UNIVERSITY OF EDINBURGH

ML & AI applications in Experimental Particle Physics

Christos Leonidopoulos

Higgs Maxwell Workshop

February 2025

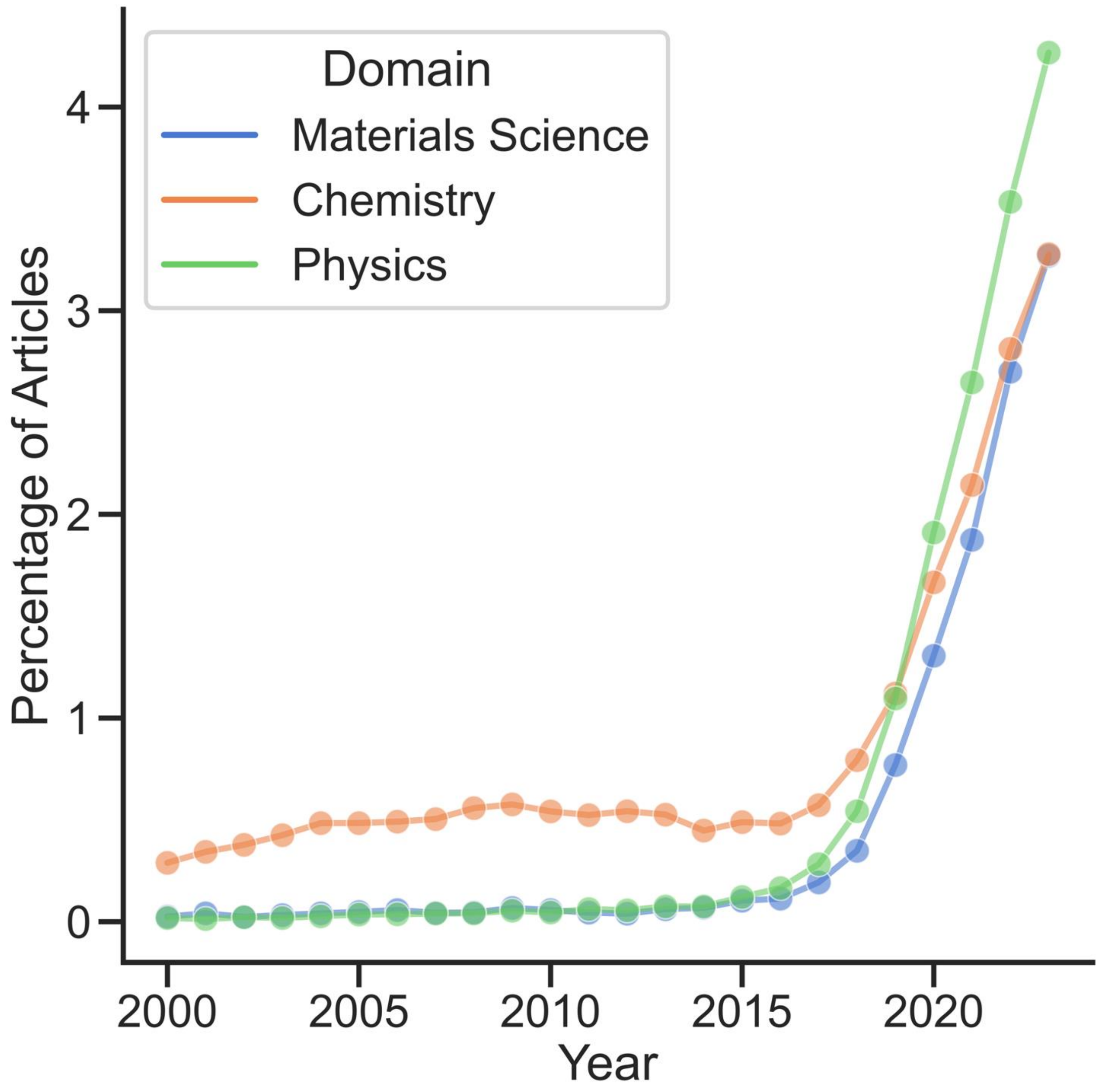
CHRISTOS LEONIDOPOULOS

Executive Summary

- Powerful new technologies & methodologies
- But also new ways to produce new ideas

Executive Summary

CHRISTOS LEONIDOPOULOS



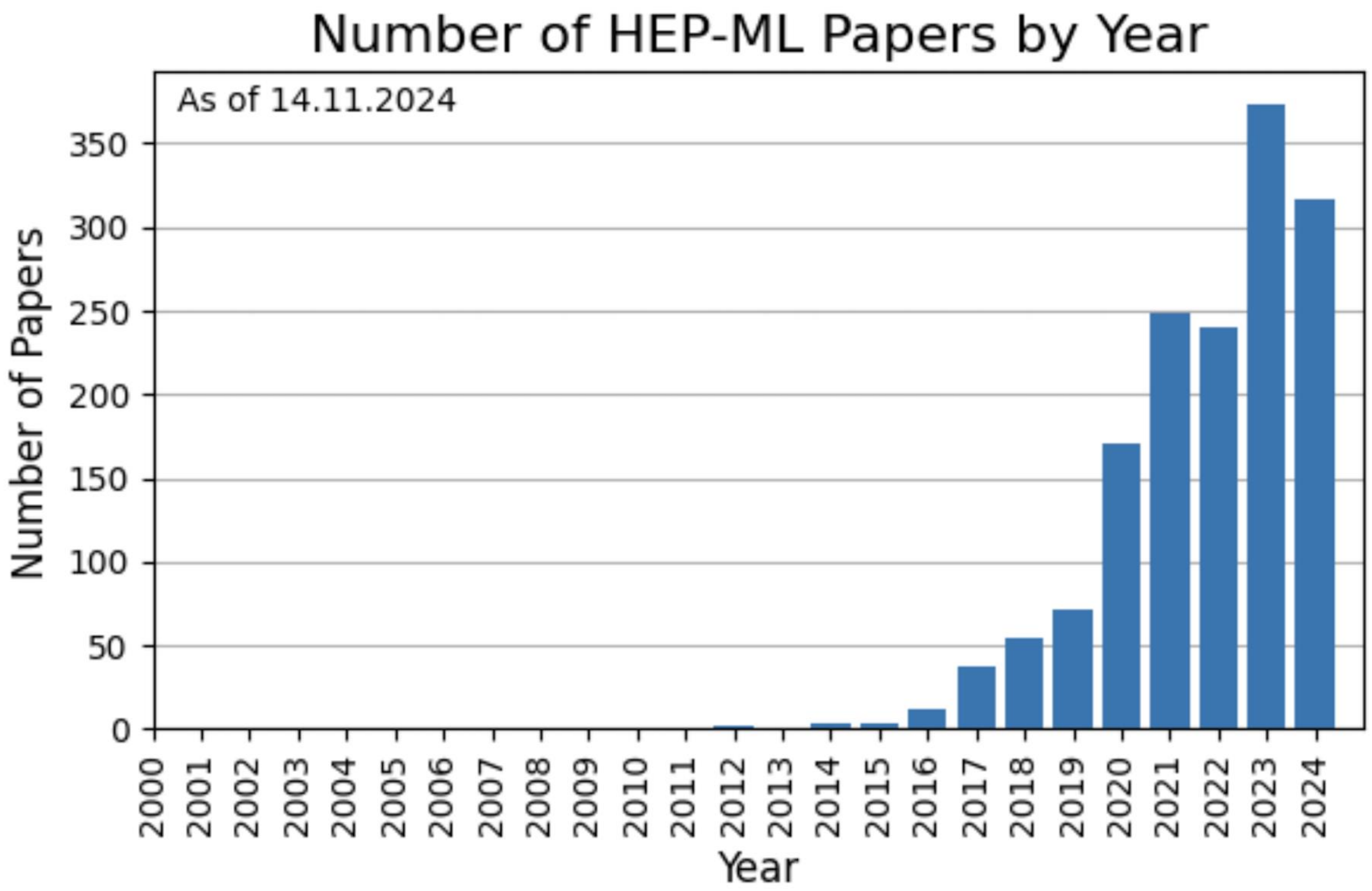
Methodology

- Web of Science topic matching and matching of domain. Exact queries are provided in the data directory.

ID	Service	Query
1	Web of Science	TS=("machine learning" OR "informatics" OR "deep learning" OR "cheminformatics" OR "artificial intelligence" OR "chemoinformatics" OR "QSAR" OR "QSPR") AND WC="{Domain}"

Fraction of Physics papers involving ML in recent years

doi:10.5281/zenodo.7057437



Number of HEP papers involving ML in recent years

https://iml-wg.github.io/HEPML-LivingReview/

Disclaimer:

- Impossible to do a thorough review of all interesting techniques & results
- Focus on newer methods to show breadth of applications & future directions

CHRISTOS LEONIDOPOULOS

07

Terminology

Difference between ML & AI

Machine Learning

Can learn from data, identify patterns, make decisions with minimal human intervention using statistical methods

Artificial Intelligence

Broader concept: encompasses not only ML but also logic and rule-based algorithms which can mimic human thought process & reasoning; includes language understanding, decision-making, problem-solving

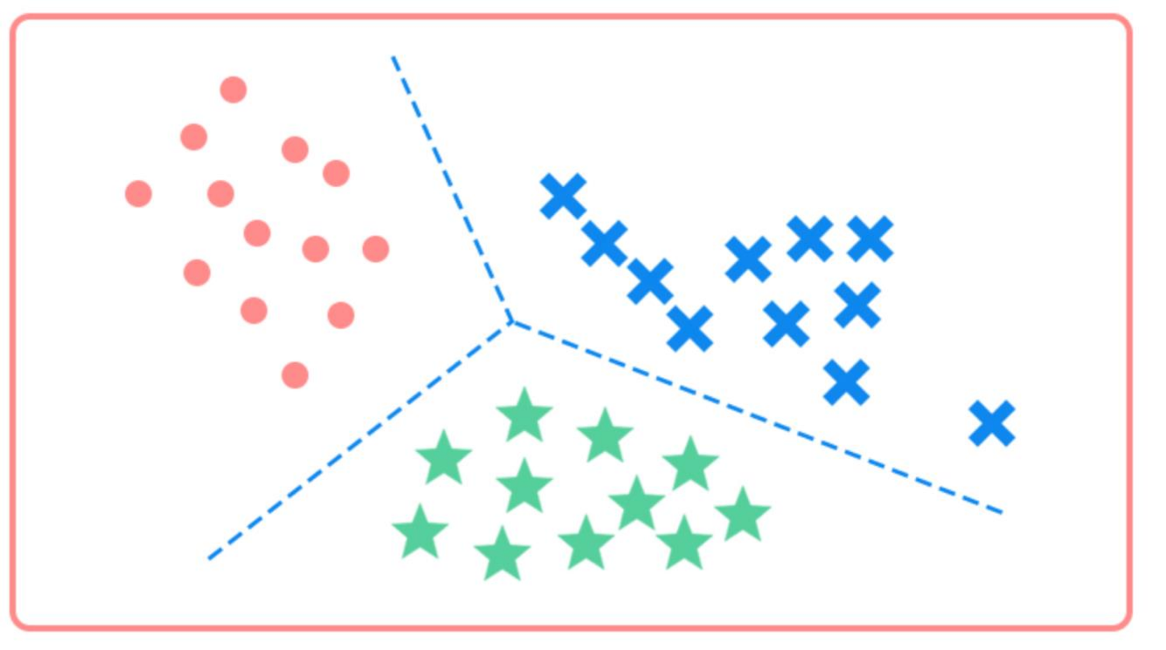
CHRISTOS LEONIDOPOULOS
THE UNIVERSITY OF EDINBURGH

09

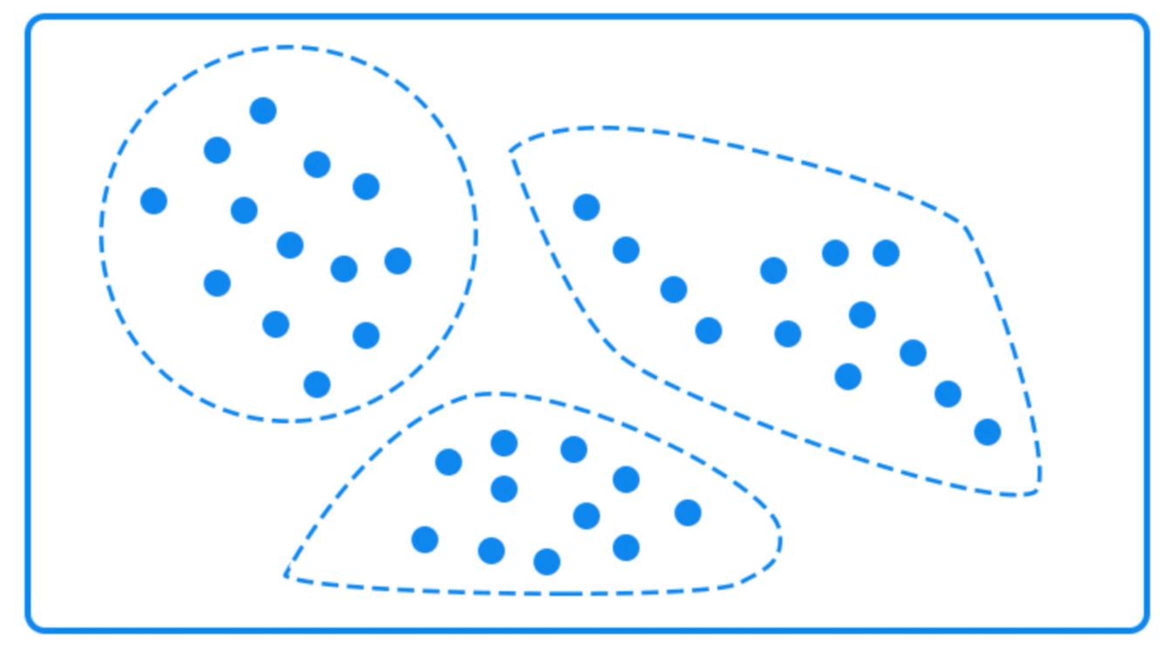
CHRISTOS LEONIDOPOULOS

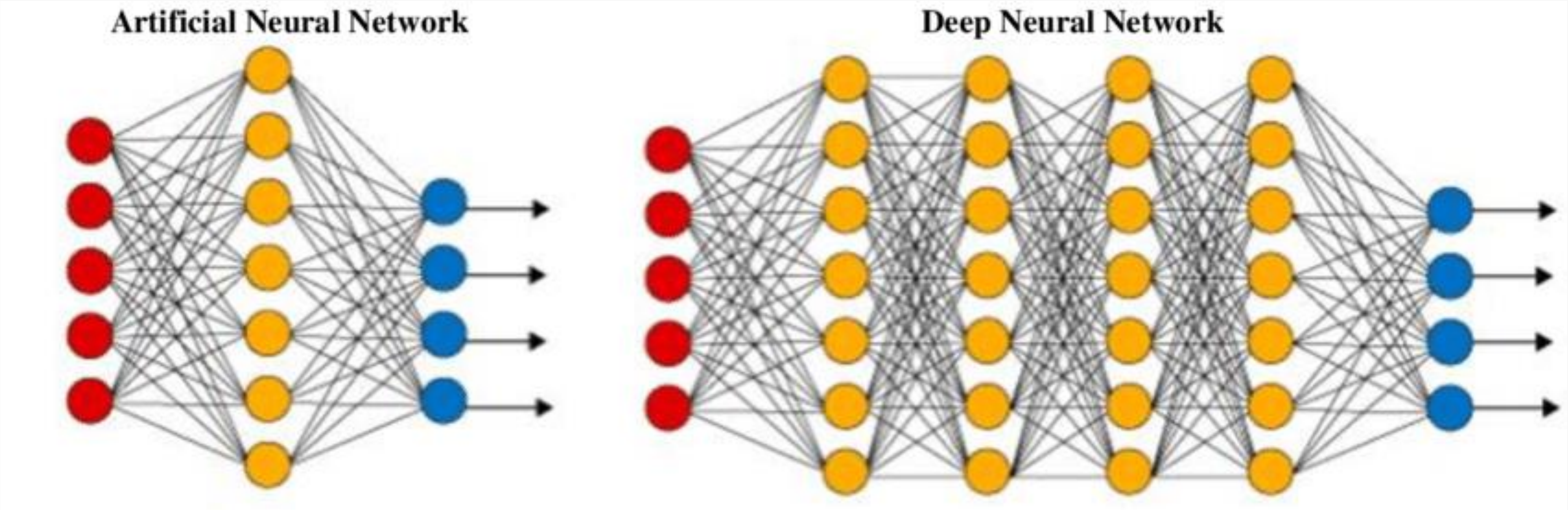
Supervised & unsupervised learning

- Supervised learning: classification (data has labels)



- Unsupervised learning: clustering (data has no labels)





Neural Networks & Deep Neural Networks

"Deep": complexity & hierarchy in data

Reconstruction

Neutrino reconstruction

PhysRevD.111.L041301

YU, KAMP, and ARGÜELLES PHYS. REV. D **111,** L041301 (2025)

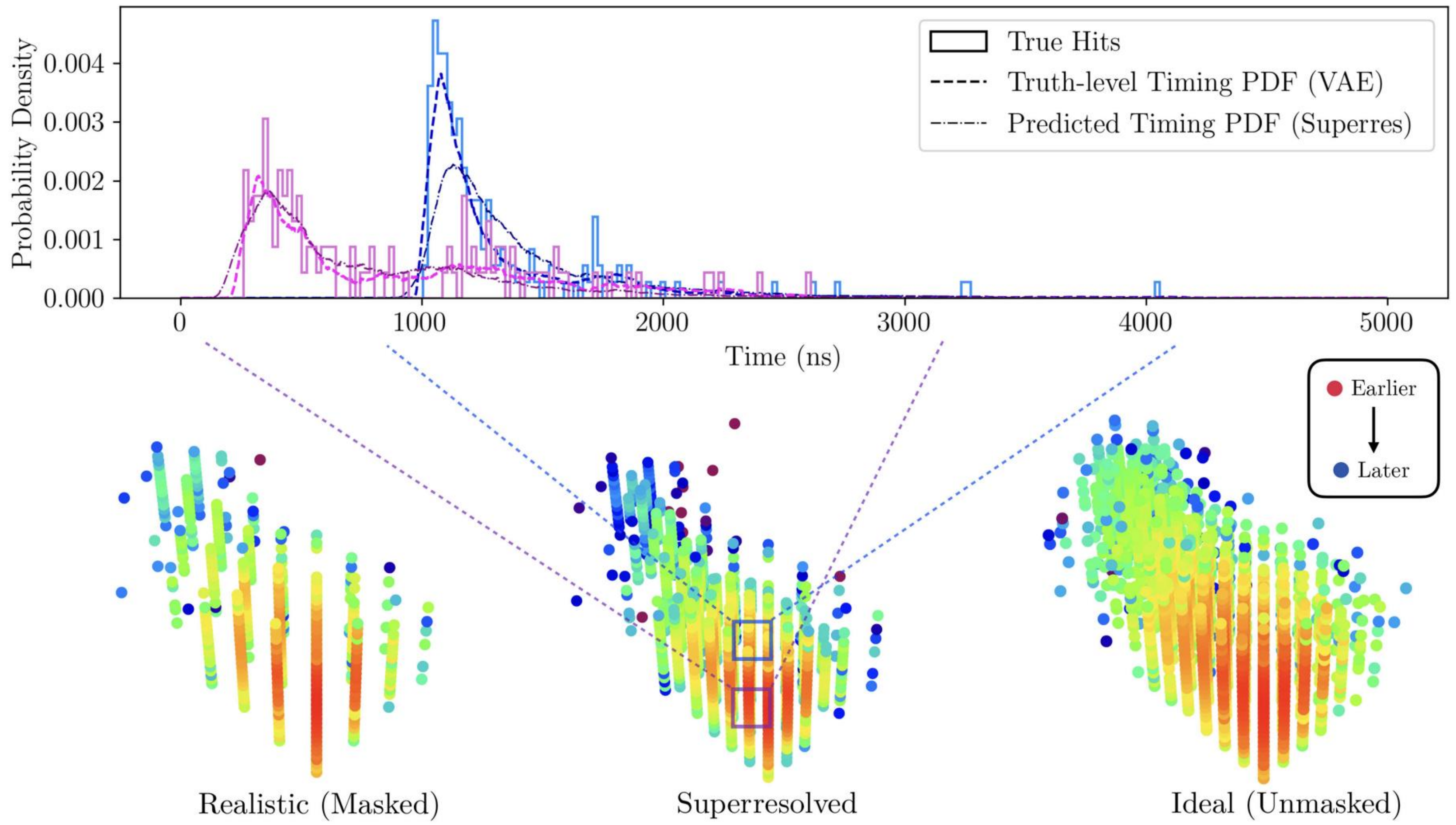


FIG. 1. Event displays, showing the masked, superresolved, and unmasked event. The unmasked and masked events are obtained from simulation, representing ideal and realistic detector configurations, while the superresolution network attempts to enhance the masked event into the unmasked. The top plot shows the photon arrival time series from the superresolution network and the pretrained VAE on two particular virtual OMs in the superresolved event.

