Event Reconstruction Techniques

Maurizio Martinelli
University of Milano Bicocca

CN HPC Spoke2 Kickoff meeting
13.10.2022

Subject Areas

Trigger Infrastructure

Vertexing and Tracking

Particle Identification

Events' Classification

Trigger Infrastructure

Trigger Infrastructure

Anomaly Detection and Model-Independent Trigger

- Dini, Gennai, Govoni (MIB) + Pazzini, Tosi, Zanetti (PD)

Heterogeneous and Portable Event Reconstruction

- Di Florio, Pompili (BA)

Anomaly Detection and Model-Independent Trigger

Dini, Gennai, Govoni (MIB)
Tosi, Pazzini, Zanetti (PD)

Trigger for HL-LHC (CMS-oriented)

- **Ultimate aim: a trigger for anomaly detection**
- **FPGA vs GPU: lower performance, but also much lower power consumption**
- **Reasonable aim in the 3 years of the CN HPC: develop expertise on running AutoEncoders on a FPGA infrastructure**
  Learn how to write FPGA firmware and to deal with the FPGA hardware
- **Project: create a minimal testbed of a FPGA cluster of ~3 FPGA to test connections, data transfers and analysis**
  Possibly on site @MIB for developing the hardware expertise
- **Requirements: ~3 FPGA and manpower (1 TD) to create the testbed and write the firmware**
  Training on firmware writing is also needed (WP4?)

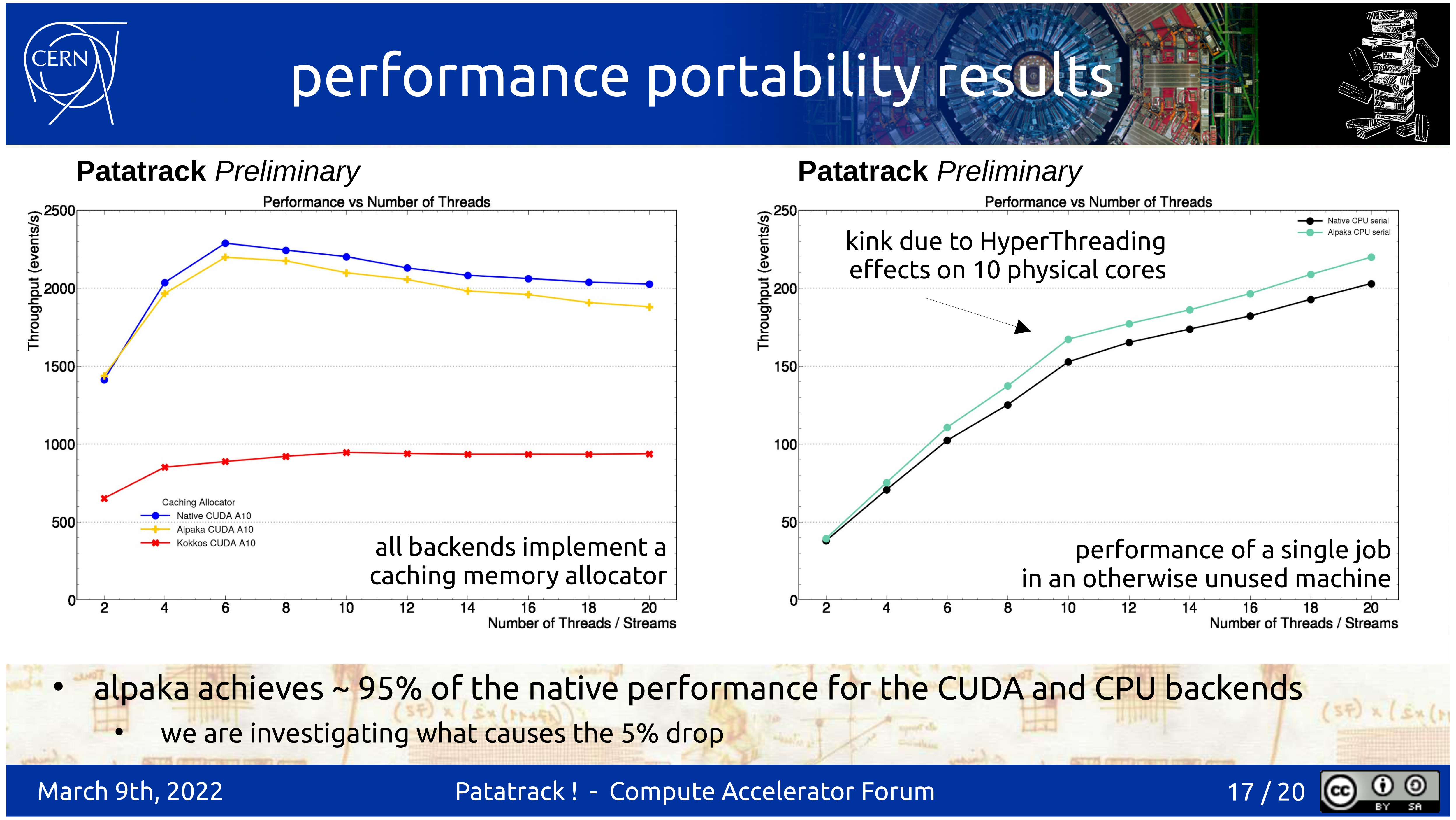
Heterogeneous and Portable Event Reconstruction

Di Florio, Pompili (BA)

Heterogeneous Trigger Reconstruction

- **Aim: 50% of CMS reconstruction code on GPU within 3 years**
- **Use a portability layer to simplify and accelerate the transition**

  Development ongoing on Alpaka, a portability layer that allows to run the same code on various platforms without loss of performance
- **Project: Port most of CMS trigger reconstruction to Alpaka**
- **Testbed already in place**
- **Requirements: manpower (1 TD?)**



Vertexing and Tracking

Vertexing and Tracking

GNN for Pattern Recognition

- Pazzini, Tosi, Zanetti (PD)

