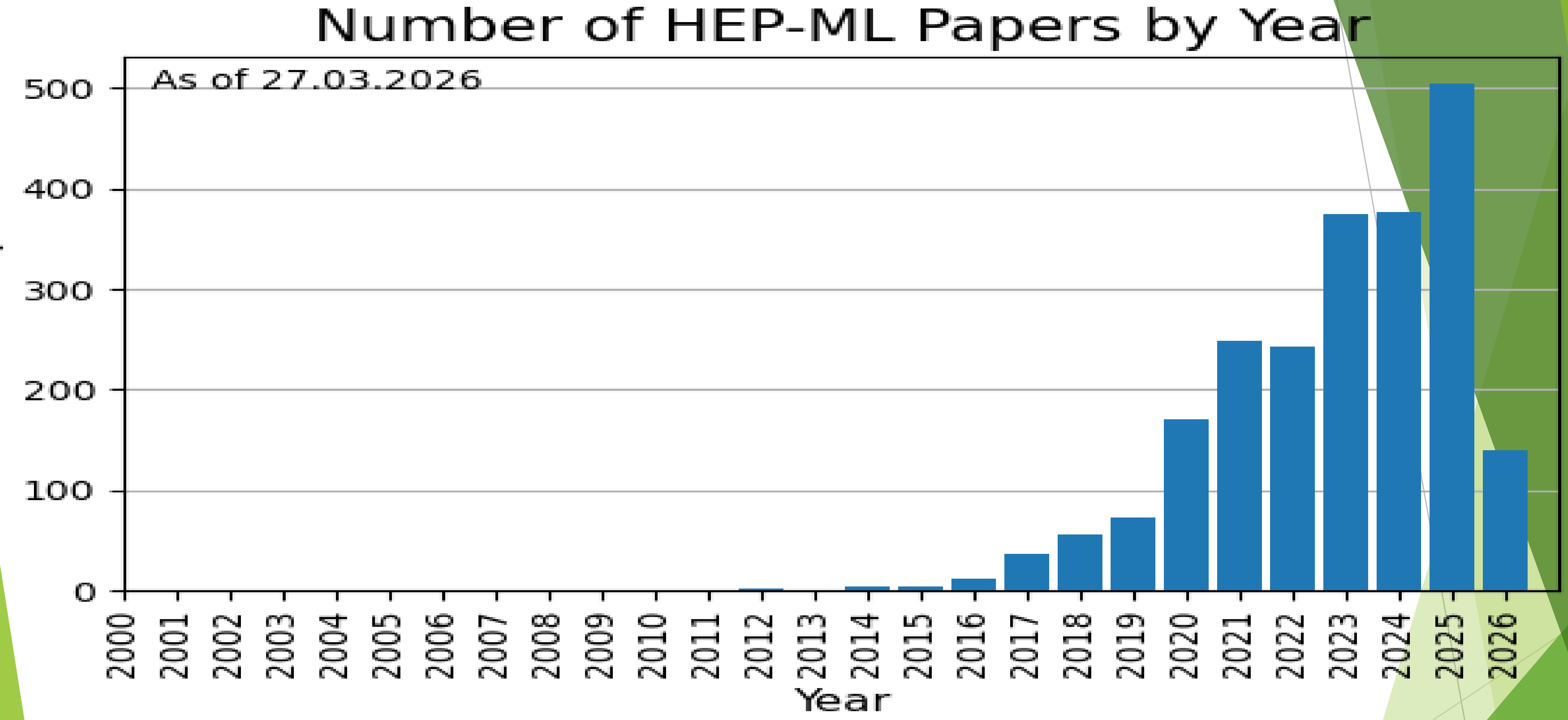
Machine Learning in High Energy Physics

Myagkov A.

NRC KI - IHEP

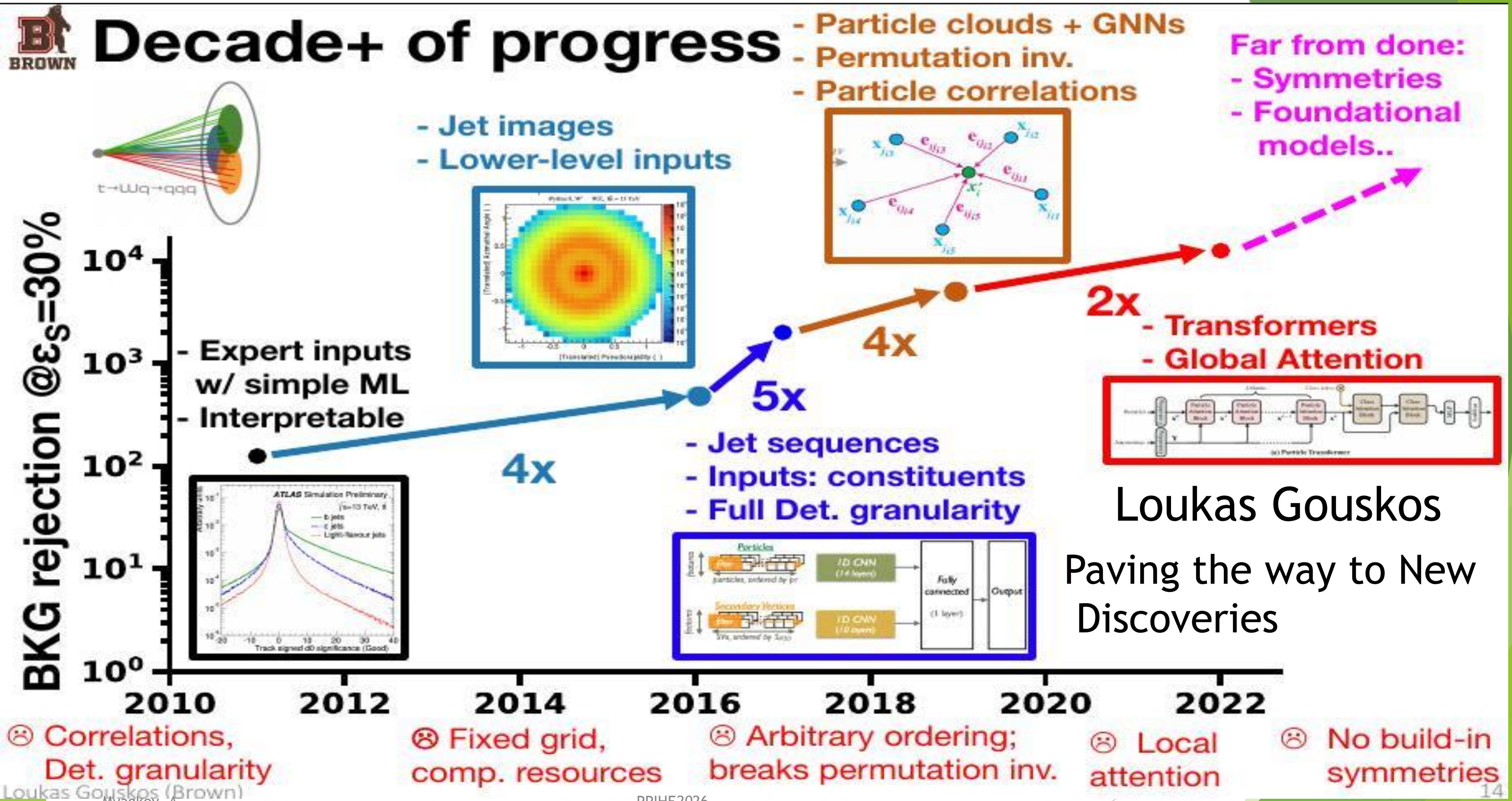


https://iml-wg.github.io/HEPML-LivingReview/

M. Feickert and B. Nachman, A Living Review of Machine Learning for Particle Physics, 2102.02770. 1

ML: main goals & methods

- Main algotithms: Supervised, Unsupervised, Semi-Supervised, and Reinforcement Learning
- Goals: Classification, Regresion ,Generation
- Input types: vector, sequence,picture,graph … text,music,movie…
- Architectures: DNN, CNN,RNN,GNN,Transformers



Decade+ of progress

- Particle clouds + GNNs
- Permutation inv.
- Particle correlations

Far from done:
- Symmetries
- Foundational models..

