

ML4Jets2025

Sunday 17 August 2025 - Saturday 23 August 2025

California Institute of Technology

Book of Abstracts

Contents

DeepSub: Deep Learning for Thermal Background Subtraction in Heavy-Ion Collisions	1
Machine Learning Approaches for Investigating Jet Quenching in Quark-Gluon Plasma via Jet Substructures Analysis	1
1 Particle - 1 Qubit: Particle Physics Data Encoding for Quantum Machine Learning	1
Real-Time event reconstruction for Nuclear Physics Experiments using Artificial Intelli- gence	2
Quantum-Enhanced Inference for Four-Top-Quark Signal Classification at the LHC Using Graph Neural Networks	3
IAFormer: Interaction-Aware Transformer network for collider data analysis	3
Representation Learning of Jets with Physics-Informed Self-Distillations	4
Multi-scale Optimal Transport for Complete Collider Events	4
DLScanner and LeStrat-Net: Machine learning for improved Monte Carlo exploration	5
Anomaly detections in 3 lepton channel using AutoEncoders #35	5
Data-Driven High Dimensional Statistical Inference with Generative Models	5
Machine Learning for Dark Matter searches at the LHC	6
Fast and Precise Track Fitting with Machine Learning	6
Comparative Analysis of FPGA and GPU Performance for Machine Learning-Based Track Reconstruction at LHCb	7
Event-level Observables based on Optimal Transport for Resonant Anomaly Detection	7
MaskFormers for Reconstruction Tasks in High Energy Physics	8
Physics-guided Machine Learning in Cosmology	8
CMS FlashSim: an end-to-end ML approach speeds up simulation in CMS	9
Anomaly Detection Results from CMS	9
Neural autoregressive flows for data-driven background estimation in a search for four-top quark production in the all-hadronic final state with CMS at 13 TeV	9
Particle transformers for boosted $H \rightarrow WW$ identification	10

Autonomous Model Building with Reinforcement Learning: An Application with Lepton Flavor Symmetries	10
A Graph Neural Network Approach for General Reconstruction of Non-Helical Tracks	10
Weakly supervised anomaly detection with event-level variables	11
It's not a FAD: how to use Flows for Anomaly Detection on FPGAs	11
Search for keV-scale Sterile Neutrinos with TRISTAN at KATRIN Using a Neural Network-Based Approach	12
A High-Dimensional, Unbinned Standard Model Measurement with the ATLAS Experiment	12
Higgs Signal Strength Estimation with a Dual-Branch GNN under Systematic Uncertainties	13
DECADE: Selecting the unexpected with decorrelated anomaly triggers	13
Neural autoregressive flows for data-driven background estimation in a search for four-top quark production in the all-hadronic final state with CMS at 13 TeV	14
Simultaneous reconstruction of boosted, resolved, and semi-boosted top-quark events with symmetry-preserving attention networks	14
Observable Optimization for Precision Theory: Machine Learning Energy Correlators	15
Boosting HH(4b) beyond boosted HH(4b): a calibratable full-particle search framework	15
A comparison of self-supervised pre-training methods for foundation models in jet physics	16
Heavy-Flavour Frontier: Tagging at ATLAS with GN3	16
Improving the model agnostic sensitivity of weakly supervised anomaly detection	16
Blooming LHC analyses with all-inclusive pretrained boosted-jet models	17
Testing the Robustness of Via Machinae Stellar Stream Detections Using Resonant Anomaly Detection	17
Anomaly Detection in High-Energy Particle Collisions at the LHC	18
Going HyPER: Enhancing collider measurements with hypergraph learning	18
Debiasing Ultrafast Anomaly Detection with Posterior Agreement	19
Giving machine learning a boost towards respecting (approximate) symmetries	19
Towards a Self-Driving Trigger: Adaptive Response to Anomalies in Real Time	19
Transformer-based tagger for boosted Higgs	20
Fragmentation tagging	20
wifi Ensembles for Simulation-Based Inference with Systematic Uncertainties	21

Anomaly Detection applied to the Quality Control of new detector components	21
The Pareto Frontier of Resilient Jet Tagging	21
Fair Universe HiggsML Uncertainty Challenge: Benchmark for Uncertainty-Aware Machine Learning in High Energy Physics	22
Explicit versus implicit physics priors for separating nearly identical classes	23
Forward folding versus unfolding in the age of ML	23
Unbinned inclusive cross-section measurements with machine-learned systematic uncertainties	24
Stay Positive: Neural Refinement of Simulated Event Weights	24
Discriminative versus Generative Approaches to Simulation-based Inference	24
Incorporating Physical Priors into Weakly-Supervised Anomaly Detection	25
A Novel Anomaly Detection Approach for Primary Vertex Selection at the HL-LHC . . .	25
On focusing statistical power for searches and measurements in particle physics	26
Integrating Energy Flow Networks with Jet Substructure Observables for Enhanced Jet Quenching Studies	26
Deep Learning Methods for Jet Tagging and Process Classification Using Image Processing	27
EveNet: Towards a Generalist Event Transformer for Unified Understanding and Generation of Collider Data	27
Machine Learning-Assisted Measurement of Lepton-Jet Azimuthal Angular Asymmetries and of the complete final state in Deep-Inelastic Scattering at HERA	28
Isolating Unisolated Upsilons with Anomaly Detection in CMS Open Data	28
Machine Learning Neutrino-Nucleus Cross Sections	29
HEP-JEPA: Towards a found model for high energy physics using joint embedding predictive architecture	29
Optimal Transport for e/π^0 Particle Classification in LArTPC Neutrino Experiments . .	29
Convolutional Neural Networks for pile-up suppression in the ATLAS Global Trigger . .	30
Jet tagging with the Lund Jet Plane	30
Efficient Transformers for Jet Tagging	30
A novel loss function to optimise signal significance in particle physics	31
Tackling interpretability with physical baselines for Integrated Gradients	31
Generator Based Inference (GBI)	32
Jet calibration with in-stiu pileu suppression for the L1 trigger	32

Real-Time Compression of CMS Detector Data Using Conditional Autoencoders	32
GELATO: A Generic Event-Level Anomalous Trigger Option for ATLAS	33
Fast Jet Tagging with MLP-Mixers on FPGAs	33
Machine Learning Symmetries in Physics from First Principles	34
Progresses on AI-based jet tagging	34
End-to-end particle reconstruction for current and future colliders	34
AI for gravitational waves	34
Uncertainty quantification in machine learning: A selective overview	35
AI for particle accelerators	35
Foundation models for astrophysics & cosmology	35
Likelihood free inference	35
AI-driven detector design	35
AI-based end-to-end simulation	35
AI at the extreme edge	35
Closing remarks	36
Comparing Continuous and Tokenized Jet Generation Approaches for Precision Modeling	36
Get Your Jets in Shape: Conditioning Heads and Backbones	36
AI in HEP Theory	37
An Energy Correlation Function tagger for gluon-gluon resonances	37

Reconstruction and Analysis / 2

DeepSub: Deep Learning for Thermal Background Subtraction in Heavy-Ion Collisions

Author: Umar Sohail Qureshi¹

Co-author: Rithya Kunawalkam Elayavalli¹

¹ *Vanderbilt University*

Corresponding Author: umar.sohail.qureshi@vanderbilt.edu

Jet reconstruction in an ultra-relativistic heavy-ion collision suffers from a notoriously large thermal background. Traditional background subtraction methods struggle to remove this soft background while preserving the jet's hard substructure. In this talk, we present DeepSub, the first machine learning-based approach for full-event background subtraction.

DeepSub employs Swin Transformer-based layers to denoise jet images and effectively disentangle hard jets from the heavy-ion background. Unlike existing machine learning approaches that predict some low-dimensional quantity, e.g. jet p_T , DeepSub captures the full constituent substructure.

Finally, DeepSub significantly outperforms traditional subtraction techniques on key observables, achieving sub-percent to percent level non-closure on distributions of jet p_T , mass, girth, and the energy correlation functions. As such, DeepSub paves the way for precision QCD measurements in heavy-ion collisions.

Jet Physics / 3

Machine Learning Approaches for Investigating Jet Quenching in Quark-Gluon Plasma via Jet Substructures Analysis

Authors: Leonardo Lima Da Silva¹; Marcelo Gameiro Munhoz¹

¹ *Universidade de Sao Paulo (USP) (BR)*

This project investigates jet quenching phenomena observed in relativistic heavy-ion collisions by applying machine learning techniques to modifications in jet substructures resulting from interactions between jets and the quark-gluon plasma (QGP). A robust dataset is generated using the Jet Evolution With Energy Loss (JEWEL) framework, and the findings are compared with results obtained from the more realistic hydrodynamical medium represented by the v-USPhydro model. In the initial phase, we calculate the jet substructures and study their correlation to choose the most adequate set of observables. Then we explore traditional supervised algorithms, such as Random Forest, to study the performance of a classifier based on multiple decision trees. Subsequently, we employ a neural network known as Long Short-Term Memory (LSTM), which is well-suited for managing sequential data and effectively capturing the dynamic evolution of jets as they traverse the medium. Furthermore, unsupervised techniques, such as k-means, the Gaussian Mixture Model (GMM), and an autoencoder, are implemented to explore the underlying patterns of energy suppression more deeply and to outperform the limitations of supervised techniques. We find that most of the substructures are highly correlated, and that a small set of observables can capture the information and are able to train the models effectively. We then repeat the whole process of training using the Principal Component Analysis (PCA) procedure to extract the most relevant information in the whole dataset, and then compare it with the initial features. Our first analysis indicates that variables related to the angularities are more effective in discriminating the jet samples in both supervised models and that the Random Forest classifier outperforms the neural network. For the unsupervised case, we found that results are very sensitive to the chosen parameters and they do not perform as well as the supervised ones. In general, using PCA variables to train the models yields results slightly better.

Quantum / 4

1 Particle - 1 Qubit: Particle Physics Data Encoding for Quantum Machine Learning

Authors: Aritra Bal¹; Benedikt Maier²; Eric Pezone³; Markus Klute⁴; Melik Oughton³; Michael Spannowsky⁵

¹ *KIT - Karlsruhe Institute of Technology (DE)*

² *Imperial College (GB)*

³ *Imperial College*

⁴ *Karlsruhe Inst. of Technology (GER)*

⁵ *University of Durham (GB)*

We introduce 1P1Q, a novel quantum data encoding scheme for high-energy physics (HEP), where each particle is assigned to an individual qubit, enabling direct representation of collision events on quantum circuits without classical compression. We demonstrate the effectiveness of 1P1Q in quantum machine learning (QML) through two applications: a Quantum Autoencoder (QAE) for unsupervised anomaly detection and a Variational Quantum Circuit (VQC) for supervised classification of top quark jets. Our results show that the QAE successfully distinguishes signal jets from background QCD jets, achieving superior performance compared to a classical autoencoder while utilizing significantly fewer trainable parameters. Similarly, the VQC achieves competitive classification performance, approaching state-of-the-art classical models despite its minimal computational complexity. Furthermore, we validate the QAE on real experimental data from the CMS detector, establishing the robustness of quantum algorithms in practical HEP applications. These results demonstrate that 1P1Q provides an effective and scalable quantum encoding strategy, offering new opportunities for applying quantum computing algorithms in collider data analysis.

Fast ML / 5

Real-Time event reconstruction for Nuclear Physics Experiments using Artificial Intelligence

Author: Gagik Gavalian¹

¹ *Jefferson National Lab*

Corresponding Author: gavalian@jlab.org

Charged track reconstruction is a critical task in nuclear physics experiments, enabling the identification and analysis of particles produced in high-energy collisions. Machine learning (ML) has emerged as a powerful tool for this purpose, addressing the challenges posed by complex detector geometries, high event multiplicities, and noisy data. Traditional methods rely on pattern recognition algorithms like the Kalman filter, but ML techniques, such as neural networks, graph neural networks (GNNs), and recurrent neural networks (RNNs), offer improved accuracy and scalability. By learning from simulated and real detector data, ML models can identify and classify tracks, predict trajectories, and handle ambiguities caused by overlapping or missing hits. Moreover, ML-based approaches can process data in near-real-time, enhancing the efficiency of experiments at large-scale facilities like the Large Hadron Collider (LHC) and Jefferson Lab (JLAB). As detector technologies and computational resources evolve, ML-driven charged track reconstruction continues to push the boundaries of precision and discovery in nuclear physics.