ML for 4D Vertexing @ HL-LHC

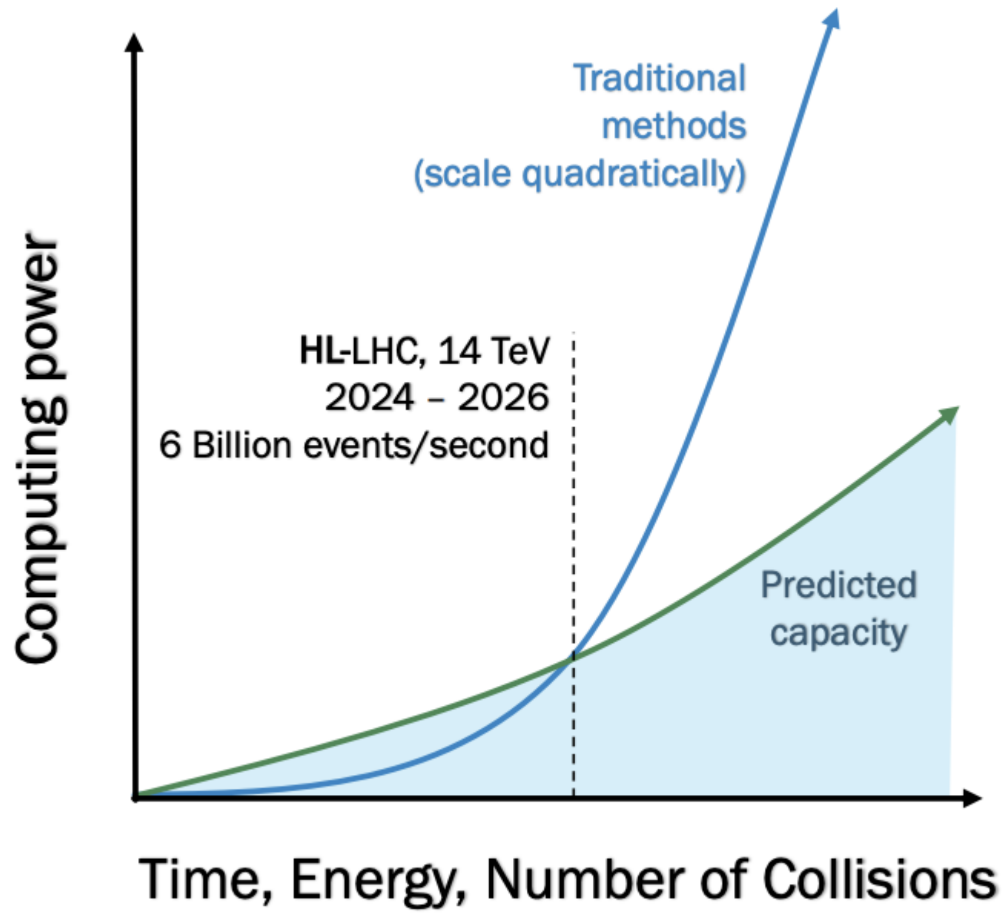
- Candelise, Della Ricca, Zaccolo (TS)

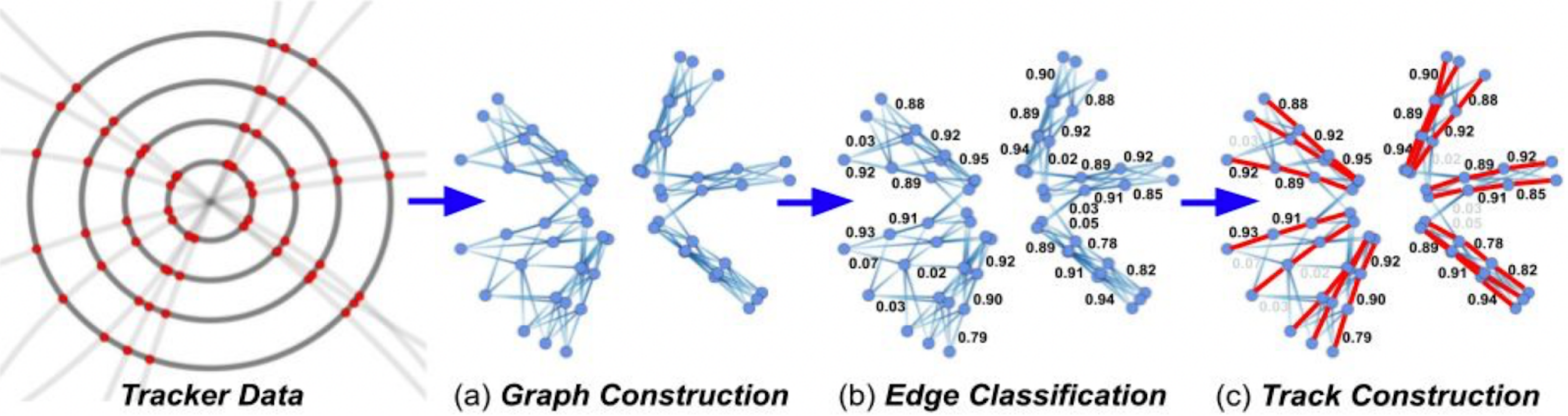
GNN for Pattern Recognition

Tosi, Pazzini, Zanetti (PD)

Trending Topic in HEP

- **Promising performance of GNN in HEP already shown**
  GNNs shown successful in associating hits. They can capture the inherent sparsity of much physics data and their manifold and relational structure
  Challenge is mainly from the very large size of the training graphs ⟶ filtering
- **Goal: test GNN performance on CMS Phase2 simulation constructing the graph by using edge classifier and clustering techniques**
- **Current testbed: lxplus-gpu (T4 92GB RAM)**
  Studies on load and throughput to be done;
  Probably the larger the RAM the better





nodes ↔ hits

edges ↔ tracks

ML for 4D Vertexing @ HL-LHC

Candelise, Della Ricca, Zaccolo (TS)

High Pileup in Various Reconstruction Algorithms

- **Adding the time coordinate to hits may mitigate the problem by reducing combinatorics and secondary hits**
- **Various studies ongoing in ALICE and CMS**
  Multi-charm hadrons and exotic states in high-multiplicity and high-density conditions (ALICE)
  Heavy-flavour tagging of beauty and charm quarks (CMS)
- **Project: Extend tracking and vertexing ML algorithms by adding the time information at track level, in particular for cleaning secondary vertices from pile-up**
- **Testbed: currently 1.1k CPUs @ INFN-TS; considerations ongoing about GPU**
- **Requirements: 1 RTDA + 1PhD to be hired in 2023**
- **Synergies: De Filippis (PoliBA) (tracking @ HL-LHC)**

Particle Identification

Particle Identification

ML for TOP reconstruction in Belle II

- Gaz, Stroili (PD)

ML for Particle ID in Heavy Ion Experiments

- Volpe (BA)

