tracks and do not allow for an overfitting to simulations. In our benchmark, we increase the statistical power of flavour-tagged samples by 10% with respect to the usage of deep sets, thus demonstrating the value of interaction-aware and domain-invariant representation learning.

AI keywords:

set transformers; unsupervised domain adaptation; multiple-instance learning; dataset shift; representation learning

51

Learning to Optimize Cosmic Initial Conditions with Non-Differentiable Structure Formation Models

Author: Ludvig Doeser[1]

[1] *Stockholm University*

Corresponding Author: ludvig.doeser@fysik.su.se

Traditional gradient-based optimization and statistical inference methods often rely on differentiable models, making it challenging to optimize models with non-differentiable components. In this talk, I'll introduce Learning the Universe by Learning to Optimize (LULO), a novel deep learning-based framework designed to fit non-differentiable simulators at non-linear scales to data. By employing a neural optimizer in an iterative scheme while keeping full physics simulations in the loop, LULO ensures both scalability and reliability. In particular, I will demonstrate how LULO accurately reconstructs the 3D cosmological initial conditions and corresponding late-time structures from mock data generated by a non-differentiable and non-linear simulator pipeline. Our method provides a

promising path forward for performing detailed field-level inference with next-generation galaxy clustering survey data, without the need for differentiable models.

AI keywords:

Learning-to-optimize; neural optimizers; convolutional neural networks; physics-in-the-loop

52

AI-enabled Insights Into Galaxy Evolution

Author: Sinan Deger[1]

Co-authors: Boris Leistedt [2]; Daniel Mortlock [3]; Hiranya Peiris [1]; Joel Leja [4]; Justin Alsing [5]; Stephen Thorp [5]

[1] *Institute of Astronomy, University of Cambridge*

[2] *Astrophysics Group, Imperial College London*

[3] *Astrophysics Group & Department of Mathematics, Imperial College London*

[4] *Department of Astronomy & Astrophysics, The Pennsylvania State University*

[5] *Oskar Klein Centre, Stockholm University*

Corresponding Author: sinan.deger@ast.cam.ac.uk

Understanding the properties of galaxy populations and their evolution is directly linked to the success of large-scale surveys such as The Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST). Galaxy spectral energy densities (SEDs) encode these properties, but SED observations for a broad wavelength range via spectroscopy is a time consuming practice. LSST will perform photometric observations measuring the integrated light through a handful of bandpasses, effectively trading information loss for observational speed. An early obstacle we need to overcome is the inference of accurate distances (redshifts) from these limited photometric observations. We addressed this challenge with a forward modeling framework, pop-cosmos, utilizing AI to infer redshifts jointly with constraining galaxy properties. The complex space of galaxy properties is calibrated by fitting a population model parametrized by a diffusion model to photometric data. This high-dimensional fitting, complete with data-driven noise modeling and flexible selection effects, is achieved via a novel use of simulation-based inference. We fit this model to photometric data from one of the deepest extragalactic surveys, COSMOS2020. As a result samples from our trained model are realizations of the galaxy population within the survey limiting magnitude and up to a redshift of 4, capturing 90% of cosmic time. Therefore pop-cosmos unlocks a medium for the study of galaxy evolution that was not possible before. I will be presenting population-level results, specifically a comprehensive look at the star formation histories of galaxies inferred from our model. This will entail a detailed picture of the stellar mass assembly of galaxies together with the evolution of their metallicity and colors. I will showcase key results like the cosmic star formation rate density we derive using our model, and how these compare with results from the extensive literature.

AI keywords:

simulation-based inference, diffusion models, generative models

54

Anomaly Detection with Machine Learning in Time Series

Author: Andrea Adelfio[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: andrea.adelfio@pg.infn.it

Multimessenger astrophysics relies on multiple observational data channels, necessitating efficient methods for analyzing events of astrophysical origin. With the continuous increase in both volume and complexity of data from modern observatories, advanced Machine Learning techniques have become very useful for identifying and classifying signals effectively.

My project aim at developing a framework using Machine Learning techniques to analyze time series data. The use case that will be presented regards the data from the Anti-Coincidence Detector (ACD) onboard the Fermi Gamma-ray Space Telescope. The primary objective is to enhance the detection of high-energy transient events, such as Gamma-Ray Bursts (GRBs) and other astrophysical signals. An ensamble of Neural Networks models may be employed to model and predict the temporal structure of the ACD background data. The network's predictions serve as a baseline for implementing a triggering algorithm designed for anomaly detection. By identifying significant deviations from the predicted background, the system effectively flags potential astrophysical transients in the ACD time series data.

In addressing challenges such as noise variability, this work explores advanced approaches to refine anomaly detection thresholds, by characterizing the noise amplitude in the data. Bayesian Neural Networks (BNNs) are highlighted as a promising method to dynamically adapt thresholds based on the noise characteristics of the data, offering a robust alternative to traditional fixed-threshold methods. These developments demonstrate a robust and adaptable framework for signal detection, applicable across various datasets and observatories in multimessenger astrophysics.

AI keywords:

Anomaly detection; bayesian neural networks; time series predictions;

55

Simulation-based inference for parameter estimation of high-redshift sources with the Einstein telescope

Author: Filippo Santoliquido[1]

Co-authors: Jacopo Tissino [2]; Jan Harms [2]; Marica Branchesi [2]; Ulyana Dupletsa [2]

[1] *Gran Sasso Science Institute (GSSI)*

[2] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: filippo.santoliquido@gssi.it

The Einstein Telescope (ET) will be a key instrument for detecting gravitational waves (GWs) in the coming decades. However, analyzing the data and estimating source parameters will be challenging, especially given the large number of expected detections—between 10^4 and 10^5 per year —which makes current methods based on stochastic sampling impractical. In this work, we use DingoIS to perform Neural Posterior Estimation (NPE), a simulation-based inference technique that leverages normalizing flows to approximate the posterior distribution of detected events. After training, inference is fast, requiring only a few minutes per source, and accurate, as validated through importance sampling. We process 1000 randomly selected injections and achieve an average sample efficiency of $\sim 13\%$, which increases to $\sim 18\%$ ($\sim 20\%$) if we consider only sources merging at redshift $z > 4$ ($z > 10$). To confirm previous findings on ET ability to estimate parameters for high-redshift sources, we compare NPE results with predictions from the Fisher information matrix (FIM) approximation. We find that FIM underestimates sky localization errors by a factor of > 8, as it does not capture the multimodalities in sky localization introduced by the geometry of the triangular detector. On the contrary, FIM overestimates the uncertainty in luminosity distance by a factor of ~ 3 on average when the injected luminosity distance $d_{\mathrm{L}}^{\mathrm{inj}} > 10^5$ Mpc, further confirming that ET will be particularly well suited for studying the early Universe.

AI keywords:

simulation-based inference, normalizing flows, GPUs

56

Accurate and computationally inexpensive 21 cm maps with diffusion models

Author: Satvik Mishra[1]

Co-authors: Matteo Viel [2]; Roberto Trotta [3]

[1] *SISSA*

[2] *Istituto Nazionale di Fisica Nucleare*

[3] *SISSA, Italy*

Corresponding Author: samishr@sissa.it

Modeling the distribution of neutral hydrogen is essential for understanding the physics of structure formation and the nature of dark matter, but accurate numerical simulations are computationally expensive. We describe a novel Variational Diffusion Model (VDM), built on a 3D CNN attention U-Net architecture, which we use in concert with the CAMELS simulation suite to generate accurate 21 cm intensity maps from computationally inexpensive dark-matter-only N-body simulations. Our model delivers both halo mass density and neutral hydrogen maps in large cosmological boxes of 25 $(\mathrm{Mpc}/h)^3$, from which we derive highly accurate 21 cm power spectra up to scales as small as $k = 10\ h\ \mathrm{Mpc}^{-1}$. We discuss the strengths of this method compared to existing approaches (e.., its ease of training) and highlight potential applications in the upcoming SKA era for cosmological studies and for distinguishing between cold and warm dark matter models.

AI keywords:

Diffusion models, emulation, conditional generation.

57

Synthetic Data Generation with Lorenzetti for Time Series Anomaly Detection in High-Energy Physics Calorimeters

Author: Laura Boggia[1]

Co-author: Bogdan MALAESCU [2]

[1] *LPNHE*

[2] *CNRS, LPNHE*

Corresponding Author: laura.boggia@cern.ch

Anomaly detection in multivariate time series is crucial to ensure the quality of data coming from a physics experiment. Accurately identifying the moments when unexpected errors or defects occur is essential but challenging, as the types of anomalies are unknown beforehand and reliably labeled data is scarce. Additionally, the multi-dimensional nature of time series data adds to the problem's complexity, as the correlations between different dimensions must be considered.

To address the lack and unreliability of labeled data, we produce synthetic data with the Lorenzetti Simulator, a general-purpose framework simulating a high energy experiment, where we introduce

artificial anomalies in the calorimeter. By introducing artificial anomalies in the calorimeter, we can systematically evaluate the effectiveness and sensitivity of anomaly detection methods, including transformer-based and other deep learning models. The approach employed here is generic and can be adapted to various detector architectures and their potential defects.

AI keywords:

anomaly detection, transformers, unsupervised learning, dataset creation

58

Transformer-based compression of big data

Author: James Smith[1]

[1] *University of Manchester*

Corresponding Author: jl.smith@cern.ch

The storage, transmission and processing of data is a major challenge across many fields of physics and industry. Traditional generic data compression techniques are lossless, but are limited in performance and require additional computation.

BALER [1,2] is an open-source autoencoder-based framework for the development of tailored lossy data compression models suitable for data from multiple disciplines. BALER models can also be used in FPGAs to compress live data from detectors or other sources, potentially allowing for massive increases in network throughput.

This presentation will introduce BALER and discuss recent developments and results. These include the development and analysis of new transformer-based autoencoder models, the application of BALER to particle physics analyses and the resulting affect on discovery significance, and the evaluation of the energy consumption and sustainability of differing autoencoder compression models.

BALER is developed by a cross-disciplinary team of physicists, engineers, computer scientists and industry professionals, and has received substantial contributions from a large number of master' s and doctoral students. BALER has received support from industry both in providing datasets to develop BALER, and to transfer industry best practices.

[1] https://arxiv.org/pdf/2305.02283.pdf
[2] https://github.com/baler-collaboration/baler

AI keywords:

compression; sustainability; transformers; autoencoders; big data

59

Efficient combination of Likelihood-Based and Simulation-Based Inference: application to Planck+Euclid analyses

Author: Guillermo Franco Abellán[1]

Co-authors: Christoph Weniger [2]; Oleg Savchenko [1]

[1] *GRAPPA Institute, University of Amsterdam*

[2] *GRAPPA, University of Amsterdam*

Corresponding Author: g.francoabellan@uva.nl

Simulation-based inference (SBI) allows amortized Bayesian inference for simulators with implicit likelihoods. However, some explicit likelihoods cannot easily be reformulated as simulators, hindering its integration into combined analyses within the SBI framework. One key example in cosmology is given by the Planck CMB likelihoods. In this talk, I will present a simple method to construct an effective simulator for any explicit likelihood using posterior samples from a previously converged MCMC run. To illustrate this method, I conduct a joint cosmological analysis that combines the full Planck CMB likelihoods with a simulator for an Euclid-like galaxy survey. This result opens up the possibility of performing massive global scans combining explicit and implicit likelihoods in a hyper-efficient way.

AI keywords:

Simulation-based inference; Likelihood-based inference; AI for handling nuisance parameters

61

Advancing b-Decay Reconstruction via Probability-Weighted Message Passing in Heterogeneous GNNs

Author: William Sutcliffe[1]

Co-authors: Abhijit Mathad [2]; Jonas Eschle ; Julián García Pardiñas [3]; Marta Calvi [4]; Nicola Serra [1]; Simone Capelli [4]

[1] *University of Zurich*

[2] *University of Warwick*

[3] *University of Milano-Bicocca*

[4] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: william.sutcliffe@physik.uzh.ch

Graph neural networks (GNNs) have become state-of-the-art tools across diverse scientific disciplines due to their ability to model complex relationships in datasets that lack simple spatial or sequential structures. In this talk, we present recent advancements in the deep full event interpretation (DFEI) framework [García Pardiñas, J., et al. Comput. Softw. Big Sci. 7 (2023) 1, 12]. The DFEI framework leverages a novel GNN-based hierarchical reconstruction of b-hadron decays within the hadronic collision environment of the LHCb experiment. We will discuss significant performance improvements achieved through a novel end-to-end node and edge pruning GNN architecture that employs a novel probability-weighted message passing to exploit the intrinsic structure of decay graphs. Finally, we introduce a more flexible heterogeneous GNN approach with multi-task learning that not only enhances reconstruction performance but also supports additional critical tasks simultaneously, such as precisely associating reconstructed b-hadrons with their corresponding primary vertices.

AI keywords:

Graph Neural Network, Heterogeneous GNN, Multi-task learning, Message passing

62

Can SBI unlock the LISA global fit?

Author: James Alvey[1]

[1] *University of Cambridge*

Corresponding Author: jbga2@cam.ac.uk

The Laser Interferometer Space Antenna (LISA) will provide an unprecedented window into the gravitational wave sky. However, it also presents a serious data analysis challenge to separate and classify various classes of deterministic sources, instrumental noise, and potential stochastic backgrounds. This "global fit" problem presents an extremely high-dimensional inference task that sits right at the limits of traditional likelihood-based methods. In this talk, I will explore how SBI could transform our approach to the LISA global fit.

AI keywords:

simulation-based inference, normalising flows, bayesian inference

63

Cause-mic Universe : Causal Approaches probing Solar and Astrophysical Variability

Author: Nachiketa Chakraborty[1]

[1] *University of Reading*

Corresponding Author: n.chakraborty@reading.ac.uk

Astrophysical sources vary across vast timescales, providing insight into extreme dynamical phenomena, from solar outbursts to distant AGNs and GRBs. These time-varying processes are often complex, nonlinear, and non-Gaussian, making it difficult to disentangle underlying causal mechanisms, which may act simultaneously or sequentially. Using solar variability and AGNs as examples, we demonstrate how causal inference and graphical models, supported by synthetic time-series data, help unravel these processes. In solar variability, understanding solar wind and flares is crucial not only for physics but also for space weather forecasting, where supervised and unsupervised machine learning methods have made significant progress. Causal diagnostics enhance interpretability and feature selection. In extragalactic sources like AGNs and GRBs, causal measures enable insights into variability mechanisms and fundamental physics, such as Lorentz Invariance, through precise lag estimates.

AI keywords:

causal inference, graphical models, time-series, interpretable forecasting, simulations

64

astroLLM: An AI-Powered Multi-Agent Research Assistant for Multiwavelength and Multimessenger Astrophysics

Author: Narek Sahakyan[1]

Co-authors: Damien Begue [2]; Mher Khachatryan [3]

[1] *ICRANet Armenia IO*

[2] *Bar Ilan University*

[3] *ICRANet-Armenia*

Corresponding Author: narek.sahakyan@icranet.org

The rapid increase in astrophysical data and scientific literature poses a significant challenge for researchers seeking to efficiently process, analyze, and extract meaningful insights. While traditional Large Language Models (LLMs) primarily focus on text-based tasks, there is a pressing need for advanced AI-driven frameworks that seamlessly integrate literature review, data retrieval, and theoretical modeling. To address this, we introduce astroLLM, an AI-powered multi-agent research assistant, in which specialized agents handle literature retrieval, data analysis, and theoretical modeling, collectively providing domain-specific knowledge while enabling access to extensive datasets and novel modeling tools. astroLLM leverages Retrieval-Augmented Generation (RAG) to synthesize information from extensive scientific literature and employs chain-of-thought reasoning to enhance its analytical capabilities and domain-specific applications. It interacts with external computational tools and astrophysical databases, including the Markarian Multiwavelength Data Center (MMDC), providing seamless retrieval of multiwavelength and multimessenger data from diverse catalogs. The framework also facilitates theoretical modeling through Convolutional Neural Networks (CNNs), which are trained on outputs from leptonic and lepto-hadronic models, thus enabling accurate spectral energy distribution (SED) modeling, parameter estimation, and computational optimization. As a first step, astroLLM v1.0 is tailored for blazar research, with future expansions planned to progressively include other astrophysical source classes. This presentation will discuss the architecture of astroLLM, its integration with external astrophysical resources, and its key applications in high-energy astrophysics. We will also highlight the advantages of AI-driven frameworks in multiwavelength and multimessenger studies and explore the broader impact of LLM-assisted research in astrophysics.

AI keywords:

Foundation Models; Multi-Agent Systems; Retrieval-Augmented Generation; Explainable AI

65

Nested sampling neural ratio estimation for gravitational waves

Author: Metha Prathaban[1]

[1] *University of Cambridge*

Corresponding Author: myp23@cam.ac.uk

Gravitational wave astronomy in the era of third-generation (3G) detectors will pose significant computational challenges. While standard parameter estimation methods may remain technically feasible, the demand for more efficient inference algorithms is on the rise. We present a sequential neural simulation-based inference algorithm that merges neural ratio estimation (NRE) with nested sampling (NS) for advanced analysis in the field of gravitational waves. Building upon the principles of PolySwyft, which introduced the NSNRE algorithm, this framework leverages the power of JAX and seamlessly integrates the NRE algorithm with a state-of-the-art blackjax implementation of nested sampling. This integrated approach enables efficient and accurate parameter estimation, and will enable the continuation of breakthrough science in the 3G era and beyond.

AI keywords:

simulation-based inference

66

Design of the SHiP's Muon Shield with Machine Learning

Authors: Luis Felipe Cattelan[1]; Shah Rukh Qasim[1]

Co-authors: Guglielmo Frisella ; Melvin Liebsch [2]; Nicola Serra [1]; Patrick Owen [1]; massimiliano ferro-luzzi [2]

[1] *University of Zurich*

[2] *CERN*

Corresponding Author: luis.felipe.cattelan@physik.uzh.ch

The SHiP experiment is a proposed fixed-target experiment at the CERN SPS aimed at searching for feebly interacting particles beyond the Standard Model. One of its main challenges is reducing the large number of muons produced in the beam dump, which would otherwise create significant background in the detector. The muon shield, a system of magnets designed to deflect muons away from the detector acceptance, must be optimized to achieve the best possible background suppression while keeping costs low. In this work, we extend previous optimization efforts by explicitly incorporating geometric constraints and the on-the-fly simulation of field maps, providing a more realistic evaluation of the shield's performance. This optimization includes both the case of using normal-conducting magnets and the potential for incorporating superconducting magnets, which may offer significant advantages in terms of performance for the muon shielding. We start by re-optimizing the parameters of the magnets through merging Bayesian Optimization (BO) and local optimization with surrogate models. Moreover, to address the computational cost of full-scale simulations, we train the surrogate model on a carefully selected reduced sample, allowing it to predict results for the full sample without sacrificing accuracy. Building upon this, we take a step further by generalizing the problem as a dynamic magnet and veto allocation task. Given the sequential nature of this challenge, we propose to apply reinforcement learning (RL) to iteratively select and arrange magnets in an optimal configuration, offering a more flexible and adaptive solution. We compare both approaches and evaluate their impact on the muon shield's performance and computational efficiency.

AI keywords:

Bayesian Optimization, surrogate models, gaussian processes, reinforcement learning

67

Enhanced Gravitational Wave Detection with Normalizing-Sequential Flow

Author: Huifang Lyu[1]

Co-authors: Christoph Weniger [1]; Uddipta Bhardwaj [1]

[1] *University of Amsterdam*

Corresponding Author: h.lyu@uva.nl

Gravitational waves coming from mergers of binary compact objects observed by detectors like LIGO and Virgo have profoundly transformed our understanding of the universe. However, as future detectors become more sensitive, it becomes increasingly difficult to effectively identify and characterize fainter and more complex signals hidden in noisy data. To address this problem, we use a machine learning approach, Normalizing-Sequential Flow (NSFlow), which leverages sequential normalizing flows to efficiently model complex posterior distributions and enhance inference speed. This approach provides faster sampling, more accurate uncertainty quantification, and better scalability for large-scale gravitational wave datasets. We will use this method to reconstruct each parameter posterior for binary black hole mergers and demonstrate its accuracy and robustness in direct comparison with established likelihood-based methods.

AI keywords:

simulation-based inference, normalizing-sequential flow, probability density estimate

69

Tracking Transformer synthesis for low-latency FPGA deployment

Authors: Amir Yousefzadeh[1]; Arjan Blankestijn[1]; Uraz Odyurt[1]

[1] *University of Twente*

Corresponding Author: uodyurt@nikhef.nl

The Transformer Machine Learning architecture has been gaining considerable momentum in recent years. Computational High-Energy Physics tasks such as jet tagging and particle track reconstruction (tracking), have either achieved proper solutions, or reached considerable milestones using Transformers. On the other hand, the use of specialised hardware accelerators, especially FPGAs, is an effective method to achieve online, or pseudo-online latencies.

The development and integration of Transformer-based machine learning on FPGAs is still ongoing, and the support from current tools is very limited. Additionally, FPGA resources present a significant constraint. Considering the model size alone, while smaller models can be deployed directly, larger models are to be partitioned in a meaningful and ideally automated way. We aim to develop methodologies and tools for monolithic, or partitioned Transformer synthesis, specifically targeting inference. Our primary use-case involves machine learning models for tracking, derived from the TrackFormers project. We strive for lower latencies compared to GPU deployments.

AI keywords:

FPGA deployment; Transformer synthesis; Inference latency

70

Reinforcement Learning for background determination in particle physics

Author: Guillermo Hijano[1]

Co-authors: Alex Marshall [2]; andrea mauri [1]; Davide Lancierini [1]; Hanae Tilquin [3]; Konstantinos Petridis [2]; Mitesh Patel [3]; Nicola Serra [1]; Patrick Owen [1]; Shah Rukh Qasim [1]; William Sutcliffe [1]

[1] *University of Zurich*

[2] *University of Bristol*

[3] *Imperial College London*

Corresponding Author: guillermo.hijano@cern.ch

Experimental studies of ⊠-hadron decays face significant challenges due to a wide range of backgrounds arising from the numerous possible decay channels with similar final states. For a particular signal decay, the process for ascertaining the most relevant background processes necessitates a detailed analysis of final state particles, potential misidentifications, and kinematic overlaps which, due to computational limitations, is restricted to the simulation of only the most relevant backgrounds. Moreover, this process typically relies on the physicist's intuition and expertise, as no systematic

method exists. This work presents a novel approach that utilises Reinforcement Learning to overcome these challenges by systematically determining the critical backgrounds affecting ⊠-hadron decay measurements. Our method further incorporates advanced Artificial Intelligence models and techniques to enhance background identification accuracy: a transformer model is employed to handle token sequences representing decays, a Graph Neural Network is used for predicting Branching Ratios (BRs), and Genetic Algorithms are utilised as an auxiliary tool to efficiently explore the action space, among others.

AI keywords:

Reinforcement Learning; transformers; Graph Neural Networks; Genetic Algorithms;

71

Pulse pile-up reconstruction using 1D-CAE for signal discrimination in nuclear experiments

Author: Jose Manuel Deltoro Berrio[1]

Co-authors: Alain Goasduff [2]; Andres F. Gadea Raga [3]; Grześ Jaworski [4]; Vicente Gonzalez [5]

[1] *IFIC (Instituto de Física Corpuscular)*

[2] *INFN Laboratori Nazionali di Legnaro*

[3] *IFIC CSIC-University of Valencia*

[4] *Heavy Ion Laboratory, University of Warsaw, Poland*

[5] *ETSE - Universitat de València*

Corresponding Author: jose.m.deltoro@uv.es

Pulse pile-up is a common issue in nuclear spectroscopy and nuclear reaction studies, degrading energy and timing accuracy in particle identification. This work presents a novel method for reconstructing pile-up events using a one-dimensional convolutional autoencoder (1D-CAE). The method effectively separates and reconstructs overlapping pulses, enabling acceptance of these events and allowing for signal discrimination. This technique successfully reconstructs over 80% of the pile-up events acquired on a 64-bit processor. The proposed reconstruction technique has been successfully implemented on the 32-bit microprocessor of the PINQZ2 platform (ARM Cortex-A9), achieving similar performance as implementations on 64-bit processors. Furthermore, ongoing work is adapting the model to the Artix-7 FPGA of the PINQZ2 using the HLS4ML Python package. The results will be presented at the conference. This advancement paves the way for real-time pulse reconstruction in high-rate nuclear experiments, significantly enhancing data acquisition and analysis accuracy.