In this talk, we highlight advancements in charged track identification leveraging Artificial Intelligence within the CLAS12 detector, achieving a notable enhancement in experimental statistics compared to traditional methods. Additionally, we showcase real-time event reconstruction capabilities, including the inference of charged particle properties such as momentum, direction, and species identification, at speeds matching data acquisition rates. These innovations enable the extraction of physics observables directly from the experiment in real-time.

Quantum / 6

Quantum-Enhanced Inference for Four-Top-Quark Signal Classification at the LHC Using Graph Neural Networks

Author: Syed Haider Ali¹

Co-authors: Ashfaq Ahmad²; Muhammad Saiel³; Nadeem Shaikat⁴

¹ Department of Physics & Applied Mathematics, Pakistan Institute of Engineering and Applied Sciences (PIEAS), P. O. Nilore 45650, Islamabad

² Exp. High Energy Physics Department (EHEPD), National Centre for Physics (NCP), Shahdra Valley Road, 44000, Islamabad, Pakistan

³ Department of Physics, Quaid-e-Azam university, Shahdra Valley Road, 45320, Islamabad, Pakistan.

⁴ Center for Mathematical Sciences, Pakistan Institute of Engineering and Applied Sciences (PIEAS), P. O. Nilore 45650, Islamabad

Corresponding Author: syedhaider.ali2021453@gmail.com

Rare event classification in high-energy physics (HEP) plays a crucial role in probing physics beyond the Standard Model (BSM). Such processes serve as indirect searches for new physics by testing deviations from SM predictions in extreme kinematic regimes. The production of four top quarks in association with a (W^-) boson at ($\sqrt{s} = 13$) TeV is an exceptionally rare SM process with a next-to-leading-order (NLO) cross-section of $(6.6^{+2.4}_{-2.6} ab)$. In its fully hadronic decay mode, with intricate jet topology and overwhelming QCD background, demands advanced techniques for signal extraction, making it a prime candidate for new physics probes like anomalous top-quark interactions or EFT deviations. Identifying this process in the fully hadronic decay channel is particularly challenging due to overwhelming backgrounds from $t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$, and triple-top production processes. This study introduces **CrossQuantumPhysGNN (CQPGNN)**, a quantum-classical hybrid graph neural network (GNN) designed to tackle rare event classification. CQPGNN integrates GINEConv layers for particle-level features, a quantum circuit employing angle encoding and entanglement for global feature processing, and cross-attention fusion to combine local and quantum-enhanced global representations. Physics-informed losses enforce momentum conservation and jet multiplicity constraints derived from the event decay dynamics, making a faster physics informed convergence. Benchmarked against conventional methods, CQPGNN achieves a signal significance ($S/\sqrt{S+B}$) of $0.174 \pm 0.05\%$, recall of 0.957, and ROC-AUC of 0.961, surpassing BDTs ($0.148 \pm 0.04\%$, 0.914, 0.908) and Xgboost ($0.149 \pm 0.04\%$, 0.924, 0.913). The classification models are trained on parametrized Monte Carlo (MC) simulations of the CMS detector, with events normalized using cross-section-based reweighting to reflect their expected contributions in a dataset corresponding to $350 fb^{-1}$ of integrated luminosity. This ensures that significance calculations accurately reflect realistic collider conditions. The proposed method is benchmarked against conventional machine learning approaches, with results demonstrating improved classification significance. This quantum enhanced approach offers a novel framework for precision event selection at the LHC, leveraging high dimensional statistical learning and quantum-enhanced inference to tackle fundamental HEP challenges, aligning with cutting-edge ML developments.

Jet Physics / 7

IAFormer: Interaction-Aware Transformer network for collider data analysis

Author: Ahmed Hammad¹

¹ KEK, Japan

Corresponding Author: ahammad115566@gmail.com

I will introduce IAFormer, a novel Transformer-based architecture that efficiently integrates pairwise particle interactions through a dynamic

sparse attention mechanism. By leveraging sparsity, IAFormer dynamically prioritizes relevant particle tokens while reducing computational overhead associated with less informative ones. This approach significantly lowers the model complexity without compromising performance. Despite being an order of magnitude more computationally efficient than the Particle Transformer network, IAFormer achieves state-of-the-art performance in classification tasks on the Top and quark-gluon datasets. Our findings highlight the need to sparse attention in any Transformer analysis which reduce the network size while improving its performance.

Jet Physics / 9

Representation Learning of Jets with Physics-Informed Self-Distillations

Authors: Zichun Hao¹; Raghav Kansal²; Abhijith Gandrakota³; Javier Mauricio Duarte⁴; Chang Sun⁵; Jennifer Ngadiuba⁶; Haoyang Li⁴; Zihan Zhao⁷; Maria Spiropulu¹

¹ *California Institute of Technology*

² *Caltech / Fermilab*

³ *Fermi National Accelerator Lab. (US)*

⁴ *Univ. of California San Diego (US)*

⁵ *California Institute of Technology (US)*

⁶ *FNAL*

⁷ *UCSD*

In high-energy physics (HEP) experiments, jets are key concepts in physics analysis. While supervised learning approaches have demonstrated success in tasks such as jet classification and mass regression, they often require large amounts of labeled data and rely on the accuracy of computationally expensive simulations. Self-supervised learning (SSL) has shown promising results for developing robust representations with successes in natural language processing and computer vision, with some recent progress in HEP. Contrastive learning, a popular family of SSL techniques, relies on carefully chosen augmentations to create different views of the same data, but the quality of learned representations depends on the choice of augmentations. However, jets display complex substructures that make standard computer vision augmentations such as cropping and Gaussian smearing non-trivial to apply meaningfully and effectively. We present a physics-informed self-distillation approach for jets that employs a teacher-student model while incorporating hierarchical clustering methods based on physical properties to better capture jet substructure, demonstrating better quality representations than simply smearing or removing particles as augmentations in a few quantitative measures. Our results suggest the possibility of training directly on unlabeled experimental data, potentially reducing the heavy dependence on simulations.

Theory / 10

Multi-scale Optimal Transport for Complete Collider Events

Authors: Katy Craig¹; Lynn Lin^{None}; Nathaniel Craig^{None}; Tianji Cai^{None}

¹ *UCSB*

Corresponding Author: xil244@ucsd.edu

Building upon the success of optimal transport metrics defined for single collinear jets, we develop a multi-scale framework that models entire collider events as distributions on the manifold of their constituent jets, which are themselves distributions on the ground space of the calorimeter. This hierarchical structure of optimal transport effectively captures relevant physics at different scales. We

demonstrate the versatility of our method in two event classification tasks, which respectively emphasize intra-jet substructure and inter-jet spatial correlations. Our results highlight the relevance of a nested structure of manifolds in the treatment of full collider events, broadening the applicability of optimal transport methods in collider analyses.

Event Generation and Detector Simulation / 11

DLScanner and LeStrat-Net: Machine learning for improved Monte Carlo exploration

Author: Raymundo Ramos¹

¹ *Korea Institute for Advanced Study*

In this talk we present two recent proposals to use neural networks to improve Monte Carlo sampling and exploration of simulations that employ CPU expensive calculations. The main idea in the discussed methods is to employ a neural network to distinguish points likely to yield relevant results. This is achieved by training the neural network with previously obtained points that have been labeled according to their importance to the study being performed.

Next is where the two proposals diverge. First we talk about DLScanner, a method to use neural networks to explore high dimensional parameter spaces in theoretical models in search for regions that meet certain criteria. These criteria could be, for example, experimental constraints and theoretical conditions, that can be formulated to output 0 and 1 labels or a continuous value related to likelihood. Then we train a network to predict this output from a large set of points and determine probabilities that these points will be relevant for a full calculation. Second, we present LeStrat-Net, a proposal to use multiple classes to separate the domain of a function according to importance. The domain is divided according values of the function rather than ranges of the variables, thus creating regions of arbitrary shape. A neural network is trained to recognize multiple regions and used to take care of stratification of the domain. We show that this stratification can be used to perform Monte Carlo integration and event generation.

Anomaly Detection / 12

Anomaly detections in 3 lepton channel using AutoEncoders #35

Author: Alfredo Castaneda¹

¹ *Universidad de Sonora (MX)*

The use of autoencoders for anomaly detection has been extended to many fields of science. Their application in high energy physics is particularly relevant, as a trained model can be used to identify experimental failures, data fluctuations, or—most interestingly—signs of new physics phenomena. In this study, we focus on analyzing event topologies with three leptons, aiming to identify potential signal processes. Specifically, we consider a signal W' boson decaying into a WZ pair, resulting in a final state with three leptons and a neutrino ($\ell\ell\ell\nu$), while taking into account Standard Model background processes. The framework for data processing, model training, and preliminary results will be discussed

Unfolding & Inference / 13

Data-Driven High Dimensional Statistical Inference with Generative Models

Authors: Manuel Szwec^{None}; Oz Amram¹

¹ *Fermi National Accelerator Lab. (US)*

Crucial to many measurements at the LHC is the use of correlated multi-dimensional information to distinguish rare processes from large backgrounds.

Since the rise of machine learning in the last decade, it is now standard for analyses to employ multivariate classifiers trained on simulation to distinguish signal and background. Such classifiers significantly increase the statistical power of the analysis, but come with a huge loss of interpretability and render reliable background estimation difficult. Additionally, because they collapse the high-dimensional space into a single observable, fits to classifier scores are sub-optimal for the simultaneous estimation of multiple parameters, a crucial drawback in many present and future LHC measurements.

In this talk we introduce an alternative approach, where instead of dimensionality reduction through classification, a generative ML model is instead trained to learn the signal and background distributions in the high-dimensional space.

The background generative model can be trained directly on data, reducing reliance on simulation as compared to previous ‘simulation based inference’ approaches.

These generative models are then used in place of traditional histograms in a template fit to extract the relevant parameters.

Systematic uncertainties on these models can be parameterized by standard template morphing methods and profiling over bootstrapped ensembles.

We show that this approach can offer comparable or better sensitivity to the classifier-based approach for single parameter fits, while being much more robust and interpretable. This approach also naturally scales to optimally perform multi-parameter inference as well.

Reconstruction and Analysis / 14

Machine Learning for Dark Matter searches at the LHC

Author: Rafal Maselek^{None}

Co-authors: Kazuki Sakurai¹; Mihoko Nojiri²

¹ *King's College London*

² *Theory Center, IPNS, KEK*

Corresponding Author: r.maselek@uw.edu.pl

Dark Matter remains one of the most intriguing mysteries in modern physics. A promising strategy to uncover its nature is through a possible production at the Large Hadron Collider (LHC). One of the key signatures for such searches is the monojet channel, characterised by one or a few high-energy jets recoiling against the large missing transverse momentum and no isolated leptons. However, the analysis of this channel poses significant challenges for traditional methods, particularly due to the need to capture jet substructure and correlations among jet constituents. To address this, we propose a novel search strategy based on Graph Neural Networks (GNNs) to enhance sensitivity to Dark Matter in the monojet channel.

We focus on a simplified supersymmetric scenario where the lightest neutralino serves as the Dark Matter candidate. Our GNN architecture incorporates information at the particle-, jet-, and event-levels to optimise classification performance. We evaluate our method for Wino-, Higgsino-, and Bino-like neutralinos, and derive discovery prospects for Run 3 and the High-Luminosity LHC. Finally, we interpret the trained network to identify which physical features it relies on to distinguish between Standard Model backgrounds and BSM signals.

Reconstruction and Analysis / 15**Fast and Precise Track Fitting with Machine Learning****Authors:** Daniel Whiteson¹; Makayla Vessella ¹¹ *University of California Irvine (US)*

Accurate and efficient particle tracking is a crucial component of precise measurements of the Standard Model and searches for new physics. This task consists of two main computational steps: track finding, the identification of a subset of all hits that are due to a single particle; and track fitting, the extraction of crucial parameters such as direction and momenta. Novel solutions to track finding via machine learning have recently been developed. However, track fitting, which traditionally requires searching for the best global solutions across a parameter volume plagued with local minima, has received comparatively little attention.

