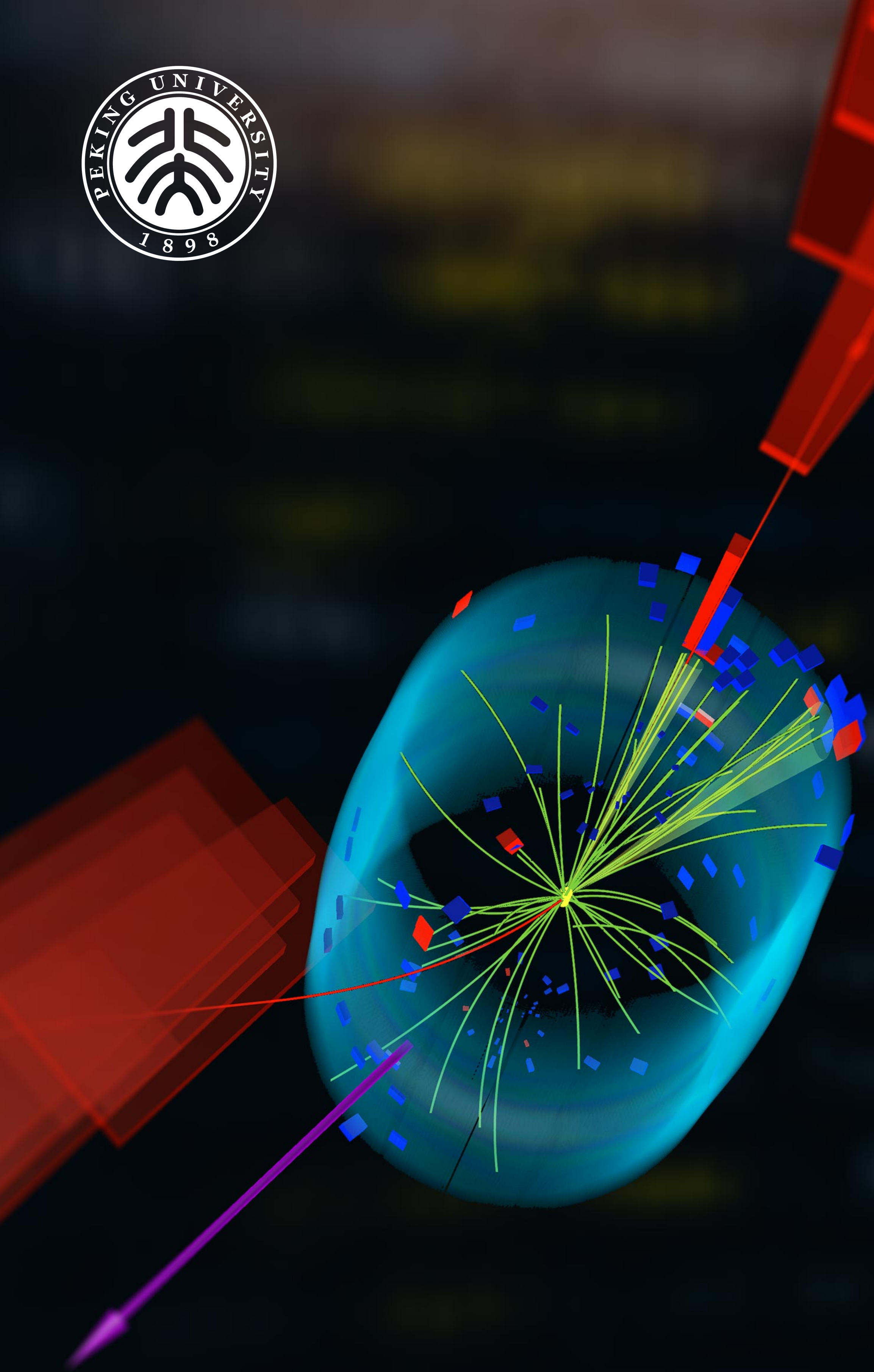


AI in HEP Experiment: Status and Prospects

— With Personal Insights and Reflections

Congqiao Li (李聪乔) *(Peking University)*

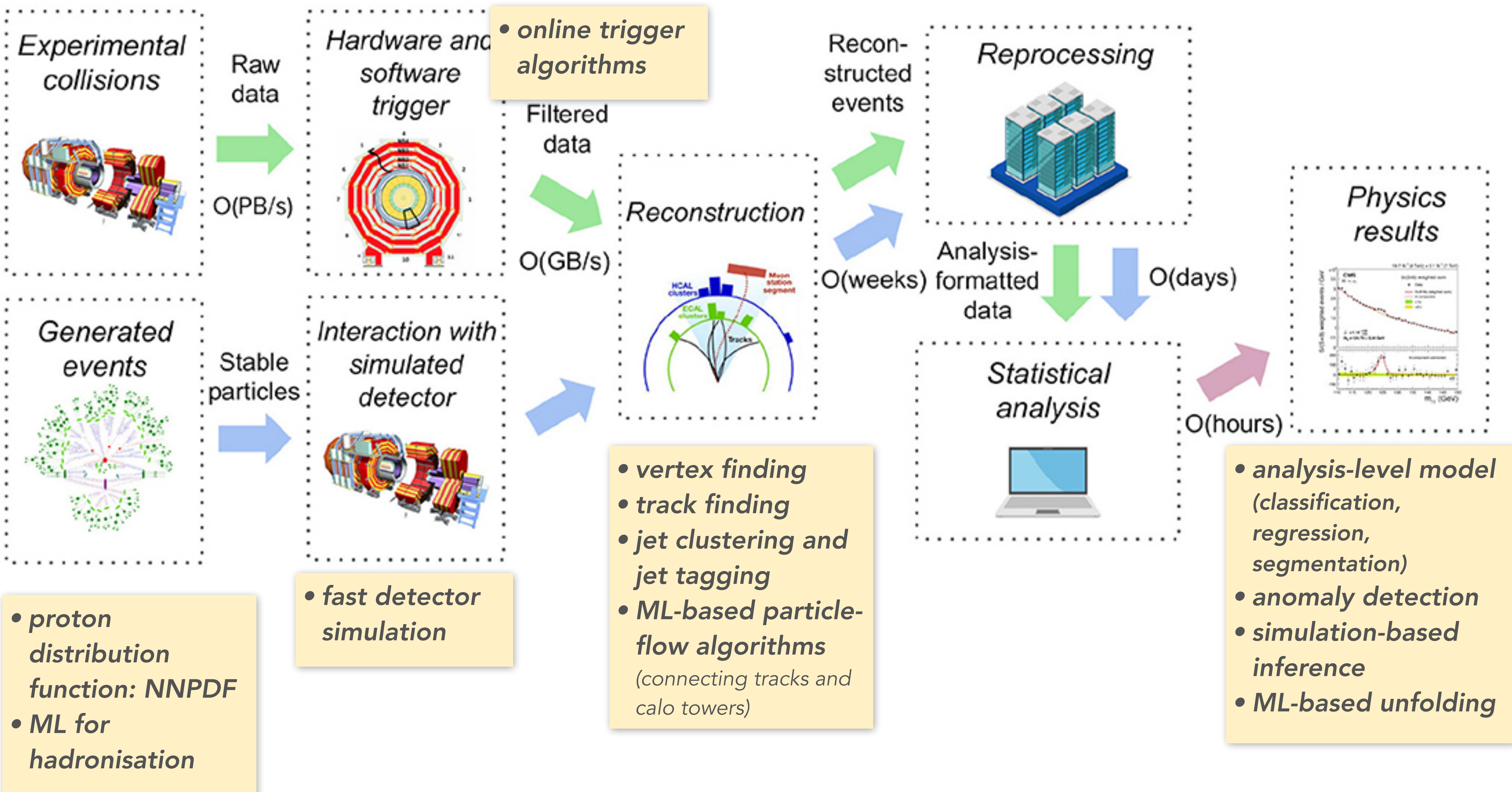
Quantum Computing and Machine Learning Workshop 2026 · *Shenzhen*

17 August, 2026

AI in HEP: activities

HEP experiments at LHC as an example

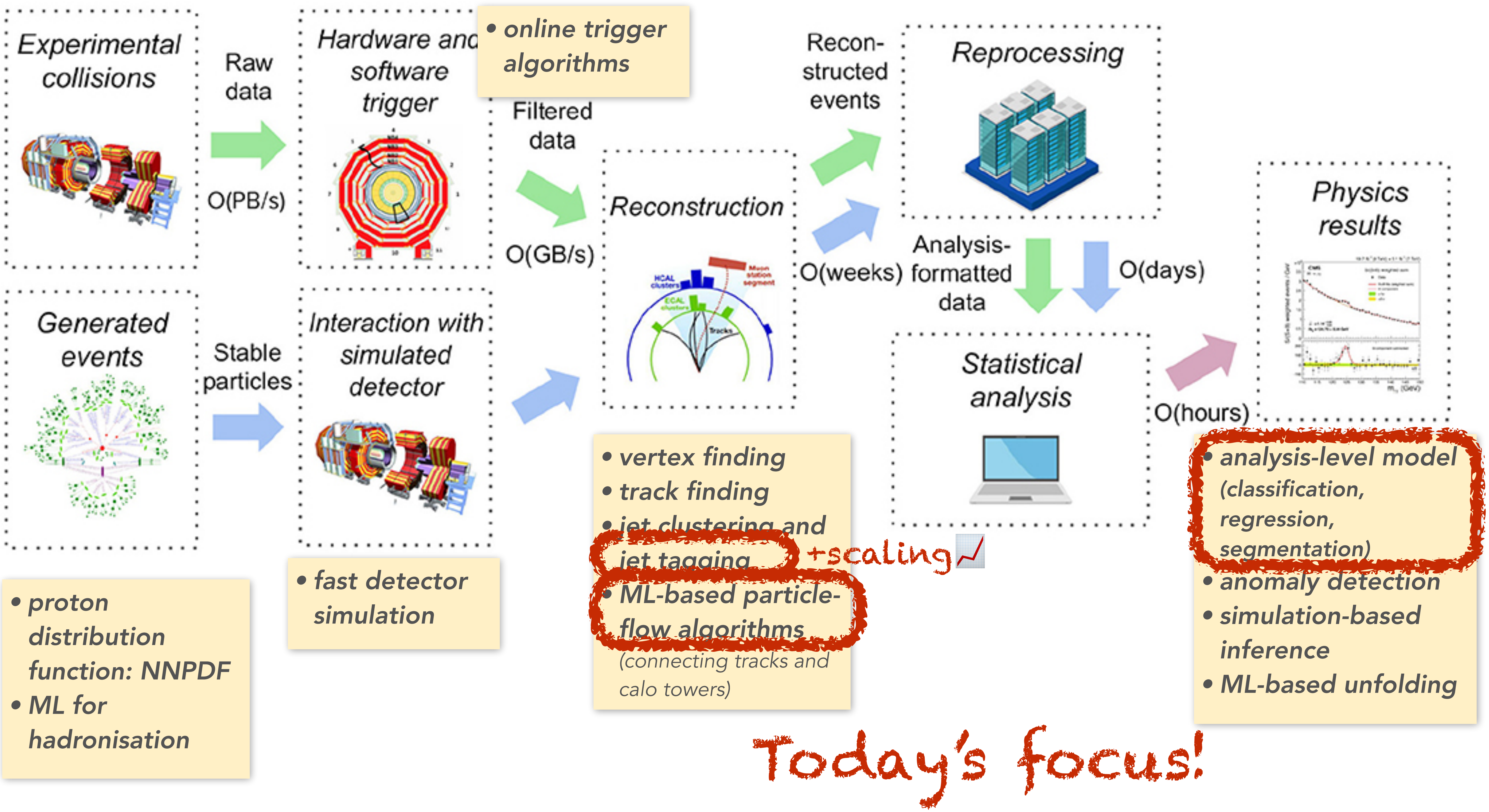
Deep learning appears in every stage of the experimental workflow!



AI in HEP: activities

HEP experiments at LHC as an example

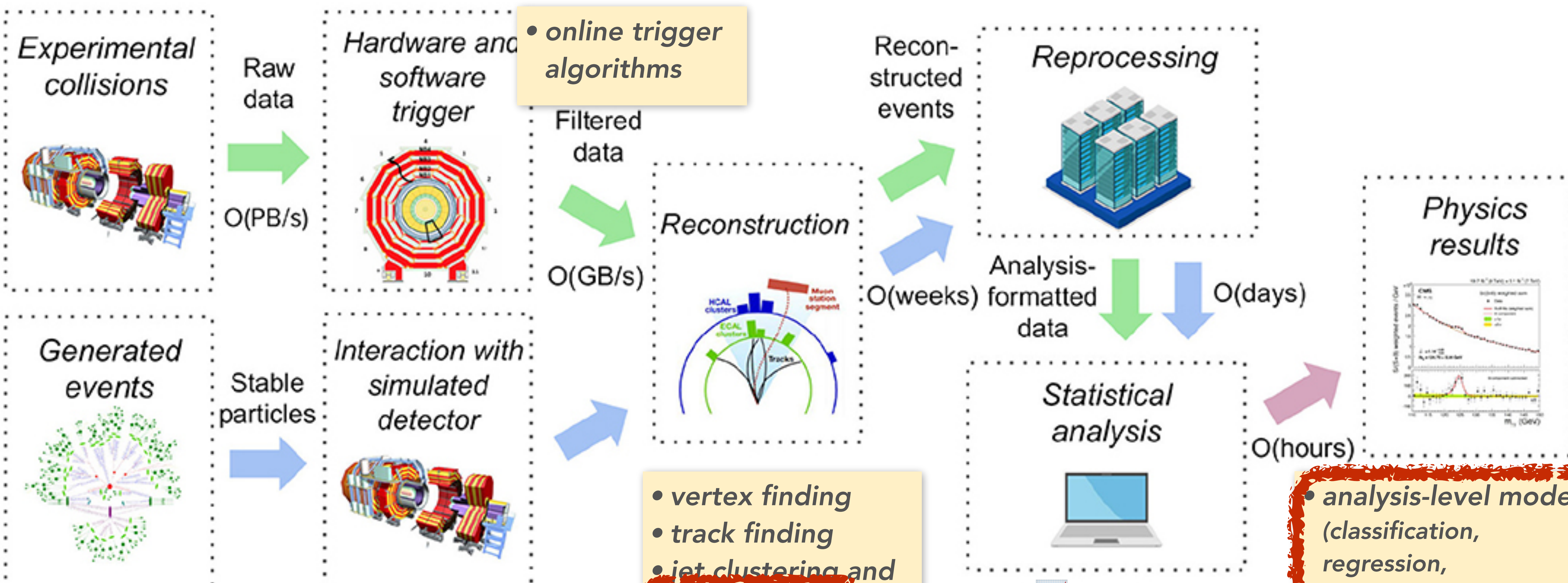
Deep learning appears in every stage of the experimental workflow!



AI in HEP: activities

HEP experiments at LHC as an example

Deep learning appears in every stage of the experimental workflow!



- *online trigger algorithms*

- *vertex finding*
- *track finding*
- *jet clustering and jet tagging* +scaling📈
- *ML-based particle-flow algorithms* *(connecting tracks and calo towers)*

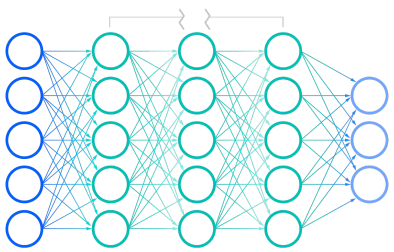
- ***analysis-level model** (classification, regression, segmentation)*
- *anomaly detection*
- *simulation-based inference*
- *ML-based unfolding*

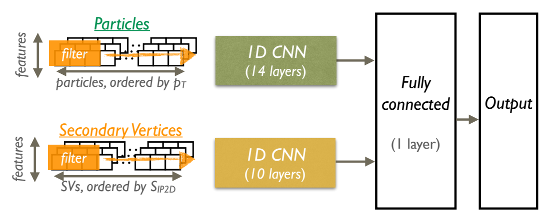
Today's focus!

