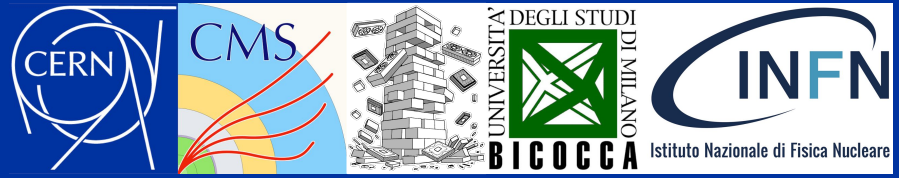
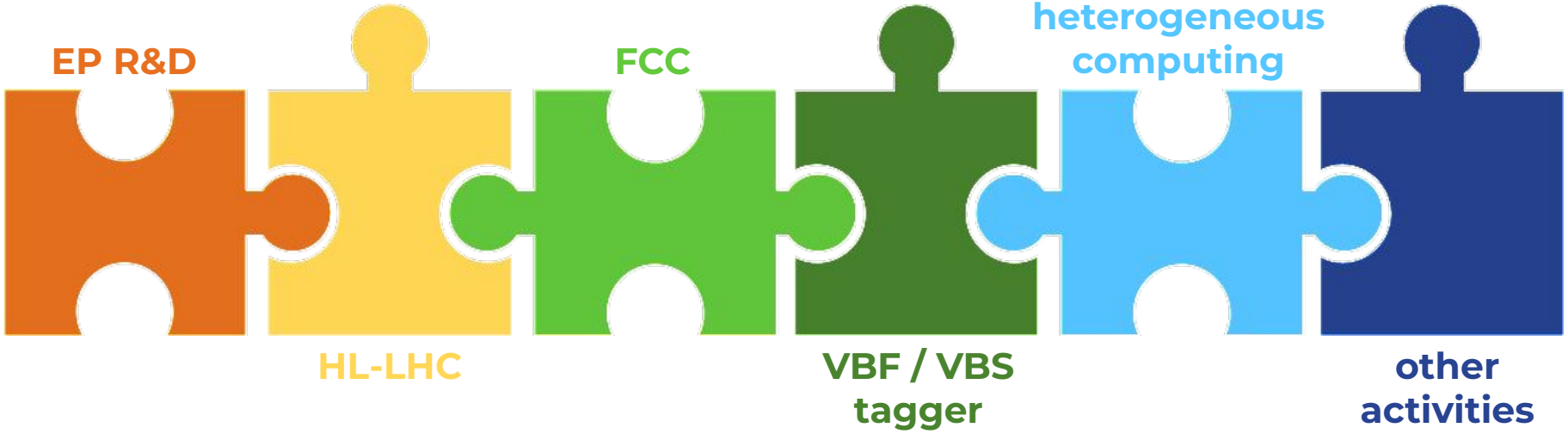


Time information in event reconstruction at HL-LHC and future colliders and VBS/VBF tagger