Here, we propose a novel machine learning solution to track fitting. The per-track optimization task of traditional fitting is transformed into a single learning task optimized in advance to provide constant-time track fitting via direct parameter regression. This approach allows us to optimize directly for the true targets, precise and unbiased estimates of the track parameters. This is in contrast to traditional fitting, which optimizes a proxy based on the distance between the track and the hits. In addition, our approach removes the requirement of making simplifying assumptions about the nature of the noise model. Most crucially, in the simulated setting described here, it provides more precise parameter estimates at a computational cost 100 times smaller. Potential applications of more accurate, computationally cheaper, track fitting include improved track finding in iterative algorithms, rapid tracking in the trigger setting, improved estimation of particle momenta, more accurate particle ID, vertex finding and jet substructure.

Fast ML / 16**Comparative Analysis of FPGA and GPU Performance for Machine Learning-Based Track Reconstruction at LHCb****Authors:** Bertrand Granado¹; Fotis Giasemis²; Vava Gligorov²; Vladimir Loncar³¹ *sorbonne universite*² *Centre National de la Recherche Scientifique (FR)*³ *CERN*

In high-energy physics, the increasing luminosity and detector granularity at the Large Hadron Collider are driving the need for more efficient data processing solutions. Machine Learning has emerged as a promising tool for reconstructing charged particle tracks, due to its potentially linear computational scaling with detector hits. The recent implementation of a graph neural network-based track reconstruction pipeline in the first level trigger of the LHCb experiment on GPUs serves as a platform for comparative studies between computational architectures in the context of high-energy physics. This paper presents a novel comparison of the throughput of ML model inference between FPGAs and GPUs, focusing on the first step of the track reconstruction pipeline—an implementation of a multilayer perceptron. Using HLS4ML for FPGA deployment, we benchmark its performance against the GPU implementation and demonstrate the potential of FPGAs for high-throughput, low-latency inference without the need for an expertise in FPGA development and while consuming significantly less power.

Anomaly Detection / 17**Event-level Observables based on Optimal Transport for Resonant Anomaly Detection**

Authors: Aditya Bhargava^{None}; Ben Nachman¹; Tianji Cai^{None}

¹ *Lawrence Berkeley National Lab. (US)*

Corresponding Author: aditya_bhargava@berkeley.edu

We introduce a novel class of event-level observables based on Optimal Transport (OT) and demonstrate their efficacy in collider anomaly detection. Under the weakly supervised Classification Without Labels (CWoLa) framework, we evaluate the discriminative power of OT-derived observables on the LHC Olympics dataset, benchmarking their performance against standard high-level features, with both deep neural networks and boosted decision trees as classifiers. Models incorporating OT observables not only significantly outperform the baseline approach but also achieve results comparable to the full phase space method, even at lower levels of signal injection. Our study highlights the utility of OT-based event-level observables for anomaly detection and encourages further explorations of optimal transport applications in collider data analysis.

Reconstruction and Analysis / 18

MaskFormers for Reconstruction Tasks in High Energy Physics

Author: Max Hart¹

¹ *University College London (GB)*

Mask Transformers, or MaskFormers, have emerged as the current state of the art in a wide range of image and point cloud segmentation tasks. We present the application of this architecture to various reconstruction tasks in high energy physics with the aim of tackling both problem scale and complexity. We consider the popular track reconstruction algorithm benchmark dataset TrackML, which represents a problem scale comparable to what will be encountered by the next generation of detectors at the HL-LHC. Combined with an encoder-only transformer to pre-filter hits, a maskformer is able to reconstruct tracks with ~99% efficiency down to a p_T of 600 MeV, while maintaining a low fake rate, and an inference time comparable to other machine learning tracking approaches. To test the model in complex environments, we consider track reconstruction in highly boosted hadronic ROIs in ATLAS simulation data. In such ROIs, increased collimation leads to high amounts of sharing of clusters between tracks, which leads to markedly reduced tracking performance when using traditional approaches. By using a maskformer, we are not only able to mitigate this, increasing efficiency at high p_T by 40%, but also simplify, unify, and improve the process of identifying and regressing the local positions of particles on shared clusters. Finally, we discuss ongoing efforts to use a maskformer to perform global particle flow reconstruction, where tracker, calorimeter, and muon hit assignment with particle regression is performed using one global model, providing a proof-of-concept for single-step global reconstruction.

Uncertainties & Interpretability / 19

Physics-guided Machine Learning in Cosmology

Author: Leonora KARDUM^{None}

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 losing 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.

Event Generation and Detector Simulation / 20

CMS FlashSim: an end-to-end ML approach speeds up simulation in CMS

Author: CMS Collaboration^{None}

Detailed event simulation at the LHC is taking a large fraction of computing budget. CMS developed an end-to-end ML based simulation framework, called FlashSim, that can speed up the time for production of analysis samples of several orders of magnitude with a limited loss of accuracy. We show how this approach achieves a high degree of accuracy, not just on basic kinematics but on the complex and highly correlated physical and tagging variables included in the CMS common analysis-level format, the NANOAOB. We prove that this approach can generalize to processes not seen during training. Furthermore, we discuss and propose solutions to address the simulation of objects coming from multiple physical sources or originating from pileup. Finally, we present a comparison with full simulation samples for some simplified analysis benchmarks. The simulation of particle jets, as well as that of other objects, takes as input relevant generator-level information, e.g. from PYTHIA, while outputs are directly produced in the NANOAOB format. The underlying models being used are state-of-the-art continuous FLOws, trained through Flow Matching.

With this work, we aim to demonstrate that this end-to-end approach to simulation is capable of meeting experimental demands, both in the short term and in view of HL-LHC.

Anomaly Detection / 21

Anomaly Detection Results from CMS

Author: CMS Collaboration^{None}

Anomaly detection has emerged as a new paradigm for physics analyses enabled by machine learning. This talk will overview the latest results from CMS featuring anomaly detection, highlighting the machine learning techniques employed and the achieved physics performance.

Reconstruction and Analysis / 22

Neural autoregressive flows for data-driven background estimation in a search for four-top quark production in the all-hadronic final state with CMS at 13 TeV

Author: CMS Collaboration^{None}

A novel machine-learning based background estimation technique using normalizing flows to estimate background distributions from data in control regions is described in detail. This is demonstrated in four-top quark production in the all hadronic channel in Run II [1], which was facilitated by this method to estimate dominant QCD multijets and $t\bar{t}$ backgrounds. The flow is able to

reliably estimate the distribution of complex variables in signal regions by learning the transformation from input simulated distributions to target data distributions in control regions inspired by the “ABCD” method. This “ABCDnn” method as applied to the four-top all hadronic method is described, including discussion of selections and choice of control regions, validation of the method, estimation of uncertainties, and discussion of closure tests. A public git repository for training the network is also included. This background estimation method is applicable to other hadronic-dominated analyses where QCD multijets simulation is unreliable or statistically limited. This work is expected to be included in an anticipated Run 3 iteration of the all hadronic four-top analysis.

[1] A. Tumasyan et al. [CMS], “Evidence for four-top quark production in proton-proton collisions at $s=13\text{TeV}$,” *Phys. Lett. B* 844, 138076 (2023) doi:10.1016/j.physletb.2023.138076 [arXiv:2303.03864 [hep-ex]].

Jet Physics / 23

Particle transformers for boosted $H \rightarrow WW$ identification

Author: CMS Collaboration^{None}

We present a novel deep neural network classifier, the “particle transformer”, ParT, for identifying highly Lorentz-boosted, multi-pronged jets for measurements and searches with the CMS detector at the LHC. Based on a self-attention architecture, ParT is trained on a wide variety of topologies, notably demonstrating strong performance for the first time on boosted Higgs boson decays to vector bosons. The ParT algorithm achieves a tagging efficiency of $\approx 60\%$ for such jets at a background efficiency of $\approx 1\%$, while maintaining decorrelation from the jet mass. This performance is calibrated on data using the primary Lund jet planes of individual subjects. The impact of ParT is illustrated with the first all-hadronic search for Higgs boson pair production in the two bottom quark and two vector boson channel.

Theory / 25

Autonomous Model Building with Reinforcement Learning: An Application with Lepton Flavor Symmetries

Authors: Aishik Ghosh¹; Daniel Whiteson²; Jake Rudolph¹; Jason Baretz¹; Max Fieg²; Victoria Knapp-Perez¹; Vijay Ganesh³

¹ *UC Irvine*

² *University of California Irvine (US)*

³ *Georgia Tech*

To explain Beyond the Standard Model phenomena, a physicist has many choices to make in regards to new fields, internal symmetries, and charge assignments, collectively creating an enormous space of possible models. We describe the development and findings of an Autonomous Model Builder (AMBer), which uses Reinforcement Learning (RL) to efficiently find models satisfying specified discrete flavor symmetries and particle content. Aside from valiant efforts by theorists following their intuition, these theory spaces are not deeply explored due to the vast number of possibilities and the time-consuming nature of building and fitting a model for a given symmetry group and particle assignment. The lack of any guarantee of continuity or differentiability prevents the application of typical machine learning approaches. We describe an RL software pipeline that interfaces with newly optimized versions of physics software, and apply it to the task of neutrino model building. Our agent learns to find fruitful regions of theory space, uncovering new models in commonly analyzed symmetry groups, and exploring for the first time previously unexamined symmetries.

Reconstruction and Analysis / 26**A Graph Neural Network Approach for General Reconstruction of Non-Helical Tracks****Authors:** Levi Condren¹; Daniel Whiteson¹¹ *University of California Irvine (US)*

Tracking algorithms typically assume helical trajectories to simplify the task of reconstruction. However, numerous theories predict interactions which lead to non-helical tracks. Graph neural networks can split the task of finding and fitting tracks, allowing them to find non-helical tracks from physics beyond the Standard Model, such as quirks. Yet, particles could exhibit behavior beyond what theory has predicted. With only model dependent search strategies, we can only find physics that has already been anticipated by theory. A model-agnostic reconstruction technique would afford us the opportunity to make single event discoveries, free of background, and would not require predictions from theory. We present a method of training the GNN4ITK pipeline to reconstruct a broad general set of non-helical tracks. Our work shows that the pipeline has a high aptitude to operate as a generalized track finder and presents itself as a promising approach for making background free single event discoveries.

Anomaly Detection / 27**Weakly supervised anomaly detection with event-level variables****Author:** Liam Brennan¹**Co-authors:** Louis Moureaux ²; Manuel Sommerhalder ²; Oz Amram ³; Petar Maksimovic ⁴; Sanjana Sekhar ⁴; Tamas Almos Vami ¹; Tianji Cai ; Yuta Takahashi ⁵¹ *Univ. of California Santa Barbara (US)*² *Hamburg University (DE)*³ *Fermi National Accelerator Lab. (US)*⁴ *Johns Hopkins University (US)*⁵ *University of Zurich (CH) / University of Florida (US)*

We introduce a new topology for weakly supervised anomaly detection searches, di-object plus $\tilde{X}$. In this topology, one looks for a resonance decaying to two standard model particles produced in association with other anomalous event activity ($\tilde{X}$). This additional activity is used for classification. We demonstrate how anomaly detection techniques which have been developed for di-jet searches focusing on jet substructure anomalies can be applied to event-level anomaly detection in this topology. To robustly capture event-level features of multi-particle kinematics, we employ new physically motivated variables derived from the geometric structure of a collision's phase space manifold. As a proof of concept, we explore the application of this approach to several benchmark signals in the di- τ plus $\tilde{X}$ final state. We demonstrate that our anomaly detection approach can reach discovery-level significances for signals that would be missed in a conventional bump-hunt approach.

Fast ML / 28**It's not a FAD: how to use Flows for Anomaly Detection on FPGAs****Author:** Francesco Vaselli¹

¹ *Scuola Normale Superiore & INFN Pisa (IT)*

Corresponding Author: francesco.vaselli@cern.ch

We present an FPGA implementation of a Normalizing Flow (NF) for Anomaly Detection (AD) of new physics in realistic high-rate trigger systems of large HEP experiments. To the best of our knowledge, this marks the first operation of such an algorithm on FPGA, demonstrating anomaly detection performance and latency comparable to existing FPGA-based ML solutions.

We train a continuous NF model through the Flow Matching routine on a dataset consisting of realistic low-level features of physics object from a SM signature. Conventionally, NFs map input data to a latent Gaussian space, and the likelihood under this prior serves as an anomaly score. Points deviating from the SM training data are expected to map to low-likelihood regions of the prior.

However, as this procedure is too complex to port efficiently to FPGA, we rely on the intuition that anomalous points will have to be displaced more than training ones to fall outside of the bulk of the prior; thus, we can effectively use the output of the NF forward pass as our anomaly score. We show that this simple approach is still effective on a variety of different beyond-SM signatures, all while requiring minimal resources and running with a latency of less than 100 ns, making it suitable for applications in high-throughput scenarios such as modern trigger systems.