IceCube Neutrino Observatory:

- Reconstruction algorithms limited by sparsity of optical modules for detection of photons
- ML interpolates photon path and calculates hypothetical hits assuming “virtual” optical modules

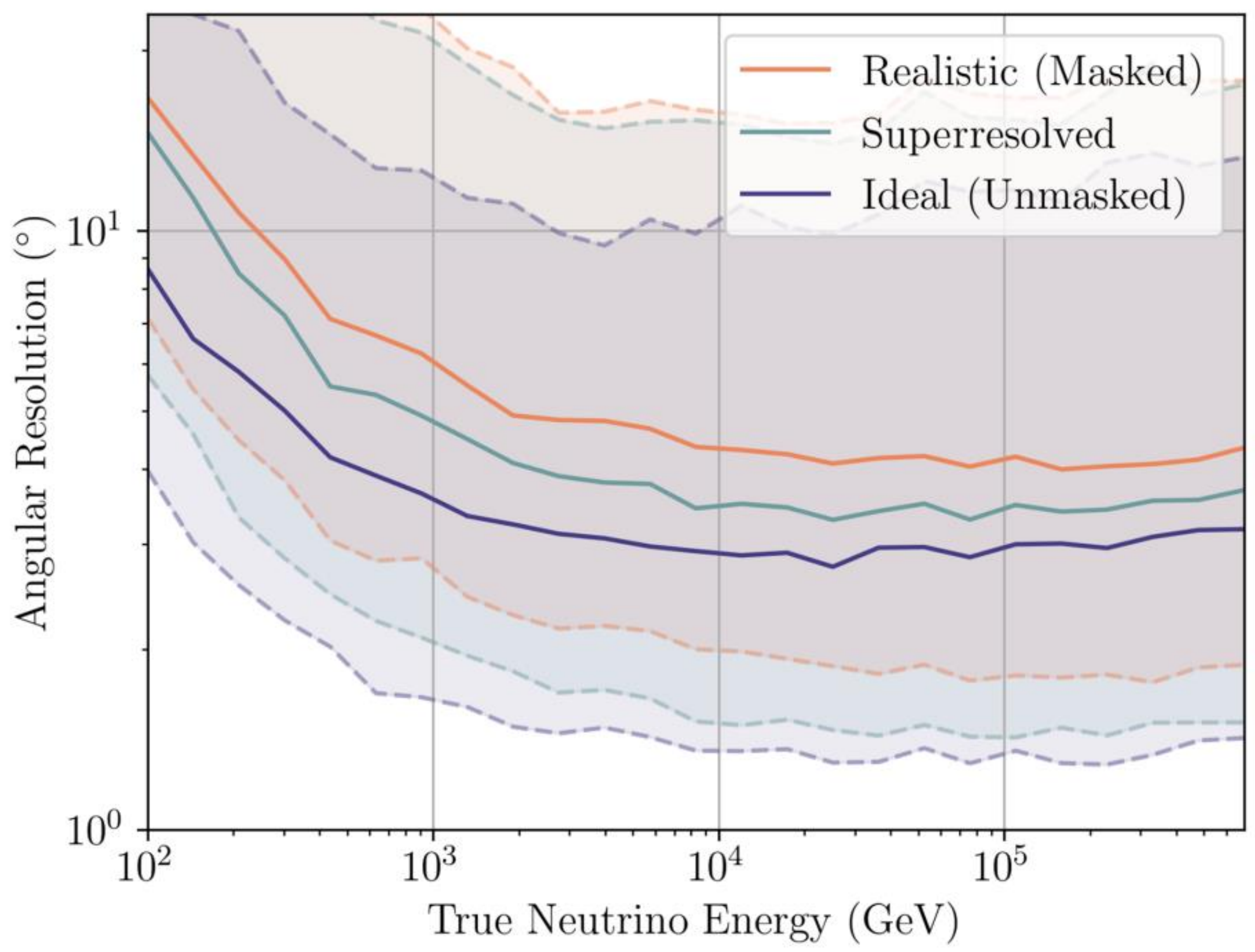


FIG. 4. Log-scale angular resolution. The median lines are drawn in solid color as a function of the true neutrino energy, produced by a baseline SSCNN method. The 20 and 80 percentiles are denoted by the dashed lines and shaded regions.

PHYSICAL REVIEW D **111,** L041301 (2025)

Letter

Enhancing events in neutrino telescopes through deep-learning-driven superresolution

Felix J. Yu,[1,2,*] Nicholas Kamp,[2,†] and Carlos A. Argüelles[1,2,‡]
[1]*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[2]*Department of Physics and Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 3 September 2024; accepted 6 January 2025; published 5 February 2025)

Recent discoveries by neutrino telescopes, such as the IceCube Neutrino Observatory, relied extensively on machine learning (ML) tools to infer physical quantities from the raw photon hits detected. Neutrino telescope reconstruction algorithms are limited by the sparse sampling of photons by the optical modules due to the relatively large spacing (10–100 m) between them. In this Letter, we propose a novel technique that learns photon transport through the detector medium through the use of deep-learning-driven superresolution of data events. These “improved” events can then be reconstructed using traditional or ML techniques, resulting in improved resolution. Our strategy arranges additional “virtual” optical modules within an existing detector geometry and trains a convolutional neural network to predict the hits on these virtual optical modules. We show that this technique improves the angular reconstruction of muons in a generic ice-based neutrino telescope. Our results readily extend to water-based neutrino telescopes and other event morphologies.

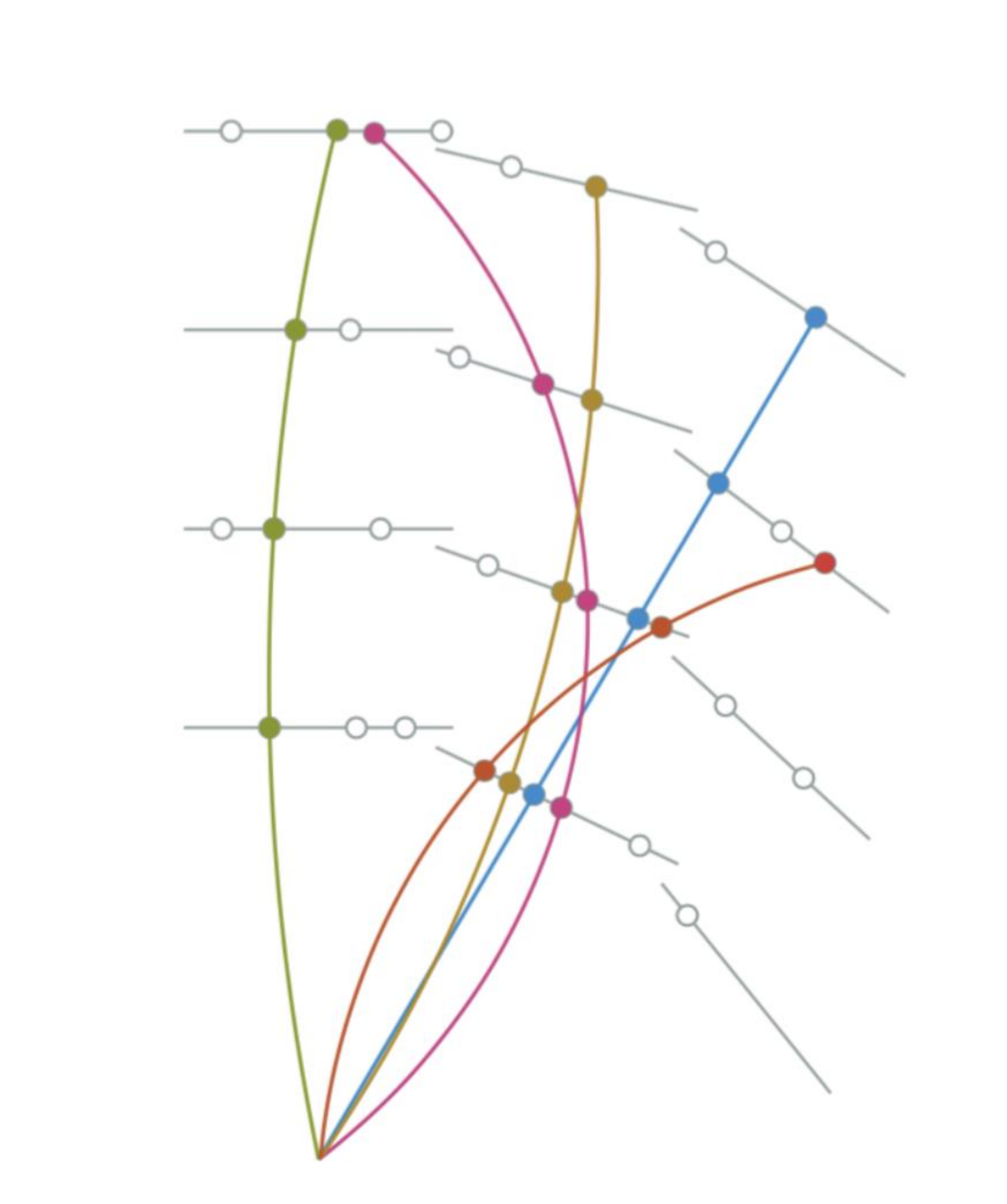
DOI: 10.1103/PhysRevD.111.L041301

Neutrino reconstruction

PhysRevD.111.L041301

CHRISTOS LEONIDOPOULOS

Particle Tracking



- Track reconstruction is a clustering problem: start with a (large) number of 3D points, identify set of points belonging to a common trajectory

- Combinatorial Kalman Filter considers all combinations of hits:
    - Excellent efficiency & purity
    - CPU performance scales very badly with increasing # of hits
    - Solutions for LHC (1k hits per event) not applicable later (HL-LHC, FCC)

Particle Tracking: The Problem

CHRISTOS LEONIDOPOULOS

Connecting The Dots 2023

10 Oct 2023, 09:00 → 13 Oct 2023, 18:00 Europe/Zurich

Toulouse

Description

8th International Connecting The Dots Workshop

The Connecting The Dots workshop series brings together experts on track reconstruction and other problems involving pattern recognition in sparsely sampled data. While the main focus will be on High Energy Physics (HEP) detectors, the Connecting The Dots workshop is intended to be inclusive across other scientific disciplines wherever similar problems or solutions arise.

The 2023 edition will be hosted in Toulouse (France). It is the 8th in the series after: Berkeley 2015, Vienna 2016, Orsay 2017, Seattle 2018, Valencia 2019, virtual in 2020 and Princeton 2022.

The workshop is plenary sessions only, with a mix of invited talks and submitted contributions. There will also be a Poster session.

CTD 2023 is organised as an **in-person conference** and no remote presentation is foreseen. We expect all presenters to register.

The last day, Friday 13 October, is dedicated to a satellite **mini-workshop on Real time Tracking : triggering events with tracks,** see the dedicated indico page. Registration to the mini-workshop are free and independent of the main CTD conference (and if you register to CTD, you are not automatically registered to the mini-workshop).

Important dates

Abstract submission: **26 May - ~~30 June~~ 14 July 2023 (The call for abstracts is now closed)**
Registration deadlines **: Early-bird 1st September,** otherwise 22 September 2023**.**

Fees
- Standard **350€**
- Early Bird (up to 01/09/2023) **315€**
- Students **220€**

This fee covers local support, morning and afternoon coffee breaks, lunches, the welcome reception and workshop dinner.

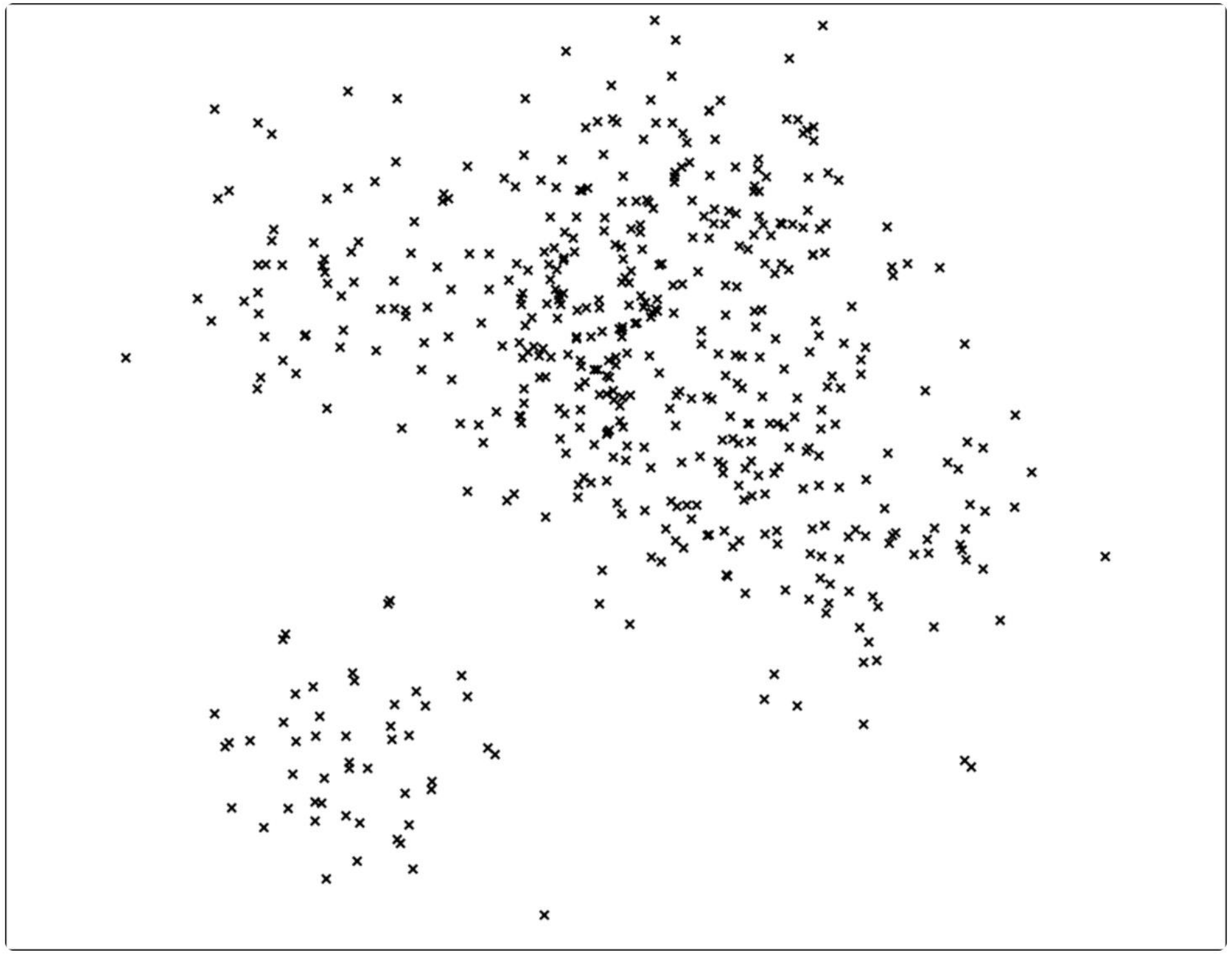
Institut national de physique nucléaire et de physique des particules