- Jet images
- Lower-level inputs

- Expert inputs w/ simple ML
- Interpretable

- Jet sequences
- Inputs: constituents
- Full Det. granularity

- Transformers
- Global Attention

☹ Correlations, Det. granularity

☹ Fixed grid, comp. resources

☹ Arbitrary ordering; breaks permutation inv.

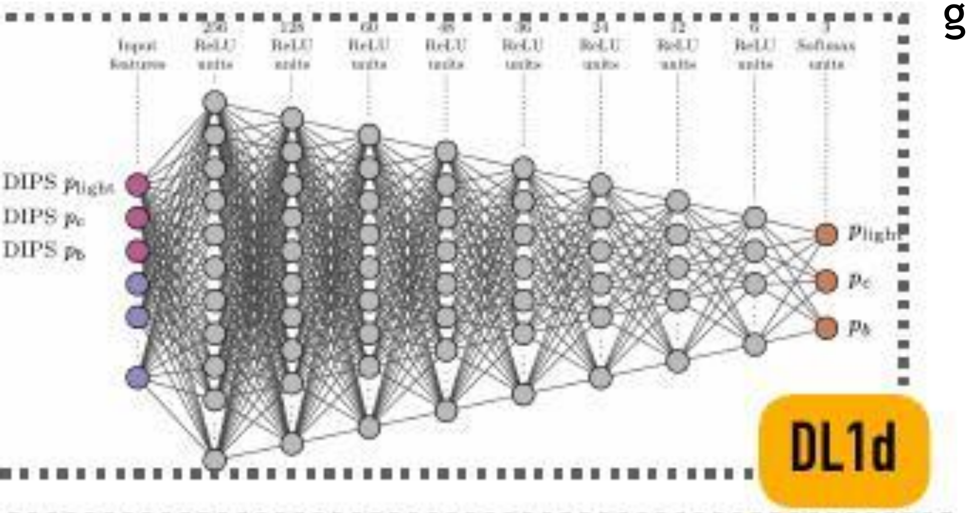
☹ Local attention

☹ No build-in symmetries

Loukas Gouskos

Paving the way to New Discoveries

BDT and Deep Neural Networks (DNN)

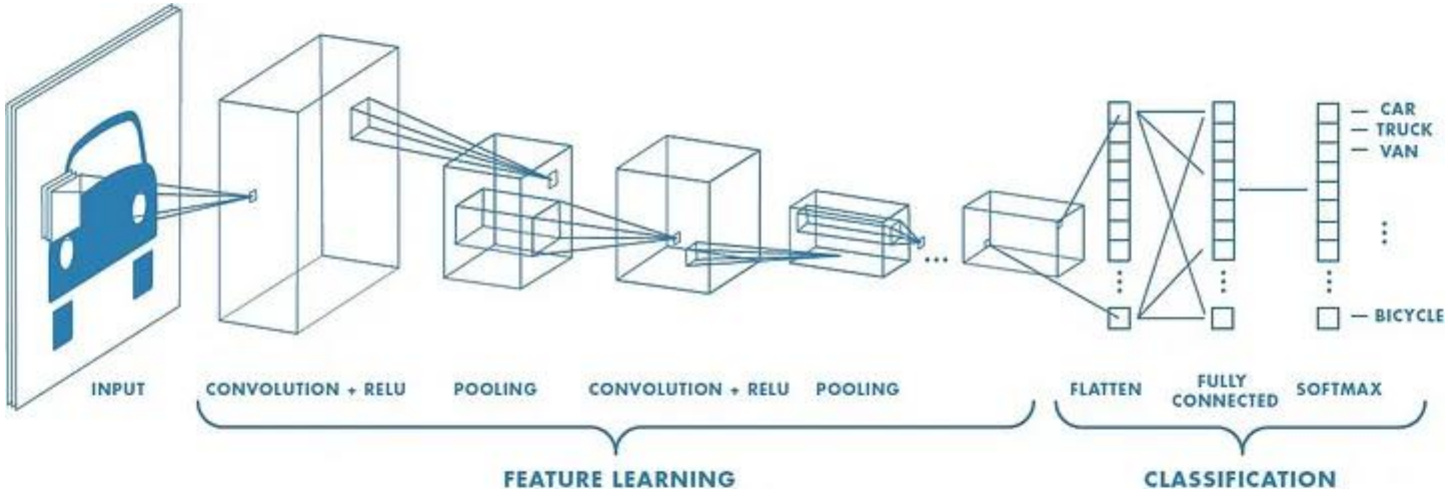


The MV1 (ATLAS-CONF-2014-046) algorithm was designed to combine the outputs of the two low-level likelihood-based taggers exploiting IP and SV information.

MV2 (ATL-PHYS-PUB-2015-022) is built on a BDT-based architecture. The key distinction between MV1 and MV2 lies in the handling of input IP and SV information. MV1 utilizes the outputs of intermediate low-level tagging algorithms, whereas MV2 directly takes as input the variables employed as inputs to the low-level taggers.

DL1 (ATL-PHYS-PUB-2017-013) architecture consists of 8 fully-connected hidden layers, each with a decreasing number of nodes ranging from 78 to 6, and 3 Max-out layers

Convolutional Neural Network CNN

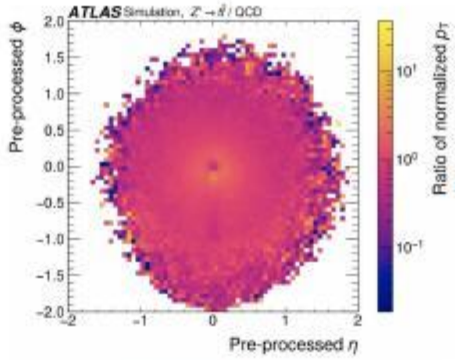


Jets as Images

Constituents binned in (η, φ) → “jet images”.

- Enables 2D pattern recognition of energy flow using CNNS (as ResNet50).