ML for Particle ID @ FCC-ee

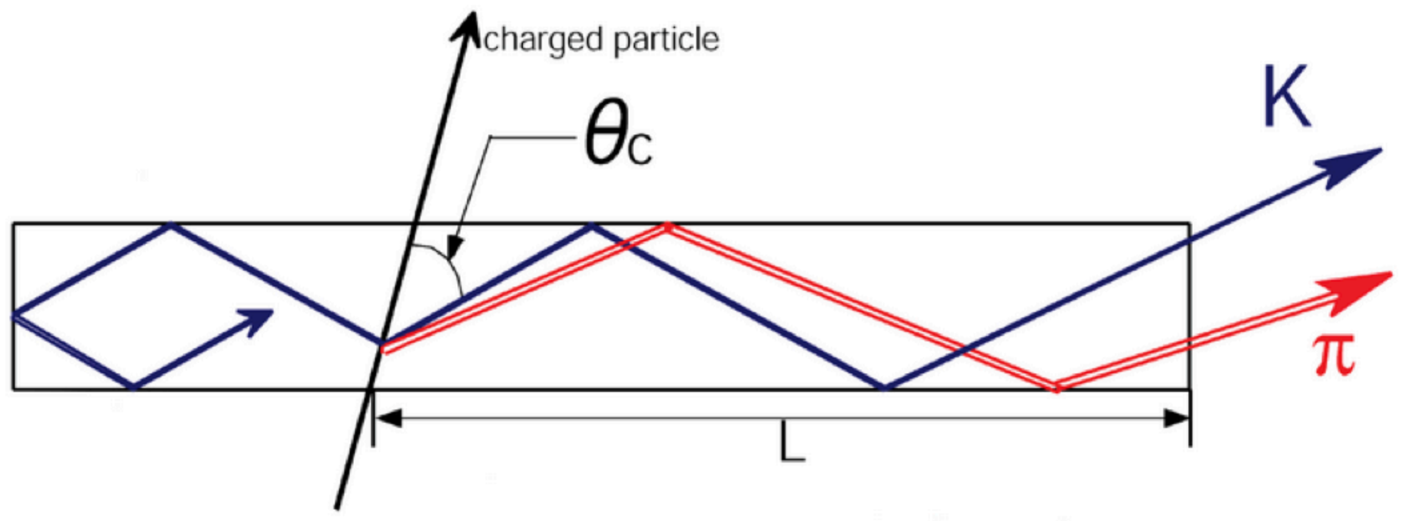
- De Filippis (BA), Gorini (SA)

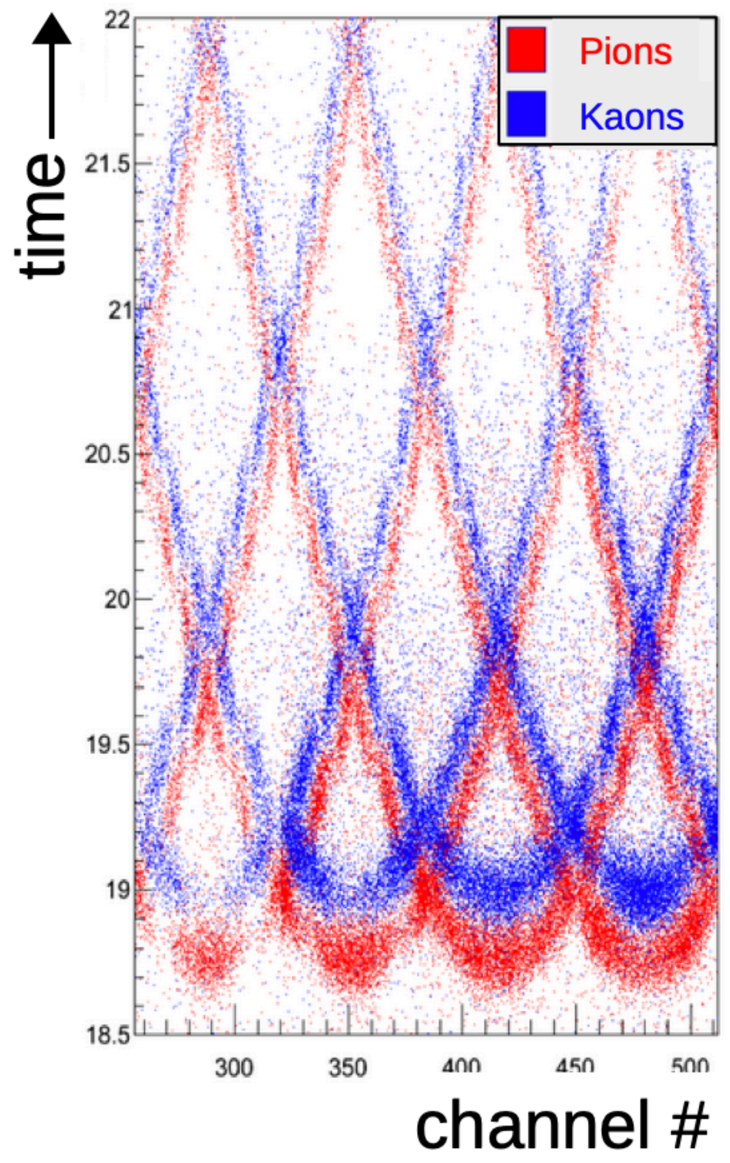
ML for TOP Reconstruction in Belle II

Gaz, Stroili (PD)

Pattern Recognition of Cherenkov Photons

- **TOP in Belle II: a Cherenkov detector**
- **Currently using likelihood to identify particle**
  Based on perfect knowledge of the detector geometry
- **Project: use ML for the pattern recognition**
  The algorithm can learn on data and track changes over time
  Preliminary studies showed potential but extreme sensitivity to training
- **Testbed and Production: cloud resources in PD**
- **Request: 1 PhD**





ML for Particle ID in Heavy-Ion Experiments

Volpe (BA)

