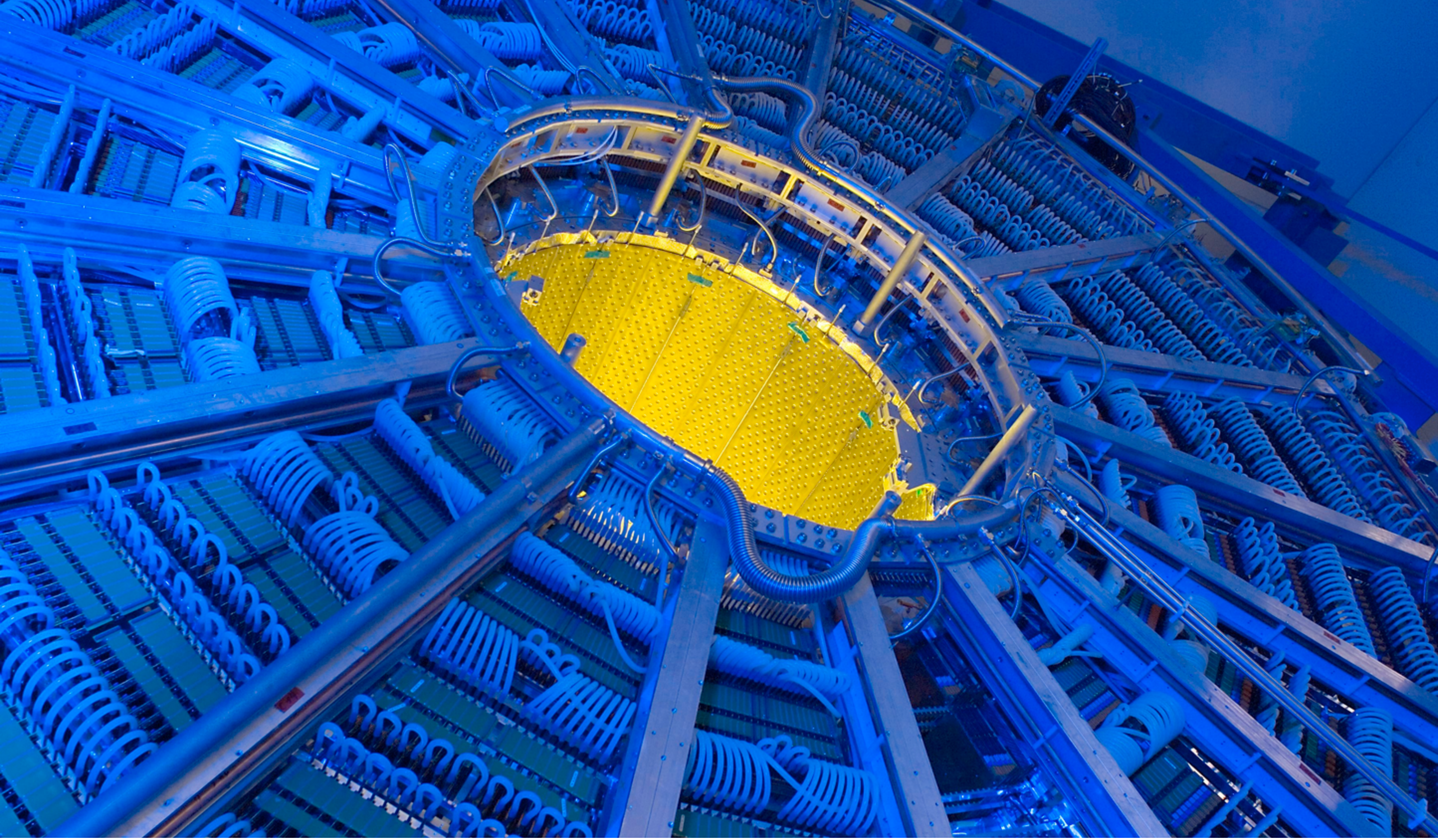


Deep Learning In Low Signal/Background Environments

Michele Floris (CERN)
OpenLab ML&DA workshop
April 27, 2017

Intro

- **Deep neural networks** → breakthroughs in a number of classification and regression problems (e.g. images)
- Physics analysis routinely deals with **classification problems**
- **Non-deep machine learning** often used in these cases (based on high-level expert features)
- Obvious questions: can we **improve significantly** w/ Deep Learning? (representation learning?)
- Some **attempts in the literature**, initially mostly toy but getting more realistic
  - To my knowledge, deep learning not yet applied in published physics analyses
  - Need input from data scientists / industry!
- Discussed here: examples from the literature and from the LHC experiments
  - **Jet Classification** (based on images or tracks)
  - $\boldsymbol{\Lambda_c} \rightarrow \pi K p$
  - **Beyond toy** studies

DISCLAIMER: Not a comprehensive review, examples biased towards my interests

Low signal/background

Many **signals of interests** are (relatively) **rare**

Embedded in events with **large number of particles**

Low signal/background

Many **signals of interests** are (relatively) **rare**

Embedded in events with **large number of particles**

Pile-up: within a single "bunch crossing" multiple collisions are possible

In run 2 average number of collisions per bunch crossing μ: 20 → 50

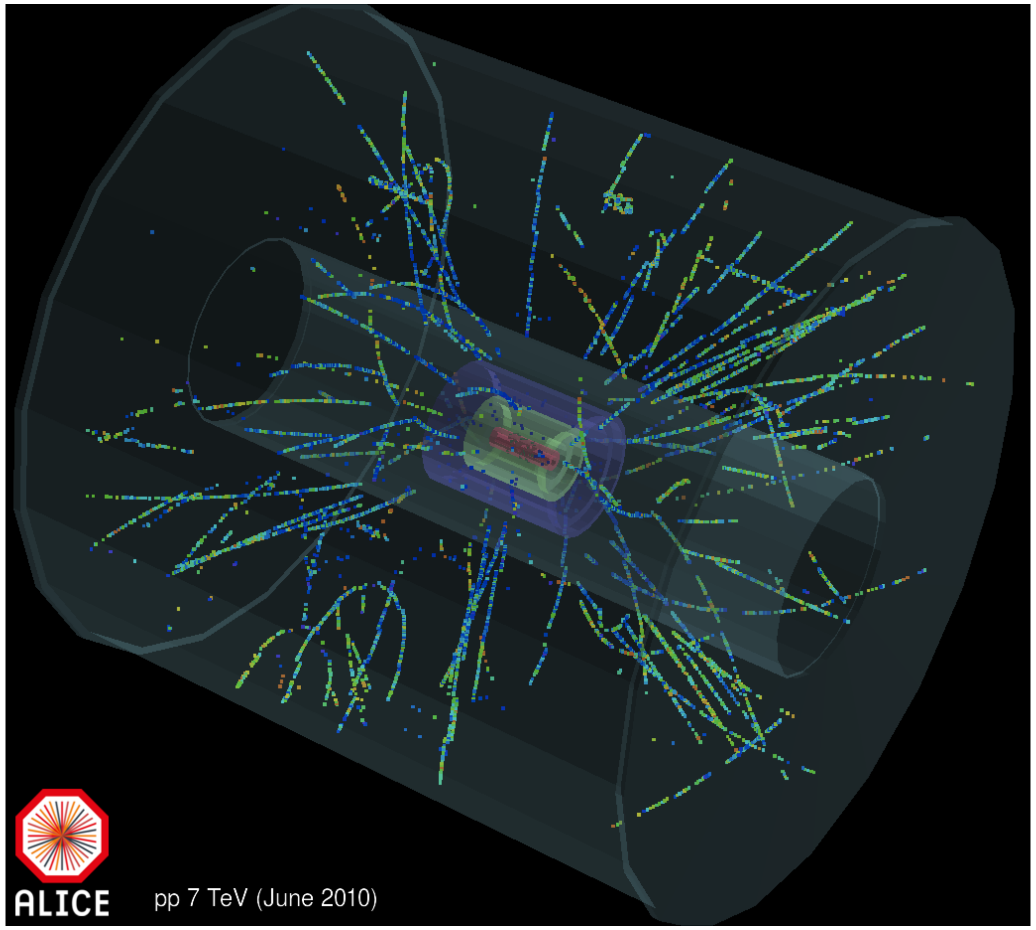
After the upgrades up to $\mu = 140$

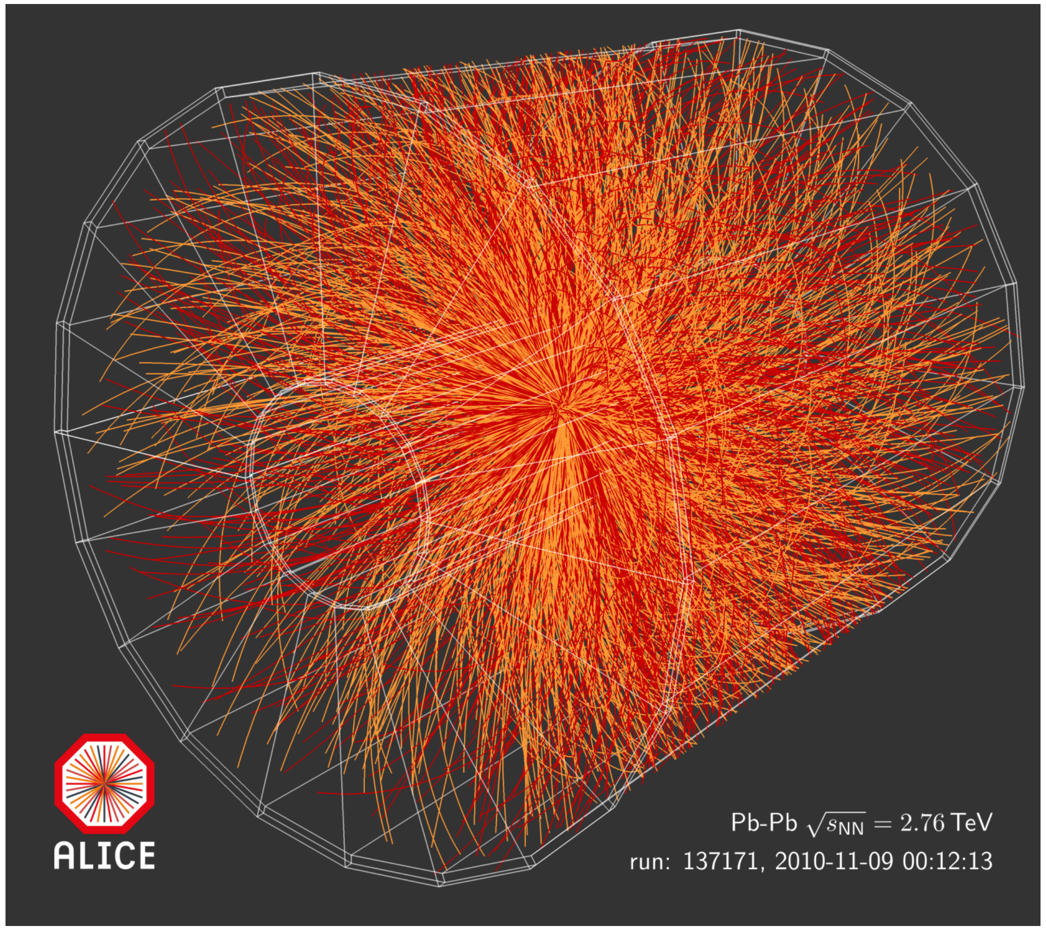
Low signal/background

Many **signals of interests** are (relatively) **rare**
Embedded in events with **large number of particles**

Pile-up: within a single "bunch crossing" multiple collisions are possible
In run 2 average number of collisions per bunch crossing μ: 20 → 50
After the upgrades up to μ = 140

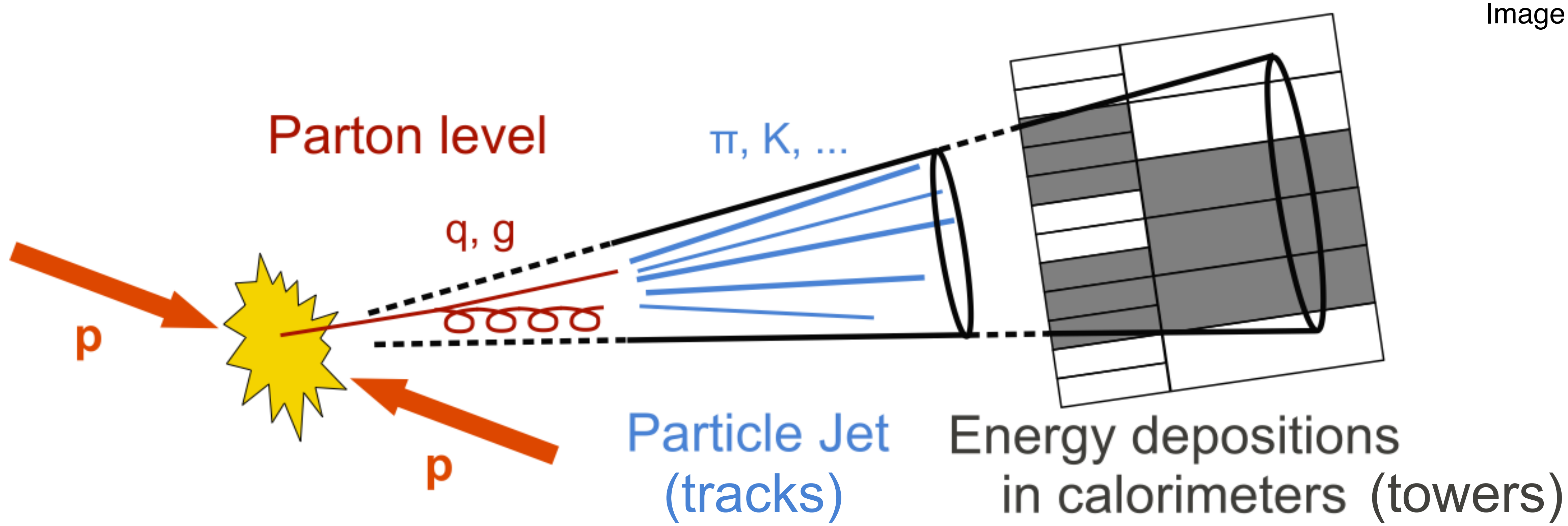
Colliding systems: LHC studies pp, p–Pb, Pb–Pb collisions
Head-on Pb–Pb collision (5% most central): **multiplicity > 200 x pp**