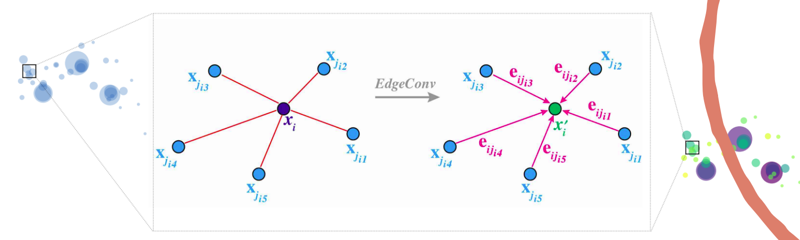
- ✦ Not covering everything (and maybe not necessary; similar talks are given yearly)
- ✦ Will only cover some ***emerging new progress/insights*** in HEP-ex with AI (personal reflection in orange)

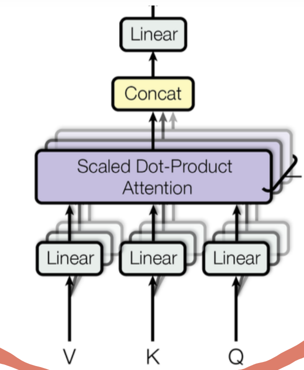
Jet models in the Transformer era

feed-forward NNs *(high-level inputs)* ··▸·· ***1D/2D CNNs, RNNs*** *(low-level inputs)* ··▸·· ***graph NNs*** ··▸·· ***Transformers*** ··▸·· *??*

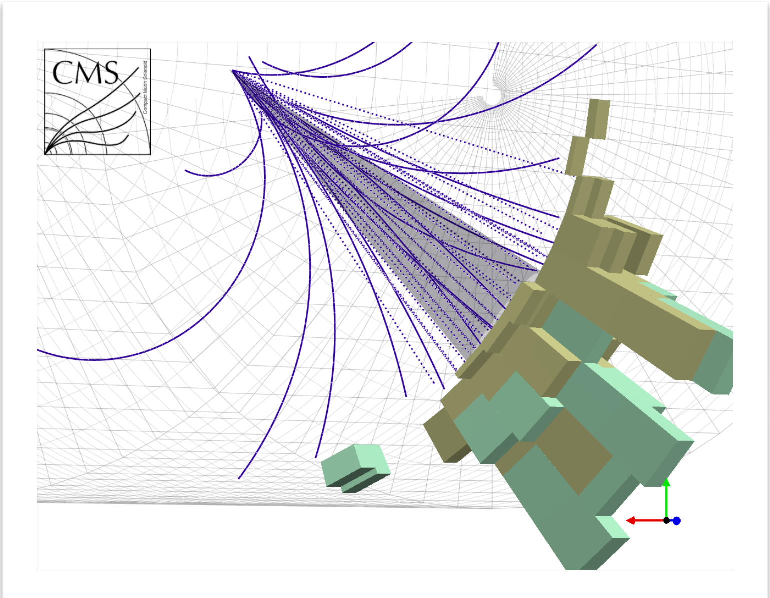








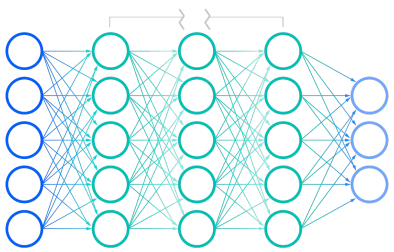
➔ Jet tagging has always been at the AI forefront @LHC

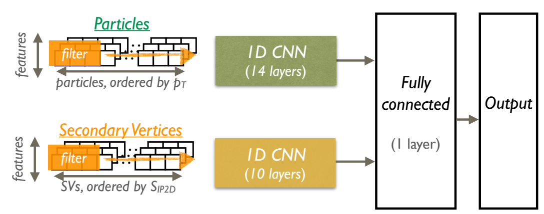


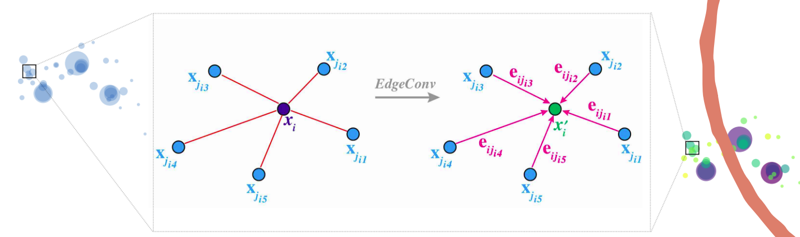
Jet tagging: determine the origin of a jet

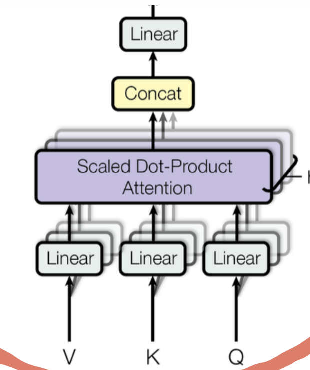
Jet models in the Transformer era

feed-forward NNs (high-level inputs) → *1D/2D CNNs, RNNs* (low-level inputs) → *graph NNs* → *Transformers* → ??





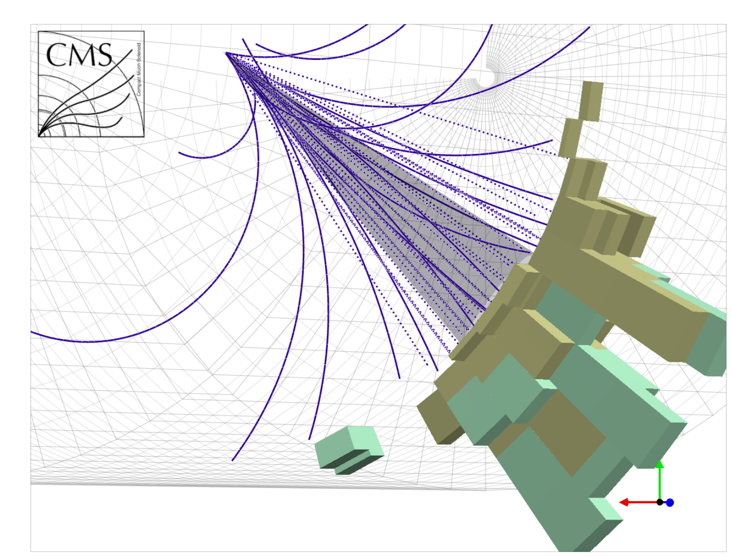




➔ Jet tagging has always been at the AI forefront @LHC

➔ After ~15 years of evolution, **"attention is now all you need"**

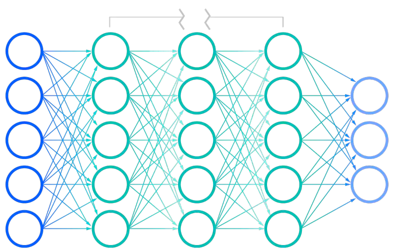
- Both ATLAS and CMS flagship taggers are **Transformer-based**
- notably: CMS has adopted Particle Transformer since 2022 as the backbone for both small-*R* and large-*R* jet tagging *H. Qu, CL, S. Qian. ICML 2022*

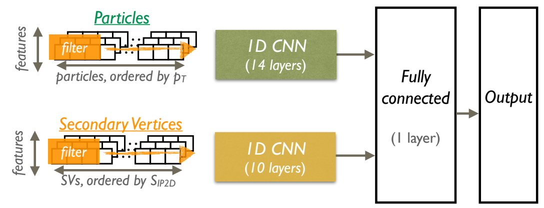


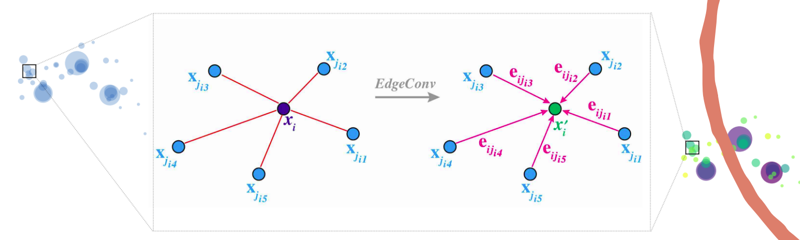
Jet tagging: determine the origin of a jet

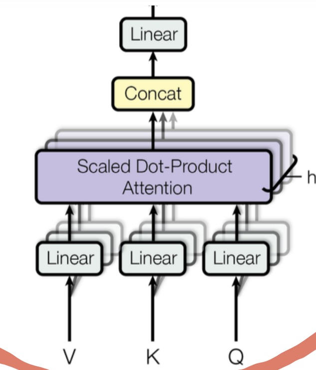
Jet models in the Transformer era (ATLAS)

feed-forward NNs (high-level inputs) → *1D/2D CNNs, RNNs* (low-level inputs) → *graph NNs* → *Transformers* → ??

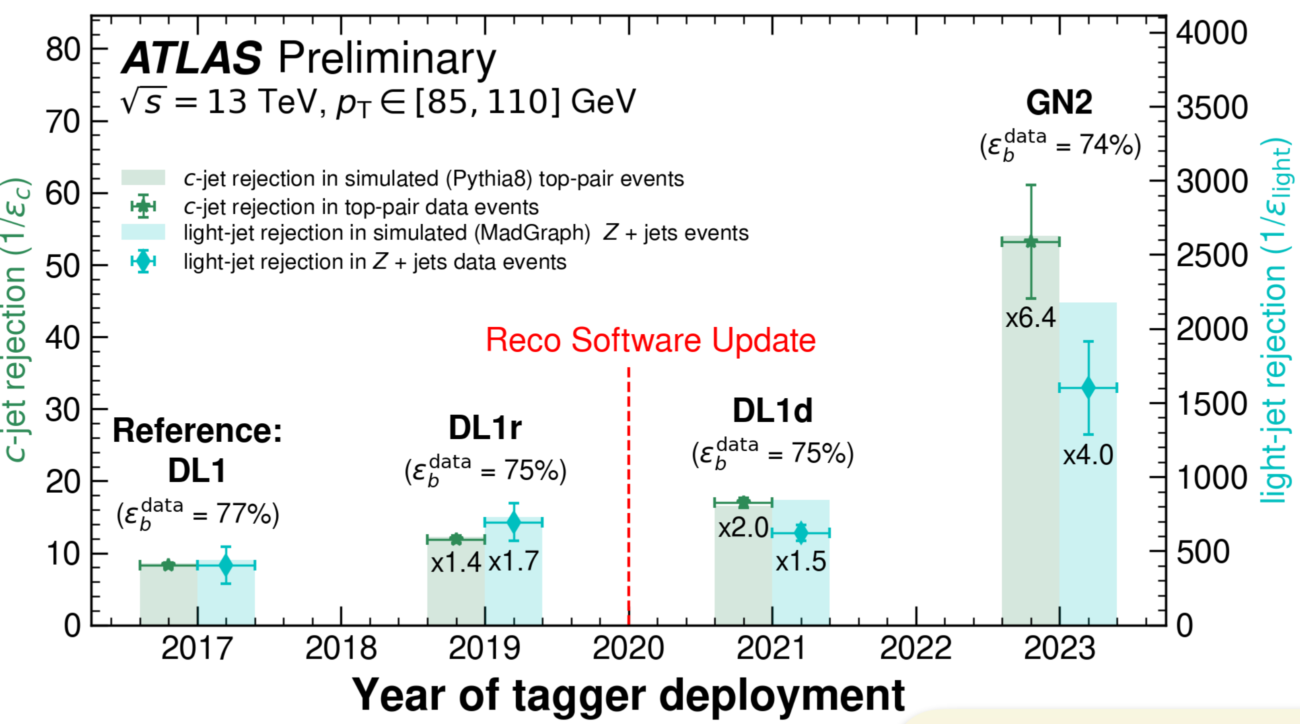






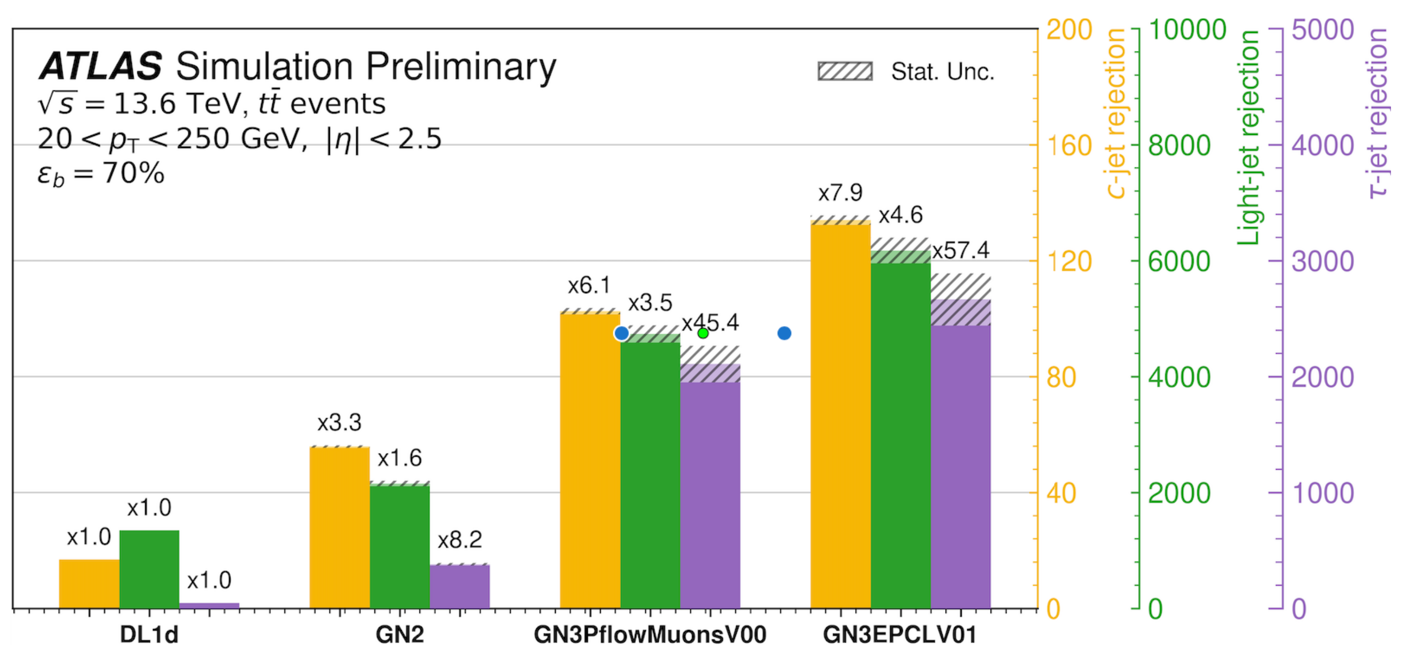


ATLAS's tagger revolution roadmap:



ATLAS FTAG-2023-07

ATLAS GN3



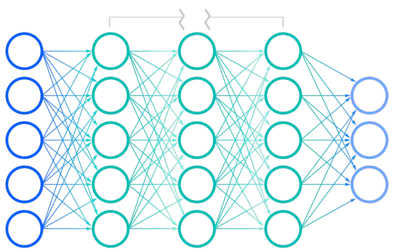
ATLAS FTAG-2026-07

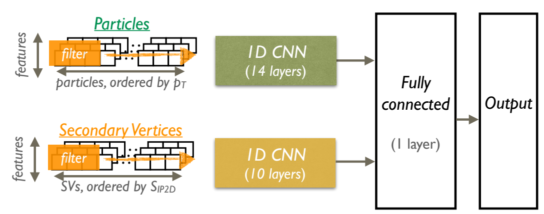
NEW

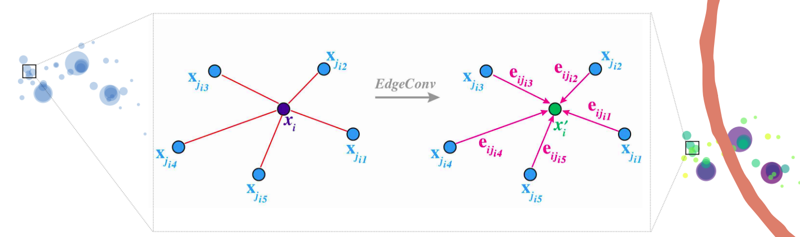
DL1 series: fully connected NN-based
GN2/GN3 (new!): transformer based
Significantly improved bkg (light/c/tau jet) rejection!

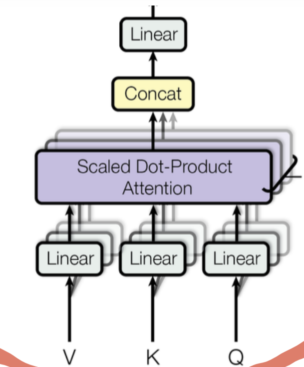
Jet models in the Transformer era (CMS)

feed-forward NNs (*high-level inputs*) ··▸·· *1D/2D CNNs, RNNs* (*low-level inputs*) ··▸·· *graph NNs* ··▸·· *Transformers* ··▸·· ??