Cherenkov Photons in High Multiplicity Environment

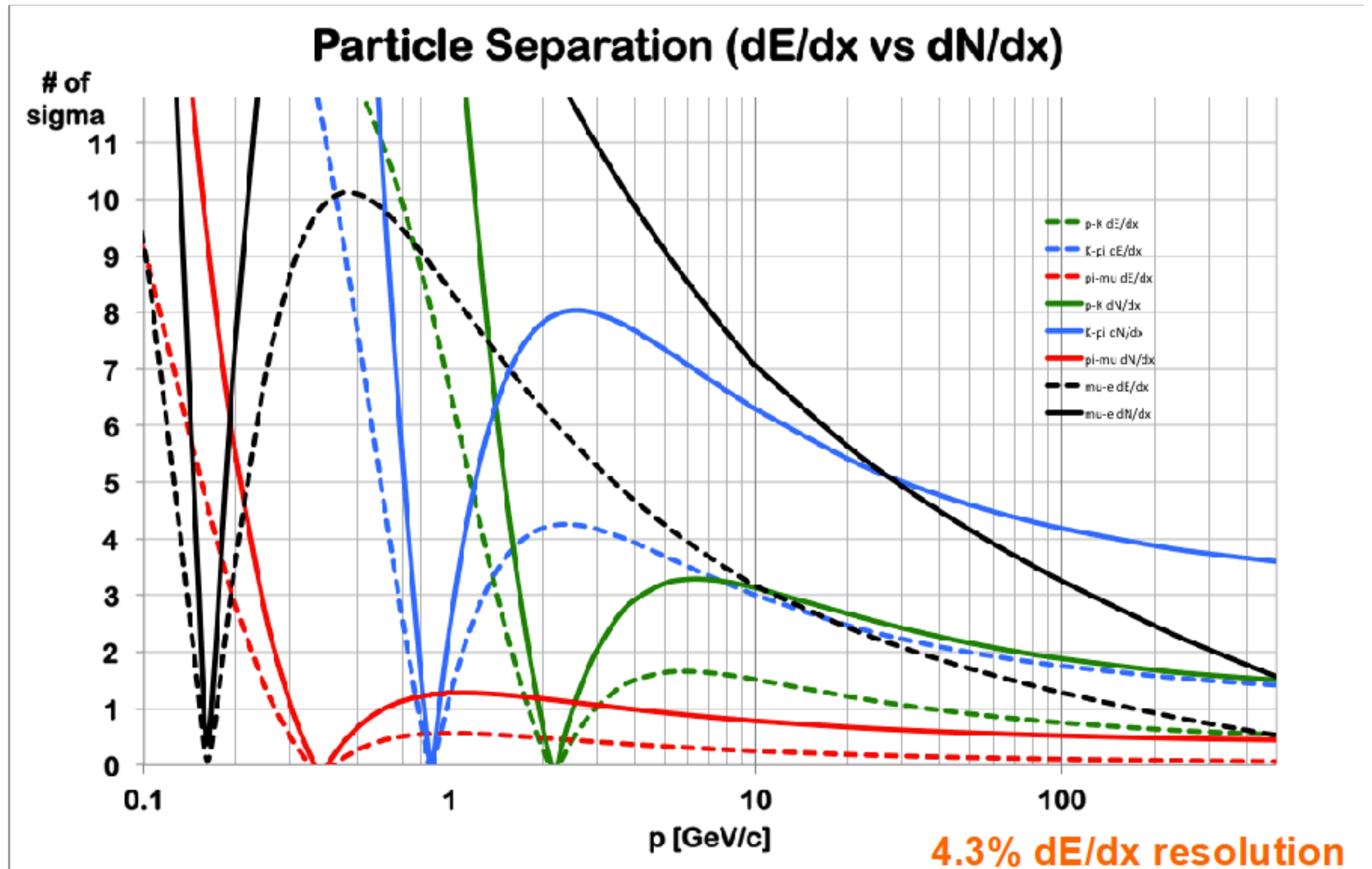
- Particle Identification from momentum and Cherenkov photon emission angle is very challenging in high-multiplicity environments
- Goal: develop a ML approach to Cherenkov rings reconstruction and association
- Testbed: ReCaS (BA) with CPU and GPU
- Personnel: TBD

ML for Particle ID @ FCC-ee

De Filippis (BA), Gorini (SA)

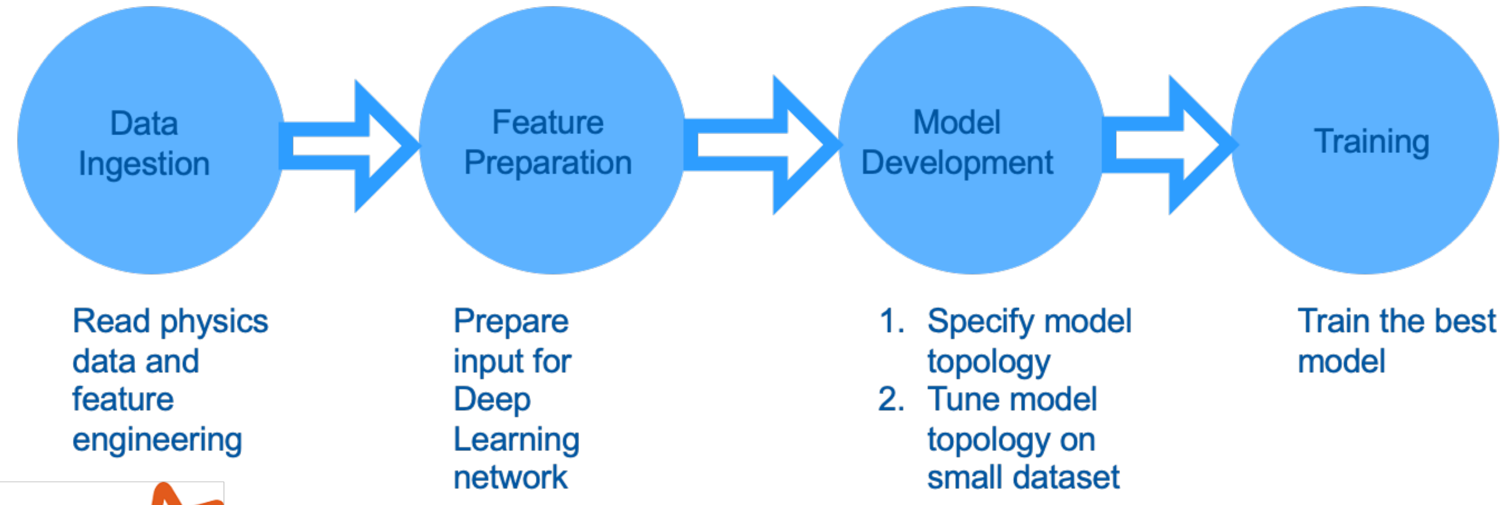
ML on Energy Deposit

- **Goal: develop a charged particle identification for pions, kaons, protons, muons and electrons, based on the sub-detectors response (mainly a drift chamber for IDEA experiment at FCC-ee), by using machine learning techniques (one-vs-rest, one-vs-one and multi-classification)**
  - DNN and GNN could be the best tools to combine the information from various detectors effectively, to be compared with the cluster counting technique (dN/dx)
  - the performance of the machine learning techniques for PID are also measured in terms of signal to background discrimination in the context of physics analysis for heavy flavor measurements
- **Pipeline and testbed:**
  - deep learning pipeline to be setup on a testbed
  - cluster of 5 machine handled by Apache Spark + Analytics Zoo/BigDL for hyperparameter optimization (using AnalyticsZoo / Big DL)
  - 5 TB data as an input



4.3% dE/dx resolution
80% cluster counting efficiency

Deep Learning Pipeline for Physics Data



Built with Apache Spark + Analytics Zoo + Python Notebooks

Events' Classification

Events' Classification

ML Regression for H$\rightarrow\tau\tau$

- Di Nardo, Gennai (MIB)