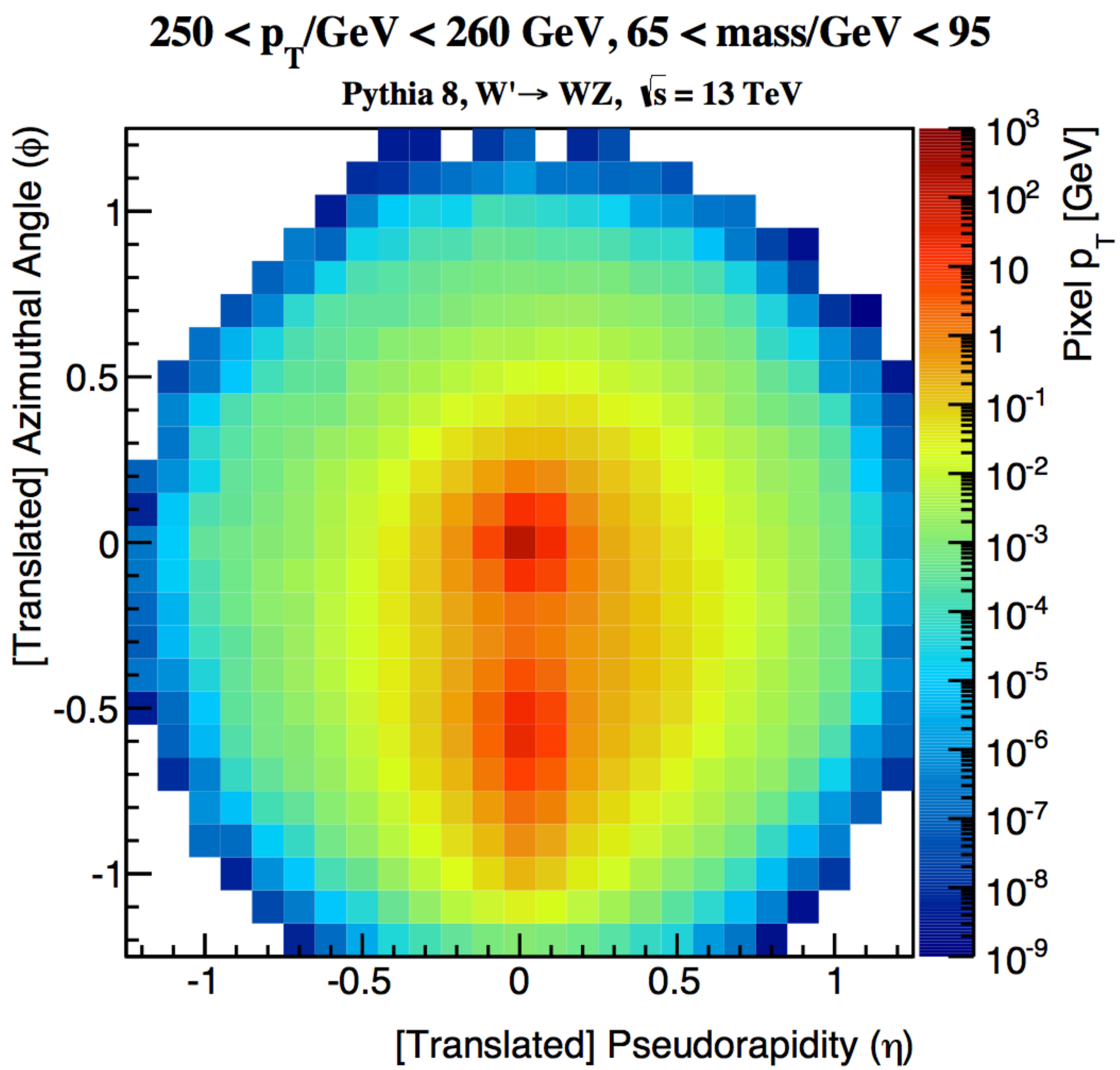
What is a jet?

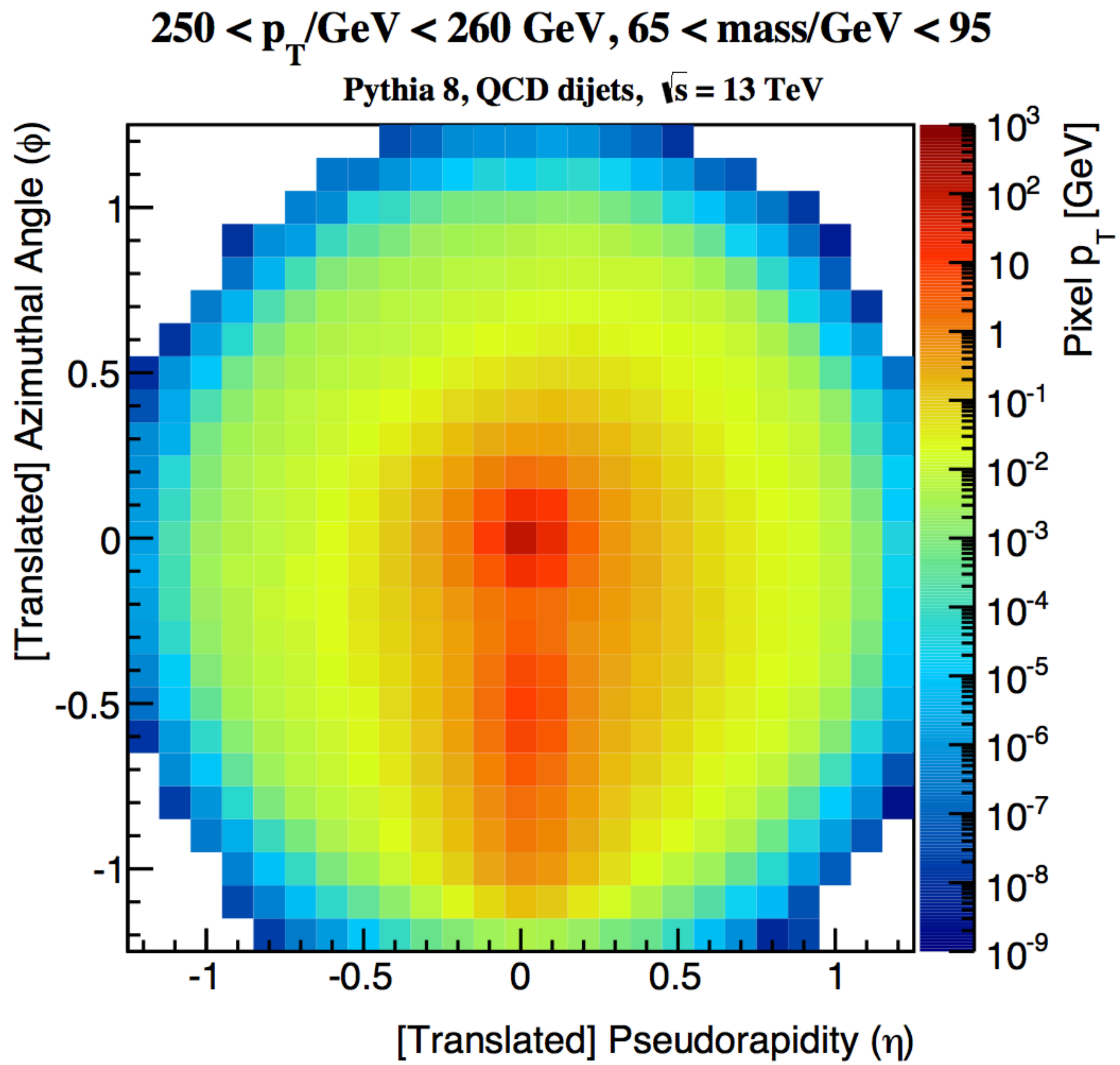
Image from CMS



- Hard collisions between protons constituents generate energetic partons (q,g)
- Partons cannot exist as free particles, produce **"spray" of particles**
- Nature produces "**particles**",
  experiments measure "**tracks**" and calorimetric "**towers**"
- Experimentally: tracks and towers (**constituents**) **clustered** by specialized algorithms to **reconstruct jets**
- **The problem**: determine the nature of the **object** which **created the jet**
  - Several interesting cases (**boosted objects, b-jets, quark vs gluon, …**)
- Standard approach: compute (expert) **high-level features** from constituents
  - Possibly combine several features using machine learning

Jet images





Idea: treat **jets** as "**images**" in η (= f(θ)) and ϕ, where each pixel is a calorimeter tower and intensity is proportional to energy deposition

Single jet images are **sparse** (**5-10%** of pixels)

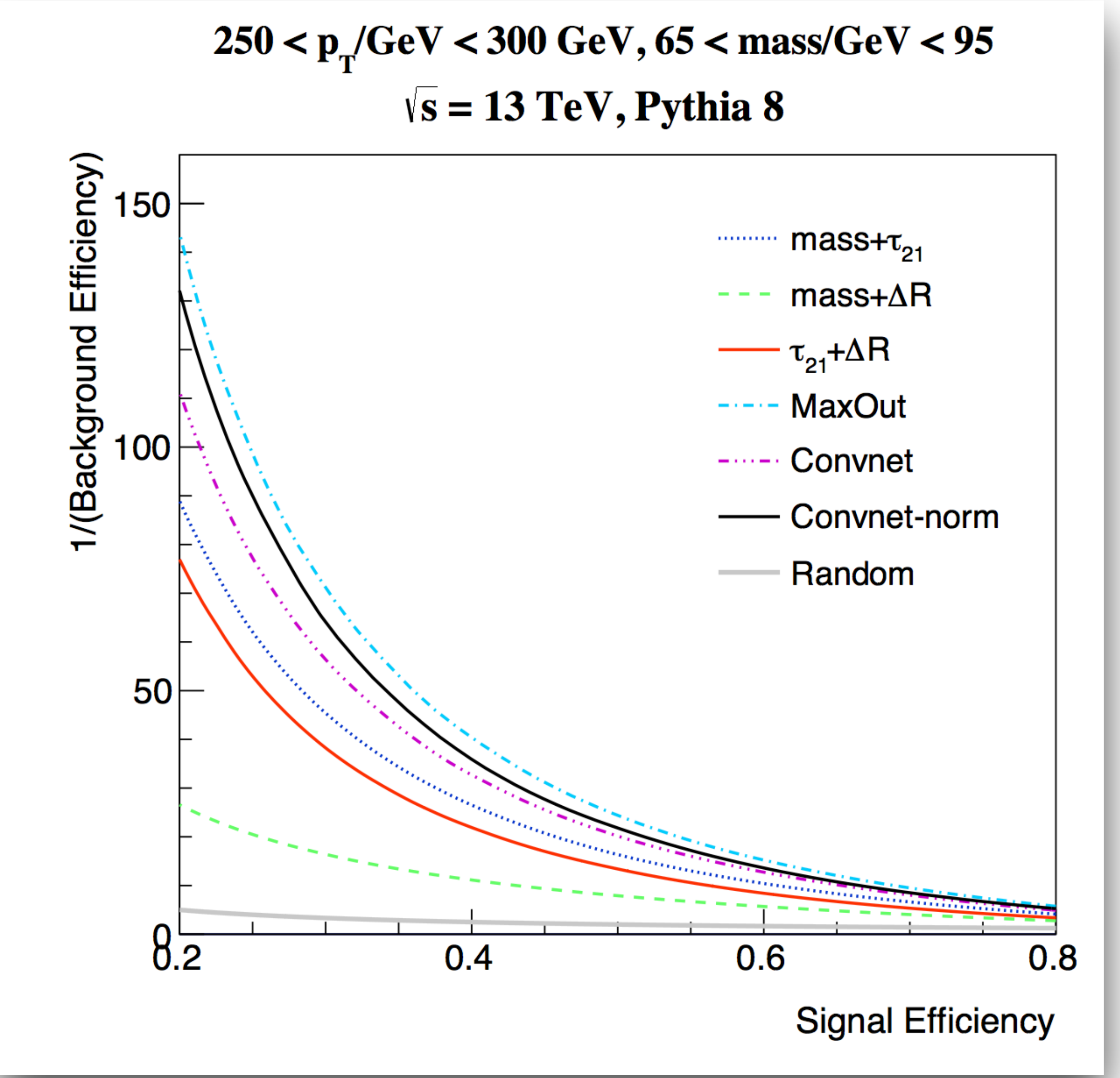
Use (almost) standard CV machinery (**Deep** or **Conv NN**)