Reconstruction and Analysis / 29

Search for keV-scale Sterile Neutrinos with TRISTAN at KATRIN Using a Neural Network-Based Approach

Author: Luca Fallböhmer¹

Co-authors: Anthony Onillon¹; Susanne Mertens¹

¹ *Max-Planck-Institute for Nuclear Physics*

Following the completion of its neutrino mass measurement program at the end of 2025, the KATRIN experiment aims to probe keV-scale sterile neutrinos by analyzing the full tritium beta decay spectrum with a novel detector system, TRISTAN. Leveraging KATRIN's high source activity, this search is sensitive to mixing amplitudes at the parts-per-million level. However, extracting a potential sterile neutrino signature is challenging, as it relies on detailed modeling of the differential tritium spectrum and requires computationally intensive Monte Carlo simulations. To address this challenge, we explore a neural network-based approach to identify sterile neutrino signatures directly from the spectral data. In this poster, we will present the expected sensitivity of this method and evaluate its robustness against key experimental and modeling uncertainties. In addition, we demonstrate how flow-matching techniques can be leveraged to overcome the limitations of traditional Monte Carlo simulations, enabling more efficient modeling of the experimental spectra.

Unfolding & Inference / 30

A High-Dimensional, Unbinned Standard Model Measurement with the ATLAS Experiment

Author: Mariel Pettee¹

Co-authors: Ben Nachman¹; Dag Gillberg²; Laura Miller³

¹ *Lawrence Berkeley National Lab. (US)*

² *Carleton University*

³ *TRIUMF (CA)*

Traditional approaches to precise Standard Model (SM) measurements of fundamental particles at the LHC generally restrict the format of these results to just one or two properties at a time in predetermined histogram bins. The ATLAS Experiment recently published such a measurement in a notable new format for LHC experiments: high-dimensional and unbinned datasets that can be used for a wide range of scientific applications. This precision measurement of high-momentum Z boson events uses a neural network strategy to reduce detector distortions and therefore facilitate direct comparison with theoretical QCD predictions. Physicists can easily configure the datasets to produce traditional binned measurements of any of the measured properties, or arbitrary combinations of them, with full uncertainty covariances and customized binning.

Unfolding & Inference / 31

Higgs Signal Strength Estimation with a Dual-Branch GNN under Systematic Uncertainties

Authors: Claudius Krause¹; Daohan Wang²; Minxuan He³

¹ *HEPHY Vienna (ÖAW)*

² *HEPHY ÖAW*

³ *University of Chinese Academy of Sciences*

We present a graph-neural-network (GNN) framework that delivers precise extractions of the Higgs-boson signal strength while coherently propagating systematic uncertainties. The architecture combines a deterministic branch, which processes kinematic features immune to nuisance parameters, with an uncertainty-aware branch that ingests systematics-modulated inputs and applies gated-attention message passing; their representations are merged through skip connections and learnable gates to form a single signal-background discriminant. Training proceeds by explicitly scanning the nuisance-parameter space, injecting the corresponding systematic variations and updating the network weights at each scan point, so the classifier learns a smooth, differentiable dependence on all systematics. After training, classifier scores are recomputed for shifted nuisance parameters, the resulting signal and background score distributions are interpolated via template morphing, and surrogate likelihoods are built in which nuisance parameters are profiled out numerically. Large-scale pseudo-experiments demonstrate that the method yields unbiased estimates of signal strength with 68 % confidence intervals that match or surpass the coverage and precision of traditional binned-template techniques, all while retaining the full event-level information.

Fast ML / 32

DECADE: Selecting the unexpected with decorrelated anomaly triggers

Author: Noah Clarke Hall¹

¹ *University College London*

At ATLAS and CMS, the rate of proton collisions far exceeds the rate at which data can be recorded. A real-time event selection process, or trigger, is needed to ensure that the data recorded contains the highest possible discovery potential. In the absence of hoped-for anomalies such as SUSY, there is increasing motivation to develop dedicated anomaly detection triggers. A common approach is to use unsupervised machine learning to predict an event-by-event anomaly score based on the 4-momenta and multiplicity of reconstructed objects. We show that such anomaly scores often exhibit high mutual information with existing trigger variables, duplicating the acceptance of current triggers rather than accessing underexplored regions of phase space. We introduce DECORRELATED Anomaly DETECTION (DECADE), in which the resulting anomaly score is decorrelated from existing trigger variables. By minimising the mutual information between the anomaly score and the primary triggers, DECADE prioritises acceptance in regions of phase space not captured by existing

trigger strategies. We benchmark two approaches to decorrelation, each suited to deployment in hardware (FPGA) and software (CPU/GPU), and compare physics performance.

33

Neural autoregressive flows for data-driven background estimation in a search for four-top quark production in the all-hadronic final state with CMS at 13 TeV

Authors: Javier Mauricio Duarte¹; Jet Zhang Yue¹; Melissa Quinnan¹; Su Yong Choi²

¹ Univ. of California San Diego (US)

² Korea University (KR)

Corresponding Author: melissa.quinnan@cern.ch

A novel machine-learning-based background estimation technique using normalizing flows to estimate background distributions from data in control regions is described in detail. This is demonstrated in four-top quark production in the all hadronic channel in Run II [1], which was facilitated by this method to estimate dominant QCD multijets and $t\bar{t}b\bar{b}$ backgrounds. The flow is able to reliably estimate the distribution of complex variables in signal regions by learning the transformation from input simulated distributions to target data distributions in control regions inspired by the “ABCD” method. This “ABCDnn” method as applied to the four-top all hadronic method is described, including discussion of selections and choice of control regions, validation of the method, estimation of uncertainties, and discussion of closure tests. A public git repository for training the network is also included. This background estimation method is applicable to other hadronic-dominated analyses where QCD multijets simulation is unreliable or statistically limited. This work is expected to be included in an anticipated Run 3 iteration of the all-hadronic four-top analysis.

[1] A. Tumasyan et al. [CMS], “Evidence for four-top quark production in proton-proton collisions at $s=13\text{TeV}$,” Phys. Lett. B 844, 138076 (2023) doi:10.1016/j.physletb.2023.138076 [arXiv:2303.03864 [hep-ex]].

Reconstruction and Analysis / 34

Simultaneous reconstruction of boosted, resolved, and semi-boosted top-quark events with symmetry-preserving attention networks

Authors: Daniel Primosch¹; Haoyang Li¹; Javier Mauricio Duarte¹; Maria Spiropulu²; Melissa Quinnan¹; Thomas Coulter Sievert²

¹ Univ. of California San Diego (US)

² California Institute of Technology (US)

Corresponding Author: tsievert@caltech.edu

The production of multiple top quarks at the CERN LHC provides both a rich environment to probe the Standard Model for signs of new physics and also serves as a major background in many searches. A substantial fraction of these events result in fully hadronic final states, where each top quark decays into a bottom quark and a W boson, with the latter further decaying into two light quarks. In the case of multiple hadronic tops this yields a multi-jet final state, posing a combinatorial challenge known as the jet assignment problem: assigning reconstructed jets to top quark candidates. Symmetry-preserving attention networks (SPA-Nets) have been developed to tackle such problems by leveraging permutation-invariant representations of jet sets. However, the complexity increases

when accounting for “resolved” topologies—where all decay products are reconstructed as separate small-radius jets—, “boosted” topologies—where high-momentum top quarks are reconstructed as large-radius jets with substructure, and “semi-boosted” topologies—where some decay products are merged into large-radius jets while others are resolved. Traditional top tagging approaches typically target either boosted or resolved topologies exclusively, or select top candidates preferentially for these topologies, leading to reduced efficiency in detecting hadronic tops. In this work, we extend the SPA-Net framework to simultaneously consider resolved, boosted and semi-boosted topologies, enabling unambiguous classification of hadronic tops as “fully resolved”, “fully boosted”, or “semi-boosted”. Our method significantly improves the top quark reconstruction purity and assignment efficiency compared to baseline techniques, depending on the event topology, and thus has great potential for improving the sensitivity of searches involving multiple hadronic tops at the LHC. A full code repository containing a general library, the specific configuration used, and a complete dataset release are included.

Theory / 35

Observable Optimization for Precision Theory: Machine Learning Energy Correlators

Author: Katherine Fraser¹

Co-authors: Arindam Bhattacharya ; Matthew Schwartz

¹ *Harvard University*

Corresponding Author: kfraser@berkeley.edu

The practice of collider physics typically involves the marginalization of multi-dimensional collider data to one-dimensional observables. In many cases, the observable can be arbitrarily complicated, such as the output of a neural network. However, for precision measurements, the observable must correspond to something computable. In this work, we demonstrate that precision-theory-compatible observable space exploration can be systematized by using neural simulation-based inference techniques from machine learning. We illustrate this approach by exploring the space of marginalizations of the energy 3-point correlation function to optimize sensitivity to the top quark mass.

Reconstruction and Analysis / 36

Boosting HH(4b) beyond boosted HH(4b): a calibratable full-particle search framework

Authors: Congqiao Li¹; Tianyi Yang¹

¹ *Peking University (CN)*

We demonstrate that the successful techniques developed for boosted HH(4b) analyses can be effectively extended to the resolved regime through advanced deep learning engineering. By leveraging O(100M) training samples, employing efficient state-of-the-art architectures and training frameworks, and analyzing objects containing O(100) particles, we can replicate the capabilities of Xbb taggers from the boosted regime across a broader phase space, significantly enhancing $\kappa\lambda$ measurements. To achieve this, we present a comprehensive calibratable experimental strategy. Our approach involves training a universal classifier to distinguish $X \rightarrow Y_1 Y_2 \rightarrow \text{bbbb}$ signals from QCD and $t\bar{t}$ multijet backgrounds across a wide range of X and $Y_{1,2}$ mass values, while simultaneously estimating the $Y_{1,2}$ masses via a multiclass classification technique. This discriminant is first calibrated using “fake ZZ(4b) events” generated through an event hemisphere mixing technique from a distinct di-muon triggered phase space, then validated in a search for genuine ZZ(4b) events that is capable of reaching observation ($>5\sigma$) sensitivity with Run2+3 datasets. We demonstrate that the

HH(4b) sensitivity of this method is comparable to the HL-LHC projection, holding great promise to accelerate the pace of HH searches at the LHC.

Jet Physics / 37

A comparison of self-supervised pre-training methods for foundation models in jet physics

Authors: Anna Hallin¹; Gregor Kasieczka²; Joschka Birk²

¹ *University of Hamburg*

² *Hamburg University (DE)*

Corresponding Author: joschka.birk@uni-hamburg.de

Over the last few years, different pre-training strategies for foundation models in HEP have been proposed. Some of them, like generative pre-training (used in OmniJet- α) and Masked Particle Modeling (MPM), rely on self-supervised pre-training, allowing models to be pre-trained on unlabelled data collected by experiments.

We present studies that compare those two self-supervised methods with straightforward supervised pre-training for jet tagging. Furthermore, we investigate how combinations of those different methods affect the quality of the pre-trained representations.

While the original OmniJet- α model used exclusively tokenized input, leading to a loss of information that could decrease the performance on downstream tasks such as jet tagging, we here use a hybrid setup with continuous input features for all tasks and tokenized particle representations as next-token-prediction target (OmniJet- α) or masked-token-prediction target (MPM), respectively.

Jet Physics / 38

Heavy-Flavour Frontier: Tagging at ATLAS with GN3

Authors: Alexander Froch¹; Dan Guest²; Nicole Michelle Hartman³; Valentina Vecchio⁴

¹ *Universite de Geneve (CH)*

² *Humboldt University of Berlin*

³ *TUM (DE)*

⁴ *University of Manchester*

Accurate identification of jets that originate from heavy-flavour hadrons is pivotal for many ATLAS analyses, from Higgs-boson and top-quark measurements to searches for new physics. We present GN3, the newest heavy-flavour tagger, which introduces a full-transformer architecture tailored to the environment of Run 2 and Run 3.

GN3 processes low-level track, vertex, neutral particle, and muon information to extract correlations between the inputs and infer the origin of the jet. Compared with the current Run 3 baseline, GN2, the new model achieves better separation of b - and c -jets from light-flavour jets across a wide kinematic phase-space.