Anomaly Detection and Graph Neural Networks

- Ippolito (RM1)

Heavy-Flavor Tagging in ATLAS

- Fazio, Meoni, Tassi (CAL)

ML Regression for H→ττ

Di Nardo, Gennai (MIB)

ML Regression for H→ττ

- **Challenging final state for Higgs and BSM**
  Large di-jet background contamination; non-leptonic triggers at low momenta
- **Goal: develop a network based on kinematics output to reconstruct faster the events; future goal to add more low-level quantities in the network**
- **Testbed: actual training on CPU (Tier3-CMS @MIB); GPU may be needed for adding low-level quantities, but they are available on site**
- **Production: available resources on site.**

Anomaly Detection and Graph Neural Networks

Ippolito (RM1)

Finding What We Are Not Searching For

- **Use only background sample for training and identify events that are different from the rest**
- **Ongoing Project: Train GNNs on toys from LHC Olympics dataset (topocluster in events with at least two large-radius jets)**
  In the network each node is a topocluster and connections are mad by edges that are weighted by the relative distance among the nodes
- **Goal:**
  1) develop an algorithm to discriminate between {1,2,multi}-prong jets
  2) classify the event by adding further info to the topocluster (e.g. E_T)
  Will need porting the proof of principle to ATLAS data
- **Resources: currently working on GPU RTX 3090 24GB on site**
  Dataset: 1M events, 5GB; limited by I/O: processing 50 events/s with 700 topoclusters
- **Testbed: 1 GPU more powerful; improved I/O (optimised software layer)**
- **Production: Much larger datasets**

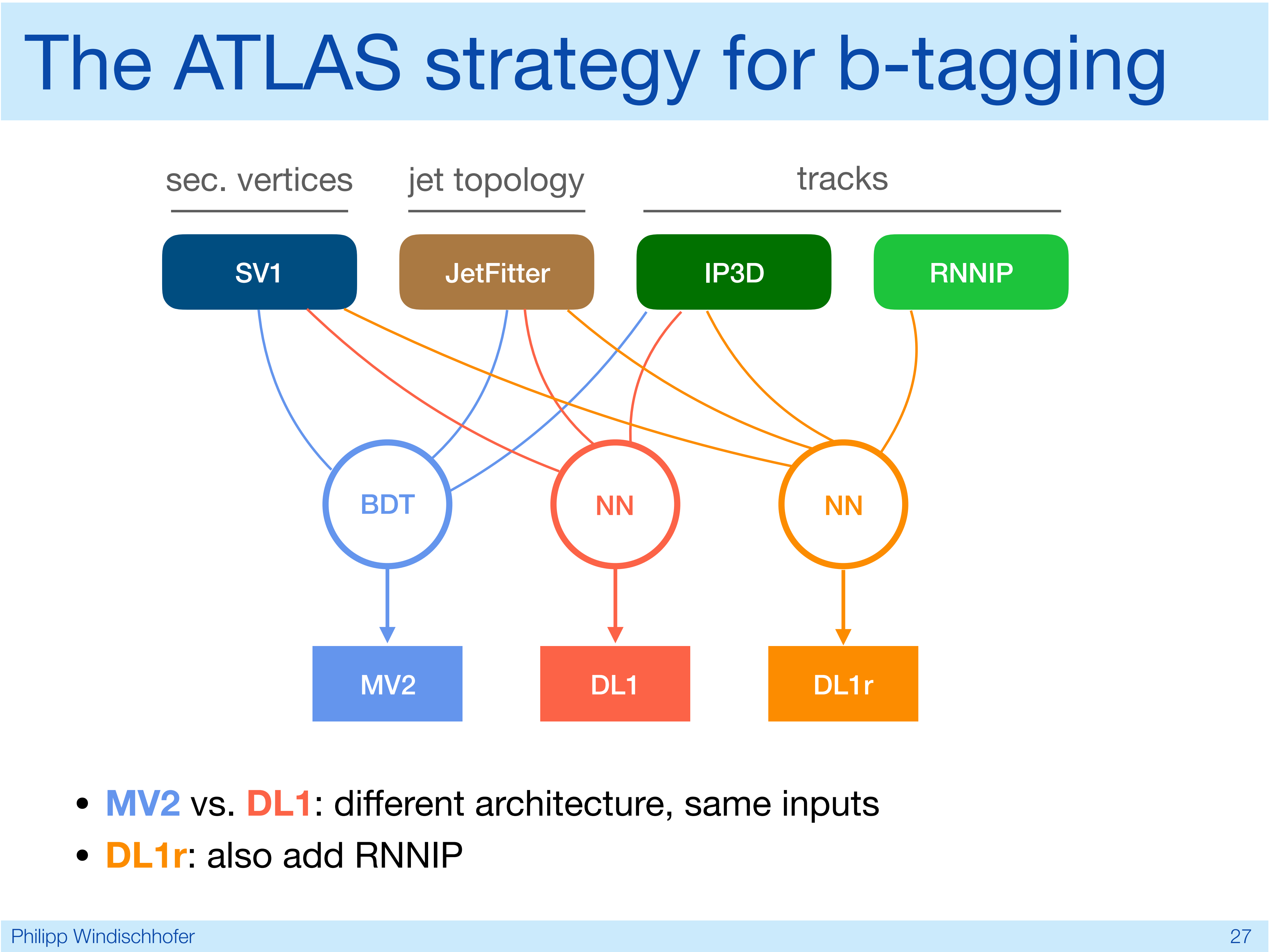
UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

Heavy Flavor Tagging ATLAS

Fazio, Meoni, Tassi (CAL)

Develop b-tagging Algorithms

- **Improve DL1, DL1r (latest algorithms by ATLAS)**
  High-level taggers built on top of ATLAS low-level taggers in a NN
  Alternatively work on the "Deep Sets" approach
- **Goal: Validate different NN topologies; use more low-level information; be as much as possible detector-independent**
- **Testbed: currently ReCaS Cosenza available (~3.5k CPU); may need GPU**
- **Request: Manpower (1 TD?)**



Wrap Up

Requests for Test Bed Activities

Topic	GPU	FPGA	DISK	Personnel
Anomaly Detection and Model-Independent Trigger	-	~3	-	1 TD
Heterogeneous and Portable Event Reconstruction	-	-	-	1 TD
GNN for Pattern Recognition	1	-	?	-
ML for 4D Vertexing @ HL-LHC	1	-	?	1 TD + 1 PhD
ML for TOP Reconstruction in BELLE II	-	-	-	1 PhD
ML for PID in Heavy-Ion Experiments	-	-	-	TBD
ML for PID @ FCC-ee	-	-	5 TB	-
ML Regression for $H\rightarrow\tau\tau$	-	-	-	-
Anomaly Detection and GNN	1	-	100 GB	-
Heavy Flavor Tagging in ATLAS	1	-	5 TB	1 TD