In this paper: boosted W, pixelation mimics detector

de Oliveira et al. JHEP 1607 (2016) 069

Results from jet images

de Oliveira et al. JHEP 1607 (2016) 069

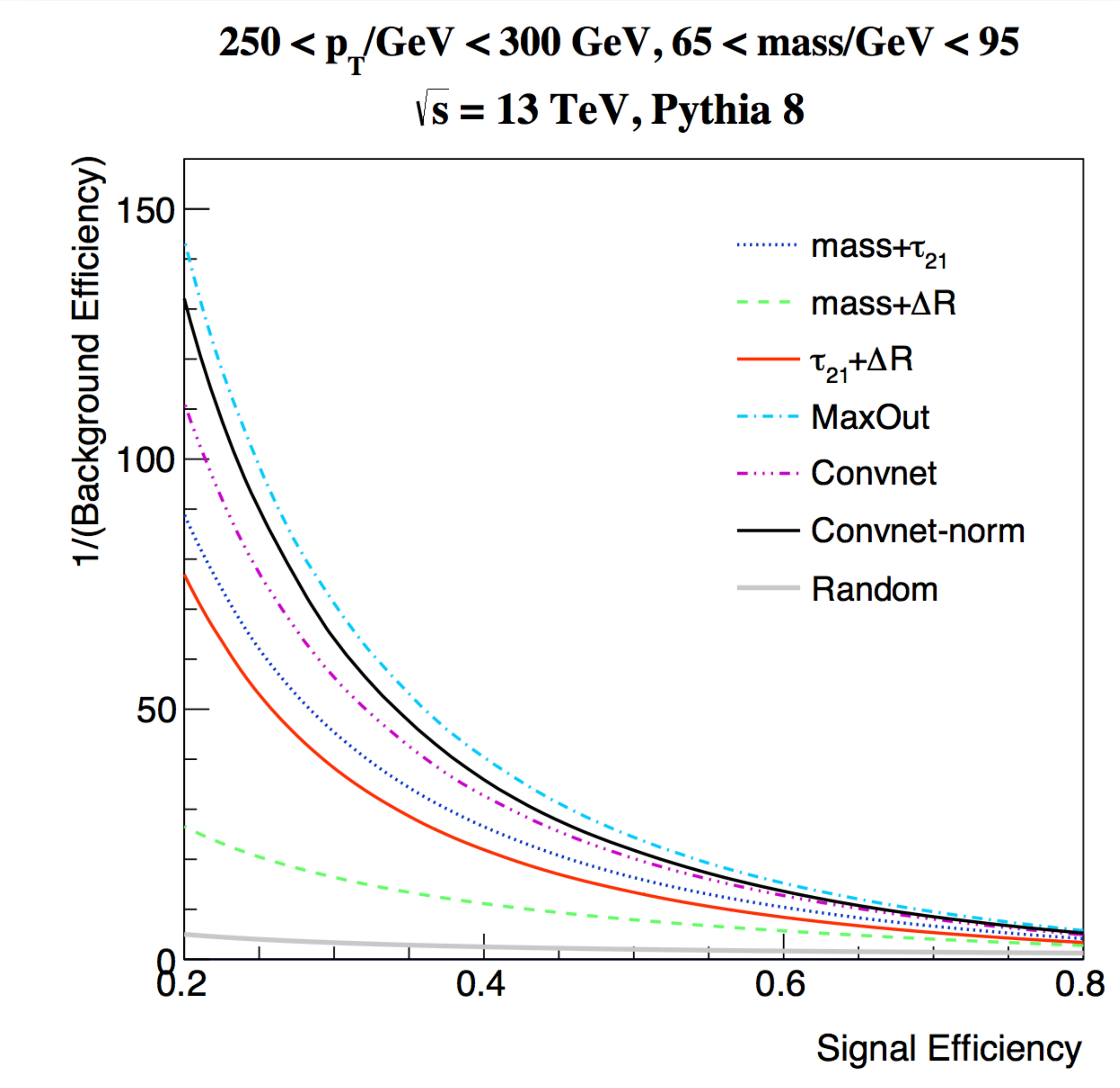


Deep NN **outperform** selection based on physics-inspired features
Fully connected better than conv (sparsity?)

Similar approach:
Kasieczka et al, arXiv:1701.08784v1
Komiske et al,. JHEP01(2017)110
Barnard et al, PRD 95, 014018 (2017)

Results from jet images

de Oliveira et al. JHEP 1607 (2016) 069



Deep NN **outperform** selection based on physics-inspired features
Fully connected better than conv (sparsity?)

Baldi et al, PRD 93, 094034 (2016)

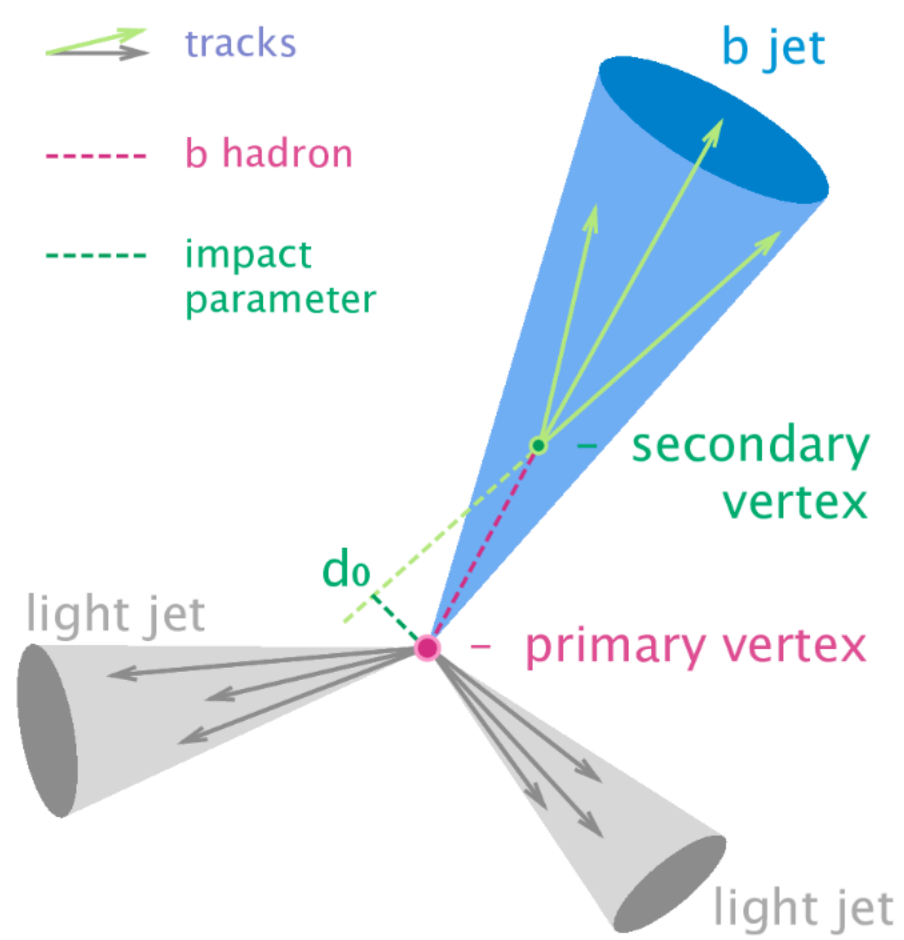
Technique	Signal efficiency at background rejection = 10	AUC
	No pileup	
BDT on derived features	86.5%	95.0%
Deep NN on images	87.8% $_{(0.04\%)}$	95.3% $_{(0.02\%)}$
	With pileup	
BDT on derived features	81.5%	93.2%
Deep NN on images	84.3% $_{(0.02\%)}$	94.0% $_{(0.01\%)}$

DNN still performs better when **detector effects** (Delphes) and **pile-up** are taken into account

Similar approach:
Kasieczka et al, arXiv:1701.08784v1
Komiske et al,. JHEP01(2017)110
Barnard et al, PRD 95, 014018 (2017)

b-jets tagging

- Alternative approach: use **array of constituents** instead of images or high level features
    - Allows **avoiding pixelation**
    - Can go even **lower-level** than jet images
- Well suited for **b-jets tagging**
    - B-hadrons decay after **finite length** ($c\tau$ ~ 500 μm)
    - Traditional approach: high level features based on the identification of **secondary vertices**
    - Can one use **individual tracks** as input and let the ML method find **(better) high level features**?

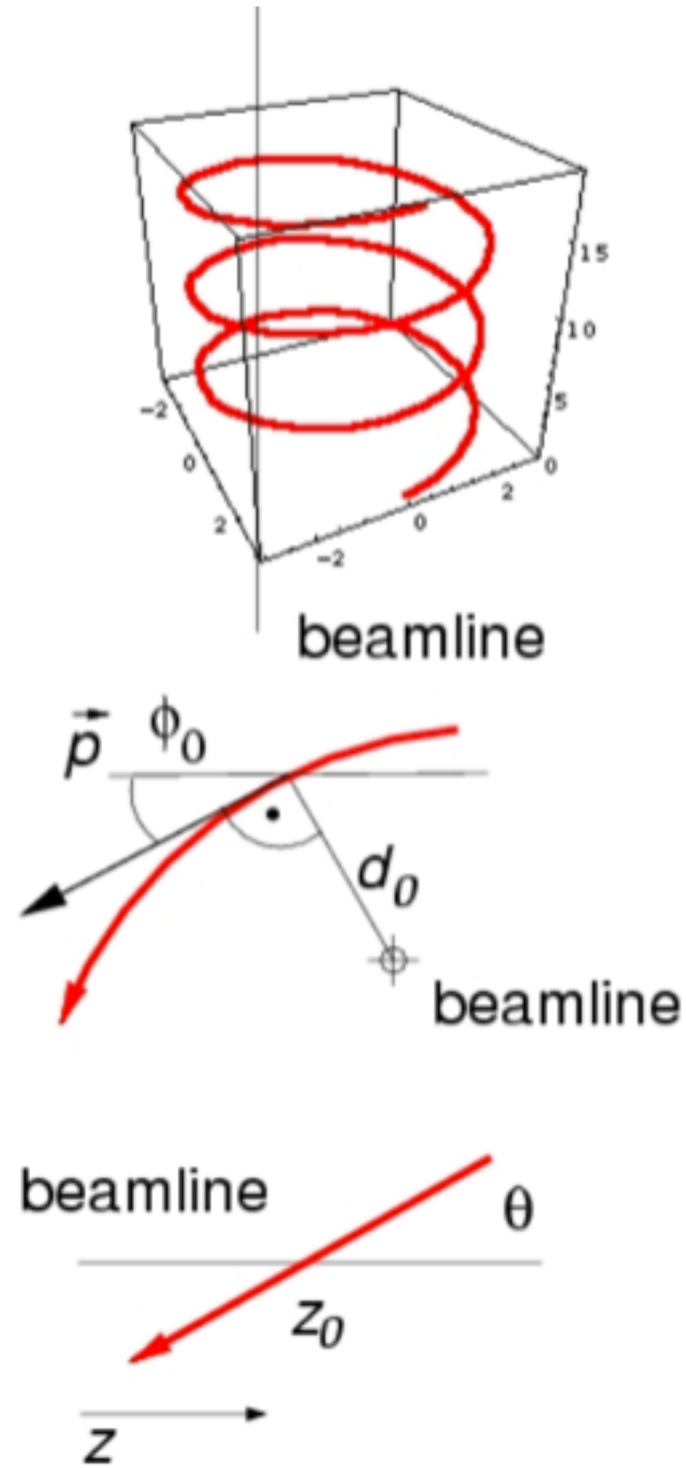


http://bartosik.pp.ua/hep_sketches/btagging

for a study on top tagging using constituents, see Pearkes et al, arXiv:1704.02124v1

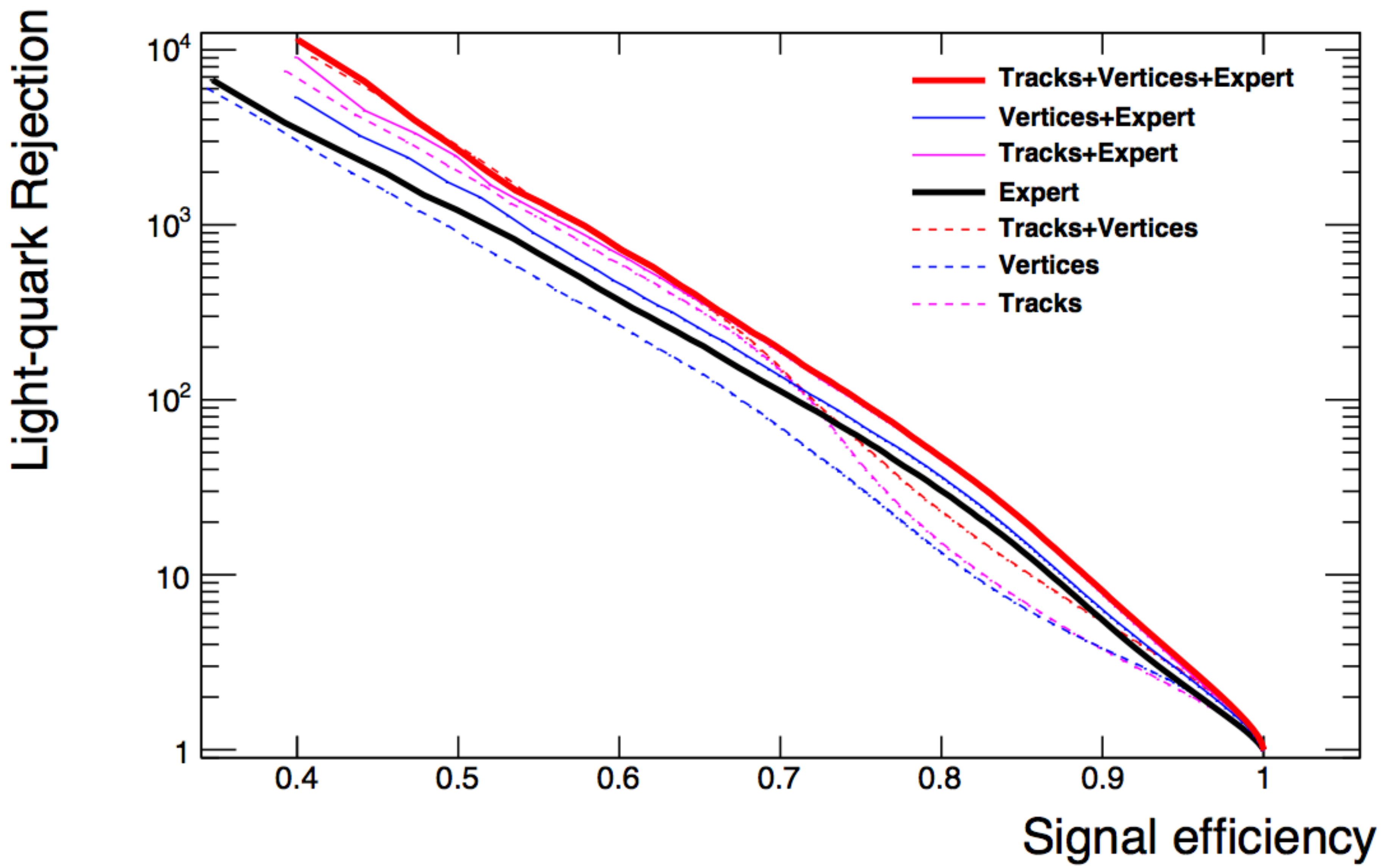
What is a track?