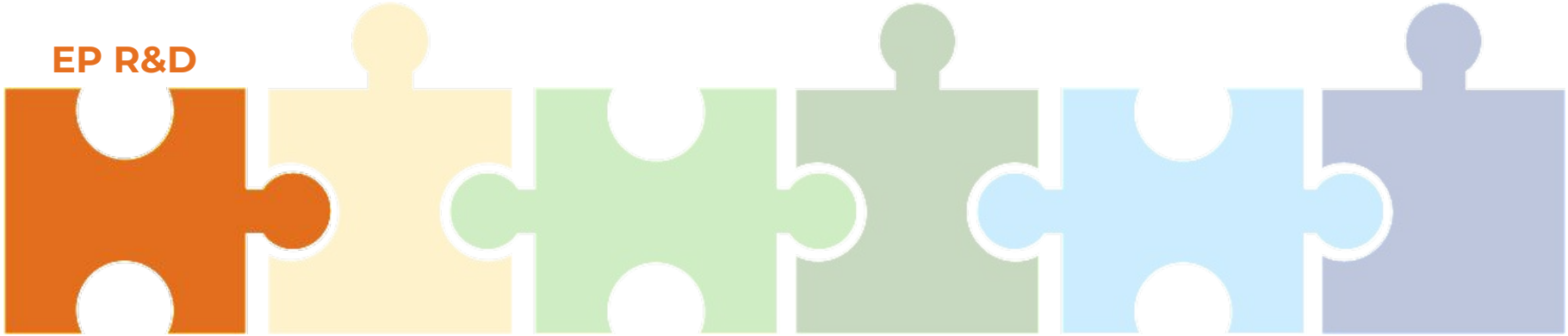
Aurora Perego

End of year PhD report - 20/09/24

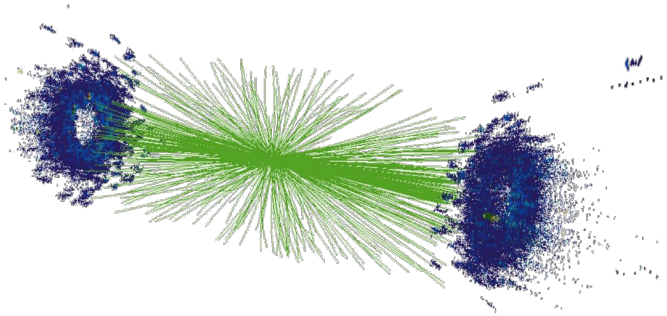
Outline


EP R&D
HL-LHC
FCC
VBF / VBS tagger
heterogeneous computing
other activities

EP R&D

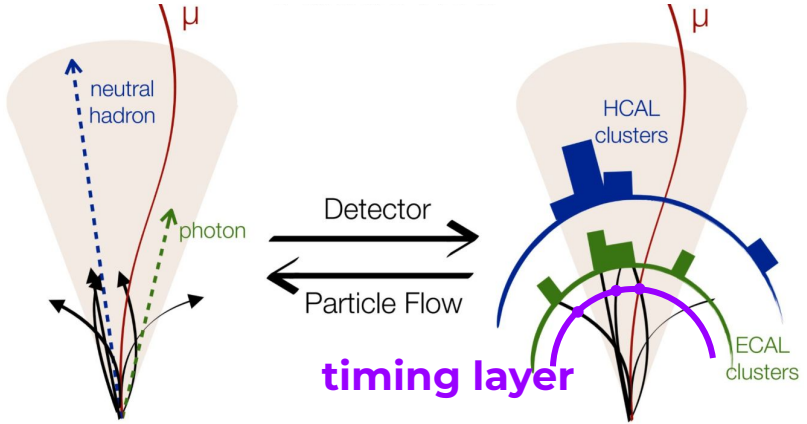


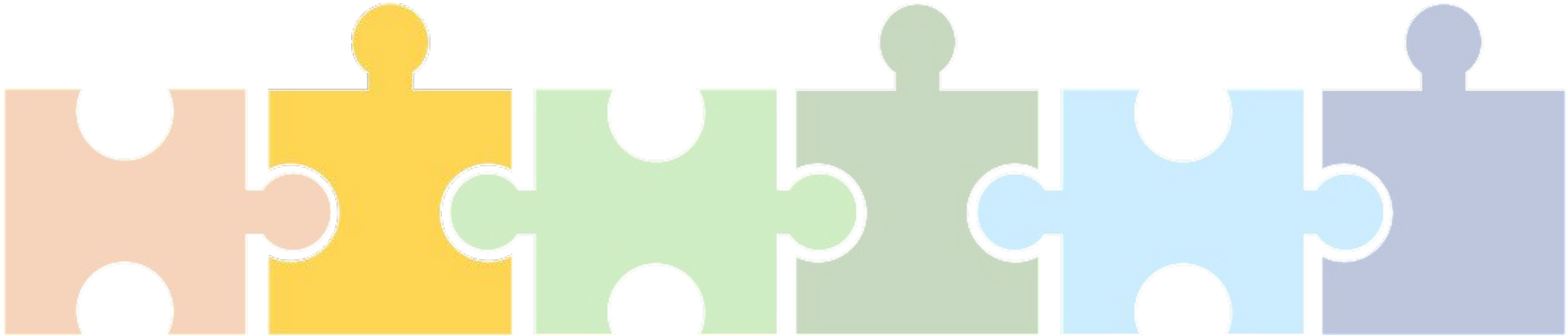
PhD project within EP R&D at CERN



- My project at CERN is part of the **EP R&D** Phase 2 effort, specifically in the software group
- The goal of EP R&D is to address the technological challenges of future experiments
- My task is **calorimetry** reconstruction with the usage of **timing** information

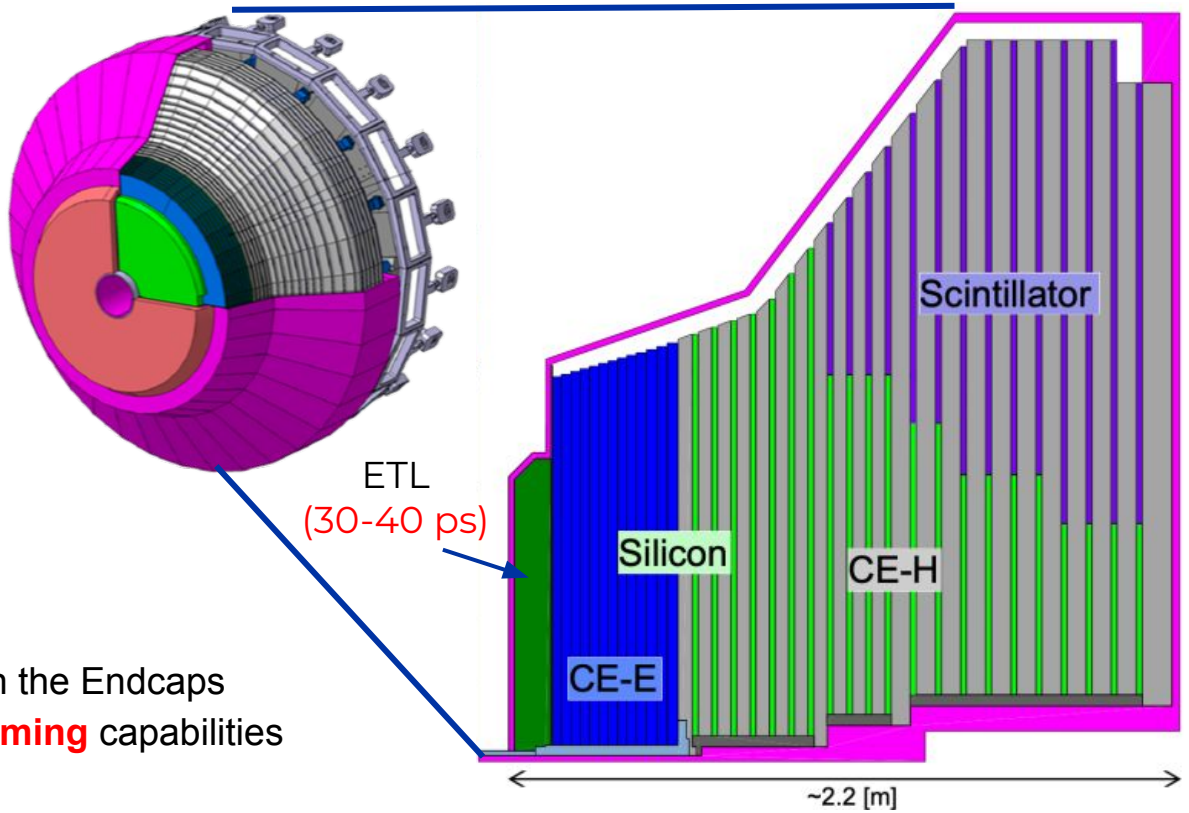
- Both **HL-LHC** and **future colliders** are covered in the project, with developments made for the former that can be adapted for the latter
- **Combination** with **tracker** information and **heterogeneous** software are also in the plan





HL-LHC

HL-LHC: CMS experiment

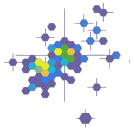


HGCAL
(floor precision of 20ps on ToA)

47 layers
(26 in CE-E
+ 21 in CE-H)
~31k modules
~**6M** channels

Phase-2 **upgrades** in the Endcaps
New detectors with **timing** capabilities

TICLv5 structure



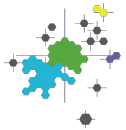
RecHits

heterogeneous

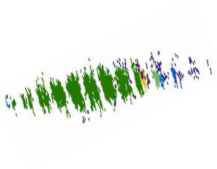
CLUE 2D CE-E

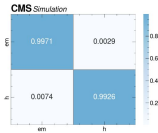
CLUE 2D CE-H

CLUE 2D Scintillator

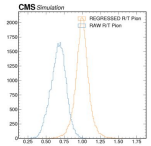


CLUE 3D Global iteration





PID



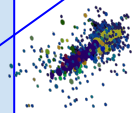
Energy regression and PID

Tracksters link

FastJet (KT,AKT,CA)