JINST 19 (2024) P08018



JINST 19 (2024) P08018

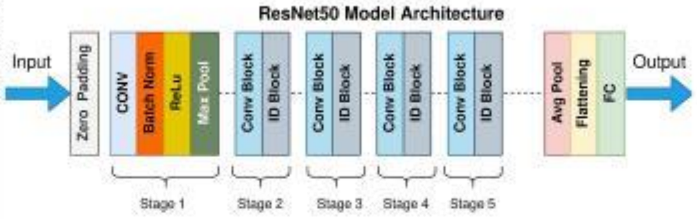
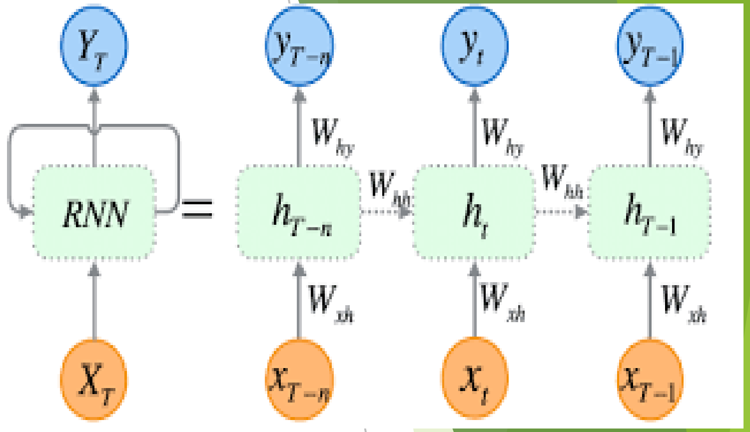


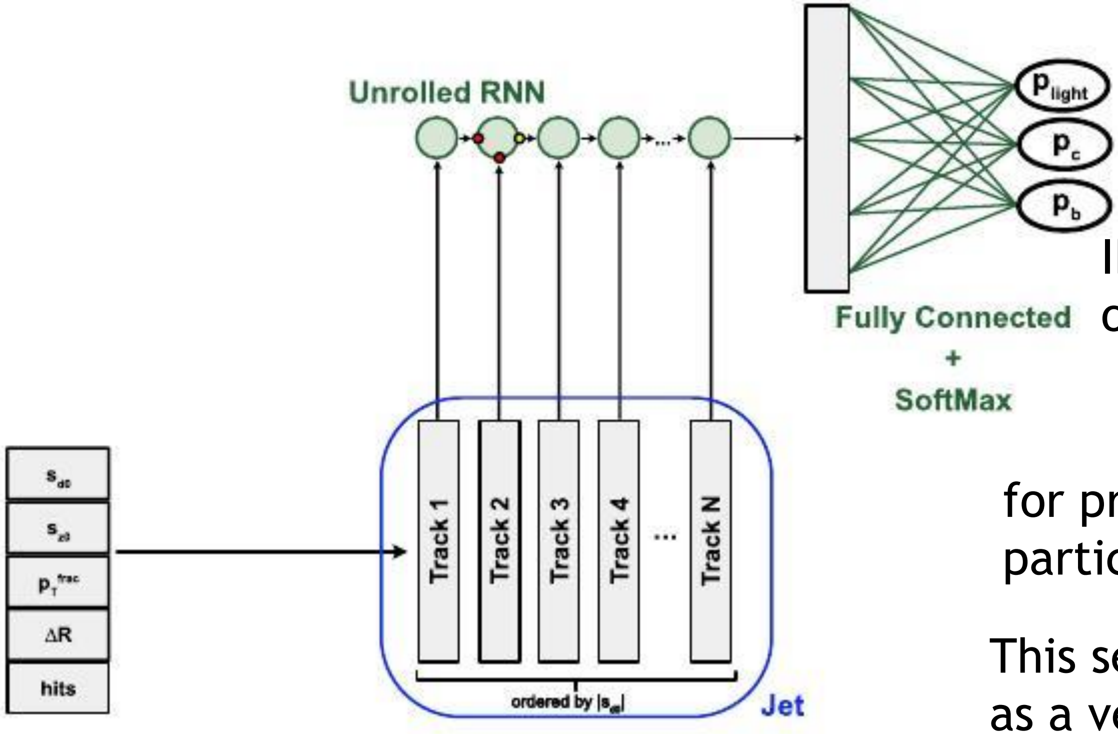
Image Credits

Recurrent Neural Network (RNN)

RNNIP Eur. Phys. J. C (2023) 83:681

ATLAS flavour-tagging algorithms for the LHC Run 2dataset



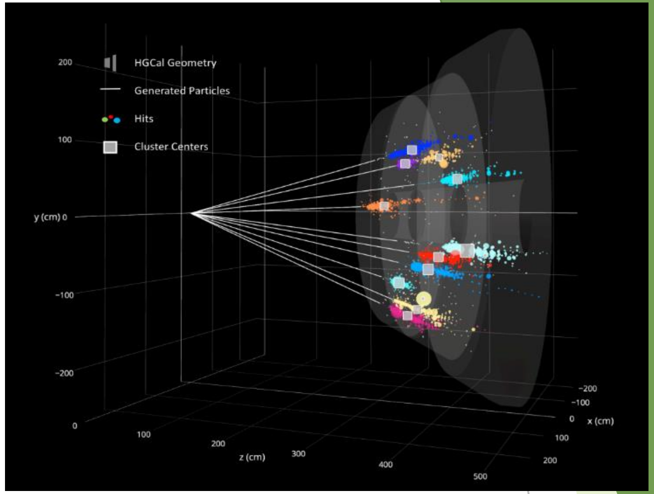


IP-based b-tagging algorithms assume that the properties of each track in a jet are independent of all other tracks

for processing sequential data like jet constituents from particle collisions,

This sequence is then passed to the neural network cells as a vector of the ordered-track features.

Graph Neural Networks (GNN)



- GNN - Many problems involve data represented as unordered sets of elements with rich relations and interactions with one another, and can be naturally expressed as graphs.

Supports variable size and permutation invariance.

GN1(ATL-PHYS-PUB-2022-027)