In this talk, we will discuss the architecture and training workflow, as well as the newest results from GN3. Furthermore, we will highlight the new capabilities added to GN3 which complement the excellent tagging performance.

Anomaly Detection / 39

Improving the model agnostic sensitivity of weakly supervised anomaly detection

Authors: Alexander Mück¹; David Shih^{None}; Gregor Kasieczka²; Louis Moureaux²; Lukas Lang^{None}; Marie Hein¹; Michael Kramer^{None}; Radha Mastandrea³; Ranit Das⁴; Tobias Quadfasel²

¹ RWTH Aachen University

² Hamburg University (DE)

³ University of California, Berkeley

⁴ Rutgers University

Corresponding Author: marie.hein@rwth-aachen.de

Weakly supervised anomaly detection can detect new physics at lower cross sections and improve limits without placing many constraints on signal models. For optimal sensitivity across different BSM scenarios, it's important to choose suitable classification architectures and feature sets that offer sensitivity to a wide range of signals. In this study, we explore how to set up such analyses effectively. To properly evaluate how model agnostic our findings are, we introduce a new dataset containing several signal models based on signals from a recent CMS analysis, which can be used alongside the LHC Olympics R&D dataset background.

Jet Physics / 40

Blooming LHC analyses with all-inclusive pretrained boosted-jet models

Authors: Antonis Agapitos¹; Congqiao Li¹; Cristina Ana Mantilla Suarez²; Dawei Fu¹; Gregor Kasieczka³; Huilin Qu⁴; Javier Mauricio Duarte⁵; Leyun Gao¹; Louis Moureaux³; Qiang Li¹; Raghav Kansal⁶

¹ Peking University (CN)

² University of Virginia (US)

³ Hamburg University (DE)

⁴ CERN

⁵ Univ. of California San Diego (US)

⁶ Caltech / Fermilab

Corresponding Author: congqiao.li@cern.ch

I will present recent advances in the development of inclusive, large-scale pretrained models for Lorentz-boosted jets at the LHC's general-purpose experiments. These models significantly enhance the LHC physics program by (1) extending the sensitivity reach of model-specific analyses, and (2) substantially improving model-agnostic strategies, thereby unlocking previously unreachable physics potential. I will focus on the *Sophon* model, trained on realistic Delphes-simulated datasets as a representative benchmark, and extend to the concept of the *Global Particle Transformer (GloParT)* models developed within real experimental contexts. I will also provide insights into the underlying deep learning methodologies and discuss future directions in this rapidly evolving field.

Anomaly Detection / 41

Testing the Robustness of Via Machinae Stellar Stream Detections Using Resonant Anomaly Detection

Authors: David Shih^{None}; Rafael Porto^{None}

Corresponding Author: rafael.porto@rutgers.edu

We build upon the results of the Via Machinae stream-finding algorithm, which uses the ANODE method for resonant anomaly detection to search for stellar streams in *Gaia* data, by employing new tests to identify the stream candidates most likely to represent real stellar streams. We measure the consistency with which candidates are discovered across multiple retrainings of the ANODE neural density estimators and find that classifying candidates based on this metric reduces the expected rate of false positive discoveries by a factor of roughly 2 while increasing the number of stream candidates classified as real by more than 20%. As an independent test, we apply an automated orbit-fitting algorithm to determine whether each candidate lies along a physical orbit integrated in a model of the Milky Way gravitational potential. We present a list of candidates that pass both these tests and merit follow-up observations, some of which are to our knowledge previously unknown.

Anomaly Detection / 42

Anomaly Detection in High-Energy Particle Collisions at the LHC

Authors: Ben Nachman¹; Dennis Daniel Nick Noll²; Runze Li³

¹ *Lawrence Berkeley National Lab. (US)*

² *Lawrence Berkeley National Lab (US)*

³ *Yale University (US)*

Corresponding Author: runze.li@yale.edu

This contribution presents a novel approach to model-independent anomaly detection in LHC collisions, targeting physics beyond the Standard Model. Leveraging machine learning techniques, we identify potential signals without relying on specific signal models. The analysis employs a machine learning driven background estimation in different signal regions, with weakly supervised classifiers trained to differentiate the background estimate from actual data. We analyze the invariant mass spectrum of the selected events for potential local excesses, which could indicate the presence of new physics. The method and its latest results are discussed, highlighting the potential of machine learning in enhancing the search for new physics.

Reconstruction and Analysis / 43

Going HyPER: Enhancing collider measurements with hypergraph learning

Author: Ethan Lewis Simpson¹

¹ *The University of Manchester (GB)*

Hypergraph learning extends traditional graph learning techniques by exploring higher-order correlations on graphs, leading to powerful and expressive representations of collider events. The HyPER model employs hypergraph learning to tackle the reconstruction of short-lived particles, and the separation of signal events from backgrounds. HyPER has been tested on top quark kinematic reconstruction, where it demonstrates improved performance over existing ML techniques while using far fewer learnable parameters. The model is applicable to arbitrary topologies: we showcase new HyPER results in rare top quark processes, including interfacing HyPER with neutrino reconstruction techniques. The impact of training using jet constituent information is also studied, and by augmenting the learning task, HyPER can leverage reconstruction information to classify signal and background events.

Anomaly Detection / 44

Debiasing Ultrafast Anomaly Detection with Posterior Agreement

Authors: Denis-Patrick Odagiu¹; João Sá Carvalho²; Thea Aarrestad¹; Victor Jimenez²

¹ *ETH Zurich (CH)*

² *ETH Zurich*

The Level-1 Trigger system of the CMS experiment at CERN makes the final decision on which LHC collision data are stored to disk for later analysis. One algorithm used with this scope is an anomaly detection model based on an autoencoder architecture. This model is trained self-supervised on measured data, but its performance is typically evaluated on simulated datasets of potential anomalies. Since the true nature of anomalies in the real collision data is unknown, such a validation strategy inherently biases the model towards the characteristics of the simulated cases. We propose an alternative validation criterion: maximizing the mutual information between latent spaces produced by models that are obtained using different data sources. Thus, we explicitly quantify the bias introduced through the current model selection procedure by computing the mutual information between latent spaces of autoencoders in a cross-validation setting with different subsets of 25 simulated potential anomaly datasets. Additionally, we investigate how our metric can be used as a model selection criterion at training time, circumventing the reliance on simulated anomaly datasets. Therefore, using our method not only exposes the existing validation bias in the current level-1 anomaly detector, but also yields new models whose anomaly definitions are both robust and broadly informative, ensuring that the trigger remains sensitive to genuinely unexpected novel physics in LHC collisions.

Theory / 45

Giving machine learning a boost towards respecting (approximate) symmetries

Authors: Pradyun Hebbar^{None}; Thandikire Madula¹; Vinicius Massami Mikuni²; Ben Nachman²; Nadav Outmeuguine^{None}; Inbar Savoray³

¹ *University of London (GB)*

² *Lawrence Berkeley National Lab. (US)*

³ *UC Berkeley*

In the physical sciences, symmetries provide powerful inductive biases from theoretical insights. Incorporating these constraints into the training of machine learning models is expected to improve robustness and lead to more data-efficient models that are easier to interpret. However, fully equivariant models can be difficult to train and implement. Moreover, real-world experiments often result in broken symmetries due to imperfections and finite detector resolution. In this work, we explore an alternative to create symmetry-aware machine learning models through soft constraints. We investigate two complementary approaches, one that penalizes the model based on sampled transformations of the inputs and one inspired by group theory and tracks infinitesimal variations. We implement those ideas for the case of Lorentz invariance, which is of particular importance in particle physics. We find that the addition of the soft constraints may improve performance while requiring negligible changes to current state-of-the-art models.

Fast ML / 46

Towards a Self-Driving Trigger: Adaptive Response to Anomalies in Real Time

Authors: Abhijith Gandrakota¹; Cecilia Tosciri²; Christian Herwig³; David Miller²; Giovanna Salvi³; Jennifer Ngadiuba¹; Nhan Tran¹; Shaghayegh Emami³; Yuxin Chen²; Zixin Ding²

¹ *FNAL*

² *University of Chicago*

³ *University of Michigan*

Machine learning has opened new possibilities for detecting anomalous signatures in high-energy physics data. While most approaches have focused on offline use, there is growing interest in applying these tools directly at the trigger level to enhance discovery potential. In this work, we present a novel framework for autonomous triggering that not only detects anomalous patterns in real time but also determines how to respond to them. We develop and benchmark a self-driving trigger system that integrates anomaly detection with real-time control strategies, dynamically adjusting trigger thresholds and resource allocations in response to changing beam conditions. Using CMS Open Data and a realistic benchmarking setup, our system employs feedback-based control and resource-aware optimization that accounts for trigger bandwidth and compute constraints to maintain stable trigger rates while enhancing sensitivity to rare or unexpected signals. This approach represents a step toward adaptive, intelligent trigger systems for high-throughput experimental environments.

Jet Physics / 47

Transformer-based tagger for boosted Higgs

Author: Edoardo Critelli¹

¹ *UCL (GB)*

Flavour tagging, the identification of jets originating from b- and c-quarks, is a critical component of the physics programme of the ATLAS experiment at the Large Hadron Collider. In recent years, ATLAS introduced new machine learning algorithms based on the transformer architecture, which use information from charged particle tracks within a jet to predict the jet flavour without the need for intermediate low-level algorithms.

This is particularly relevant for hadronic boosted decays, such as high pT Higgs Bosons produced at the LHC. For these, dedicated taggers for the $H \rightarrow b\bar{b}$ decay topology are developed, based on the same transformer architecture but optimized and trained for highly boosted particle decays captured in a single large-radius jet.

In this talk, we present recent developments in such Transformer-based tagging algorithms for boosted Higgs. These include the incorporation of Particle Flow inputs, the introduction of new signal classes (e.g., $H \rightarrow \tau\tau$), additional background subclasses such as granular QCD decays, and general architectural advancements. We discuss the performance gains achieved with these upgrades and highlight the broader implications for flavour tagging in the boosted regime.

Jet Physics / 48

Fragmentation tagging

Author: Yevgeny Kats¹

¹ *Ben-Gurion University*

Corresponding Author: yevgeny.kats@cern.ch

Many ML tools tackle the problem of jet tagging (or flavor tagging), namely determining what particle type gave rise to the jet. We point out that another task to which ML tools can be applied is “fragmentation tagging”, which is a question about the hadronization process that occurred in

a given jet. For example, one may ask whether a given b-jet contained a b-meson or a b-baryon. This can in principle be useful for inclusive measurements of fragmentation functions (or at least fragmentation fractions), especially at high p_T where reliance on clean decay modes is statistically limited. Fragmentation tagging can also help reduce background for certain analyses. For example, in b-quark polarization and spin correlation studies using semileptonic Λ_b decays, it can help suppress the large background from semileptonic B -meson decays.

Unfolding & Inference / 49

wifi Ensembles for Simulation-Based Inference with Systematic Uncertainties

Author: Sean Benevedes¹

¹ *Massachusetts Institute of Technology*

Corresponding Author: seanbenevedes@gmail.com

Neural ratio estimation provides a means of performing frequentist simulation-based inference (SBI), but uncertainties on the estimated ratios of probability densities must be taken into account in order to yield reliable confidence intervals on the inferred parameters. We examine the role of these uncertainties on estimated density ratios in the context of the FAIR Universe HiggsML Uncertainty Challenge dataset. We use $w_i f_i$ ensembles, our recently proposed framework for obtaining frequentist uncertainties on estimated density ratios using ensembles of neural networks, to perform SBI in the presence of parameterized systematic uncertainties. The uncertainties on the inferred density ratios contribute to the overall uncertainty budget of the downstream SBI, and larger training datasets reduce this contribution at the expense of additional required computational resources. We examine the size of this contribution as a function of training dataset size to quantify this tradeoff.

Anomaly Detection / 50

Anomaly Detection applied to the Quality Control of new detector components

Author: Louis Vaslin¹

Co-authors: Daniela Bortoletto ²; Yu Nakahama ³

¹ *KEK High Energy Accelerator Research Organization (JP)*

² *University of Oxford (GB)*

³ *High Energy Accelerator Research Organization (KEK)*

Corresponding Author: lovaslin@post.kek.jp

In High Energy Physics (HEP), new discoveries depend on the development and deployment of new experimental setups using cutting edge detector technologies. Ensuring the quality of these new detectors is required to ensure the success of such experiments. We propose a tool based on Computer Vision algorithms to improve the reliability and efficiency of the Visual Inspection of new detector components, which is a major part of the Quality Control procedures. The tool is being developed in the context of the production of the new Inner Tracker (ITk) detector that will be installed in the ATLAS experiment for the High Luminosity phase of the LHC. We will present the models we are developing, as well as the first defect detection and software integration results. We also want to open discussions on possible extensions and applications of our tool, including beyond HEP experiments.