Recovery + Skeletons

SuperClustering DNN



Tracks and MTD (configurable)

PFCandidates

TICLCandidates

General interpretation

EGamma interpretation

E/Gamma GSF Tracking

The Iterative CLustering (TICL)

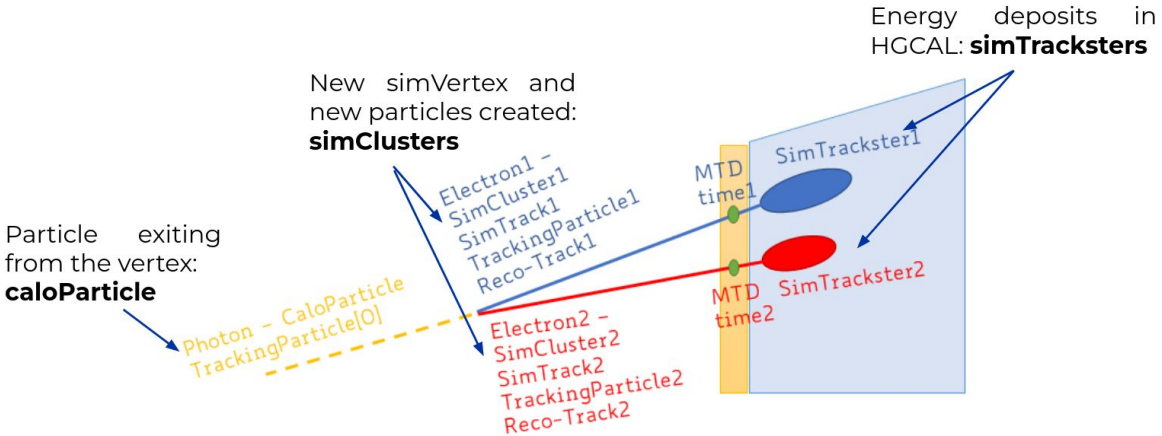
TICL is a modular software framework integrated in CMSSW used to reconstruct particle showers in HGCAL from hits to the final particle flow candidates

TICLv4
- pure tracksters
- room for improvement in hadron reconstruction

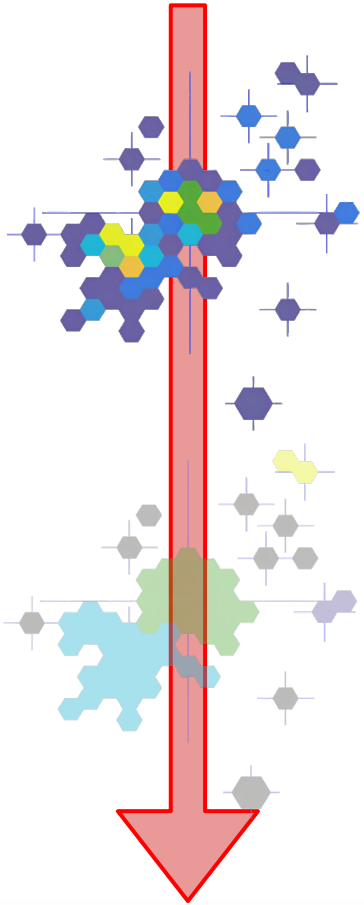
TICLv5
- improved framework
- better usage of timing
- improved e/gamma and hadron reconstruction (new linking+recovery algorithm)
- split of algorithms parameters in different regions of HGCAL
- new ML models (DNN for superclustering and CNN for energy regression and particle Id)

Simulation and validation

- For each particle that leaves hits in HGCAL a **simTICLCandidate** is created
  - represents the **best reconstruction** that can be achieved
  - contains the **reconstructed track** that is best associated with the simulated particle
  - knows the **true** energy and pid of the particle, but also the **reconstructed energy** in HGCAL
- hit by hit associators between the simulated and reconstructed TICLCandidates
- used to develop a dedicated validation to monitor the performance of the HGCAL reconstruction and linking with the track