Requests for Production

No Requests

- **A few projects have already the necessary resources for the testbed and production phase**
  ML for TOP Reconstruction in Belle II; ML for PID in Heavy-Ion Experiments;
  ML Regression for H→$\tau\tau$
- **The trigger infrastructure ones do not really need a production environment since the testbed will prove the technology to use in the experiment at CERN**

Extrapolations

- **All the other projects will need a production system with much larger capabilities than the testbed to produce the final result**
- **The common denominator of these projects is the use of GPU, therefore I would naively suggest that a reasonably large cluster with a few powerful GPUs (5? 10?), fast disks (2TB/GPU?) and a reasonably large data storage (50TB?) will be enough.**

Summary

Many thanks to everybody that has provided feedback on my requests and helped me draft this presentation

I’ve learned a lot about many different and promising areas of research at the boundary of HEP and ML

I hope that it will help the WP coordinators to get a more consistent picture of what we need

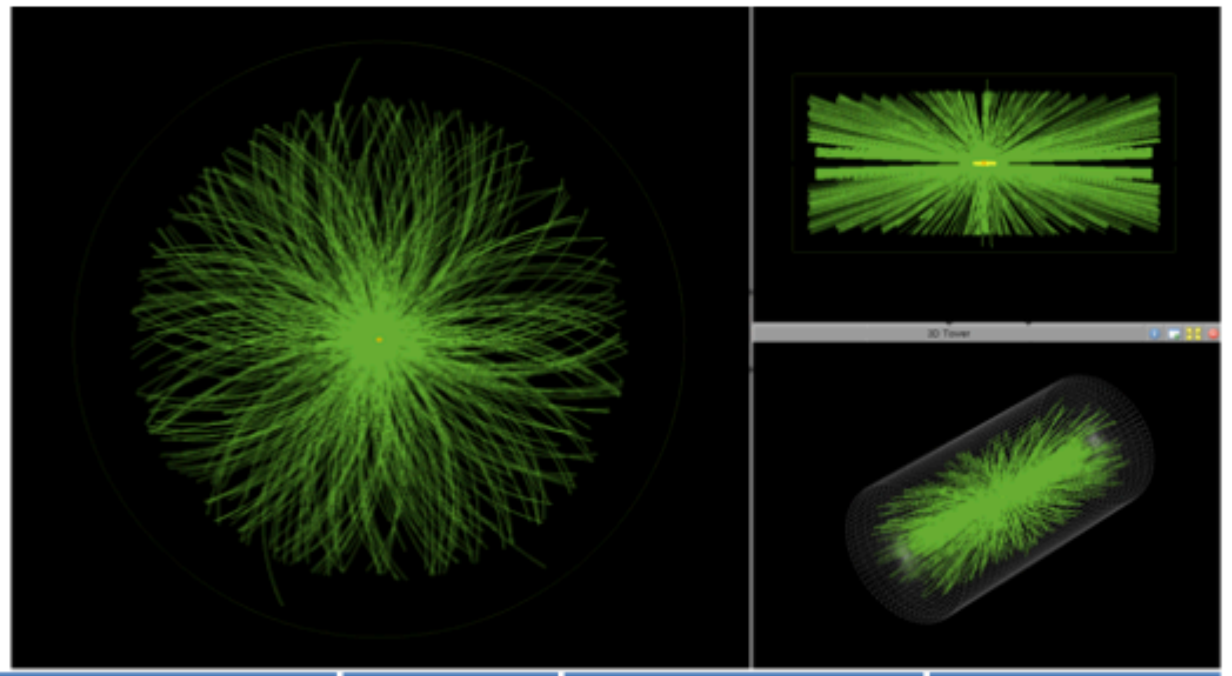
UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

Spares

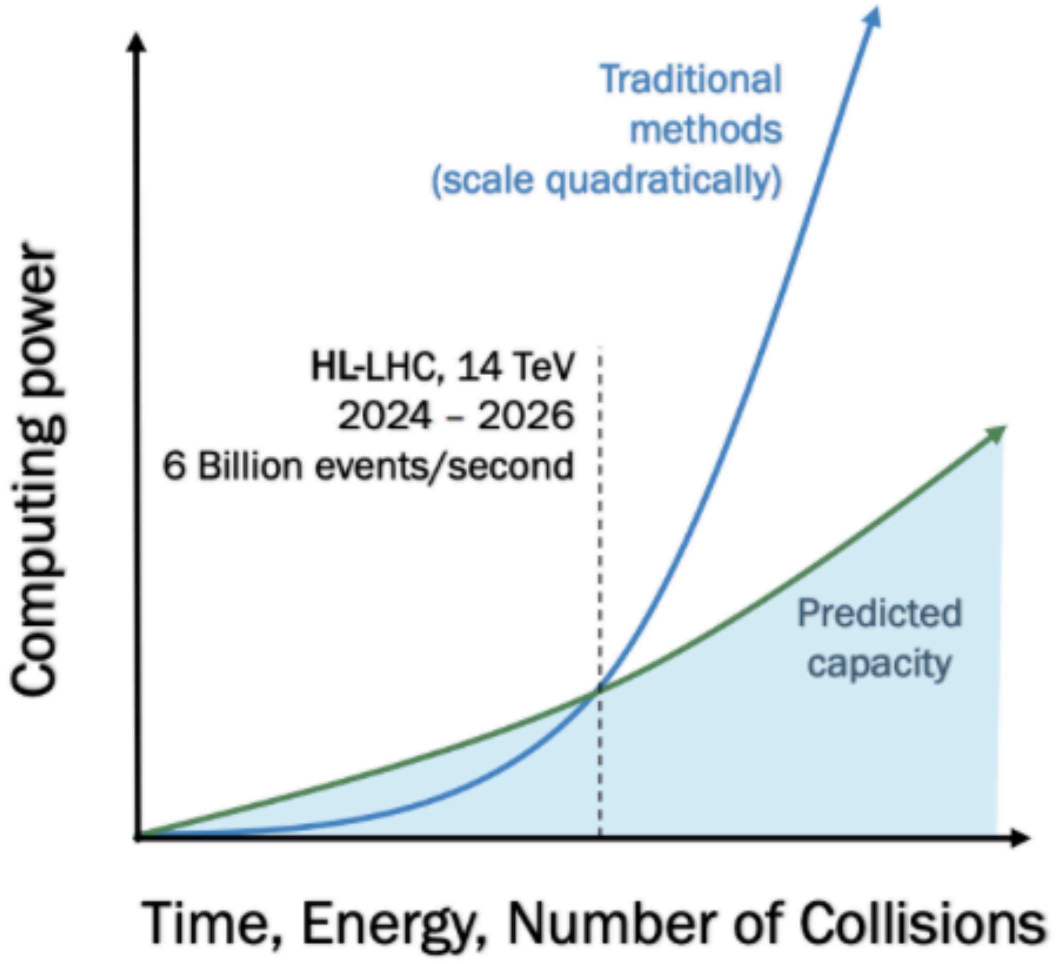
GNN for track reconstruction

Courtesy Mia Tosi (PD)