Jet Physics / 51

The Pareto Frontier of Resilient Jet Tagging

Author: Rikab Gambhir¹

Co-authors: Matt LeBlanc²; Yuanchen Zhou²

¹ MIT

² Brown University (US)

Corresponding Author: rikab@mit.edu

Jet tagging using information extracted from the kinematics of particles inside jets is a common task in high-energy collider physics. Often, event classifiers are designed by targeting the best performance in terms of accuracy, AUC, or similar metrics, and many classifiers have been developed that score high on these metrics by training on simulations such as PYTHIA. However, optimizing these metrics can come at the expense of physics and systematics that are important for downstream tasks, such as simulation dependence, which leads to increased uncertainties in physics analyses. We show the Pareto Front of performance versus model-dependence for jet classification tasks, illustrating how “simpler” models with lower accuracy also have less modeling dependence, and more sophisticated ML models have much greater modeling dependence. We also explore methods to improve the performance of classifiers while minimizing their model dependence.

Uncertainties & Interpretability / 52

Fair Universe HiggsML Uncertainty Challenge: Benchmark for Uncertainty-Aware Machine Learning in High Energy Physics

Author: Po-Wen Chang^{None}

Co-authors: Aishik Ghosh¹; Alexandre Michon²; Benjamin Nachman³; Benjamin Sluijter⁴; Benjamin Thorne⁵; Chris Harris⁵; David Rousseau⁶; Elham Khoda⁷; Ihsan Ullah⁸; Isabelle Guyon⁸; Jordan Dudley⁹; Mathis Reymond¹⁰; Paolo Calafiura¹¹; Peter Nugent⁵; Ragansu Chakkappai⁶; Rémy Lyscar²; Sascha Diefenbacher¹¹; Shih-Chieh Hsu¹²; Steven Farrell⁵; Wahid Bhimji; Yuan-Tang Chou⁷; Yulei Zhang⁷

¹ University of California Irvine (US)

² Université Paris-Saclay, CNRS/IN2P3, IJCLab

³ Stanford University

⁴ Universiteit Leiden

⁵ Lawrence Berkeley National Laboratory

⁶ IJCLab-Orsay

⁷ University of Washington (US)

⁸ ChaLearn

⁹ University of California, Berkeley

¹⁰ Université Paris-Saclay

¹¹ Lawrence Berkeley National Lab. (US)

¹² University of Washington Seattle (US)

Measurements and observations in particle physics fundamentally depend on one’s ability to quantify their uncertainty and, thereby, their significance. Therefore, as machine learning (ML) methods become more prevalent in high energy physics, being able to determine the uncertainties of an ML method becomes more important. A wide range of possible approaches has been proposed, however, there has not been a comprehensive comparison of individual methods. To address this, the Fair Universe project organized the HiggsML Uncertainty Challenge, which took place from September 2024 to March 2025, and the dataset and performance metrics of the challenge will serve as a

permanent benchmark for further developments. Additionally, the Challenge was accepted as an official NeurIPS2024 competition. The goal of the challenge was to measure the Higgs to $\tau^+\tau^-$ signal strength, using a dataset of simulated pp collision events observed in LHC. Participants were evaluated on both their ability to precisely determine the correct signal strength, as well as on their ability to report correct and well-calibrated uncertainty intervals. In this talk, we present an overview of the competition itself and of the infrastructure that underpins it. Further, we present the winners of the competition and discuss their winning uncertainty quantification approaches.

The HiggsML Uncertainty Challenge itself can be found under <https://www.codabench.org/competitions/2977/>. And more details are available as <https://arxiv.org/abs/2410.02867>

Theory / 53

Explicit versus implicit physics priors for separating nearly identical classes

Author: Tanvi Wamorkar¹

Co-authors: Ben Nachman¹; Kevin Thomas Greif²; Tilman Plehn; Víctor Bresó Pla³

¹ *Lawrence Berkeley National Lab. (US)*

² *University of California Irvine (US)*

³ *University of Heidelberg*

Machine learning in high energy physics has been accelerated due to two key developments: equivariant models which encode prior knowledge about the symmetries present in high energy physics datasets and models that are pretrained to perform similar tasks on large datasets, encoding useful domain knowledge. In this work, we explore the fundamental tradeoff between explicitly incorporating physics constraints through network architecture versus implicitly encoding knowledge through pretraining when distinguishing nearly identical classes. We focus on two challenging prototypical scenarios: likelihood-ratio estimation via unfolding (data versus simulation) and weakly-supervised anomaly detection (data versus reference sample), and study the impact of both developments.

Unfolding & Inference / 54

Forward folding versus unfolding in the age of ML

Authors: Aishik Ghosh¹; Daniel Whiteson¹; Kevin Thomas Greif¹

¹ *University of California Irvine (US)*

Corresponding Author: kgreif@uci.edu

In most measurements in particle physics, correcting for the imperfect resolution of the detector used to observe the events is a necessary step to extract a parameter of interest. This can be done through forward folding, in which the theoretical predictions are adjusted by running a simulation of the detector, or through unfolding, in which detector effects are removed from the experimental data. Recently, machine learning methods have allowed either approach to make use of high dimensional correlations in the data to better constrain the parameter of interest, possibly altering conventional wisdom about which approach should be used under which circumstances. In this talk, machine learning based methods for both approaches are applied to a toy dataset, so that their performance can be compared to each other and to traditional methods. Conclusions from this case study will inform decisions on which approach to utilize in future machine learning driven measurements by particle physics experiments at the LHC.

Uncertainties & Interpretability / 55

Unbinned inclusive cross-section measurements with machine-learned systematic uncertainties

Authors: Ang Li¹; Claudius Krause²; Cristina Giordano¹; Daohan Wang³; Dennis Schwarz¹; Lisa Benato¹; Maryam Shooshtari¹; Robert Schoofbeck¹

¹ *Austrian Academy of Sciences (AT)*

² *HEPHY Vienna (ÖAW)*

³ *HEPHY ÖAW*

Corresponding Author: claudius.krause@oeaw.ac.at

We introduce a novel methodology for addressing systematic uncertainties in unbinned inclusive cross-section measurements and related collider-based inference problems. Our approach incorporates known analytic dependencies on parameters of interest, including signal strengths and nuisance parameters. When these dependencies are unknown, as is frequently the case for systematic uncertainties, dedicated neural network parametrizations provide an approximation that is trained on simulated data. The resulting machine-learned surrogate captures the complete parameter dependence of the likelihood ratio, providing a near-optimal test statistic. As a case study, we perform a first-principles inclusive cross-section measurement of $H \rightarrow \tau \tau$ in the single-lepton channel, utilizing simulated data from the FAIR Universe Higgs Uncertainty Challenge. Results in Asimov data, from large-scale toy studies, and using the Fisher information demonstrate significant improvements over traditional binned methods. Our computer code “Guaranteed Optimal Log-Likelihood-based Unbinned Method”(GOLLUM) for machine-learning and inference is publicly available.

Our submission won first place ex aequo in the FAIR Universe Higgs Uncertainty Challenge and is available at <https://arxiv.org/abs/2505.05544>.

Event Generation and Detector Simulation / 56

Stay Positive: Neural Refinement of Simulated Event Weights

Authors: Benjamin Nachman¹; Dennis Daniel Nick Noll²

¹ *Stanford University*

² *Lawrence Berkeley National Lab (US)*

Corresponding Author: dennis.noll@cern.ch

Monte Carlo simulations are an essential tool for data analysis in particle physics. Simulated events are typically produced alongside weights, that redistribute the production rate of a simulated process across the phase space. The presence of latent degrees of freedom can lead to a distribution of weights with negative values, often complicating analyses, especially if they involve machine learning methods. Traditional post-hoc reweighting methods aim to approximate the average weight as a function of phase space. In contrast, we propose a novel approach that refines the initial weights to eliminate negative values through a scaling transformation, utilizing a phase space dependent factor. Our method uses neural networks to process high-dimensional and unbinned phase spaces. We will show that our neural weight refinement method achieves comparable or superior accuracy to existing reweighting schemes, and demonstrate its behavior on realistic and synthetic examples.

Unfolding & Inference / 57

Discriminative versus Generative Approaches to Simulation-based Inference

Authors: Benjamin Nachman¹; Benjamin Sluijter^{None}; Sascha Diefenbacher²; Wahid Bhimji^{None}

¹ *Stanford University*

² *Lawrence Berkeley National Lab. (US)*

Most of the fundamental, emergent, and phenomenological parameters of particle and nuclear physics are determined through parametric template fits. Simulations are used to populate histograms which are then matched to data. This approach is inherently lossy, since histograms are binned and low-dimensional. Deep learning has enabled unbinned and high-dimensional parameter estimation through neural likelihood(-ratio) estimation. We compare two approaches for neural simulation-based inference (NSBI): one based on discriminative learning (classification) and one based on generative modeling. These two approaches are directly evaluated on the same datasets, with a similar level of hyperparameter optimization in both cases. In addition to a Gaussian dataset, we study NSBI using a Higgs boson dataset from the FAIR Universe Challenge. We find that both the direct likelihood and likelihood ratio estimation are able to effectively extract parameters with reasonable uncertainties. For the numerical examples and within the set of hyperparameters studied, we found that the likelihood ratio method is more accurate and/or precise. Both methods have a significant spread from the network training and would require ensembling or other mitigation strategies in practice.

Anomaly Detection / 58

Incorporating Physical Priors into Weakly-Supervised Anomaly Detection

Authors: Benjamin Nachman¹; Chi Lung Cheng^{None}; Gurpreet Singh²

¹ *Stanford*

² *University of California, Berkeley*

We propose a new machine-learning-based anomaly detection strategy for comparing data with a background-only reference (a form of weak supervision). The sensitivity of previous strategies degrades significantly when the signal is too rare or there are many unhelpful features. Our Prior-Assisted Weak Supervision (PAWS) method incorporates information from a class of signal models to significantly enhance the search sensitivity of weakly supervised approaches. As long as the true signal is in the pre-specified class, PAWS matches the sensitivity of a dedicated, fully supervised method without specifying the exact parameters ahead of time. On the benchmark LHC Olympics anomaly detection dataset, our mix of semi-supervised and weakly supervised learning is able to extend the sensitivity over previous methods by a factor of 10 in cross section. Furthermore, if we add irrelevant (noise) dimensions to the inputs, classical methods degrade by another factor of 10 in cross section while PAWS remains insensitive to noise. This new approach could be applied in a number of scenarios and pushes the frontier of sensitivity between completely model-agnostic approaches and fully model-specific searches.

Anomaly Detection / 59

A Novel Anomaly Detection Approach for Primary Vertex Selection at the HL-LHC

Authors: Ariel Gustavo Schwartzman¹; Sanha Cheong¹; Wasikul Islam²

¹ *SLAC National Accelerator Laboratory (US)*

² *University of Wisconsin-Madison (US)*

Corresponding Author: wasikul.islam@cern.ch

Vertex selection plays a crucial role in the identification of the hard-scatter primary vertex in high-energy collisions at the Large Hadron Collider (LHC). The high pileup environment at the HL-LHC presents significant challenges, particularly in accurately selecting the hard-scatter vertex. This study investigates an anomaly detection approach, specifically an autoencoder model, for primary vertex selection based on pileup vertex properties, independent of physics signal dependence. The model is trained on pileup vertices directly from data, providing a significant advantage over traditional methods reliant on Monte Carlo (MC) simulations. We present preliminary findings of selecting Primary vertex at the ATLAS experiment, serving as a proof-of-concept, and discuss results with multiple final states.