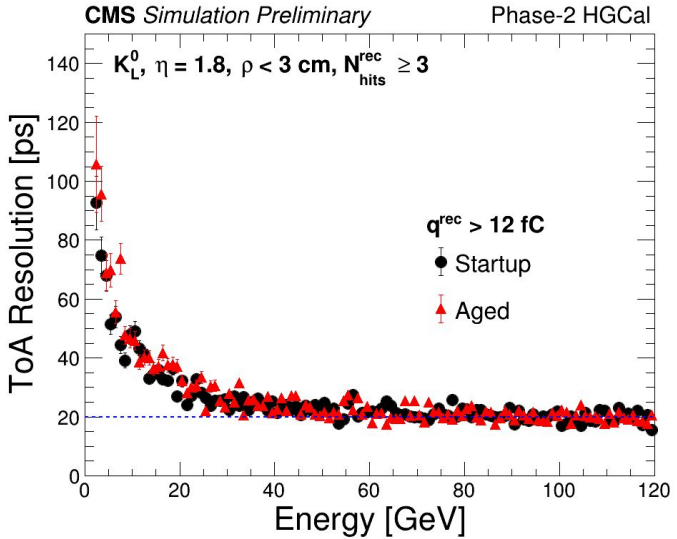


Time information in TICL

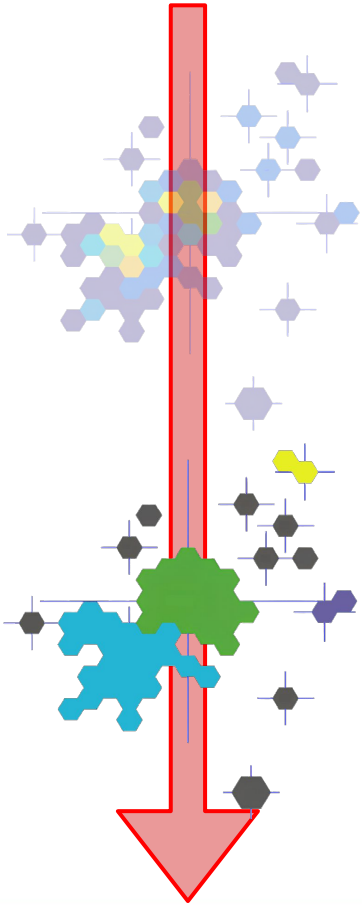


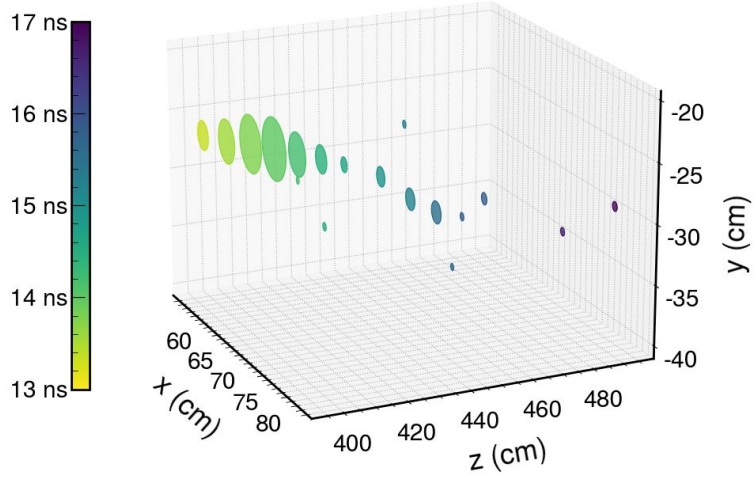
- Each reconstructed **hit** with a charge **deposit above** a certain **threshold** has a **time** associated with it
- The **error** on the time corresponds to the expected resolution based on its **S/N ratio** and is **inversely proportional** with respect to the rechit **energy**

$$\sigma_t = \sigma_{jitter} \oplus \sigma_{floor} = \frac{A}{S/N} \oplus C$$



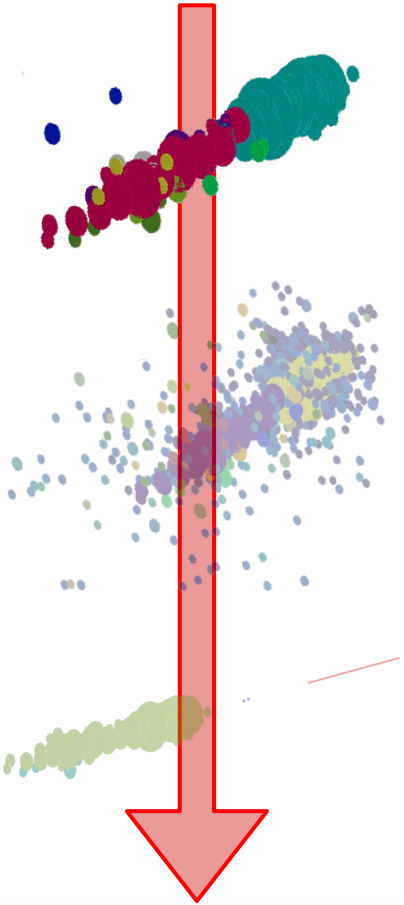
Time information in TICL



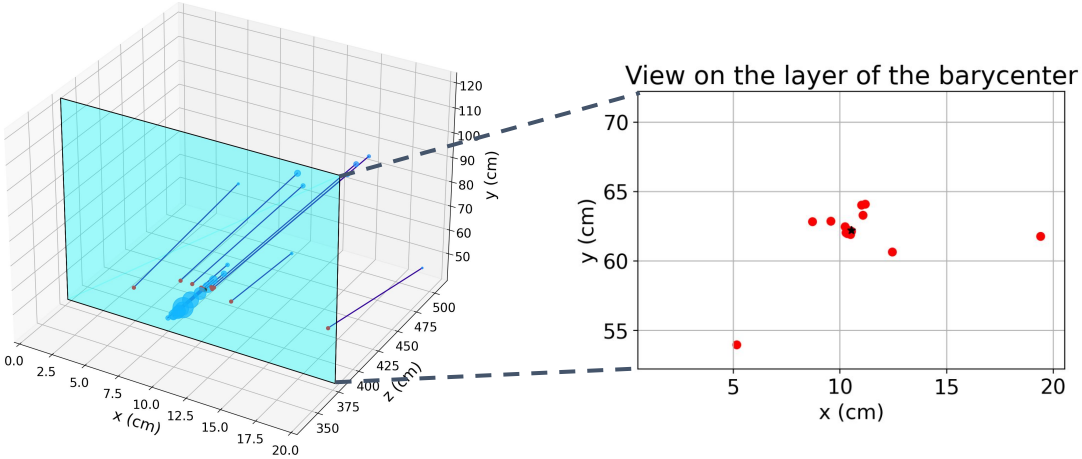


- The time for **layer clusters** with at least three hits with time information is computed as the **average** of the **time** of the **rechits** weighted on their error (a time window with the highest density of hits is used)
- Again, the **higher** the layer cluster **energy**, the **better** the **resolution** on the time

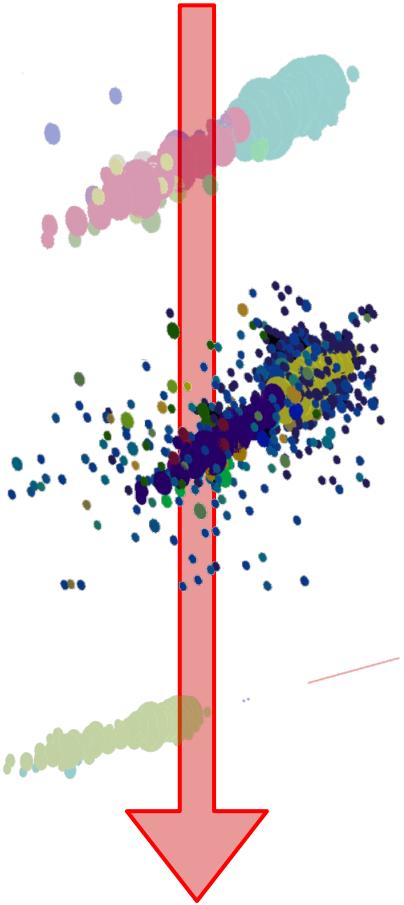
Time information in TICL



- The layer clusters times are projected onto a **common surface** (HGCAL layer where the barycenter is) and their **weighted average** is computed to obtain the trackster time. To improve the resolution, only clusters within 3 cm from the shower axis are considered.