- Track is an **approximately helix** trajectory in 3D, described by
  - **5 parameters**
  - their **covariance** matrix (15 parameters)
- Physics analysis often uses only **momenta** (p_T, η, ϕ), implicitly assuming a common origin for all particles
- Secondary vertex finding requires **propagating tracks** along their trajectory
- Standard **workflow**:
  - Tracks → Vertices → High Level Features



b-jet tagging, Delphes study

Guest et. al, PRD **94**, 112002



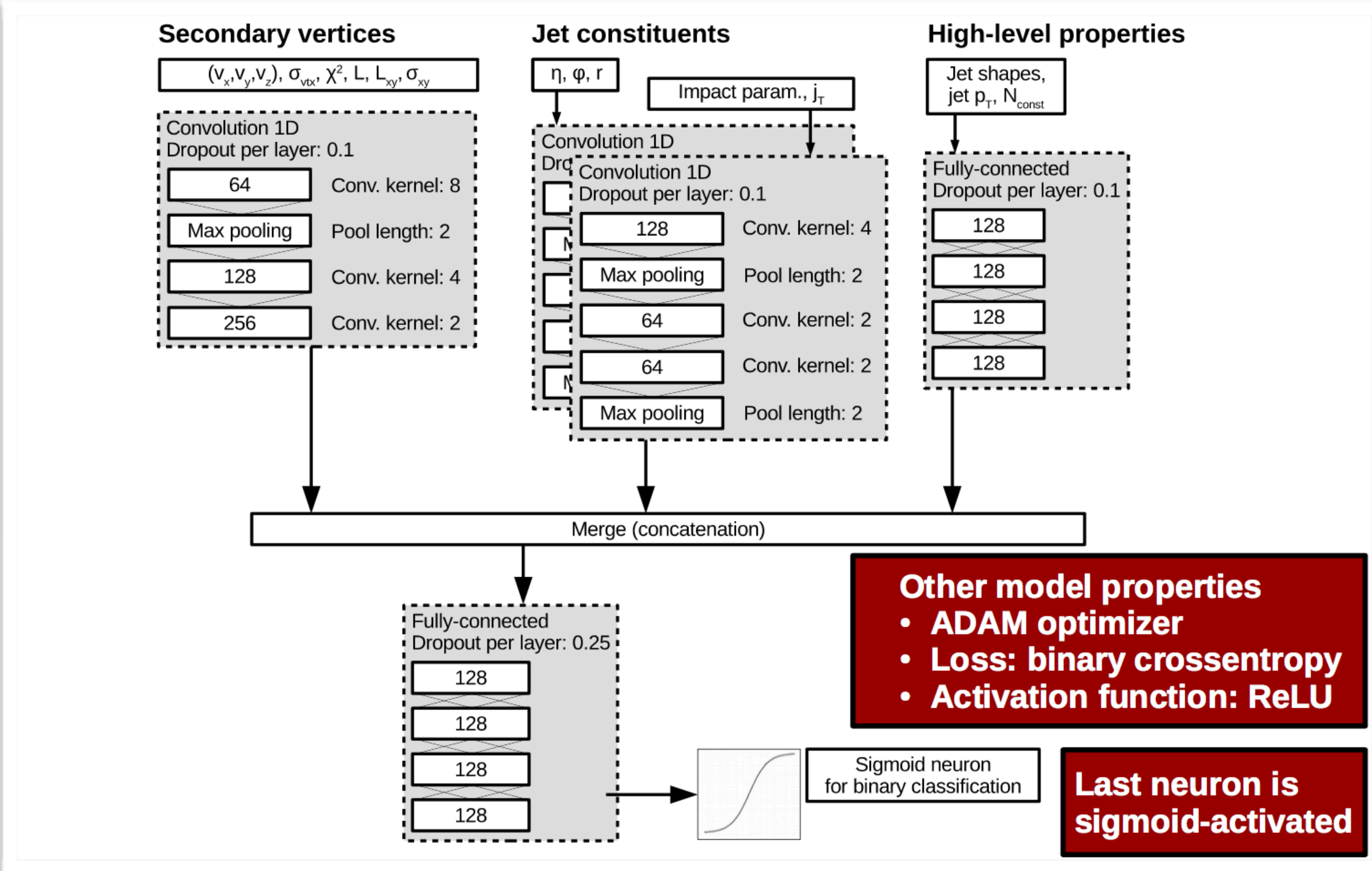
Semi-realisitic **detector simulation** (Delphes)

Uses full info on **track parameters + covariance**

Tracks or vertices alone under-perform expert features

Track+Vertices or Tracks+Vertices+Expert **outperform expert**

Various architecture: **feed-forward** (better), LSTM, Outer recursive



Full detector simulation, **p-Pb** collisions

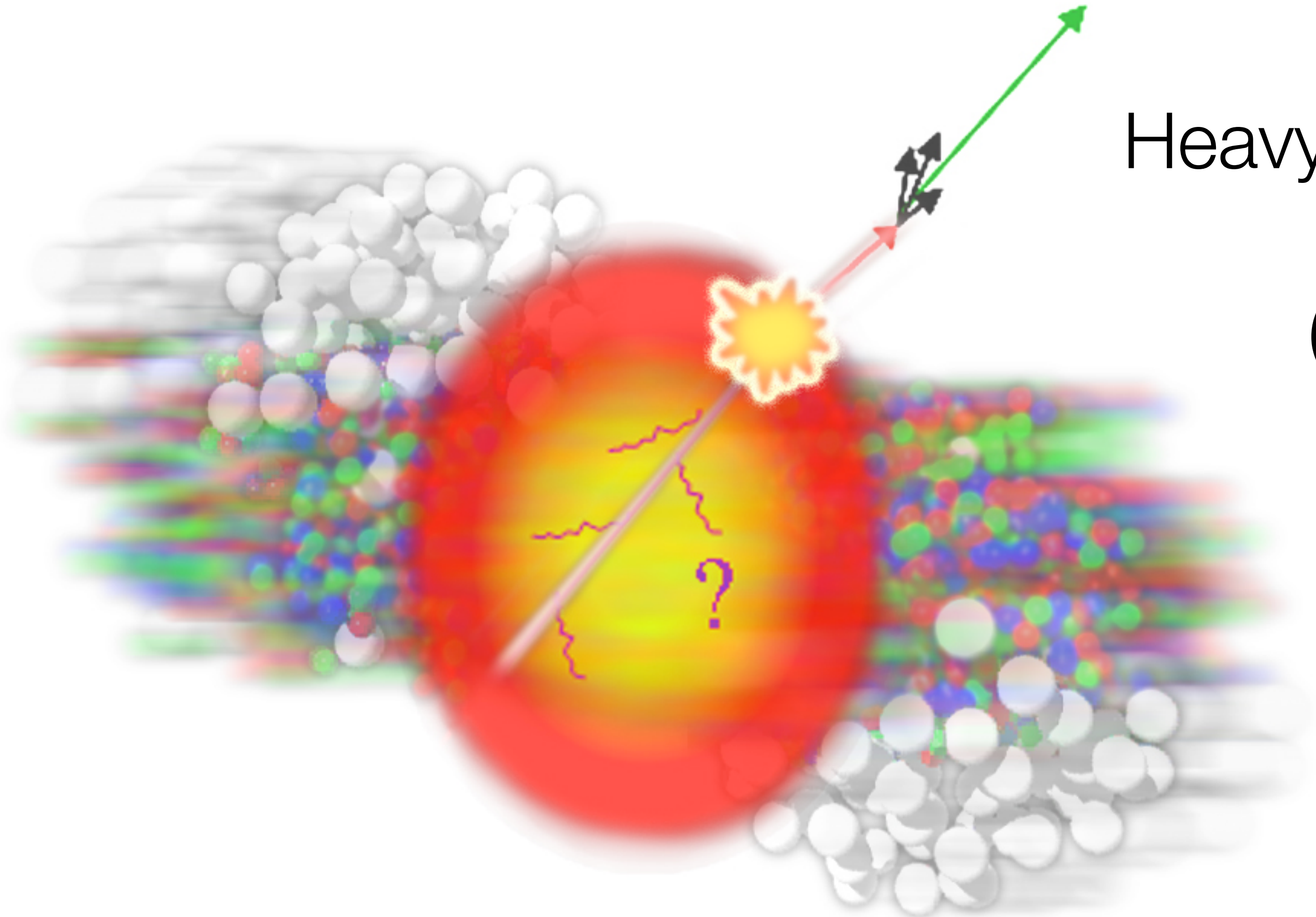
Not "as **low-level**" as previous study

Several other **architectures** studied

(LSTMs, 2D convolutional networks on jet images, …)

R. Haake for ALICE, IML Workshop

Heavy ions issues: a primer on jet quenching



Heavy ion collisions goal: study hot and dense QCD matter (quark gluon plasma – **QGP**)

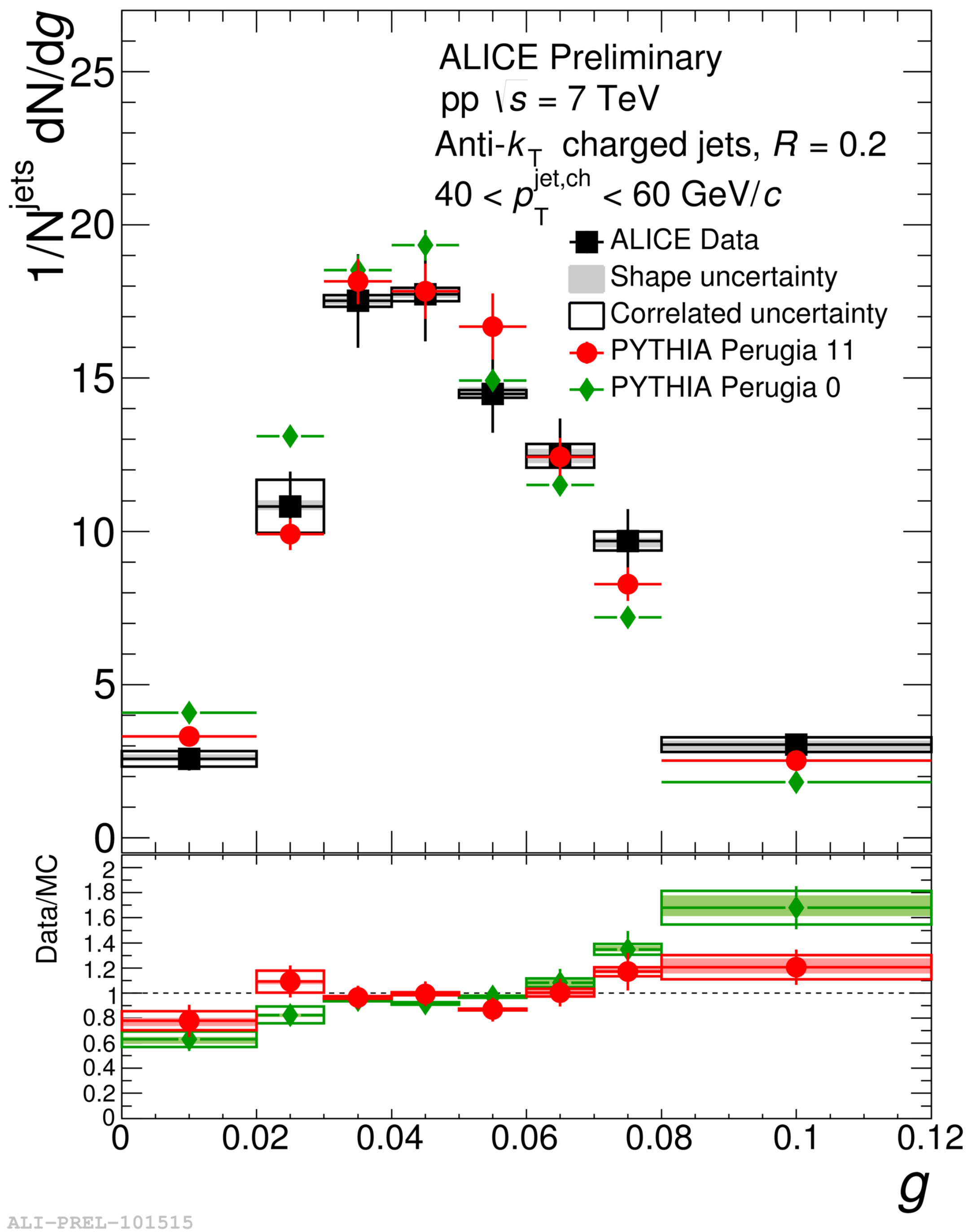
Jets lose energy when traversing the QGP
Different partons → Different energy loss

Distinguishing heavy quark, light quark and gluon:
→ **microsopic process** of energy loss
→ Information on QGP

Problem:
classifier trained on **pp-like** jets → **mis-tag Pb-Pb** quenched jets?