Input : track+jet param's, Using concatenated i

jet and associated tracks directly

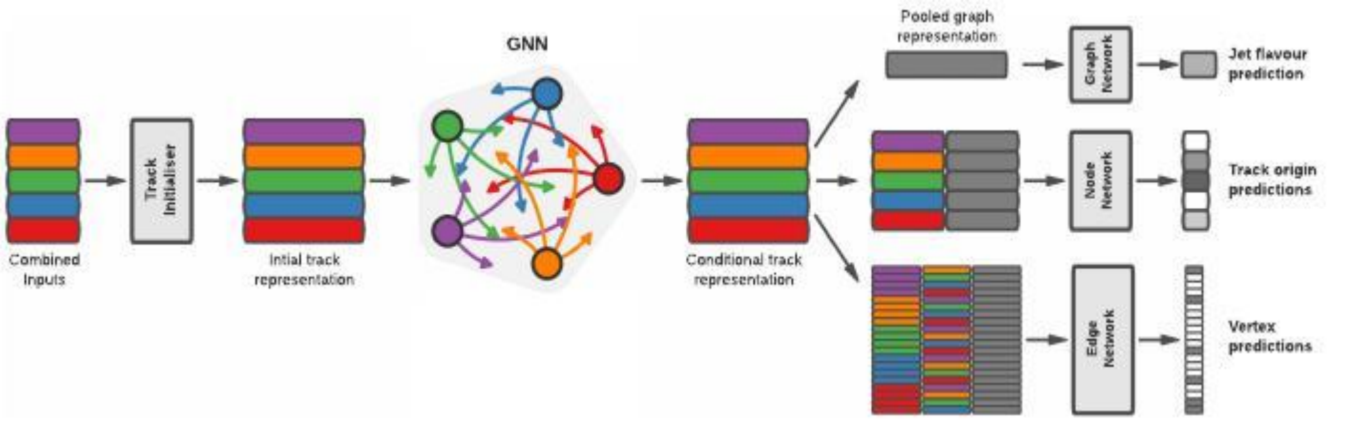
NN: input->latent space

Vertex in graph= track

GNN: full connected

Output : Prob's (b,c,light),

Auxiliary tasks: Predict track origin and vertices



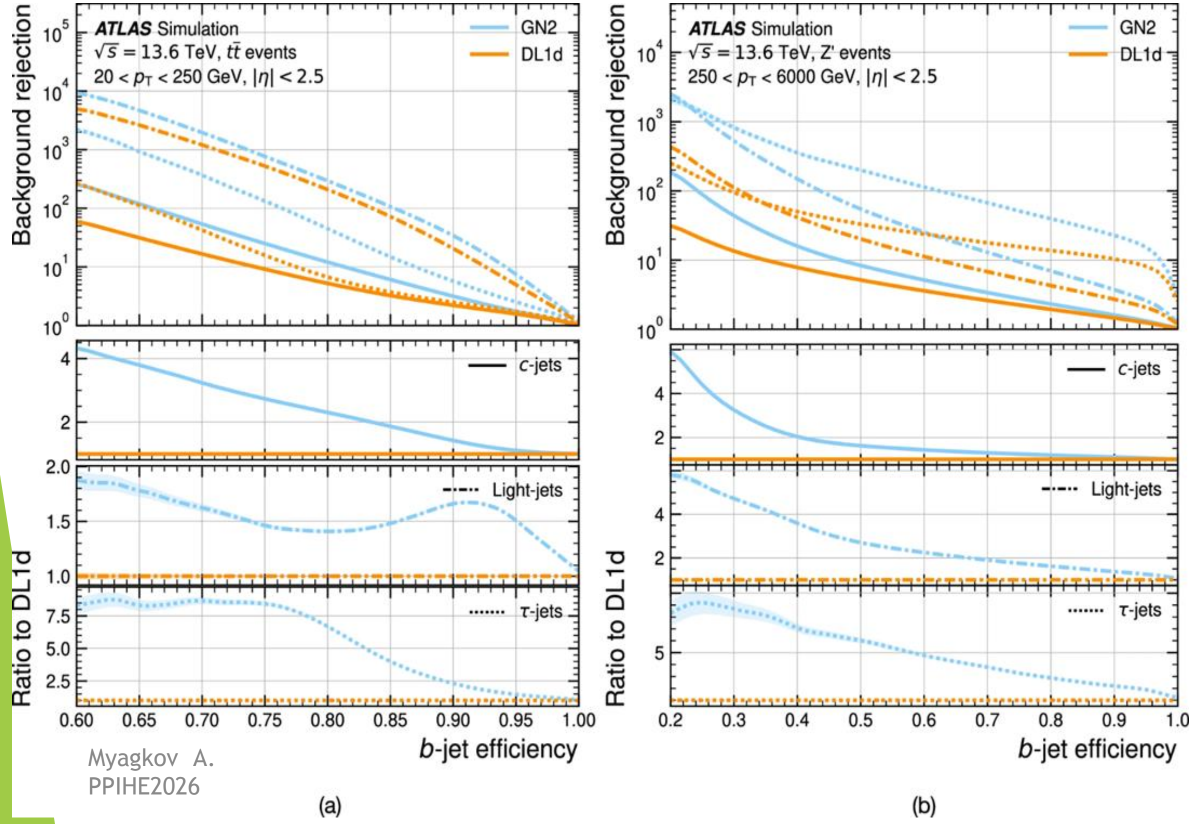
Transformer

Transforming jet flavour tagging at ATLAS
Nature Communications volume 17, Article number: 541 (2026)

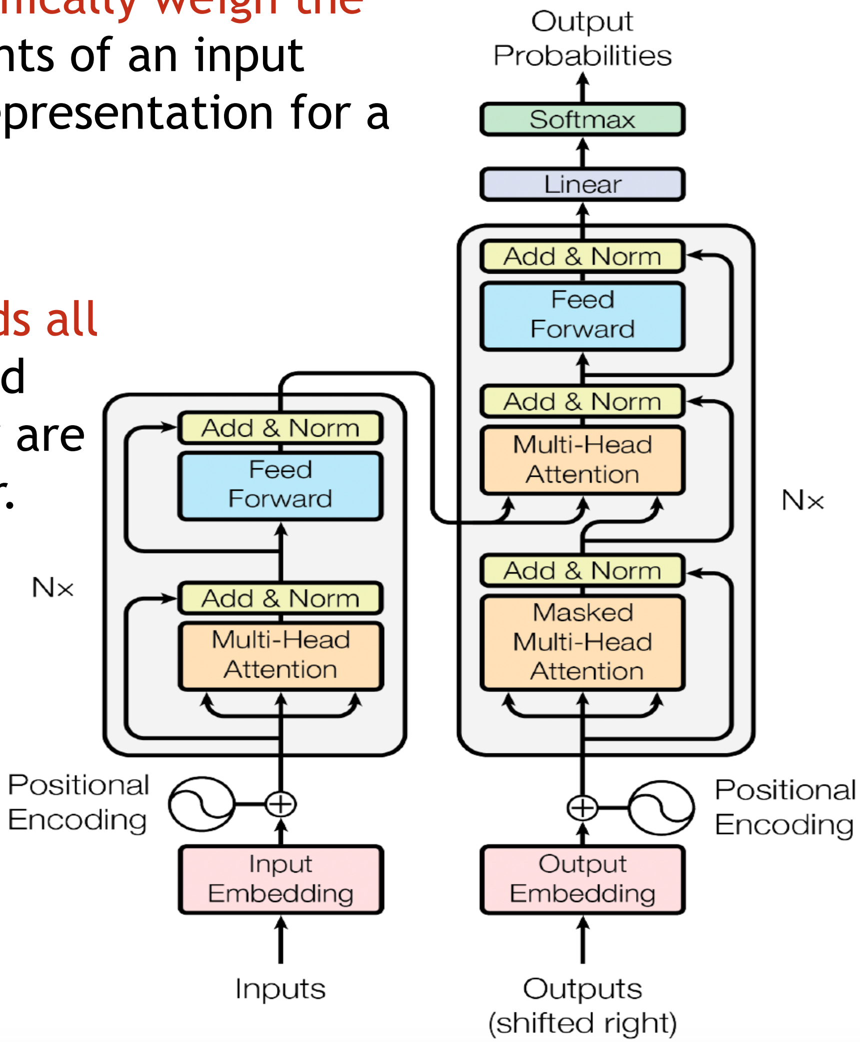
Attention Is All You Need
arxiv:1706.03762
Cited: 250 K times

GN2: Like GN1, but GNN-> Transformer
More layers
1.5 M parameters

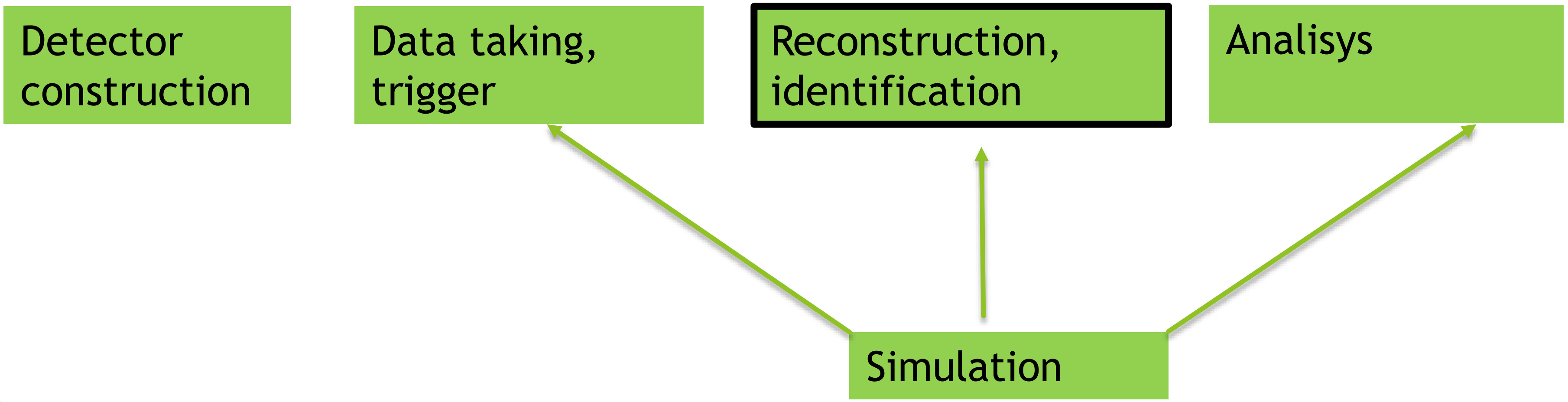
The ability of a model to dynamically weigh the importance of different elements of an input sequence when computing a representation for a particular element



The Transformer reads all the words at once and determines how they are related to each other.