Université de Toulouse

Université Toulouse III Paul Sabatier

Particle Tracking: connecting the dots

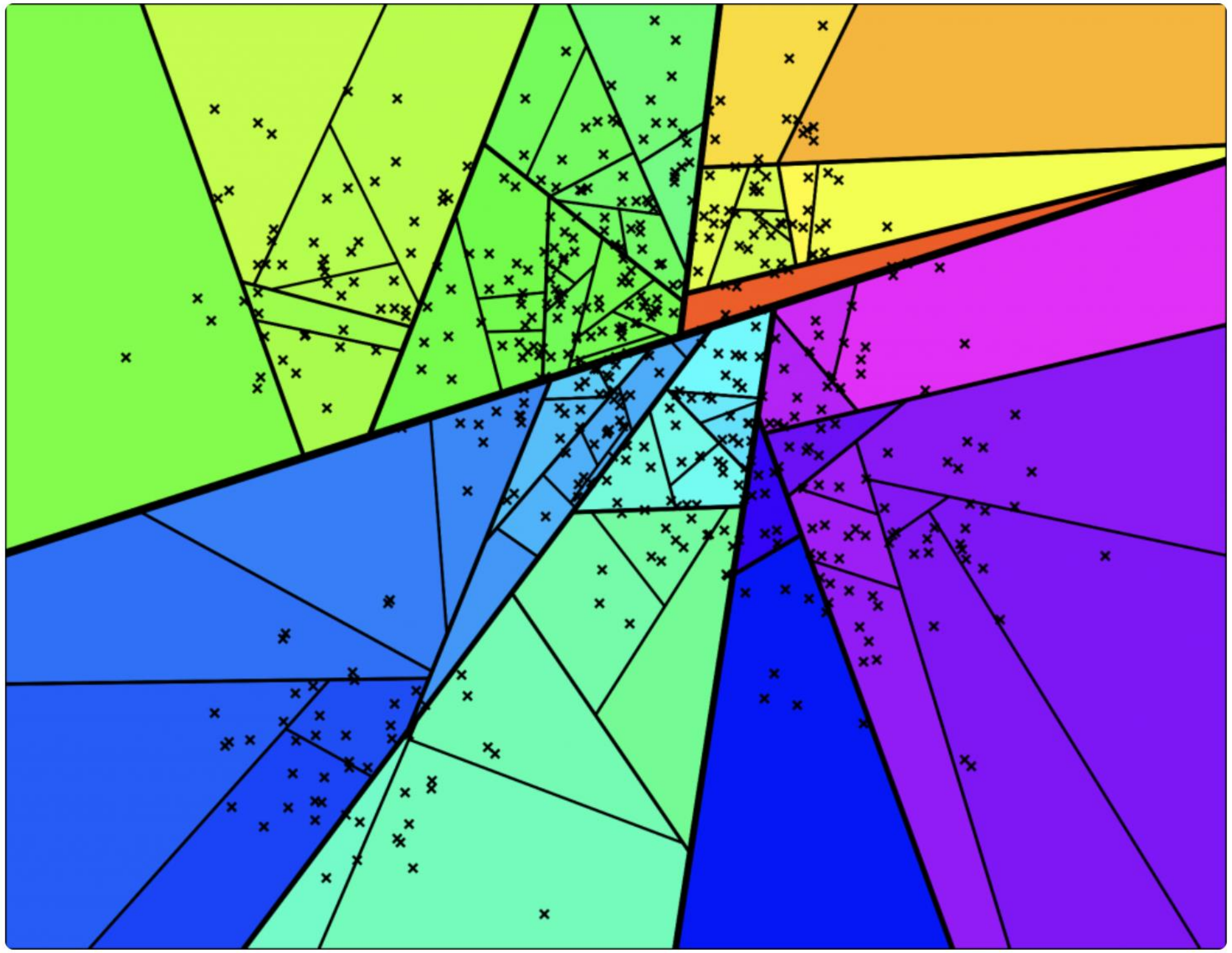
CHRISTOS LEONIDOPOULOS



Particle Tracking: Nearest Neighbour Search

CHRISTOS LEONIDOPOULOS

CHRISTOS LEONIDOPOULOS

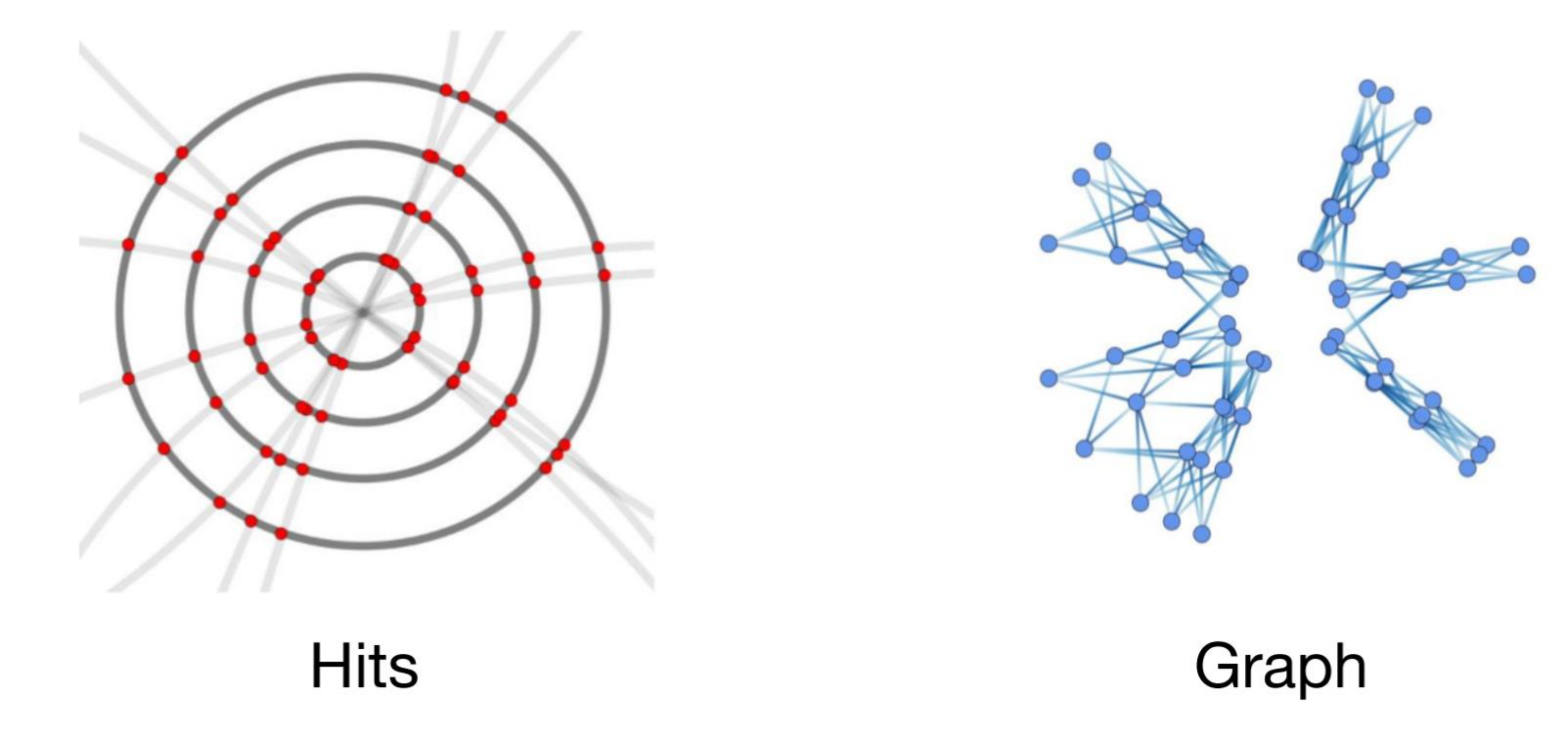


credit

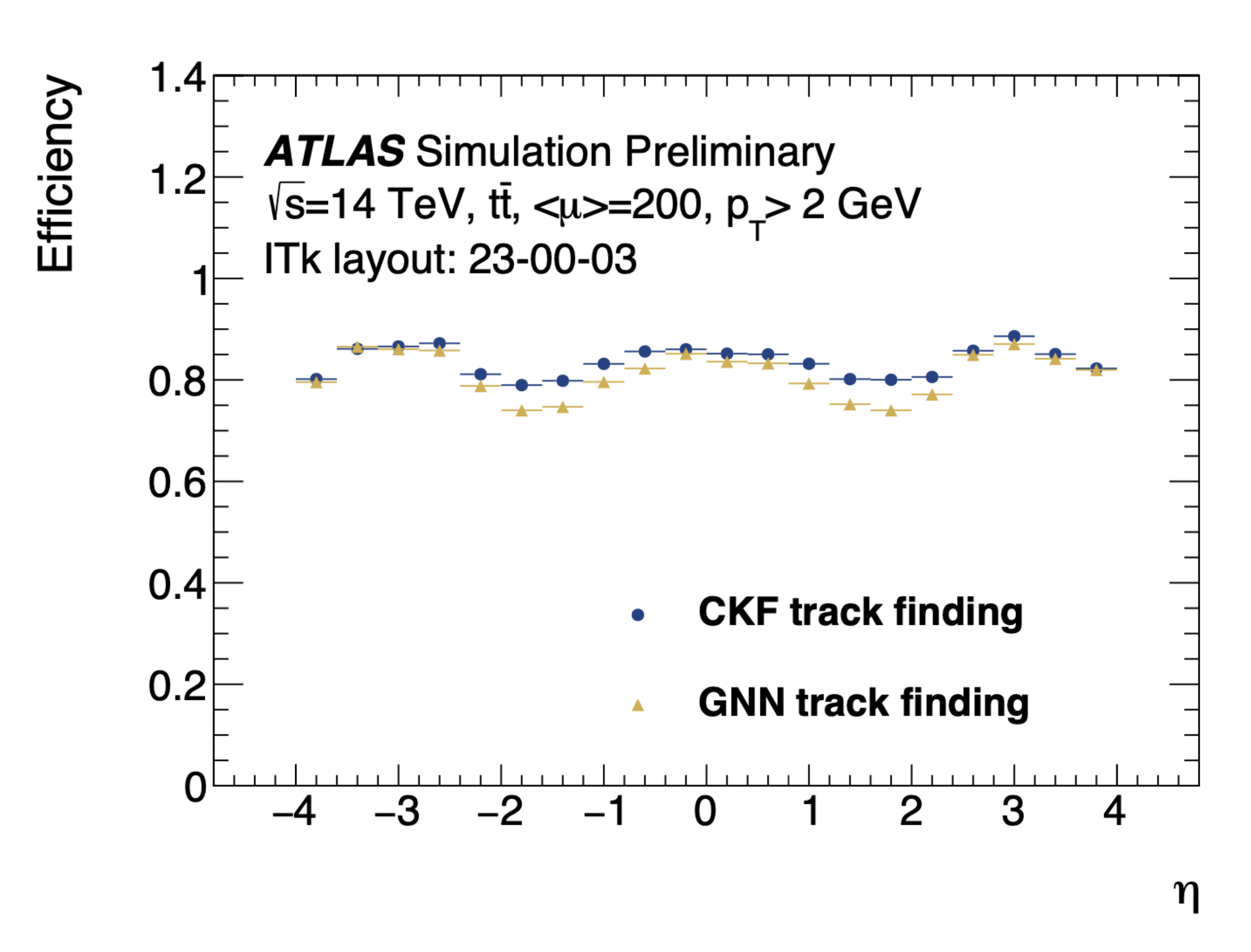
Particle Tracking: Nearest Neighbour Search

Particle Tracking: Graph Neural Networks

ATL-ITK-PROC-2022-006



credit



IDTR-2023-06

- GNN converts collection of hits into a graph: nodes correspond to hits, edges to probability that hits belong to same track

Particle Tracking: Recurrent Neural Networks

arXiv:2212.02348

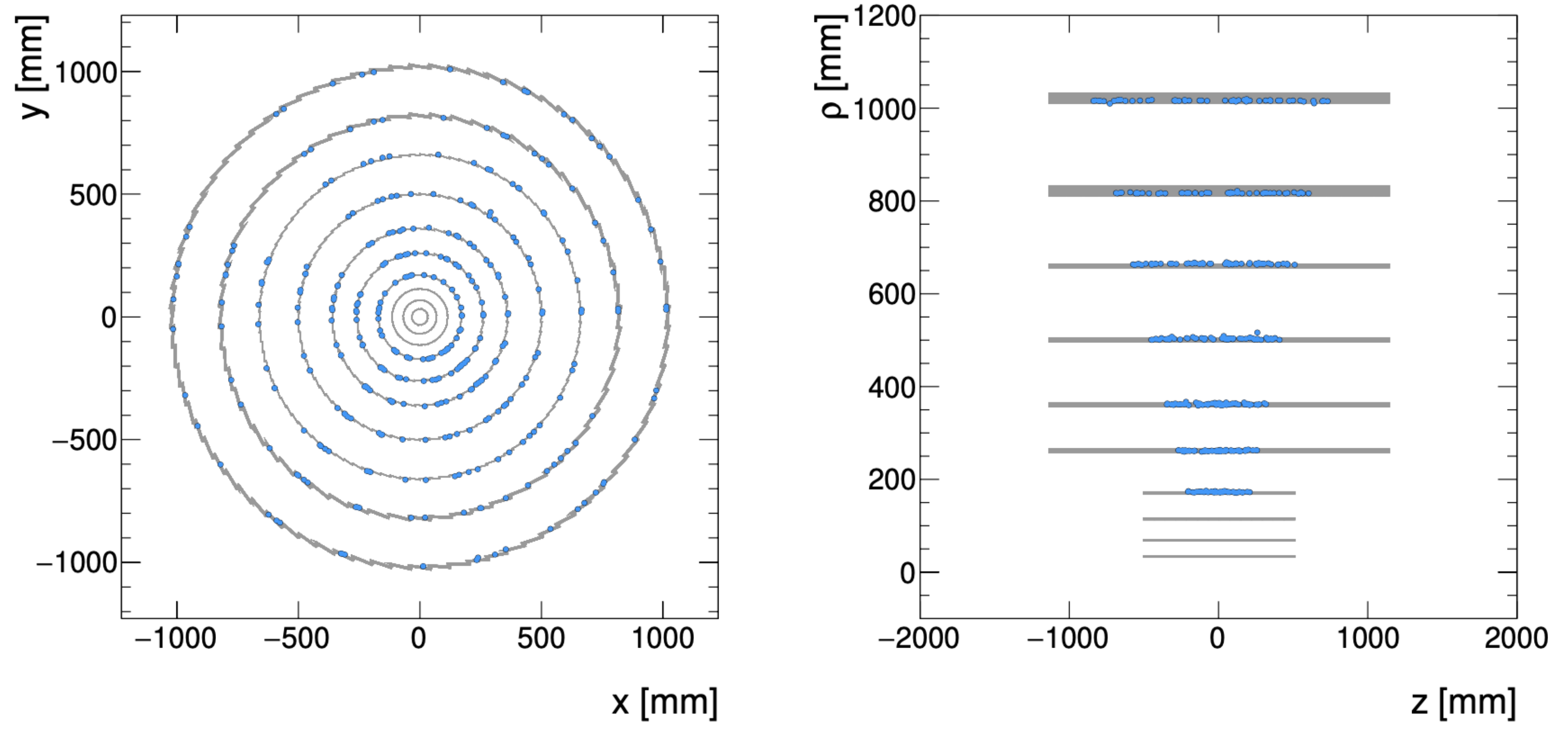
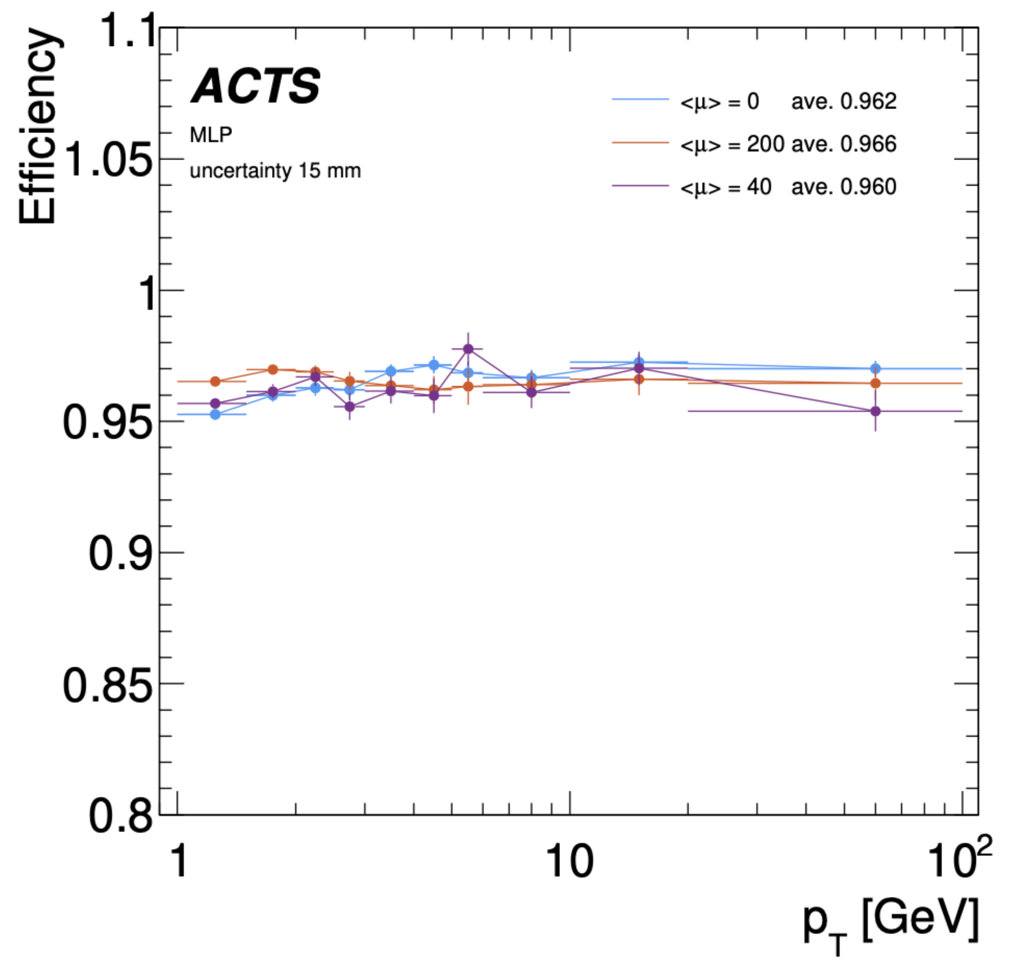


Figure 2. The predicted hit coordinates output by an RNN where the preceding three true hits are used as input, are overlaid on the ACTS detector. Tracks are required to have at least 8 hits within the barrel.



RNN

- Designed to handle sequential data
- Suited for tasks where the order and context of data points are crucial
- Represent a significant leap in our ability to model sequences in data

CHRISTOS LEONIDOPOULOS

Particle Tracking: Recurrent Neural Networks

arXiv:2212.02348

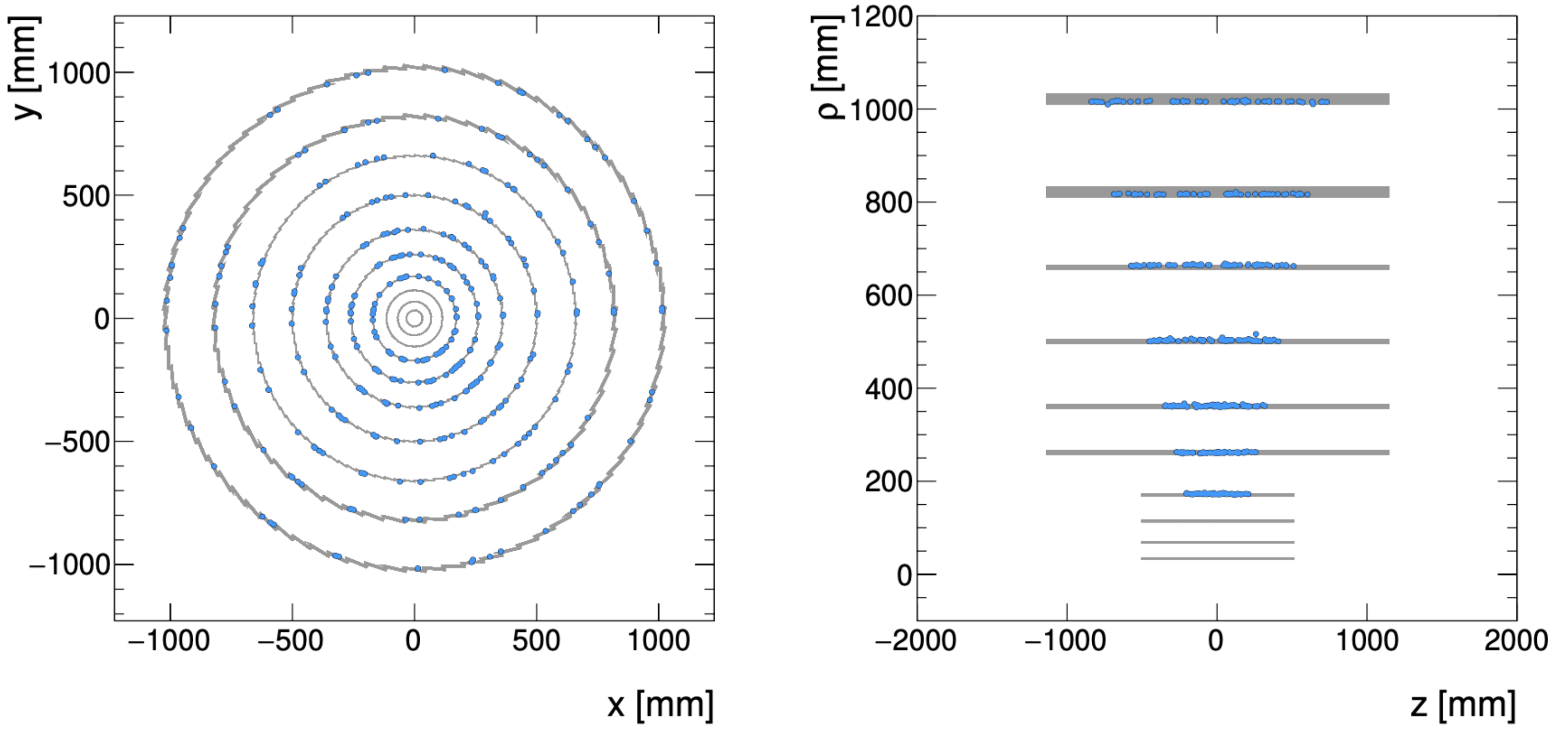


Figure 2. The predicted hit coordinates output by an RNN where the preceding three true hits are used as input, are overlaid on the ACTS detector. Tracks are required to have at least 8 hits within the barrel.

Charged Particle Tracking with Machine Learning on FPGAs

H. Abidi,[a] A. Boveia,[b] V. Cavaliere,[a] D. Furletov,[c] A. Gekow,[b] C. W. Kalderon,[a] S. Yoo[a]

[a] *Brookhaven National Laboratory*
[b] *Ohio State University*
[c] *William and Mary*

E-mail: sabidi@bnl.gov, antonio.boveia@cern.ch, vcavaliere@bnl.gov, denis.furletov@gmail.com, gekow.1@osu.edu, william.kalderon@cern.ch, sjyoo@bnl.gov

Abstract: The determination of charged particle trajectories (tracking) in collisions at the CERN Large Hadron Collider (LHC) is one of the most important aspects for event reconstruction at hadron colliders. This is especially true in the high conditions expected during the future high-luminosity phase of the LHC (HL-LHC) where the number of interactions per beam crossing will increase by a factor of five. Deep learning algorithms have been successfully applied to this task for offline applications. However, their study in hardware-based trigger applications has been limited . In this paper, we study different algorithms for two different steps of tracking and show that such algorithms can be run on field-programmable gate arrays (FPGAs).

Classification

flavour-tagging & event classification

CHRISTOS LEONIDOPOULOS

Event Classification: Boosted Decision Trees

Eur. Phys. J. C 73 (2013) 2677

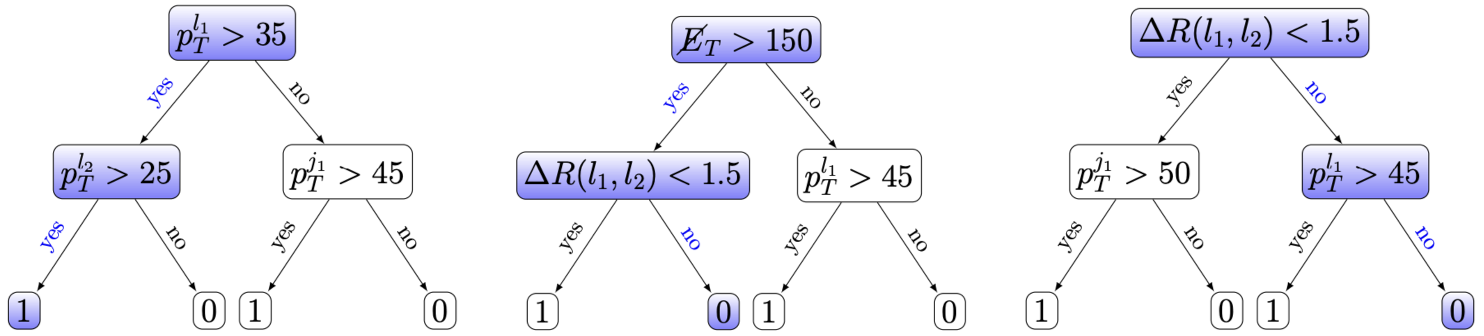


Figure 14: A sample forest of three trees made from the training dataset given in Table. 3. The prediction for the test data (Table. 3) and its direction of flow is shown in blue color.