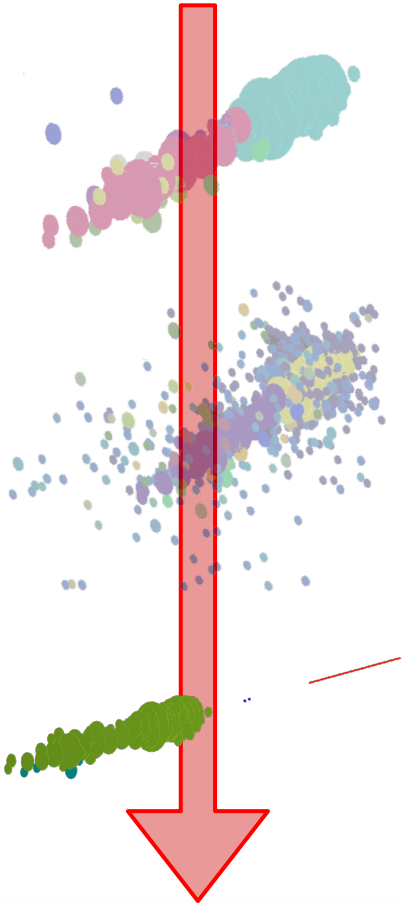


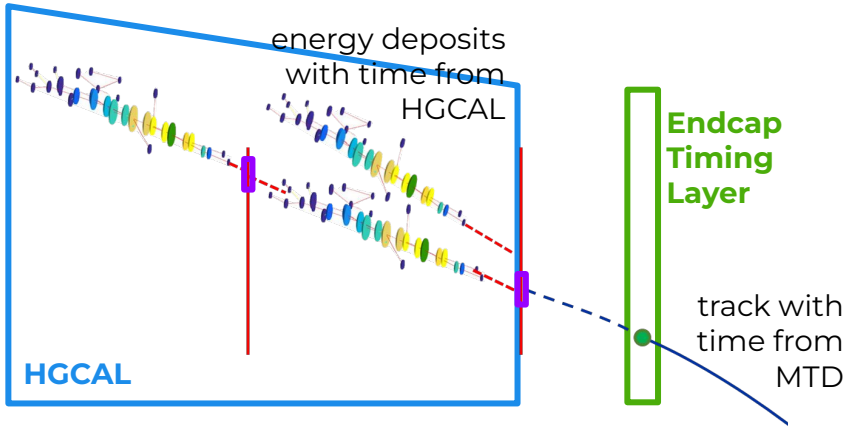
Time information in TICL



- The same happens for the tracksters obtained with the linking algorithm (skeletons and recovery tracksters)

Time information in TICL

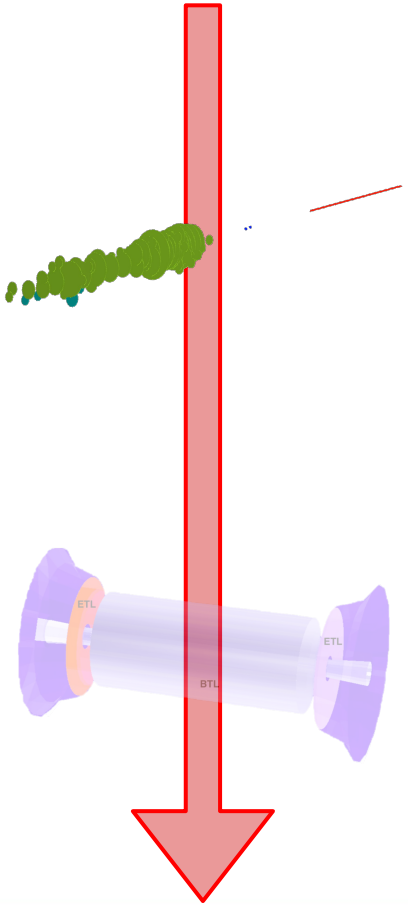




- During the linking with the track a time compatibility check is performed between the track and the trackster

$$\left| \frac{d_{\mathrm{MTD,trackster}}}{c} - (t_{MTD} - t_{trackster}) \right| < 3 \cdot \sqrt{\sigma^2_{t_{MTD}} + \sigma^2_{t_{trackster}}}$$

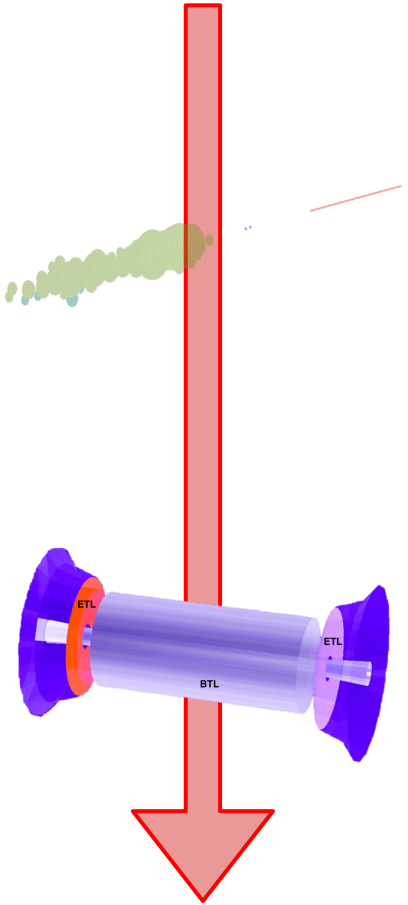
Time information in TICL



- The time of the TICLCandidate is obtained **propagating** the time of the trackster(s) back to:
  - the **origin** in a straight line if no track has been linked
  - the **point of closest approach** to the beamspot following the track if present

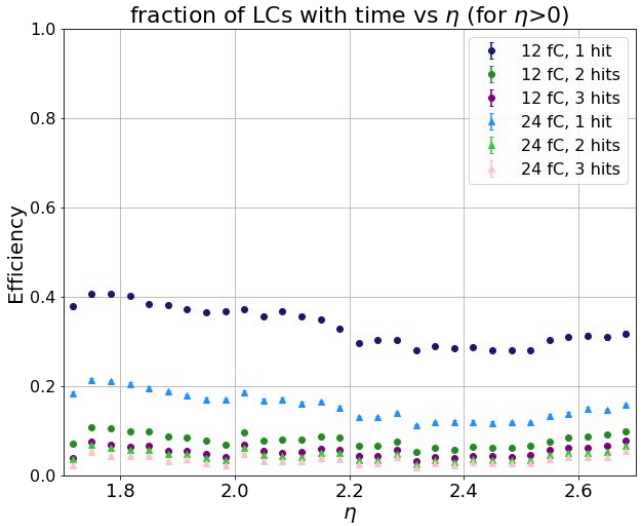
  at the speed of:
  - the particle, computed with MTD information
  - light, if MTD information is not available