Main ML applications in HEP



Detector construction

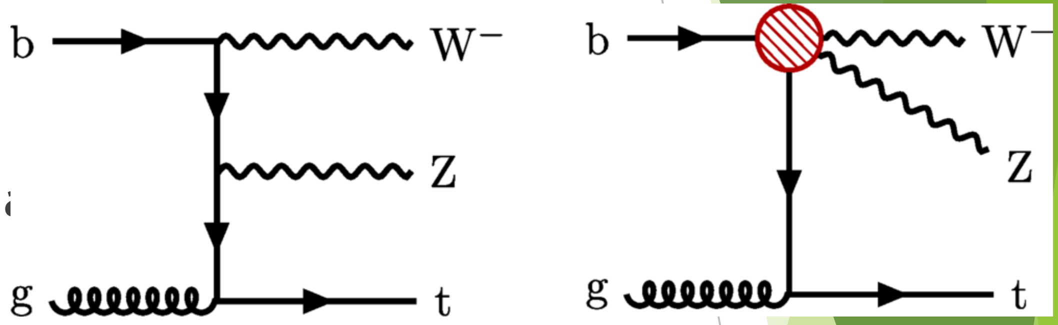
Data taking, trigger

Reconstruction, identification

Analisys

Simulation

Observation of tWZ production at the CMS experiment
Phys. Rev. Lett. 136 (2026) 081802



- **The first observation of single top quark production in**

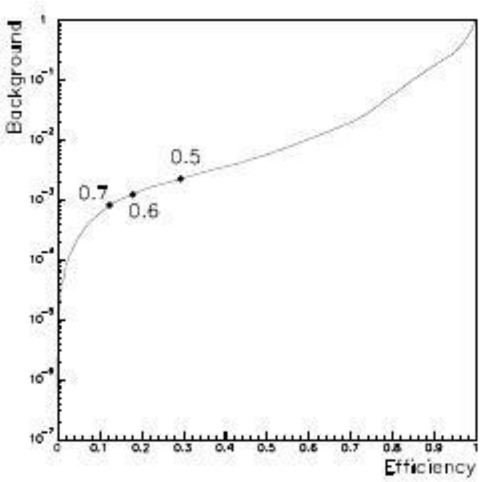
with a W and a Z boson in pp collisions

small production cross section of the tWZ process compared to backgrounds with similar signature

- The signal is established with **a statistical significance of 5.8 σ**
- **σ = 248 ± 52 fb**
- **A transformer ML algorithm** based on the particle transformer PART architecture is employed to discriminate between signal and background processes

Observation of $K^+ \to \pi^+\pi^0\pi^0\gamma$ decay
The OKA collaboration 2310.19652

The major background source, the decay $K^+ \to \pi^+\pi^0\pi^0$, decay is ≈ 5000 times more frequent (1.76%) than the radiative decay. The extra "ghost" γ easily emerges due to the fluctuations of π^+ hadronic shower in GAMS e.m. cal. -> presenting a major challenge in this analysis. The RBFN neural network is employed to suppress the background down to Signal:Noise≈ 1 : 1 level.



Neural net performance, the thresholds used in this analysis shown with bullets.

$$\mathrm{BR}\left(K^+ \to \pi^+\pi^0\pi^0\gamma\right) =$$
$$(4.1 \pm 0.9(\mathrm{stat}) \pm 0.4(\mathrm{syst})) \times 10{-}6\ (E_\gamma * > 10\,\mathrm{MeV})$$

Neural posterior estimation of the neutrino direction in IceCube using transformer-encoded normalizing flows on the sphere

2604.19846

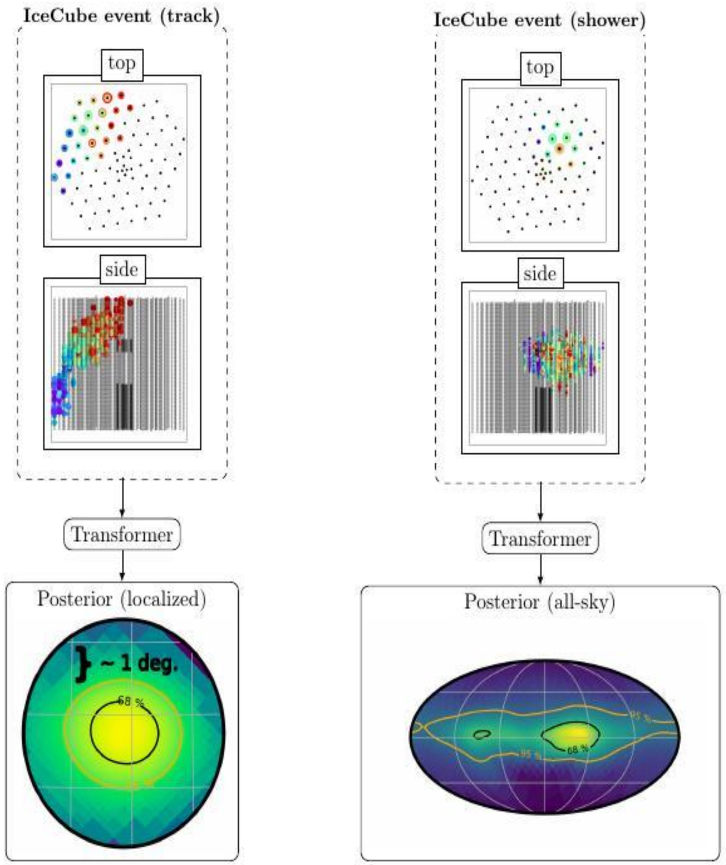
IceCube is a cubic-kilometer-scale neutrino detector

- A precise directional reconstruction of IceCube neutrinos is vital for associations with astronomical objects.

New version: GNN->Transformer

All-sky scans can be performed within seconds rather than hours

The median angular resolution improves by a factor of 1.3 for throughgoing tracks, by a factor of 1.7 for showers and by a factor of 2.5 for starting tracks compared to state-of-the art likelihood reconstructions based on B-splines



(a) track example

(b) shower example

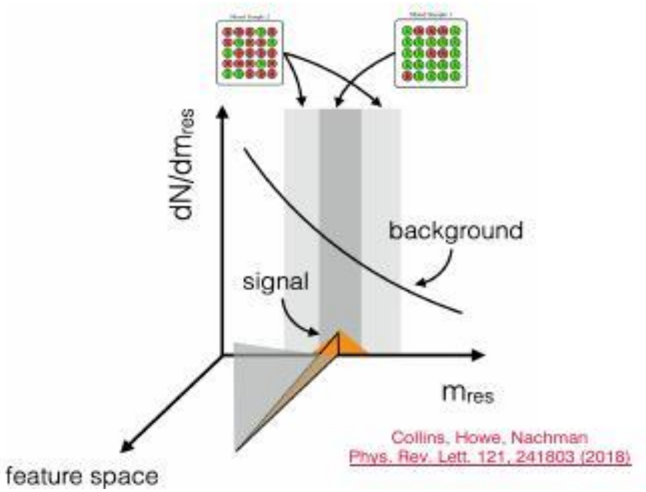
Anomaly Detection

In high-energy physics, anomaly detection solves a fundamental problem:
How do you find something if you don't know what you're looking for?