AI keywords:

Pile-up, real-time, FPGAs, one-dimensional convolutional autoencoder, machine learning

72

Probing the Parameter Space of Axion-Like Particles Using Simulation-Based Inference

Author: Pooja Bhattacharjee[None]

Corresponding Author: pooja.bhattacharjee@ung.si

Axion-like particles (ALPs) appear in various extensions of the Standard Model and can interact with photons, leading to ALP-photon conversions in external magnetic fields. This phenomenon can introduce characteristic energy-dependent “wiggles”in gamma-ray spectra. The Cherenkov Telescope Array Observatory (CTAO) is the next-generation ground-based gamma-ray observatory, designed to enhance sensitivity and energy coverage (20 GeV –300 TeV) over current Imaging Atmospheric Cherenkov Telescopes (IACTs) and offers an excellent opportunity to study such effects.

In this work, we employ Simulation-Based Inference (SBI) to explore the parameter space of ALPs. Additionally, we investigate whether this inference method can produce accurate ALP exclusion limits comparable to those reported in previous studies that use the classical likelihood-ratio approach. Through this approach, we seek to yield robust constraints on ALP-photon interactions and make substantial advancements in this field.

AI keywords:

Simulation Based Inference, Neural Ratio Estimation, Bayesian Neural Networks, Posterior Estimation

73

Point Cloud Machine Learning for Cell-to-Track Association: Enhancing Event Reconstruction in High Energy Physics

Author: Luca Clissa[1]

Co-authors: Joshua Himmens [2]; Iacopo Vivarelli [1]; Maximilian Swiatlowski [2]

[1] *University of Bologna; Istituto Nazionale di Fisica Nucleare*

[2] *TRIUMF*

Corresponding Author: luca.clissa2@unibo.it

The ATLAS detector at CERN’s Large Hadron Collider (LHC) is a complex system composed of multiple subdetectors, each designed to capture complementary aspects of particle interactions. Thus, accurate understanding of the physical phenomena under study requires effectively combining information from these components.

This work focuses on the key challenge of associating data from the inner tracker with the corresponding energy deposits in the calorimeter.

Current approaches tackle this problem in a modular fashion. First, hits in the tracker and calorimeter are reconstructed separately into tracks and topo-clusters, respectively. Second, tracks are iteratively associated with topo-clusters to improve particle identification and energy calibration in subsequent reconstruction steps.
However, this strategy relies on rigid algorithms tuned to address reasonably well the most common known scenarios. Moreover, they fail to fully exploit the complementary information provided by the two subsystems. Consequently, errors in track and topo-cluster reconstruction are propagated to later stages, hampering accurate association.

To overcome these limitations, we propose a PointNet model for cell-to-track association. This approach enables the direct integration of tracking information into energy reconstruction, facilitates fine-grained association between tracks and individual calorimeter cells rather than whole topo-clusters, and leverages efficient point cloud data representations.
This methodology demonstrates promising results for enhancing offline reconstruction, particularly relevant for addressing the increased detector occupancy and event complexity anticipated in the High-Luminosity LHC era.

AI keywords:

PointNet, Point Cloud Segmentation Task; Class Imbalance; Robust Loss and Metrics

74

Anomaly preserving contrastive neural embeddings for end-to-end model-independent searches at~the~LHC

Authors: Ekaterina Govorkova[1]; Gaia Grosso[2]; Kyle Metzger[3]; Lana Xu[1]; Mia Sodini[1]; Philip Harris[1]; Thea Aarrestad[3]

[1] *MIT*

[2] *IAIFI/MIT/Harvard*

[3] *ETH Zurich*

Corresponding Author: kmetzger@student.ethz.ch

Anomaly detection —identifying deviations from Standard Model predictions —is a key challenge at the Large Hadron Collider due to the size and complexity of its datasets. This is typically addressed by transforming high-dimensional detector data into lower-dimensional, physically meaningful features. We tackle feature extraction for anomaly detection by learning powerful low-dimensional representations via contrastive neural embeddings. This approach preserves potential anomalies indicative of new physics and enables rare signal extraction using novel machine learning-based statistical methods for signal-independent hypothesis testing. We compare supervised and self-supervised contrastive learning methods, for both MLP- and Transformer-based neural embeddings, trained on the kinematic observables of physics objects in LHC collision events. The learned embeddings serve as input representations for signal-agnostic statistical detection methods in inclusive final states, achieving up to a ten fold improved detection performance over the original feature representation and up to 35% improvement over using a physics-informed selections of the same dimensionality. We achieve significant improvement in discovery power for both rare new physics signals and rare Standard Model processes across diverse final states, demonstrating its applicability for efficiently searching for diverse signals simultaneously. We study the impact of architectural choices, contrastive loss formulations, supervision levels, and embedding dimensionality on anomaly detection performance. We show that the optimal representation for background classification does not always maximize sensitivity to new physics signals, revealing an inherent trade-off between background structure preservation and anomaly enhancement. Our findings demonstrate that foundation models for particle physics data hold significant potential for improving neural feature extraction, enabling scientific discovery in inclusive final states at collider experiments.

AI keywords:

contrastive learning, anomaly detection, data representation, goodness of fit, neural embeddings

75

Enhancing low energy reconstruction and classification in KM3NeT/ORCA with transformers

Author: Ivan Mozun Mateo[1]

[1] *LPC CAEN / CNRS / IN2P3*

Corresponding Author: mozun@lpccaen.in2p3.fr

KM3NeT is a new research infrastructure housing the next generation of neutrino telescopes in the Mediterranean deep sea. This facility comprises two detectors: KM3NeT/ARCA and KM3NeT/ORCA, consisting of vertically-arranged detection units, 230 and 115, respectively, each equipped with 18

digital optical modules. The photomultipliers within each optical module detect Cherenkov light emitted by charged particles propagating in the seawater. KM3NeT/ARCA is optimized for the search of astrophysical neutrino sources in the range of TeV to PeV; whereas KM3NeT/ORCA is used to study the neutrino oscillation phenomena in the 1-100 GeV energy range.

The current KM3NeT/ORCA telescope, with 24 deployed detection units, is still under construction and has not yet reached its full potential in neutrino reconstruction capability. When training any deep learning model, no explicit information about the physics nor the detector is provided, nor is it already embedded in the data, thus remaining unknown to the model. This study demonstrates the efficacy of transformer models, as large representation models, on retaining valuable information from the simulations of the complete detector when evaluating data from various smaller KM3NeT/ORCA configurations. The study leverages the strengths of transformers, with respect to other models, by incorporating attention masks inspired by the physics and detector design. This allows to filter out irrelevant background light pulses and focusing on those resulting from a neutrino interaction, at the same time it captures the physics measured on the telescope.

AI keywords:

transformers, transfer learning, multi-task inference

76

Reconstructing source motion from gravitational wave strain

Author: Joe Bayley[None]

Corresponding Author: joseph.bayley@glasgow.ac.uk

The identification of burst gravitational wave signals can be challenging due to the lack of well-defined waveform models for various source types. In this study, we propose a novel approach to understanding the mass dynamics of the system that produced the burst signal by reconstructing the possible motions of masses that could generate the detected waveform within certain constraints.

Our method involves training a normalising flow on random motions of masses, which allows us to explore a wide range of possible mass configurations and their corresponding waveforms. By employing this technique, we can reconstruct all feasible mass dynamics that may have contributed to the observed gravitational wave signal.

We present some results from our analysis, as well as discuss its limitations and potential extensions. These methods could provide valuable insights into the nature of burst gravitational wave sources and point towards a better understanding of their astrophysical origins.

AI keywords:

simulation-based-inference: transformers: generative AI: normalising flows

77

Conditional Deep Generative Models for Simultaneous Simulation and Reconstruction of Entire Events

Authors: Benjamin Nachman[1]; Dmitrii Kobylianskii[2]; Eilam Gross[2]; Etienne Dreyer[2]; Vinicius Mikuni[3]

[1] *Lawrence Berkeley National Laboratory*

[2] *Weizmann Institute of Science*

[3] *LBNL*

Corresponding Author: dmitry.kobylyansky@weizmann.ac.il

We extend the Particle-flow Neural Assisted Simulations (*Parnassus*) framework of fast simulation and reconstruction to entire collider events. In particular, we use two generative Artificial Intelligence (genAI) tools, conditional flow matching and diffusion models, to create a set of reconstructed particle-flow objects conditioned on hadrons from CMS Open Simulations. While previous work focused on jets, our updated methods now can accommodate all particle-flow objects in an event along with particle-level attributes like particle type and production vertex coordinates. This approach is fully automated, entirely written in Python, and GPU-compatible. Using a variety of physics processes at the LHC, we show that the extended *Parnassus* is able to generalize beyond the training dataset and outperforms the standard, public tool *Delphes*.

AI keywords:

transformers;flow matching;flow;diffusion;generative models

78

Gaussian-Process Bayesian Optimization for the Design of a Precision Particle Physics Experiment

Author: Johannes Alexander Jaeger[1]

[1] *Paul Scherrer Institute*

Corresponding Author: alexander.jaeger@psi.ch

Monte-Carlo (MC) simulations are essential for designing particle physics experiments, as they enable us to evaluate and optimize key objectives—such as enhancing experimental sensitivity and performance. Since exhaustively sampling the full parameter space of experimental configurations is computationally prohibitive, sample-efficient methods to identify promising configurations are necessary. Bayesian optimization (BO) provides a principled framework for optimizing black-box functions by constructing a surrogate model and adaptively selecting promising parameter configurations. Many BO methods have theoretical convergence guarantees, typically relying on assumptions such as Lipschitz continuity or smoothness of the objective function. While these assumptions may not strictly hold in all cases, BO provides a sample-efficient and effective strategy for optimization.

In this study, we explore the use of BO with Gaussian processes [Chowdhury et al., 2019] in the design of an experiment searching for the electric dipole moment (EDM) of the muon at Paul Scherrer Institute [Adelmann et al., 2025]. We aim to optimize the injection of muons into the experiment to maximize the sensitivity to an EDM. To that end, we simulate our experiment using G4Beamline [Roberts & Kaplan, 2007], an MC framework for particle physics built on Geant4. We begin optimization with an initial set of parameter configurations, sampled using a quasi-random method to ensure broad coverage of the search space. We then use a Gaussian process as a surrogate model for the expensive simulation, and apply acquisition functions to choose new parameter combinations to simulate. By iteratively refining the surrogate model, we strategically sample parameter configurations that are likely to yield improvements for the sensitivity of the experiment.

Our study investigates the use of Bayesian optimization as an efficient method for parameter tuning in particle physics experiments. Using surrogate modeling and adaptive sampling, this approach offers a promising direction for optimizing complex experimental configurations.

References:

Chowdhury, S.R., and Gopalan, A. (2019). On Batch Bayesian Optimization. ArXiv, abs/1911.01032.

Adelmann, A. et al. (2025). A compact frozen-spin trap for the search for the electric dipole moment of the muon. submitted to EPJC, arXiv preprint arXiv:2501.18979.

Roberts, T. J. and Kaplan, D. M. (2007), G4beamline simulation program for matter-dominated beamlines. IEEE Particle Accelerator Conference (PAC)

AI keywords:

Bayesian optimization; Gaussian processes; surrogate modeling; experimental design

79

Mixture of Expert Graph Transformer for Particle Collision Detection

Author: Donatella Genovese[1]

Co-authors: Alessandro Sgroi [1]; Alessio Devoto [1]; Cristiano Sebastiani [2]; Lennox Wood [3]; Monica D'Onofrio [3]; Samuel Valentine [3]; Simone Scardapane [4]; Stefano Giagu [5]

[1] *Sapienza Università di Roma*

[2] *Roma1 - INFN*

[3] *Liverpool University*

[4] *Istituto Nazionale di Fisica Nucleare*

[5] *Sapienza Università di Roma and Istituto Nazionale di Fisica Nucleare*

Corresponding Author: donatella.genovese@uniroma1.it

The Large Hadron Collider (LHC) at CERN generates vast amounts of data from high-energy particle collisions, requiring advanced machine learning techniques for effective analysis. While Graph Neural Networks (GNNs) have demonstrated strong predictive capabilities in high-energy physics (HEP) applications, their "black box" nature often limits interpretability. To address this challenge, we propose a novel Mixture-of-Experts Graph Transformer (MGT) model that enhances both predictive performance and interpretability in collision event classification.

Our approach combines a Graph Transformer architecture with Mixture-of-Experts (MoE) layers, allowing for specialized expert subnetworks that improve classification while maintaining transparency. Attention maps from the Graph Transformer provide insights into key physics-driven features, while the MoE structure enables an analysis of expert specialization, highlighting the model's decision-making process. We evaluate our model using simulated collision events from the ATLAS experiment, focusing on distinguishing rare Supersymmetric (SUSY) signal events from Standard Model background processes. Results show that our model achieves competitive classification accuracy while offering interpretable outputs aligned with known physics principles.

This work underscores the significance of explainable AI in HEP, ensuring greater trust in machine learning-driven discoveries. By embedding interpretability into the model architecture, we provide a powerful and transparent framework for analyzing complex particle collision data, paving the way for more reliable AI-assisted research in fundamental physics.

AI keywords:

Graph Transformers; Mixture-of-Experts; Explainability

80

Data–Driven Dark Energy: Probing w(a) with Flexknots

Author: Adam Ormondroyd[1]

Co-author: Will Handley [1]

[1] *University of Cambridge*

Corresponding Author: ano23@cam.ac.uk

Recent cosmological surveys have opened a new window onto the nature of dark energy. In our work we reconstruct the dark energy equation of state using a "flexknot"parameterisation that represents $w(a)$ as a linear spline with free–moving nodes. By combining the latest DESI Baryonic Acoustic Oscillation measurements with Pantheon+ supernovae data—and cross–checking our results with an independent Cobaya–based pipeline—we obtain posterior distributions for $w(a)$ that reveal an unexpected W–shaped structure. Although the Bayesian evidence does not ultimately favour dynamical dark energy over the standard ΛCDM model, our analysis shows that even non–CMB datasets can indicate deviations from a constant $w = -1$.
We also generalise dataset tension statistics to marginalise over multiple models, ensuring our unexpected results are not driven by inter-dataset disagreement.
We finish with a brief discussion of the pedagogical advantages and computational benefits of developing an analysis pipeline in-house, which, in addition to increasing efficiency, allows us to analytically marginalise nuisance parameters and provides confidence that these features are genuinely driven by the data.

AI keywords:

Bayesian inference; dataset tension; model selection

81

End-to-End Optimization of Generative AI for Robust Background Estimation

Authors: Christina Reissel[1]; Gaia Grosso[None]; Giada Badaracco[2]; Philip Harris[1]; Thea Aarrestad[2]

[1] *MIT*

[2] *ETH Zurich*

Corresponding Author: gbadaracco@student.ethz.ch

As searches at the LHC probe increasingly rare signals against an overwhelming background of Standard Model events, progressively tighter selection criteria are applied to enhance signal-rich regions. Simulated background samples serve as the basis for hypothesis testing, enabling comparisons between observed data and expected Standard Model backgrounds. However, this approach becomes challenging when the available background statistics are insufficient. This talk presents an end-to-end framework for estimating background models endowed with uncertainties. We train a generative model, explore different approaches to attribute a shape uncertainty and check its compatibility with the underlying ground truth using NPLM, a machine learning-based goodness-of-fit test. This procedure allows us to assess to which extent generative AI models are safe for sampling. By incorporating well-defined uncertainties, we ensure the framework can perform effectively even in data-limited scenarios to provide robust and reliable anomaly detection.

AI keywords:

generative AI uncertainty estimation, oversampling, anomaly detection

82

Anomaly Detection with the ATLAS Trigger at CERN

Author: Shreya Saha[1]

[1] *University of Adelaide*

Corresponding Author: shreya.saha@cern.ch

The application of machine learning techniques in particle physics has accelerated the development of methodologies for exploring physics beyond the Standard Model. This talk will present an overview of anomaly detection and its potential to enhance the detection of new physics within data collected by the ATLAS detector at CERN. The talk will discuss the adaptation and real-time deployment of anomaly detection algorithms, integrated into the detector's trigger system. Additionally, a novel analysis strategy will be outlined for detecting non-resonant anomalies in a model-agnostic manner, utilizing autoencoders, which are based on deep neural networks. The presentation will include signal sensitivity studies for anomalous events, along with background estimation studies, based on the performance of the autoencoders.

AI keywords:

Anomaly detection, auto encoders, neural networks, FPGAs

83

Machine learning methods for studying cosmology with upcoming SKA data

Author: Sambit Giri[1]

[1] *University of Groningen*

Corresponding Author: s.k.giri@rug.nl

The formation of the first galaxies was a pivotal period in cosmic history that ended the cosmic dark ages and paved the way for present-day galaxies such as our Milky Way. This period, characterised by distinct conditions—such as the absence of crucial metals necessary for efficient gas cooling—poses a frontier in cosmology and astrophysics, offering opportunities to discover novel physics. Emitting energetic radiation into the intergalactic medium (IGM), these early galaxies initiated "cosmic reionisation", a process gradually heating and ionising the gas. The 21-cm signal, produced by the neutral hydrogen, is a unique signal that can probe the evolution of gas in the IGM. Although current radio experiments, such as Low-frequency Array (LOFAR), Hydrogen Epoch of Reionization Array (HERA) and Murchison Widefield Array (MWA), are attempting to detect this signal, the upcoming low-frequency component of the Square Kilometre Array (SKA) will revolutionise the field by producing images of the distribution of this signal on the sky at various epochs of reionisation. This data is a treasure trove of information about the evolution of our universe.
In this talk, I will present a U-Net-based framework designed to detect physical patterns in 21-cm images, extending beyond the Gaussian information captured by the power spectrum. I will address the challenges of image reconstruction posed by telescope systematics, including noise, resolution limitations, and foreground residuals, and introduce a machine learning-based approach optimised for mitigating these effects in 21-cm images. Additionally, I will discuss a simulation-based inference framework aimed at constraining the physics of the early universe.

AI keywords:

pattern recognition, image reconstruction, simulation-based inference

84

Efficient Graph Coloring with Neural Networks: A Physics-Inspired Approach for Large Graphs

Authors: Andrea Cacioppo[1]; Andrea Ciardiello[2]; Federico Ricci Tersenghi[3]; Lorenzo Colantonio[4]; Maria Chiara Angelini[5]; Stefano Giagu[6]

[1] *Sapienza University of Rome and INFN Roma*

[2] *Istituto Superiore Sanità and INFN Roma*

[3] *Sapienza Università di Roma*

[4] *Sapienza University of Rome and INFN Rome*

[5] *Sapienza University of Rome*

[6] *Sapienza Università di Roma and INFN Roma*

Corresponding Author: lorenzo.colantonio@uniroma1.it

The graph coloring problem is an optimization problem involving the assignment of one of q colors to each vertex of a graph such that no two adjacent vertices share the same color. This problem is computationally challenging and arises in several practical applications. We present a novel algorithm that leverages graph neural networks to tackle the problem efficiently, particularly for large graphs. We propose a physics-inspired approach that leverages tools
used in statistical mechanics to improve the training and performances of the algorithm. The scaling of our method is evaluated for different connectivities and graph sizes. Finally, we demonstrate the effectiveness of our method on a dataset of Erdős–Rényi graphs, showing its applicability also in hard-to-solve connectivity regions, where traditional methods struggle.

AI keywords:

Optimization:Graph Neural Networks:Physics Informed Models:Statistical Mechanics

85

Hybrid Quantum-Classical Diffusion Models for Generative Tasks

Authors: Andrea Cacioppo[1]; Lorenzo Colantonio[2]; Simone Bordoni[3]; Stefano Giagu[4]

[1] *Sapienza University of Rome and INFN Roma*

[2] *Sapienza University of Rome and INFN Rome*

[3] *Sapienza universita di Roma and INFN Roma*

[4] *Sapienza Università di Roma and Istituto Nazionale di Fisica Nucleare*

Corresponding Author: andrea.cacioppo@uniroma1.it

Generative models based on diffusion processes have recently emerged as powerful tools in artificial intelligence, enabling high-quality sampling in a variety of domains. In this work, we propose a novel hybrid quantum-classical diffusion model, where artificial neural networks are replaced with parameterized quantum circuits to directly generate quantum states. To overcome the limitations of quantum circuits in approximating complex probability distributions, we integrate them with classical neural networks, forming a hybrid architecture that increases expressivity and improves the ability to represent complex functions in Hilbert space. This hybrid model effectively learns and samples quantum states, opening new directions for generative quantum machine learning.
We assess the performance of our model using a combination of quantitative metrics and qualitative evaluations. Furthermore, we explore its applications in fundamental science, demonstrating its potential in domains where quantum generative models can provide computational and conceptual advantages over classical approaches.

AI keywords:

Quantum Machine Learning:Diffusion Models:Hybrid Models

86

Field-level inference of primordial non-Gaussianity in upcoming galaxy redshift surveys

Author: Adam Bengt Johansson Andrews[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: johansso@bo.infn.it

One of the most outstanding questions in modern cosmology concerns the physical processes governing the primordial universe and the origin of cosmic structure. The detection and measurement of (local) primordial non-Gaussianity would provide insights into the shape of the potential of the inflaton field, the hypothetical particle driving cosmic inflation. In the coming years, the next generation of galaxy surveys, e.g., Euclid, will observe the Universe, with the scientific goal of constraining the nonlinearity parameter, fnl, to the precision required to identify viable inflationary models. In my talk, I will describe my approach to constraining primordial non-Gaussianity using a field-level inference algorithm and demonstrate a proof of concept. The method is based on forward-modelling the initial conditions and exploring the joint posterior in a Bayesian hierarchical framework. By utilizing the full field, the method naturally and fully self-consistently accounts for all stochastic uncertainties and systematic effects associated with selection effects, galaxy biasing, and survey geometries. The current and preliminary results, which I will showcase, demonstrate that we are able to achieve competitive constraints on f_{NL} in existing and upcoming galaxy redshift surveys

AI keywords:

Field-level inference, emulators, Hamiltonian Monte Carlo

87

Real-Time Motion Correction in Magnetic Resonance Spectroscopy: AI Solutions Inspired by fundamental science

Authors: Alberto Annovi[1]; Alessandra Maiuro[2]; Alessandro Lonardo[1]; Andrea Ciardiello[3]; Benedetta Argiento[1]; Biagio Rossi[1]; Carlo Mancini Terracciano[1]; Federico Giove[4]; Stefano Giagu[5]; Matteo Cacioppo[6]; Cecilia Voena[1]; Cristian Rossi[1]; Francesca Lo Cicero[1]; Elvira Rossi[1]; Francesco Simula[7]; Marco Montuori[8]; Mario Merola[1]; Emilia Nistico[6]; Pierpaolo Perticaroli[1]; Roberto Coccurello[9]; Silvia Capuani[10]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *Sapienza Università di Roma and INFN Roma*

[3] *Istituto Superiore Sanità and INFN Roma*

[4] *CREF*

[5] *Sapienza Università di Roma and Istituto Nazionale di Fisica Nucleare*

[6] *Sapienza Università di Roma*

[7] *ROMA1*

[8] *Consiglio Nazionale delle Ricerche, Roma*

[9] *Consiglio Italiano delle Ricerche, Roma*

[10] *CNR ISC UOS Roma Sapienza*

Corresponding Author: benedetta.argiento@na.infn.it

Magnetic Resonance Spectroscopy is a powerful, non-invasive tool for in vivo biochemical and metabolic tissue analysis, yet its widespread clinical application remains hindered by susceptibility to motion artifacts. Traditional retrospective corrections struggle with real-time constraints, limiting diagnostic precision in key medical scenarios such as neurodegenerative disease monitoring.
The Recentre project founded by the Italian Ministry of University and Research, pioneers an AI-driven, real-time motion correction system that dynamically adapts MR acquisition sequences using deep learning models deployed on fast hardware accelerators. By leveraging know-how developed in real-time inference techniques developed for high-energy physics experiments, particularly fast trigger systems used in particle detection at CERN, we introduce optimized deep neural networks capable of low latency corrections. In the talk we'll explore the synergy between fundamental physics and application in medical imaging, highlighting how machine learning techniques developed for particle physics data pipelines can enhance real-time diagnostics in biomedical applications.