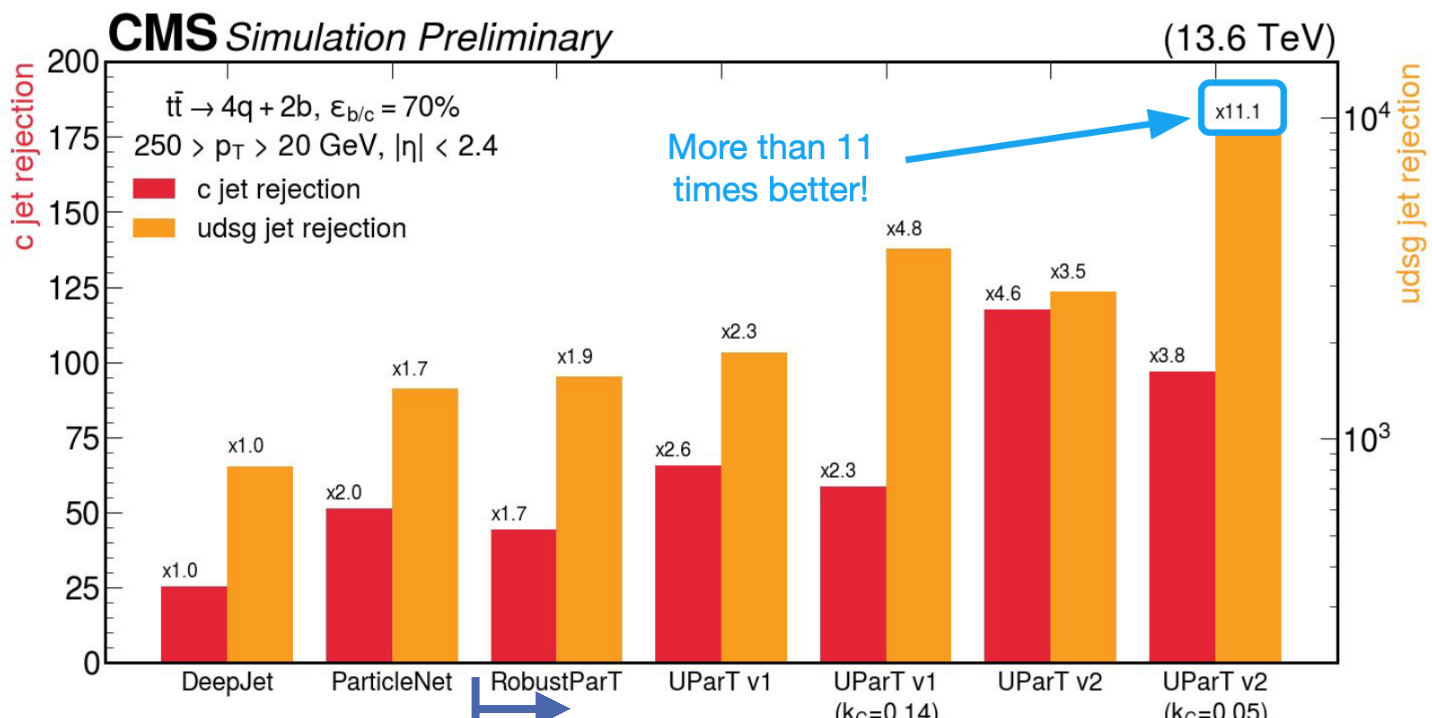








CMS's tagger revolution roadmap:



CNN/RNN *GNN* → *transformers*

CMS-DP-2024-066

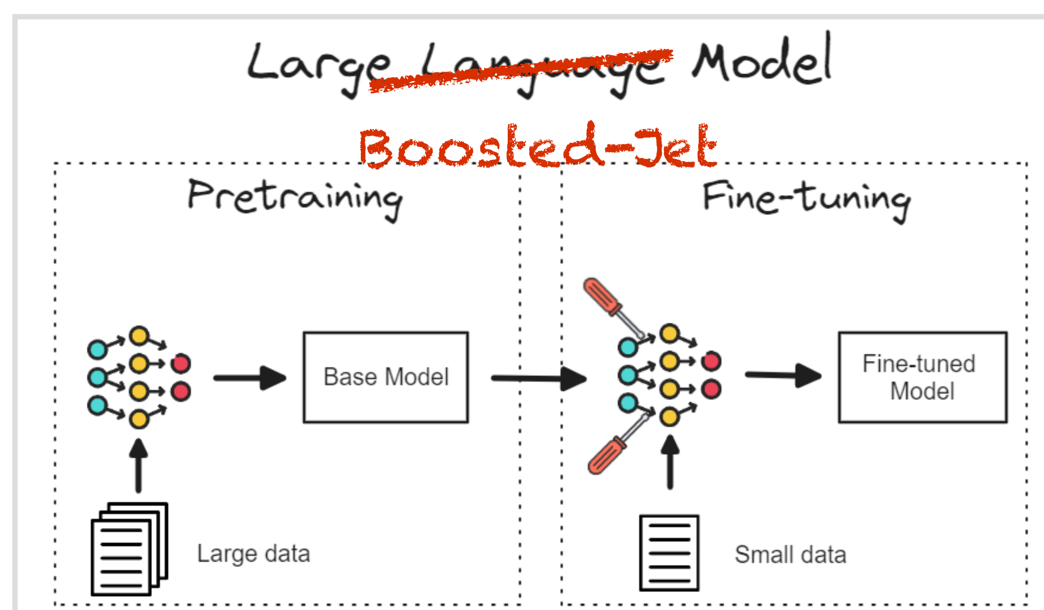
CMS UParT v2

(state-of-the-art small-*R* jet tagger)

NEW *CMS-DP-2026-104*

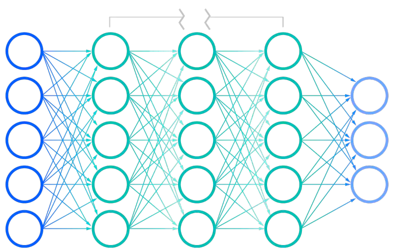
CMS GloParT v3

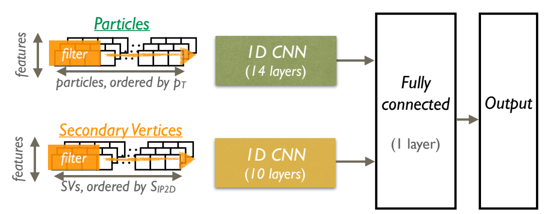
(best large-*R* jet tagger covering 300+ classes; paradigm of pretraining + fine-tuning)

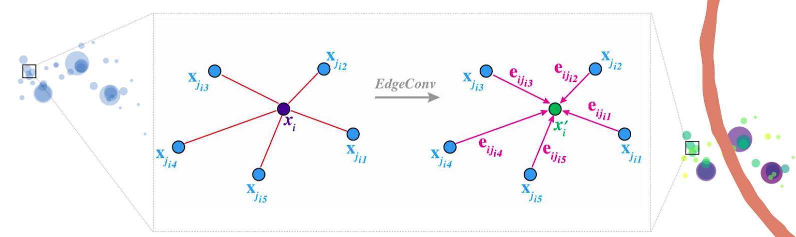


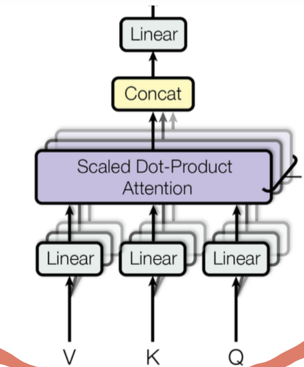
Jet models in the Transformer era (ATLAS+CMS)

feed-forward NNs (high-level inputs) · · ▸ · · *1D/2D CNNs, RNNs* (low-level inputs) · · ▸ · · *graph NNs* · · ▸ · · *Transformers* · · ▸ · · ??



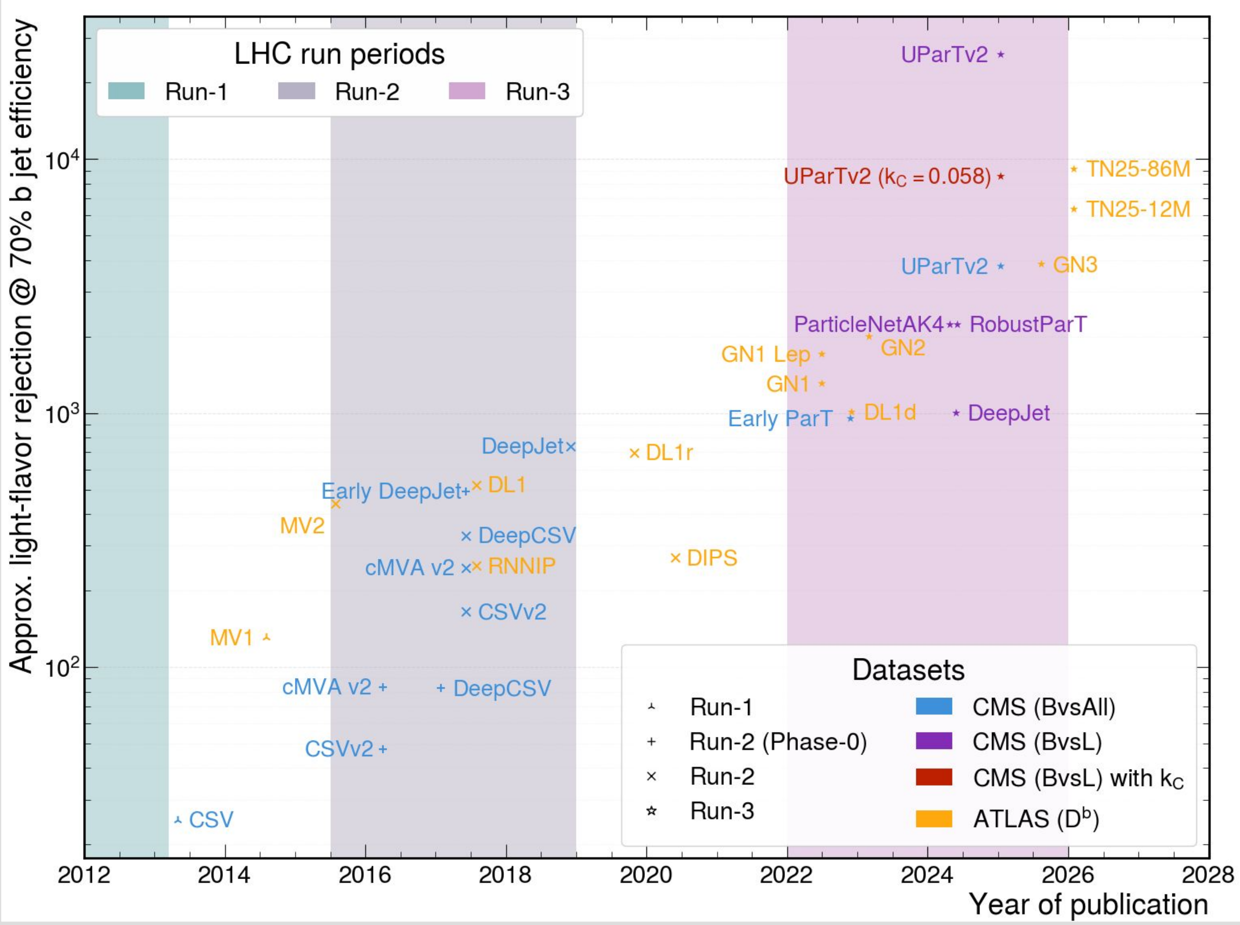






ATLAS + CMS

Big money plot [BEST TO OUR KNOWLEDGE]



Old data: 2404.01071
GN3: ATL-PHYS-PUB-2026-001
TN25: ATL-SOFT-PUB-2026-002
UParTv2: CMS-DP-2025-081

$$\mathrm{BvsAll} = \frac{\mathrm{prob(b)}}{\mathrm{prob(b)} + \mathrm{prob(c)} + \mathrm{prob(udsg)}}$$

$$\mathrm{BvsL} = \frac{\mathrm{prob(b)}}{\mathrm{prob(b)} + \mathrm{prob(udsg)}}$$

$$\mathrm{BvsC} = \frac{\mathrm{prob(b)}}{\mathrm{prob(b)} + \mathrm{prob(c)}}$$

$$\mathrm{BvsAll\ weighted} = \frac{\mathrm{prob(b)}}{k_c \cdot \mathrm{prob(c)} + (1 - k_c) \cdot \mathrm{prob(udsg)}}$$

- The light mistag for a b-jet efficiency at 70%.
- The color of markers indicated usage of different discriminators.

Is It possible to compare ATLAS/CMS b-tagging performance in one chart?

credit to Pavlo Kashko's talk in FTAG workshop 2026

✳ Take it with a grain of salt (synchronization still needs confirmation)

✳ Good thing is that CMS results are reproducible with the Delphes (fast simulation) card (backup)

Future models – what drives the frontier

Richer inputs

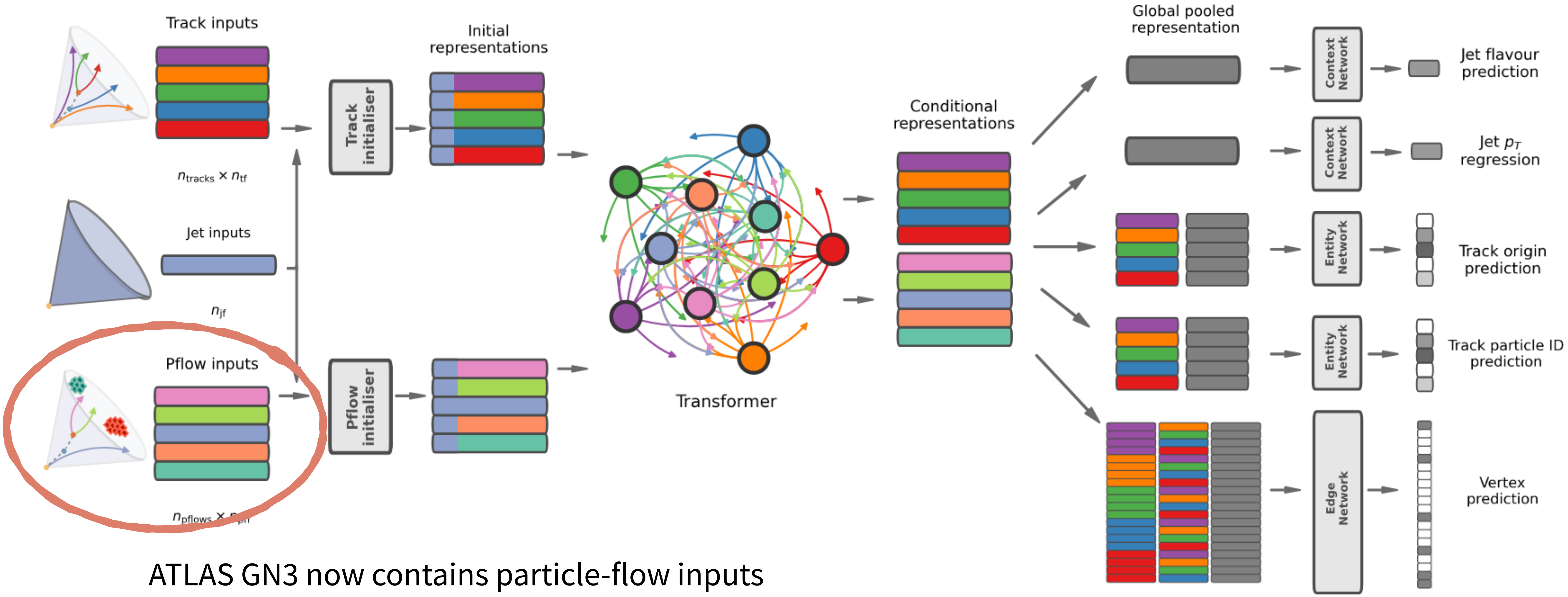
Higher token (particle) multiplicities

Scaling up

Future models – what drives the frontier (I)

Richer inputs | Higher token (particle) multiplicities | Scaling up

1. More low-level features



ATLAS GN3 now contains particle-flow inputs

- Combine information from both tracker and calorimeter; no overlap removal
- CMS jet models largely uses particle-flow inputs, with additional secondary vertices + lost tracks

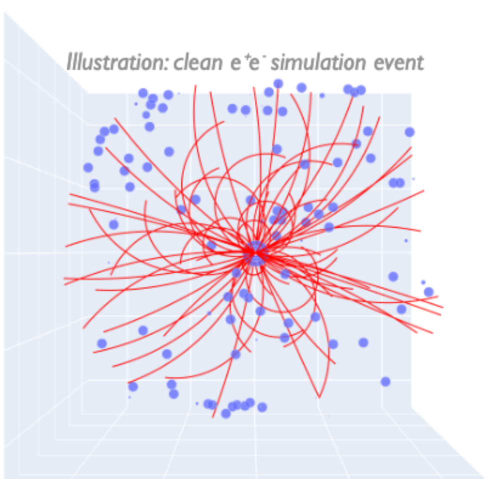
GN3: *ATL-PHYS-PUB-2026-001*
GN2: *ATLAS. Nature Commun. 17 (2026) 541*

Future models – what drives the frontier (I)

Richer inputs | Higher token (particle) multiplicities | Scaling up

2. Promising frontier: including latent features?

Input (PF elements)



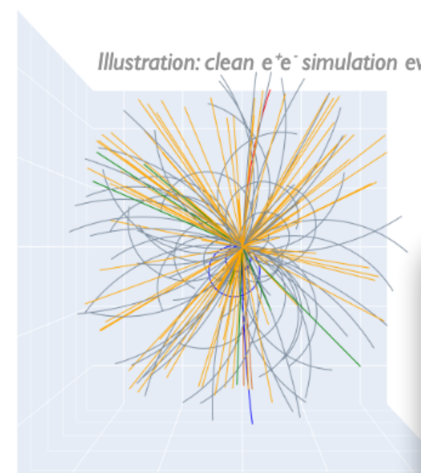
$x_{\text{track}} = (p_{\mathrm{T}}, \eta, \phi, d_0, \Omega, \ldots)$

$x_{\text{cluster}} = (E, \eta, \phi, x, y, z, \ldots)$

Transformer (set-to-set model)

Transformer architecture (≈ 5M parameters)

Output (PF candidates)



$y_i = (\text{PID}, p_{\mathrm{T}}, \eta, \phi, E)$

*Per-particle loss**

Target set of particles

- **CMS has implemented MLPF (Machine-Learned Particle Flow)**
  - reconstructs particle records directly from lower-level tracks and clusters - an ML-based PFlow
- MLPF is an upstream task for jet tagging