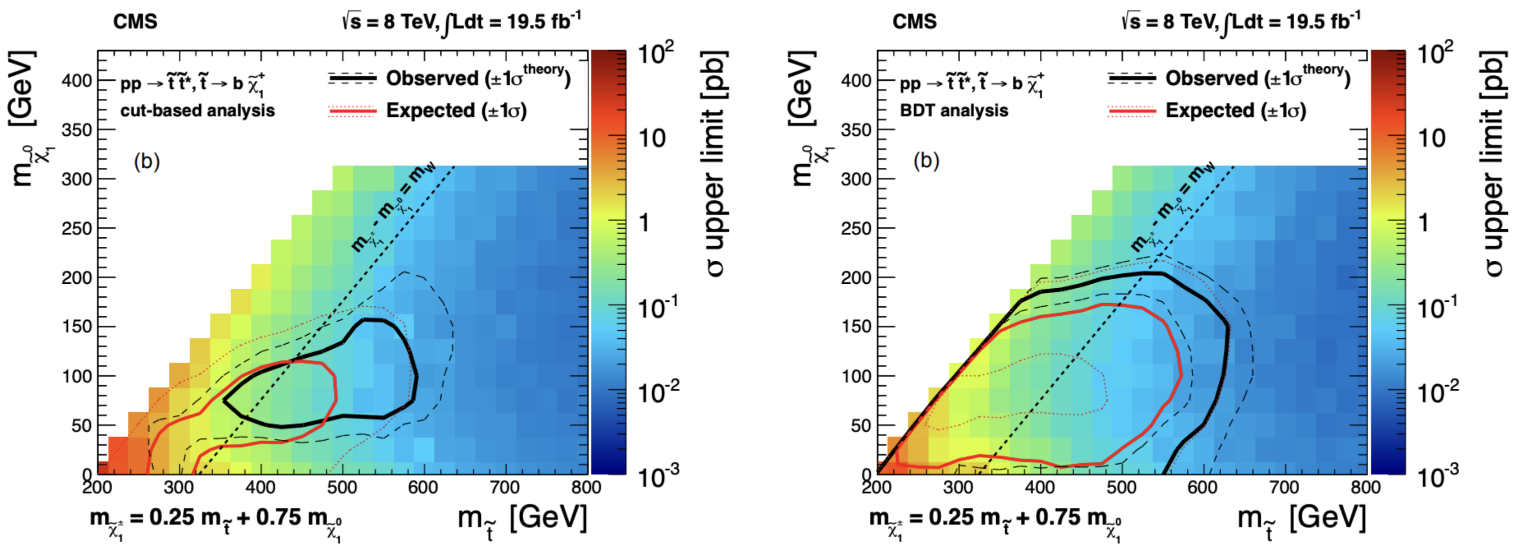


Figure 7: The exclusion plot in $m_{\tilde{t}_1} - m_{\tilde{\chi}^0_1}$ plane for $\tilde{t}_1 \to b\tilde{\chi}^\pm_1$ decay from cut-based (left panel) and BDT method (right panel) respectively [117].

BDT: “a series of if-then-else statements”

…which makes a big difference

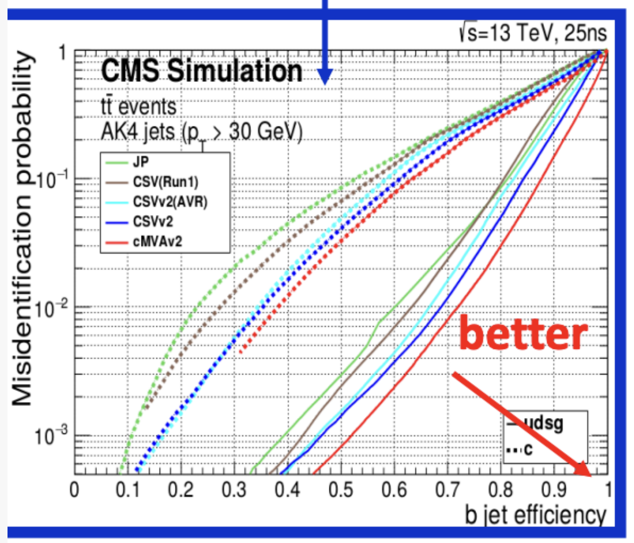
Review of a decade [CMS]

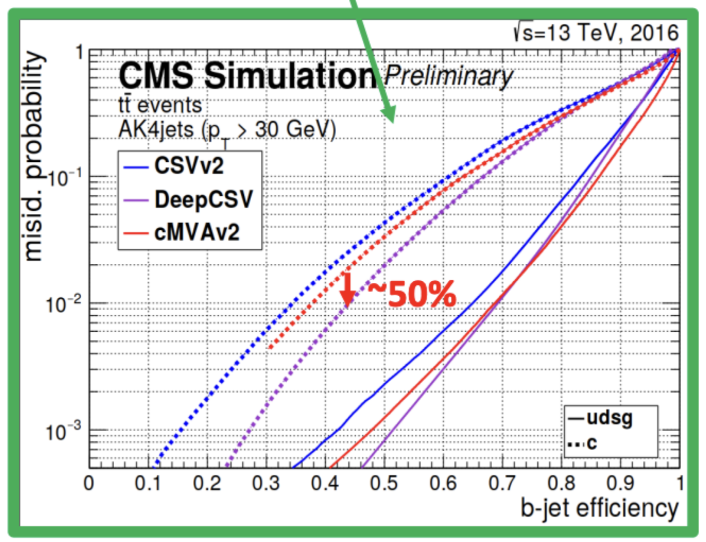
- Enormous progress over the last few years:

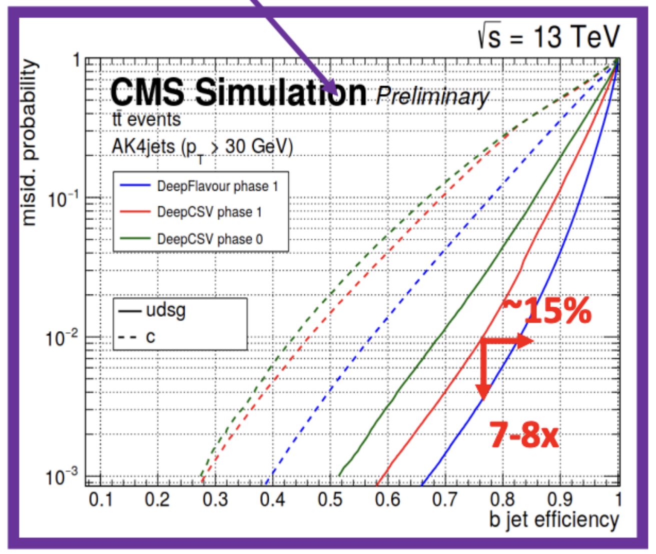
e.g., b-tagging

The early days:	**Early Run 2:**	**The "Game changer":**	**Pushing limits further:**
Human – inspired high-level variables + Cut-based selection	Human – inspired high-level variables + Simple ML	Inputs: low-level info [as particle sequences] + Advanced ML (CNN, RNN)	Inputs: low-level info [as unordered sets] + Graph Neural Nets

Run 1 **Run 2** **Run 3**







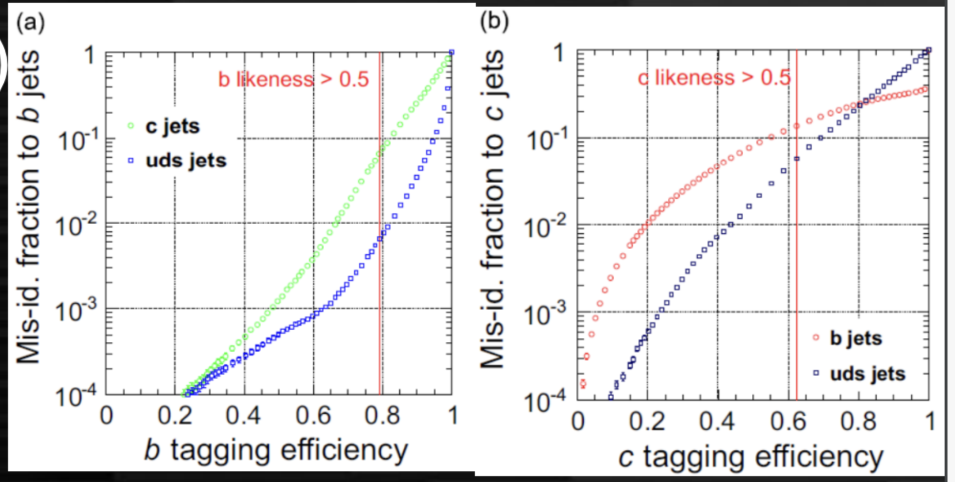
[Disclaimer: Focus on CMS results; similar methods developed by the other LHC experiments]

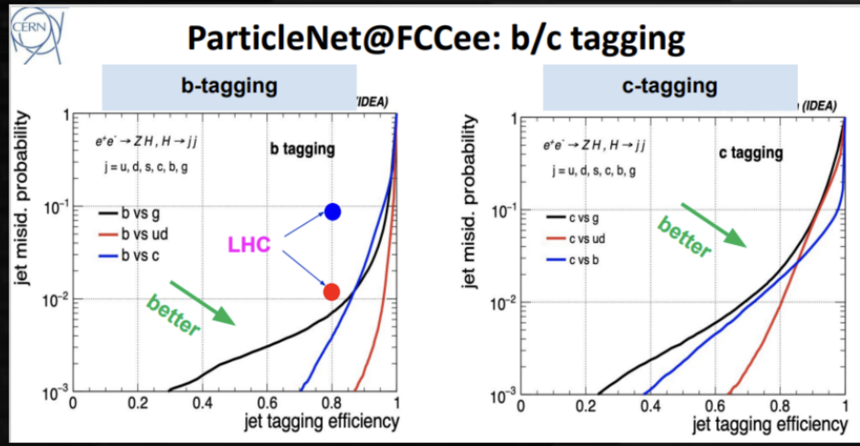
Flavour Tagging: Convolutional/Recurrent NN

Flavor tagging for Higgs factories

- Jet flavor tagging is essentially important for Higgs studies (including self coupling)
- LCFIPlus (published 2013) was long used for flavor tagging
  - All physics performance in ILD/SiD/CLIC are based on LCFIPlus
- FCCee reported >10x better rejection using ParticleNet (GNN) in 2022
  - Delphes is used for simulation
- We studied DNN-based flavor tag with ILD full simulation to confirm it
  - Using latest algorithm: Particle Transformer (ParT)

LCFIPlus performance plots





Taikan Suehara et al., 3rd ECFA WS on Higgs/Top/Electroweak Factories, 9 Oct. 2024, page 2

credit

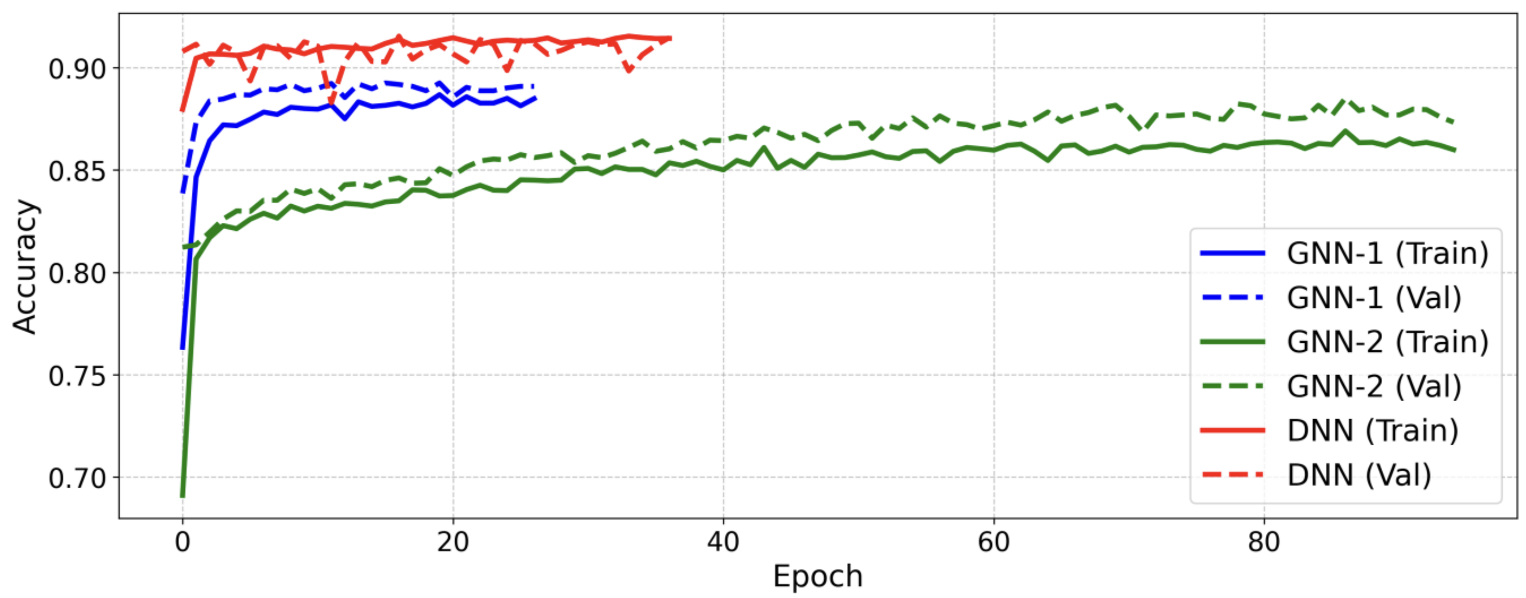
x10 performance improvement in 9-10 years

CHRISTOS LEONIDOPOULOS

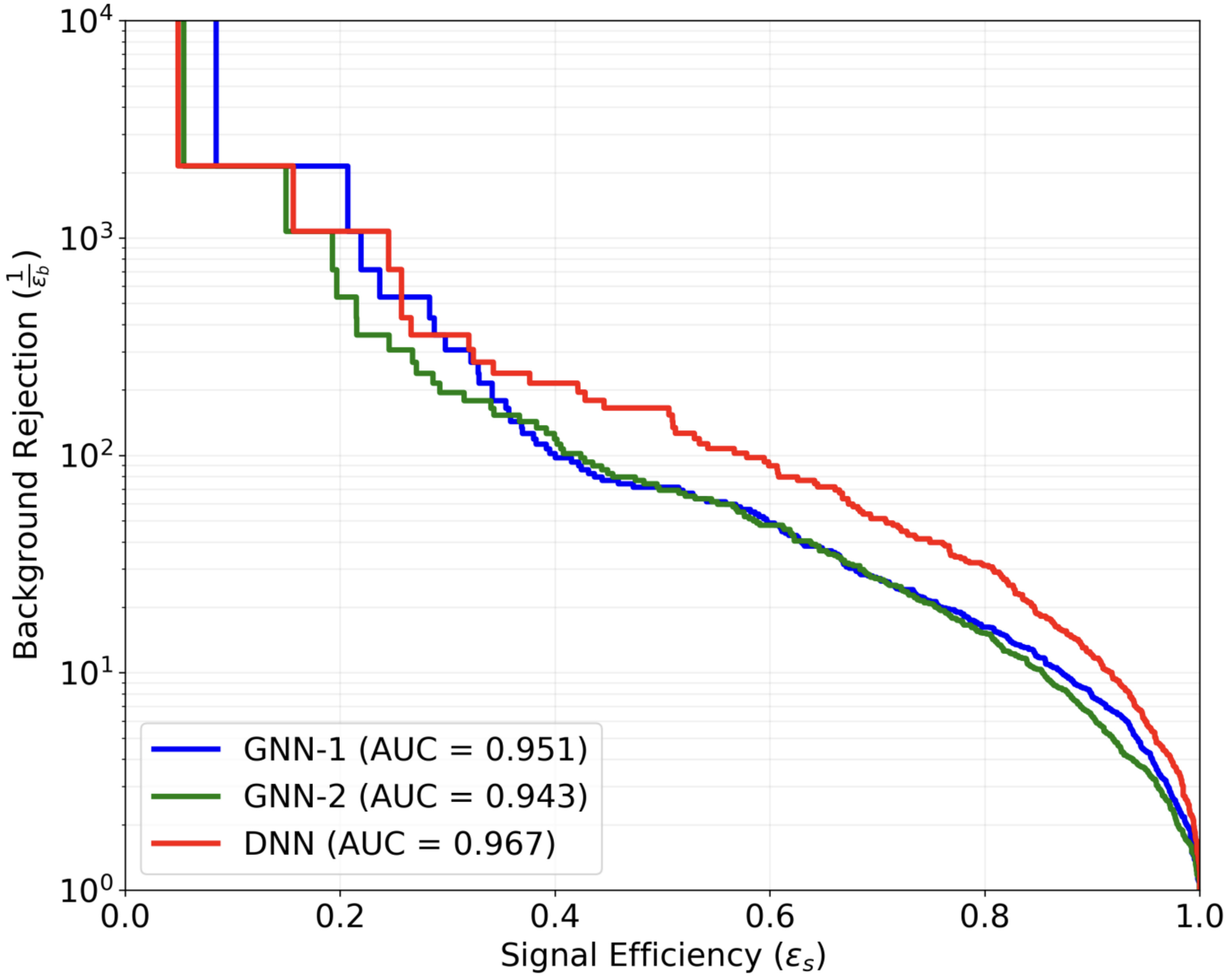
Flavour Tagging: Graphical Neural Networks

EPJ C 82 646 (2022)

...the quest for the best algorithm continues



Event classification for BSM searches



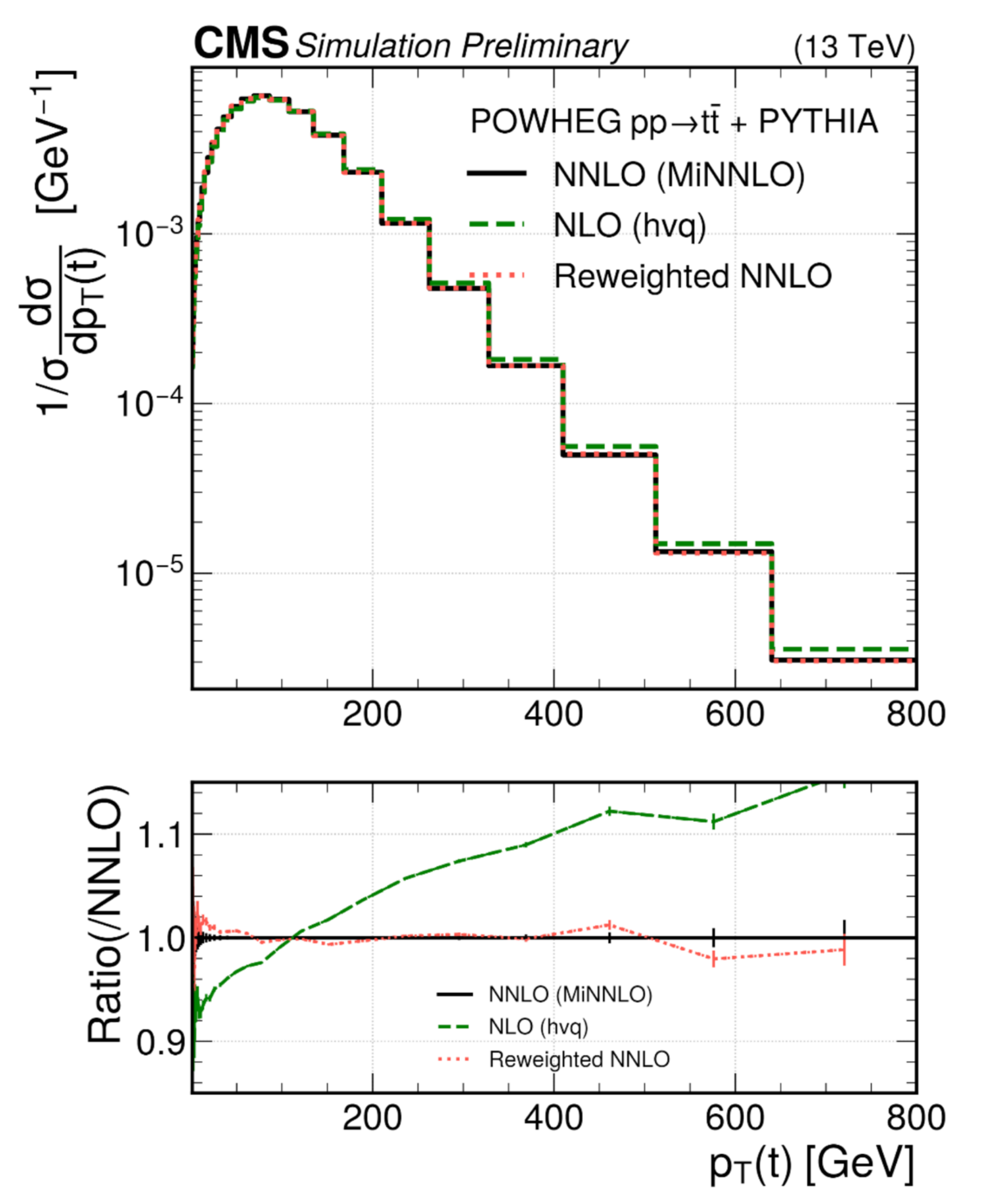
CHRISTOS LEONIDOPOULOS

Event Classification: GNN vs DNN

arXiv:2411.06487

Simulation

Using DNN for tuning and reweighting simulated samples



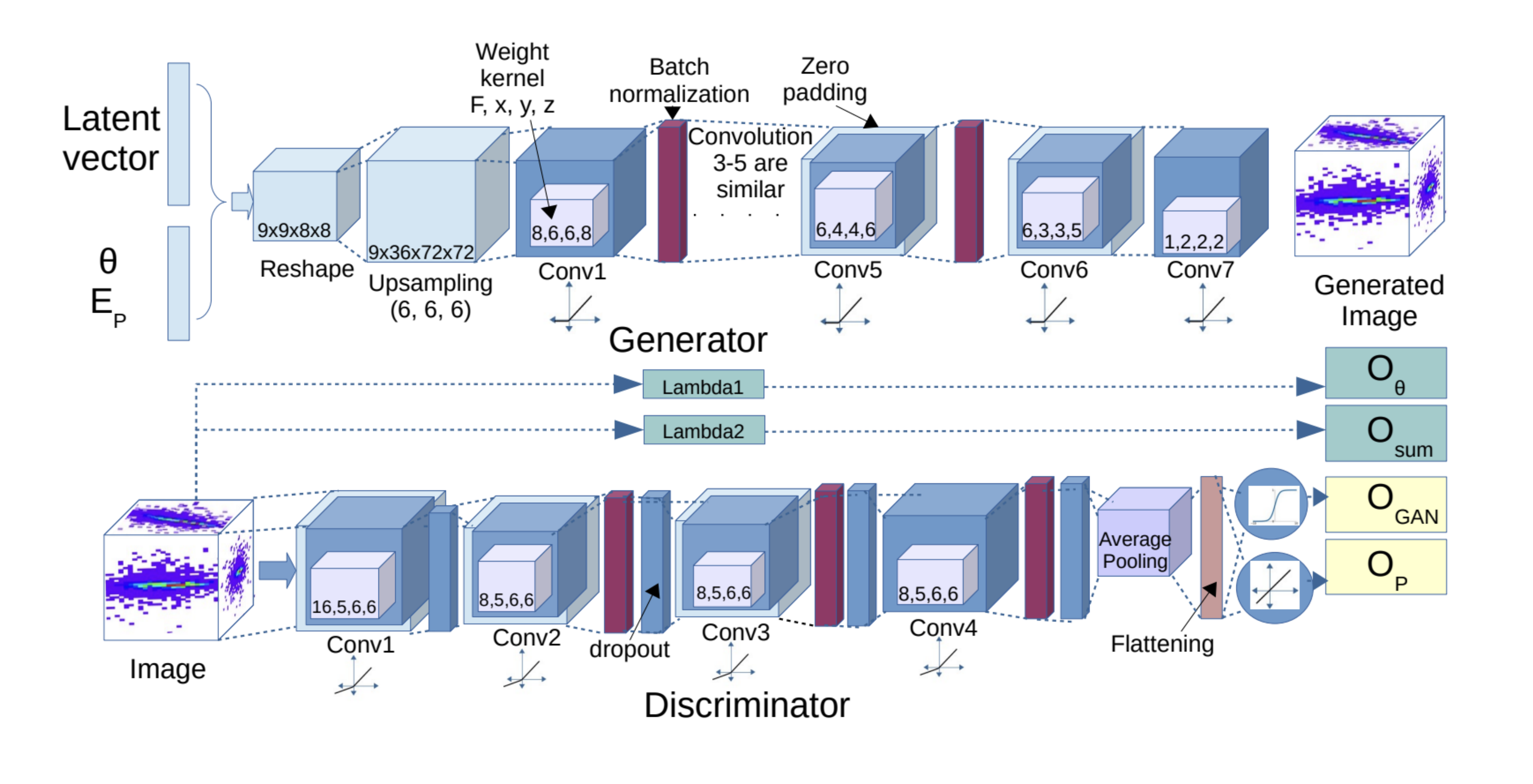
- One large sample of simulated data is produced with detailed detector simulation (very CPU expensive) using a given physics model (e.g. NLO)
- Additional, smaller samples are produced based on different models (e.g. NNLO) in order to train the ML algorithm w/o detailed detector simulation and event reconstruction
- ML computes weights for each simulated event, applies them on large simulated sample
- Update existing simulation using new theoretical progress (or a systematic effect) fast & reliably