AI keywords:

motion correction:magnetic resonance:real-time AI

88

Graph Neural Network Acceleration on FPGAs for Fast Inference in the Future ATLAS Muon Trigger Update

Authors: Davide Fiacco[1]; Giuliano Gustavino[1]; Graziella Russo[1]; Martino Errico[2]; Stefano Giagu[3]; Valerio Ippolito[1]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *Sapienza Università di Roma and INFN Roma*

[3] *Sapienza Università di Roma and Istituto Nazionale di Fisica Nucleare*

Corresponding Author: martino.errico@roma1.infn.it

The upcoming High Luminosity phase of the Large Hadron Collider will require significant advancements in real-time data processing to handle the increased event rates and maintain high-efficiency trigger decisions. In this work, we extend our previous studies on deploying compressed deep neural networks on FPGAs for high-energy physics applications by exploring the acceleration of graph neural networks for fast inference in the ATLAS muon trigger system.
Graph-based architectures offer a natural way to represent and process detector hits while preserving spatial and topological information, making them particularly suitable for muon reconstruction in a noisy and sparse environment. This work contributes to the broader goal of integrating AI-driven solutions into HEP triggering systems and represents a step forward in realizing fast, hardware-optimized, graph-based inference in experimental physics.

AI keywords:

fast-inference:FPGA:trigger:LHC

89

Development of a Central Trigger Processor board for the Advanced SiPM based camera of the CTAO Large-Sized Telescopes

Authors: Alejandro Pérez Aguilera[1]; Juan Abel Barrio[2]; Luis Ángel Tejedor[2]; María Molina Delicado[2]

[1] *UCM - IPARCOS*

[2] *IPARCOS-UCM*

Corresponding Author: alejpe15@ucm.es

Current Imaging Atmospheric Cherenkov Telescopes (IACT) use combined analog and digital electronics for their trigger systems, implementing simple but fast algorithms. Such trigger techniques are needed in order to cope with the extremely high data rates and strict timing requirements. In recent years, in the context of the Advanced camera design for the Large-Sized Telescopes (LSTs) of the Cherenkov Telescope Array Observatory (CTAO) based on Silicon PhotoMultipliers (SiPM), a new fully digital trigger system incorporating Machine Learning (ML) algorithms is being developed. The main concept is to implement those algorithms in Field Programmable Gate Arrays (FPGAs) and take advantage of the higher camera resolution, in order to improve the ability to distinguish between low-energy gamma ray showers and noise. Thus, the sensitivity to low-energy gamma rays will be improved, while being able to fulfill the previous constraints of CTAO-LST. We will describe the project to develop a Central Trigger Processor (CTP) board for the Advanced LST SiPM-CAM, with the aim to run such advanced trigger algorithms and additionally perform a hardware stereo trigger among LSTs. We will present the CTP conceptual design, as well as the prototypes built so far.

AI keywords:

FPGAs; Trigger; Real-Time; ML; Image Processing

90

Track Inference of the Ion-optics of WASA-FRS based on machine learning models

Author: David Calonge[None]

Co-authors: Christophe Rappold [1]; For the WASA-FRS/Super-FRS Experiment Collaboration ; Samuel Escrig [2]

[1] *Instituto de Estructura de la Materia - CSIC*

[2] *IEM- CSIC*

Corresponding Author: david.calonge@csic.es

The WASA-FRS HypHI Experiment focuses on the study of light hypernuclei by means of heavy-ion induced reactions in 6Li collisions with 12C at 1.96GeV/u. It is part of the WASA-FRS experimental campaign, and so is the eta-prime experiment [1]. The distinctive combination of the high-resolution spectrometer FRagment Separator (FRS) [2] and the high-acceptance detector system WASA [3] is used. The experiment was successfully conducted at GSI-FAIR in Germany in March 2022 as a component of the FAIR Phase-0 Physics Program, within the Super-FRS Experiment Collaboration. The primary objectives of this experiment are twofold: to shed light on the hypertriton puzzle [4] and to investigate the existence of the previously proposed $nn\Lambda$ bound state [5]. Currently, the data from the experiment is under analysis.

Part of the data analysis is to provide a precise ion-optics of the measurement of the fragment originated from the mesonic weak decay of the hypernuclei of interest. The reconstruction the ion-optics of fragments is based on the calibration run of FRS optics. We have proposed to implement machine learning models and neural networks to represent the ion-optics of FRS: While the current state of the problem involves solving equations of motion of particles in non-ideal magnetic fields - which leads to the application of approximations in the calculations - the implementation of data-driven models allows us to obtain accurate results with possible better momentum and angular resolution.

In this presentation, we will show the current status of the R&D in machine learning model of the ion-optics and the prospect of the inference of the track parameters of the fragments based on

the calibration data recorded during the WASA-FRS experimental campaign of 2022. Our model selection optimization follows this approach: we utilize AutoML environments [6], to determine the best pipeline for our data. Once identified, this optimized pipeline is implemented in a PyTorch model.

The results of this study demonstrate a robust reconstruction of the track angles in the FRS mid-focal plane, achieving an improvement of up to a ~40%. A resolution of 0.65 mrad and 0.46 mrad was achieved for the horizontal and vertical angular track plane, respectively. Additionally, the reconstruction of the magnetic rigidity in the final focal plane attained a resolution Δp/p of 5 10^{-4}. From these results, we demonstrated that a data-driven model of non-linear ion optics is feasible. We also observed that training the full model can be achieved very quickly, paving the way for online training during data collection at the FRS. This capability will enable more accurate real-time analysis of fragment identification and improve the quality of the exotic beam obtained from the fragment separator.

[1] Y.K. Tanaka et al., J. Phys. Conf. Ser. 1643 (2020) 012181.
[2] H. Geissel et al., Nucl. Instr. and Meth. B 70 (1992) 286-297.
[3] C. Bargholtz et al., Nucl. Instr. and Meth. A 594 (2008) 339-350.
[4] T.R. Saito et al., Nature Reviews Physics 3 (2021) 803-813.
[5] C. Rappold et al., Phys. Rev. C 88 (2013) 041001.
[6] M. Feurer et al., JMLR 23 261 (2022) 1-61.

AI keywords:

Automated Machine learning;optimal pipeline;online-data analysis

91

Towards foundation model for astrophysical source detection: An End-to-End Gamma-Ray Data Analysis Pipeline Using Deep Learning

Authors: Dmitry Malyshev[1]; Gabrijela Zaharijas[2]; Giacomo Principe[3]; Judit Pérez-Romero[2]; Roberto Ruiz de Austri[4]; Saptashwa Bhattacharyya[2]; Sascha Caron[5]; Zoja Rokavec[6]

[1] *ECAP*

[2] *CAC/UNG*

[3] *Istituto Nazionale di Fisica Nucleare*

[4] *IFIC UV-CSIC*

[5] *Radboud University and Nikhef*

[6] *University of Ljubljana*

Corresponding Author: judit.perez@ung.si

The increasing volume of gamma-ray data from space-borne telescopes, like *Fermi*-LAT, and the upcoming ground-based telescopes, like the Cherenkov Telescope Array Observatory (CTAO), presents us with both opportunities and challenges. Traditional analysis methods based on likelihood analysis are often used for gamma-ray source detection and further characterization tasks. A key challenge to analysing data from these telescopes arises due to background contamination; consistent of interstellar emission for *Fermi*-LAT and cosmic ray background for CTAO data, which obscures the faint source population.
Here we will present our results from an end-to-end Deep Learning (DL) based pipeline for detection, localization and further characterization tasks of gamma-ray sources. We extend our AutoSourceID (ASID) pipeline, a DL-based pipeline initially tested with Fermi-LAT simulated data and optical data (MeerLICHT), to include results for CTAO simulated data for Galactic Plane Survey (GPS) observation. We will also present a pre-processing step designed with Deep Neural Net for denoising tasks which can potentially decrease the source detection threshold. Training on data from different telescopes to capture a broad representational space of the gamma-ray sky, this end-to-end pipeline -

from denoising to detection and characterization - could potentially serve as a foundational model for gamma-ray astronomy by offering a generalizable framework for other surveys.

AI keywords:

Deep Neural Nets; Pattern Identification; Foundational Model

92

E Pluribus Unum: Ensemble Methods for Improved PDF and Uncertainty determination

Author: Juan M. Cruz Martinez[1]

[1] *CERN*

Corresponding Author: juan.cruz.martinez@cern.ch

The determination of parton distribution functions (PDFs) is core problem in hadron colliders physics, as PDFs model the initial state of the colliding partons. Traditionally, these distributions have been obtained by fitting a functional form to a subset of experimental data, making them susceptible to biases from these choices.
The NNPDF collaboration introduced an alternative approach in which the PDF is modeled using neural networks, removing the bias inherent to the choice of functional form and enabling a fully data-driven determination.

In this talk, I explore the power of ensemble regression to take this approach one step further by eliminating the choice of a specific machine learning architecture. By sampling from a space of possible models that can accommodate the collider data. This new approach, which incorporate the (hyper)parametric uncertainty directly into the determination has been made feasible by recent developments in hardware acceleration and opens the door to more robust extractions of PDFs.

AI keywords:

uncertainty estimation, ensemble regression, model sampling, data-driven models

94

Machine Learning-Based Energy Reconstruction for the ATLAS Tile Calorimeter at HL-LHC

Author: AUGUSTO CERQUEIRA[1]

[1] *Universidade Federal de Juiz de Fora*

Corresponding Author: augusto@cern.ch

The HL-LHC project is driving significant upgrades to the ATLAS experiment to enhance data processing and maintain its discovery potential under high-luminosity conditions. A key aspect of this upgrade is the replacement of the readout electronics for the ATLAS Tile Hadronic Calorimeter. The new Tile PreProcessor (TilePPr) system, equipped with Kintex Ultrascale FPGAs, serves as the interface between the front-end electronics and the first level of the future ATLAS Trigger system. The TilePPr will perform real-time signal reconstruction, delivering calibrated data for each bunch crossing at 40 MHz with a fixed and low-latency path. This contribution will focus on the design, implementation, and performance evaluation of Machine Learning-based reconstruction algorithms

within the TilePPr, designed to meet the HL-LHC requirements. Different neural network architectures are being explored to achieve accurate and efficient energy reconstruction while keeping computational and storage demands low. Given the constraints of real-time processing, special emphasis is placed on model optimization strategies, ensuring fast inference on FPGAs without loss of precision.

AI keywords:

Estimation, Machine Learning, FPGA, Neural Network

95

Learning Optimal and Interpretable Summary Statistics of Galaxy Catalogs with SBI

Author: Kai Lehman[1]

Co-authors: Sven Krippendorf [2]; Jochen Weller [1]; Klaus Dolag [1]

[1] *LMU Munich*

[2] *DAMTP Cambridge*

Corresponding Author: kai.lehman@physik.lmu.de

How much cosmological information can we reliably extract from existing and upcoming large-scale structure observations? Many summary statistics fall short in describing the non-Gaussian nature of the late-time Universe in comparison to existing and upcoming measurements. We demonstrate that we can identify optimal summary statistics and that we can link them with existing summary statistics. Using simulation based inference (SBI) with automatic data-compression, we learn summary statistics for galaxy catalogs in the context of cosmological parameter estimation. By construction these summary statistics do not require the ability to write down an explicit likelihood. We demonstrate that they can be used for efficient parameter inference. These summary statistics offer a new avenue for analyzing different simulation models for baryonic physics with respect to their relevance for the resulting cosmological features. The learned summary statistics are low-dimensional, feature the underlying simulation parameters, and are similar across different network architectures. To link our models, we identify the relevant scales associated to our summary statistics and we are able to match the summary statistics to underlying simulation parameters across various simulation models.

AI keywords:

simulation-based inference; graph neural networks; informative summary statistics

96

Refinement of calorimeter showers simulated with normalizing flow model

Author: Vera Maiboroda[1]

Co-authors: Corentin Allaire [1]; David Rousseau [2]

[1] *IJCLab, CNRS*

[2] *IJCLab-Orsay*

Corresponding Author: vera.maiboroda@cern.ch

The simulation of calorimeter showers is computationally expensive, leading to the development of generative models as an alternative. Many of these models face challenges in balancing generation quality and speed. A key issue damaging the simulation quality is the inaccurate modeling of distribution tails. Normalizing flow (NF) models offer a trade-off between accuracy and speed, making them a promising approach. This work builds on the CaloINN NF model and introduces a set of post-processing modifications of analysis-level observables aimed at improving the accuracy of distribution tails. We used CaloChallenge datasets as well as simulations produced with Open Data Detector (ODD) to validate the method. The results show that introduced refinements enhance overall performance, achieving accuracy comparable to the most precise calorimeter shower models while maintaining the simulation speed of NF models. The study is conducted as part of the interTwin project, which develops Digital Twins for applications in physics and earth observation, demonstrating the use of the intertwin platform for calorimeter simulation.

AI keywords:

normalizing flow, generative model, post-processing

97

Machine learning for muon identification at the CMS experiment

Author: CMS Speaker[None]

Muon objects are an important component of the CMS physics program, with several analyses focusing on final states involving muons. Efficient identification of prompt and non-prompt muons are essential for many physics studies. To enhance muon identification beyond the standard cut-based criteria used in Run 2, new identification strategies based on machine learning techniques have been developed. Specific multivariate algorithms have been optimized to select respectively signatures with prompt muons with medium transverse momentum (greater than 10 GeV), for rejecting misidentified muons in the same kinematic regime and finally for selecting "soft"muons such as muons from heavy-flavor hadrons decays against misidentified muons and muons from long-lived light-flavor hadron decays. This presentation will highlight the performance of the new classifiers, compared with the cut-based identification criteria in data and simulation.

AI keywords:

Signal identification reconstruction, inference, ML classifiers

98

Anomaly detection for automated Data Quality Monitoring and Certification at the CMS experiment

Author: CMS Speaker[None]

Maintaining high data quality in large HEP experiments like CMS at the LHC is essential for obtaining reliable physics results. The LHC high-luminosity phase will introduce higher event rates, requiring more sophisticated monitoring techniques to promptly identify and address potential issues. The CMS protocols for Data Quality Monitoring (DQM) and Data Certification (DC) rely on significant human intervention and have limited time granularity, which may lead to transient anomalies going undetected. To address these challenges, unsupervised machine learning techniques have been deployed for anomaly detection with per-lumisection granularity. Given the complexity and diversity of CMS subdetectors, multiple tools are being developed in parallel and maintained by

subsystem experts. In this contribution, we discuss the development of automated workflows with per-lumisection granularity for online DQM and DC across different CMS subdetectors, and their integration into a common interface.
Speakers

AI keywords:

Signal identification reconstruction, inference, ML classifiers

99

Identification of Lorentz-boosted jets in the CMS experiment

Author: CMS Speaker[None]

A fundamental aspect of CMS researches concerns the identification and characterisation of jets originating from quarks and gluons produced in high-energy proton-proton collisions. Electroweak scale resonances (Z/W bosons) and Higgs bosons are often produced with high Lorentz-boosts, where their products become highly collimated large and massive jets, usually reconstructed as AK8 jets. Therefore, the identification of the particle initiating the jet plays a crucial role in distinguishing between boosted bosons from the QCD background. In this talk, an overview of the usage of boosted jet taggers within CMS will be given. It will highlight the most recent AK8 tagging algorithms, which make use of sophisticated machine learning techniques, optimised for performance and efficiency.

AI keywords:

Signal identification pattern reconstruction, inference, ML classifiers

100

Computing the Matrix Element Method with generative machine learning

Author: CMS Speaker[None]

The Matrix Element Method (MEM) is a powerful technique for computing the event-by-event likelihood for a given theory hypothesis, simultaneously accounting for both detector effects and theoretical models. Despite its strong theoretical foundation, MEM is seldom used in analyses involving final states with many jets due to the complex, multi-dimensional integrals required to accurately model detector reconstruction effects through precise transfer functions. In this contribution, we implement a new approach leveraging Transformers and generative ML models to sample parton-level events for efficient numerical MC integration and to encode the complex transfer function which models showering, hadronization, and detector reconstruction. We demonstrate this strategy on a complex final state, the ttH(bb) process in the semileptonic decay channel, using the full CMS detector simulation and reconstruction. Furthermore, we show that this method can be generalized and applied to multiple physics processes using a single ML model.

AI keywords:

Transformers; generative ML models;simulation-based inference

101

Using AI on FPGAs for the CMS Overlap Muon Track Finder for the HL-LHC

Author: CMS Speaker[None]

In view of the HL-LHC, the Phase-2 CMS upgrade will replace the entire trigger and data acquisition system. The L1T system has been designed to process 63 Tb/s input bandwidth with state-of-the-art commercial FPGAs and high-speed optical links reaching up to 28 Gb at a fixed latency below 12.5 µs. In view of the upgraded trigger system and in preparation for the HL-LHC, a GNN has been trained to reconstruct displaced muon signatures in the transition region between barrel and endcap and its implementation in FPGAs within the strict latency requirements imposed by the system will be discussed in this presentation. The process of adapting such machine learning models to hardware will also be described in detail. Additionally, other recent muon reconstruction algorithms developed for the HL-LHC will be reviewed.

AI keywords:

FPGAs;real-time signal identification,ultra fast processing

102

Wasserstein normalized autoencoders for detecting anomalous jets in CMS

Author: CMS Speaker[None]

Unsupervised machine learning algorithms are powerful tools for identifying potential new physics at the LHC, enabling the separation of standard model (SM) background events from anomalous signal events without relying on predefined signal hypotheses. Autoencoders (AEs) are frequently employed in such tasks, but their effectiveness is often hindered by the reconstruction of outliers. In this work, we present the Wasserstein Normalized Autoencoder (WNAE), an improved version of the normalized autoencoder (NAE), designed to mitigate these challenges. We apply the WNAE to a search for semivisible jets (SVJs) in the CMS experiment, demonstrating its enhanced ability to distinguish signal from background events. Our results highlight the potential of WNAE-based anomaly detection to improve sensitivity in LHC searches for beyond-the-SM physics.

AI keywords:

Autoencoders; anomaly detection

103

Multi-Scale Transformer Encoder for Di-Tau Invariant Mass Reconstruction

Author: CMS Speaker[None]

With the discovery of the Standard Model (SM) Higgs boson (H) by the CMS and ATLAS experiments in 2012, the last missing elementary particle predicted by the SM was identified. Since then, extensive measurements of the Higgs boson's properties have been performed across various decay channels. One of the most important is its decay into a pair of tau leptons, the second-most frequent fermionic decay mode after the decay into bottom quarks. In such analyses, the reconstructed invariant mass of the di-tau system plays a crucial role in distinguishing signal (H) from background events.

However, due to the presence of neutrinos in the final state, a portion of the energy is lost, leading to an underestimation of the invariant mass. This work proposes a novel Deep Learning (DL) model designed to enhance the reconstruction of the invariant mass by estimating the full four-momentum of each tau lepton before decay, rather than directly regressing the mass. The approach allows for a more precise kinematic characterization of the parent particle. The implemented model is a custom transformer encoder —an advanced DL architecture originally developed for Natural Language Processing, now proving its versatility in diverse domains. It takes as input the information from the di-tau products, the reconstructed properties of the two tau leptons, the missing transverse energy (MET), and other key event variables relevant for invariant mass reconstruction. Through learned embeddings, the model extracts meaningful features from each input source, which are then combined and processed using self-attention layers within the transformer encoder. This architecture enables the model to effectively capture correlations between inputs and recover missing contributions from neutrinos, leading to a more accurate invariant mass estimation. The performance of the proposed algorithm is benchmarked against the current standard method used in CMS (SVFit) using simulated $H \longrightarrow \tau\tau$ events and main background processes. The results demonstrate the potential of this novel approach in improving mass reconstruction and enhancing Higgs boson analyses.

AI keywords:

Deep Learning model; transformer encoder; signal identification

104

Advancing the CMS Level-1 Trigger: Jet Tagging with DeepSets at the HL-LHC

Author: CMS Speaker[None]

At the Phase-2 Upgrade of the CMS Level-1 Trigger (L1T), particles will be reconstructed by linking charged particle tracks with clusters in the calorimeters and muon tracks from the muon station. The 200 pileup interactions will be mitigated using primary vertex reconstruction for charged particles and a weighting for neutral particles based on the distribution of energy in a small area. Jets will be reconstructed from these pileup-subtracted particles using a fast cone algorithm. For the first time at the CMS L1T, the particle constituents of jets will be available for jet tagging. In this work we present a new multi-class jet tagging neural network (NN). Targeting the L1T, the NN is a small DeepSets architecture, and trained with Quantization Aware Training. The model predicts the classes: light jet (uds), gluon, b, c, tau_h+, tau_h-, electron, muon. The model additionally predicts the pT of the object. The new model enhances the selection power of the L1T for important processes for CMS at the High Luminosity LHC such as di-Higgs and Higgs production via Vector Boson Fusion. We present the model including its performance at object tagging and deployment into the L1T FPGA processors, and showcase the improved trigger capabilities enabled by the new tagger.

AI keywords:

multi-class jet tagging neural network; Quantization Aware Training; DeepSets architecture

105

Design and deployment of a fast neural network for measuring the properties of muons originating from displaced vertices in the CMS Endcap Muon Track Finder

Author: CMS Speaker[None]

We report on the development, implementation, and performance of a fast neural network used to measure the transverse momentum in the CMS Level-1 Endcap Muon Track Finder. The network aims to improve the triggering efficiency of muons produced in the decays of long-lived particles. We implemented it in firmware for a Xilinx Virtex-7 FPGA and deployed it during the LHC Run 3 data-taking in 2023. The new displaced muon triggers that use this algorithm broaden the phase space accessible to the CMS experiment for searches that look for evidence of LLPs that decay into muons.

AI keywords:

FPGAs;real-time signal identification,ultra fast processing

106

B-hadron identification in b-jets using novel deep learning technique in pp collisions in CMS

Author: CMS Speaker[None]

Understanding the substructure of jets initiated by heavy quarks is essential for quantum chromodynamics (QCD) studies, particularly in the context of the dead-cone effect and jet quenching. The kinematics of b-hadron decays present a challenge for substructure measurements with inclusive b-jets. We propose an approach using geometric deep learning to extract the optimal representation of the b-hadron decays utilizing the charged decay products of the jet represented as a point cloud and identify tracks associated with the b-hadrons while simultaneously tagging the b-jets. The method is demonstrated in simulations of p-p and Pb-Pb collisions at $\sqrt{s} = 5.02$ TeV with the CMS detector and compared with previous approaches based on boosted decision trees.

AI keywords:

geometric deep learning; point cloud identification; signal reconstruction

107

Reconstructing the Initial Density Field of the Universe with Multi-Tracer Line Intensity Mapping

Authors: Anchal Saxena[None]; Daan Meerburg[1]

[1] *University of Groningen*

Corresponding Author: anchal.saxena.009@gmail.com

Reconstructing the initial density field of the universe is crucial for improving cosmological parameter constraints. In this work, we employ a U-Net architecture to reconstruct the initial density field from simulated 21-cm and CO line intensity maps from the Epoch of Reionization (EoR). These tracers provide complementary information, with 21-cm maps capturing low-density neutral regions and CO maps tracing high-density star-forming regions. By combining these two maps, we achieve accurate reconstructions across different ionization fractions. Our results demonstrate that the network effectively recovers both large- and small-scale features of the initial density field, achieving a high cross-correlation with the true field (≥ 0.75 for $k \leq 1\mathrm{Mpc}^{-1}$). To extract cosmological information, we apply Marginal Neural Ratio Estimation (MNRE) to perform simulation-based parameter inference using the reconstructed density field. We find that post-reconstruction information significantly improves constraints on cosmological parameters, reducing uncertainties in σ_8 and n_s by a factor of 2-3. Additionally, we assess the impact of Gaussian random noise on the low-resolution input maps, showing that while the network remains robust in recovering large-scale features, small-scale

structures are more affected. Our results highlight the potential of combining multi-tracer intensity mapping with deep learning and neural inference techniques to enhance cosmological constraints from the EoR.