Jet Shapes, results

$$g = \sum_{i \in \text{jet}} \frac{p_T^i}{p_T^{jet}} |r_i|$$

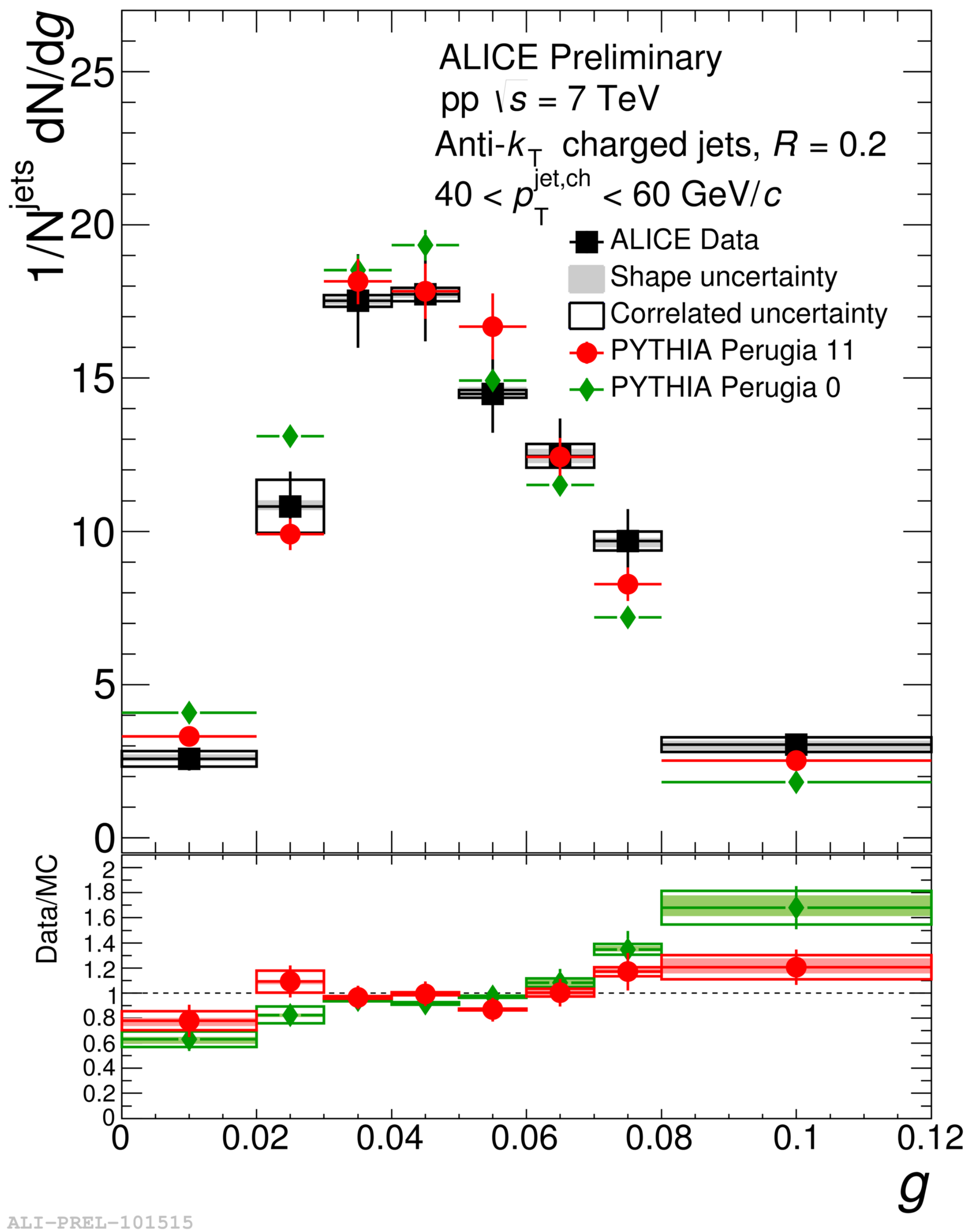


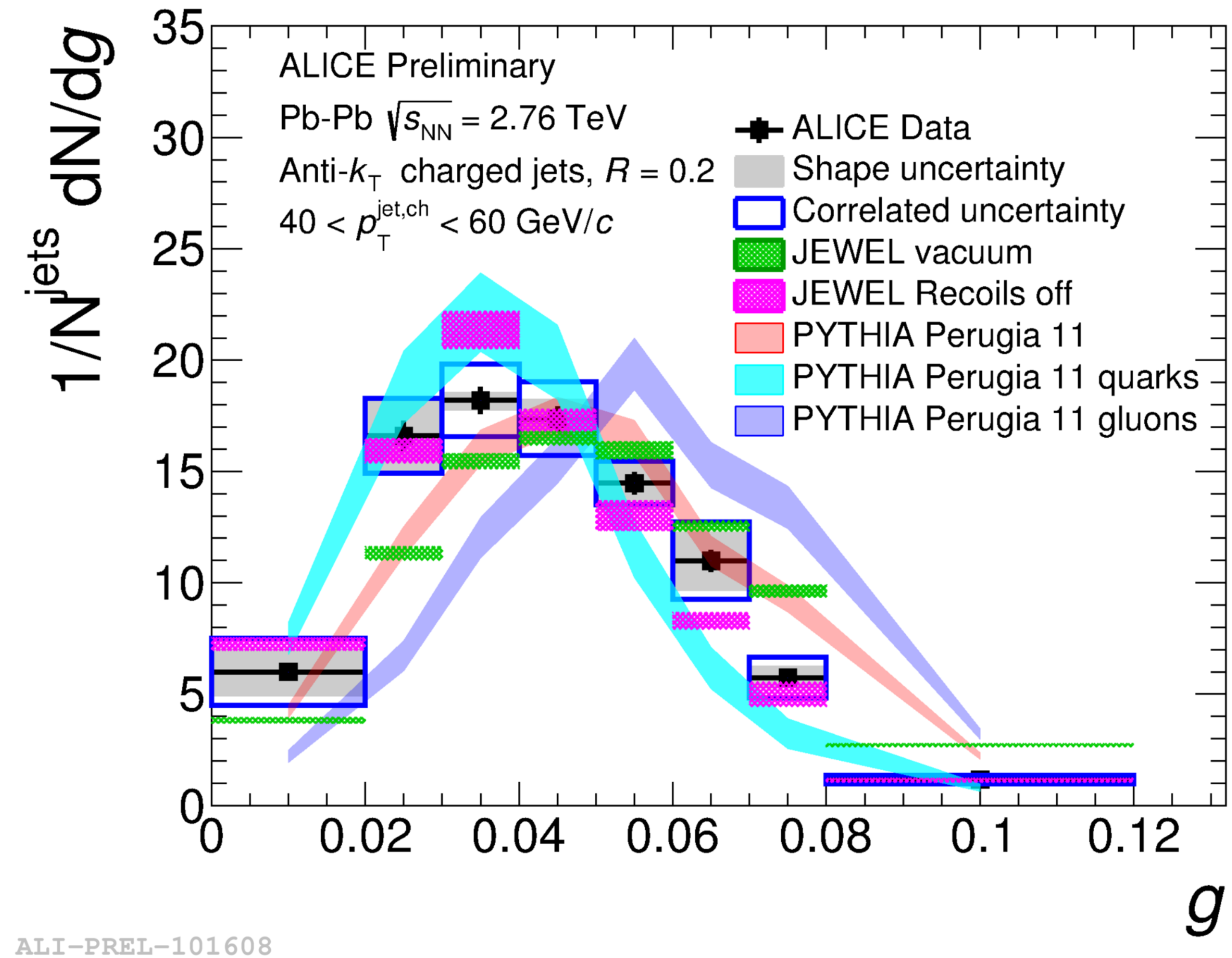
Pythia reproduces jet **shapes**

(e.g. girth) in pp collisions

Jet Shapes, results

$$g = \sum_{i \in \text{jet}} \frac{p_T^i}{p_T^{jet}} |r_i|$$





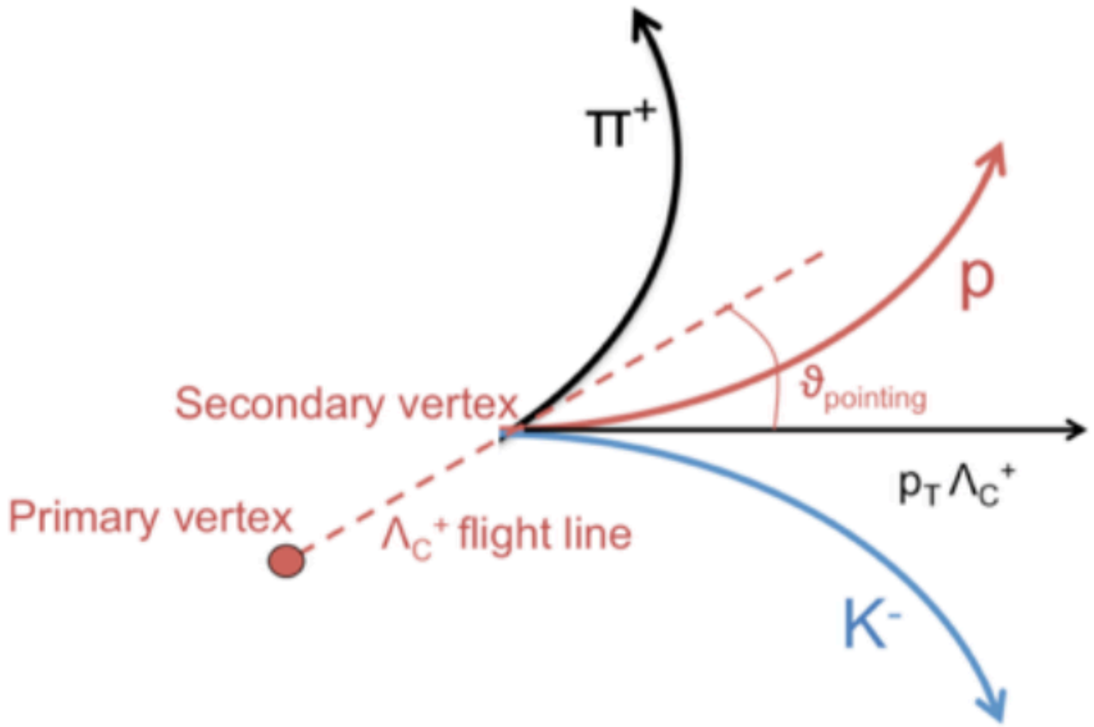
Pythia reproduces jet **shapes**
(e.g. girth) in pp collisions

Shapes change in Pb-Pb,
more "**quark like**"
Different suppression of q and g?
Modification of fragmentation?
Can **higher-dimensional** data still distinguish? **unsupervised** methods?

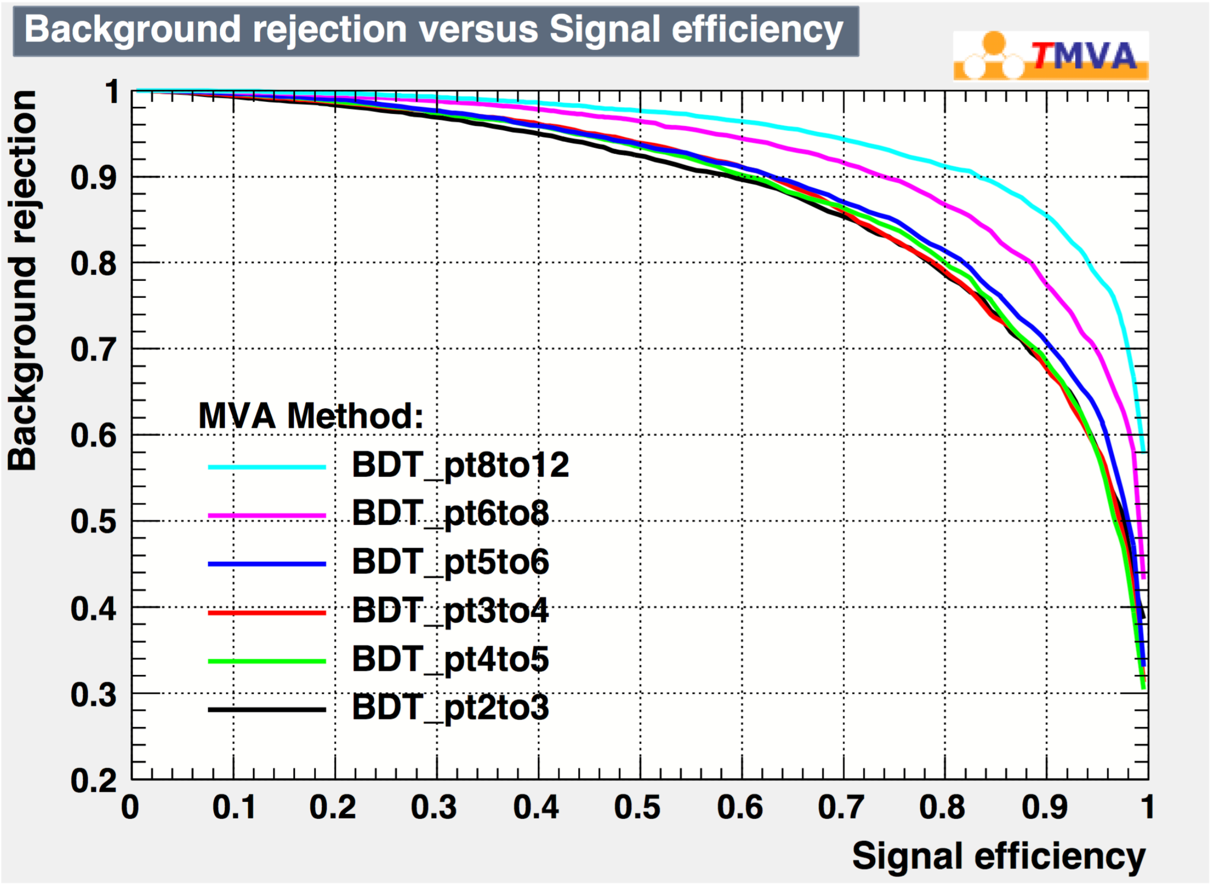
Finding a decay, $\Lambda_C \rightarrow \pi K p$