Time information in TICL



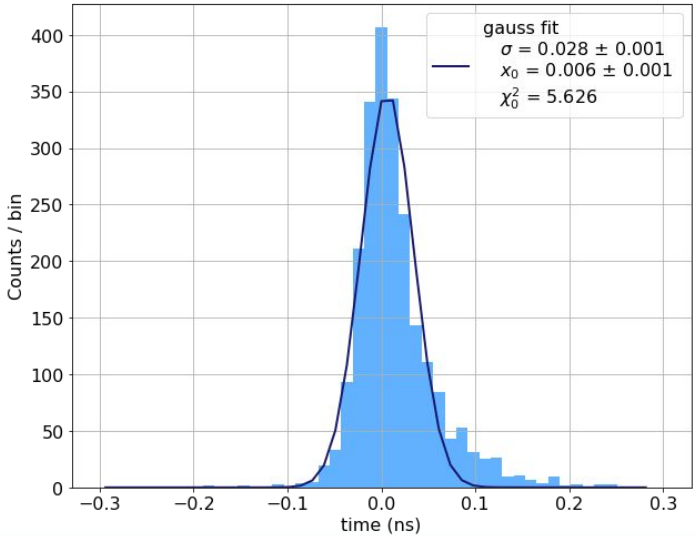
- If **MTD** time is available and the **combination** is enabled the final TICLCandidate time and the MTD time at the vertex are combined
  - `true` by default, but configurable because the MTD time may not always be available (e.g. at the HLT)

Studies on tracksters time

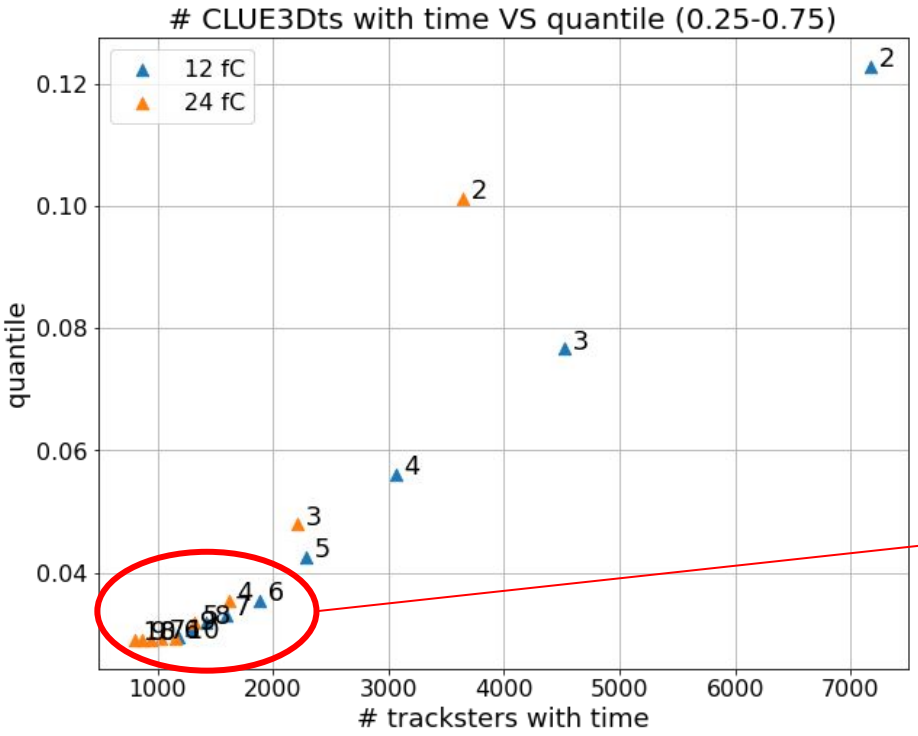


- The threshold to trigger the **Time of Arrival** (ToA) in a sensor now is 12 fC, but it may change in the future
- The requirement on at least three **hits with time** in a cluster to compute the cluster's time could be optimized

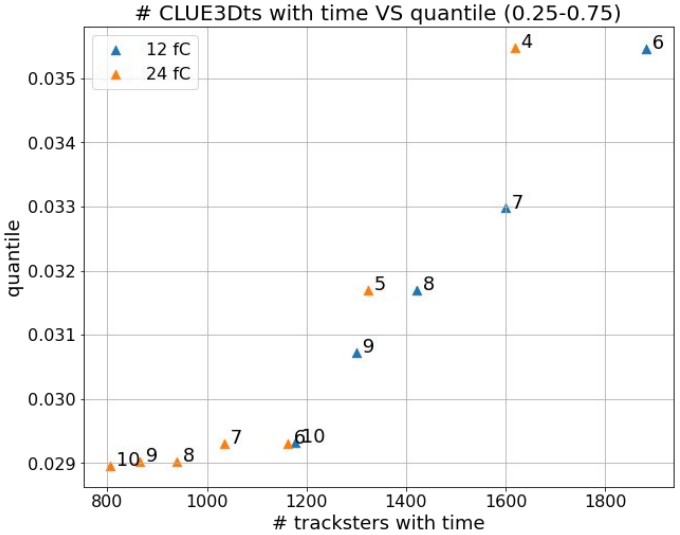
- Considering the CLUE3D tracksters, the difference between the trackster time and the simulated time has been computed for each combination of ToA (12 - 24 fC) and minimum number of hits (1 to 10)