credit

Simulation: Deep Neural Networks

CHRISTOS LEONIDOPOULOS

GAN: emulate interaction of particles with high-granularity calorimeter



- "Fake" event created by GAN, compare with simulated event
- Discriminator tries to guess if event is real or not
- Discriminator guessing correctly pushes GAN to adjust generation strategy (e.g. sampling of features) to improve "deception"
- Repeat till discriminator can no longer tell difference

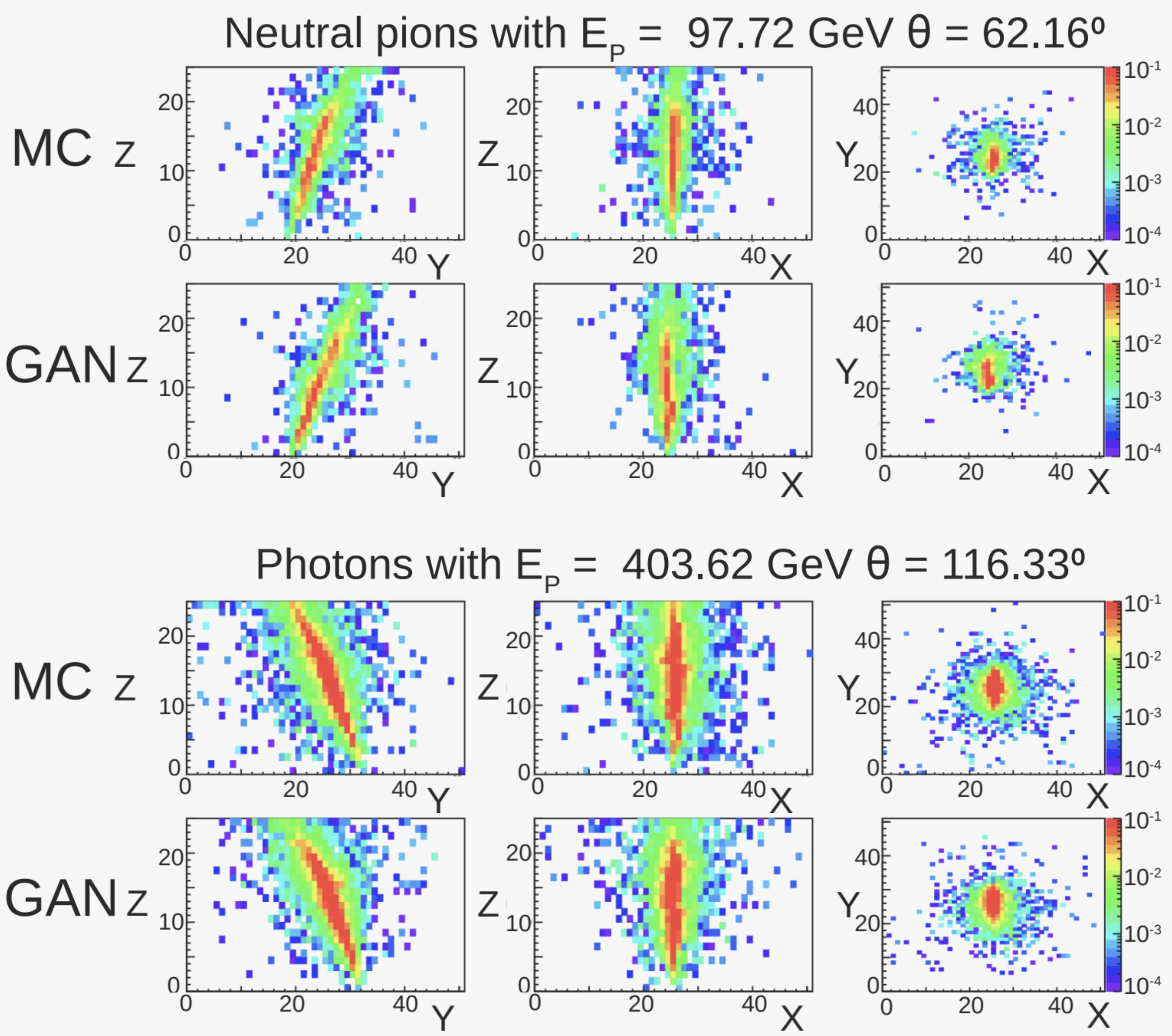
Simulation: Generative Adversarial Networks

arXiv:2109.07388

Can you guess which of these images are synthetic and which ones are real?

Simulation: Generative Adversarial Networks

GAN: emulate interaction of particles with high-granularity calorimeter



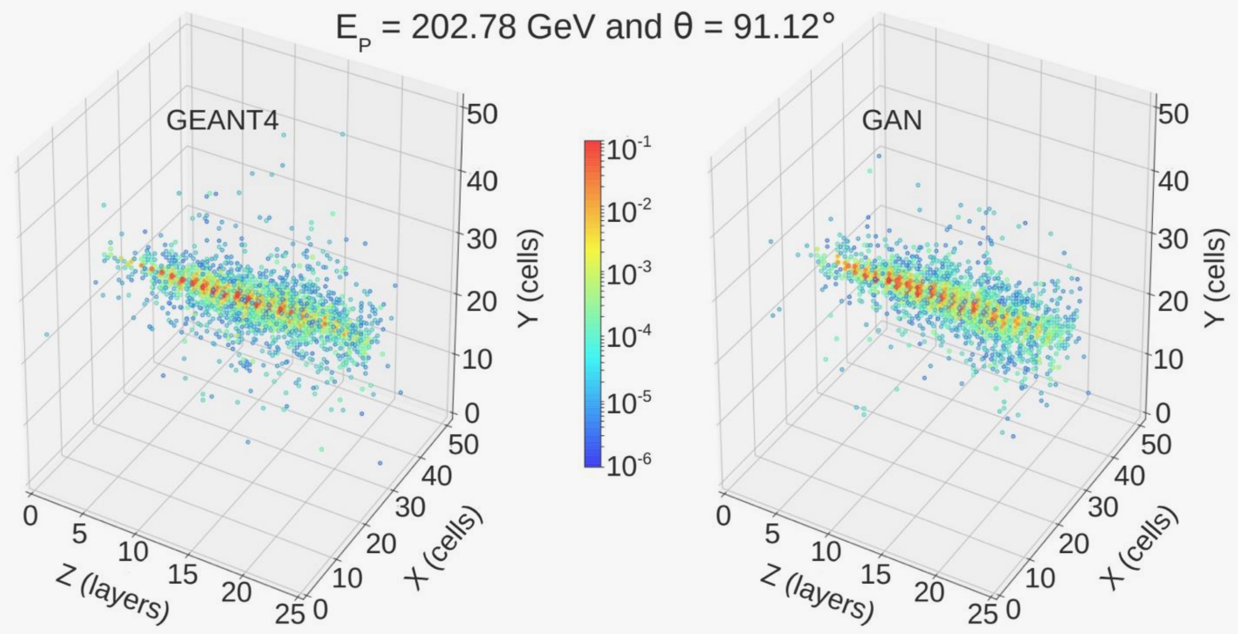


Figure 6. GEANT4 vs. GAN electrons showers with $E_p = 202.78$ GeV and $\theta = 91.12°$.

Showers generated by GAN present accuracy within 10% of Monte Carlo for a diverse range of physics features, with three orders of magnitude speedup

Simulation: Generative Adversarial Networks

arXiv:2109.07388

CHRISTOS LEONIDOPOULOS

End-to-end simulation of collision events

End-to-end simulation of particle physics events with Flow Matching and generator Oversampling

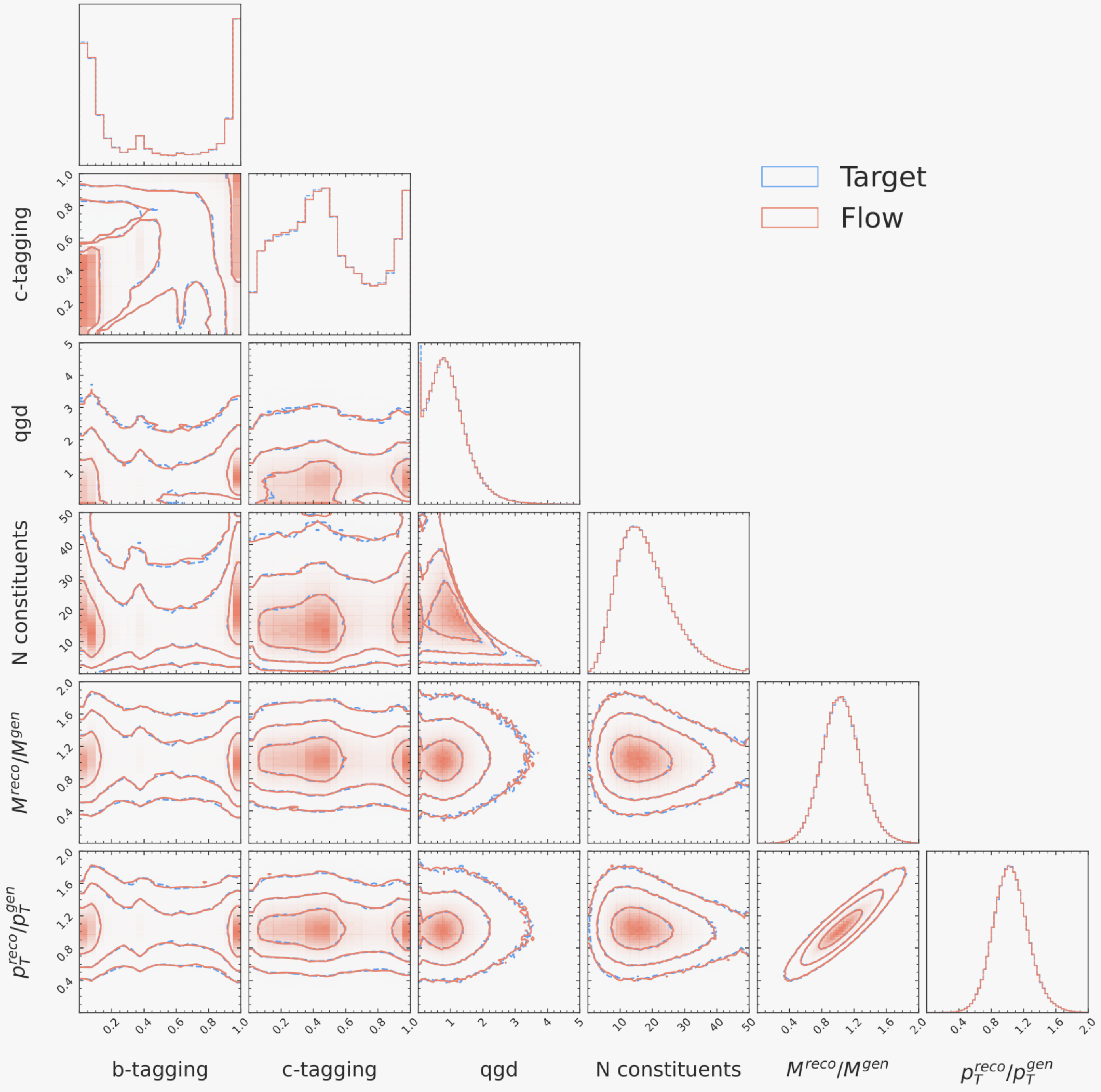
F Vaselli † [1,2] , F Cattafesta [1,2], P Asenov [2,3] and A Rizzi [2,3]

[1] Scuola Normale Superiore, Pisa
[2] Istituto Nazionale di Fisica Nucleare, Pisa
[3] Università di Pisa

E-mail: francesco.vaselli@cern.ch, filippo.cattafesta@cern.ch, patrick.asenov.asenov@cern.ch, andrea.rizzi@cern.ch

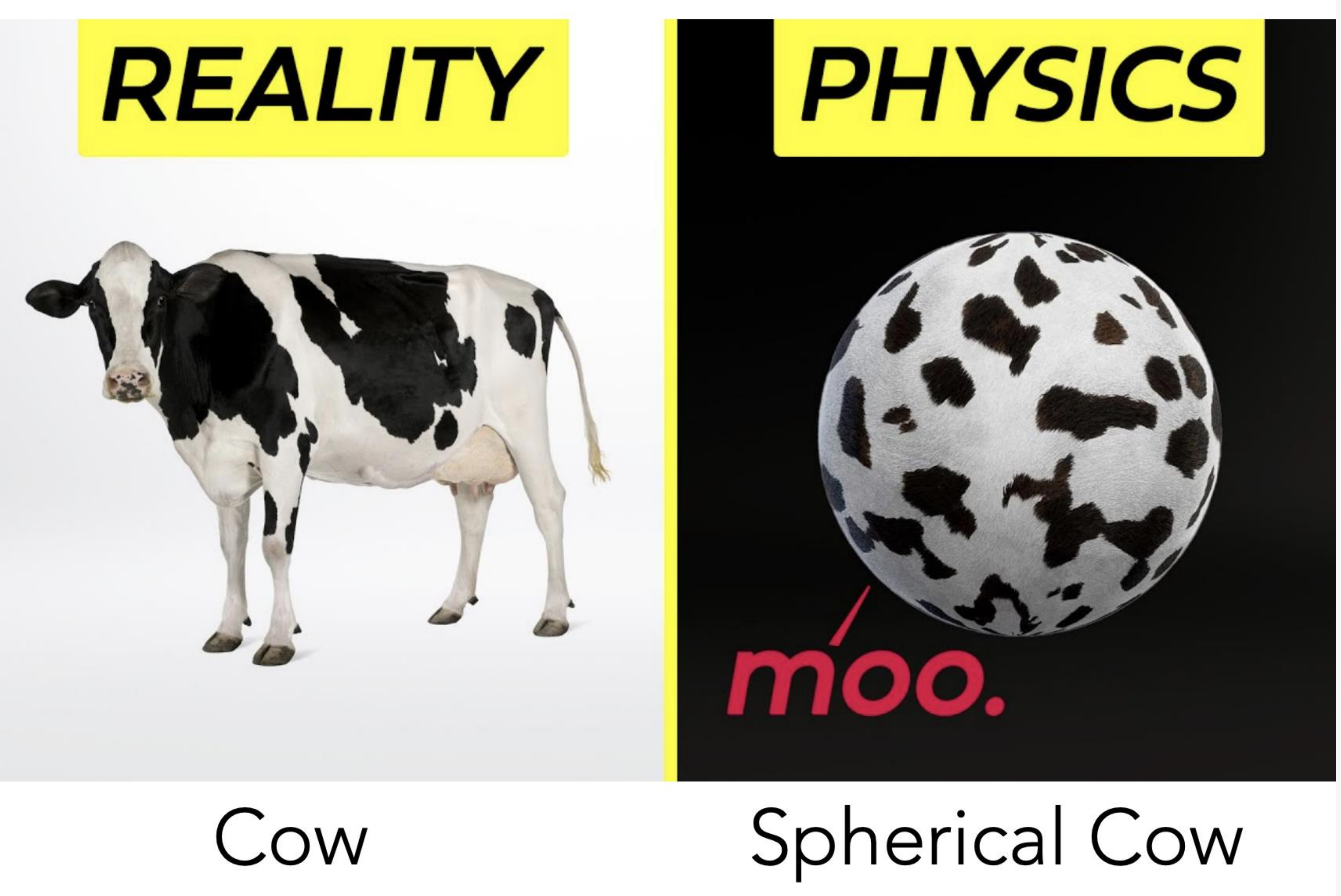
February 2024

Abstract. The simulation of high-energy physics collision events is a key element for data analysis at present and future particle accelerators. The comparison of simulation predictions to data allows looking for rare deviations that can be due to new phenomena not previously observed. We show that novel machine learning algorithms, specifically Normalizing Flows and Flow Matching, can be used to replicate accurate simulations from traditional approaches with several orders of magnitude of speed-up. The classical simulation chain starts from a physics process of interest, computes energy deposits of particles and electronics response, and finally employs the same reconstruction algorithms used for data. Eventually, the data are reduced to some high-level analysis format. Instead, we propose an end-to-end approach, simulating the final data format directly from physical generator inputs, skipping any intermediate steps. We use particle jets simulation as a benchmark for comparing both *discrete* and *continuous* Normalizing Flows models. The models are validated across a variety of metrics to identify the most accurate. We discuss the scaling of performance with the increase in training data, as well as the generalization power of these models on physical processes different from the training one. We investigate sampling multiple times from the same physical generator inputs, a procedure we name *oversampling*, and we show that it can effectively reduce the statistical uncertainties of a dataset. This class of ML algorithms is found to be capable of learning the expected detector response independently of the physical input process. Their speed and accuracy, coupled with the stability of the training procedure, make them a compelling tool for the needs of current and future experiments.