Image credit to Farouk Mokhtar's FTAG 2026 talk

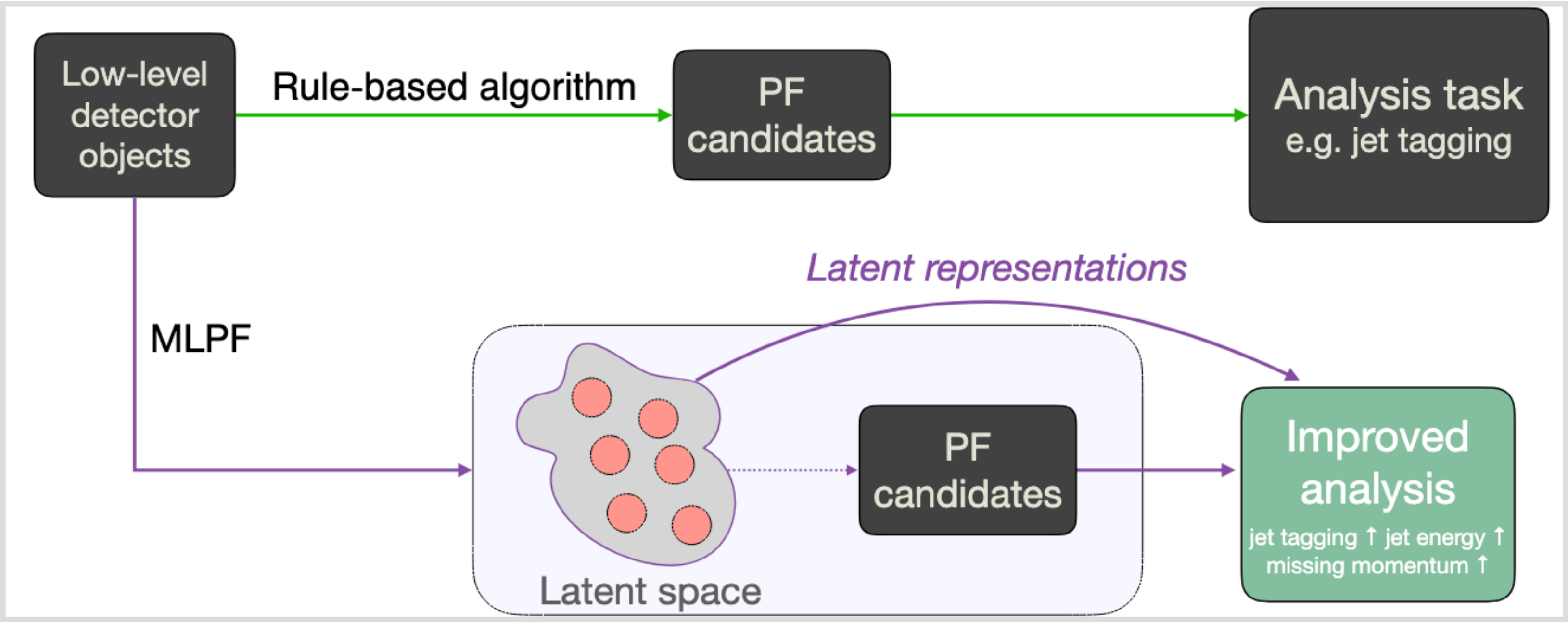
Future models – what drives the frontier (I)

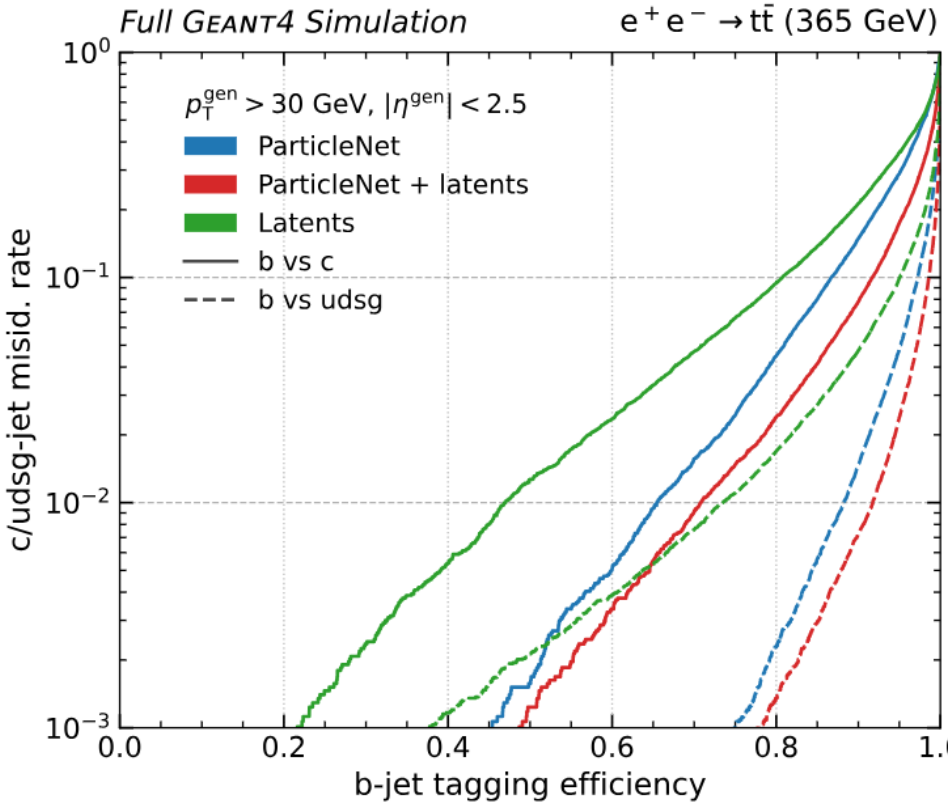
Richer inputs | Higher token (particle) multiplicities | Scaling up

2. Promising frontier: including latent features?

- Latent features from upstream models can be more expressive than hand-crafted low-level features

Mokhtar et al. 2606.14373





- Improved performance when adding latent as inputs

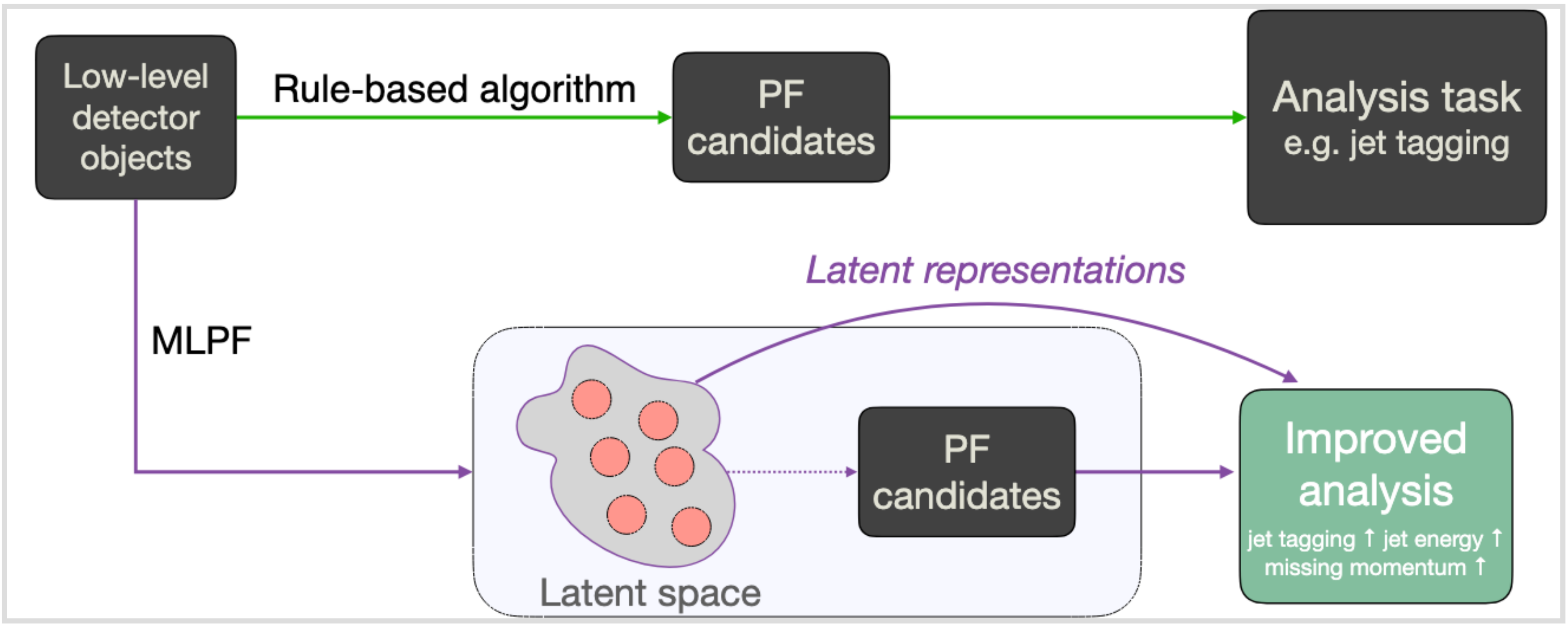
Future models – what drives the frontier (I)

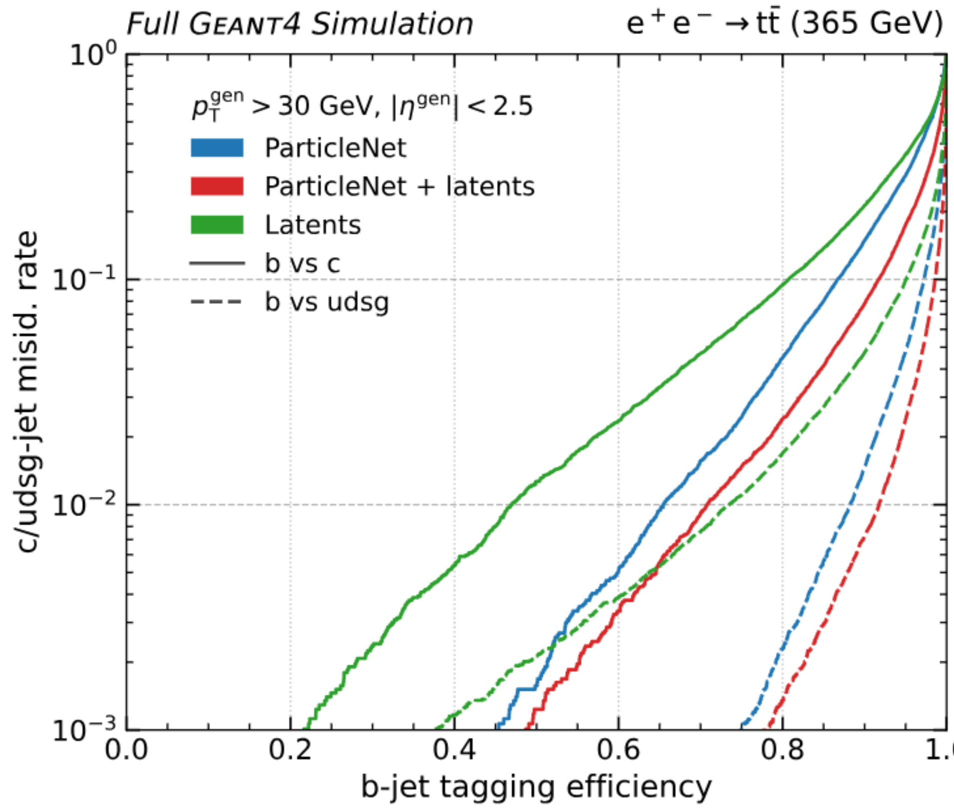
Richer inputs | Higher token (particle) multiplicities | Scaling up

2. Promising frontier: including latent features?

- Latent features from upstream models can be more expressive than hand-crafted low-level features

Mokhtar et al. 2606.14373





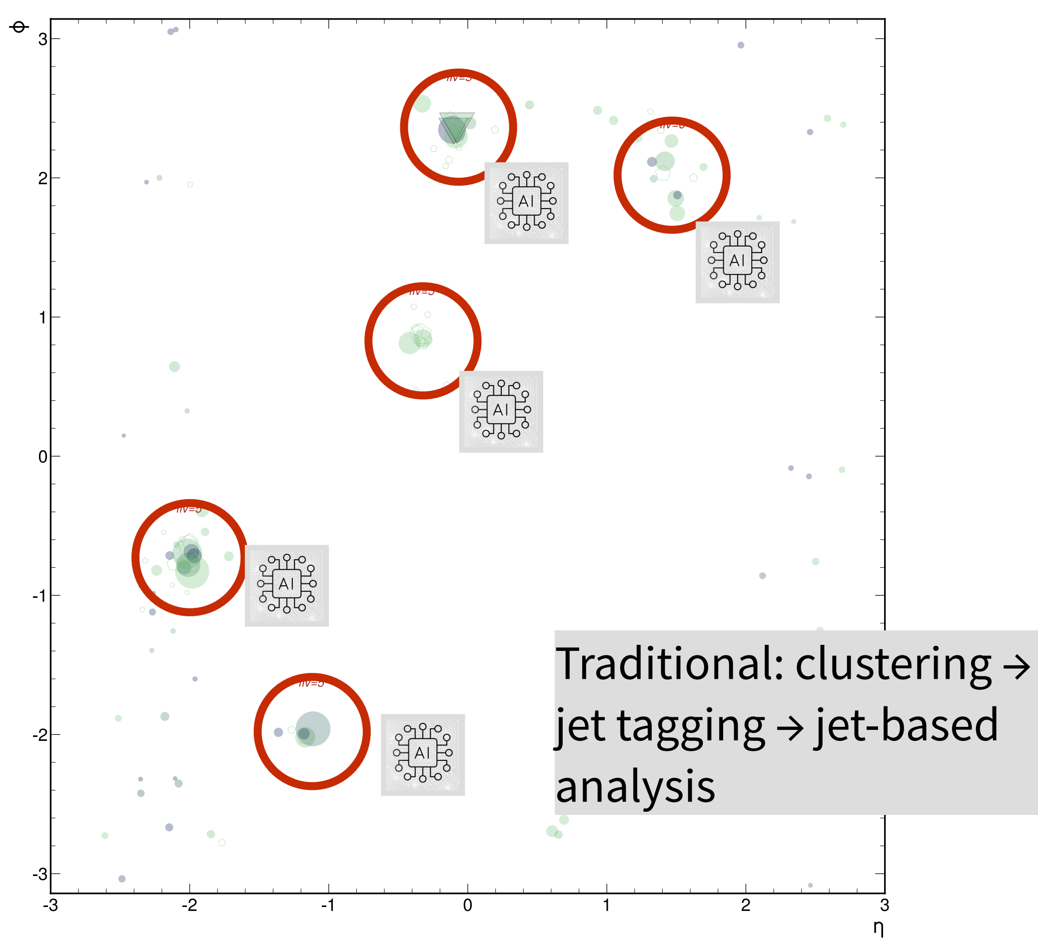
-> It's very promising, but hints that fully end-to-end training (if we dare) may ultimately perform best. An example will follow

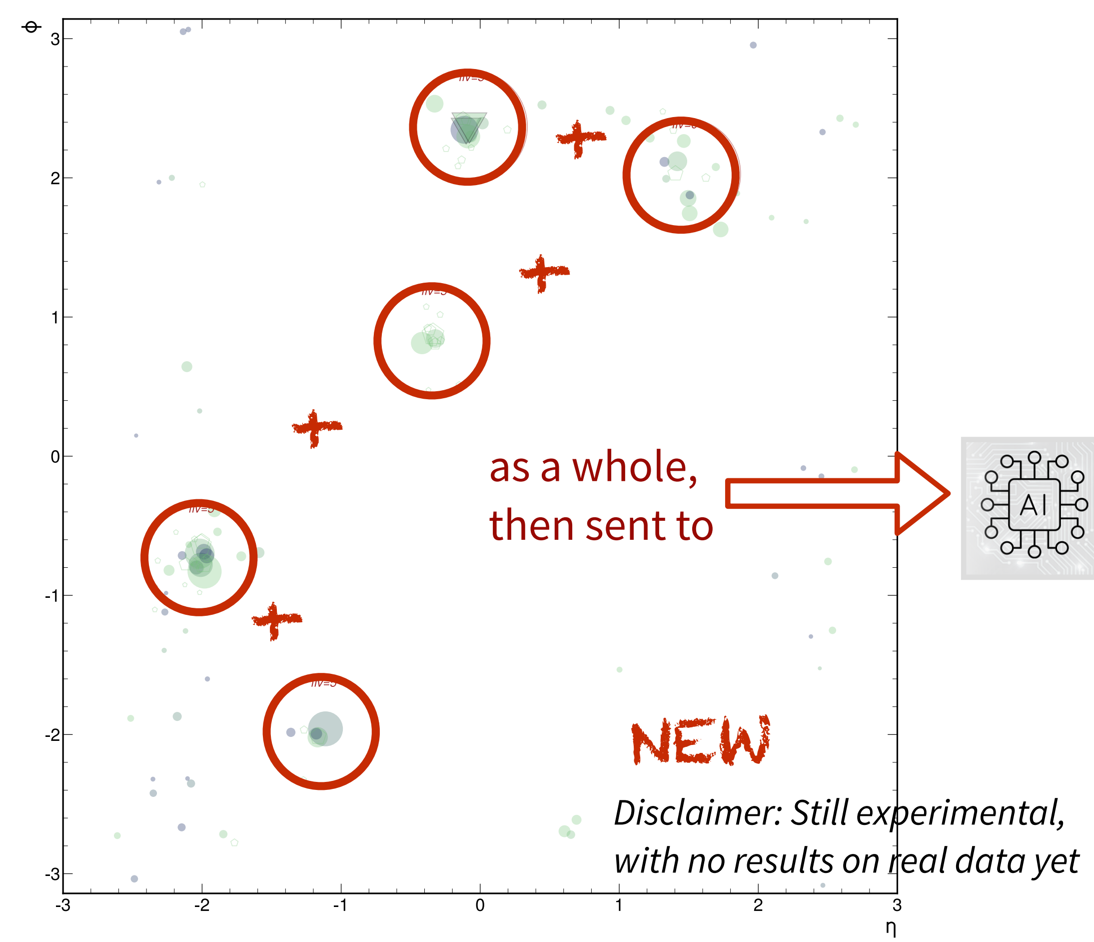
- Improved performance when adding latent as inputs

Future models – what drives the frontier (II)