- Model agnostic/independent search
- Looking for deviations from background only
- Less sensitive to any specific model, but can look for multiple different models

Input → Encoder → Latent Space → Decoder →Reconstruction
Learning from background events
New physics events are difficult to reconstruct → high reconstruction error
Examples: variational autoencoders (VAEs), convolutional autoencoders

Classification Without Labels (CWoLa) 1805.02664



Collins, Howe, Nachman
Phys. Rev. Lett. 121, 241803 (2018)

1. Identify an observable m_res in which a signal is expected to be resonant, and a set of auxiliary variables Y that are to be used for signal selection

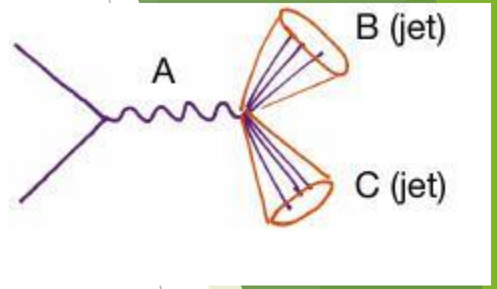
-3. Define a signal region in a window around m_res and sideband regions

. Use a cross-validation procedure to separate training samples from test amples.

) **Train a classifier to discriminate training events drawn from the sideband regions**

) **Select a fraction of the most signal-like test events as determined by the classifiers.**

.Perform a statistical test for the presence of an excess in the signal egion of the m_res distribution **after the cut has been applied**,



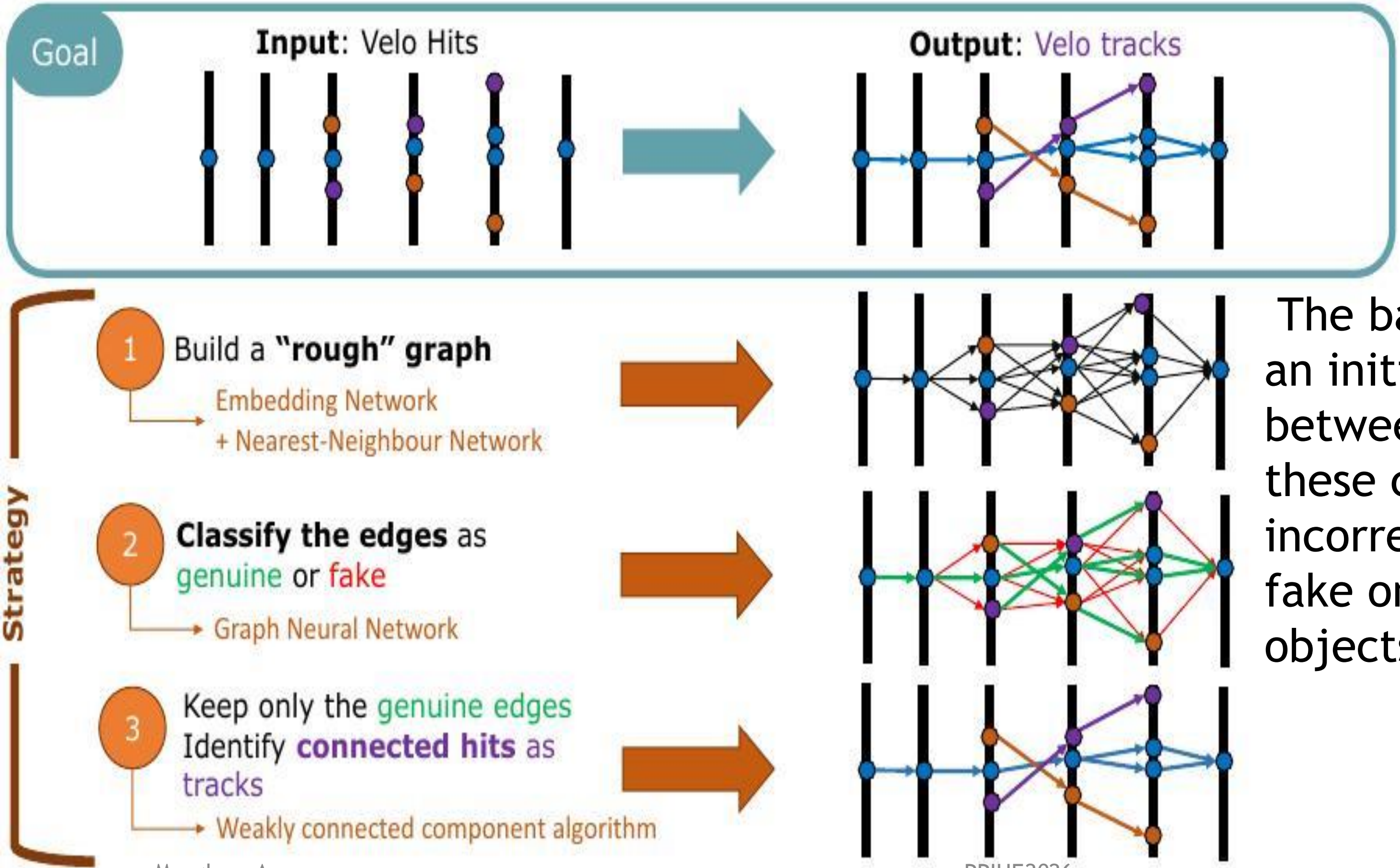
Di-jet (large-R jets) resonance search

➡ pp → A → BC → JJ

➡ Training classifiers on data, with no labels

it is enough to have two datasets at hand with an unequal fraction of signal instances in each set. A standard classifier can then be trained to discriminate between the two mixed datasets, and this can be shown to be the optimal classifier to discriminate between signal and background instances.