Unfolding & Inference / 60

On focusing statistical power for searches and measurements in particle physics

Author: James Carzon¹

Co-authors: Aishik Ghosh ²; Ann Lee ¹; Daniel Whiteson ²; Luca Masserano ¹; Rafael Izbicki ³

¹ *Carnegie Mellon University*

² *University of California Irvine (US)*

³ *Federal University of Sao Carlos*

Corresponding Author: jcarzon@andrew.cmu.edu

Particle physics experiments rely on the (generalised) likelihood ratio test (LRT) for searches and measurements. This is not guaranteed to be optimal for composite hypothesis tests, as the Neyman-Pearson lemma pertains only to simple hypothesis tests. An improvement in the core statistical testing methodology would have widespread ramifications across experiments. We discuss an alternate test statistic that provides the data analyzer an ability to focus the power of the test in physics-motivated regions of the parameter space. We demonstrate the improvement from this technique compared to the LRT on a Higgs $\rightarrow \tau\tau$ dataset simulated by the ATLAS experiment and a dark matter dataset inspired by the LZ experiment. This technique also employs machine learning to efficiently perform the Neyman construction that is essential to ensure valid confidence intervals.

Jet Physics / 61

Integrating Energy Flow Networks with Jet Substructure Observables for Enhanced Jet Quenching Studies

Author: João A. Gonçalves¹

¹ *LIP - IST*

Corresponding Author: jgoncalves@lip.pt

The phenomena of Jet Quenching, a key signature of the Quark-Gluon Plasma (QGP) formed in Heavy-Ion (HI) collisions, provides a window of insight into the properties of the primordial liquid. In this study, we evaluate the discriminating power of Energy Flow Networks (EFNs), enhanced with substructure observables, in distinguishing between jets stemming from proton-proton (pp) and jets stemming from HI collisions. This work is a crucial step towards separating HI jets that were quenched from those with little or no modification by the interaction with the QGP on a jet-by-jet

basis. We trained simple Energy Flow Networks (EFNs) and further enhanced them by incorporating jet observables such as N-Subjettiness and Energy Flow Polynomials (EFPs). Our primary objective is to assess the effectiveness of these approaches in the context of Jet Quenching, exploring new phenomenological avenues by combining these models with various encodings of jet information. Initial evaluations using Linear Discriminant Analysis (LDA) set a performance baseline, which is significantly enhanced through simple Deep Neural Networks (DNNs), capable of capturing non-linear relations expected in the data. Integrating both EFPs and N-Subjettiness observables into EFNs results in the most performant model over this task, achieving state-of-the-art ROC AUC values of approximately 0.84. This significant performance is noteworthy given that both medium response and underlying event contamination effects on the jet are taken into account. These results underscore the potential of combining EFNs with jet substructure observables to advance Jet Quenching studies and adjacent areas, paving the way for deeper insights into the properties of the QGP. Results on a variation of EFNs, Moment EFNs (MEFNs), which can achieve comparable performance with a more manageable and, in turn interpretable, latent space, will be presented.

Jet Physics / 62

Deep Learning Methods for Jet Tagging and Process Classification Using Image Processing

Authors: Alexandre Alarcon Do Passo Suaide¹; Daniel Neves Fachieri^{None}; Jhoao Gabriel Martins Campos Almeida Arneiro¹

¹ *Universidade de Sao Paulo (USP) (BR)*

The use of neural networks in high-energy physics has rapidly expanded, particularly in jet tagging applications. This study explores a convolutional neural network (CNN) based approach to classify jets produced in high-energy collisions by differentiating between heavy quark (charm, bottom), light quark (up, down, strange), and gluon jets. The method constructs image-like representations based on the kinematics of charged decay products using detector-level variables, which allow CNNs to identify visual patterns characteristic of each jet type. This approach demonstrate strong classification performance, highlighting the versatility of CNN architectures in jet tagging and advancing our understanding of jet substructures.

Event Generation and Detector Simulation / 63

EveNet: Towards a Generalist Event Transformer for Unified Understanding and Generation of Collider Data

Authors: Bai-Hong Zhou¹; Ben Nachman²; Haoran Zhao³; Qibin Liu⁴; Shih-Chieh Hsu⁵; Shu Li⁶; Ting-Hsiang Hsu⁷; Vinicius Massami Mikuni²; Wei-Po Wang³; Yuan-Tang Chou³; Yue Xu³; Yulei Zhang³

¹ *Tsung-Dao Lee Institute (CN)*

² *Lawrence Berkeley National Lab. (US)*

³ *University of Washington (US)*

⁴ *Tsung-Dao Lee Institute (CN) & Shanghai Jiao Tong University (CN)*

⁵ *University of Washington Seattle (US)*

⁶ *Tsung-Dao Lee Institute, Shanghai Jiao Tong Univ. (CN)*

⁷ *National Taiwan University (TW)*

Corresponding Author: yulei.zhang@cern.ch

With the increasing size of the machine learning (ML) model and vast datasets, the foundation model has transformed how we apply ML to solve real-world problems. Multimodal language models like

chatGPT and Llama have expanded their capability to specialized tasks with common pre-train. Similarly, in high-energy physics (HEP), common tasks in the analysis face recurring challenges that demand scalable, data-driven solutions. In this talk, we present a foundation model for high-energy physics. Our model leverages extensive simulated datasets in pre-training to address common tasks across analyses, offering a unified starting point for specialized applications. We demonstrate the benefit of using such a pre-train model in improving search sensitivity, anomaly detection, event reconstruction, feature generation, and beyond. By harnessing the power of pre-trained models, we could push the boundaries of discovery with greater efficiency and insight.

Reconstruction and Analysis / 64

Machine Learning-Assisted Measurement of Lepton-Jet Azimuthal Angular Asymmetries and of the complete final state in Deep-Inelastic Scattering at HERA

Authors: Benjamin Henry Hooberman¹; Fernando Torales Acosta²; Miguel Arratia^{None}; Stefan Schmitt³; Vinicius Massami Mikuni²

¹ Univ. Illinois at Urbana Champaign (US)

² Lawrence Berkeley National Lab. (US)

³ Deutsches Elektronen-Synchrotron (DE)

Deep-inelastic positron-proton scattering at high momentum transfer Q^2 is an ideal place to study QCD effects. The H1 collaboration presents two such studies based on data collected in ep collisions at $Q^2 > 150 \text{ GeV}^2$. The data are unfolded (corrected for detector effects) using advanced machine learning methods. This results in parallel and unbinned measurements of several observables, hence it is possible to measure quantities such as moments or variables with poor resolution. One such example are moments of the lepton-jet azimuthal angular asymmetry, sensitive to subtle gluon radiation effects which have to be pinned down accurately in order to be able measure TMDs from similar observables. The moments are presented as a function of the total transverse momentum of the lepton-jet system,

$l_{\text{vert}} \vec{q}_{\perp}$

r_{vert} . Another analysis is targeting a simultaneous measurement of all final state particles in these high Q^2 events, such that complex studies such as comparisons of different jet algorithms, or jet substructure measurements can be performed on the unfolded data, free of restrictions on the choice of observables or other technicalities such as bin boundaries. The unfolded dataset is projected for validation purposes onto a few example observables which have been measured earlier. New measurements such as comparisons of jet algorithms or energy-energy correlators are also presented.

arxiv:2412.14092, Submitted to PLB
H1prelim-25-031

Anomaly Detection / 65

Isolating Unisolated Upsilon's with Anomaly Detection in CMS Open Data

Authors: Benjamin Nachman¹; Jesse Thaler²; Radha Mastandrea³; Rikab Gambhir²

¹ Stanford University

² MIT

³ LBNL

Proton collisions at the Large Hadron Collider. Using a machine learning (ML)-based anomaly detection strategy, we “rediscover” the Y in 13 TeV CMS Open Data from 2016, despite overwhelming anti-isolated backgrounds. We elevate the signal significance to 6.4σ using these methods, starting from 1.6σ using the dimuon mass spectrum alone. Moreover, we demonstrate improved sensitivity from using an ML-based estimate of the multi-feature likelihood compared to traditional “cut-and-count” methods. Our work demonstrates that it is possible and practical to find real signals in experimental collider data using ML-based anomaly detection, and we distill a readily-accessible benchmark dataset from the CMS Open Data to facilitate future anomaly detection developments.

Theory / 67

Machine Learning Neutrino-Nucleus Cross Sections

Author: Karla Tame-Narvaez¹

¹ *Fermilab National Accelerator Laboratory*

Neutrino-nucleus scattering cross sections are critical theoretical inputs for long-baseline neutrino oscillation experiments. However, robust modeling of these cross sections remains challenging. For a simple but physically motivated toy model of the DUNE experiment, we demonstrate that an accurate neural-network model of the cross section—leveraging Standard Model symmetries—can be learned from near-detector data. We then perform a neutrino oscillation analysis with simulated far-detector events, finding that the modeled cross section achieves results consistent with what could be obtained if the true cross section were known exactly. This proof-of-principle study highlights the potential of future neutrino near-detector datasets and data-driven cross-section models.

Jet Physics / 68

HEP-JEPA: Towards a found model for high energy physics using joint embedding predictive architecture

Authors: Jai Bardhan^{None}; Abhiram Tilak^{None}; Radhikesh Agrawal^{None}; Cyryn Neera^{None}; Subhadip Mitra^{None}

We present HEP-JEPA, a transformer architecture-based foundation model for tasks at high-energy particle colliders such as the Large Hadron Collider. We pre-train the model on particle jets using a self-supervised strategy inspired by the Joint Embedding Predictive Architecture on the large-scale JetClass dataset containing 100M jets. We evaluate and compare HEP-JEPA to other foundation models on several downstream tasks, such as jet classification and jet observable prediction. HEP-JEPA outperforms or matches the performance of contemporary approaches on these benchmarks.

Reconstruction and Analysis / 69

Optimal Transport for e/π^0 Particle Classification in LArTPC Neutrino Experiments

Authors: Chuyue “Michaelia” Fang^{None}; David Caratelli¹; Jessica N. Howard²; Nathaniel Craig^{None}

¹ *UC Santa Barbara*

² *Kavli Institute for Theoretical Physics*

Corresponding Author: jnhoward@ucsb.edu

The efficient classification of electromagnetic activity from π^0 and electrons remains an open problem in the reconstruction of neutrino interactions in Liquid Argon Time Projection Chamber (LArTPC) detectors. We address this problem using the mathematical framework of Optimal Transport (OT), which has been successfully employed for event classification in other HEP contexts and is ideally suited to the high-resolution calorimetry of LArTPCs. Using an open simulated dataset from the MicroBooNE collaboration, we show that OT methods achieve state-of-the-art reconstruction performance in e/π^0 classification. The success of this first application indicates the broader promise of OT methods for LArTPC-based neutrino experiments.

Fast ML / 70

Convolutional Neural Networks for pile-up suppression in the ATLAS Global Trigger

Author: Stephen Hillier¹

¹ *University of Birmingham (GB)*

We describe a pile-up suppression algorithm for the ATLAS Global Trigger, using a convolutional neural network (CNN) architecture. The CNN operates on cell towers and exploits both shower topology and E_T to correct for the contribution of pile-up. The algorithm is optimised for firmware deployment and demonstrates high throughput and low resource usage. The small size of the input and lightweight implementation enable a high degree of scalability and parallelisation. The physics performance of the algorithm is benchmarked against a range of existing algorithms by reconstructing and calibrating small- R central jets. We find that the CNN allows for the lowest thresholds for multi-jet, jet H_T and H_T^{miss} trigger signatures and, correspondingly, gives the highest acceptance for key signal processes like $HH \rightarrow b\bar{b}b\bar{b}$ and $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ in the simulation.

Jet Physics / 71

Jet tagging with the Lund Jet Plane

Authors: Carlos Sandoval Usme¹; Jean Pierre Cifuentes Salazar¹; Rafael Andrei Vinasco Soler¹; Toni Mlinarević²; Yuanda Zhu³

Co-authors: Jad Mathieu Sardain⁴; Luis Felipe De La Ossa Mayorga; Mario Campanelli⁵; Reina Coromoto Camacho Toro⁶

¹ *Universidad Nacional de Colombia*

² *UCL (GB)*

³ *University of London (GB)*

⁴ *University of Arizona (US)*

⁵ *UCL*

⁶ *LPNHE-Paris CNRS/IN2P3*

The Lund jet plane is a representation of the emissions within a jet, where each point corresponds to an emission. Hard and soft emissions, as well as collinear and wide-angle emissions, correspond to different regions of the Lund plane and are populated differently by jets with different origins. This means the Lund plane can be used for jet tagging.

We present previous studies done in ATLAS on W and top tagging using the Lund plane and recent efforts to improve performance and to reduce background model dependence.