- Some particles identified through their decay products
- Reconstruction of **2- and 3-prong decays** in heavy ion collisions is challenging: **large combinatorics**
  - (remember:
    several thousand particles/event)
- **Example**: $\Lambda_C \rightarrow \pi K p$
  - Loop over all possible triplets
  - Find distance of closest approach
  - Compute geometrical + Particle ID quantities (18)
  - Decide if this is a viable candidate

https://www.flickr.com/photos/mayaevening/138372058



High Level Features classification

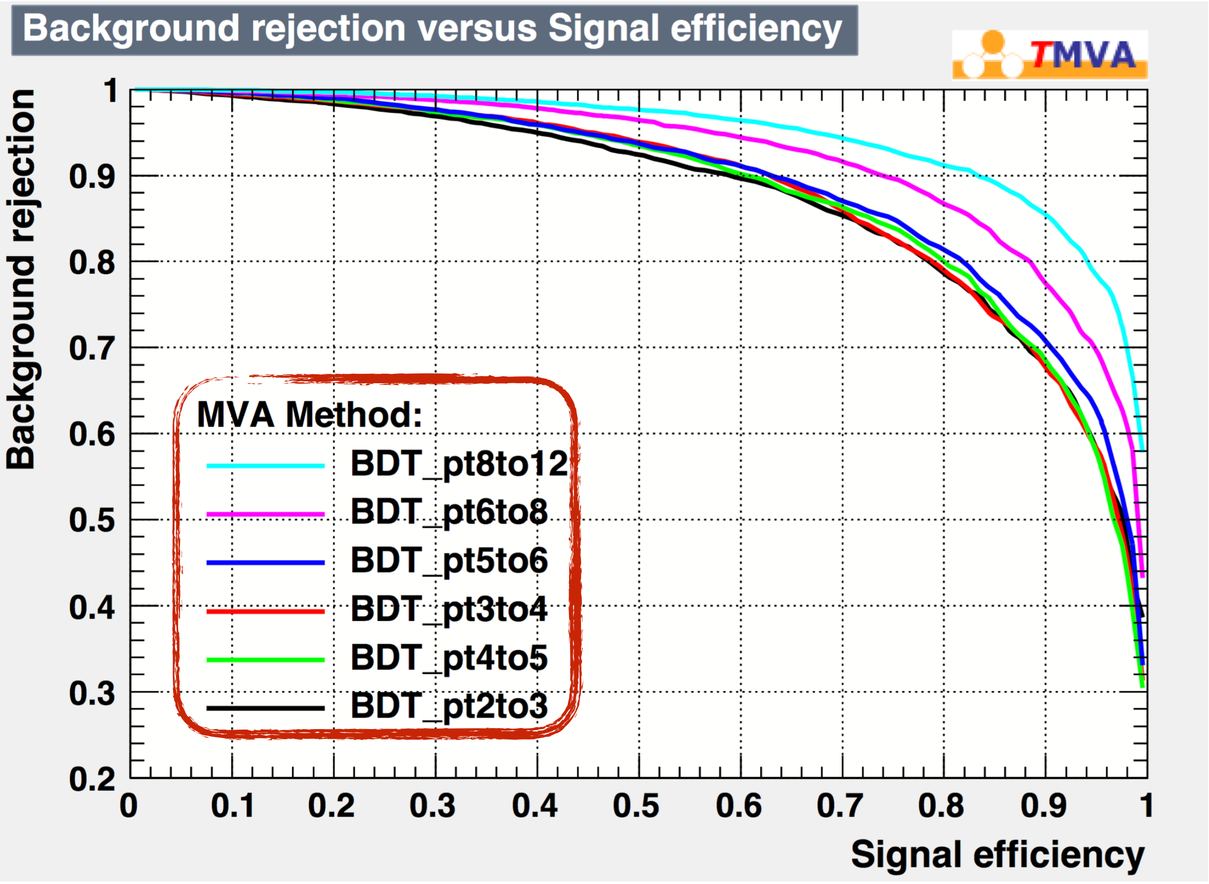


Full detector simulation, p-Pb collisions

BDT, based on **18 "high level features"**

AUC depends on **momentum bin**

High Level Features classification

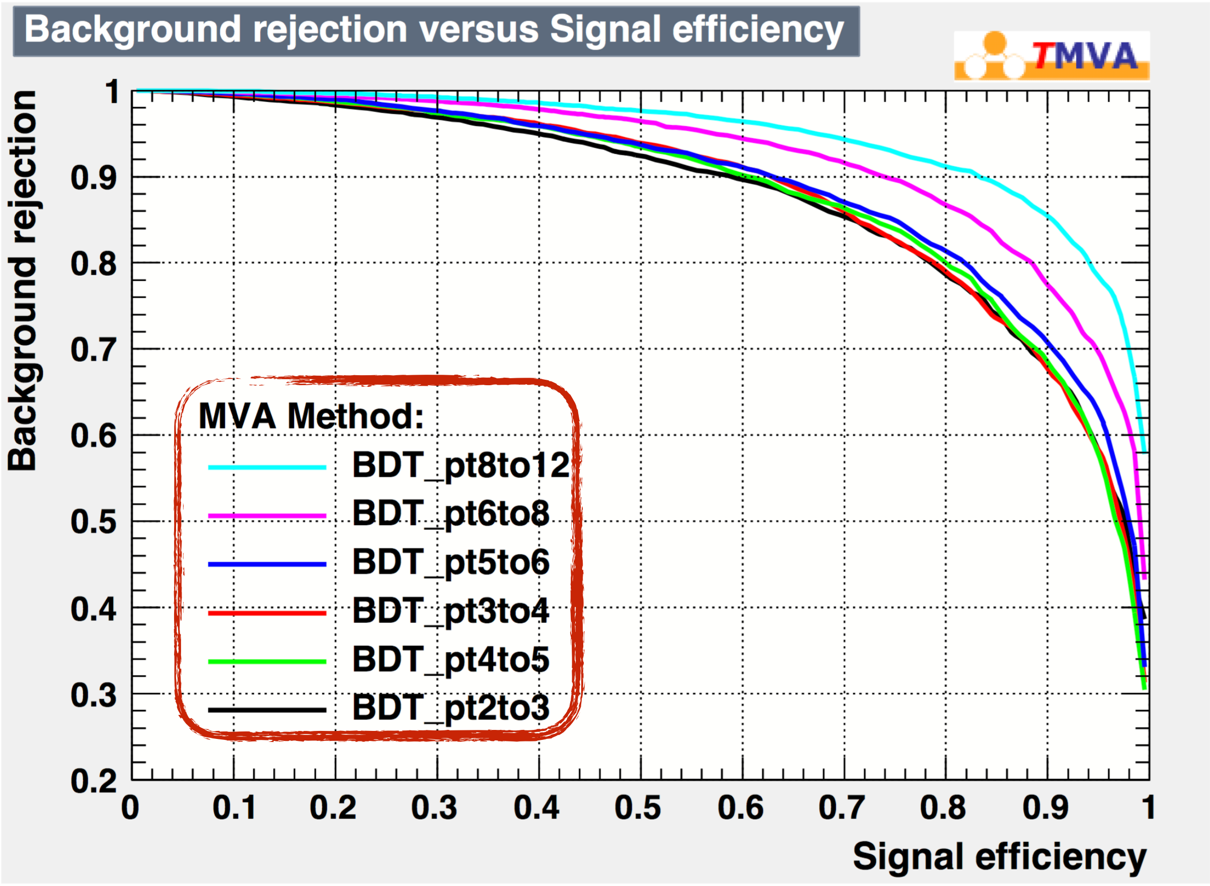


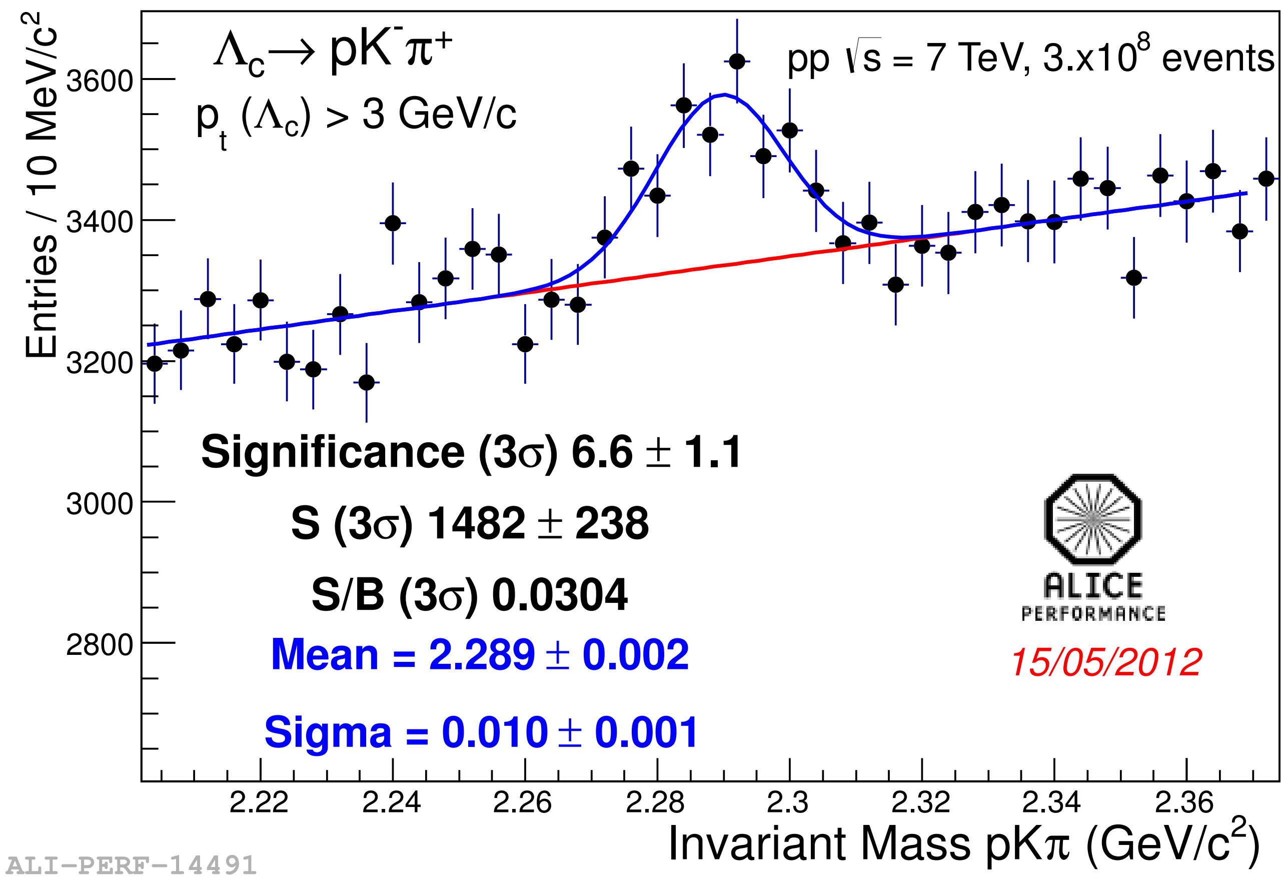
Full detector simulation, p-Pb collisions