AI keywords:

Simulation-based inference; U-Net; Pattern-recognition

108

Semi-Supervised Density Estimation for Suppressing ^{42}Ar/^{42}K Surface Beta Events in LEGEND

Author: Niko Nanda Putra Nila Lay[1]

Co-authors: Andreas Leonhardt [1]; Baran Hashemi [1]; Brennan Hackett [2]; Béla Majorovits [2]; Christoph Vogl [1]; Konstantin Gusev [1]; Lukas Heinrich [1]; Mario Schwarz [1]; Moritz Neuberger [1]; Nadezda Rumyantseva [1]; Patrick Krause [1]; Stefan Schönert [1]; Tommaso Comellato [1]

[1] *Technical University of Munich*

[2] *Max Planck Institute for Physics*

Corresponding Author: nikonandap.lay@tum.de

The LEGEND experiment aims to detect neutrinoless double-beta ($0\nu\beta\beta$) decay using high-purity germanium detectors (HPGes) enriched in ^{76}Ge, immersed in instrumented liquid argon (LAr). Atmospheric LAr contains the cosmogenically activated isotope ^{42}Ar, whose decay progeny, ^{42}K, can undergo beta decay ($Q_\beta = 3.5$ MeV) on the HPGe surface. Without the baseline mitigation strategy —using underground-sourced LAr (UGLAr) depleted in ^{42}Ar—this decay would become the dominant background at the $0\nu\beta\beta$ Q-value ($Q_{\beta\beta} = 2.039$ MeV) in LEGEND-1000. Given the non-negligible risk that UGLAr may not be available in time for LEGEND-1000, alternative approaches are being explored, such as optically active enclosures combined with machine-learning-based pulse-shape discrimination (ML-PSD), to distinguish between $0\nu\beta\beta$ signals and background events. To develop and evaluate novel ML-PSD techniques, we operated high-purity germanium (HPGe) detectors in ^{42}Ar-enriched LAr at the SCARF LAr test facility at TU Munich, generating a dataset enriched in ^{42}K surface beta events.

In this work, we investigate the construction of a latent representation of raw HPGe waveform data using variational inference. Unlike conventional PSD parameters, the latent vectors are designed to fully utilize the high-level features of the waveforms. By constraining the latent space with a predefined prior, we estimate the data density corresponding to a signal proxy derived from ^{228}Th calibration data. This is achieved by first employing a classifier neural network to estimate the posterior probability of class labels for samples drawn from both the latent prior and the signal-proxy distribution and then applying Bayes'rule to compute the likelihood of the data under the signal-like hypothesis.

We use the resulting density estimate to classify events as signal- or background-like at a specified significance level. Our evaluation demonstrates promising suppression of ^{42}K surface beta events, providing a pathway for density-based PSD that utilizes the complete raw waveform information from HPGe detectors.

This work was supported by the Cluster of Excellence ORIGINS (EXC 2094-39078331), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy, and by the DFG Collaborative Research Center SFB1258-283604770.

AI keywords:

variational inference, anomaly detection, pattern recognition

109

Machine-Learned Fixed-Point Actions in Four-Dimensional SU(3) Gauge Theory

Authors: Kieran Holland[1]; Andreas Ipp[2]; David I. Müller[3]; Urs Wenger[4]

[1] *University of the Pacific*

[2] *TU Wien, Austria*

[3] *TU Wien*

[4] *Institute for Theoretical Physics, University of Bern*

Corresponding Author: andreas.ipp@tuwien.ac.at

Extracting continuum properties from discretized quantum field theories is significantly hindered by lattice artifacts. Fixed-point (FP) actions, defined via renormalization group transformations, offer an elegant solution by suppressing these artifacts even on coarse lattices. In this work, we employ gauge-covariant convolutional neural networks to parameterize an FP action for four-dimensional SU(3) gauge theory. We show that the gradient flow of the FP action is formally free of lattice artifacts at tree level, enabling the extraction of continuum physics with improved accuracy. Furthermore, our enhanced parameterizations facilitate efficient Hybrid Monte Carlo simulations, effectively mitigating challenges such as critical slowing down and topological freezing. Our results underscore the potential of machine learning techniques to advance lattice QCD studies with reduced discretization errors, a critical step toward precision tests of the Standard Model.

AI keywords:

Gauge equivariant neural networks; L-CNNs; renormalization group learning; physics-informed machine learning

110

Investigating Explainable Jet Tagging with Pretrained Vision Transformers and Attention Mechanisms

Authors: Mariagrazia Monteleone[1]; Federico Camponovo[2]; Lorenzo Cederle[3]; Francesca Camagni[3]; Pietro Govoni[3]; Simone Gennai[3]; Chiara Paganelli[4]

[1] *Politecnico di Milano*

[2] *INFN sezione di Milano –Bicocca, Milano, Italy*

[3] *Istituto Nazionale di Fisica Nucleare*

[4] *Dipartimento di Elettronica,Informazione e Bioingegneria, Politecnico di Milano*

Corresponding Author: mariagrazia.monteleone@polimi.it

Background: In High Energy Physics (HEP), jet tagging is a fundamental classification task that has been extensively studied using deep learning techniques. Among these, transformer networks have gained significant popularity due to their strong performance and intrinsic attention mechanisms. Furthermore, pre-trained transformer models are available for a wide range of classification tasks.
Aim of the work: In this work, we investigate the use of a pre-trained Vision Transformer (ViT) for jet tagging, leveraging reconstructed images from the JETCLASS dataset. Leveraging on its attention mechanism and analyzing attention maps, we provide insights into the model's decision-making process, addressing the challenge of interpretability in deep learning models often seen as "black

boxes.”
Methods: Two jet tagging tasks were selected: the first distinguishing between two very different processes with different number of jets in the final space and the second differentiating between two very similar processes with the same number of jets and similar properties in the final state.
To assess the generalization of the pretrained model, the fine-tuning was performed on a small dataset of 300 images per class, updating only the final two encoder layers while freezing the others. This approach refined high-level features while preserving pretrained representations.
Cumulative attention maps were generated by averaging attention weights across all heads and layers, incorporating residual connections for normalization. Model interpretability was assessed by computing the centroids of each jet within the image and defining regions of interest (ROIs) around these centroids. The attention fraction, quantifying the proportion of attention concentrated within the ROI relative to the total attention across the entire map, was computed to analyze the model's decision-making process.
Conclusions: This work aims to evaluate whether pretrained networks can optimize computational resources without compromising performance and whether attention-based methods enhance interpretability and explainability analysis. Additionally, it explores the potential of the HEP domain as a framework for technical validation, leveraging high-quality data and well-understood causal structures that could be applied to other scientific fields.

AI keywords:

Explainability, Vision Transformers, Physics-informed AI, Jet Tagging, Attention Mechanism

111

Inference optimization with Memory Management and GPU Acceleration in TMVA SOFIE

Author: Sanjiban Sengupta[1]

Co-author: Lorenzo Moneta [1]

[1] *CERN*

Corresponding Author: lorenzo.moneta@cern.ch

Within ROOT/TMVA, we have developed SOFIE - System for Optimized Fast Inference code Emit - an engine designed to convert externally trained deep learning models—such as those in ONNX, Keras, or PyTorch formats—into optimized C++ code for fast inference. The generated code features minimal dependencies, ensuring seamless integration into the data processing and analysis workflows of high-energy physics experiments.
SOFIE now supports a comprehensive range of machine learning operators as defined by the ONNX standard, and also supports the translation and inference of Graph Neural Networks trained in DeepMind's Graph Nets.
Recent advancements in SOFIE include memory optimizations that enable efficient reuse of in- termediate tensor data during inference, significantly reducing memory overhead. Additionally, SOFIE now incorporates enhanced GPU acceleration, supporting stacks such as SYCL, which have abstractions over platforms like CUDA and ROCm. These improvements result in a runtime- efficient and user-friendly machine learning inference engine, competitive with other state-of-the- art solutions.
This work highlights the latest developments in SOFIE, focusing on its memory optimization capabilities and GPU acceleration enhancements, which collectively deliver efficient inference performance for HEP applications.

AI keywords:

Fast ML Inference; ML Software;‘ Next Generatiuon Trigger Project; GPU

112

Scalable Bayesian Inference for Third-Generation Gravitational Wave Data with Normalizing Flows

Author: Thibeau Wouters[1]

[1] *Utrecht University*

Corresponding Author: t.r.i.wouters@uu.nl

Third-generation (3G) gravitational wave (GW) observatories will unveil a cosmic orchestra, detecting thousands of sources annually. However, their increased detection rate poses a major challenge for data analysis. Existing, widely used techniques to obtain the source parameters are prohibitively expensive, creating a bottleneck for extracting scientific insights from 3G detector data. We present ongoing developments of an efficient data analysis pipeline that leverages normalizing flows and hardware accelerators. As an example, we demonstrate Bayesian inference of GW data from binary neutron star mergers, their electromagnetic counterparts, and their implications for nuclear physics, reducing the computational cost from months to a couple of hours. Moreover, our approach enables joint parameter estimation of overlapping GW signals within a few hours. Our methods hold strong promise in meeting the scalability demands of 3G GW detectors, enabling efficient and comprehensive data analysis for future observatories.

AI keywords:

normalizing flows;generative AI;Bayesian inference;deep learning;automatic differentiation

113

Advanced deep-learning applications in neutrino physics

Author: Saul Alonso Monsalve[1]

[1] *ETH Zurich*

Corresponding Author: saul.alonso.monsalve@cern.ch

Deep learning is playing an increasingly important role in particle physics, offering powerful tools to tackle complex challenges in data analysis. This talk presents a range of advanced deep-learning techniques applied to neutrino physics, with a particular focus on the T2K experiment. The discussion includes the use of cutting-edge models such as transformers, domain adaptation strategies like contrastive learning, and anomaly detection methods. These approaches have been employed to improve neutrino interaction identification and enhance the reconstruction of particle kinematics. By integrating these techniques, we aim to refine data analysis pipelines, boost measurement precision, and gain deeper insights into neutrino properties.

AI keywords:

transformers; domain adaptation; anomaly detection

114

Exploiting the latent space of deep AutoEncoders for the identification of signal pulses in noisy time-series

Authors: Gioacchino Alex Anastasi[1]; Sebastiano Albergo[2]; Marzio De Napoli[2]; Noemi Pino[3]; Sebastiana Maria Puglia[2]; Alessia Rita Tricomi[2]; Paules Zakhary[4]

[1] *Università di Catania & INFN Catania*

[2] *Istituto Nazionale di Fisica Nucleare*

[3] *Istituto Nazionale di Fisica Nucleare - Laboratori Nazionali del Sud*

[4] *Università di Catania & INFN Catania & CSFNSM*

Corresponding Author: gialex.anastasi@dfa.unict.it

In this contribution we propose a data-driven technique based on self-supervised deep neural networks, specifically convolutional and variational autoencoders (AE), developed to improve the sensitivity to signal pulses over a significant background in long waveforms.
The dataset consists of synthetic waveforms with around 10,000 samples; each time-series is composed of non-gaussian noise, with the addition of a log-normal shaped signal pulse in a fraction of the events. The AE model is set up to heavily compress the input waveform in a 4-dimensional latent space, allowing a direct study of the features in such a reduced representation.
After a training of about 100 epochs on 7000 waveforms, a region in the latent space where the network encodes time-series presenting only background noise clearly emerges, allowing in turn to tag as signal candidates those falling outside this range. When applied on a test dataset of freshly generated waveforms, such a procedure correctly labels 100% of the events with a large signal, and the fraction of successful identifications only decreases for signal peak amplitudes comparable with the accidental pulses in the background.
This approach was designed to fully exploit the measurements in dual-phase Liquid Argon Time Projection Chambers, as the one of the Recoil Directionality (ReD) experiment, a R&D apparatus built in the context of the Darkside project. The goal is the identification of delayed electroluminescence (S2) signals in gas, produced by very low energy (~ a few keV) nuclear recoils, with a sensitivity at least comparable to the conventional reconstructions. Furthermore, we aim to export this technique to other distinct experimental settings in the field of astroparticle physics.

This work is supported by ICSC –Italian National Research Centre for High Performance Computing, Big Data and Quantum Computing, funded by European Union –NextGenerationEU, and it has been carried out within the Spoke 2 ("Fundamental Research and Space Economy") as part of the activities in the Working Group 3 ("Applications for experimental astroparticle physics and gravitational waves experiments") under the use-case DAIDREAM ("DAta-driven IDentification of Rare Events in Astroparticle physics through Machine learning techniques").

AI keywords:

autoencoders; time-series; latent space; unsupervised learning; pulse identification

115

A Differentiable Bayesian Anomaly Detection Framework for Robust SALT3 Parameter Estimation and Supernova Distance Calibration Using JAX

Author: Sam Leeney[1]

Co-authors: Eloy de Lera Acedo [1]; Harry Bevins [1]; Will Handley [1]

[1] *University of Cambridge*

Corresponding Author: sakl2@cam.ac.uk

We present a novel Bayesian anomaly detection framework, applied to supernova analysis, that exploits a custom-built, differentiable, and highly parallelisable JAX implementation of the commonly used SNcosmo framework. In our framework, each supernova's flux is modelled via the SALT3

formalism, with the core computation—integrating the model flux over observational bandpasses—being fully differentiable and highly parallelisable.

We implement our Bayesian data cleaning strategy, where contaminated (or anomalous) data points are not simply excised but are instead managed by imposing an Occam penalty within the likelihood. This leads to a robust estimation of the SALT parameters (e.g., brightness scaling, stretch, and colour) even when subtle anomalies are present. In addition, we integrate a JAX-based Nested Sampling engine into our toolkit.

Following the methodology in Leeney et al. (2024), we compute a piecewise likelihood from:
$\mathrm{P(D}|\theta) = \prod_i \left([L_i(1-p_i)]^{\epsilon_{\max,i}} \left[\frac{p_i}{\Delta}\right]^{(1-\epsilon_{\max,i})} \right)$

which yields a masked chi-squared–like term that distinguishes between reliable and corrupted data. Notably, the condition $\log L_i + \log \Delta >$
$operatornamelogit(p)$
$relatesdirectlytothelogitfunction—acommonactivationfunctioninmachinelearningusedforbinaryclassific$

More robust modelling of the SALT parameters—particularly brightness scaling, stretch, and colour—directly translates to more precise distance measurements. By incorporating Bayesian anomaly detection, our framework not only flags but quantitatively down-weights anomalous data rather than discarding it outright. This comprehensive treatment minimises systematic biases in the fitted SALT parameters, reducing the scatter in the Hubble diagram. As a consequence, the inferred distance moduli are more accurate, which tightens the calibration of the SN Ia distance ladder. In turn, the improved precision in distance measurements can lead to significantly tighter constraints on the Hubble constant, potentially addressing current tensions in cosmological parameter estimation.

AI keywords:

anomaly detection; differentiable algorithms; AI tooling

116

Continuous normalizing flows in lattice QFT

Author: Mathis Gerdes[1]

[1] *University of Amsterdam*

Corresponding Author: m.gerdes@uva.nl

Recent advances in generative models have demonstrated the potential of normalizing flows for lattice field theory, particularly in mitigating critical slowing down and improving sampling efficiency. In this talk, I will discuss the role of continuous normalizing flows (CNF) or neural ODEs in learning field theories, highlighting their advantages and challenges compared to discrete flow architectures. CNFs enable expressive and scalable transformations while naturally incorporating symmetries of the theory. I will focus on the challenges and importance of equivariance and architectural choices, drawing from applications to both scalar and gauge theories.

AI keywords:

equivariance; neural ODE; normalizing flows; sampling problem

117

Machine Learning for Enhanced 0νββ Searches in LEGEND

Author: Marta Ewelina Babicz[1]

[1] *University of Zurich*

Corresponding Author: marta.babicz@physik.uzh.ch

The LEGEND experiment aims to push the sensitivity of neutrinoless double beta decay ($0\nu\beta\beta$) searches by minimizing backgrounds while leveraging the exceptional energy resolution of high-purity germanium (HPGe) detectors. A key challenge is improving background rejection, particularly through pulse shape discrimination (PSD). Machine learning provides powerful tools to enhance this effort.
We present a transformer-based waveform classification framework designed to distinguish signal-like from background-like events with high accuracy. While simulations play a crucial role in training such models, they inevitably differ from real experimental data, as no simulation can capture every subtle detail of a detector's response. To mitigate potential biases arising from these differences, we integrate Domain Adversarial Neural Networks (DANN), which improve generalization by aligning simulated and real waveforms in a shared feature space.
Beyond waveform classification, we employ semi-supervised data cleaning with Affinity Propagation to cluster waveform structures and train an SVM-based classifier, adapting to evolving detector conditions. These techniques not only strengthen LEGEND's $0\nu\beta\beta$ search but also have broad applications in dark matter detection, neutrino oscillation experiments, and rare event searches at colliders, where precise event classification and background suppression are critical.
This talk will provide insight into how AI-driven methods are reshaping event selection in fundamental physics, offering a pathway toward next-generation discovery.

AI keywords:

Transformers; Domain Adaptation; Simulation-Based Inference; Semi-Supervised Learning

118

Differentiable modeling for calorimeter simulation using diffusion models

Authors: Federico Nardi[1]; Julien Donini[2]; Long Chen[3]; Nicolas R. Gauger[3]; Tommaso Dorigo[4]; Xuan Tung Nguyen[4]

[1] *Università di Padova, LPCA Clermont*

[2] *Université Clermont Auvergne (FR)*

[3] *University of Kaiserslautern-Landau*

[4] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: xuantung.nguyen@pd.infn.it

The design of calorimeters presents a complex challenge due to the large number of design parameters and the stochastic nature of physical processes involved. In high-dimensional optimization, gradient information is essential for efficient design. While first-principle based simulations like GEANT4 are widely used, their stochastic nature makes them non-differentiable, posing challenges in gradient-based optimization. To address this, we propose a machine learning-based approach where we train a conditional diffusion denoising probabilistic model (CDDPM) as a differentiable surrogate for these simulations. The CDDPM not only predicts particle showers based on different particle types and incoming energy levels but also conditions on different detector design variables. Furthermore, we explore post-training adaptation techniques, such as adapter-based fine-tuning, to efficiently specialize the model for new calorimeter conditions without requiring full retraining. This allows for flexible optimization across different calorimeter configurations while maintaining computational efficiency. We evaluate the predictive accuracy of the model and assess its gradient output to demonstrate its potential for the future detectors design and optimization.

AI keywords:

machine learning; diffusion model; gradient-based optimization

119

Fast, accurate, and precise detector simulation with vision transformers

Authors: Andrea Giammanco[1]; Ayodele Ore[2]; Luigi Favaro[None]; Sofia Palacios Schweitzer[3]; Tilman Plehn[2]

[1] *UCLouvain, CP3*

[2] *Heidelberg University, ITP*

[3] *ITP, Heidelberg University*

Corresponding Author: luigi.favaro@uclouvain.be

The speed and fidelity of detector simulations in particle physics pose compelling questions about LHC analysis and future colliders. The sparse high-dimensional data combined with the required precision provide a challenging task for modern generative networks. We present solutions with different tradeoffs, including accurate and precise Conditional Flow Matching and faster coupling-based Normalizing Flow networks. Vision transformers, including autoregressive elements, allow us to reliably simulate the energy deposition in the detector phase space starting from the detailed Geant4 detector response. We also study dimension reduction with latent networks and faster flow-matching generation with bespoke samplers. We evaluate the networks using high-level observables, neural network classifiers, and sampling timings, showing minimum deviations from Geant4 while achieving faster generation. Our results use public benchmark datasets for easier reproducibility and further development.

AI keywords:

conditional generation; surrogate models; vision transformers; flow matching; point cloud;

120

Classification of Radio Sources Through Self-Supervised Representation Learning

Author: Nicolas Baron Perez[1]

[1] *Universität Hamburg*

Corresponding Author: nicolas.baron.perez@uni-hamburg.de

Modern radio telescopes are detecting a large number of radio sources that will be impossible to analyze individually. In particular, the morphological classification of radio galaxies remains a difficult computational challenge.
In this study, we use contrastive learning to classify radio galaxies from the LOFAR Two-meter Sky Survey Data Release 2 (LoTSS-DR2) and propose a new classification procedure.
We have developed a five-step pipeline: (i) Self-supervised training of the encoder. (ii) Search for clusters or high-density regions in the representation space. (iii) Manual cluster curation. (iv) Fine-tuning of the trained encoder with the cluster labels. (v) Deep ensemble training. To ensure the morphological relevance of the representations, we have designed a new random augmentation.
Our results show that the obtained representations encode morphological properties like source extension, the number of source components, the relative intensity of radiation peaks and source bending.
Furthermore, we show that by training a deep ensemble, we are able to provide corresponding class

probabilities increasing the scientific usability of the results.
Finally, we analyse the radio sources in LoTSS-DR2 with a peak flux $F_{\text{peak}} >$ $SI0.75mJy/beam$ in two different largest angular size (LAS) bins: $SI30arcsec \leq$ LAS $\leq$ $SI60arcsec$ and LAS $>$ $SI60arcsec$. We present the morphological classes we found in both LAS bins and discuss their properties.
We demonstrate that radio galaxies can be classified in a semi-supervised manner, enabling a fast analysis and the discovery of data-driven classification schemes.

AI keywords:

Representation learning; Contrastive Learning; Self-supervised Learning

121

lsbi: linear simulation based inference

Authors: Nicolas Mediato-Diaz[1]; Will Handley[None]

[1] *University of Oxford*

Corresponding Authors: nicolas.mediatodiaz@stx.ox.ac.uk, wh260@cam.ac.uk

https://arxiv.org/abs/2501.03921

Simulation-based inference is undergoing a renaissance in statistics and machine learning. With several packages implementing the state-of-the-art in expressive AI [mackelab/sbi] [undark-lab/swyft], it is now being effectively applied to a wide range of problems in the physical sciences, biology, and beyond.

Given the rapid pace of AI/ML, there is little expectation that the implementations of the future will resemble these current first generation neural network-based approaches. This talk will present a new framework for simulation-based inference, linear simulation-based inference (lsbi), which abstracts the core principles of SBI from the specific details of machine learning, implementing a plug-and-play framework of linear and mixture models.

lsbi has several use-cases:

1. It is pedagogically helpful to separate out the general principles of SBI from the specific details of neural networks (particularly for ML skeptics).
2. It is practically useful for producing expressive examples with known ground truths.
3. It is pragmatically useful, since in many cases, lsbi is competitive with neural approaches in terms of accuracy, whilst being faster and more interpretable.

An evolving code-driven PyPI/conda research package is available at:
https://github.com/handley-lab/lsbi

AI keywords:

simulation-based-inference

122

Simulation-based inference for extreme mass ratio inspirals

Author: Philippa Cole[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: philippa.cole@unimib.it

Extreme mass ratio inspirals are a key target for next generation space-based gravitational wave detectors because they have a rich phenomenology that could offer new astrophysics and fundamental physics insights. However, their dynamics are complicated to model, and they will be buried amongst a large population of other sources in the milliHertz frequency band, with a background of non-stationary and non-Gaussian noise. Searching for these systems and measuring their parameters therefore presents a difficult challenge.

Simulation-based inference methods could offer solutions to some of these challenges. I will show parameter estimation results for extreme mass ratio inspiral systems achieved using sequential simulation-based inference, specifically truncated marginal neural ratio estimation. I will highlight the benefits of this approach with respect to traditional likelihood-based methods, and discuss the broader context in which such a pipeline will need to be embedded.

AI keywords:

simulation-based inference

123

Challenges and Innovations in Learning from Heterogeneous Data in Fundamental Physics

Authors: Catherine Biscarat[1]; Jan Stark[1]; Sylvain Caillou[1]

[1] *L2I Toulouse, CNRS/IN2P3, Université de Toulouse*

Corresponding Authors: stark@in2p3.fr, sylvain.caillou@l2it.in2p3.fr

Learning from heterogeneous data is one of the major challenges for AI in the coming years, particularly for what are called 'foundation models'. In fundamental physics, the heterogeneity of data can come from the instruments used to acquire them (the subsystems of large detectors in particle physics or the crossing of modalities in multi-messenger, multi-instrument systems in astrophysics, for example) or from the data themselves when the signal is a superposition of many sources of different nature (as is the case in the forthcoming LISA detector for gravitational waves that will observe the superposition of signals from a large, a priori unknown number of sources of different types). Models capable of learning from these heterogeneous data will need to be able to integrate a common representation of these data in shared latent spaces. We discuss the problems posed by such learning and why it is crucial to solve them for future AI-assisted research in fundamental physics. We provide an overview of the significant work in this area, in particular different integration architectures considered and techniques for latent alignment in AI in general and in fundamental physics in particular. We will cite current projects and summarise the main contributions made and key messages identified during the workshop "Heterogeneous Data and Large Representation Models in Science" [1] held as part of the "Artificial Intelligence for the two infinites [2]" initiative of the AISSAI centre [3] of the French National Centre for Scientific Research (CNRS). The results obtained, key questions and challenges for enabling these models to develop and become fully operational in the near future will be discussed.