Richer inputs | **Higher token (particle) multiplicities** | Scaling up

➔ Simply **breaking the "particle clustering → jet tagging → jet-based analysis" tradition** and analyzing more particles from the beginning brings huge gains





Disclaimer: Still experimental, with no results on real data yet

Future models – what drives the frontier (II)

Richer inputs | **Higher token (particle) multiplicities** | Scaling up

→ Simply **breaking the "particle clustering → jet tagging → jet-based analysis" tradition** and analyzing more particles from the beginning brings huge gains

→ Related pheno paper (hep-ex context) in this direction

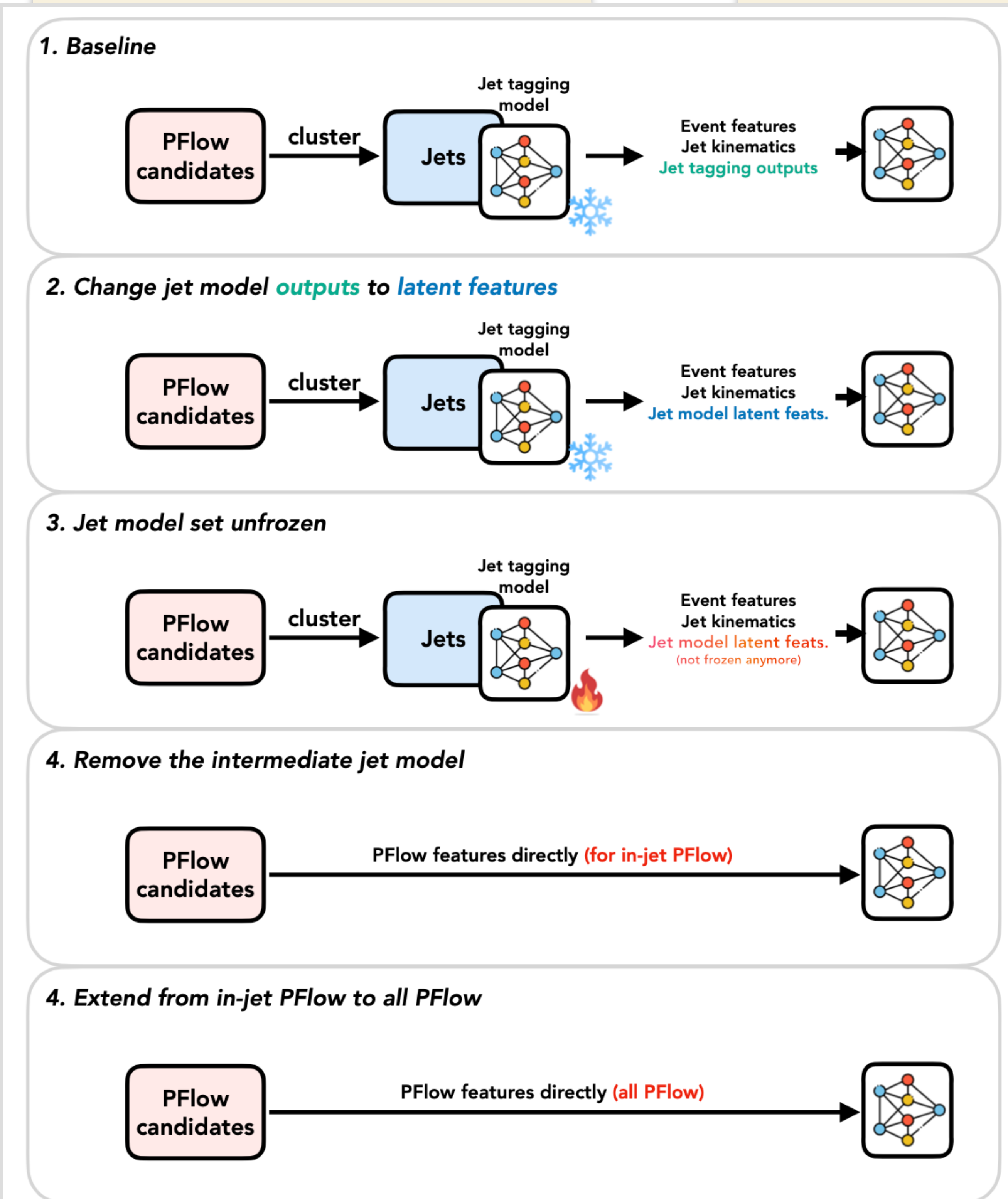
- Mondal et al. 2311.11011 (JHEP). **PAIReD & PAIReDEllipses**: process all particles within dijet/ within dijet ellipse
- Zhu, Wang, Ruan et al. 2506.11783 **"Holistic approach"** -> process particles within full events, under CEPC context
- Yang & Li. 2508.15048. **Jet-free HH4b**: process particles within full events directly
- Qureshi et al. 2601.16417. **PanopTag**: process all particles from all jets

Full-event particle inputs + **scaling** can then deliver very striking gains (e.g. ~5× in HH→4b sensitivity -> Yang & Li. 2508.15048)

Future models – what drives the frontier (II)

n (particle)
licities

Scaling up

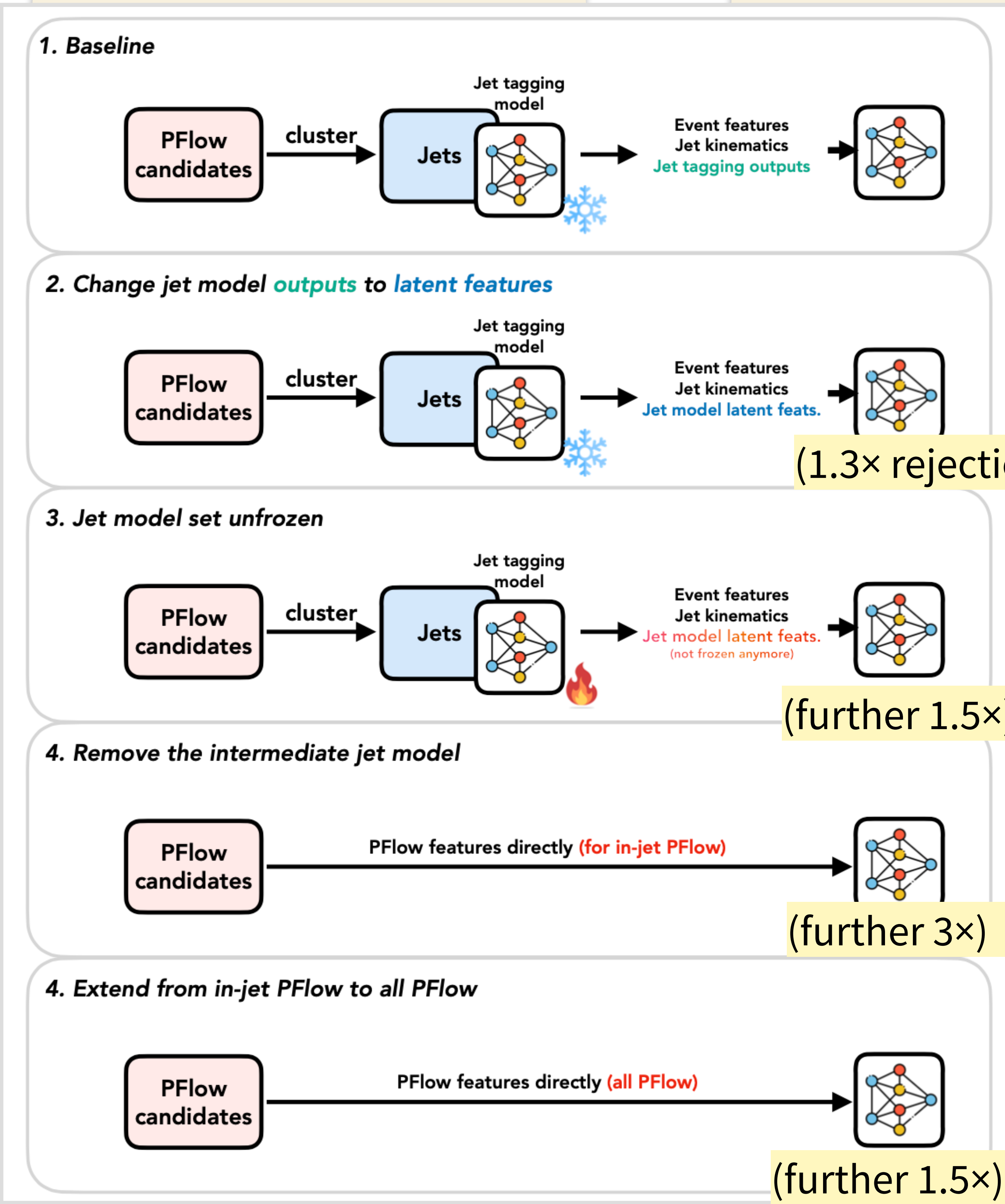


<u>“Full-event particle inputs” as the ultimate solution</u>
→ build a step-by-step comparison

Future models – what drives the frontier (II)

n (particle)
licities

Scaling up



(1.3× rejection)

(further 1.5×)

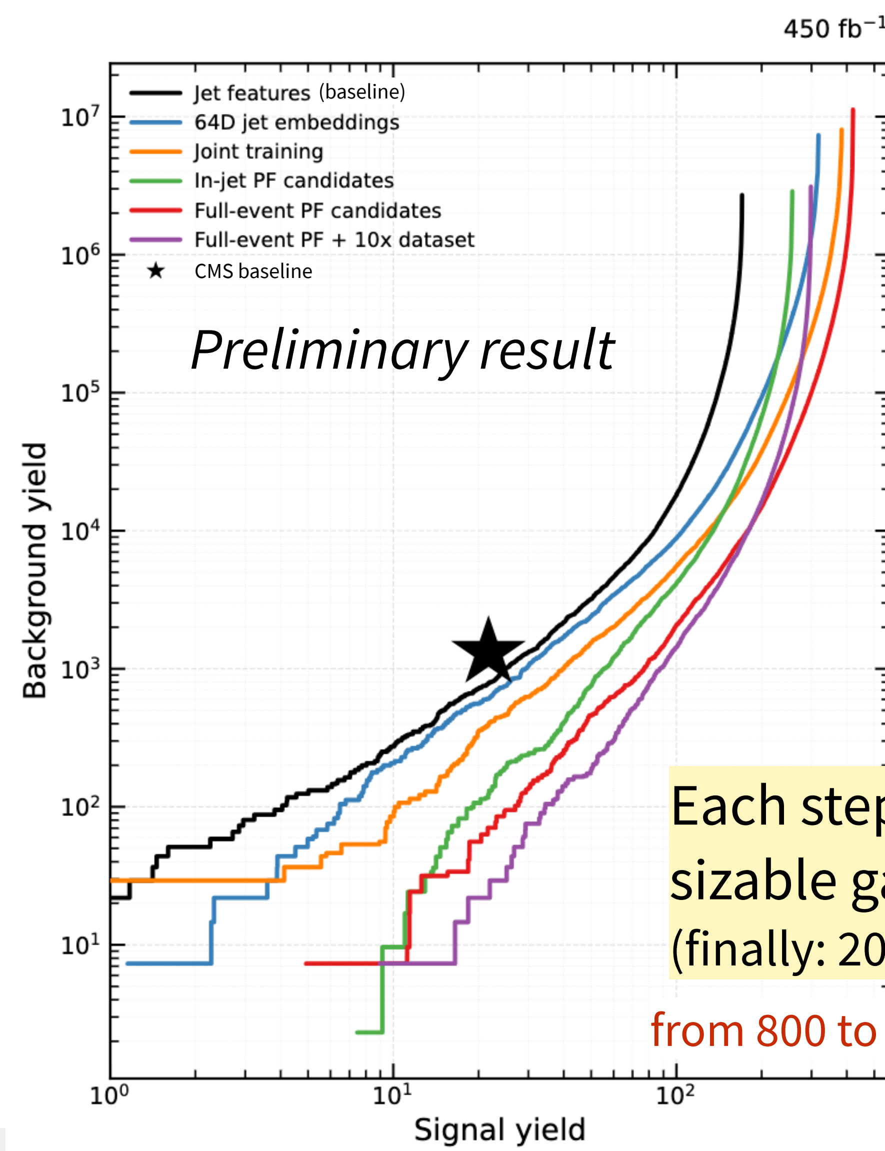
(further 3×)

(further 1.5×)

Final curve: just scale up → 2×

“Full-event particle inputs” as the ultimate solution

→ build a step-by-step comparison



Each step brings a sizable gain

(finally: 20× bkg. rejection)

from 800 to 40 backgrounds

Future models – what drives the frontier (III)

Richer inputs

Higher token (particle) multiplicities

Scaling up

- ➔ *The bitter lesson* in HEP: simply scaling up (model, data size) can beat extensive optimization of jet architectures/inputs
- ➔ Since 2026: see a growing number of studies on HEP's scaling law
  - ❖ ATL-SOFT-PUB-2026-002 (small-*R* jet tagging, ATLAS context)
  - ❖ Vigl, Hartman, Kagan, Heinrich. 2602.15781 (for large-*R* jet tagging)
  - ❖ Amram et al. 2605.28940 (for jet generation)

Basics: In the context of **Chinchilla scaling laws** (Hoffmann et al. (DeepMind), 2022)

$$L(N, D) = L_\infty + \frac{a}{N^\alpha} + \frac{b}{D^\beta}$$

Future models – what drives the frontier (III)

Richer inputs | Higher token (particle) multiplicities | **Scaling up**

- *The bitter lesson* in HEP: simply scaling up (model, data size) can beat extensive optimization of jet architectures/inputs
- Since 2026: see a growing number of studies on HEP's scaling law
  - ATL-SOFT-PUB-2026-002 (small-*R* jet tagging, ATLAS context)
  - Vigl, Hartman, Kagan, Heinrich. 2602.15781 (for large-*R* jet tagging)
  - Amram et al. 2605.28940 (for jet generation)

Basics: In the context of **Chinchilla scaling laws** (Hoffmann et al. (DeepMind), 2022)

$$L(N, D) = L_\infty + \frac{a}{N^\alpha} + \frac{b}{D^\beta}$$

Help us understand:

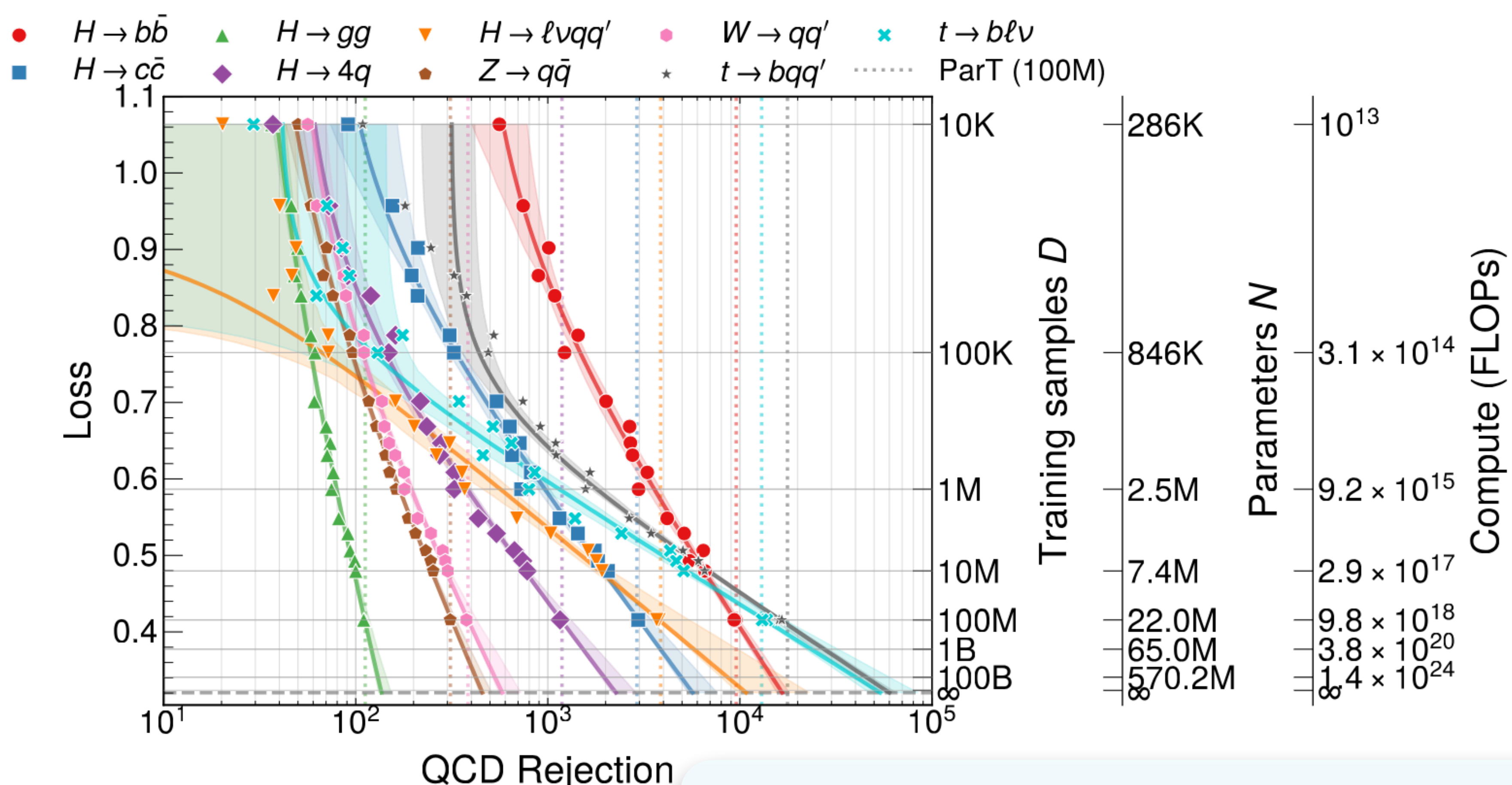
- Where is the irreducible loss floor L_∞? → How far are we from the theoretical sensitivity limit?
- How should the optimal model and dataset sizes scale?
  (this goes beyond Chinchilla law due to multi-epoch training effects, but has been studied in these papers)