BDT, based on **18 "high level features"**

AUC depends on **momentum bin**

High Level Features classification





Full detector simulation, p-Pb collisions

BDT, based on **18 "high level features"**

AUC depends on **momentum bin**

Invariant mass distribution to judge quality of the selection

Important: **avoid "sculpting" a peak** in the background

Deep topologies

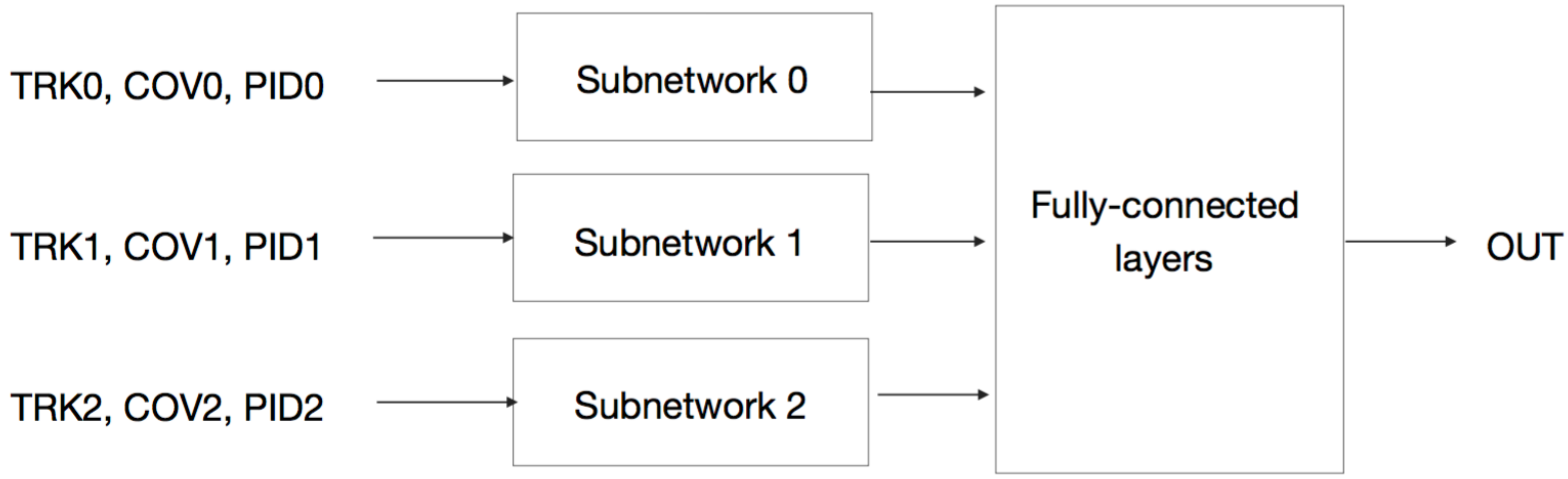
Summer student report

Deep topologies

1. Fully connected (10 layers)

Summer student report

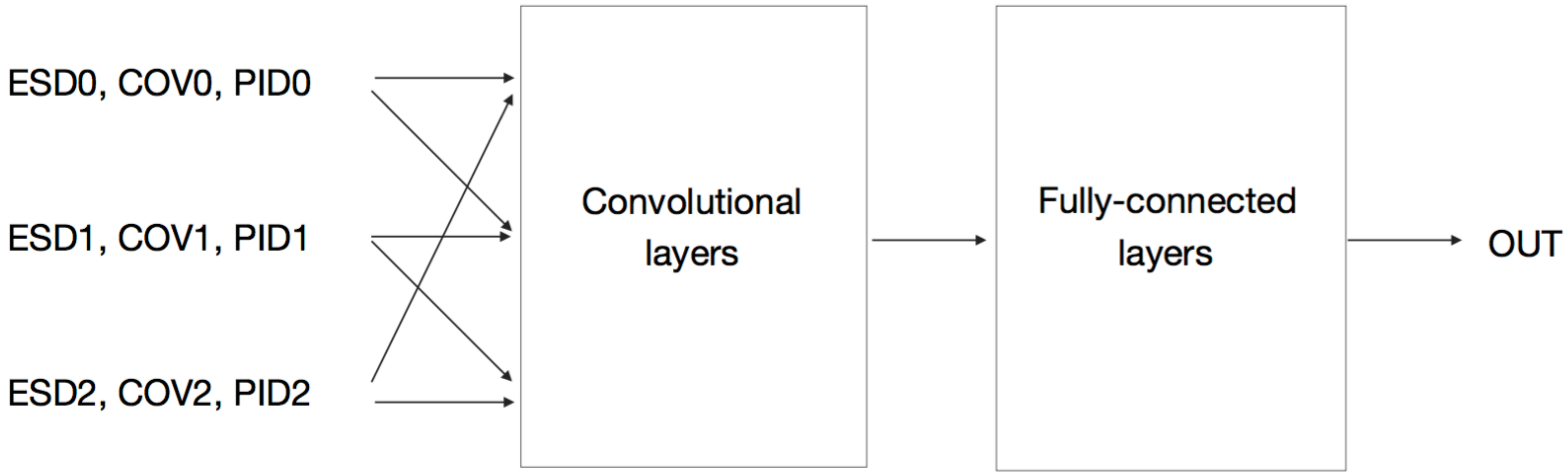
Deep topologies



1. Fully connected (10 layers)
2. Per-track subnetwork (5+5 layers)

Summer student report

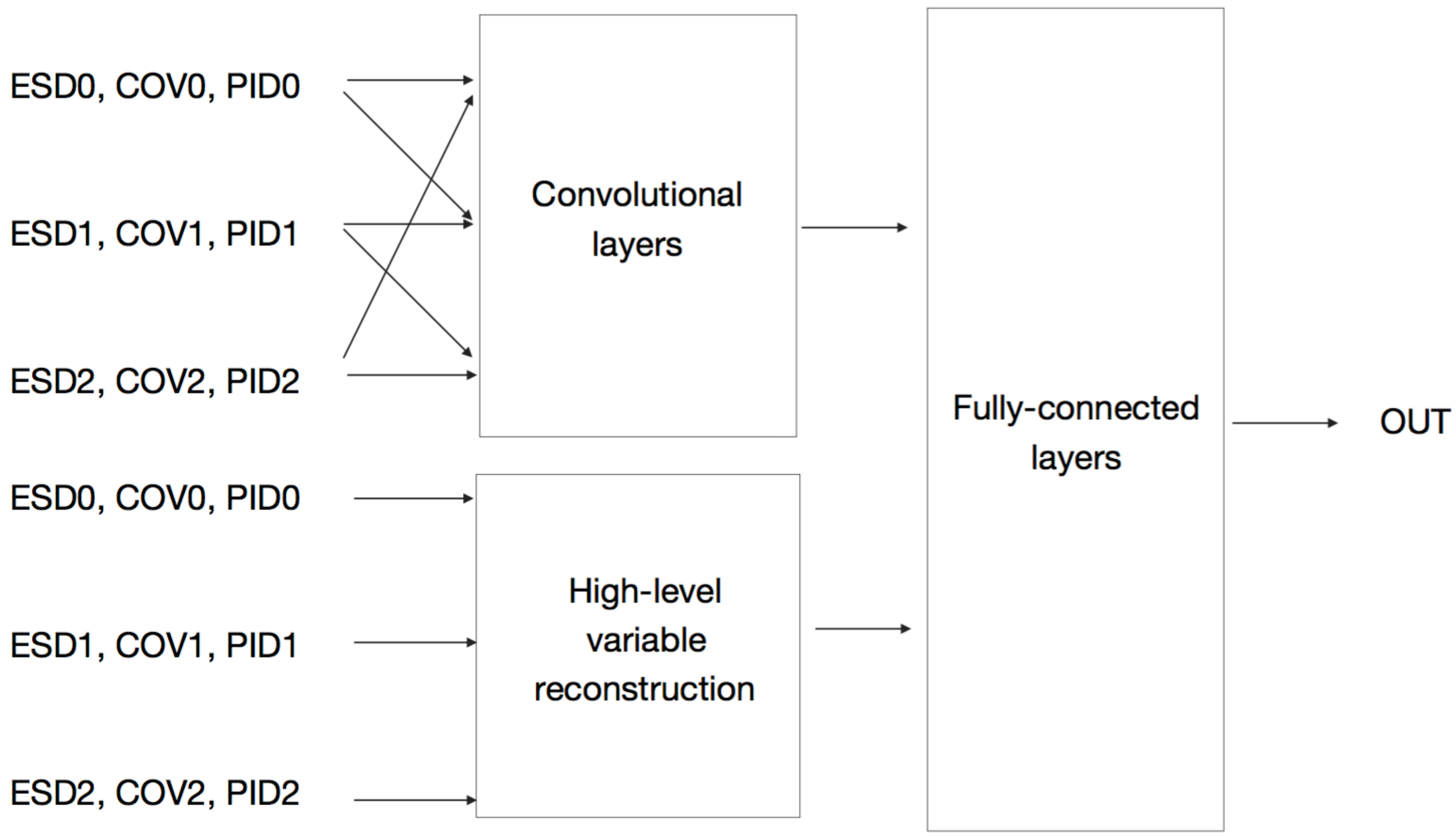
Deep topologies



1. Fully connected (10 layers)
2. Per-track subnetwork (5+5 layers)
3. Track pairs convolution (2+5 layers)

Summer student report

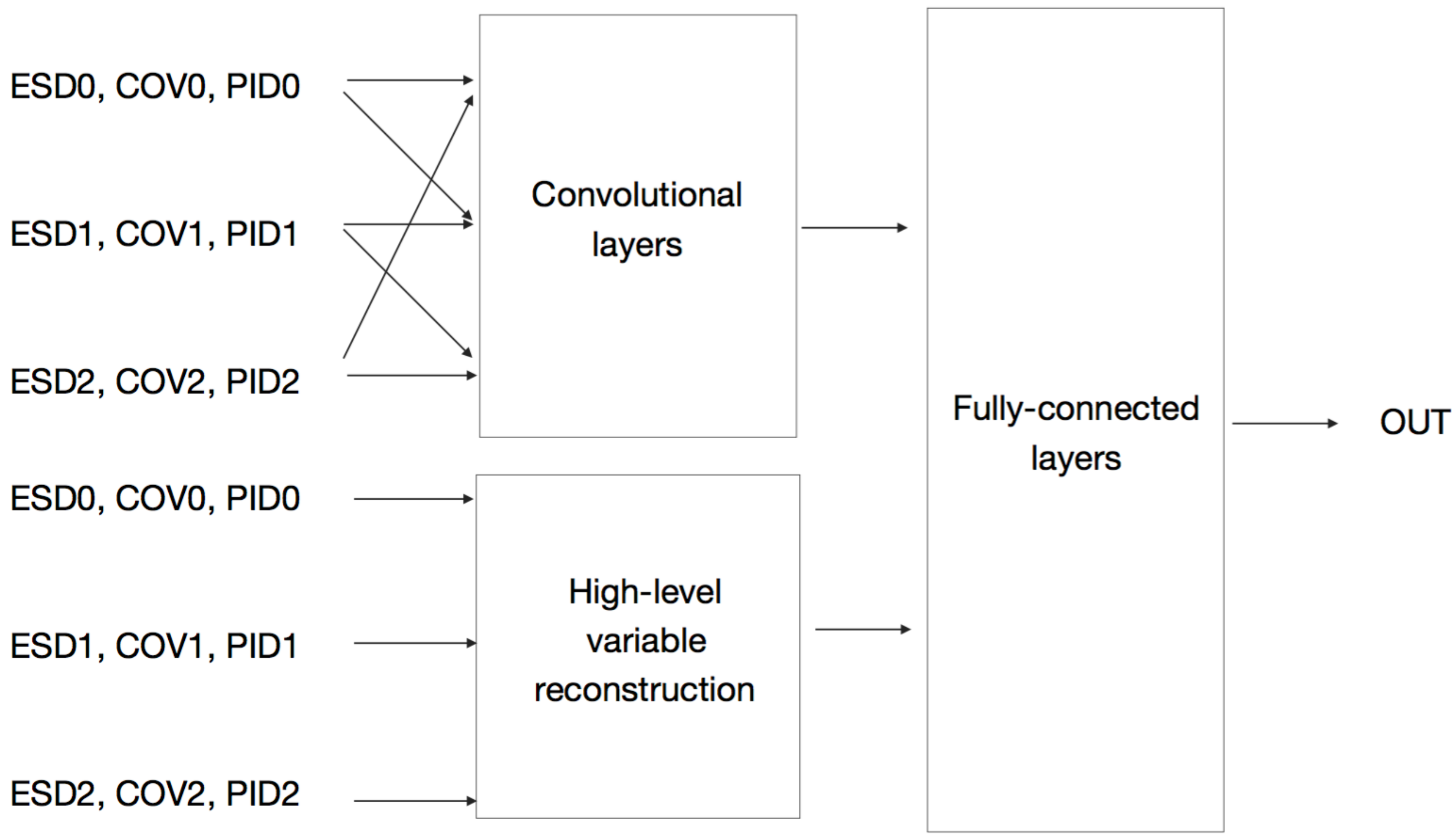
Deep topologies



1. Fully connected (10 layers)
2. Per-track subnetwork (5+5 layers)
3. Track pairs convolution (2+5 layers)
4. High level filter

Summer student report

Deep topologies

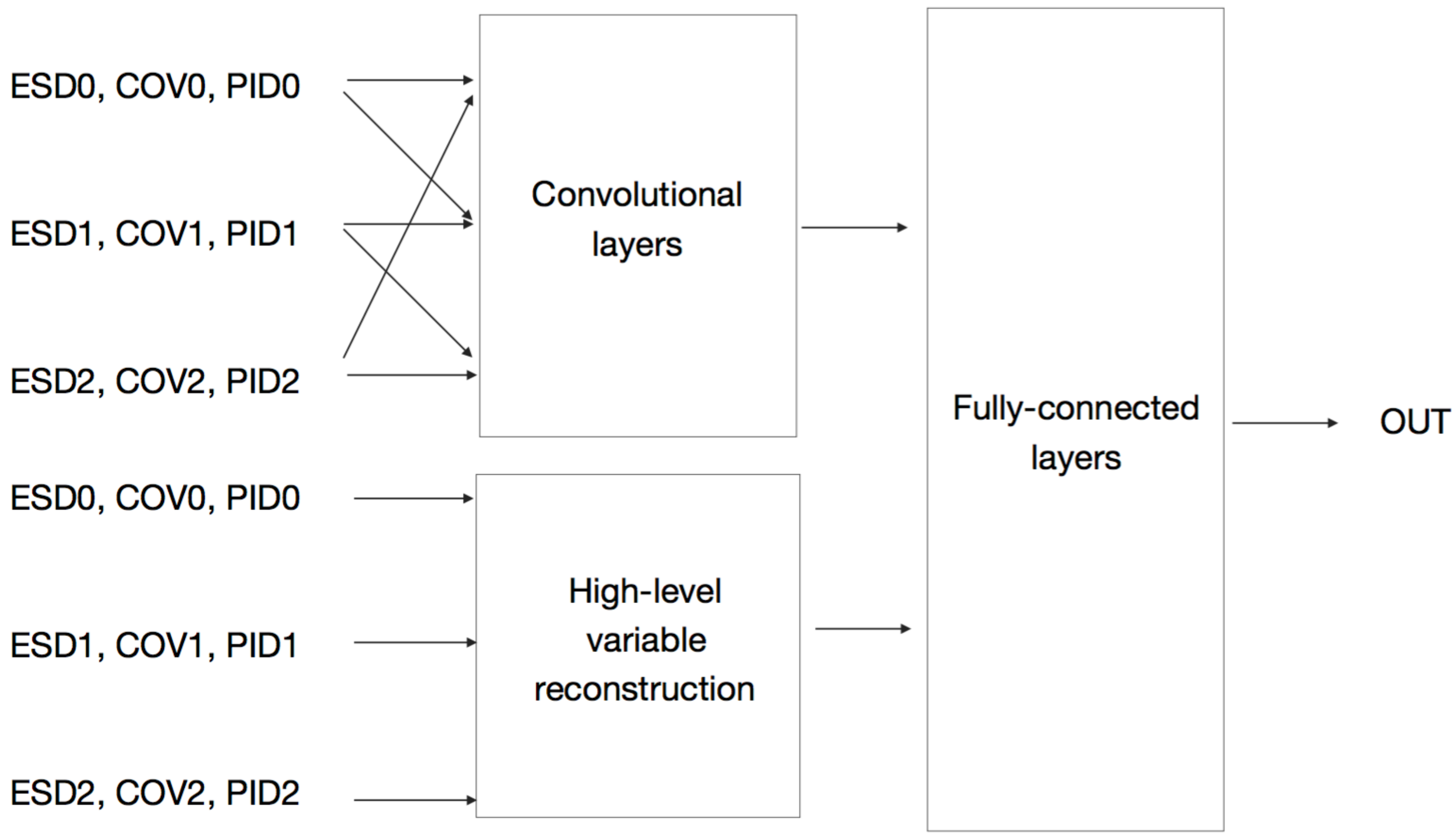


1. Fully connected (10 layers)
2. Per-track subnetwork (5+5 layers)
3. Track pairs convolution (2+5 layers)
4. High level filter

None outperforms "shallow" methods

Summer student report

Deep topologies



1. Fully connected (10 layers)
2. Per-track subnetwork (5+5 layers)
3. Track pairs convolution (2+5 layers)
4. High level filter

None outperforms "shallow" methods

Best DNN: AUC ~ 0.906 vs 0.920 (but margins for improvement!)

Summer student report

Outlook: Beyond the toys

- Training data come from **Monte Carlo**, imperfect description of real data
  - Avoiding over-relying on Monte Carlo (**domain adaptation**, see e.g. here)
  - DL requires very large training samples, may get **expensive** (GANs?)
- Jet **images are sparse**, some CV technique need to be adapted (e.g. CNN)
- Effect of **detector reconstruction** and **pile-up**
  - Partially investigated in current studies
- Large **backgrounds** (from pp pile-up or heavy ion):
  - Fake jets or spurious constituents attached to jets
  - Fluctuations in the jet background ⇒ smearing in jet energy
  - Use **DL for regression** to handle backgrounds?
- **Heavy ion specific**:
  - Training based on pp distributions, jets **modified** in Pb-Pb
  - **Semi-supervised** or **unsupervised** approaches? Domain adaptation?
- Aggressive classification may result in “**background sculpting**”, serious issue for decay studies (Adversarial decorrelation? see e.g. here)

Summary

- **Deep learning** methods being investigated in the HEP community for various classification problems
  - Initial performance **promising**
- Going deep by itself seems to improve some problems
  - Symmetries and **peculiarities of physics datasets** not yet fully exploited
    - (see also arxiv:1702.00748)
  - **Expert knowledge** leads to gains also when combined with deep methods
- Some **real-life issues** to be addressed (studies ongoing)
- **Infrastructure** for large-scale application of deep-learning (GPUs?)
- Will surely benefit from **collaboration with industry**!