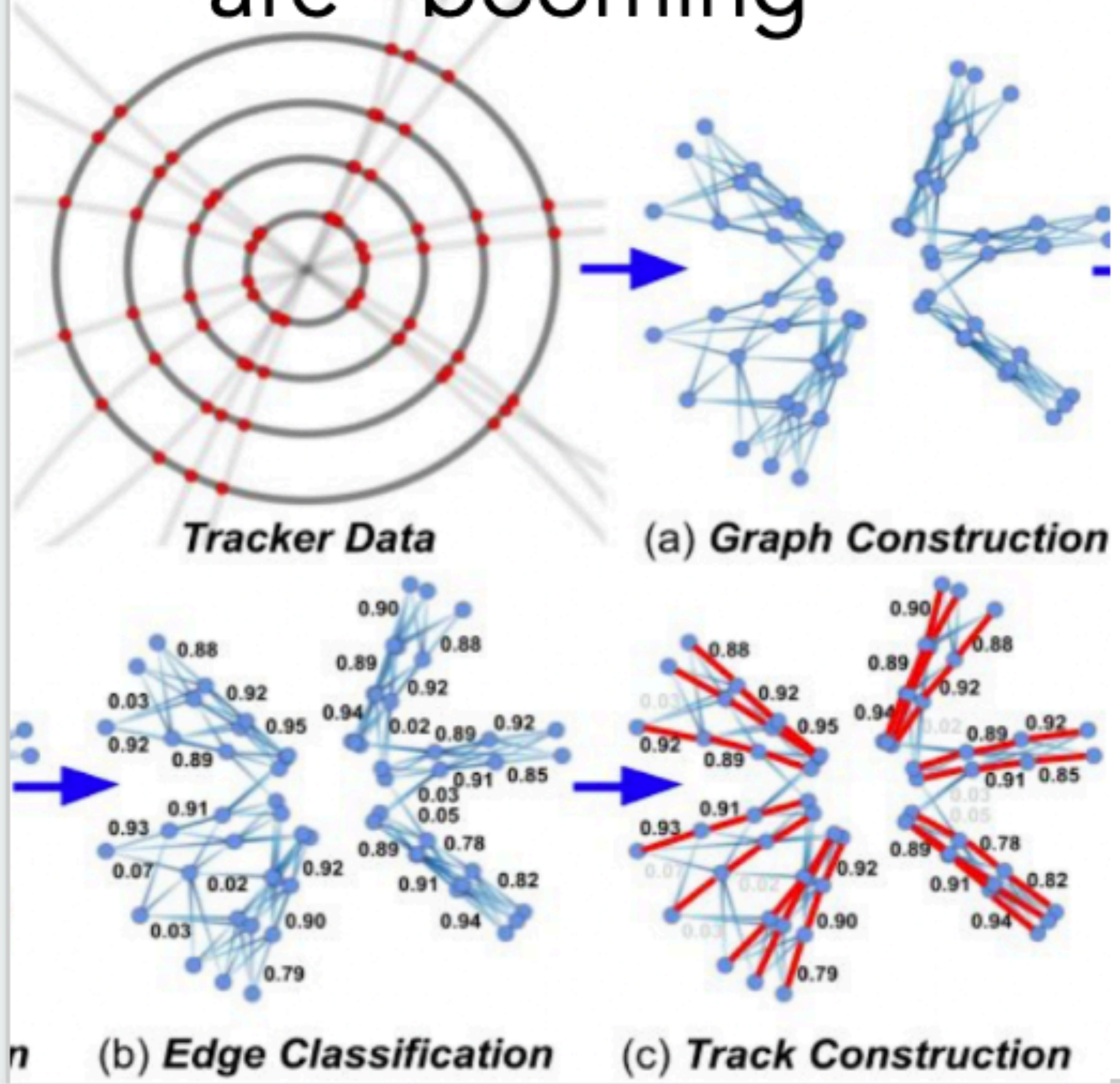
HL-LHC brings many computational challenges
→ the number of tracks increases by x10
→ the vertex density by x5
and the current –traditional– approach scales almost quadratically !



	Run1	Run2-Run3	Run4
<PU>	20	30-60	150-200
#tracks/event	~500	~1200	~12000



Graph Neural Networks are "booming"



graphs can capture
- inherent **sparsity** of much physics data
- the **manifold** and **relational structure** of much physics data

⇒ conversion to and from graphs can allow **manipulation of dimensionality**

track properties may be estimated using ML and GNNs
- graph nodes → tracker hits
- graph edges → tracks

⇒ current work is focused on graph construction using edge classifier and clustering techniques

0

ultimate goals

Courtesy Mia Tosi (PD)

computational performance :
it would be desirable to have at least a $\times\mathbf{1.5-2}$ **improvement**

"but current gains are unknown
and not possible to estimate without a detailed study" –O&C R&D snapshot

currently, server equipped
w/ T4 and 92 GB of RAM

physics performance :
it would be desirable to
- keep the current efficiency and possibly reduce the fake rate
- consider the displaced tracks as well
- exploit the time information

questions :
- which is the cost in terms of RAM ?
- which is the robustness w.r.t. the data taking conditions ?
  (both the bad components and the misalignment effects)