Future models – what drives the frontier (III)

Richer inputs

Higher token (particle) multiplicities

Scaling up



Vigl, Hartman, Kagan, Heinrich. 2602.15781

- **What we possibly see:**

1. State-of-the-art ATLAS/CMS training is already at O(100M) events, yet datasets can still scale further
2. Our ability to scale models may lag behind our ability to scale data — potentially closely related to observations from foundation models (see next)

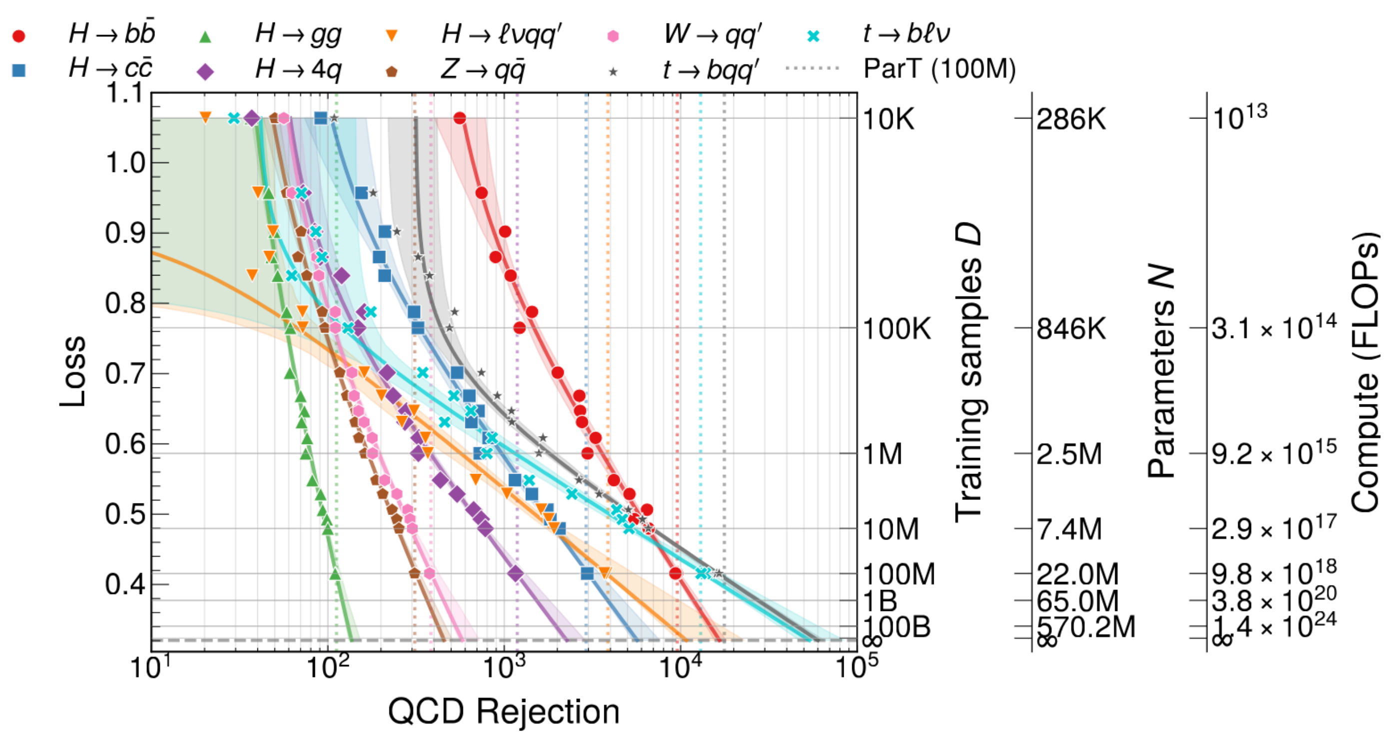
Future models – what drives the frontier (III)

Richer inputs

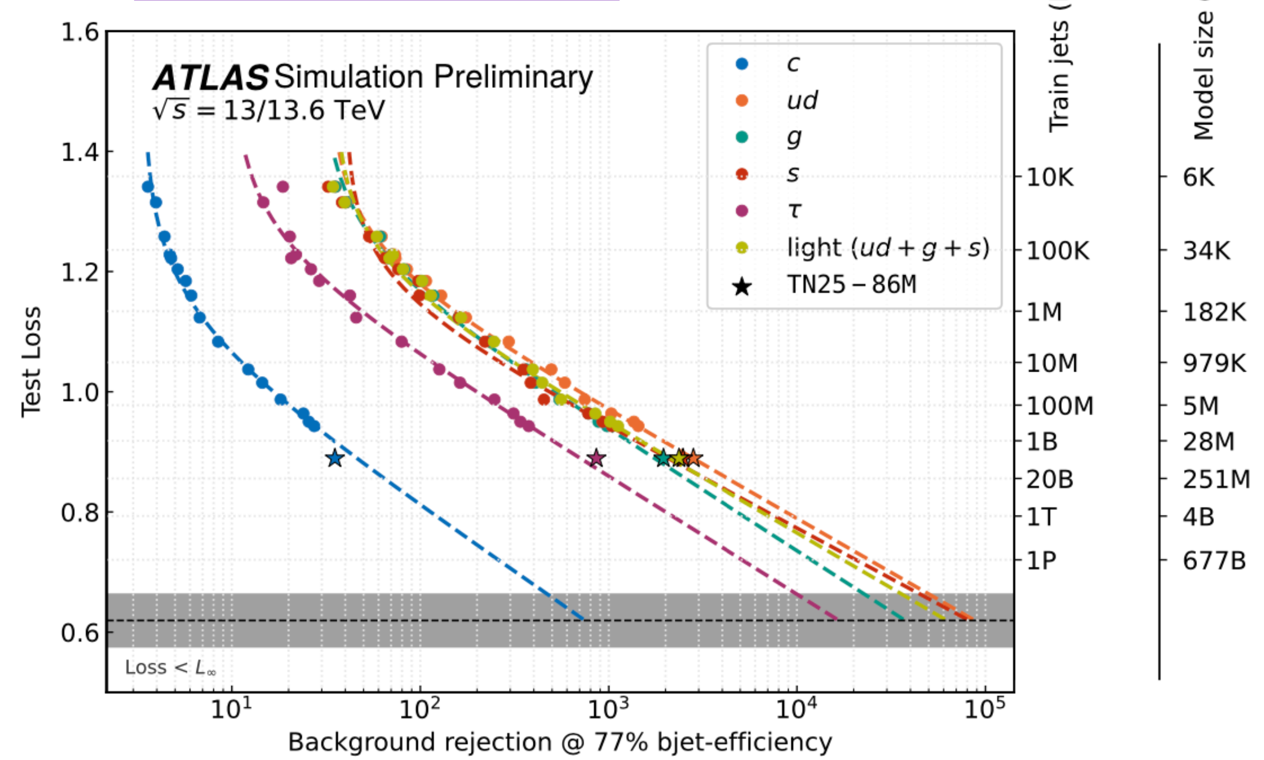
Higher token (particle) multiplicities

Scaling up

Vigl, Hartman, Kagan, Heinrich. 2602.15781



ATL-SOFT-PUB-2026-002



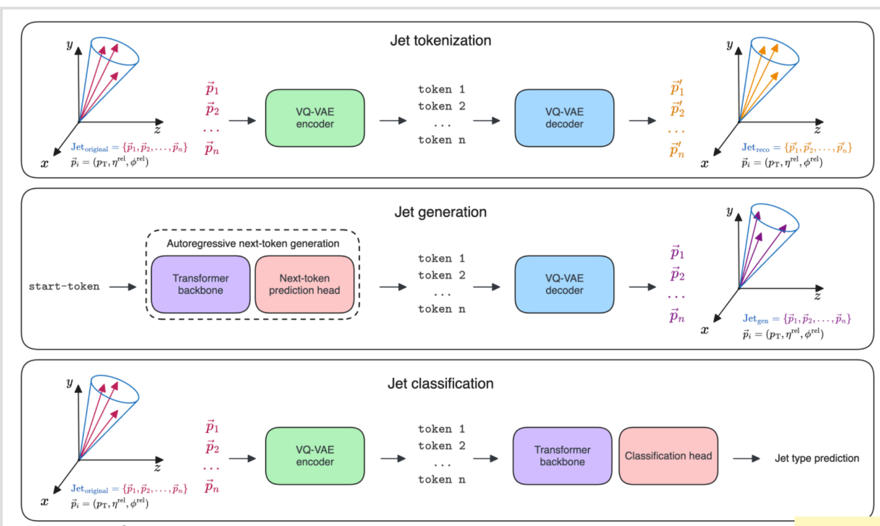
- **What we possibly see:**

1. State-of-the-art ATLAS/CMS training is already at O(100M) events, yet datasets can still scale further
2. Our ability to scale models may lag behind our ability to scale data — potentially closely related to observations from foundation models (see next)
3. **Comparing ATLAS (right) vs fast sim (left): we are far away from saturation; is fast-sim becoming the bottleneck?**

Future models – what drives the frontier (III)

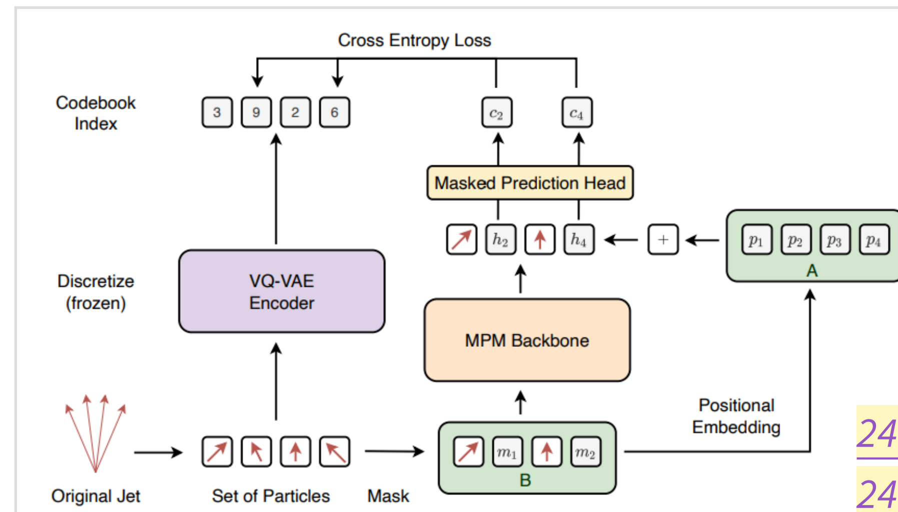
Richer inputs | Higher token (particle) multiplicities | **Scaling up**

➔ On the "Foundation Model" Fever (typical examples)



OmniJet-α 2403.05618 (MLST)

(GPT-like, next-token prediction to learn jet properties)

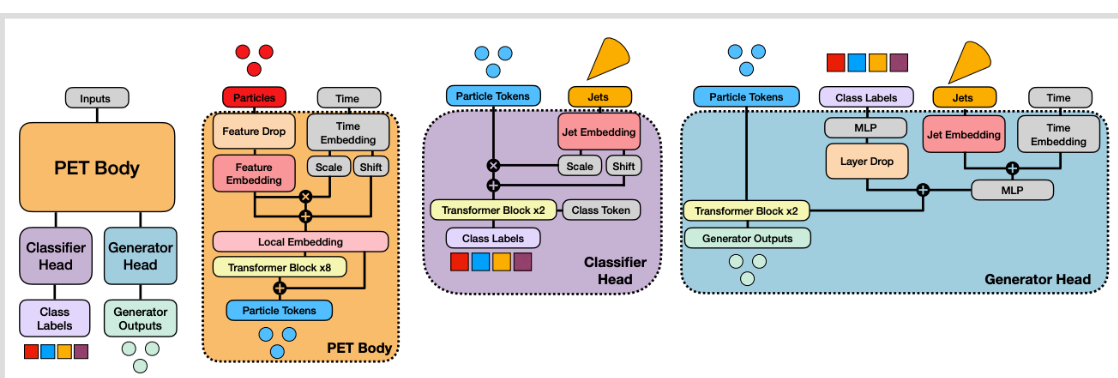


2401.13537 (MLST)

2409.12589 (MLST)

Masked Particle Modelling

(SSL with Masked autoencoder (MAE) style)



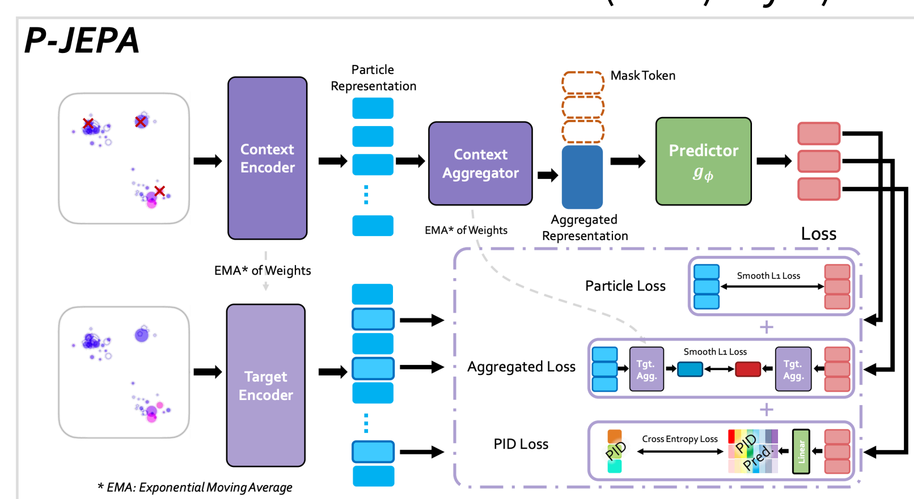
OmniLearn + OmniLearned

(trained on simultaneous classification + diffusion-based generative tasks)

2404.16091 (PRD)

2502.14652 (PRD)

2510.24066 (PRD)



p-jepa jet embedding prediction (particle-level)

see e.g. H. Qu's talk

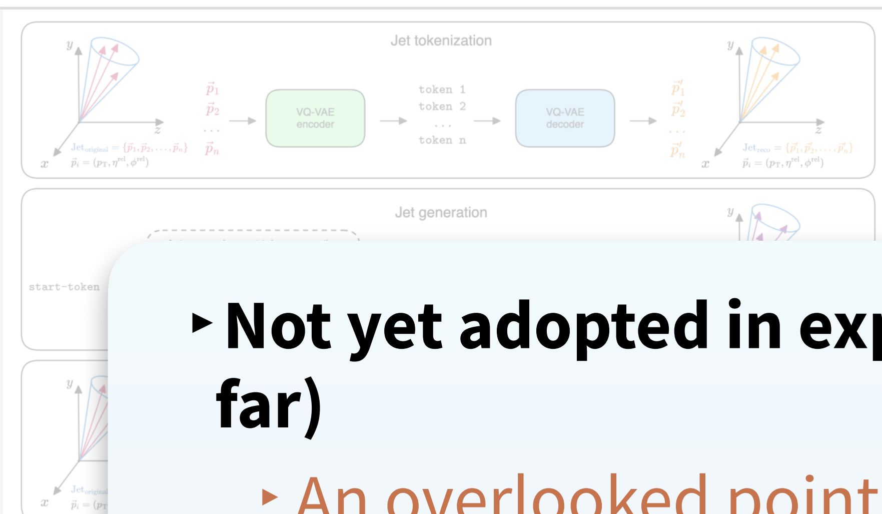
Future models – what drives the frontier (III)

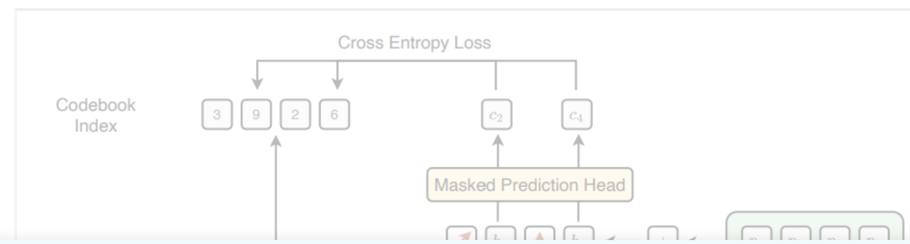
Richer inputs

Higher token (particle) multiplicities

Scaling up

➔ On the "Foundation Model" Fever (typical examples)





- **Not yet adopted in experiments (with no compelling motivation so far)**
  - An overlooked point: ATLAS/CMS already train on O(100M) samples, pushing "pure classification" close to the limits
  - A different story in computer vision: on ImageNet (~1M images), self-supervised methods emerged partly to overcome weak "pure classification" baselines
  - In HEP: find model scaling harder, but data scaling much easier

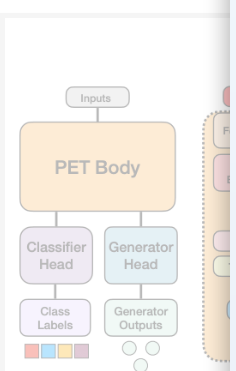
Future direction is unclear, but this reflection may give a hint

Omn

(GPT-

LST)

LST)



Om

(trained on simultaneous classification + diffusion-based generative tasks)

2502.14652 (PRD)

2510.24066 (PRD)

p-jepa jet embedding prediction (particle-level)

see e.g. H. Qu's talk

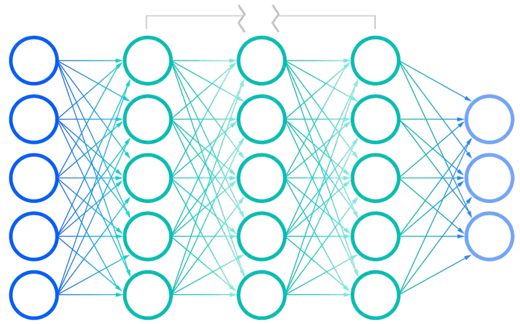
Conclusion

- ➔ This talk highlights three recent insights into how we build models
  - ❖ Not covering all AI topics, but these ideas may generalize (jet model, end-to-end model, PFlow reconstruction, scaling behavior…)
- ➔ **Richer inputs:** connect upstream & downstream models in HEP — has become important now
- ➔ **Higher multiplicities:** time to seriously explore game-changing paradigms (e.g. analysis with full event particles)
- ➔ **Scaling capability:** help us cut through the fog: where is the final sensitivity limit, and how do we reach it faster?