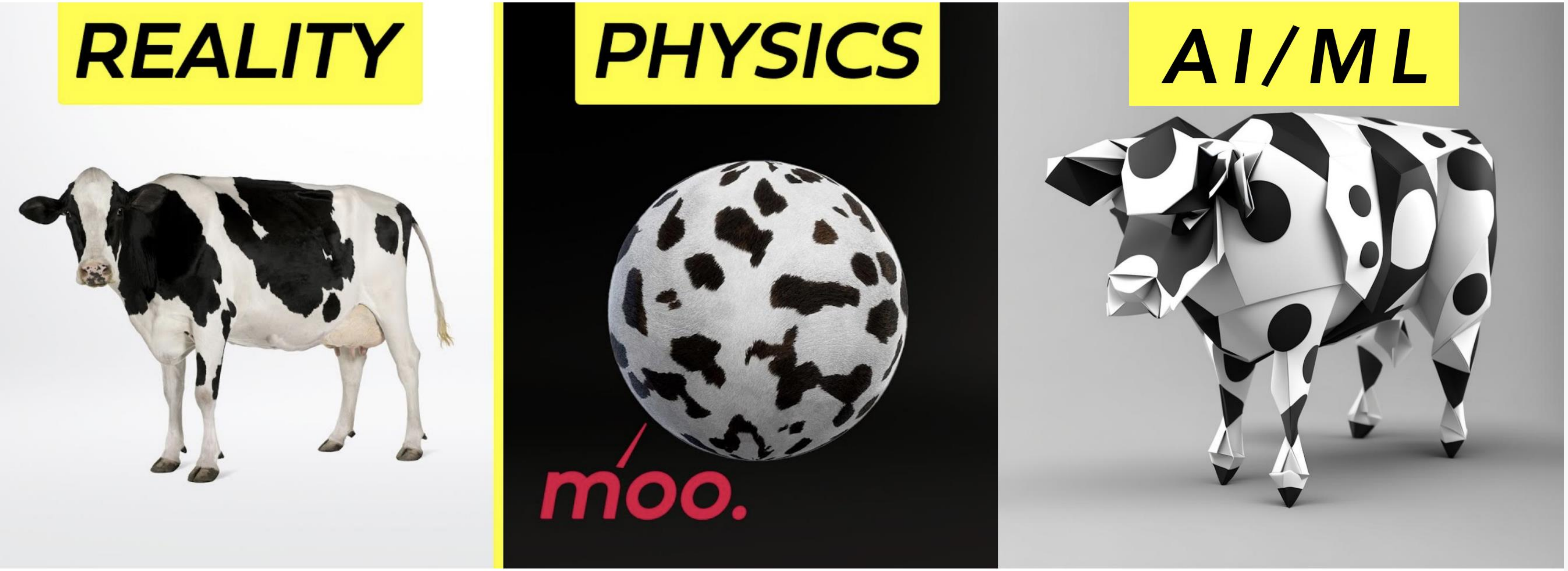


Simulation: Normalizing Flows

arXiv:2402.13684



Simulation with ML



Cow

Spherical Cow

ReLU Cow

Simulation with ML

Anomaly Detection

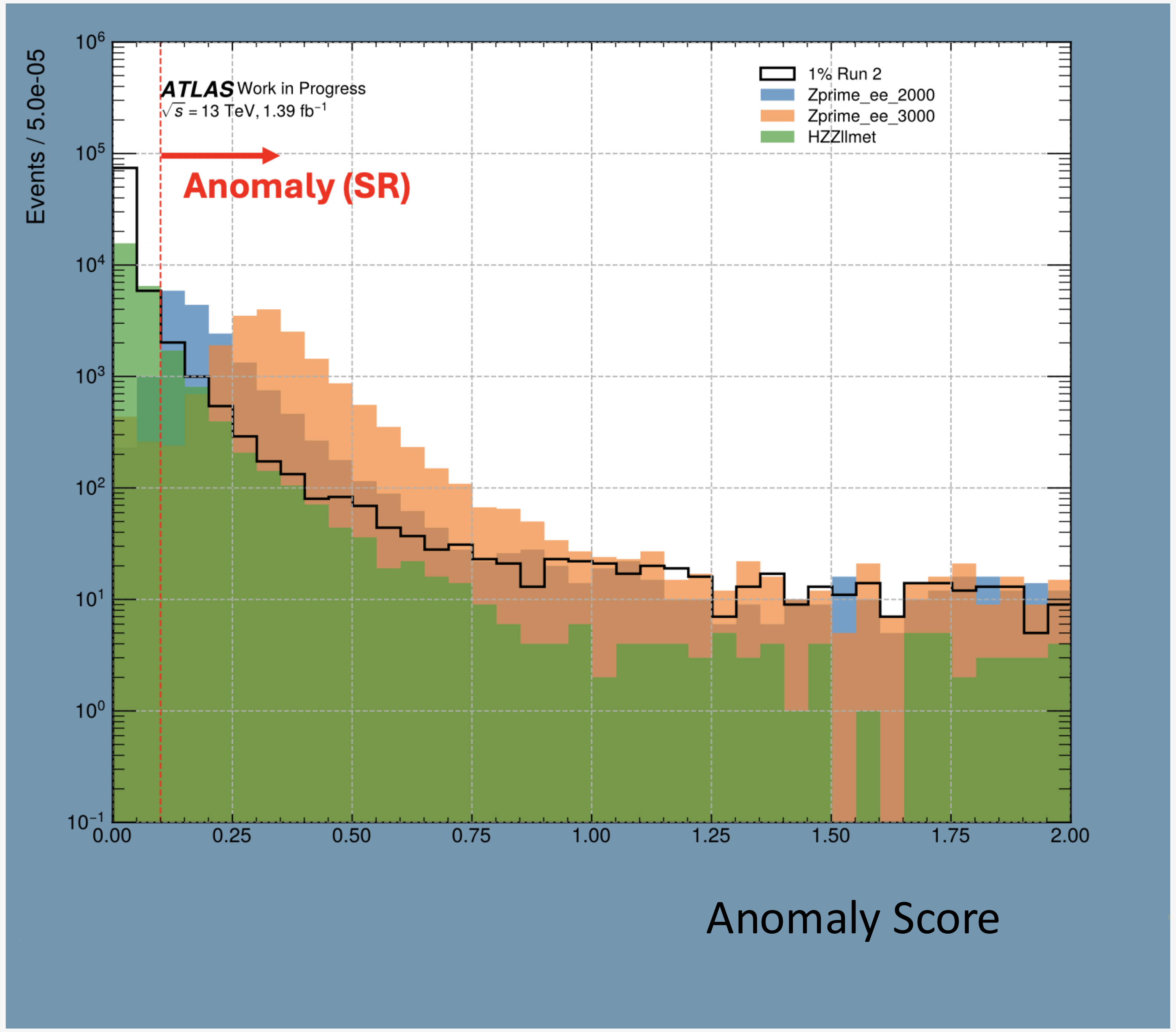
Supervised learning: dog or cat?

Unsupervised learning: dogs only!

Exploit vast datasets to search for any strange behaviour or unusual signs that could point to New Physics without relying on specific theoretical models

Anomaly Detection

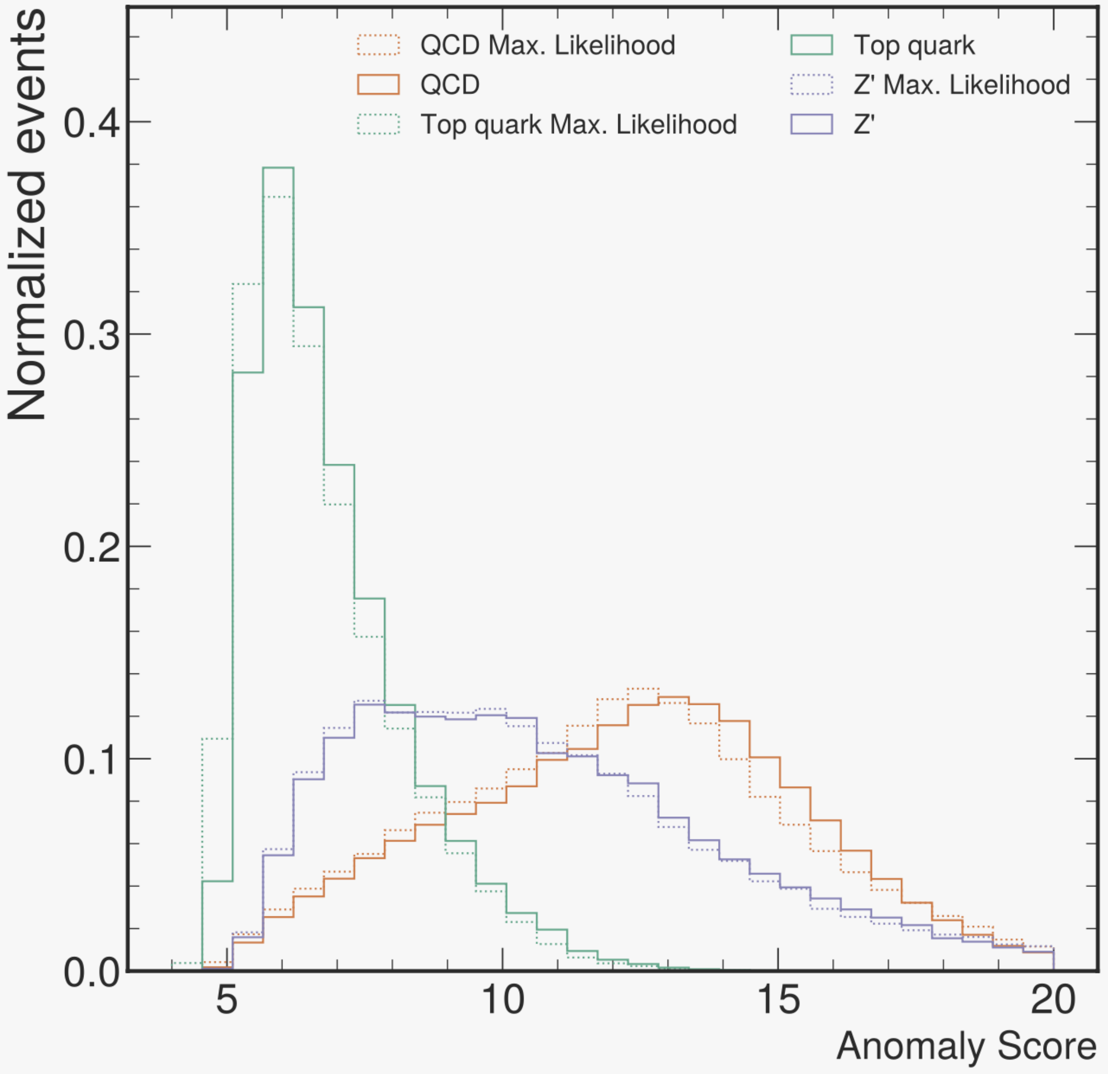
CHRISTOS LEONIDOPOULOS



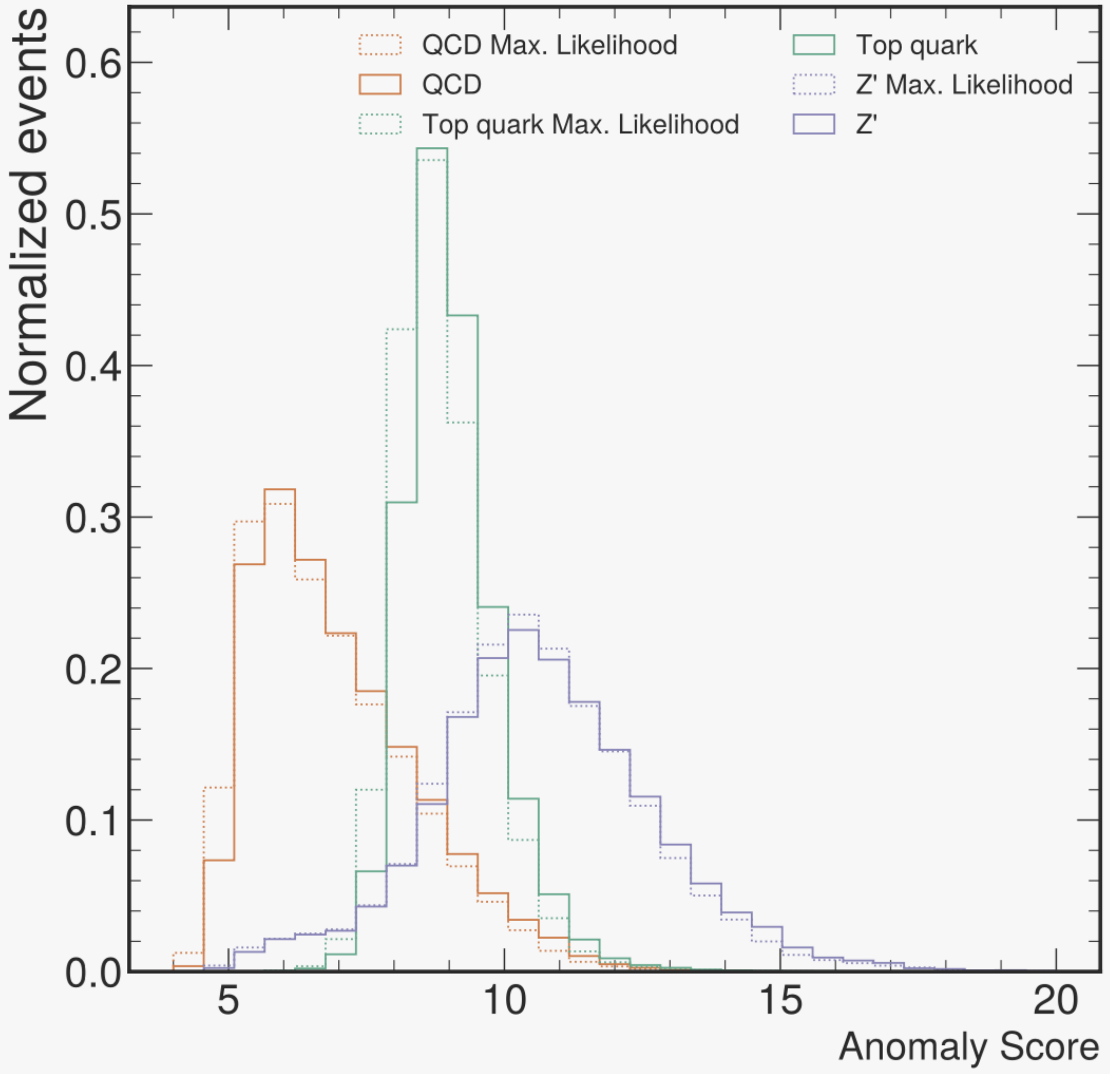
AD: BSM searches in model-agnostic way

Anomaly Detection: Auto-Encoders

Flavour-tagging with Anomaly Detection



Trained on top quarks



Trained on QCD

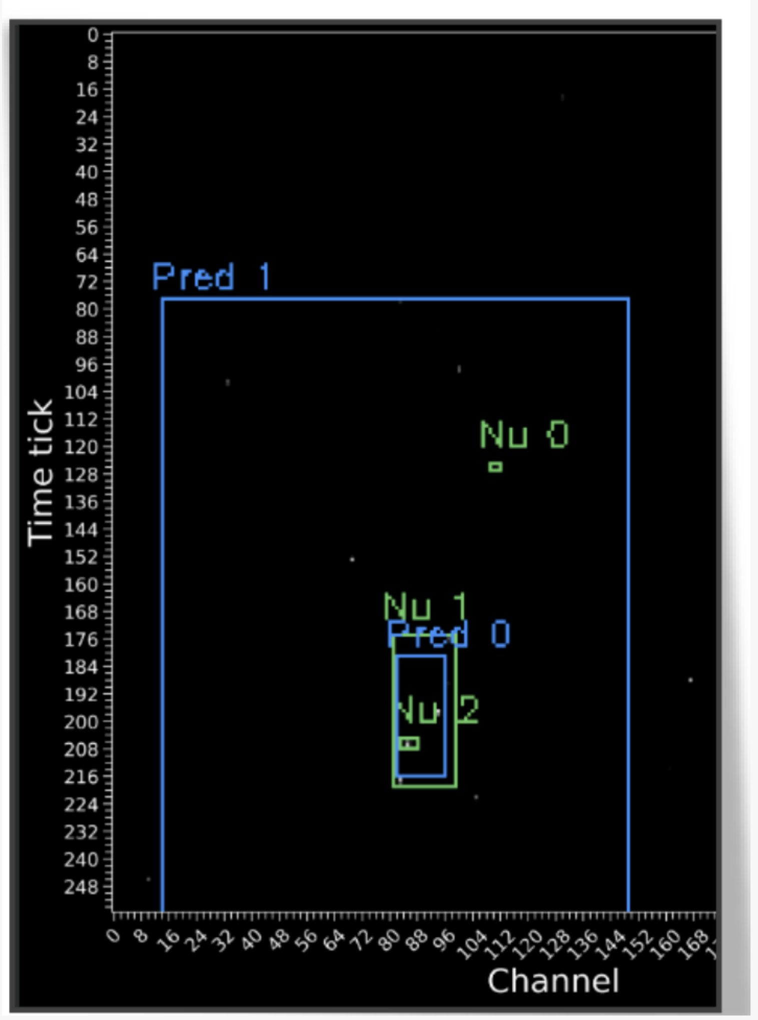
Anomaly Detection: Deep Sets

SciPost Phys. 16, 062 (2024)

CHRISTOS LEONIDOPOULOS

Real-time processing

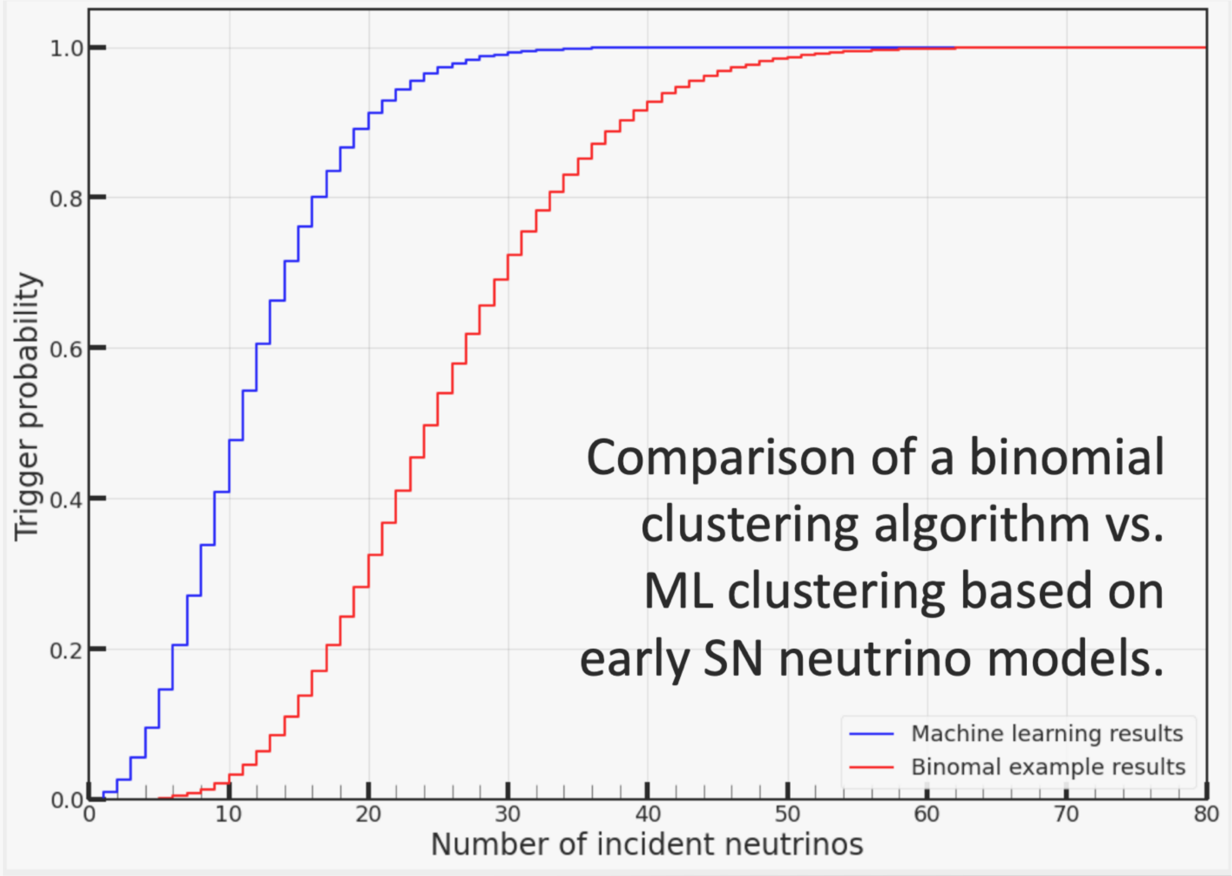
DUNE triggering on neutrinos from Galactic Supernova burst



Standard clustering algorithms
- Binomial clustering efficiencies prob = 33%
- Background rate: 0.14 Hz

ML clustering creates “boxes” with neutrino prediction
- Average multiplicity: 1.2
- Background rate: 0.45 Hz

credit



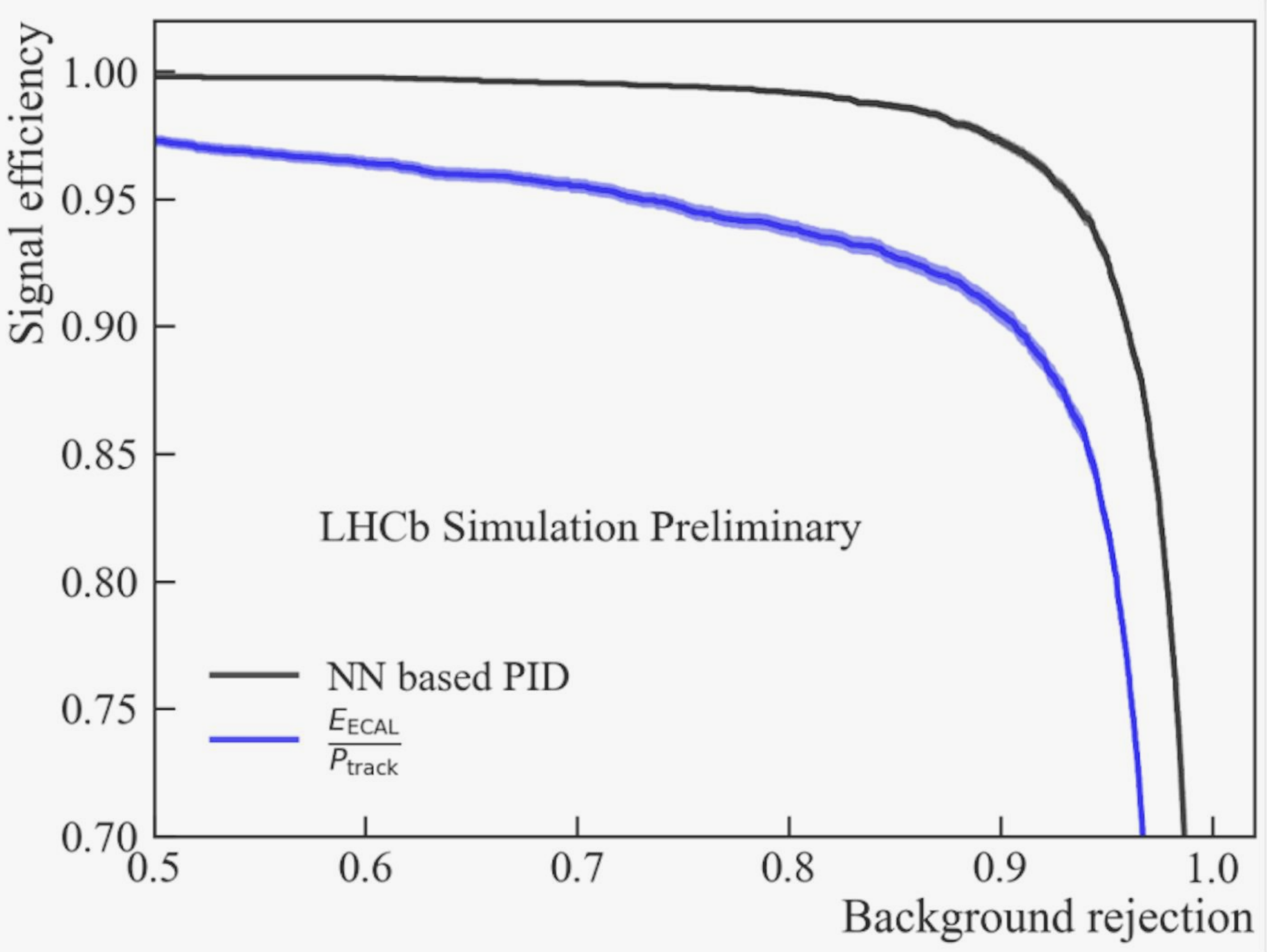
Comparison of a binomial clustering algorithm vs. ML clustering based on early SN neutrino models.