Fast ML / 72

Efficient Transformers for Jet Tagging

Authors: Aaron Wang¹; Abhijith Gandrakota²; Elham Khoda³; Javier Mauricio Duarte⁴; Jennifer Ngadiuba²; Vivekanand Gyanchand Sahu⁵; Zihan Zhao⁴

¹ *University of Illinois Chicago (US)*

² *FNAL*

³ *University of Washington (US)*

⁴ *Univ. of California San Diego (US)*

⁵ *University of California San Diego*

Particle Transformer has emerged as a leading model for jet tagging, but its quadratic scaling with sequence length presents significant computational challenges, especially for longer sequences. This inefficiency is critical in applications such as HL LHC trigger systems where rapid inference is essential. To overcome these limitations, we evaluated several Transformer variants and identified the Linformer as a very promising alternative. Our tests on both small and large models using the JetClass and HLS4ML datasets show that the Linformer dramatically reduces inference time and computational demands measured in FLOPs while nearly matching the performance of the Particle Transformer. We also examined the impact of the input sequence order by testing various strategies, including those based on physics motivated projection matrices, to further improve performance. Finally, we employed interpretability methods such as analyzing the attention matrices and examining the embeddings to gain deeper insights into the model operation.

Theory / 73

A novel loss function to optimise signal significance in particle physics

Authors: Jai Bardhan^{None}; Cyrin Neera^{None}; Subhadip Mitra^{None}; Tanumoy Mandal¹

¹ *IISER Thiruvananthapuram*

We construct a surrogate loss to directly optimise the significance metric used in particle physics. We evaluate our loss function for an event classification task and show that it produces decision boundaries that change according to the cross sections of the processes involved. We find that the models trained with the new loss have higher signal efficiency for similar values of estimated signal significance compared to ones trained with a cross-entropy loss, showing promise to improve the sensitivity of particle physics searches at colliders.

Uncertainties & Interpretability / 74

Tackling interpretability with physical baselines for Integrated Gradients

Authors: Jai Bardhan^{None}; Cyrin Neera^{None}; Mihir Rawat^{None}; Subhadip Mitra^{None}

Machine learning methods have seen a meteoric rise in their applications in the scientific community. However, little effort has been put into understanding these “black box” models. We show how one can apply integrated gradients (IGs) to understand these models by designing different baselines, by taking an example case study in particle physics. We find that the zero-vector baseline does not provide good feature attributions and that an averaged baseline sampled from the background events provides consistently more reasonable attributions.

Unfolding & Inference / 75**Generator Based Inference (GBI)**

Authors: Alkaid Cheng¹; Benjamin Nachman²; David Shih^{None}; Gurpreet Singh³; Radha Mastandrea³; Ranit Das⁴; Runze Li⁵; Vinicius Massami Mikuni⁶

¹ *University of Wisconsin Madison (US)*

² *Stanford University*

³ *University of California, Berkeley*

⁴ *Rutgers University*

⁵ *Yale University (US)*

⁶ *Lawrence Berkeley National Lab. (US)*

Corresponding Author: chi.lung.cheng@cern.ch

Statistical inference in physics is often based on samples (from a ‘forward model’) that emulate experimental data and depend on parameters of the underlying theory. Modern machine learning has supercharged this workflow to enable high-dimensional and unbinned analyses to utilize much more information than ever before. We propose a general framework for describing the integration of machine learning with forward models called Generator Based Inference (GBI). A well-studied special case of this setup is Simulation Based Inference (SBI) where the forward model is a physics-based simulator. In this work, we examine other methods within the GBI toolkit that use data-driven methods to build the forward model. In particular, we focus on resonant anomaly detection, where the forward model describing the background is learned from sidebands. We show how to perform machine learning-based parameter estimation in this context with data-derived forward models. This transforms the statistical outputs of anomaly detection to be directly interpretable and the performance on the LHC community benchmark dataset establishes a new state-of-the-art for anomaly detection sensitivity.

Fast ML / 76**Jet calibration with in-situ pileup suppression for the L1 trigger**

Author: Ben Carlson¹

Co-authors: Stephen Roche²; Tae Min Hong³

¹ *Westmont College*

² *University of Pittsburgh*

³ *University of Pittsburgh (US)*

Corresponding Author: bcarlson@cern.ch

We present a method to suppress pileup and calibrate hadronic jet energy at L1 triggers using boosted decision trees for regression and classification. The fwX platform is used for implementation of BDTs on FPGA within the necessary timing and resource constraints. The in-situ pileup suppression can improve trigger performance in the high pileup environment of the HL-LHC.

Fast ML / 77**Real-Time Compression of CMS Detector Data Using Conditional Autoencoders**

Author: Zachary Baldwin¹

Co-authors: Danny Noonan²; Erdem Yigit Ertorer³; Jim Hirschauer²; Mariel Peczak; Matteo Cremonesi³; Nate Woodward; Nhan Tran²; Peter Eduard Meiring³; Philip Coleman Harris⁴; Simon Rothman⁴

¹ *Carnegie Mellon University*

² *Fermi National Accelerator Lab. (US)*

³ *Carnegie-Mellon University (US)*

⁴ *Massachusetts Inst. of Technology (US)*

Corresponding Author: zbaldwin@andrew.cmu.edu

The upcoming high-luminosity upgrade to the LHC will involve a dramatic increase in the number of simultaneous collisions delivered to the Compact Muon Solenoid (CMS) experiment. To deal with the increased number of simultaneous interactions per bunch crossing as well as the radiation damage to the current crystal ECAL endcaps, a radiation-hard high-granularity calorimeter (HGCAL) will be installed in the CMS detector. With its six million readout channels, the HGCAL will produce information on the energy and position of detected particles at a rate of 5 Pb/s. These data rates must be reduced by several orders of magnitude in a few microseconds in order to trigger on interesting physics events. We explore the application of machine learning for data compression performed by the HGCAL front-end electronics. We have implemented a conditional autoencoder which compresses data on the ECON-T ASIC before transmission off-detector to the rest of the trigger system.

Fast ML / 78

GELATO: A Generic Event-Level Anomalous Trigger Option for ATLAS

Author: Sagar Addepalli¹

¹ *SLAC National Accelerator Laboratory (US)*

The absence of beyond-Standard-Model physics discoveries at the LHC suggests that new physics may evade conventional trigger strategies. The existing ATLAS triggers are required to control data collection rates with high energy thresholds and target signal topologies specific to only certain models. Unsupervised machine learning enables the use of anomaly detection, presenting a unique model-agnostic way to search for anomalous signatures that deviate from Standard Model expectations. We present a new trigger sequence using fast anomaly detection algorithms in both the hardware and software triggers implemented for ATLAS Run-3 data-taking. The design and performance of the triggers will be described along with their integration and commissioning strategy with an emphasis on rate stability and operational robustness. First results from analysis of data collected through this new trigger stream, focusing on validating the trigger response, will be shown. This first anomaly detection trigger for ATLAS provides a framework for future machine learning implementations in the trigger system. The approach offers potential for novel sensitivity to a broad spectrum of new physics signatures in Run-3 and beyond.

Jet Physics / 79

Fast Jet Tagging with MLP-Mixers on FPGAs

Authors: Chang Sun¹; Jennifer Ngadiuba²; Maria Spiropulu¹; Maurizio Pierini³

¹ *California Institute of Technology (US)*

² *FNAL*

³ CERN

We explore the innovative use of MLP-Mixer models for real-time jet tagging and establish their feasibility on resource-constrained hardware like FPGAs. MLP-Mixers excel in processing sequences of jet constituents, achieving state-of-the-art performance on datasets mimicking Large Hadron Collider conditions. By using advanced optimization techniques such as High-Granularity Quantization and Distributed Arithmetic, we achieve unprecedented efficiency. These models match or surpass the accuracy of previous architectures, reduce hardware resource usage by up to 97%, double the throughput, and half the latency. Additionally, non-permutation-invariant architectures enable smart feature prioritization and efficient FPGA deployment, setting a new benchmark for machine learning in real-time data processing at particle colliders.

Theory / 80

Machine Learning Symmetries in Physics from First Principles

Author: Konstantin Matchev¹¹ University of Alabama (US)

Symmetries are the cornerstones of modern theoretical physics, as they imply fundamental conservation laws. The recent boom in AI algorithms and their successful application to high-dimensional large datasets from all aspects of life motivates us to approach the problem of discovery and identification of symmetries in physics as a machine-learning task. In a series of papers, we have developed and tested a deep-learning algorithm for the discovery and identification of the continuous group of symmetries present in a labeled dataset. We use fully connected neural network architectures to model the symmetry transformations and the corresponding generators. Our proposed loss functions ensure that the applied transformations are symmetries and that the corresponding set of generators is orthonormal and forms a closed algebra. One variant of our method is designed to discover symmetries in a reduced-dimensionality latent space, while another variant is capable of obtaining the generators in the canonical sparse representation. Our procedure is completely agnostic and has been validated with several examples illustrating the discovery of the symmetries behind the orthogonal, unitary, Lorentz, and exceptional Lie groups.

Invited Plenaries / 81

Progresses on AI-based jet tagging

Corresponding Author: huilin.qu@cern.ch

Invited Plenaries / 82

End-to-end particle reconstruction for current and future colliders

Corresponding Author: eilam.work@gmail.com

Invited Plenaries / 83

AI for gravitational waves

Corresponding Author: philip.coleman.harris@cern.ch

Invited Plenaries / 84

Uncertainty quantification in machine learning: A selective overview

Corresponding Author: prasanth@fnal.gov

Invited Plenaries / 85

AI for particle accelerators

Corresponding Author: edelen@slac.stanford.edu

Invited Plenaries / 87

Foundation models for astrophysics & cosmology

Invited Plenaries / 88

Likelihood free inference

Corresponding Author: aishik.ghosh@cern.ch

Invited Plenaries / 89

AI-driven detector design

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

Invited Plenaries / 90

AI-based end-to-end simulation

Corresponding Author: andrea.rizzi@cern.ch

Invited Plenaries / 91

AI at the extreme edge

Corresponding Author: jannicke.pearkes@cern.ch

92

Closing remarks

Jet Physics / 93

Comparing Continuous and Tokenized Jet Generation Approaches for Precision Modeling

Authors: Darius Faroughy^{None}; David Shih^{None}; Gregor Kasieczka¹; Ian Pang^{None}; Ranit Das²

¹ *Hamburg University (DE)*

² *Rutgers University*

Corresponding Author: ian.pang@physics.rutgers.edu

In this work, we explore two distinct jet generation approaches—continuous flow matching models and tokenized autoregressive models—to assess their effectiveness in precision generative modeling of jets. We examine the strengths and limitations of each approach, providing insights into how they inform the theoretical limits of jet tagging and the trade-offs between continuous and tokenized representations. Additionally, we demonstrate how flow matching models enable accurate density estimation and further improve existing tokenized jet generative models enhancing their performance for jet physics applications.

Jet Physics / 94

Get Your Jets in Shape: Conditioning Heads and Backbones

Authors: Anna Hallin¹; David Shih^{None}; Gregor Kasieczka²; Humberto Reyes-González³; Ian Pang^{None}; Joschka Birk²; Michael Kramer^{None}; Oz Amram⁴

¹ *University of Hamburg*

² *Hamburg University (DE)*

³ *RWTH Aachen*

⁴ *Fermi National Accelerator Lab. (US)*

Corresponding Author: ian.pang@physics.rutgers.edu

In recent years, foundation models for jet physics, often based on transformer architectures, have been increasingly applied to tasks such as jet classification, jet generation, and anomaly detection. Conditioning on high-level jet features (HLFs) is crucial for many of these tasks. However, pre-training on all possible HLFs is impractical due to their vast variety. To address this, we propose a conditional finetuning framework that adapts pretrained models to specific tasks without retraining from scratch. Our approach introduces task-relevant features during finetuning, either by modifying the output head to accept task-specific variables or by adding auxiliary control networks to the intermediate layers. The latter allows for localized conditioning of activations, guiding the pretrained model with task-specific information while preserving its general capabilities. This method enhances

model adaptability and performance across diverse jet-based applications, offering an efficient way to specialize without exhaustive retraining.

Invited Plenaries / 95

AI in HEP Theory

Theory / 96

An Energy Correlation Function tagger for gluon-gluon resonances

Author: Connor Moore¹

¹ *University of Notre Dame (US)*

In this talk, we present a gluon-gluon resonance tagger using Energy Correlation Functions evaluated on large-radius jets. We discuss the performance in simulation and also assess the ability of simulation to accurately model high-point correlator functions in data.