[1] https://indico.in2p3.fr/e/AISSAI-TLS

[2] the infinitely small (particle physics) and the infinitely large (cosmology)

[3] https://aissai.cnrs.fr/en/

AI keywords:

heterogeneous data; latent representations, and their alignment; model architectures; training techniques

124

Applications of Machine Learning in Constraining Multi-Scalar Models

Author: Darius Jurčiukonis[1]

[1] *Vilnius University (LT)*

Corresponding Author: darius.jurciukonis@tfai.vu.lt

Machine learning techniques are used to predict theoretical constraints—such as unitarity, boundedness from below, and the potential minimum—in multi-scalar models. This approach has been demonstrated to be effective when applied to various extensions of the Standard Model that incorporate additional scalar multiplets. A high level of predictivity is achieved through appropriate neural network architectures, learning algorithms, and well-prepared training datasets. Machine learning offers a significant computational advantage by enabling faster computations compared to other numerical methods, such as scalar potential minimization. This research investigates the potential of machine learning as an alternative approach for predicting these constraints, potentially improving upon traditional numerical techniques.

AI keywords:

dataset creation; training; pattern recognition; integration of physics and ML

125

Harnessing AI and ML Innovations for High-Luminosity LHC: Transitioning from R&D to Production

Authors: Catherine Biscarat[1]; Jan Stark[1]; Sylvain Caillou[1]

[1] *L2I Toulouse, CNRS/IN2P3, Université de Toulouse*

Corresponding Authors: sylvain.caillou@l2it.in2p3.fr, stark@in2p3.fr

With the upcoming High-Luminosity Large Hadron Collider (HL-LHC) and the corresponding increase in collision rates and pile-up, a significant surge in data quantity and complexity is expected. In response, substantial R&D efforts in artificial intelligence (AI) and machine learning (ML) have been initiated by the community in recent years to develop faster and more efficient algorithms capable of managing this deluge of data. Several projects focused on triggering and offline reconstruction are currently underway. These initiatives have demonstrated highly promising results, offering physics performance comparable to existing algorithms but with the ability to run on modern hardware architectures ("not just CPU") with substantially reduced computation times. As we approach the start of the HL-LHC, the time is ripe to transition these models from the R&D phase to prototyping and eventually to full-scale deployment within the production systems of the experiments. This presentation will discuss ongoing integration efforts that address new questions and challenges within an environment with high throughput and stringent constraints on timing, energy consumption and total cost per collision event. These include both the high-level trigger and offline reconstruction, aiming for the acceleration of model inferences and the execution of complex pipelines on hybrid heterogeneous CPU/GPU and CPU/inference server architectures. Successful

integration of these advances will capitalise on recent AI and ML R&D efforts, helping experiments to efficiently process their data during the HL-LHC era.

AI keywords:

deployment; fast inference; high throughput; production; heterogeneous computing environments

126

Enhancing the development of Cherenkov Telescope Array control software with Large Language Models

Author: Dmitriy Kostunin[1]

Co-authors: Alexandre Strube [2]; Sergo Golovachev [3]; Vladimir Sotnikov [3]

[1] *DESY*

[2] *FZJ*

[3] *JetBrains*

Corresponding Author: dmitriy.kostunin@desy.de

We develop AI agents based on instruction-finetuned large language models (LLMs) to assist in the engineering and operation of the Cherenkov Telescope Array Observatory (CTAO) Control and Data Acquisition Software (ACADA). These agents align with project-specific documentation and codebases, understand contextual information, interact with external APIs, and communicate with users in natural language. We present our progress in integrating these features into CTAO pipelines for operations and offline data analysis. The fast-evolving ACADA codebase is embedded in a vector database and linked to an open-source LLM. This integration enables advanced search, automated code and data model generation, and quality assurance tailored to the project's needs.

AI keywords:

LLM; RAG; code generation; agents

127

Sequential simulation-based inference for cosmological initial conditions

Authors: O Savchenko[None]; Guillermo Franco Abellán[1]; Christoph Weniger[2]; Noemi Anau Montel[3]; Florian List[4]

[1] *GRAPPA Institute, University of Amsterdam*

[2] *GRAPPA, University of Amsterdam*

[3] *Max Planck Institute for Astrophysics*

[4] *University of Vienna*

Corresponding Author: o.savchenko@uva.nl

Knowledge of the primordial matter density field from which the present non-linear observations formed is of fundamental importance for cosmology, as it contains an immense wealth of information about the physics, evolution, and initial conditions of the universe. Reconstructing this density

field from galaxy survey data is a notoriously difficult task, requiring sophisticated statistical methods, advanced cosmological simulators, and exploration of a multi-million-dimensional parameter space. In this talk, I will discuss how sequential simulation-based inference allows us to tackle this problem and simultaneously obtain data-constrained realisations of the primordial dark matter density field together with constraints on the cosmological parameters in a simulation-efficient way for general non-differentiable simulators. In addition, I will describe our novel adaptive learning training strategy and how our results compare to those obtained with classical likelihood-based methods such as Hamiltonian Monte Carlo.

AI keywords:

simulation-based inference, adaptive learning, uncertainty quantification, field-level inference

128

Characterizing, Not Just Detecting: Bayesian Neural Networks for Gravitational-Wave Physics

Author: Costantino Pacilio[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: costantino.pacilio@unimib.it

Gravitational waves provide a powerful means to perform null tests of strong-gravity physics. Statistical methods based on hierarchical inference, adapted from population studies, have been developed to confidently identify potential signatures of new physics. While these methods are well-suited for detection, they provide limited insight into how exotic physics depends on standard degrees of freedom, such as the mass and spin of an observed black hole. In this talk, we present an extension of hierarchical tests that enables the modeling of such dependencies in a flexible and theory-agnostic manner using fully connected neural networks. Additionally, we incorporate Bayesian neural networks and variational inference to model epistemic uncertainty in the network weights, optimizing the hierarchical population likelihood. Finally, we also discuss an alternative optimization strategy based on Gaussian Process Regression.

AI keywords:

variational inference; bayesian neural networs; gravitational waves; hierarchical inference; tests of general relativity.

129

Using Artificial Intelligence to Scan Beyond Standard Model Parameter Spaces

Author: Fernando Souza[1]

[1] *LIP*

Corresponding Author: abreurochasouza@gmail.com

In High Energy Physics, when testing theoretical models of new physics against experimental results, the customary approach is to simply sample random points from the parameter space of the model, calculate their predicted values for the desired observables and compare them to experimental data. However, due to the typically large number of parameters in these models, this process is highly

time consuming and inefficient. We propose a solution to this by adopting optimization algorithms which make use of Artificial Intelligence methods in order to improve the efficiency of this validation task.

A first study compared the performance of three different optimization algorithms (Bayesian, evolutionary and genetic algorithms) at constraining conventional Supersymmetry realizations, when confronted against Higgs mass and Dark Matter relic density constraints and the results show an increase in up to 3 orders of magnitude in sampling efficiency when compared to random sampling.

In a much more challenging scenario, a follow-up analysis was implemented for the scotogenic model, this time using an evolutionary multi-objective optimization algorithm assisted by a novelty detection (ND) algorithm, confronted against experimental constraints coming from the Higgs and neutrinos masses, lepton flavor violating decays, neutrino mixing and the anomalous magnetic moment of the muon. Results show at least 6 orders of magnitude increase in sampling efficiency as well as a better coverage of the parameter space due to the inclusion of a multi-objective cost function. Lastly, the use of ND improved the exploratory capacities of the algorithm, leading to new phenomenology.

AI keywords:

simulation-based inference, anomaly detection, evolutionary algorithms

130

Next generation cosmological analysis with a re-usable library of machine learning emulators across a variety of cosmological models and datasets

Author: Dily Duan Yi Ong[1]

Co-authors: Harry Bevins [1]; Will Handley

[1] *University of Cambridge*

Corresponding Author: dlo26@cam.ac.uk

In recent years, disparities have emerged within the context of the concordance model regarding the estimated value of the Hubble constant H0 [1907.10625] using Cosmic Microwave Background (CMB) and Supernovae data (commonly referred to as the Hubble tension), the clustering σ8 [1610.04606] using CMB and weak lensing, and the curvature ΩK [1908.09139, 1911.02087] using CMB and lensing/BAO, and between CMB datasets. The study of these discrepancies between different observed datasets, which are predicted to be in agreement theoretically by a cosmological model, is called tension quantification.
We approach this problem by producing a re-usable library of machine learning emulators across a grid of cosmological models through detecting cosmological tensions between datasets from the DiRAC allocation (DP192). This library will be released at this conference as part of the pip-installable package unimpeded (https://github.com/handley-lab/unimpeded) and serve as an analogous grid to the Planck Legacy Archive (PLA), but machine learning enhanced and expanded to enable not only parameter estimation (currently available with the MCMC chains on PLA), but also allowing cosmological model comparison and tension quantification. These are implemented with piecewise normalising flows [2305.02930] as part of the package margarine [2205.12841], though alternative density estimation methods can be used. The combination of nested sampling and density estimation allows us to obtain the same posterior distributions as one would have found from a full nested sampling run over all nuisance parameters, but many orders of magnitude faster. This allows users to use the existing results of cosmological analyses without the need to re-run on supercomputers. Currently, a systematic coverage of nine cosmological models and 30 datasets (to be extended) are easily accessible via the unimpeded package using a few lines of code. Hyperparameter tuning for

cosmological normalising flows are explored across a grid of datasets and models with different combinations of architecture (number of hidden layers), learning rate/scheduling and activation function (e.g. sigmoid, tanh and ReLU) for the best performance.

We believe this work represents a significant step forward in cosmological data analysis, providing a versatile, efficient, and user-friendly platform to address current observational tensions and advance our understanding of the Universe.

AI keywords:

anomaly detection, emulators, neural network, normalising flow, machine learning enhanced bayesian statistics

131

A QGP Trigger based on Convolutional Neural Network for the CBM Experiment

Authors: Artemiy Belousov[1]; Ivan Kisel[None]

[1] *Goethe University Frankfurt*

Corresponding Author: i.kisel@compeng.uni-frankfurt.de

The field of heavy-ion experiments, particularly those like the upcoming Compressed Baryonic Matter (CBM) experiment at the Facility for Antiproton and Ion Research (FAIR), requires high-performance algorithms capable of efficient real-time data analysis. The incorporation of machine learning, especially artificial neural networks, into these experiments is a major breakthrough. This report highlights the use of a specialized neural network package, ANN4FLES, tailored for high-performance computing environments, focusing on its application in the CBM experiment to identify and classify events that could indicate the creation of Quark-Gluon Plasma (QGP).

Our study introduces an innovative method using ANN to create a QGP detection system within the First Level Event Selection (FLES) framework, a key component of the CBM experiment. We explore the effectiveness of both fully-connected and convolutional neural networks by training and evaluating them on simulated collision data (Au+Au collisions at 31.2A GeV) created with the Parton-Hadron-String Dynamics (PHSD) model. The findings show that convolutional neural networks significantly outperform their fully-connected counterparts, achieving an impressive accuracy of over 95% on the test data.

This report provides an in-depth look at why the neural network excels in accurately identifying QGP-related events, touching on the complex physics involved. It also covers the critical aspects of neural networks, particularly their relevance to analyzing heavy-ion collisions where detecting quark-gluon plasma is crucial.

AI keywords:

Fully-Connected Neural Network (FCNN), Convolutional Neural Network (CNN), Shapley Additive Explanations

132

Scaling Normalizing Flows for Lattice Gauge Theories

Authors: Christopher Kirwan[1]; Sinéad Ryan[1]; Srikumar Venugopal[2]; Vassilis Vassiliadis[2]

[1] *Trinity College Dublin*

[2] *IBM Research Europe - Dublin*

Corresponding Author: kirwanch@tcd.ie

In varying action parameters in a lattice gauge theory towards a critical point, such as the continuum limit, generic Markov chain Monte Carlo algorithms incur dramatic sampling penalties. Proof-of-principle studies in applying flow-based generative models to lattice gauge theories have suggested that such methods can mitigate against critical slowing down and topological freezing. There remains a question in how normalizing flows perform in state-of-the-art calculations. In this work, we aim to quantify how well these algorithms scale as we increase the number of degrees of freedom in the system. In particular, we study the scaling behaviour of flows for SU(2) and SU(3) gauge theories in two and four spacetime dimensions while also incorporating a variety of architectural improvements to enhance sampling efficiency.

AI keywords:

Normalizing Flows; Variational Inference; Training models at scale

133

Maria Goes Nifty - Simulation, Gaussian Process-Based Reconstruction and Denoising of (Sub-)Millimetre Single-Dish Telescope Data

Authors: Jonas Würzinger[1]; Joshiwa van Marrewijk[2]; Richard Fuchs[1]; Thomas Morris[3]; Tony Mroczkowski[4]

[1] *Technical University of Munich*

[2] *Leiden Observatory, Leiden University*

[3] *Yale University & Brookhaven National Laboratory*

[4] *European Southern Observatory (ESO)*

Corresponding Author: jonas.wuerzinger@tum.de

(Sub-)millimeter single-dish telescopes offer two key advantages compared to interferometers: they can efficiently map larger portions of the sky, and they can recover larger spatial scales. Nonetheless, fluctuations in the atmosphere, the dominant noise source in ground-based observations, limit the accurate retrieval of signals from astronomical sources. We introduce maria (https://thomaswmorris.com/maria/), a versatile, python-based general-purpose simulator for microwave and radio telescopes. Maria generates location-specific weather with accurate turbulence, allowing an in-depth study of optimal removal of the atmosphere in astronomical data. We generate synthetic observations from simulations that account for both atmospheric signal and filtering effects by letting the input array scan in a pre-defined scanning pattern over the astronomical background through the simulated turbulent and evolving atmosphere. The synthetic data are used to optimize the transfer function and various calibration techniques, from which the astronomical map is reconstructed using the Numeric Information Field Theory (NIFTy, https://ift.pages.mpcdf.de/nifty/) package. The contributions from the map and atmosphere to the time series data are described by separate gaussian process models, allowing for a separation and accurate reconstruction of both components. Noise contributions from the detectors and other sources are removed in the reconstruction process. We observe that the NIFTy-based approach leads to a significantly improved map reconstruction for simulated data, while providing an uncertainty quantification of the results. The application of the NIFTy-based reconstruction method to simulated observations by the 50 m Atacama Large Aperture Submillimeter Telescope (AtLAST) is discussed, as well as the possibility to reconstruct CMB contributions.

AI keywords:

gaussian process, variational inference, Bayesian inference, uncertainty

134

Cosmological inference using gravitational waves and normalizing flows

Author: Federico Stachurski[1]

[1] *Univeristy of Glasgow*

Corresponding Author: f.stachurski.1@research.gla.ac.uk

We present a machine learning approach using normalizing flows for inferring cosmological parameters
from gravitational wave events. Our methodology is general to any type of compact binary coalescence
event and cosmological model and relies on the generation of training data representing distributions of
gravitational wave event parameters. These parameters are conditional on the underlying cosmology and
incorporate prior information from galaxy catalogues. We provide an example analysis inferring the
Hubble constant using binary black holes detected during the O1, O2, and O3 observational runs conducted
by the advanced LIGO/VIRGO gravitational wave detectors. We obtain a Bayesian posterior on the Hubble
constant from which we derive an estimate and 1σ confidence bounds of H0 = 74.51 +14.80
−13.63 km s−1 Mpc−1.
We are able to compute this result in O(1) s using our trained normalizing flow model.

AI keywords:

simulation-based inference; normalizing flows; multi layered perceptrons;

135

Continuous Calibration Classifiers via Optimal Transportation Maps

Authors: Chris Pollard[1]; Malte Algren[2]; Tobias Golling[3]

[1] *University of Warwick*

[2] *UNIGE*

[3] *University of Geneva*

Corresponding Author: malte.algren@unige.ch

A calibration of the classifier using a new calibration procedure based on optimal transportation maps is presented. Simultaneous, continuous corrections to the classification probabilities from tagging algorithms in simulation. After application of the derived calibration maps, closure between simulation and observation is achieved. A continuous calibration opens up new possibilities for the future use of jet flavor information in LHC analyses and furthermore serves as a guide for deriving high-dimensional corrections to simulation via transportation maps, an important development for a broad range of inference tasks.

AI keywords:

Optimal Transport; input-convex-neural-networks; foundation model

136

Real-time calibrations for future detectors at FAIR

Author: Valentin Kladov[1]

Co-authors: James Ritman [2]; Johan Messchendorp [3]

[1] *Ruhr-Universität Bochum*

[2] *Research Center Juelich and Ruhr-University-Bochum*

[3] *GSI/FAIR GmbH*

Corresponding Author: v.kladov@gsi.de

The real-time data processing of next-generation experiments at FAIR requires precise event topology reconstruction, which in turn depends on accurate in-situ calibration procedures. Machine learning techniques offer a promising approach to achieving fast and reliable calibrations using continuously available environmental data. In this study, we investigate a neural network-based method for calibrating the Drift Chambers of the HADES detector. By combining Long Short-Term Memory with graph convolutions, we achieve stable and accurate predictions, matching the quality of standard offline calibration across all drift chambers. Moreover, our approach significantly reduces the calibration time, making it well-suited for real-time applications within high-rate data acquisition environments.

AI keywords:

GConv-LSTM, regression, supervised/semi-supervised.

137

Physics Informed Neural Networks for design optimisation of diamond particle detectors for charged particle fast-tracking at high luminosity hadron colliders

Authors: Alessandro Bombini[1]; Alessandro Rosa[2]; Clarissa Buti[1]; Giovanni Passaleva[1]; Lucio Anderlini[1]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *INFN*

Corresponding Author: alessandro.bombini.fi@gmail.com

The next generation of tracking detectors at upcoming and future high luminosity hadron colliders will be operating under extreme radiation levels with an unprecedented number of track hits per proton-proton collision that can only be processed if precise timing information is made available together with state-of-the-art spatial resolution. 3D Diamond pixel sensors are considered as a promising technology to design such future detectors thanks to diamond radiation hardness and high charge carrier mobility, potentially resulting in excellent timing resolution.

Charge collection in 3D diamond detectors is usually obtained by using femtosecond pulses of infrared laser light to induce a phase transition of the diamond lattice to a mixture of graphite and amorphous carbon, which results in conductive electrodes although with relatively high resistivity. The description of ionising particle induced signals in such devices is complicated by the fact that the electric field in the material is modified by the signal propagation through the resistive electrodes, thus requiring an extended application of the classical Ramo-Shockley weighting potential methods. The optimisation of the design of these detectors from the point of view of space and time resolution requires therefore innovative signal simulation techniques.

We discuss a novel approach relying on the numerical resolution of a 3rd order, 3+1-dimensional partial differential equation (PDE) to describe the effect of signal propagation through the resistive electrodes as a time-dependent Ramo-Shockley weighting potential, then integrated with third-party software applications modelling the charge carrier transport across the detector material.

We discuss how such PDE can be obtained as the quasi-stationary approximation of Maxwell Equations and different approaches to their solution for the 3D geometry of the proposed diamond sensors.

Using a small portion of data obtained from an ad hoc devised Spectral Method simulation, we trained a Mixture-of-Experts Physics Informed Deep Neural Network, using Kolmogorov-Arnold Networks (KANs), MLPs and FourierNet as models, and various optimisation recipes, to obtain a meshless solver and used it to infer the effect of the electrode resistance on the time resolution of the diamond detector.

We conclude discussing how a parametric PINN may help in the optimisation of diamond detector design and, more generally, in the study of radiation detectors embedding resistive charge collection elements.

AI keywords:

Physics Informed Neural Network; Kolmogorov-Arnold Network; Mixture-of-Experts; surrogate model

138

AI-Driven Pattern Recognition for Self-tuning PMT Gain Optimization in the Forward Wall Detector

Author: Pavel Kohout[1]

[1] *Palacký University Olomouc*

Corresponding Author: pvlkohout@gmail.com

The Forward Wall (FW) detector in the HADES experiment at GSI/FAIR relies on accurate photomultiplier tube (PMT) gain tuning to ensure precise energy measurements and correct energy measurement range. Traditional calibration methods depend on iterative manual adjustments using cosmic muons, making them time-consuming and susceptible to systematic variations caused by PMT aging and environmental factors.

To automate and enhance calibration accuracy, we introduce an AI-driven self-calibrating algorithm based on pattern recognition in spectral data. A deep neural network (DNN) is trained to identify characteristic patterns in time-over-threshold (ToT) spectra, distinguishing between optimal and shifted gain states. The algorithm detects anomalies in spectral distributions and corrects deviations by predicting the necessary high-voltage (HV) adjustments. Additionally, an anomaly detection module based on unsupervised learning identifies unexpected deviations that may indicate hardware degradation or changing experimental conditions.

This approach improves calibration efficiency, reduces reliance on manual corrections, and enhances long-term detector stability. Validation with cosmic muon and heavy-ion collision data (C+C, Au+Au) demonstrates that the algorithm effectively recognizes gain shift patterns and adapts PMT settings accordingly. By leveraging AI-based pattern analysis, this method offers a scalable solution for automated calibration in large-scale scintillator-based detectors.

AI keywords:

Anomaly detection; Pattern recognition; Reinforcement learning;

139

Physics-Conditioned Diffusion Models for Lattice Gauge Field Theory

Authors: Lingxiao Wang[1]; Qianteng Zhu[None]

Co-authors: Gert Aarts [2]; Kai Zhou [3]; Wei Wang [4]

[1] *Riken*

[2] *Swansea University*

[3] *CUHK-Shenzhen*

[4] *SJTU*

Corresponding Author: qianteng.zhu@riken.jp

We develop diffusion models for simulating lattice gauge theories, where stochastic quantization is explicitly incorporated as a physical condition for sampling. We demonstrate the application of this novel sampler to U(1) gauge theory in two spacetime dimensions and find that a model trained at a small inverse coupling constant can be extrapolated to larger inverse coupling regions without encountering the topological freezing problem. Additionally, the trained model can be employed to sample configurations on different lattice sizes without requiring further training. The exactness of the generated samples is ensured by incorporating Metropolis-adjusted Langevin dynamics into the generation process. Furthermore, we demonstrate that this approach enables more efficient sampling of topological quantities compared to traditional algorithms such as hybrid Monte Carlo and Langevin simulations.

AI keywords:

Physics-informed AI; Diffusion models; Generative models;

141

Neuromorphic Readout for Hadron Calorimeters

Authors: Enrico Lupi[1]; Tommaso Dorigo[2]; Alessandro Breccia[3]

Co-authors: Abhishek [4]; Max Aehle [5]; Muhammad Awais [6]; Riccardo Carroccio [3]; Long Chen [5]; Abhijit Das [7]; Andrea De Vita [1]; Nicholas Ralph Gauger [5]; Ralf Keidel [8]; Jan Kieseler [9]; Anders Mikkelsen [7]; Federico Nardi [10]; Xuan Tung Nguyen [11]; Fredrik Sandin [12]; Kylian Schmidt [9]; Pietro Vischia [13]; Joseph Willmore [14]

[1] *INFN Padova, University of Padova*

[2] *INFN Padova, Luleå University of Technology, MODE Collaboration, Universal Scientific Education and Research Network*

[3] *University of Padova*

[4] *National Institute of Science Education and Research Jatni*

[5] *University of Kaiserslautern-Landau (RPTU), MODE Collaboration*

[6] *INFN Padova, Luleå University of Technology, MODE Collaboration*

[7] *Lund University*

[8] *Karlsruhe Institute of Technology, MODE Collaboration*

[9] *Karlsruhe Institute of Technology*

[10] *University of Padova, Laboratoire de Physique Clermont Auvergne*

[11] *INFN Padova, University of Kaiserslautern-Landau (RPTU)*

[12] *Luleå University of Technology, MODE Collaboration*

[13] *University of Oviedo, MODE Collaboration, Universal Scientific Education and Research Network*

[14] *INFN Padova*

Corresponding Author: enrico.lupi@pd.infn.it

In this work we simulate hadrons impinging on a homogeneous lead-tungstate (PbWO4) calorimeter to investigate how the resulting light yield and its temporal structure, as detected by an array of light-sensitive sensors, can be processed by a neuromorphic computing system. Our model encodes temporal photon distributions in the form of spike trains and employs a fully connected spiking neural network to regress the total deposited energy, as well as the position and spatial distribution of the light emissions within the sensitive material. The model is able to estimate the aforementioned observables in both single task and multi-tasks scenarios, obtaining consistent results in both settings. The extracted primitives offer valuable topological information about the shower development in the material, achieved without requiring a segmentation of the active medium. A potential nanophotonic implementation using III-V semiconductor nanowires is discussed.