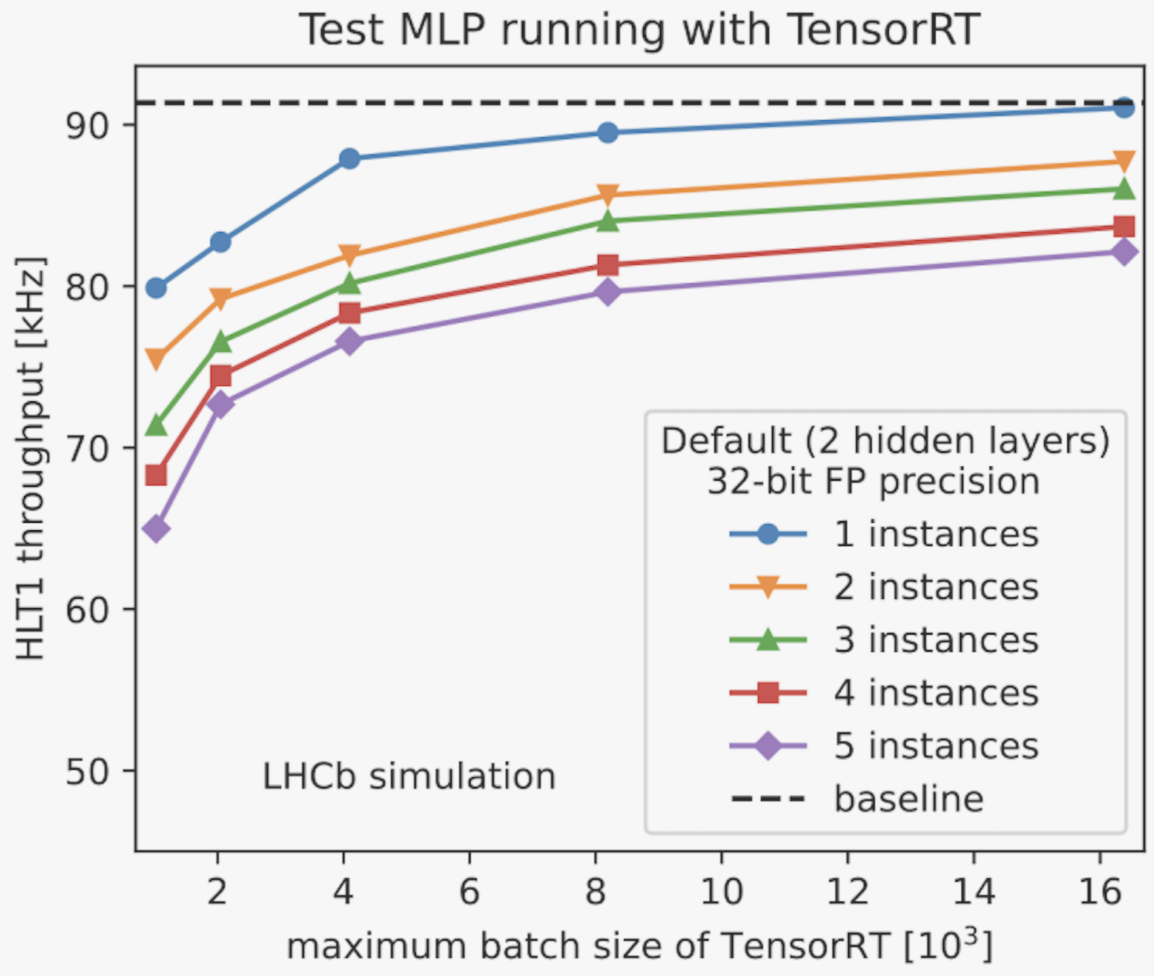
CHRISTOS LEONIDOPOULOS

Triggering on Galactic Supernova neutrinos

ML on LHCb trigger

LHCb using ML at trigger since 2015





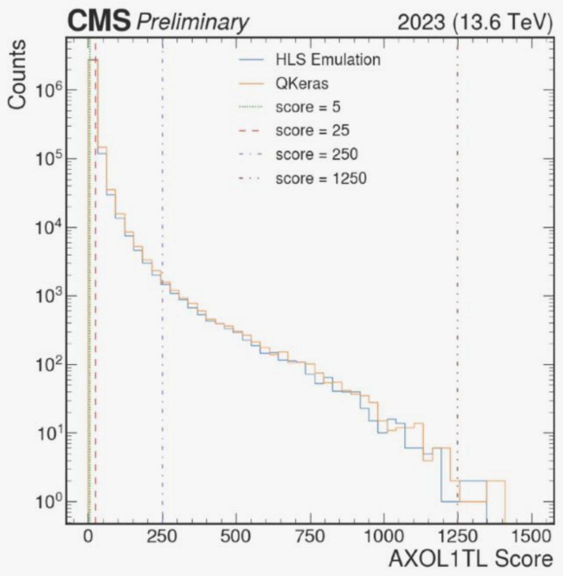
credit

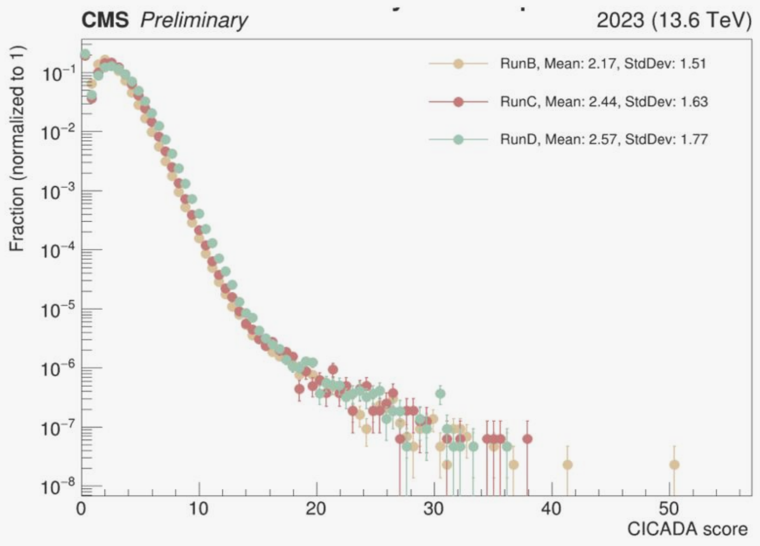
Extensive studies and usage of Lipschitz networks
- tracking and fake-track rejection
- electron ID
- Search for Long-Lived Particle signatures (arXiv:2112.00038)

ML usage extends beyond reconstruction
- Flexible and standardised pipelines for ML model serving backends
- Too tedious (and error-prone) to do manually

CHRISTOS LEONIDOPOULOS

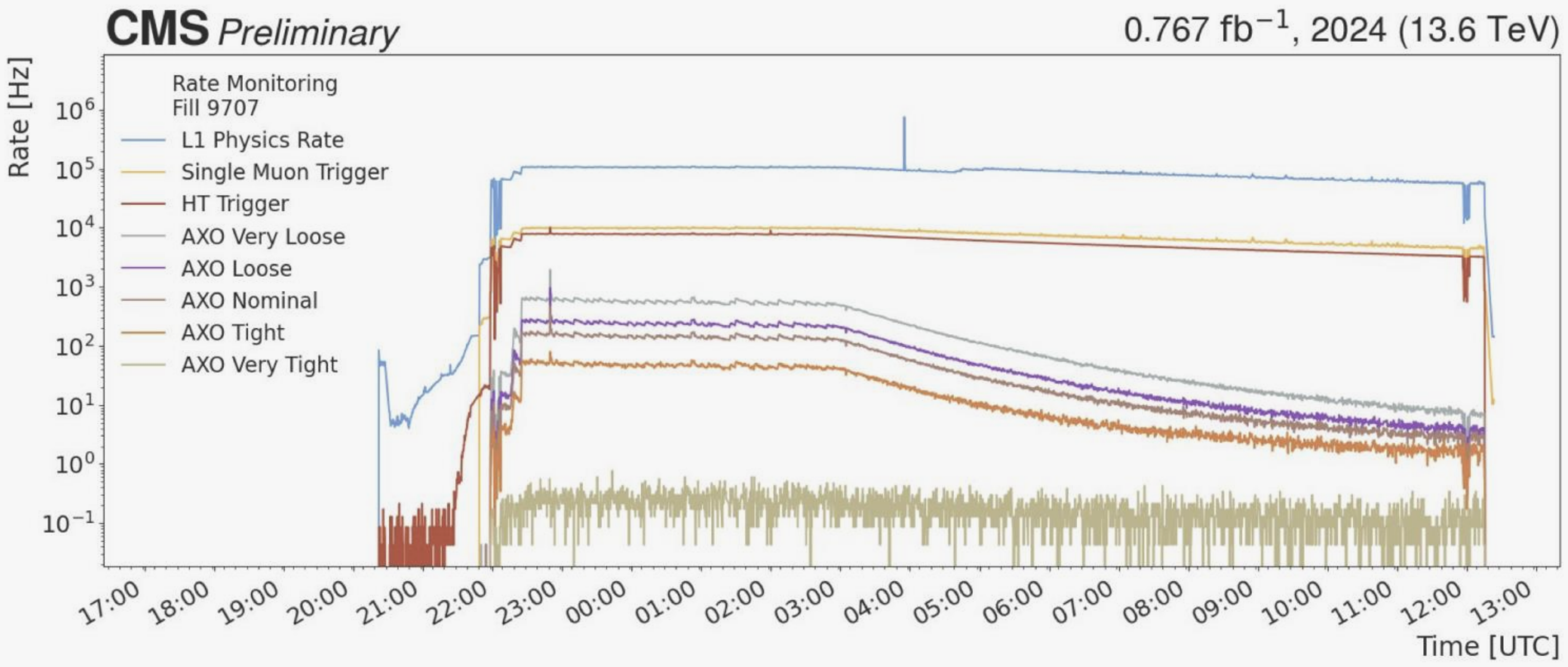
FPGA-based AD with two algorithms





2.1 Inputs to L1 anomaly detection

AXOL1TL and CICADA use different L1 reconstructions as inputs. AXOL1TL takes in 10 jets, 4 electron/photon objects, 4 muons, and transverse missing energy (MET) as reconstructed in the L1 trigger from calorimeter and muon triggers. The 3-momenta (p_T, η, ϕ) of these objects, in raw hardware integer values, are used. CICADA, by contrast, uses calorimeter region energies, which form an 18x14 image-like input. The input streams for both algorithms are shown in Fig. 1.



Anomaly Detection for CMS L1 trigger

arXiv:2411.19506

CHRISTOS LEONIDOPOULOS
THE UNIVERSITY OF EDINBURGH

Challenges

“The holy grail for almost any smart product is for it to be deployable anywhere, and require no maintenance like docking or battery replacement.”

- **Large volumes of data: petabytes of data (HL-LHC soon, FCC later?)**
- **Access to HPC resources**
- **Real-time processing: large rates, small latency**
- **Scalability: ML/AI solutions must follow evolution of operational conditions, detector upgrades, new challenges of future experiments**

Computational Demands

CHRISTOS LEONIDOPOULOS
THE UNIVERSITY OF EDINBURGH

Training data limitations:

- Supervised algorithms need labelled data (time-consuming process requiring human input: tedious & error-prone)
- Biased datasets or small statistics that introduce large errors when generalising

Data Quality

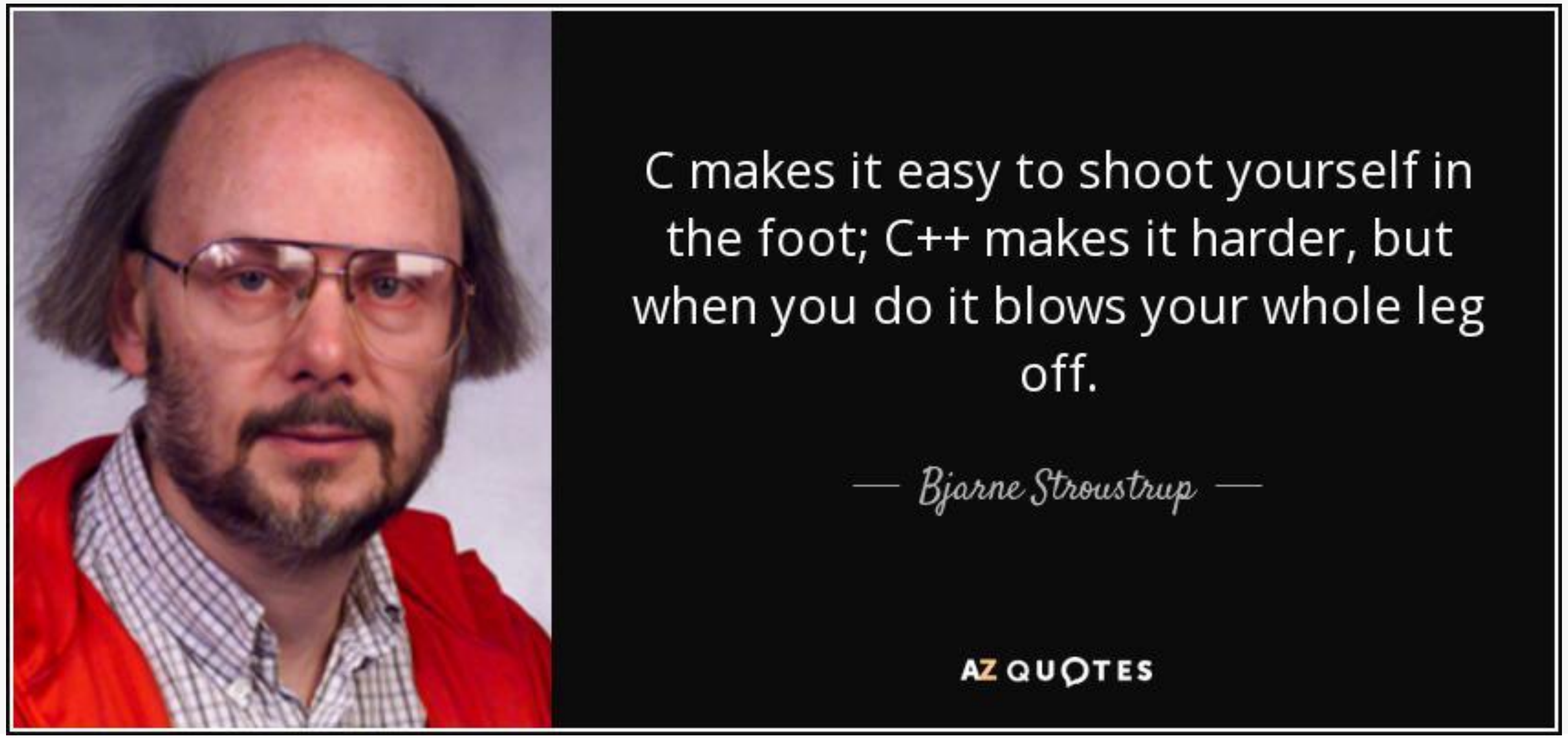
- Data comes from different sources and in different formats; many solutions being developed, but approach rather fragmented
- Lacking end-to-end physics solutions (good for job security?)

Heterogeneity

CHRISTOS LEONIDOPOULOS

- Junior people enthusiastic about adoption of new technologies
- Senior people hesitant about joining the frenzy

"Bottleneck for big advances can be either technical or sociological"



Sociological

What lies ahead

- Quantum Computing: Quantum computers offer potentially exponential speed-ups for certain types of calculations. Quantum machine learning could leverage these capabilities with quantum algorithms, enhancing speed and efficiency for CPU-intensive tasks (eg. simulation, pattern recognition, anomaly detection with massive datasets)

CHRISTOS LEONIDOPOULOS

Quantum Machine Learning

- Automated ML technologies automate parts of the ML pipeline, such as model selection and hyperparameter tuning
- Adaptive Algorithms: Algorithms that can dynamically adjust to new data or experimental conditions
- Exotic searches using automated Anomaly Detection?

Automated Machine Learning

CHRISTOS LEONIDOPOULOS

- Southern Wide-field Gamma-ray Observatory: a collaboration with an array of 6,000 particle detectors in the highlands of South America to study gamma rays from outer space.
- Collaboration assessing how to arrange and place these 6,000 detectors.
  “We have an enormous number of possible solutions; how do we pick the best one?”