GNN-based pipeline for track finding in the Velo at LHCb 2407.12119



Goal

Input: Velo Hits

Output: Velo tracks

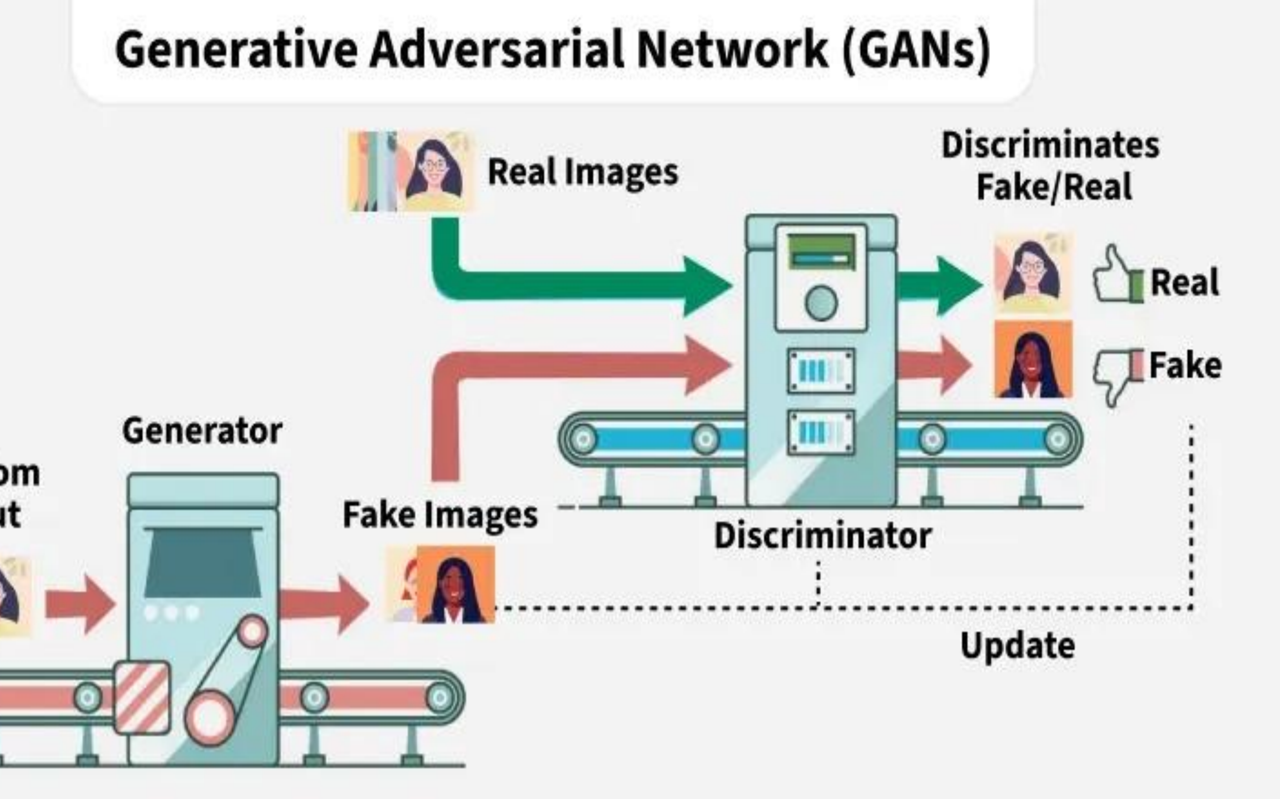
Strategy

1. Build a **"rough" graph**
   - Embedding Network + Nearest-Neighbour Network
2. **Classify the edges** as genuine or fake
   - Graph Neural Network
3. Keep only the genuine edges. Identify **connected hits** as tracks
   - Weakly connected component algorithm

On average, 150 particles in the VELO acceptance, and 2,200 hits, in each event.

The basic idea of our GNN pipeline is to build an initial graph of possible connections between hits in the detector, accurately classify these connections as correct (true/genuine) or incorrect (fake), and then, after discarding the fake ones, transform them into a set of track objects.

ML-based Fast Simulation of FARICH Responses 2605.17635



Two neural networks—the Generator and the Discriminator

The Generator: to produce data convincing enough to fool the discriminator.

The Discriminator: Evaluates the generated data against real data from an actual training dataset.

The Spin Physics Detector (SPD) at NICA to study Drell–Yan processes, J/Ψ production processes, elastic reactions, spin effects

An aerogel counter in the SPD end-cap region to provide π/K-separation below 5 GeV/c using a Focusing Aerogel Ring Imaging Cherenkov (FARICH) detector.

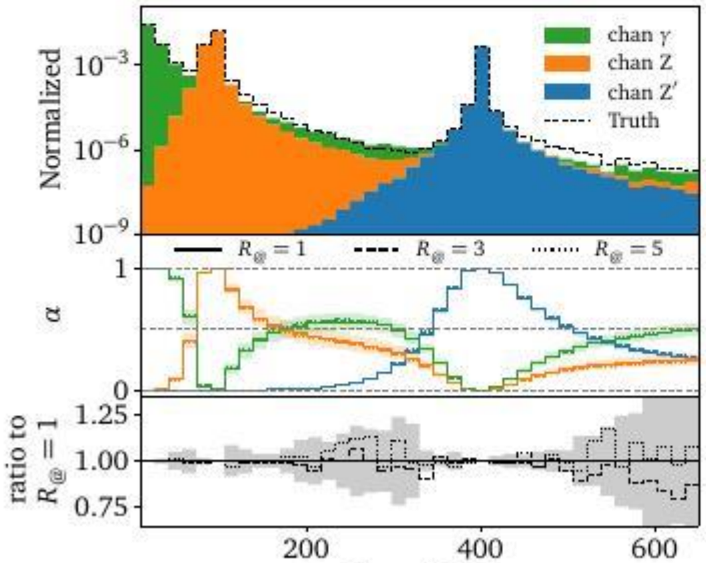
ML-based Fast Simulation of FARICH Responses

The aim is to generate realistic samples of Cherenkov photon hits on the detector matrix

Employ a conditional GAN (cGAN) with a lightweight convolutional architecture

Event Generation with “MadGraph-ready Neural Importance Sampling” (MadNIS)

- 2212.06172 2311.01548
- 2408.01486
- use neural networks to replace expensive loop amplitudes with fast and precise surrogates
- MADNIS replaces the local multi-channel weights from Sec.(2.1) with trainable channel-weight networks (CWnets)