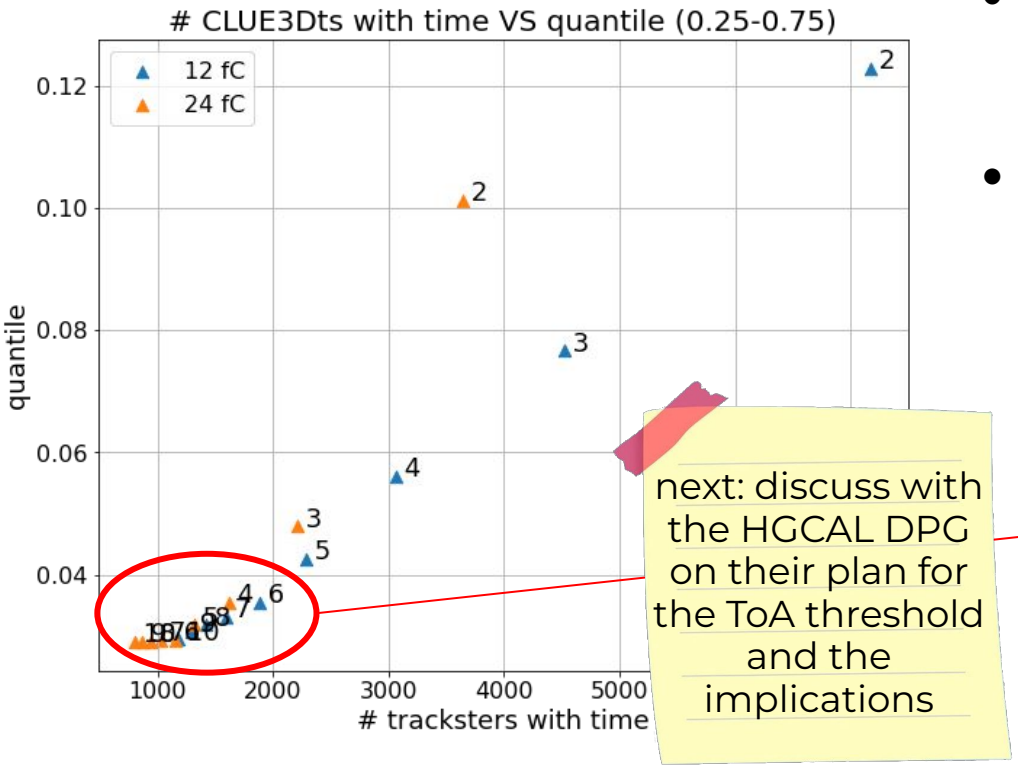
Studies on tracksters time



- number of tracksters with time against the difference between the 0.75 and 0.25 quantiles (to take into account the tails)
- the label is the minimum number of hits with time in LC

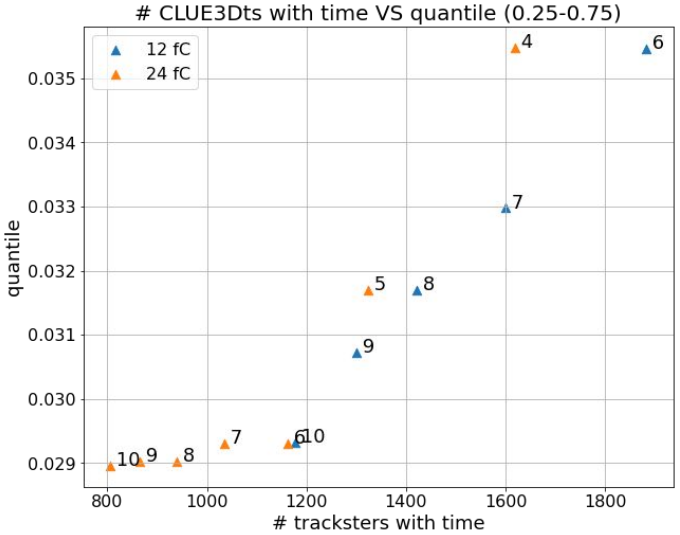


Studies on tracksters time



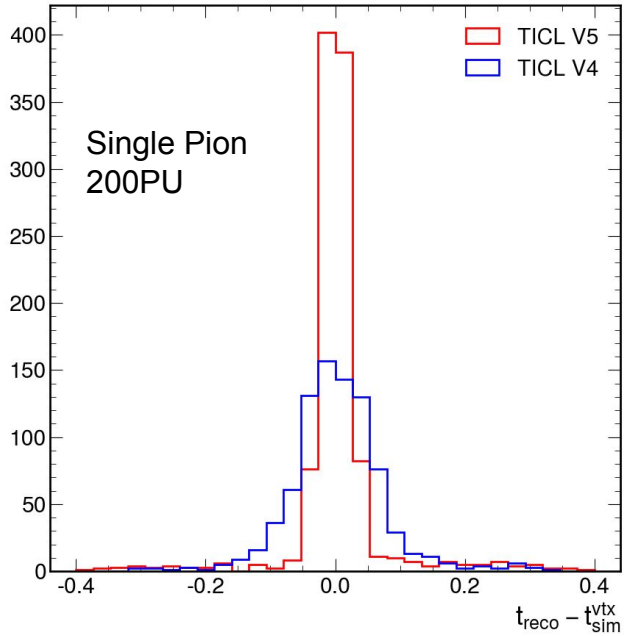
next: discuss with the HGCAL DPG on their plan for the ToA threshold and the implications

- number of tracksters with time against the difference between the 0.75 and 0.25 quantiles (to take into account the tails)
- the label is the minimum number of hits with time in LC