Backup

Evolution of jet NNs

feed-forward NN *(high-level inputs)* ▶ *1D/2D CNN, RNN* *(low-level inputs)* ▶ *graph NN, Transformers* *(low-level inputs)* ▶ *??*

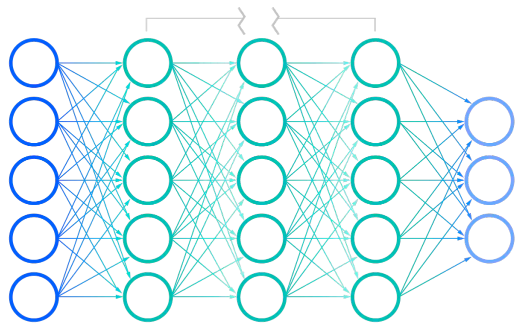


Shallow networks

✦ Using high-level features directly as input to a shallow network

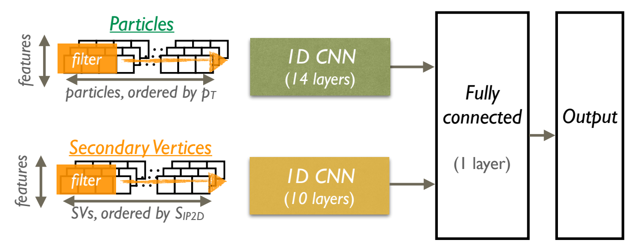
Evolution of jet NNs

feed-forward NN *(high-level inputs)* ▶ *1D/2D CNN, RNN* *(low-level inputs)* ▶ *graph NN, Transformers* *(low-level inputs)* ▶ *??*



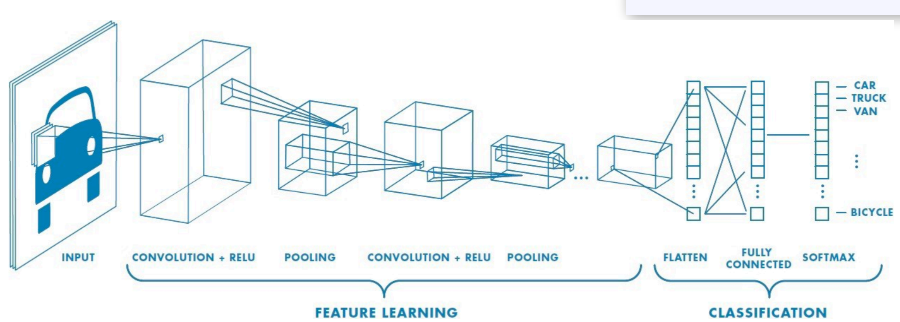
Shallow networks

- ✦ Using high-level features directly as input to a shallow network

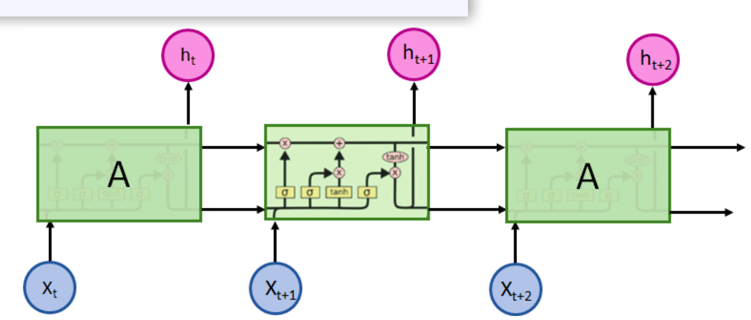


Deep NN with low-level inputs

- ✦ Using particle-level features
- ✦ Input data structure determines the type of networks
  - jet as a *image (fixed-grid data structure)*
  - jet as a *sequence* → 1D CNN or RNN



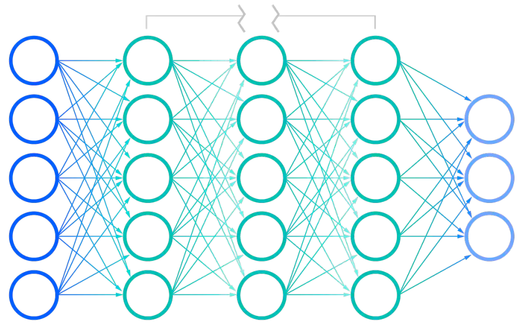
Typical CNN



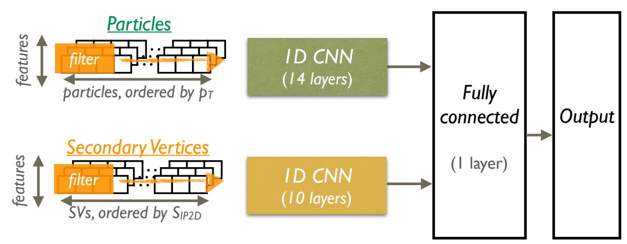
Typical RNN

Evolution of jet NNs

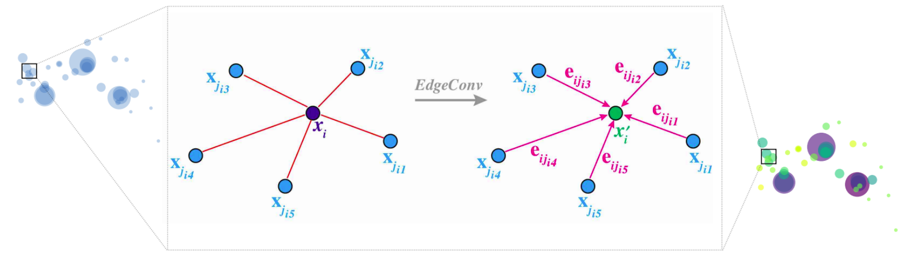
feed-forward NN (high-level inputs) ▶ *1D/2D CNN, RNN (low-level inputs)* ▶ *graph NN, Transformers (low-level inputs)* ▶ *??*



Shallow networks



Deep NN with low-level inputs



Graph structure

✦ Using high-level features directly as input to a shallow network

✦ Using particle-level features

✦ Input data structure determines the type of networks

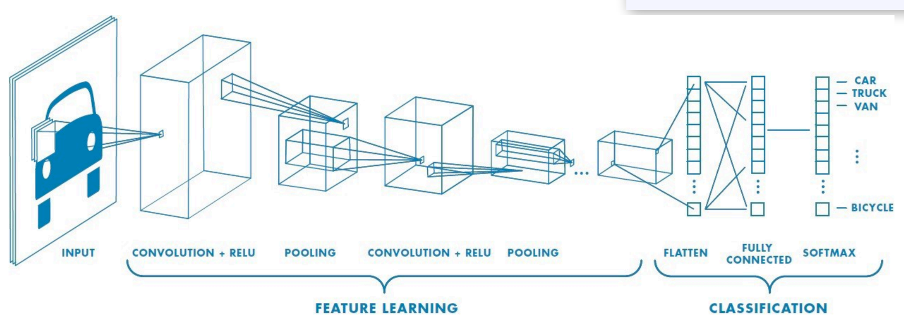
- jet as a *image (fixed-grid data structure)*
- jet as a *sequence* → 1D CNN or RNN

✦ Graph neural networks

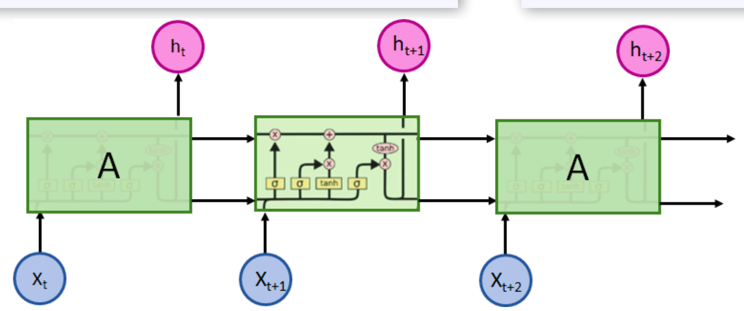
- treat a jet as a permutational-invariant set of particles (or, point cloud)
- build "edges" between particles

✦ Transformer networks

- modern architectural designs; like a full-connected graph



Typical CNN

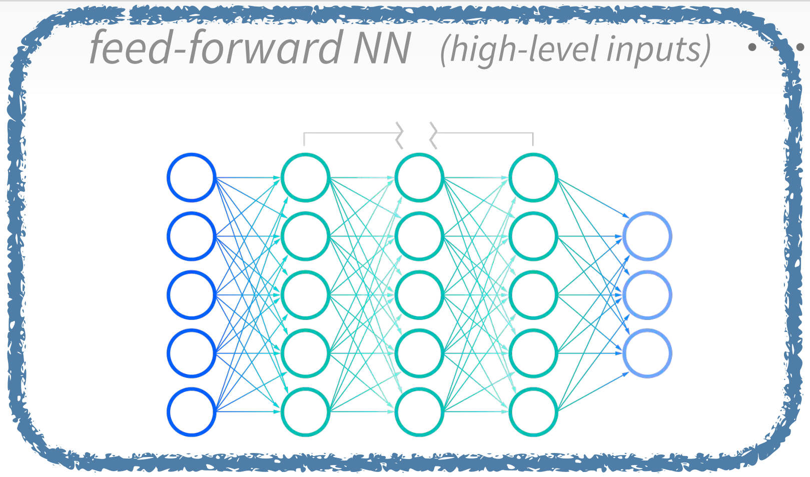


Typical RNN

Typical graph

Evolution of jet NNs (CMS)

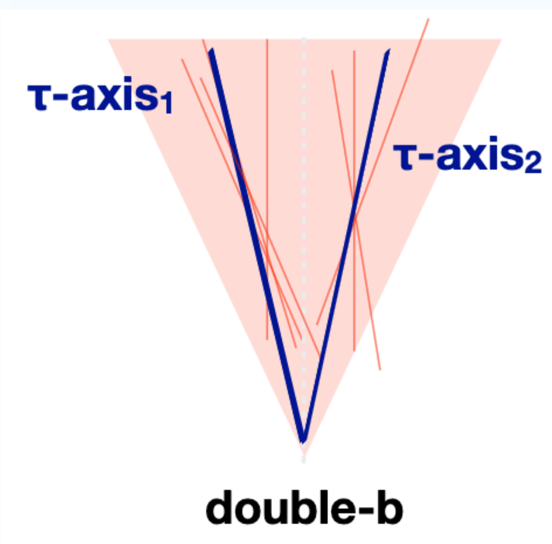
feed-forward NN (high-level inputs) → *1D/2D CNN, RNN (low-level inputs)* → *graph NN, Transformers (low-level inputs)* → ??



Shallow networks (BDTs)

Deep NN with low-level inputs

Graph structure



Double-b tagging: an early phase (~2016–17) of H→bb tagger development

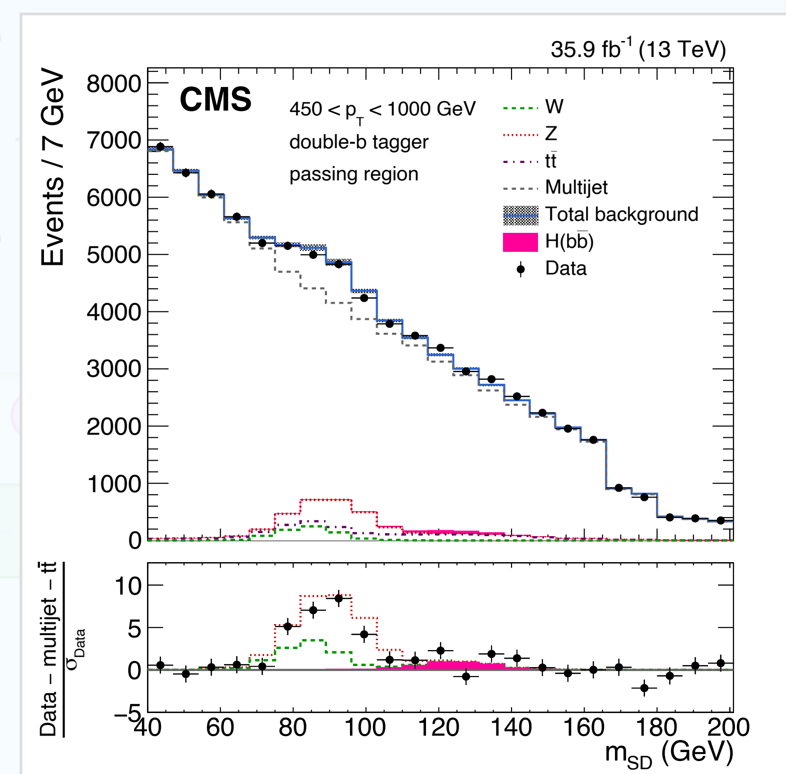
JINST 15 (2020) P06005

A BDT trained with high-level jet inputs

- selected tracks and SVs + features for the two-SV system

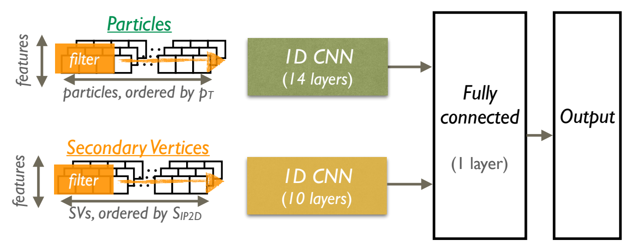
The first boosted H→bb search at the LHC

PRL 120 (2018), 071802

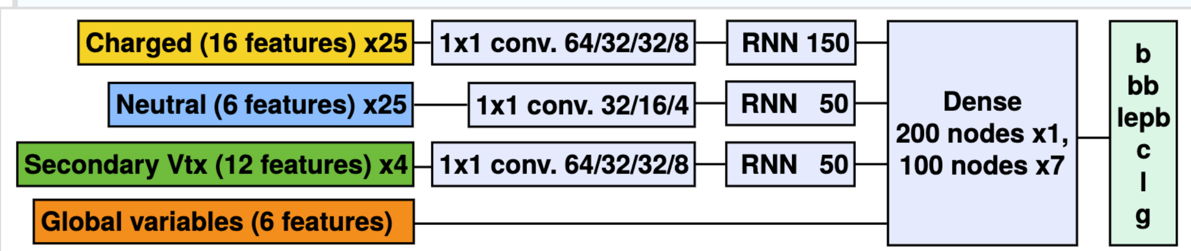


Evolution of jet NNs (CMS)

··· 1D/2D CNN, RNN (low-level inputs) ···▸··· ??



Deep NN with low-level inputs



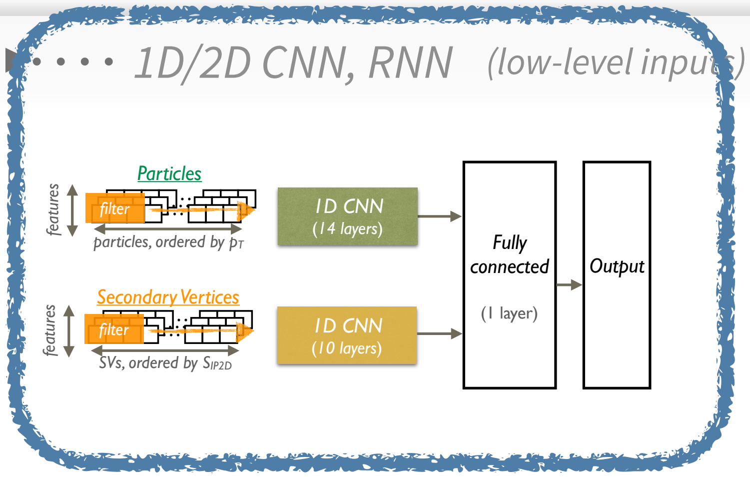
DeepJet (~2019–23):

- widely used small-R jet tagger (b/c flavour tagger) in Run 2
- processing PF candidates + SVs + global variables

JINST 15 (2020) P12012

Evolution of jet NNs (CMS)

1D/2D CNN, RNN (low-level inputs) ??