1

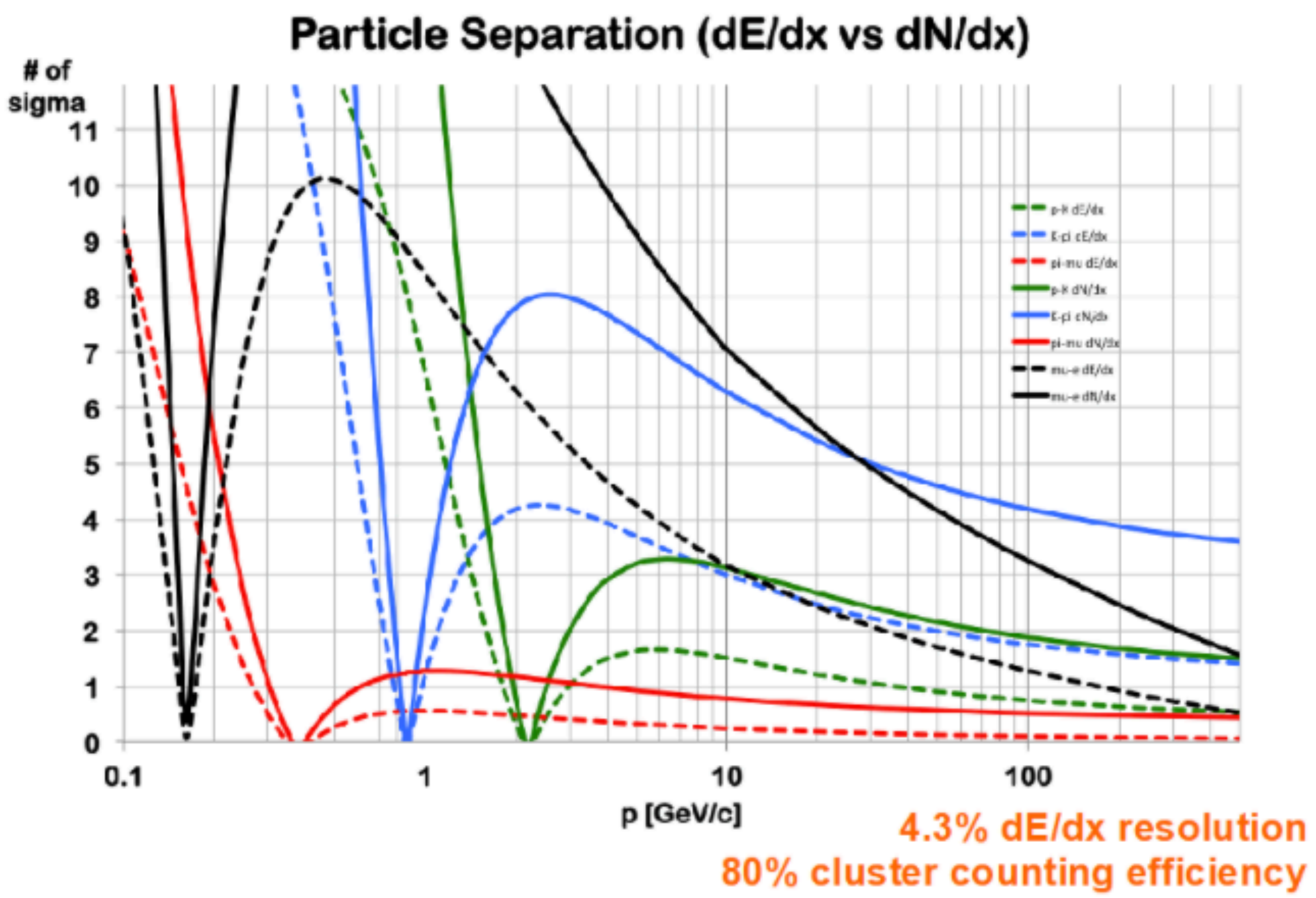
Courtesy Nicola De Filippis (BA)

ML for Particle ID @ FCC-ee

De Filippis (PoliBA), Gorini (UniSA)

Goal: develop a charged particle identification for pions, kaons, protons, muons and electrons, based on the sub-detectors response (mainly a drift chamber for IDEA experiment at FCC-ee), by using machine learning techniques (one-vs-rest, one-vs-one and multi-classification)

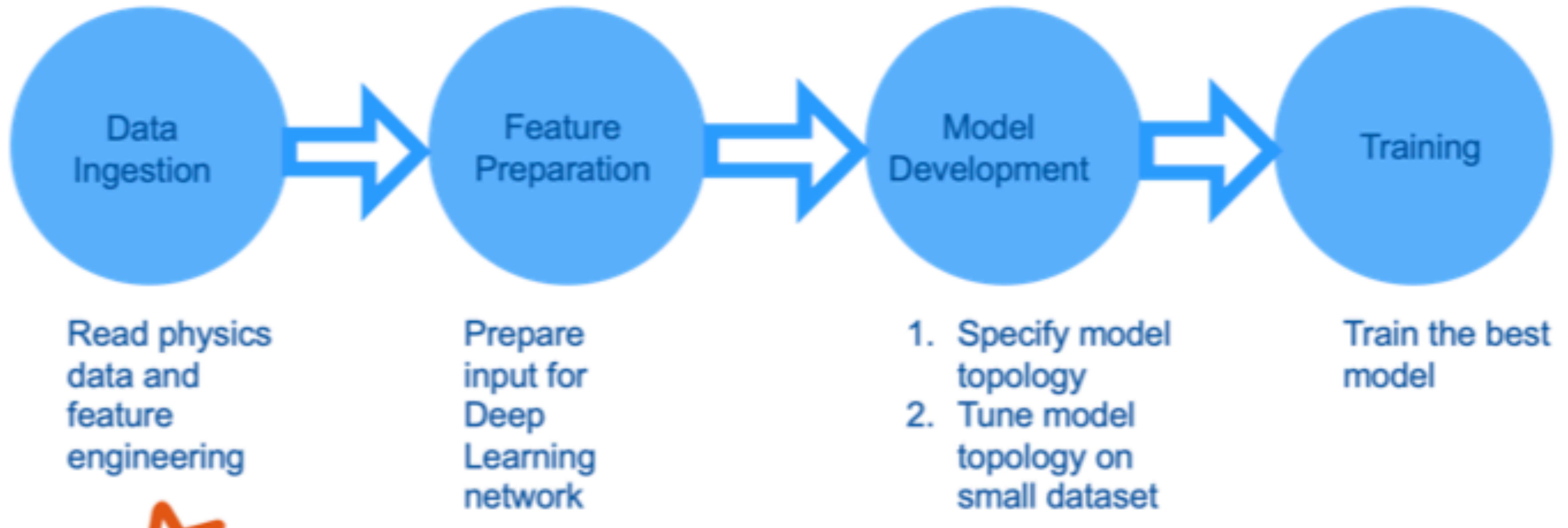
- DNN and GNN could be the best tools to combine the information from various detectors effectively, to be compared with the cluster counting technique (dN/dx)
- the performance of the machine learning techniques for PID are also measured in terms of signal to background discrimination in the context of physics analysis for heavy flavor measurements



Pipeline and testbed:

- deep learning pipeline to be setup on a testbed
- cluster of 5 machine handled by Apache Spark + Analytics Zoo/BigDL for hyperparameter optimization (using AnalyticsZoo / Big DL)
- 5 TB data as an input

Deep Learning Pipeline for Physics Data



Built with Apache Spark + Analytics Zoo + Python Notebooks