- Experiments at LHC (or future colliders): hundreds of individual (enormously complicated) sub-detectors that work together to track and measure particles. Optimizing each one’s design—not as an individual component but as a part of a complex ecosystem—is nearly impossible. “the human brain is incapable of thinking in 1,000 dimensions”

- Particle Flow algorithms: a wholistic approach in event reconstruction
- Consider physics goals, budget, and real-world limitations to choose the optimal detector design

CHRISTOS LEONIDOPOULOS

Design of new experiments

arXiv:2310.01857

CHRISTOS LEONIDOPOULOS

Machine Learning models that mimic the complex functions of the human brain

- Transformers: “attention mechanism”, aka: learning relationships / context (LLM & ChatGPT)
- Originally designed for natural language processing tasks, transformers can have an impact in various other domains, including physics, due to their ability to handle sequential data effectively (sequence prediction problems)

CHRISTOS LEONIDOPOULOS

Advancements in Neural Networks

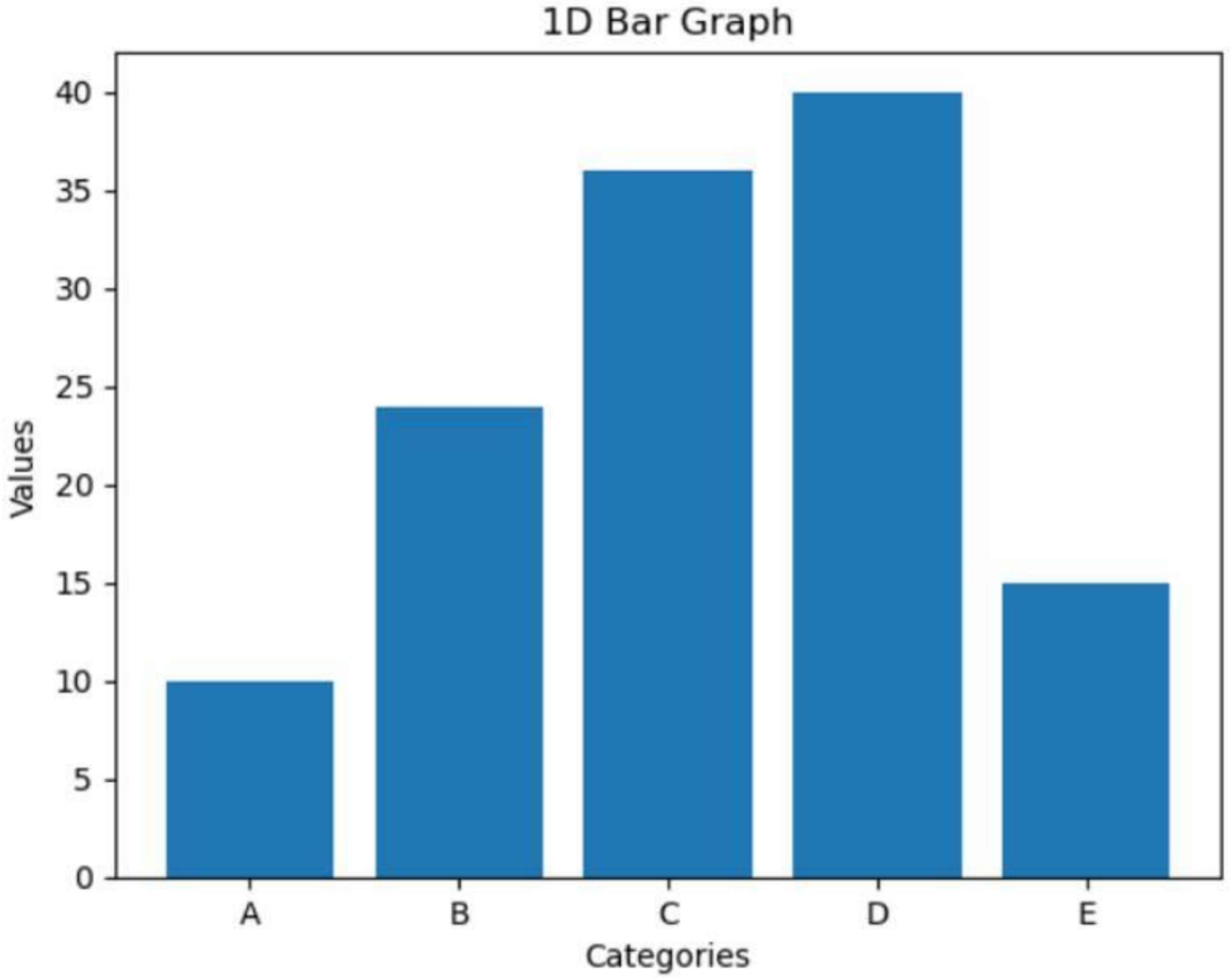
"Give me some example code for plotting a 1D graph"

cleonido

Create an example python program using the matplotlib library that creates a 1D bar graph

GitHub Copilot

```
import matplotlib.pyplot as plt

# Data
x = ['A', 'B', 'C', 'D', 'E']
y = [10, 24, 36, 40, 15]

# Create bar graph
plt.bar(x, y)

# Add title and labels
plt.title('1D Bar Graph')
plt.xlabel('Categories')
plt.ylabel('Values')

# Show graph
plt.show()
```



CHRISTOS LEONIDOPOULOS

LLM writing code

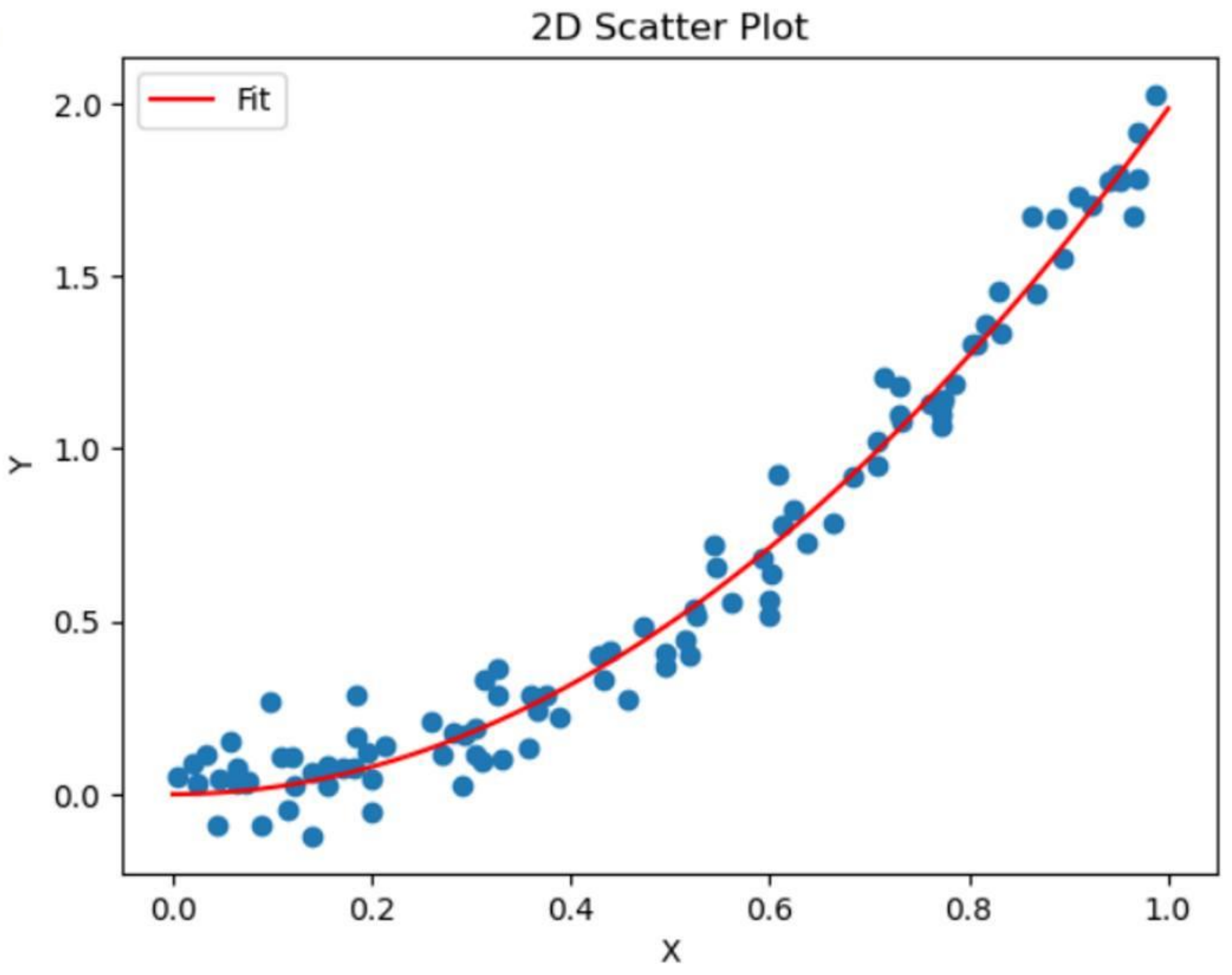
“Give me some example code for a scatter plot with 100 points, and a fit through the points that determines the parameter of the non-linear relationship”

cleonido

Create a 2D scatter plot using the relationship y=a*x^2 with a 100 random points and calculate the value of a by doing a fit

GitHub Copilot

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.optimize import curve_fit

# Define function
def func(x, a):
    return a * x**2

# Generate random data
np.random.seed(42)
x = np.random.rand(100)
y = 2 * x**2 + 0.1 * np.random.randn(100)

# Fit data to function
popt, pcov = curve_fit(func, x, y)

# Create scatter plot
plt.scatter(x, y)

# Add title and labels
plt.title('2D Scatter Plot')
plt.xlabel('X')
plt.ylabel('Y')

# Add fitted curve
xfit = np.linspace(0, 1, 100)
yfit = func(xfit, *popt)
plt.plot(xfit, yfit, 'r-', label='Fit')

# Show graph
plt.legend()
plt.show()

# Print value of a
print('a =', popt[0])
```




```
a = 1.9853600095527293
```

LLM writing code #2

Given the sequence 1, 2, 4, 7, 11, 16, what is the next term?

2 approaches:

- Numeric regression: direct prediction of the next term (like a fit through data points)
- Symbolic regression: finding a formula for the sequence
  - a closed formula: $u_n = \frac{n(n+1)}{2} + 1$
  - or a recurrence relation: $u_n = u_{n-1} + n$

d'Ascoli, Kamienny, Lample, Charton 2022

credit

Recurrent sequences: Deep symbolic regression

Constant	Approximation	Rel. error
0.3333	$(3+\exp(-6))^{-1}$	10^{-5}
0.33333	$1/3$	10^{-5}
3.1415	$2\arctan(\exp(10))$	10^{-7}
3.14159	π	10^{-7}
1.6449	$1/\arctan(\exp(4))$	10^{-7}
1.64493	$\pi^2/6$	10^{-7}
0.123456789	$10/9^2$	10^{-9}
0.987654321	$1-(1/9)^2$	10^{-11}

Expression u_n	Approximation $\hat{u}_n$
$\operatorname{arcsinh}(n)$	$\log(n+\sqrt{n^2+1})$
$\operatorname{arccosh}(n)$	$\log(n+\sqrt{n^2-1})$
$\operatorname{arctanh}(1/n)$	$\frac{1}{2}\log(1+2/n)$
$\operatorname{catalan}(n)$	$u_{n-1}(4-6/n)$
$\operatorname{dawson}(n)$	$\frac{n}{2n^2-u_{n-1}-1}$
$\mathrm{j0}(n)$ (Bessel)	$\frac{\sin(n)+\cos(n)}{\sqrt{\pi n}}$
$\mathrm{i0}(n)$ (mod. Bessel)	$\frac{e^n}{\sqrt{2\pi n}}$

credit

Recurrent sequences: Deep symbolic regression

CHRISTOS LEONIDOPOULOS

Scattering amplitudes: complex functions predicting the outcome of particle interactions
- Computed by summing Feynman diagrams of increasing complexity
- Loops: virtual particles created and destroyed in the process
- A hard problem: each loop introduces two latent variables, their integration give rise to generalized polylogarithms
- Calculated from symbols: homogeneous polynomials, degree 2L (L=loop), with integer coefficients (some of which are zero)

Cai, Merz, Nolte, Wilhelm, Cranmer, Dixon, Charton, 2023

L	number of terms
1	6
2	12
3	636
4	11,208
5	263,880
6	4,916,466
7	92,954,568
8	1,671,656,292

TABLE II. Number of terms in the symbol of $F_3^{(L)}$ as a function of the loop order L.

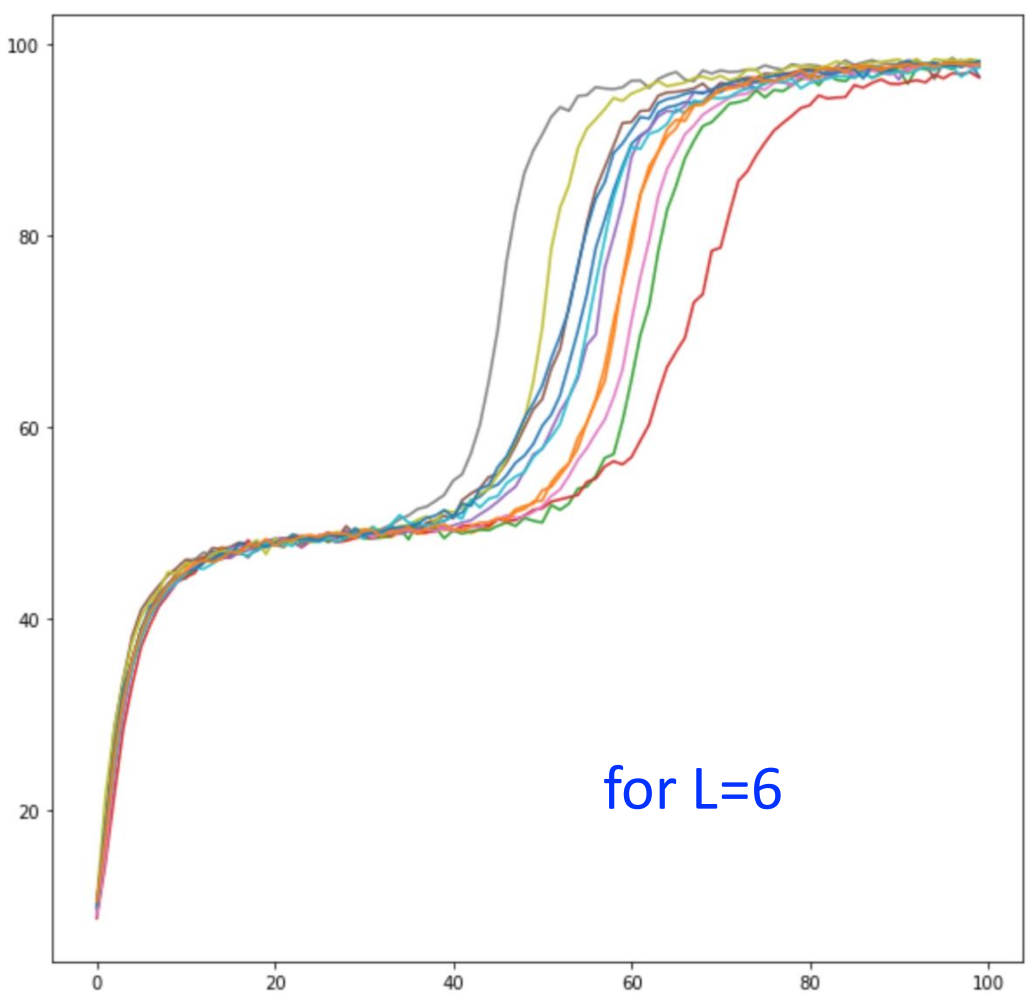
credit

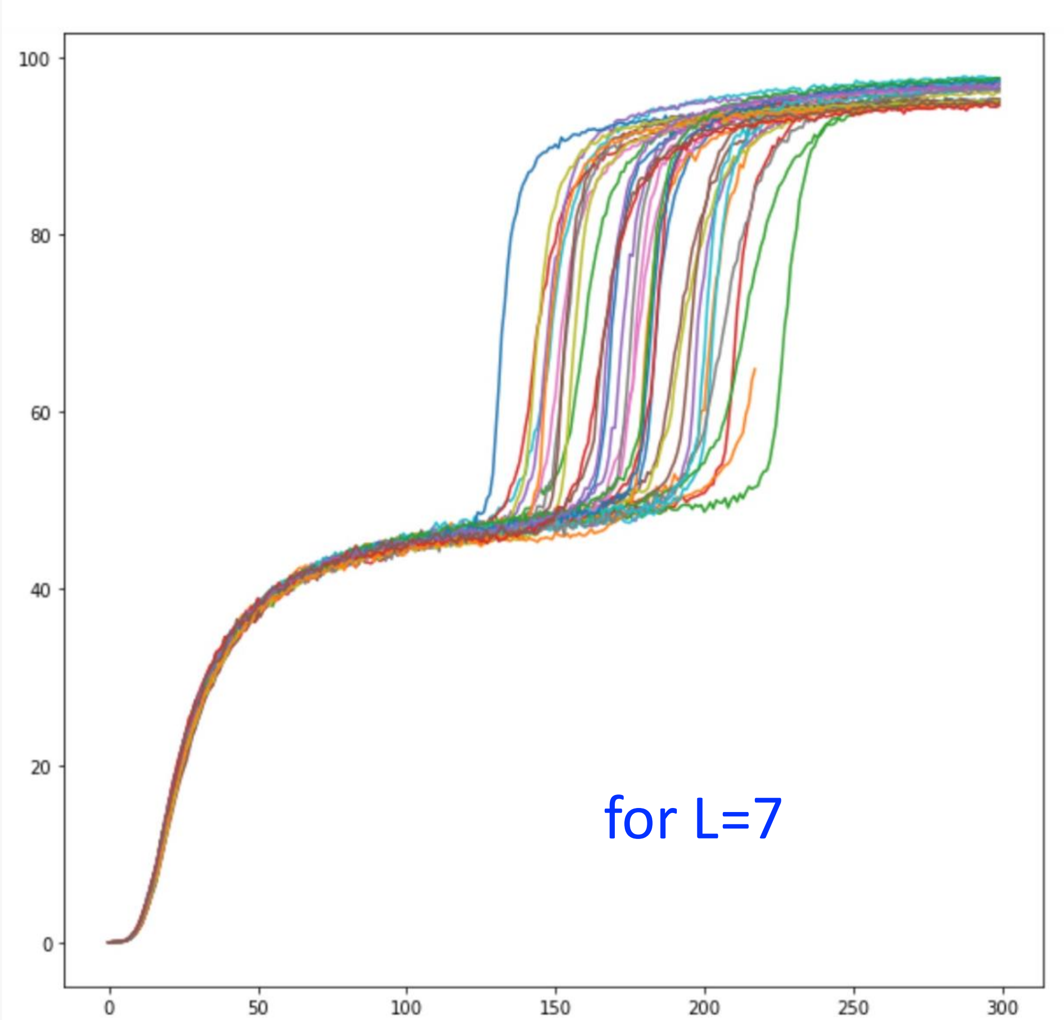
Recurrent sequences: gluon scattering amplitudes

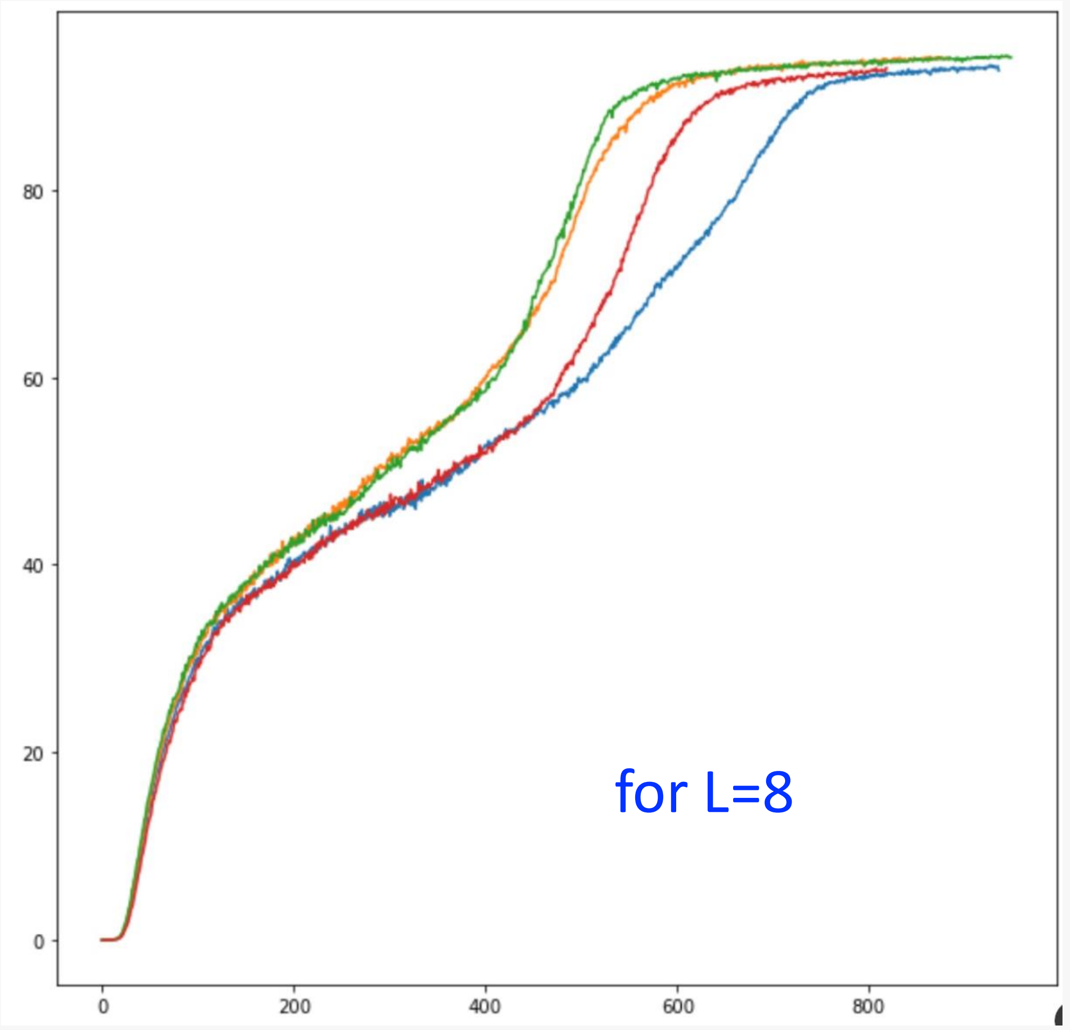
CHRISTOS LEONIDOPOULOS

Prediction accuracy of non-zero coefficients

two-step shape?







[credit]

Recurrent sequences: gluon scattering amplitudes

CHRISTOS LEONIDOPOULOS

Predict L=6 from L=5 coefficients:

- 98.4% sign accuracy
- 99.6% magnitude accuracy

Conclusions:

- There is a “function” behind this (we just don't know what it is!)
- Difficulty of learning the sign: To my theorist friends, you are not alone!

AI generating new hypotheses for us to test?

credit

Recurrent sequences: gluon scattering amplitudes

Summary

- Powerful new technologies & methodologies
- But also new ways to produce new ideas
- We are not in Kansas any more

Epilogue

- The danger is not in AI taking over the world
- The danger is in not engaging, and letting somebody else use AI to take over the world

Epilogue