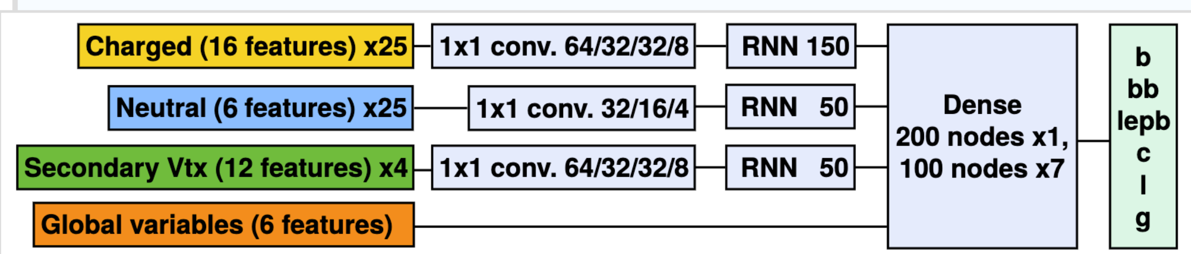


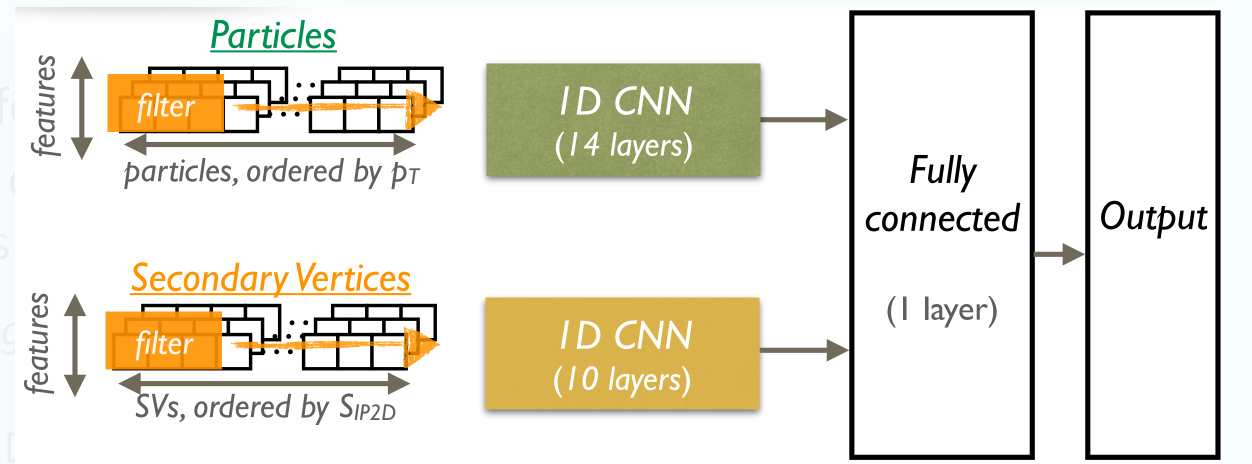
Deep NN with low-level inputs

JINST 15 (2020) P06005

DeepAK8 (~2017–20):

- early unified top/W/Z/H tagger
- processing PF candidates + SVs via 1D CNN

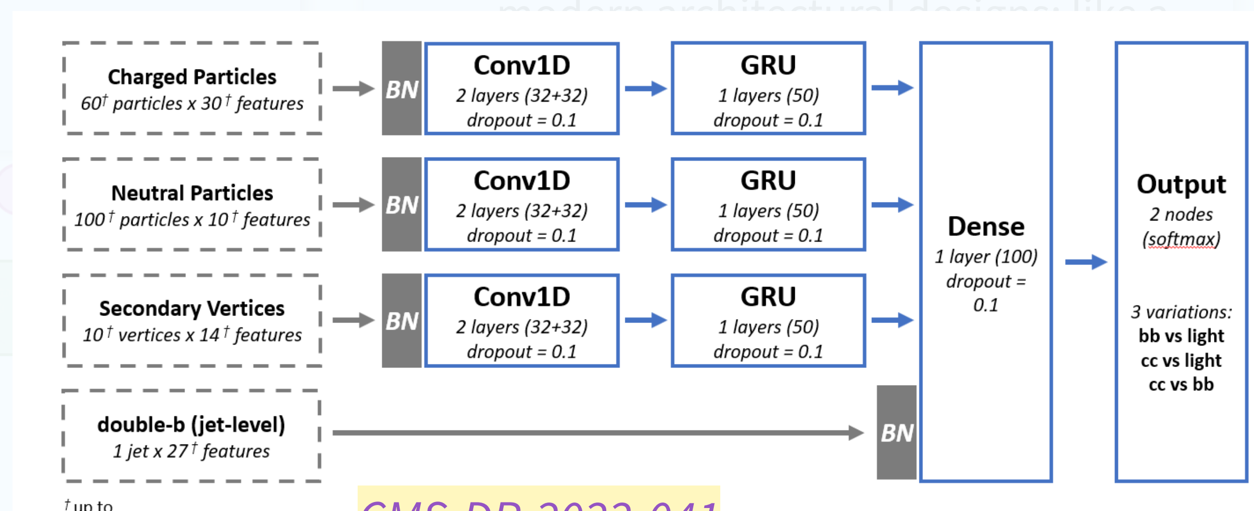




DeepJet (~2019–23):

- widely used small-R jet tagger (b/c flavour tagger) in Run 2
- processing PF candidates + SVs + global variables

JINST 15 (2020) P12012



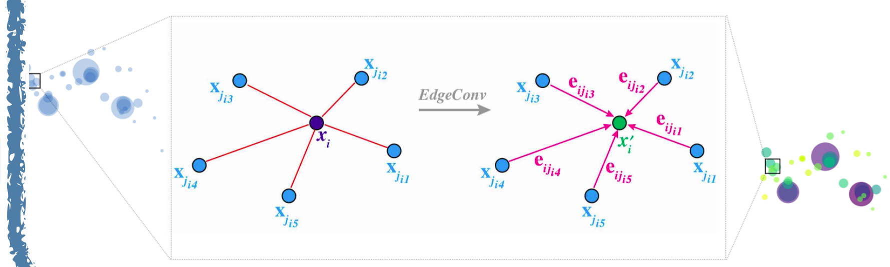
DeepDoubleX (~2018–22):

- upgraded from double-b for H→bb/cc tagging
- 1D CNN + RNN to process low-level candidates

CMS-DP-2022-041
CMS-PAS-BTV-22-001

Evolution of jet NNs (CMS)

feed-forward NN (high-level inputs) ····▶···· *1D/2D CNN, RNN (low-level inputs)* ····▶···· *graph NN, Transformers (low-level inputs)* ····▶···· *??*



Shallow networks

- ✦ Using high-level features directly as input to a shallow network

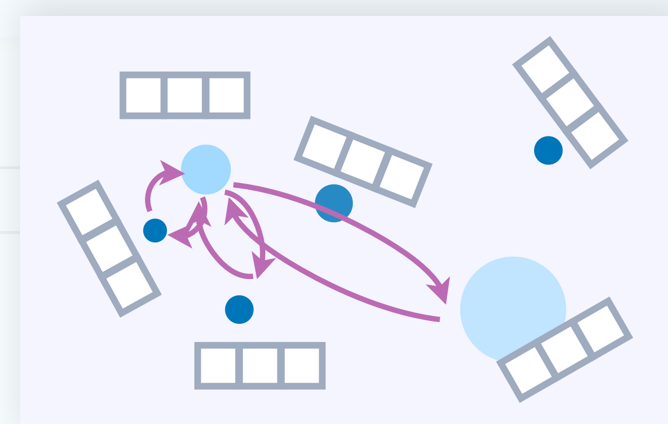
Deep NN with low-level inputs

- ✦ Using particle-level features
- ✦ Input data structure determines the type of networks
  - jet as a *image (fixed-grid data structure)*
  - jet as a *sequence* → 1D CNN or RNN

- ✦ Graph neural networks
  - treat a jet as a permutational-invariant set of particles (or, point cloud)
  - build "edges" between particles
- ✦ Transformer networks
  - modern architectural designs; like a full-connected graph

Typical CNN

Typical RNN



Global Particle Transformer

CMS-DP-2026-104

v1

v2

v3

ParT for H $\rightarrow$ WW (GloParT v1) *(development period: Sept. 2022–May 2023)*

The first-generation GloParT model was the first ParT-based tagging model, originally developed for boosted H $\rightarrow$ WW tagging, as described in Ref. [11]. The model was trained on 37 jet categories with a unified jet-mass regression output. It is designed to target variable-mass resonances decaying to $\mathrm{b\bar{b}}$, $\mathrm{c\bar{c}}$, $\mathrm{q\bar{q}}$, and $\tau\tau$, as well as H $\rightarrow$ WW signatures with fully hadronic (4q) and semileptonic ($\ell\nu$qq) final states spanning different flavour compositions, including top-quark-like decay topologies. Analyses employing this tagger include Refs. [11, 16, 17, 18].

GloParT v2 *(development period: May 2023–Dec. 2023)*

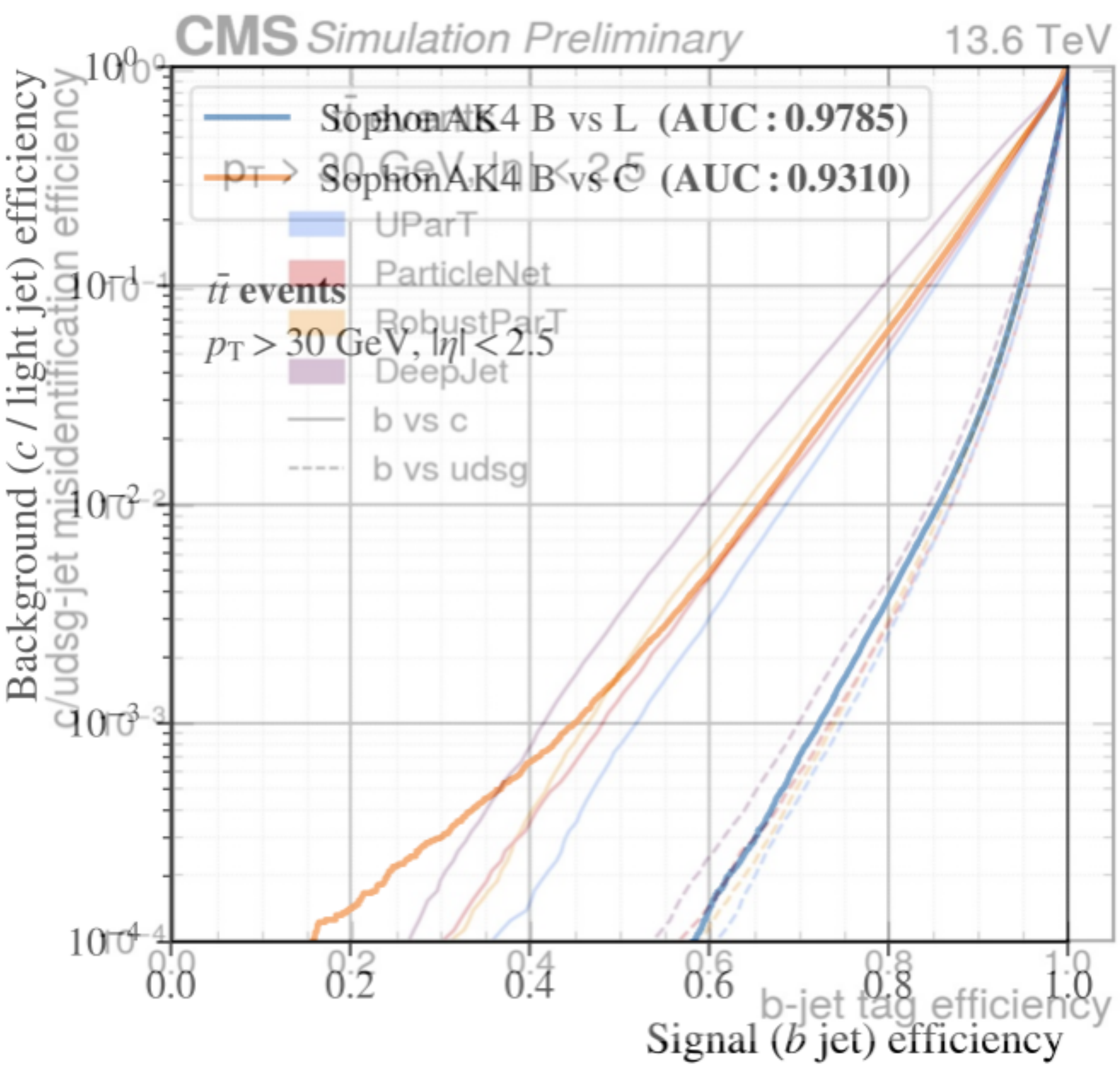
The second-generation model was developed following the large-scale classification paradigm for global jet tagging and jet-mass regression. The training dataset was substantially expanded, particularly for H $\rightarrow$ WW-like and H $\rightarrow$ ZZ-like four-prong signatures, with a significantly refined categorization scheme. The total number of jet categories was increased to 314. In addition, the input feature representation was optimized to improve model performance. Analyses using this model include Ref. [19].

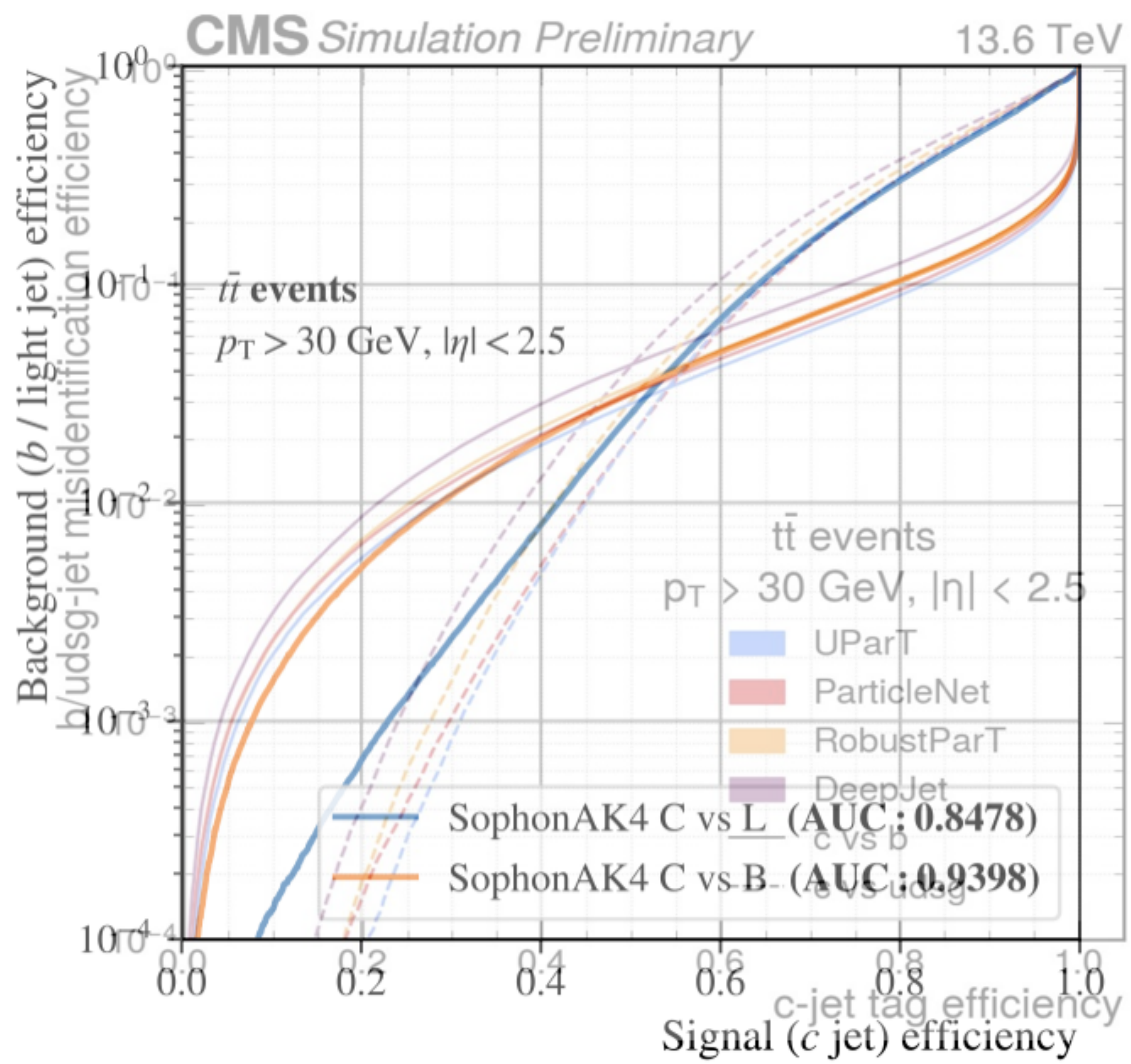
GloParT v3 *(development period: June 2024–Dec. 2024)*

The third-generation model further extended the resonance mass coverage from the previous upper limit of 250 GeV to 650 GeV, together with a substantial increase in the size of the training dataset. Several two-prong categories were further refined to distinguish the charge of charged resonances, and additional decay categories containing photons in two-, three-, and four-prong topologies were introduced, resulting in a total of 374 classes. The model implementation and hyperparameter configuration were also further optimized. Another major update was the redesign of the jet-mass regression strategy to provide class-specific regression outputs, leading to improved performance of the regression nodes. The model has been deployed within the CMS analysis workflow, and current analyses employing this model include Ref. [20].

CMS flavor tagging reproduction

➔ CMS b/c-tagging (advanced DNN based) performance is reproducible using fast simulation





(1) see SophonAK4 Huggingface release 🤗

(2) see the Appendix A here (both Sophon (for AK8) + SophonAK4)