AI keywords:

simulation-based inference; neuromorphic computing; real time processing; spiking neural network

142

Transfer Learning for Smart Background Simulation at Belle II

Author: David Giesegh[1]

Co-authors: Boyang Yu [1]; Nikolai Krug [1]; Thomas Kuhr [1]

[1] *LMU Munich*

Corresponding Author: d.giesegh@campus.lmu.de

In experimental particle physics, the development of analyses depends heavily on the accurate simulation of background processes, including both the particle collisions/decays and their subsequent interactions with the detector. However, for any specific analysis, a large fraction of these simulated events is discarded by a selection tailored to identifying interesting events to study. At Belle II, the simulation of the particle collision and subsequent particle decays is much more computationally efficient than the rest of the simulation and reconstruction chain. Thus, the computational cost of generating large simulated datasets for specific selections could be reduced by predicting which events will pass the selection before running the costly part of the simulation. Deep Neural Networks have shown promise in solving this task even when there is no obvious correlation between quantities available at this stage and after the full simulation. These models, however, must be trained on preexisting large datasets, which especially for selections with low efficiencies defeats their own purpose. To solve this issue, we present how a model, pre-trained on multiple different selections for which a lot of training data is available, can be fine-tuned on a small dataset for a new low-efficiency selection or for a known selection with different detector/software conditions.

AI keywords:

transformers; transfer learning; fine-tuning; classification

143

A Generative Geometry Foundation Model for Engineering Applications

Authors: Douglas Boubert[1]; Aamal Hussain[1]; Jamie Donnelly[1]; James-Michael Leahy[1]; Thomas Gessey-Jones[1]; Michalis Michaelides[1]

[1] *PhysicsX*

Corresponding Authors: douglas.boubert@physicsx.ai, thomas.gessey-jones@physicsx.ai

Many scientific and engineering problems are fundamentally linked to geometry, for example, designing a part to maximise strength or modelling fluid flow around an airplane wing. Thus, there is substantial interest in developing machine learning models that can not only operate on or output geometric data, but generate new geometries. Such models have the potential to revolutionise advanced industries from materials manufacturing to medical imaging. However, constructing and training these models presents a range of novel challenges. In this talk, we will discuss a generative geometry model for aircraft, using it as a case study to illustrate some of these challenges and the approaches taken to tackle them. Among other topics, we shall consider how to encode geometry for ML, the difficulties of dataset curation, the advantages of geometric foundation models, and how to predict scalar and field properties of these geometries.

AI keywords:

Generative Geometry Models; Geometric Deep Learning; Design Optimization; Functional Variational Autoencoders; Dataset Curation

144

SFTs: a scalable data-analysis framework for long-duration gravitational-wave signals

Author: Rodrigo Tenorio[1]

[1] *Università degli studi di Milano-Bicocca*

Corresponding Author: rodrigo.tenorio@unimib.it

We introduce a framework based on Short-time Fourier Transforms (SFTs) to analyze long-duration gravitational wave signals from compact binaries. Targeted systems include binary neutron stars observed by third-generation ground-based detectors and massive black-hole binaries observed by the LISA space mission, for which we present a pilot application. Leveraging differentiable and GPU-parallelizable gravitational-wave models, ours is an extremely fast, scalable, and parallelizable implementation of the gravitational-wave inner product, the key building block of all gravitational-wave data treatments. Overall, we achieve computing cost reduction in evaluating an inner product of three to five orders of magnitude, depending on the specific application, with respect to a standard approach. By speeding up this low-level element so massively, SFTs and differentiable GPU-accelerated waveform models provide an extremely promising solution for current and future gravitational-wave data-analysis problems.

AI keywords:

differentiable programming; GPUs; low-latency;

145

21cmEMU3: an emulator of 21cmFAST summary observables

Author: Daniela Breitman[None]

Corresponding Author: daniela.breitman@sns.it

The cosmic dawn (CD) of the first luminous objects and eventual reionisation (EoR) of the intergalactic medium (IGM) remain among the greatest mysteries in modern cosmology. The 21-cm line is one of the most powerful probes of these crucial moments in the history of the Universe, providing a clean window into both cosmology and astrophysics. Current 21-cm observations are upper limits on the power spectrum (PS) from instruments such as LOFAR, HERA, and MWA. Upcoming instruments, such as the SKA, will provide a detection of the 21-cm PS as well as a 3D map of the neutral hydrogen content of the Universe. A great deal of computational resources are required to improve our understanding of the CD/EoR via Bayesian inference. Past works have found that artificial neural networks significantly reduce the computational costs while accurately reproducing posteriors obtained with the simulator. While most past works focus on emulating a single summary statistic, we focus on emulating several as it is the synergy between different probes that allows us to produce the best constrains. Previously, we presented 21cmEMU1, an emulator that allows us to perform state-of-the-art inferences in over 10^5 times faster than traditional inferences. In this work, we present 21cmEMU3, an emulator of six summary statistics, including, among others, the cylindrical (2D) 21-cm power spectrum, and the ultra violet luminosity functions. This emulator is trained on more realistic simulations from 21cmFAST than 21cmEMU1. In this new database, we have a total of about 50k samples where we vary 10 astrophysical parameters and σ_8. 21cmEMU3 is the first emulator of cylindrical 21-cm PS, as well as the first score-based generative emulator of the 2D PS. 21cmEMU3 is also the first LSTM emulator of four of the remaining five summary statistics. We apply 21cmEMU3 on previous and upcoming 21-cm PS upper limits from the HERA instrument.

AI keywords:

score-based diffusion; LSTM; Bayesian inference;

146

Hadron Identification Prospects With Granular Calorimeters

Author: Andrea De Vita[1]

Co-authors: Abhishek [2]; Alessandro Breccia [3]; Enrico Lupi [1]; Federico Nardi [4]; Fredrik Sandin [5]; Jan Kieseler [6]; Joseph Willmore [7]; Kylian Schmidt [6]; Long Chen [8]; Max Aehle [8]; Muhammad Awais [9]; Nicholas Ralph Gauger [8]; Pietro Vischia [10]; Ralf Keidel [11]; Riccardo Carroccio [3]; Tommaso Dorigo [12]; Xuan Tung Nguyen [13]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *National Institute of Science Education and Research, India*

[3] *University of Padova*

[4] *University of Padova, Laboratoire de Physique Clermont Auvergne*

[5] *Luleå University of Technology, MODE Collaboration*

[6] *Karlsruhe Institute of Technology*

[7] *INFN Padova*

[8] *University of Kaiserslautern-Landau (RPTU), MODE Collaboration*

[9] *INFN Padova, Luleå University of Technology, MODE Collaboration*

[10] *University of Oviedo, MODE Collaboration, Universal Scientific Education and Research Network*

[11] *Karlsruhe Institute of Technology, MODE Collaboration*

[12] *INFN Padova, Luleå University of Technology, MODE Collaboration, Universal Scientific Education and Research Network*

[13] *INFN Padova, University of Kaiserslautern-Landau*

Corresponding Author: andrea.devita@pd.infn.it

In this work we consider the problem of determining the identity of hadrons at high energies based on the topology of their energy depositions in dense matter, along with the time of the interactions. Using GEANT4 simulations of a homogeneous lead tungstate calorimeter with high transverse and longitudinal segmentation, we investigated the discrimination of protons, positive pions, and positive kaons at 100 GeV. The analysis focuses on the impact of calorimeter granularity by progressively merging detector cells and extracting features like energy deposition patterns andtiming information. Two machine learning approaches, XGBoost and fully connected deep neural networks, were employed to assess the classification performance across particle pairs. The results indicate that fine segmentation improves particle discrimination, with higher granularity yielding more detailed characterization of energy showers. Additionally, the results highlight the importance of shower radius, energy fractions, and timing variables in distinguishing particle types. The XGBoost model demonstrated computational efficiency and interpretability advantages over deep learning for tabular data structures, while achieving similar classification performance. This motivates further work required to combine high- and low-level feature analysis, e.g., using convolutional and graph-based neural networks, and extending the study to a broader range of particle energies and types.

AI keywords:

XGBoost, Boosted Decision Trees, Deep Neural Network, Classification Task

147

PXD background generation using generative models

Authors: Fabio Novissimo[1]; Nikolai Krug[2]; Thomas Kuhr[2]

[1] *Ludwig-Maximilians-Universität München*

[2] *LMU Munich*

Corresponding Author: f.novissimo@physik.uni-muenchen.de

The Pixel Vertex Detector (PXD) is the innermost detector of the Belle II experiment. Information from the PXD, together with data from other detectors, allows to have a very precise vertex reconstruction. The effect of beam background on reconstruction is studied by adding measured or simulated background hit patterns to hits produced by simulated signal particles. This requires a huge sample of statistically independent PXD background noise hit patterns to avoid systematic biases, resulting in a huge amount of storage due to the high granularity of the PXD sensors. As an efficient way of producing background noise, we explore the idea of an on-demand PXD background generator realised using generative models. In order to evaluate the quality of generated background we measure physical quantities which are sensitive to the background in the PXD.

AI keywords:

GAN; diffusion model; generative models

148

Electron and Proton Classification with AMS ECAL Using Convolutional Vision Transformers and Domain Adaptation

Author: Berk Türk[1]

Co-authors: Bilge Demirköz [2]; Emre Akbaş [2]; Zhili Weng [3]

[1] *Middle East Technical University*

[2] *Middle East Techinal University*

[3] *CERN,MIT*

Corresponding Author: beturk@cern.ch

Alpha Magnetic Spectrometer (AMS-02) is a precision high-energy cosmic-ray experiment on the ISS operating since 2011 and has collected more than 228 billion particles. Among them, positrons are important to understand the particle nature of dark matter. Separating the positrons from cosmic background protons is challenging above 1 TeV. Therefore, we use state-of-the-art convolutional and transformer models, CoAtNet and Convolutional Vision Transformer (CvT), that employ the shower signals from the ECAL to classify the electrons/positrons in the dominant cosmic proton background. We created sets of electrons, positrons, and protons events from the ISS data and Monte Carlo Simulation in the energy range between 0.2-2 TeV by applying various data quality cuts on reconstructed variables obtained from the sub-detectors. Initially, since ECAL showers are not tunned in the AMS MC, our MC trained models show a lower proton rejection on the ISS data. To accommodate the difference between the training and test domain distributions, we implemented domain adaptation with the CoAtNet and CvT to mitigate this dataset bias/domain shift. We also trained domain adaptation with a set of well-reconstructed 1 electron charge ISS events without electron/proton labels at TeV energy order as the target dataset. We evaluated the models between 1-2 TeV energy using ISS and MC events with the proton rejection vs. electron efficiency and proton rejection vs. energy at near 90% electron efficiency plots. We performed experiments using various training and validation dataset combinations and other hyperparameters with the CvT and CoAtNet. Among them, the best models are obtained with the 1-2TeV MC events as training data and half of the labeled 1-2 TeV ISS events as validation data. Using domain adaptation with the CoAtNet, we obtained a maximum proton rejection at 88% electron efficiency on the ISS data. We also rejected all of the MC protons at higher than 99.8% electron efficiency with both CvT and CoAtNet. At 90% electron efficiency, the proton rejection power of the CvT and CoAtNet is 5 and 7 times higher than the proton rejection power of the AMS's Boosted Decision Tree and ECAL Likelihood Estimator for MC events in the 1-2 TeV range. We created another dataset in the 50-200 GeV energy range in which electrons and protons are labeled independently from the ECAL in the ISS Data. After hyperparameter tuning on the CoAtNet and applying focal loss, the models's proton rejection factors are 2 times higher than AMS ECAL LHD in the 50-200 GeV range.

AI keywords:

Convolutional Vision Transformers, domain adaptation, ECAL Shower Classification, focal loss

149

Calorimeter Reconstruction with Graph Neural Networks

Author: Amir Farbin[1]

Co-authors: Debottam Bakshi Gupta [1]; Kyle Jones [1]; Mohammadali Ghaznavi [1]

[1] *University of Texas at Arlington*

Corresponding Author: afarbin@cern.ch

Graph Neural Networks (GNNs) have emerged as powerful tools for particle reconstruction in high-energy physics experiments, particularly in calorimeters with irregular geometries, such as those used in the ATLAS experiment. In this work, we present a GNN-based approach to reconstruct particle showers, improve energy resolution, spatial localization, and particle identification. We discuss the model architecture, training strategies, and performance benchmarks, demonstrating the advantages of GNNs over conventional techniques. Our findings highlight the potential of GNNs to enhance calorimeter-based event reconstruction, paving the way for more precise measurements in future collider experiments.

AI keywords:

Graph neutral networks; clustering; reconstruction; calorimeter

150

Improved gravitational wave parameter estimation with SBI and secondary mode marginalization

Author: Chiara Anselmo[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: c.anselmo@campus.unimib.it

Simulation-based inference (SBI) has emerged as a powerful tool for parameter estimation, particularly in complex scenarios where traditional Bayesian methods are computationally intractable. In this work, we build upon a previous application of SBI, based on truncated neural posterior estimation (TNPE), to estimate the parameters of a gravitational wave post-merger signal, using real data from the GW150914 event. We extend this approach by incorporating marginalization over secondary modes, which are often neglected in standard analyses under the assumption that their contribution is weaker than noise and can therefore be absorbed into a Gaussian noise model. However, when secondary modes have a non-negligible impact, this assumption can introduce biases in the inferred parameters of the dominant mode. By explicitly accounting for these contributions within the SBI framework, we investigate whether this refinement improves both the accuracy and robustness of parameter estimation compared to standard Bayesian methods.

AI keywords:

simulation-based inference; neural posterior estimation; normalizing flows; regression

151

Background Enrichment to improve Anomaly Detection

Authors: Caterina Doglioni[1]; Deepak Kar[2]; Pratik Jawahar[None]; Sukanya Sinha[2]

[1] *University of Manchester*

[2] *University of Witwatersrand*

Corresponding Author: pratik.jawahar@cern.ch

Model-independent anomaly detection for Beyond the Standard Model (BSM) searches in high-energy physics faces significant challenges due to the lack of tractable methods to build rich background priors as well as inherent uncertainties in simulated background processes. Traditional unsupervised ML approaches to anomaly detection, commonly train models on background samples produced by a single physics generator (e.g., Pythia) using a fixed generator tune. This comes with the risk of overfitting to generator-specific features thereby increasing sensitivity to non-anomalous processes that deviate from the limited background representations. To address this, we present a novel method that enhances background modelling by aggregating samples from multiple generators (Pythia, Herwig, Sherpa) and generator tunes, capturing a broader spectrum of possible background variations. We train an unsupervised variational autoencoder (VAE) augmented with contrastive learning objectives, which enforce separation between latent space clusters corresponding to each generator and tune. This enriched background representation ensures that generator-specific features are encoded into the distinct clusters, while anomalous signals —which do not align with any generator's characteristics —are projected outside these regions. The resulting reduction in false positives thereby improves anomaly detection performance. We compare performance across different anomaly metrics, different VAE-based architectures like (but not restricted to) Normalizing

Flow augmented VAEs to arrive at a broader picture of model-agnostic, VAE-based anomaly detection for new physics searches. We also attempt an empirical study of the use of contrastive methods to build small foundation models for broader Physics tasks. We present our work through an open-source python package-style repository called BEAD, which is designed to be a modular VAE-based anomaly detection toolkit.

AI keywords:

VAE; Contrastive Methods; Anomaly Detection; Normalizing Flows

152

QCD physics exploration with regressive and generative machine learning

Author: Kai Zhou[1]

[1] *CUHK-Shenzhen*

Corresponding Author: zhoukai@cuhk.edu.cn

Recent advances in machine learning have unlocked transformative approaches to longstanding challenges in fundamental physics. In this talk, I will present our latest work that harnesses physics-driven deep learning to tackle two intertwined frontiers: solving inverse problems in Quantum Chromodynamics (QCD) and deploying generative models for statistical physics and field theory.

Inverse problems in QCD—such as the reconstruction of hadronic spectral functions and the extraction of the dense matter equation of state—are inherently ill-posed, with traditional methods often falling short of reliably capturing subtle physical details. By embedding physical priors and symmetry constraints directly into neural network architectures and leveraging automatic differentiation, our approach significantly improves the precision and stability of the extracted observables.

In parallel, we have explored the use of advanced generative models—including diffusion models and Fourier-flow techniques—as global samplers in lattice field theory. These methods recast stochastic quantization into a machine-learned framework, mitigating challenges like critical slowing down and topological freezing while enabling efficient sampling of complex quantum configurations.

Together, these innovations illustrate how machine learning can bridge data-driven insights with rigorous physics principles to decode the rich phenomenology of QCD matter under extreme conditions. This presentation will outline our methodologies, highlight key numerical and theoretical results, and discuss the promising prospects for AI-driven research in fundamental physics.

AI keywords:

simulation-based inference, generative models, automatic differentiation, diffusion models, ML based inverse problems

153

Physics-informed generative models for scattering amplitude reconstruction

Author: Gloria Montana[1]

[1] *Jefferson Lab*

Corresponding Author: gmontana@jlab.org

While cross sections are the fundamental experimental observables in scattering processes, the full quantum dynamics of the interactions are encoded in the complex-valued scattering amplitude. Since cross sections depend only on the squared modulus of the amplitude, reconstructing the complete information from nuclear and particle physics experiments becomes a challenging inverse problem. In this talk, I present a physics-informed Generative Adversarial Network (GAN) approach to reconstruct the amplitude in $2 \longrightarrow 2$ elastic pion-pion scattering, focusing on spin-0 particles to avoid the complexities from the spin of external particles or coupled channels. To address the loss of phase information, I introduce constraints from fundamental properties of scattering amplitudes in the generator's loss function: one enforcing unitarity for probability conservation and another ensuring causality through conditions on the scattering phase shifts. I will show how these constraints enable accurate amplitude reconstruction within uncertainties. I will also outline steps toward extending generative models for amplitude-level unfolding in more complex hadronic reactions.

AI keywords:

Physics-informed generative models; Generative Adversarial Networks; Amplitude-level unfolding; AI/ML for inverse problems

154

A Lorentz-Equivariant Transformer for All of the LHC

Authors: Huilin Qu[1]; Jesse Thaler[2]; Jonas Spinner[3]; Tilman Plehn[4]; Víctor Bresó Pla[3]

[1] *CERN*

[2] *MIT, IAIFI*

[3] *University of Heidelberg*

[4] *Heidelberg University, ITP*

Corresponding Author: victor.breso@gmail.com

We show that the Lorentz-Equivariant Geometric Algebra Transformer (L-GATr) yields state-of-the-art performance for a wide range of machine learning tasks at the Large Hadron Collider. L-GATr represents data in a geometric algebra over space-time and is equivariant under Lorentz transformations. The underlying architecture is a versatile and scalable transformer, which is able to break symmetries if needed. We demonstrate the power of L-GATr for amplitude regression and jet classification, and then benchmark it as the first Lorentz-equivariant generative network. For all three LHC tasks, we find significant improvements over previous architectures.

AI keywords:

Transformer; geometric deep learning; generative modeling; equivariant neural networks

155

Improving Fast Radio Burst Localizations Using Simulation-Based Inference

Author: David Krejcik[None]

Corresponding Author: d.krejcik@gmail.com

Fast radio bursts (FRBs) are extremely brief and bright flashes of radio waves originating from distant galaxies. Localizing FRBs to or within a host galaxy is key to exploring their physical origin(s) and using them as cosmological probes. However, poor uv-coverage of interferometric arrays and susceptibility to calibration errors can make FRBs exceptionally hard to localize accurately. I explored whether convolutional neural networks can extract a reliable localization estimate from a single FRB observation. I have built a Very Long Baseline radio Interferometry (VLBI) simulator for producing representative FRB maps as a training data set for Simulation-Based Inference (SBI). I use an embedded convolutional neural network to overcome the computational limitations of the fully connected SBI pipeline with parameters jointly learned during training. I train SBI on the dimensionally reduced feature latent space, sample the SBI posterior, and show that this setup can estimate localization uncertainty when trained on simulated images of FRB VLBI observations. I will comment on the limitations and challenges of this approach, discuss its implications and results of testing on real observations of the hyper-active repeating FRB~20220912A, and discuss the viability of using SBI with an embedded neural network for creating a fast and reliable pipeline for the localization of FRBs.

AI keywords:

Simulation-based inference, Convolutional networks, Uncertainty quantification

156

Deep Learning and Simulation-Based Inference for Radiation Damage Modeling in Space Telescopes: Euclid Case Study

Author: David Krejcik[None]

Corresponding Author: d.krejcik@gmail.com

The Euclid Space Telescope aims to map the geometry of the dark universe with unprecedented precision, requiring exceptional data fidelity from its Visible Instrument (VIS). However, radiation damage introduces charge transfer inefficiencies (CTI), distorting observations over time. Trap pumping is a novel technique for localizing and characterizing radiation-induced defects in the detector surface, but it remains time-consuming and often fails to identify complex defect structures. My ongoing work investigates the potential of machine learning to enhance the accuracy and efficiency of trap pumping analysis, as well as the use of simulation-based inference to bridge two interrelated physical models of radiation damage simulations which are used for the generation of training datasets.

I am utilizing convolutional neural networks (CNNs) and multi-channel architectures to localize and characterize detector defects in both simulated and in-orbit calibration data. The best estimates from the nominal trap pumping method serve as a part of the training dataset for learning a quantifiable defect representation in real and simulated data, with the goal of reducing the number of necessary calibration images for successful defect detection. I will assess the performance of different models under varying calibration time constraints and present predictions for the evolution of both the nominal and CNN-based trap detection performance throughout Euclid's mission, accounting for the expected accumulation of radiation damage.

AI keywords:

Convolutional networks, pattern recognition, Simulation-based inference

157

Density matrix estimation using autoregressive networks

Author: Tomasz Stebel[1]

Co-authors: Dawid Zapolski [1]; Piotr Bialas [1]; Piotr Korcyl [1]

[1] *Jagiellonian University*

Corresponding Author: tomasz.stebel@uj.edu.pl

The density matrix of a quantum system provides complete information about its entanglement. Using generative autoregressive networks, we show how to estimate the matrix elements for the small quantum spin chain. Using a density matrix, we calculate Renyi entanglement entropies as well as Shanon entropy at zero temperature.

AI keywords:

autoregressive neural networks, neural Monte-Carlo

158

Probabilistic DiffusionNet as a PDE surrogate endowed with mechanistic uncertainty quantification

Author: Sanmitra Ghosh[None]

Co-authors: James-Michael Leahy [1]; Aamal Hussain [1]; Bachir Djermani [1]; Michalis Michaelides [1]

[1] *PhysicsX*

Corresponding Author: sanmitra.ghosh@physicsx.ai

Geometric deep learning models are being adopted across science and engineering to estimate large-scale PDE solutions for varying boundary conditions. While accurate uncertainty quantification (UQ) is essential for better decision-making for a variety of downstream tasks like optimisation and control, these models rarely produce efficient and effective UQ. Moreover, most UQ methods focus on constructing distributions on point-wise parameters to elicit uncertainty, and so fail to exploit the useful inductive biases of geometric model architectures. To address this challenge, we propose a probabilistic modification of the DiffusionNet architecture, widely used in surface learning tasks, by introducing latent random variables that are derived from a stochastic reformulation of the underlying mechanism — the diffusion process. We demonstrate that approximate Bayesian treatment of these mechanistic latent variables yields superior UQ performance on standard datasets without sacrificing predictive performance compared to other prevalent models and UQ methods, while enjoying a lower computational cost and interpretability.