Additional links:

IML Workshop

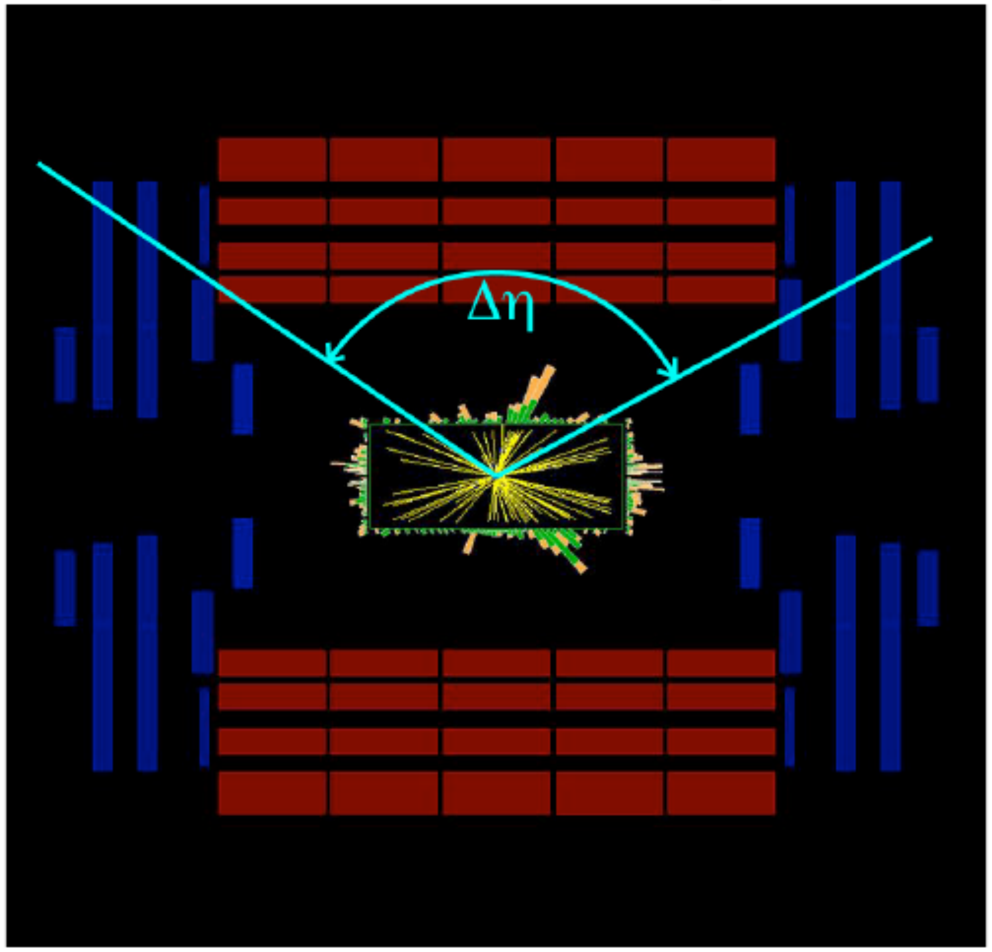
IML Meeting on Deep Learning

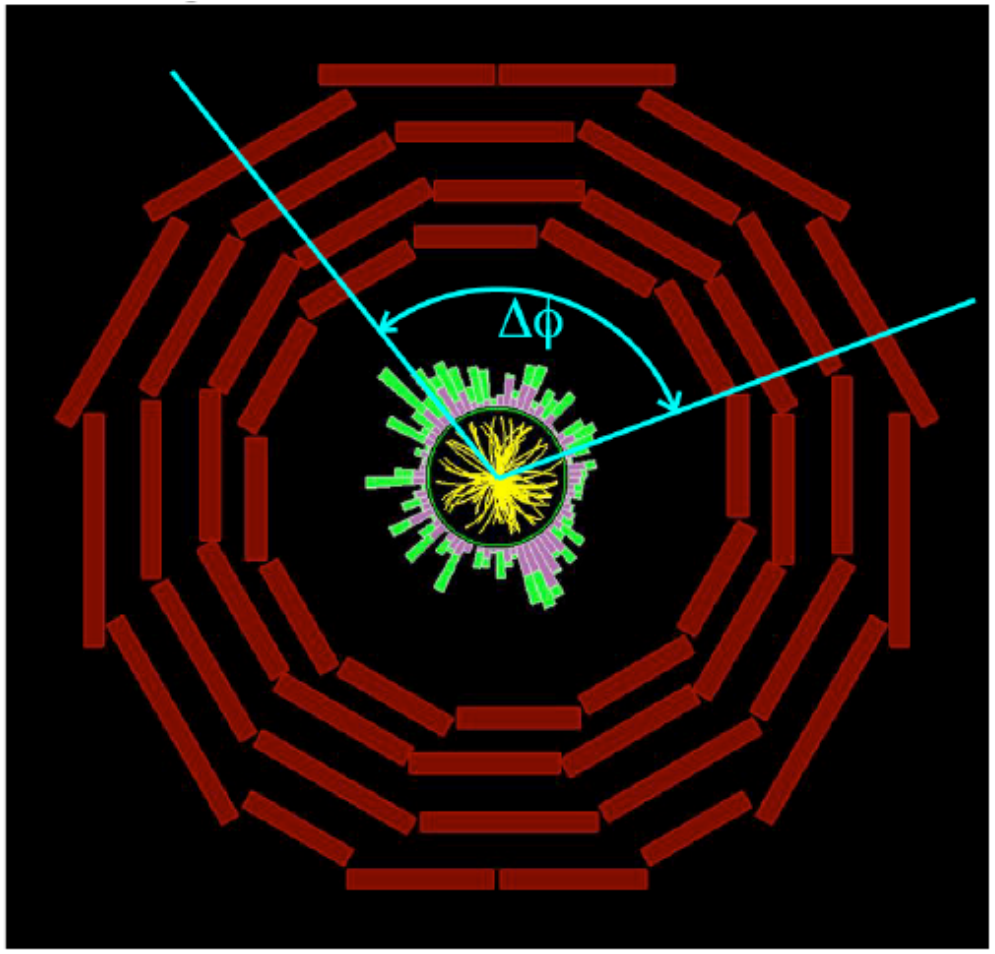
DS@HEP Series (2015, 2016, 2017)

Backup

Angular coverage

Image from here

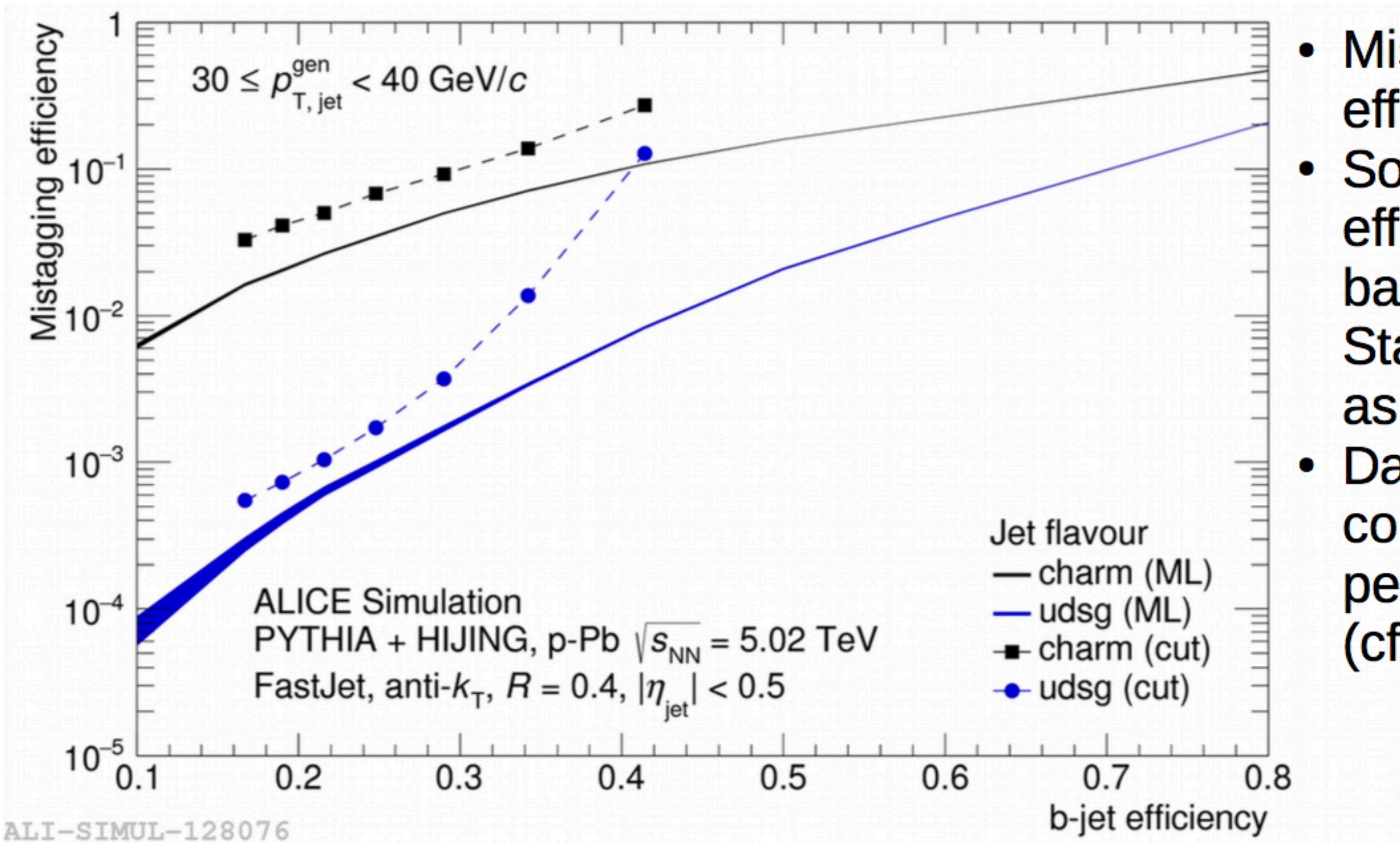




Angular coverage of the detectors, typically expressed as a function of azimuthal angle ϕ and pseudorapidity η

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$

Performance compared to non-ML

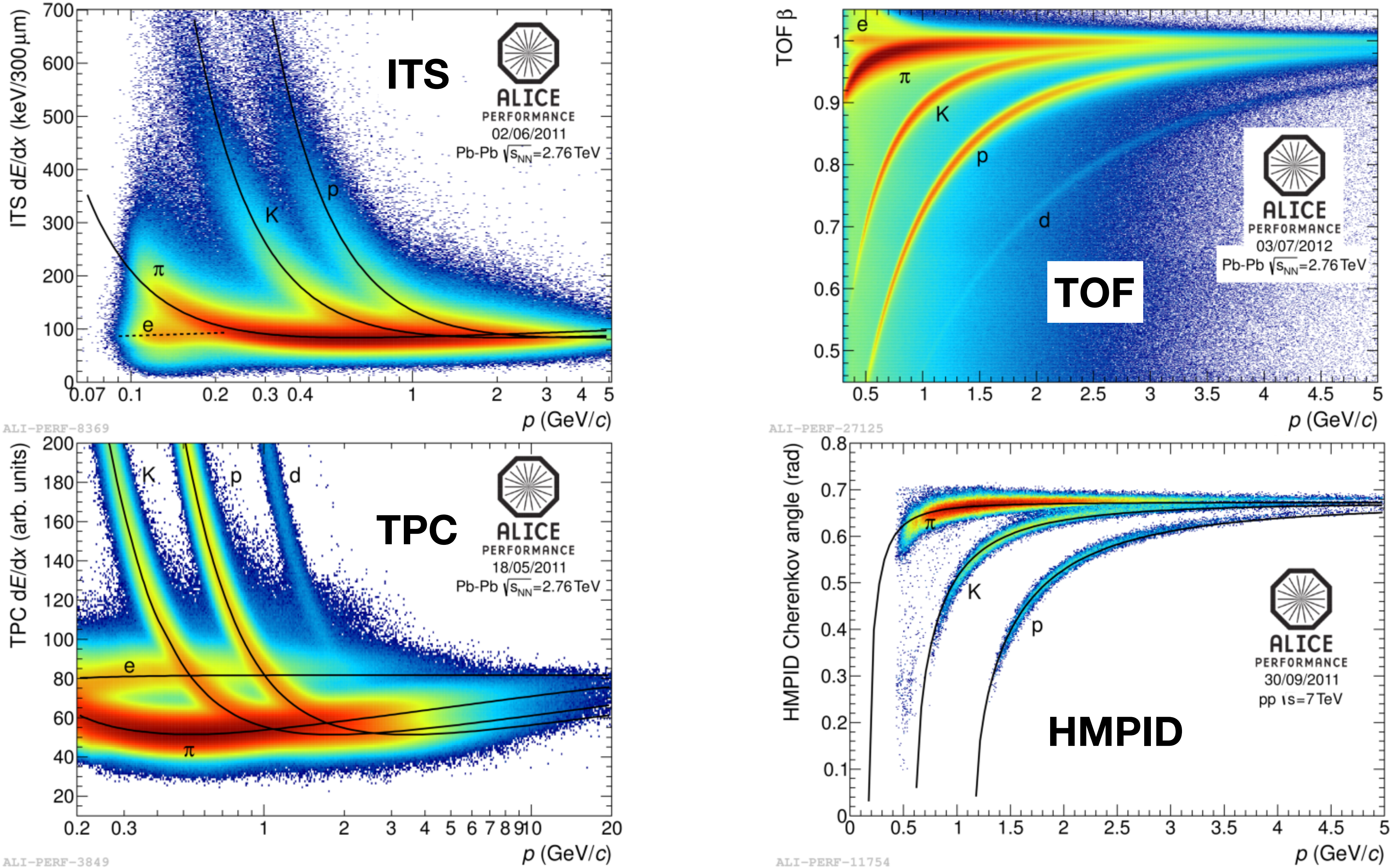


- Mistagging efficiency vs. b-jet efficiency
- Solid lines represent efficiencies with present ML-based method
  Statistical uncertainties shown as width of line
- Dashed lines show conventional, cut-based performance (cf. arXiv:1605.00143)

The present ML-assisted tagging method is very promising, compared to conventional method
– mistagging efficiency lower for c- and udsg-jets
– mistagging efficiencies rise less steep when considering higher b-jet tagging efficiency

R. Haake for ALICE, IML Workshop

Particle Identification



Identity of daughter particles not directly known

Particle identification (PID, many different techniques) correlates with particles identity