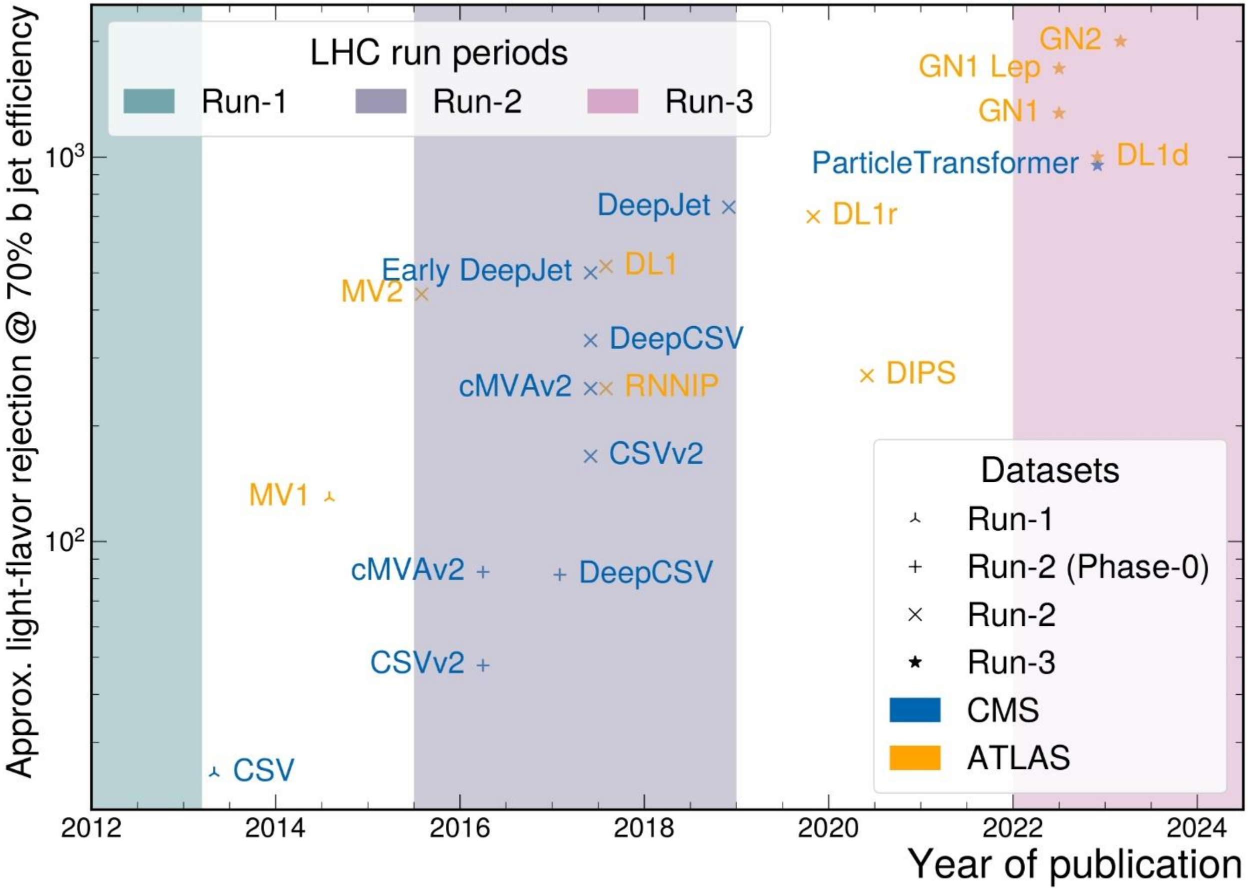


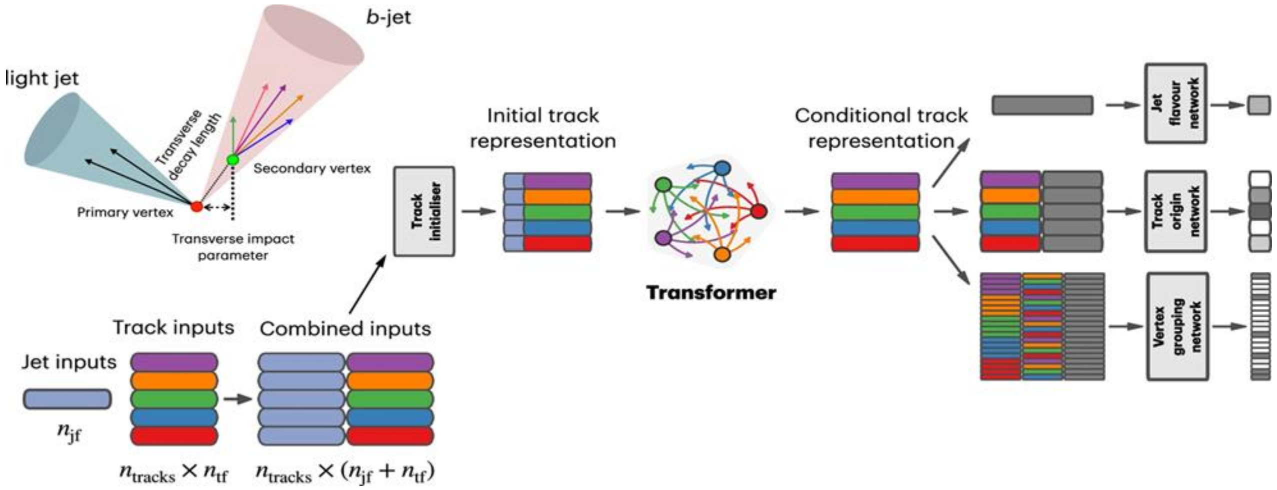
Summary

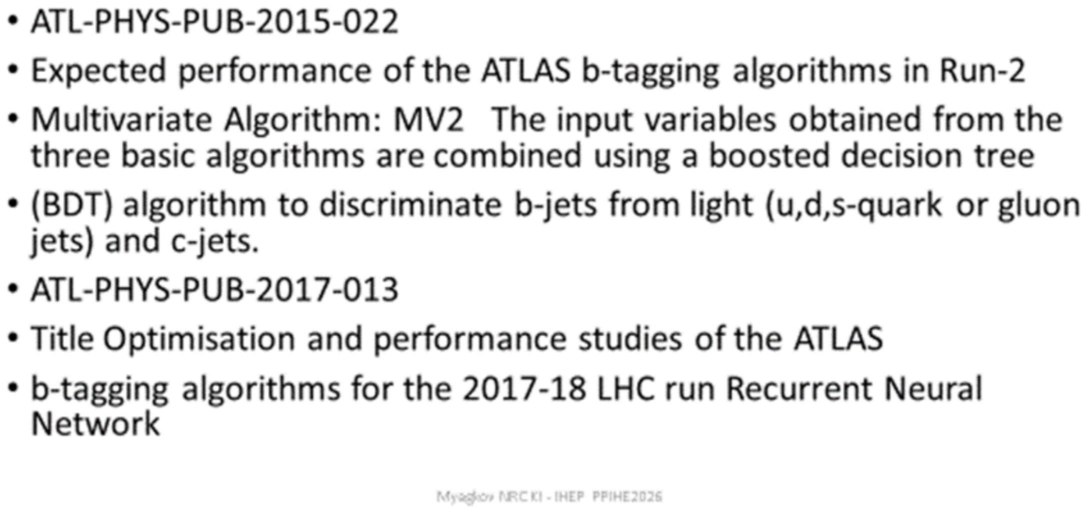
- ML has been a longstanding companion in HEP in various stages of the data

  analysis pipeline
- ML as a Toolset for HEP
- ML serves as a valuable assistant, maximising the exploration of costly collision data
- Choosing ML architectures based on the data structures to optimise efficiency
- Evolution towards unsupervised and semi-supervised learning on more generative tasks

Thank you for attention


LHC run periods
Run-1
Run-2
Run-3
Datasets
Run-1
Run-2 (Phase-0)
Run-2
Run-3
CMS
ATLAS
Approx. light-flavor rejection @ 70% b jet efficiency
Year of publication
10^3
10^2
2012
2014
2016
2018
2020
2022
2024
GN2
GN1 Lep
GN1
ParticleTransformer
DL1d
DeepJet
DL1r
Early DeepJet
DL1
MV2
DeepCSV
cMVAv2
RNNIP
DIPS
CSVv2
MV1
cMVAv2
DeepCSV
CSVv2
CSV


b-jet
light jet
Transverse decay length
Secondary vertex
Primary vertex
Transverse impact parameter
Track initialiser
Initial track representation
Transformer
Conditional track representation
Jet flavour network
Track origin network
Vertex grouping network
Track inputs
Combined inputs
Jet inputs
n_{jf}
$n_{\text{tracks}} \times n_{\text{tf}}$
$n_{\text{tracks}} \times (n_{\text{jf}} + n_{\text{tf}})$



- ATL-PHYS-PUB-2015-022
- Expected performance of the ATLAS b-tagging algorithms in Run-2
- Multivariate Algorithm: MV2 The input variables obtained from the three basic algorithms are combined using a boosted decision tree
- (BDT) algorithm to discriminate b-jets from light (u,d,s-quark or gluon jets) and c-jets.
- ATL-PHYS-PUB-2017-013
- Title Optimisation and performance studies of the ATLAS
- b-tagging algorithms for the 2017-18 LHC run Recurrent Neural Network

Myagkov NRC KI - IHEP PPIHE2026

Jets as Point Clouds

ATL-PHYS-PUB-2022-027
Graph Neural Network Jet Flavour Tagging with the ATLAS De Constituents treated as points embedded in a η–ϕ space.

- This structure allows to build local interactions.
- Graph-based models learn local + global structures.
- Naturally supports variable size and permutation invariance.
- Utilized by Particle Net (PNet) in a GNN built with EdgeConv layers.

- 80 neutrino events of TeV energy, to be within 0.18 degrees of the active galaxy NGC 1068 (M77). NGC 1068 is also the most significant astrophysical neutrino source identified from a search at the positions of 110 preselected high-energy gamma-ray sources. A search for subdominant sources in the sky map reveals evidence for two more active galaxies, PKS 1424+240 and TXS 0506+056. TXS 0506+056 had already been identified as a neutrino source from a multimessenger campaign triggered by an IceCube-detected neutrino of 290 TeV energy as well as from the observation in archival IceCube data of an earlier neutrino burst from TXS 0506+056 in 2014-15.