AI keywords:

DiffusionNet, Variational Inference, PDE surrogate, Neural Operators, Geometric Deep Learning

159

Machine Learning for Event Reconstruction in the CMS Phase-2 High Granularity Calorimeter Endcap

Author: CMS Collaboration[None]

The high-luminosity era of the LHC will offer greatly increased number of events for more precise Standard Model measurements and Beyond Standard Model searches, but will also pose unprecedented challenges to the detectors. To meet these challenges, the CMS detector will undergo several upgrades, including the replacement of the current endcap calorimeters with a novel High-Granularity Calorimeter (HGCAL). To make optimal use of this innovative detector, new and original algorithms are being devised. A dedicated reconstruction framework, The Iterative Clustering (TICL), is being developed within the CMS Software (CMSSW). This new framework is designed to fully exploit the high spatial resolution and precise timing information provided by HGCAL. Several key ingredients of the object reconstruction chain already rely on Machine Learning techniques and their usage is expected to further develop in the future. In the presentation, the existing reconstruction strategies will be presented stressing the role played by ML techniques to exploit the information provided by the detector. The areas where ML techniques are expected to play a role in the future developments will be also discussed.

AI keywords:

simulation-based inference, anomaly detection

160

Enhancing Pulse Shape Discrimination with Gradient Boosted Decision Trees and Twin Neural Networks

Author: Alain Fauquex[1]

Co-authors: Laura Baudis [1]; Marta Ewelina Babicz [1]

[1] *University of Zurich*

Corresponding Author: marta.babicz@physik.uzh.ch

High-purity germanium (HPGe) detectors play a critical role in nuclear physics experiments, including searches for neutrinoless double-beta decay. Traditional pulse shape discrimination (PSD) methods help distinguish signal from background events in such detectors. However, the performance of these traditional PSD methods declines at lower energies (500 ⊠ keV). This low-energy regime is promising for exploring beyond-standard-model physics, such as bosonic dark matter, electron decays, and sterile neutrinos. To improve sensitivity, we developed a novel machine learning pipeline for PSD in HPGe detectors. Our pipeline combines a twin neural network (TWINNN) that encodes waveforms into a 64-dimensional latent space with a gradient-boosted decision tree (GBDT) that interprets the encoded data with the help of additional human-curated input features. The algorithm was trained on the "Majorana Demonstrator Publicly Released dataset for AI/ML". Our approach leverages both the TWINNN's ability to model complex, non-linear behavior and capture temporal context, as well as the GBDTs interpretability and ability to make use of additional human-curated input features. When applied to waveforms from a Majorana Demonstrator calibration run, our approach improved classification accuracy from 87.5% to 92% and increased near-surface event identification by 13%. Two energy spectrum peaks, corresponding to single-site and multi-site events, were excluded during training, yet the model correctly retained or removed these events during full-spectrum evaluation, demonstrating generalization into unseen energy regions. Future work will involve Monte Carlo validation of the low-energy calibration spectrum, further study of latent-space topology with a variational inference model, and improved PSD generalization across a broader energy range. This work underscores the transformative potential of machine learning in nuclear physics instrumentation and data analysis while offering a new method for event characterization in rare-event searches and fundamental physics experiments.

AI keywords:

Twin Neural Network; Pulse Shape Discrimination; Latent Space; Interpretability; Gradient Boosted Decision Tree

161

Towards a Seismology Foundation Model

Author: Waleed Esmail[1]

Co-authors: Alexander Kappes [1]; Cherstine Thomas [1]; Stuart Russell [1]

[1] *University of Münster*

Corresponding Author: waleed.physics@gmail.com

We introduce a SeismoGPT, a foundation model for seismology that leverages transformer-based architectures to model seismic waveforms. Inspired by natural language processing techniques. This approach tokenizes continuous seismograms by dividing them into fixed-length patches, where each patch represents a sequence of waveform samples. These patches serve as input tokens to the transformer model, enabling the application of self-attention mechanisms to learn temporal dependencies in seismic data. The model is trained in an autoregressive manner, predicting the next patch given previous ones, making it well-suited for forecasting seismic signals.

The architecture of SeismoGPT consists of a convolutional embedding that encodes raw seismic patches into a higher-dimensional representation, followed by positional embeddings to retain temporal order. A causal transformer processes these embeddings, utilizing self-attention to capture long-range dependencies in the seismic sequence. The final prediction head generates the next patch, allowing the model to reconstruct or extend seismic waveforms iteratively. This approach enables the efficient modeling of complex seismic patterns, offering potential applications in earthquake prediction, early warning systems, and seismic noise mitigation, which is particularly relevant for third-generation gravitational wave detectors such as the Einstein Telescope (ET). Accurately predicting seismic waveforms can help in reducing the impact of seismic noise on the detector's mirrors, thereby enhancing sensitivity to gravitational waves from sources such as supermassive black hole binaries.

AI keywords:

Seismology, Gravitational Waves, Transformers, Einstein Telescope, Time series Forecasting

162

Autonomous Fabry-Perot cavity locking via deep reinforcement learning

Authors: Andrea Svizzeretto[1]; Mateusz Bawaj[1]

[1] *University of Perugia*

Corresponding Author: mateusz.bawaj@pg.infn.it

This work explores the application of Reinforcement Learning (RL) to the control of a Fabry-Perot (FP) optical cavity, a key component in interferometric gravitational-wave detectors. By leveraging RL's inherent ability to handle high-dimensional non-linear systems, the project aims to achieve robust and autonomous cavity locking—a process typically hindered by elevated finesse values, mirror velocity, and non-linear dynamics.
A custom simulator reproducing the cavity's electric field dynamics serves as the basis of a Gymnasium environment, where an RL agent is trained to acquire lock under realistic conditions. This training is optimized by rewarding effective control actions, and the environment includes strategies to mitigate the SimToReal gap by addressing latency and noise sources—thereby reducing discrepancies between simulated and physical systems. Beyond simulation, an experimental setup is proposed to verify and refine the training process.

Ultimately, the goal is to integrate simulation insights and experimental validation to enable a seamless transition of the trained RL agent from a virtual platform to real-world cavity locking. Future work will focus on enhancing control reliability and efficiency for more complex optical systems.

AI keywords:

reinforcement learning, simulation to reality gap, real-time inference

163

Open Framework for Synthetic Fraud Datasets via Generative AI: Insights from Industrial Secondment at IBM

Authors: Micol Olocco[1]; Pierre Feillet[2]

[1] *TU Dortmund*

[2] *IBM*

Corresponding Author: micol.olocco@cern.ch

This work presents an open framework for generating synthetic transactional datasets, addressing the twin challenges of data scarcity and privacy in fraud research. Conducted as an industry secondment at IBM France Lab Saclay within the SMARTHEP Network—a European project fostering collaboration between High Energy Physics and Industry—our approach leverages Generative AI Agents to simulate both legitimate and fraudulent behaviors within banking time series data. Initially, our methodology employed Markov chain-based simulations to generate baseline transactional patterns. However, to capture the nuanced dynamics of real-world activities, we transitioned to an LLM-based Chain of Thought approach, enabling the creation of adaptive fraud scenarios that more realistically mimic complex banking behaviors. The resulting synthetic dataset provides a resource for researchers to develop and benchmark fraud detection methods while mitigating issues related to proprietary data constraints. Future plans include releasing the simulator code and datasets to foster targeted collaboration on developing anomaly detection techniques for fraud detection.

In addition, exploratory thinking on applying these automation principles to High Energy Physics will be presented, particularly for automation during data taking. Drawing a conceptual parallel with previous work on the infrastructure for deployment and evaluation of LHCb trigger configurations, this idea suggests that similar strategies might be adapted to manage complex operational processes in experimental settings.

AI keywords:

Generative AI; Simulation; Automation

164

Transformers + Normalizing Flows for parameter estimation of overlapping gravitational waves in next generation detectors

Authors: Elena Cuoco[1]; Federico De Santi[2]; Ik Siong Heng[3]; Lucia Papalini[4]; Massimiliano Razzano[4]

[1] *University of Bologna*

[2] *Istituto Nazionale di Fisica Nucleare*

[3] *University of Glasgow*

[4] *University of Pisa and INFN Pisa*

Corresponding Author: federico.desanti@pi.infn.it

In the next decade, the third generation of ground-based gravitational wave detectors, such as the european Einstein Telescope, is expected to revolutionize our understanding of compact binary mergers. With a 10 factor improvement in sensitivity and an extended range towards lower frequencies, Einstein Telescope will enable the detection of longer-duration signals from binary black hole and binary neutron star coalescences, with expected rates up to ~10^5 events per year. However, the inevitable presence of overlapping signals poses a severe challenge to parameter estimation analysis pipelines.
In this talk, I will describe a foundation model for parameter estimation, leveraging the synergy of two state-of-the-art machine learning architectures: Transformers and Normalizing Flows. The Transformer component efficiently encodes the complex temporal structures of overlapping signals, capturing long-range dependencies, while Normalizing Flows provide a flexible, efficient representation of the high-dimensional posterior distributions of source parameters.
I will show how this hybrid approach enables rapid and accurate inference, even for low-SNR and highly correlated events. By significantly reducing the computational cost while maintaining accuracy, this framework represents a crucial step toward integrating machine learning-driven inference into real-data analysis pipelines for third-generation detectors. I further present performance benchmarks on simulated data, showcasing the potential for real-time parameter estimation, and discuss future developments.

AI keywords:

Transformers, Normalizing Flows, Simulation based inference

165

Identifying and Mitigating Machine Learning Biases for the Gravitational Wave Detection Problem

Author: Narenraju Nagarajan[1]

Co-author: Christopher Messenger [1]

[1] *University of Glasgow*

Corresponding Author: n.nagarajan.1@research.gla.ac.uk

Our work identifies the sources of 11 interconnected machine learning (ML) biases that hinder the generalisation of supervised learning models in the context of gravitational wave (GW) detection. We use GW domain knowledge to propose a set of mitigation tactics and training strategies for ML algorithms that aim to address these biases concurrently and improve detection sensitivity. We empirically prove that our approach leads to [i] detection sensitivities that rival current matched filtering pipelines in real noise at low false alarm rates, [ii] the ability to handle out-of-distribution noise power spectral densities, [iii] the ability to strongly reject non-Gaussian transient noise artefacts, and [iv] data efficient learning of the detection problem.

Via the injection study introduced in the Machine Learning Gravitational-Wave Search Challenge, we show that our search pipeline (Sage) detects ~11.2% more signals than the benchmark PyCBC search at a false alarm rate of 1 per month and >48% than previous machine learning based detection pipelines. In light of the identified biases, we demonstrate that existing detection sensitivity metrics are unreliable for machine-learning pipelines and discuss the trustworthiness of ML results. By studying machine-learning biases and conducting empirical investigations to understand the reasons for performance improvement/degradation, we aim to address the need for interpretability of machine-learning methods in GW science. Link to the paper

AI keywords:

AI Interpretability; Machine Learning Biases; Data Efficient Learning; Detection;

166

Latest developements in the CATHODE anomaly detection method

Authors: Chitrakshee Yede[1]; Gregor Kasieczka[1]; Louis Moureaux[1]; Tore von Schwartz[1]

[1] *Universität Hamburg*

Corresponding Author: tore.von.schwartz@uni-hamburg.de

The Standard Model of particle physics has been successful in describing fundamental particles and their interactions, yet it fails to explain concepts like dark matter or the hierarchy problem, motivating the search for physics beyond the Standard Model. Despite an extensive search program at the LHC, no hints for new physics have been found so far. Anomaly detection has been introduced as a bridge between generic searches and searches targeting a specific signal. CATHODE (Classifying Anomalies THrough Outer Density Estimation) is a two-step anomaly detection method that first uses a generative model to produce an in-situ estimate of the background and subsequently isolates signal-like events with a classifier.
We present the most recent developments to CATHODE, improving its reliability and versatility in uncovering potential new physics signals.

AI keywords:

anomaly detection; generative models; Cathode

167

Evaluating Two-Sample Tests for Validating Generators in Precision Sciences

Authors: Marco Letizia[1]; Riccardo Torre[1]; Samuele Grossi[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: samuele.grossi@ge.infn.it

Deep generative models have become powerful tools for alleviating the computational burden of traditional Monte Carlo generators in producing high-dimensional synthetic data. However, validating these models remains challenging, especially in scientific domains requiring high precision, such as particle physics. Two-sample hypothesis testing offers a principled framework to address this task. We propose a robust methodology to assess the performance and computational efficiency of various metrics for two-sample testing, with a focus on high-dimensional datasets. Our study examines tests based on univariate integral probability measures, namely the sliced Wasserstein distance, the mean of the Kolmogorov-Smirnov statistics, and the sliced Kolmogorov-Smirnov statistic. Additionally, we consider the unbiased Fréchet Gaussian Distance and the Maximum Mean Discrepancy. Finally, we include the New Physics Learning Machine, an efficient classifier-based test leveraging kernel methods. Experiments on both synthetic and realistic data show that one-dimensional projection-based tests demonstrate good sensitivity with a low computational cost. In contrast, the classifier-based test offers higher sensitivity at the expense of greater computational demands.
This analysis provides valuable guidance for selecting the appropriate approach—whether prioritizing efficiency or accuracy. More broadly, our methodology provides a standardized and efficient framework for model comparison and serves as a benchmark for evaluating other two-sample tests.

AI keywords:

Two sample test; Models evaluation; Simulation-based inference

168

Gravitational-wave posterior post-processing with normalizing flows

Author: Michael Williams[1]

Co-author: Konstantin Leyde [1]

[1] *University of Portsmouth*

Corresponding Author: michael.williams3@port.ac.uk

Bayesian inference is essential for understanding the compact binaries that produce gravitational waves detected by the LIGO-Virgo-KAGRA collaboration. Performing this inference is computationally expensive and often has to be repeated multiple times with different models, e.g. different approximations of General Relativity. These repeated analyses always start from scratch, which is highly inefficient. This process could be improved by leveraging the similarities between analyses to significantly reduce the computational cost.

In this work, we propose a novel machine-learning-based method to address this inefficiency. We use normalizing flows to post-process the posterior distribution from an initial analysis to reflect a new likelihood function. This approach produces a posterior distribution and evidence estimate that are consistent with a full reanalysis but at significantly reduced computational cost. We demonstrate the effectiveness of this method in various scenarios, including updating sky maps with galaxy catalogue information and reanalysing data with alternative waveform models.

AI keywords:

normalizing flows, Bayesian inference, neural networks

169

Linear machine-learning approaches for accurate atomistic simulations

Author: Riccardo Dettori[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: riccardo.dettori@dsf.unica.it

Molecular dynamics (MD) simulations are a fundamental tool for investigating the atomistic behavior of complex systems, offering deep insights into reaction mechanisms, phase transitions, and emergent properties in both condensed and soft matter. Recent advances in machine learning (ML) have determined a paradigm shift in atomistic simulations, allowing the development of force-fields that closely mimic quantum mechanical interactions with exceptional accuracy and efficiency—achieving this at a fraction of the computational cost of ab initio methods. However, while standard non-linear ML potentials such as Gaussian Approximation Potentials and Neural Network Potentials deliver excellent descriptions of the atomic environment, their numerous fitting parameters often restrict the size of the systems and the duration of simulations due to increased computational demands from high-dimensional parameter spaces. Furthermore, realizing these potentials is labor-intensive,

as they generally require extensive training datasets and are vulnerable to overfitting. In this presentation, I will introduce the Chebyshev Interaction Model for Efficient Simulation (ChIMES), which leverages a linear expansion in Chebyshev polynomials to accurately reproduce atomic forces, energies and stress tensors in molecular and condensed-phase systems while requiring comparatively less data. I will present overall findings and discusse future perspectives: these ChIMES force-fields for atomistic simulations are pivotal for the multiscale computational characterization of crystalline oxides like HfO_2. These materials are promising candidates for the coatings of multilayer systems in the Einstein Telescope, where a detailed understanding of their structure and atomic environment is essential for predicting performance under stress and the relationship between mechanical losses and thermoelastic properties.

AI keywords:

chebychev polynomials; linear models; active learning

170

Deep Learning for Full-Event Reconstruction in Imaging Atmospheric Cherenkov Telescopes: A Transfer Learning Approach with CTLearn

Author: Daniel Nieto Castaño[1]

[1] *Institute for Particle and Cosmos Physics - UCM*

Corresponding Author: d.nieto@ucm.es

Arrays of imaging atmospheric Cherenkov telescopes (IACTs) are exceptional instruments for probing the very-high-energy gamma-ray sky. These telescopes focus Cherenkov light, emitted from air showers initiated by very-high-energy gamma rays and cosmic rays, onto the camera plane. A high-speed camera then digitizes the longitudinal development of the air shower, capturing its spatial, temporal, and calorimetric information. From these images, the properties of the primary very-high-energy particle that initiated the air shower can be inferred: the primary particle can be classified as either a gamma ray or a cosmic ray, and its energy and incoming direction can be estimated. This process, known as full-event reconstruction, is essential for the array's sensitivity to gamma rays and can be enhanced using machine learning techniques. We present a deep-learning-driven full-event reconstruction approach applied to IACT events, where transfer learning is explored as a strategy to reduce the computational demands inherent to the technique. For this purpose, we use simulated data from the Large-Sized Telescope 1 of the Cherenkov Telescope Array Observatory and CTLearn. CTLearn is an open-source Python package that provides a backend for training deep-learning models for the reconstruction of IACT events using TensorFlow.

AI keywords:

event reconstruction, image recognition, convolutional neural networks

171

Accelerating Femtoscopic Studies with Machine Learning for Source Function Modeling

Authors: Baran Hashemi[1]; Carla Zeyn[2]; Laura Fabbietti[None]; Lukas Heinrich[1]; Maximilian Korwieser[1]

[1] *Technical University of Munich*

[2] *Technical Uniseritz of Munich*

Corresponding Authors: maximilian.korwieser@cern.ch, carla.zeyn@web.de

Femtoscopy probes the strong interaction between hadrons via two-particle correlation functions. The ALICE collaboration has recently measured these functions with unprecedented precision, including those involving strange (Λ, Ξ, Ω) and charm (D±) quarks. Extracting the final-state interactions requires solving the Schrödinger equation, with the accurate modeling of the source function—describing particles'relative emission distances—posing a key challenge. Advanced models like CECA (Common Emission in CATS) improve our understanding of emission processes but are computationally intensive, limiting simultaneous fits. For the first time, we propose leveraging machine learning (ML) to model the source. The ML model will emulate CECA, providing fast, accurate source modeling and efficient computation of correlation functions, by significantly expediting the analysis of correlation data.

AI keywords:

Normalizing Flow; simulated-based inference

172

Large Physics Models: Towards a collaborative approach with Large Language Models and Foundation Models

Authors: Kristian Gonzales Barman[1]; Sascha Caron[None]

[1] *U Gent*

Corresponding Author: kristiancampbell.gonzalezbarman@ugent.be

This work explores ideas and provides a potential roadmap for the development and evaluation of physics-specific large-scale AI models, which we call Large Physics Models (LPMs). These models, based on foundation models such as Large Language Models (LLMs) - trained on broad data - are tailored to address the demands of physics research. LPMs can function independently or as part of an integrated framework. This framework can incorporate specialized tools, including symbolic reasoning modules for mathematical manipulations, frameworks to analyse specific experimental and simulated data, and mechanisms for synthesizing theories and scientific literature. We begin by examining whether the physics community should actively develop and refine dedicated models, rather than relying solely on commercial LLMs. We then outline how LPMs can be realized through interdisciplinary collaboration among experts in physics, computer science, and philosophy of science. To integrate these models effectively, we identify three key pillars: Development, Evaluation, and Philosophical Reflection. Development focuses on constructing models capable of processing physics texts, mathematical formulations, and diverse physical data. Evaluation assesses accuracy and reliability by testing and benchmarking. Finally, Philosophical Reflection encompasses the analysis of broader implications of LLMs in physics, including their potential to generate new scientific understanding and what novel collaboration dynamics might arise in research. Inspired by the organizational structure of experimental collaborations in particle physics, we propose a similarly interdisciplinary and collaborative approach to building and refining Large Physics Models. This roadmap provides specific objectives, defines pathways to achieve them, and identifies challenges that must be addressed to realise physics-specific large scale AI models.

AI keywords:

Large Language Model, Foundation Model

173

The dangers of learning structured data with Normalising Flows: the dense matter equation of state

Author: Jessica Irwin[1]

Co-authors: Christopher Messenger [1]; Ik Siong Heng

[1] *University of Glasgow*

Corresponding Author: j.irwin.1@research.gla.ac.uk

We discuss data compression methods and evidence of learned structure in using Normalising Flows to perform the conditional mapping of nuclear equation of state data given observed parameters from gravitational wave signals of binary neutron star mergers. We use a convolutional autoencoder to compress unified high density equations of state - including data from the neutron star crust - to a lower dimensional latent representation, preserving unique features of the individual equations of state. We find that the Normalising Flow shows evidence of learning underlying structure of high density phenomenological equations of state but struggles to interpolate between training samples regardless of Flow architecture. To address this issue, we present an additional Normalising Flow method to augment data during training, mitigating inherent grid-like structure and alleviating cost associated with traditionally expensive equation of state data generation. This work promotes the rapid inference of the neutron star equation of state from multiple gravitational wave events. This is especially important for next generation ground based detectors where neutron star merger events are expected to be frequent and the assumption of a unique fixed crust is no longer valid.

AI keywords:

simulation-based inference; normalising flows; data interpretation

174

TrackFormers Part 2: Enhanced Transformer-Based Models for High-Energy Physics Track Reconstruction

Authors: Antonio Ferrer-Sanchez[1]; Evgeniy Shalugin[2]; Nadezhda Dobreva[3]; Roberto Ruiz de Austri[None]; Sascha Caron[None]; Slav Pshenov[4]; Uraz Odyurt[5]; Yue Zhao[6]; Zef Wolffs[None]

[1] *Valencia*

[2] *Radboud University*

[3] *CERN, Radboud and Nikhef*

[4] *Nikhef*

[5] *University of Twente*

[6] *SURF, NL*

Corresponding Author: yue.zhao@surf.nl

High-Energy Physics experiments are rapidly escalating in generated data volume, a trend that will intensify with the upcoming High-Luminosity LHC upgrade. This surge in data necessitates critical revisions across the data processing pipeline, with particle track reconstruction being a prime candidate for improvement. In our previous work, we introduced “TrackFormers”, a collection of Transformer-based one-shot encoder-only models that effectively associate hits with expected tracks. In this study, we extend our earlier efforts by incorporating loss functions that account for inter-hit correlations, conducting detailed investigations into (various) Transformer attention mechanisms, and a study on the reconstruction of higher-level objects. Furthermore we discuss new datasets that allow the training on hit level for a range of physics processes. These developments collectively aim to boost both the accuracy, and potentially the efficiency of our tracking models, offering a robust solution to meet the demands of next-generation high-energy physics experiments.

AI keywords:

transformers, inference, pattern recognition, foundation models

175

Agent-based code generation for the Gammapy framework

Author: Dmitriy Kostunin[1]

Co-authors: Alexandre Strube [2]; Sergo Golovachev [3]; Vladimir Sotnikov [3]

[1] *DESY*

[2] *FZJ*

[3] *JetBrains*

Corresponding Author: dmitriy.kostunin@desy.de

Software code generation using Large Language Models (LLMs) is one the most successful application of the modern AI. Foundational models are very efficient when applied to popular frameworks and libraries, which benefit from documentation, code examples, and strong community support. However, many specialized scientific libraries lack these resources and often have unstable programming interfaces under active development, making it challenging for models trained on limited or outdated data. In this work, we address these issues for the Gammapy library by developing an agent capable of writing, executing, and validating code in a specialized environment. We present a web-based demo that has been tested with early users to gather feedback. This contribution outlines our progress, describes our approach, and discusses our future plans

AI keywords:

code generation; code validation; agents; LLM

176

RoBiTE: A Foundation Model for Irregular and Sparse Time-Series Analysis

Author: Uros Zivanovic[1]

Co-authors: Andre Scaffidi [2]; Martin de los Rios [2]; Roberto Trotta [3]

[1] *University of Trieste*

[2] *SISSA*

[3] *SISSA, Italy*

Corresponding Author: uros.zivanovic@studenti.units.it

Analyzing irregular and sparse time-series is a widespread problem in fundamental physics, astronomy, climate science and many other fields. This talk presents the Rotary Bidirectional Transformer Encoder (RoBiTE), a novel Transformer-based architecture for multi-dimensional irregular time-series and sparse data, designed as a foundation model for general time-series interpolation and object classification. Our method consists of a pre-training phase, where the model is trained on interpolation, and a subsequent fine-tuning phase where the learned representation is adapted to down-stream tasks, such as classification. We highlight the performance of our algorithm on a wide variety of physics datasets including the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC) and simulated recoil events in a liquid xenon time projection chamber for

direct dark matter detection. We compare our method to other popular models for irregular time-series such as S5 and RoFormer, showing that our approach can out-compete the current state-of-the-art.

AI keywords:

transformer, irregular time-series, sparse data

177

Event Tokenization and Next-Token Prediction for Anomaly Detection at the LHC

Author: Ambre Visive[1]

Co-authors: Clara Nellist [1]; Polina Moskvitina [2]; Roberto Ruiz de Austri [3]; Sascha Caron [2]

[1] *Nikhef - University of Amsterdam*

[2] *Nikhef - Radboud University*

[3] *Valencia University, IFIC*

Corresponding Author: ambre.visive@cern.ch

Advances in Machine Learning, particularly Large Language Models (LLMs), enable more efficient interaction with complex datasets through tokenization and next-token prediction strategies. This talk presents and compares various approaches to structuring particle physics data as token sequences, allowing LLM-inspired models to learn event distributions and detect anomalies via next-token (or masked token) prediction. Trained only on background events, the model reconstructs expected physics processes. At inference, both background and signal events are processed, with reconstruction scores identifying deviations from learned patterns—flagging potential anomalies. This event tokenization strategy not only enables anomaly detection but also represents a potential new approach for training a foundation model at the LHC. The method is tested on simulated proton-proton collision data from the Dark Machines Collaboration and applied to a four-top-quark search, replicating ATLAS conditions during LHC Run 2 ($\sqrt{s} = 13$ TeV). Results are compared with other anomaly detection strategies.

AI keywords:

anomaly detection; tokenization; Large-Language Model; transformers; next-token prediction

178

GPU Accelerated Nested Sampling

Author: David Yallup[1]

Co-author: Will Handley

[1] *University of Cambridge*

Corresponding Author: dy297@cam.ac.uk

Nested Sampling is a Monte Carlo method that performs parameter estimation and model comparison robustly for a variety of high dimension and complicated distributions. It has seen widespread usage in the physical sciences, however in recent years increasingly it is viewed as part of a legacy code base, with GPU native paradigms such as neural simulation based inference coming to the fore.

In this work we demonstrate that we can effectively reformulate Nested Sampling to a form that is highly amenable to modern GPU hardware, taking unique advantage of vectorization opportunities to accelerate numerical inference to state of the art levels. We provide a public implementation of this code, and in this contribution will explore its application to a number of inference problems such as Gravitational Wave parameter estimation and CMB cosmology.

AI keywords:

GPU, MCMC, Bayesian, Sampling, Vectorization

179

Open-source, Cross-detector Comparisons for Machine Learning Reconstructions in Neutrino Telescopes

Author: Jeffrey Lazar[1]

[1] *Université Catholique de Louvain*

Corresponding Author: jlazar@icecube.wisc.edu

There are currently many Cherenkov neutrino telescopes being deployed and designed across the world. These detectors are exploring new optical sensors and geometric configurations to maximize their physics goals. Alongside detector R&D, machine learning (ML) has become established as a promising avenue for reconstructions in these detectors; however, there has not been a consistent comparison of the performance of these proposed detector geometries or existing ML-based reconstruction methods. This contribution presents a recent effort to simulate geometries comparable to existing and proposed telescopes using Prometheus, an open-source simulation library. On these datasets, we compare reconstruction performance for ML-based techniques using the open-source GraphNeT ML library. We will present the simulation sets and relative performance of each geometry across several reconstructed quantities, and summarize what this can teach us about detector design and ML-based reconstruction methods.

AI keywords:

Open-source, reconstruction, design optimization

180

Foundation models for black holes

Author: Renze Oosterhuis[1]

Co-authors: Heino Falcke [2]; Jakob Knollmüller [1]; Jesse Vos [3]; Sascha Caron

[1] *Radboud University Nijmegen*

[2] *Radboud University Nijmegen / ASTRON*

[3] *KU Leuven*

Corresponding Author: renze.oosterhuis@ru.nl

Black holes represent some of the most extreme environments in the universe, spanning vast ranges in mass, size, and energy output. Observations from the Event Horizon Telescope (EHT) have provided an unprecedented opportunity to directly image black holes, with future plans aiming to create time-resolved movies of their evolution. To fully leverage these observations, we need theoretical

models that can interpret them with high fidelity and within practical time constraints.
General Relativistic MagnetoHydroDynamic (GRMHD) simulations are the primary tool for modeling the plasma dynamics around accreting black holes. While these simulations provide detailed insights into accretion flows and jet formation, their extreme computational cost limits the range of black hole parameters that can be explored. In particular, variations in black hole spin and grid configurations require costly re-simulations, making it infeasible to construct fully comprehensive theoretical predictions. Within the ERC-funded Synergy project Blackholistic, we address this challenge by integrating advanced numerical simulations with cutting-edge deep learning techniques. Drawing inspiration from successful applications in atmospheric simulations (https://arxiv.org/abs/2405.13063), our approach aims to leverage foundation models to learn the underlying flow dynamics of accreting black hole systems, enabling rapid interpolation across simulation data. We currently base our foundation model on Universal Physics Transformers (UPTs), a type of neural operator using a uniform learning paradigm for a wide range of spatio-temporal problems where they have been shown to be applicable and well-performing (https://arxiv.org/abs/2402.12365).
By extending UPT architectures to curved spacetime, our framework promises a paradigm shift in GRMHD modeling, offering potential speedups of several orders of magnitude in computational efficiency while enhancing model flexibility and accuracy. Preliminary tests with transformer architectures have yielded encouraging results, and ongoing efforts are focused on scaling up model architectures, expanding training datasets, and harnessing increased computational resources. This work paves the way for real-time analysis and interpretation of observational data, marking a significant advance in the fusion of artificial intelligence with fundamental astrophysical research.

AI keywords:

Universal Physics Transformers; ML-based interpolation; Foundation models for PDEs; AI-accelerated fluid dynamics; Transformers

181

Unsupervised Machine Learning for Anomaly Detection in LHC Collider Searches

Authors: Adelina D'Onofrio[1]; Antonio D'Avanzo[1]; Antonio Giannini[2]; Bernardino Spisso[1]; Elvira Rossi[1]; Fabrizio Norcia[1]; Francesco Alessandro Conventi[1]; Francesco Cirotto[1]; Giuliano Gustavino[1]; Graziella Russo[1]; Michela D'Emilia[None]; Stefano Giagu[3]; Valerio Ippolito[1]

[1] *Istituto Nazionale di Fisica Nucleare*

[2] *University of Science and Technology of China*

[3] *Sapienza Università di Roma and Istituto Nazionale di Fisica Nucleare*

Corresponding Author: antonio.davanzo@na.infn.it

Searches for new physics at the LHC traditionally use advanced simulations to model Standard Model (SM) processes in high-energy collisions. These are then compared with predictions from new-physics theories like dark matter and supersymmetry. However, despite extensive research, no definitive signs of physics beyond the Standard Model (BSM) have been found since the Higgs boson's discovery.
This lack of direct discoveries has motivated the development of model-independent approaches to complement existing hypothesis-driven analyses. Unsupervised machine learning offers a novel paradigm for collider searches, enabling the identification of anomalous events without assuming a specific new-physics model or prior theoretical expectations.
Anomaly detection has become an area of increasing interest in the high-energy physics (HEP) community 1. Machine learning techniques provide a powerful framework for identifying events in data that deviate significantly from the expected background-only hypothesis. Unlike traditional supervised classification methods, which require labeled datasets where each event is assigned a known category (signal or background), anomaly detection methods operate in an unsupervised or weakly supervised manner. This allows them to learn from unlabeled data and detect deviations indicative of potential new physics.
A significant breakthrough in fully unsupervised machine learning has been reported by the ATLAS collaboration [2], where a Variational Recurrent Neural Network (VRNN) was trained directly on

recorded jet data. This method establishes an anomaly detection signal region (SR) by selecting the hypothesized X particle based purely on its structural incompatibility with background jets.
Starting from the seminal work reported by the ATLAS collaboration 1 we will review the latest efforts in anomaly detection in fully hadronic final state within the ATLAS collaboration. Two main AD techniques will be discussed: Deep Transformer and Graph Anomaly Detection (EGAT, GIN).
First results obtained with the LHC Olympics dataset are reported, along with their initial applications to search for high mass diboson resonances in fully hadronic final states using $\sqrt{s}$ = 13 TeV pp collisions with the ATLAS detector.
1 The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics, Rep. Prog. Phys. 84 (2021) 124201
[2] Phys. Rev. D 108 (2023) 052009

AI keywords:

Anomaly Detection; Graph; GNN

182

ExpertSim: Fast Particle Detector Simulation Using Mixture of Experts for Generative Models

Authors: Patryk Będkowski[1]; Jan Dubiński[2]; Filip Szatkowski[2]; Tomasz Trzciński[3]; Przemysław Rokita[1]; Kamil Deja[2]

[1] *Warsaw University of Technology*

[2] *Warsaw University of Technology; IDEAS NCBR*

[3] *Warsaw University of Technology; Tooploox*

Corresponding Authors: jan.dubinski.dokt@pw.edu.pl, patryk.bedkowski.stud@pw.edu.pl

High-energy physics experiments at the Large Hadron Collider (LHC) at CERN rely on simulations to model particle interactions and understand experimental data. These simulations, crucial for reconstructing collision events, are traditionally performed using Monte Carlo-based methods, which are highly computationally demanding. With hundreds of thousands of CPU cores dedicated to these tasks annually, the need for more efficient, high-fidelity alternatives is pressing.

Recently, generative deep learning models have emerged as a promising solution, offering faster synthetic data generation. However, applying standard generative architectures to calorimeter simulations remains challenging due to the complex and multi-modal nature of detector responses. Different particles produce diverse energy deposition patterns, making it difficult for a single model to capture the full variability without sacrificing accuracy.

In our work, we focus on simulating the Zero Degree Calorimeter (ZDC) in the ALICE experiment at CERN, which plays a critical role in measuring the energy of non-interacting nucleons in heavy-ion collisions. The ZDC responses fall into multiple distinct categories, corresponding to different types of particle interactions and energy depositions. Attempting to model all of these variations with a single generative network leads to trade-offs between fidelity and efficiency, as a single model lacks the capacity to fully represent the complex, structured nature of the data distribution.

To overcome these limitations, we propose ExpertSim, a novel Mixture-of-Generative-Experts (MoE) architecture designed specifically for ZDC simulation. Instead of relying on a single generative model, ExpertSim employs multiple specialized experts, each trained to simulate a specific subset of the data distribution. A router network dynamically assigns incoming particle events to the most appropriate expert based on their physical properties, ensuring that each expert is specialized in a particular response pattern. This division of tasks improves the overall fidelity of the simulation while maintaining computational efficiency.

A key component of our approach is the Expert Differentiation Loss, which ensures that each expert specializes in a distinct subset of the data rather than redundantly modeling overlapping distribu-

tions. This loss function penalizes similarity between experts by encouraging diversity in their generated outputs. By explicitly maximizing the difference between experts' mean energy intensities, Expert Differentiation Loss forces the model to partition the data more effectively, leading to clearer specialization and higher-quality generated responses.

In addition to the expert-based architecture, ExpertSim incorporates diversity regularization, intensity constraints, and an auxiliary regressor, which enhance the accuracy of the generated responses. The diversity regularization mitigates GAN mode collapse by encouraging variation in generated samples, while the intensity regularization ensures that the total deposited energy aligns with real detector signals. Furthermore, the auxiliary regressor aids in learning spatial correlations in energy deposition, improving the geometric consistency of the simulations.

Through extensive experiments, we demonstrate that ExpertSim outperforms existing generative models on the task of simulating the response of the ZDC. Our model achieves a 15% reduction in Wasserstein distance between the distribution of real and generated data compared to prior methods while preserving the significant computational speedup that generative models offer over Monte Carlo-based approaches.

AI keywords:

generative models, fast simulation, mixture of experts, generative adversarial networks

183

Machine Learning for \(K^0\) Event Reconstruction in the LHCf Experiment

Author: Andrea Paccagnella[1]

[1] *Istituto Nazionale di Fisica Nucleare*

Corresponding Author: andrea.paccagnella@unifi.it

The LHCf experiment aims to study forward neutral particle production at the LHC, providing crucial data for improving hadronic interaction models used in cosmic ray physics. A key challenge in this context is the reconstruction of events containing \(K^0\) mesons, which often involve multiple calorimetric hits.

To address this, we developed a machine learning pipeline that employs multiple neural networks to classify and reconstruct such events. The pipeline consists of four stages: (1) event selection, determining whether an event contains four particles, (2) photon/neutron discrimination, (3) event tagging into four specific topologies based on the distribution of photons between the two calorimeter towers, and (4) position and energy regression for each detected photon.

The model takes as input the energy deposits recorded by the two calorimetric towers of ARM2 and the energy deposits in the four pairs of silicon detectors oriented along the x and y axes. The network architecture is designed to process these heterogeneous data sources, allowing for a precise reconstruction of the event topology. Preliminary results, obtained with a dataset of 10k simulated events, show that the classification networks reach over 80% accuracy in selecting relevant events and distinguishing photon/neutron interactions. These promising results highlight the potential of deep learning techniques in enhancing event reconstruction at LHCf and lay the groundwork for further improvements with larger datasets and refined models.

AI keywords:

Neural networks, Event classification, Energy regression, Multi-modal learning, Manipulate complex graph topologies

184

Point-cloud based diffusion model for hadronic showers

Authors: Anatolii Korol[1]; Frank Gaede[2]; Gregor Kasieczka[3]; Martina Mozzanica[4]; Thorsten Buss[4]

[1] *Deutsches Elektronen-Synchrotron (DESY)*

[2] *DESY*

[3] *Universität Hamburg*

[4] *University of Hamburg*

Corresponding Author: martina.mozzanica@uni-hamburg.de

Simulating showers of particles in highly-granular detectors is a key frontier in the application of machine learning to particle physics. Achieving high accuracy and speed with generative machine learning models can enable them to augment traditional simulations and alleviate a major computing constraint.
Recent developments have shown how diffusion based generative shower simulation approach that do not rely on a fixed structure, but instead generates geometry-independent point clouds, are very efficient. We present a novel transformer-based architecture as an extension to the CaloClouds 2 architecture that was previously used for simulating electromagnetic showers in the highly granular electromagnetic calorimeter of ILD. The attention mechanism allows to generate complex hadronic showers from pions with more pronounced substructure in the electromagnetic and hadronic calorimeter together. This is the first time that ML methods are used to generate hadronic showers in highly granular imaging calorimeters.

AI keywords:

transformers, diffusion model, fast simulations, point clouds

185

A neural likelihood estimator for fast evidence computation

Author: Luca Negri[1]

[1] *Utrecht University, Nikhef*

Corresponding Author: lnegri@nikhef.nl

The computational costs of gravitational wave inference is expected to exponentially rise with the next generation of detectors: both the complexity and the amount of data itself will be much higher, requiring a complete rethinking of current parameter estimation methods to produce accurate science without prohibitive resources usage.
This work will present a novel way of dramatically reducing the computational costs of Markov-chain-monte-Carlo algorithms by approximating the analytical Bayesian likelihood with a Neural likelihood estimator. This method obtains compatible posteriors and returns the correct Bayesian evidence, requiring only a fraction of waveforms computations compared to standard methods.

AI keywords:

Simulation-based inferece; parameter estimation; bayesian statistichs; speed-up

186

End-to-end Sinkhorn AutoEncoder Latent Diffusion Model for Fast Particle Physics Simulation

Authors: Mikołaj Kita[1]; Jan Dubiński[2]; Przemysław Rokita[1]; Tomasz Trzciński[3]; Kamil Deja[2]

[1] *Warsaw University of Technology*

[2] *Warsaw University of Technology; IDEAS NCBR*

[3] *Warsaw University of Technology; Tooploox*

Corresponding Authors: mikolajkita@gmail.com, jan.dubinski.dokt@pw.edu.pl

Simulations play a crucial role in understanding the complex dynamics of particle collisions at CERN' s Large Hadron Collider (LHC). Traditionally, Monte Carlo-based simulations have been the primary tool for modeling these interactions, but their high computational cost presents significant challenges. Recently, generative machine learning models have emerged as an efficient alternative, offering the potential to drastically reduce simulation time while maintaining physical accuracy. Among these, diffusion models have demonstrated state-of-the-art performance in particle simulations, inspired by their success in computer vision. However, despite their fidelity, standard diffusion models suffer from prohibitively long generation times, limiting their practicality for High Energy Physics (HEP) applications.

To address this limitation, we turn to latent diffusion models, which accelerate generation by operating within a learned latent space rather than directly in pixel space. These models leverage powerful vision-based autoencoders to compress data, enabling significantly faster sampling without sacrificing quality. However, conventional latent diffusion models, such as Variational AutoEncoders (VAEs), impose arbitrary regularization constraints on the latent space—typically enforcing a Gaussian prior using KL-divergence. While suitable for standard computer vision tasks, such constraints can limit expressivity and accuracy when modeling complex HEP data distributions.

In this work, we propose a Sinkhorn-AutoEncoder Latent Diffusion Model (SAE-LDM), which improves upon traditional latent diffusion models by leveraging an end-to-end Sinkhorn AutoEncoder (SAE). Instead of imposing a predefined latent structure via KL-divergence, SAE directly minimizes the Wasserstein Distance between the encoded data distribution and a learned prior. This approach allows the model to better capture intricate data patterns without requiring a reparameterization trick, making it particularly well-suited for HEP simulations. By running the diffusion process within the SAE latent space, we ensure both efficient generation and high-fidelity reconstructions. Moreover, our framework jointly optimizes both the autoencoder and diffusion model using the Sinkhorn algorithm, leading to a more structured and expressive latent representation.

We evaluate our method on the task of simulating the response of the neutron Zero Degree Calorimeter (ZDC) in the ALICE experiment at CERN. Our model achieves a 50× speedup in generation time compared to Monte Carlo-based simulations, being orders of magnitude faster than pixel-space diffusion models and achieving inference speed comparable to Generative Adversarial Networks (GANs). At the same time, our approach achieves the highest simulation fidelity among all evaluated generative models on this task.

AI keywords:

generative models, fast simulation, diffusion models, autoencoders, sinkhorn loss

187

The future of cosmological inference

Author: Alessio Spurio Mancini[None]

Corresponding Author: alessio.spuriomancini@rhul.ac.uk

In this talk I will introduce a new paradigm for cosmological inference, enabled by recent advances in machine learning and its underlying technology. By combining emulation, differentiable and

probabilistic programming, scalable gradient-based sampling, and decoupled Bayesian model selection, this framework scales to extremely high-dimensional parameter spaces and enables complete Bayesian analyses—encompassing both parameter estimation and model selection—in a fraction of the time required by conventional approaches. I will demonstrate its application to various Stage IV cosmological survey configurations, tackling parameter spaces of approximately 150 dimensions that are inaccessible to standard techniques. I will also show how this framework can be used to test competing gravity theories. Finally, I will illustrate how a field-level analysis of Euclid cosmic shear data could definitively confirm or refute the recent DESI results pointing to dynamical dark energy.

AI keywords:

simulation-based inference, surrogate modelling, hardware acceleration, differentiable programming, probabilistic programming

188

A Universal Prior for Galaxy Morphologies: Building a Diffusion Model from All the Observations

Author: Sebastian Wagner-Carena[1]

[1] *Flatiron / NYU*

Corresponding Author: swagnercarena@gmail.com

A major challenge in both simulation and inference within astrophysics is the lack of a reliable prior model for galaxy morphology. Existing galaxy catalogs are heterogeneous and provide an impure representation of underlying galaxy structures due to instrument noise, blending, and other observational limitations. Consequently, priors on galaxy morphology typically rely on either simplistic analytic models or more complex models trained on a limited subset of high-resolution observations. Building on previous work that leverages diffusion models to learn priors from incomplete data, we extend this framework to jointly learn two priors: one for uncorrelated light along random sightlines and another for light associated with galaxies. This approach enables us to learn a cleaner prior on galaxy morphology and to sample from the deblended posterior of an observation. We demonstrate our method's ability to robustly infer both distributions through empirical validation. We apply our framework to build a galaxy morphology prior using images from both the Legacy Survey and the HST COSMOS catalog, showcasing its adaptability across heterogeneous datasets. Finally, we illustrate the power of this learned prior for simulation-based inference by using it to significantly tighten dark matter constraints derived from strong gravitational lensing observations. Our results highlight the potential of data-driven priors to improve astrophysical modeling and inference in the era of large-scale sky surveys.

AI keywords:

Diffusion models; Generative models; Simulation-based inference

189

Machine Learning-Driven Anomaly Detection in Dijet Events with ATLAS

Authors: Benjamin Nachman[1]; Dennis Noll[2]; Johnny Raine[3]; Kees Benkendorfer[4]; Knut Zoch[4]; Luc Le Pottier[2]; Sam Klein[3]; Sascha Diefenbacher[2]; Tobias Golling[3]

[1] *Lawrence Berkeley National Laboratory*

[2] *LBNL*

[3] *University of Geneva*

[4] *Harvard*

Corresponding Author: tobias.golling@unige.ch

This contribution discusses an anomaly detection search for narrow-width resonances beyond the Standard Model that decay into a pair of jets. Using 139 fb−1 of proton-proton collision data at sqrt(s) = 13 TeV, recorded from 2015 to 2018 with the ATLAS detector at the Large Hadron Collider, we aim to identify new physics without relying on a specific signal model. The analysis employs two machine learning strategies to estimate the background in different signal regions, with weakly supervised classifiers trained to differentiate this background estimate from actual data. We focus on high transverse momentum jets reconstructed as large-radius jets, using their mass and substructure as classifier inputs. After a classifier-based selection, we analyze the invariant mass distribution of the jet pairs for potential local excesses. Our model-independent results indicate no significant local excesses and we inject a representative set of signal models into the data to evaluate the sensitivity of our methods. This contribution discusses the used methods and latest results and highlights the potential of machine learning in enhancing the search for new physics in fundamental particle interactions.

AI keywords:

weakly supervised, anomaly detection, normalizing flow

190

Autoencoder-based time series anomaly detection for ATLAS Liquid Argon calorimeter data quality monitoring

Author: Vilius Cepaitis[1]

Co-author: Steven Schramm [1]

[1] *University of Geneva*

Corresponding Author: vilius.cepaitis@cern.ch

The ATLAS detector at the LHC has comprehensive data quality monitoring procedures for ensuring high quality physics analysis data. This contribution introduces a long short-term memory (LSTM) autoencoder-based algorithm designed to identify detector anomalies in ATLAS liquid argon calorimeter data. The data is represented as a multidimensional time series, corresponding to statistical moments of energy cluster properties. The model is trained in an unsupervised fashion on good-quality data and is evaluated to detect anomalous intervals of data-taking. The liquid argon noise burst phenomenon is used to validate the approach. The potential of applying such an algorithm to detect arbitrary transient calorimeter detector issues is discussed.

AI keywords:

anomaly detection, data quality monitoring, LSTM, autoencoder, time series

191

Predicting the trainability of deep neural networks with reconstruction entropy

Authors: Johanna Erdmenger[1]; Ro Jefferson[2]; Yanick Thurn[3]

[1] *University of Wuerzburg*

[2] *Utrecht University*

[3] *Deutsch*

Corresponding Author: yanick.thurn@uni-wuerzburg.de

An important challenge in machine learning is to predict the initial conditions under which a given neural network will be trainable. We present a method for predicting the trainable regime in parameter space for deep feedforward neural networks (DNNs) based on reconstructing the input from subsequent activation layers via a cascade of single-layer auxiliary networks. We show that a single epoch of training of the shallow cascade networks is sufficient to predict the trainability of the deep feedforward network on a range of datasets (MNIST, CIFAR10, FashionMNIST, and white noise), thereby providing a significant reduction in overall training time. We achieve this by computing the relative entropy between reconstructed images and the original inputs, and show that this probe of information loss is sensitive to the phase behaviour of the network. We further demonstrate that this method generalizes to residual neural networks (ResNets) and convolutional neural networks (CNNs). Moreover, our method illustrates the network's decision making process by displaying the changes performed on the input data at each layer, which we demonstrate for both a DNN trained on MNIST and the vgg16 CNN trained on the ImageNet dataset. Our results provide a technique for significantly accelerating the training of large neural networks.

AI keywords:

explainable AI, hyperparameter optimization, entropy