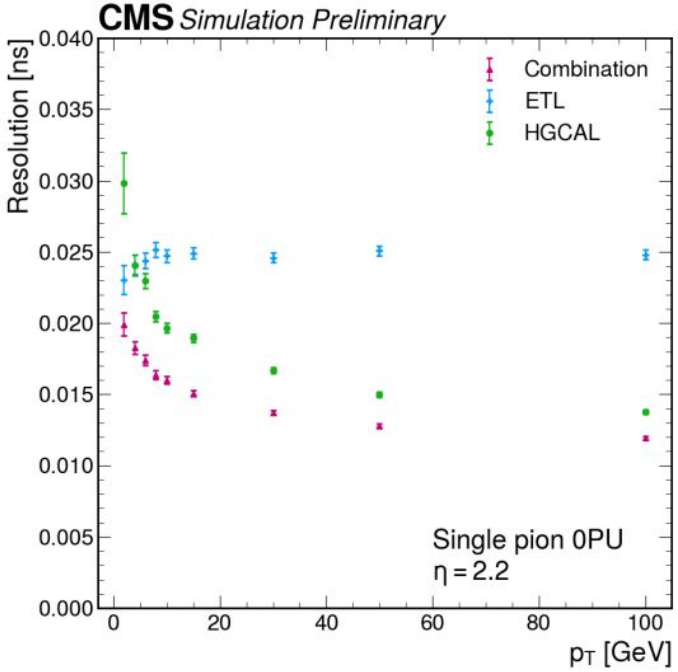


TICLCandidates time resolution

Tiziano Pauletto



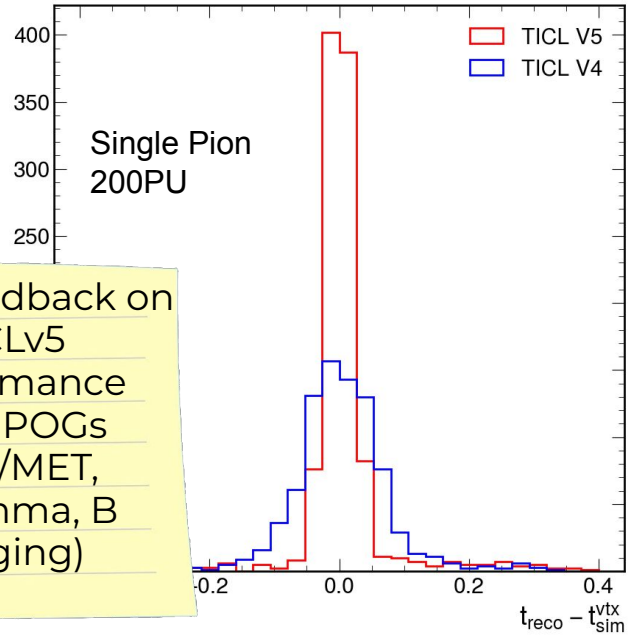
time residuals of charged candidates



time resolution as a function of the simulated particle p_T

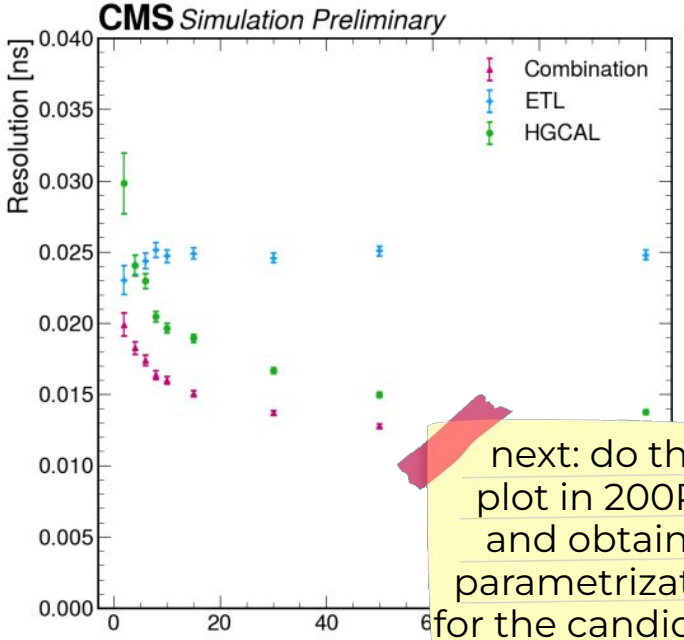
TICLCandidates time resolution

Tiziano Pauletto



soon: feedback on TICLv5 performance from POGs (Jets/MET, e/gamma, B tagging)

time residuals of charged candidates

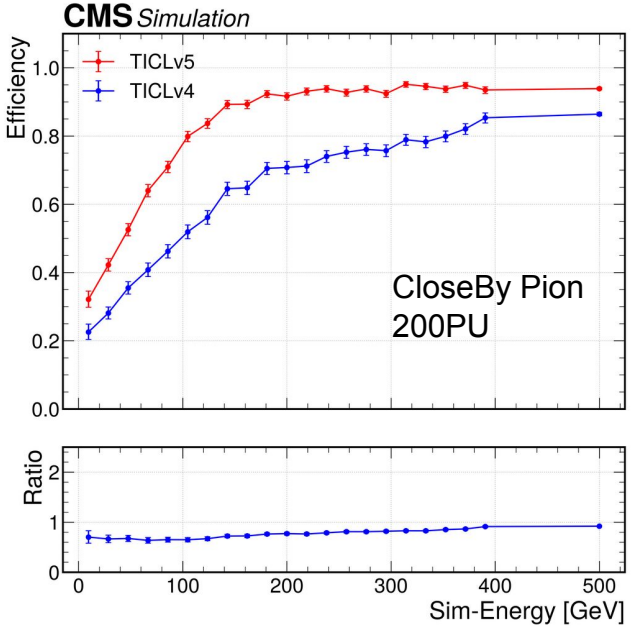


next: do this plot in 200PU and obtain a parametrization for the candidates time resolution

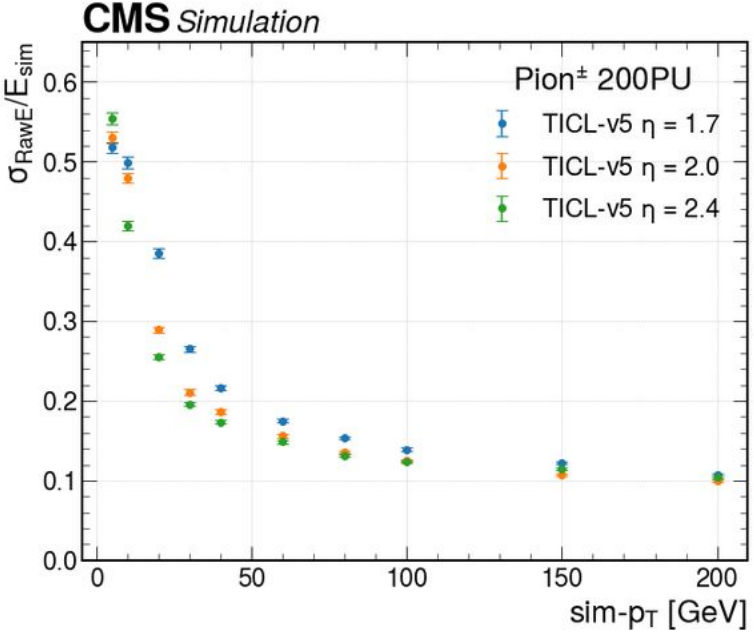
time resolution as a fun
simulated particle p_T

TICLv5 first results

Wahid Redjeb



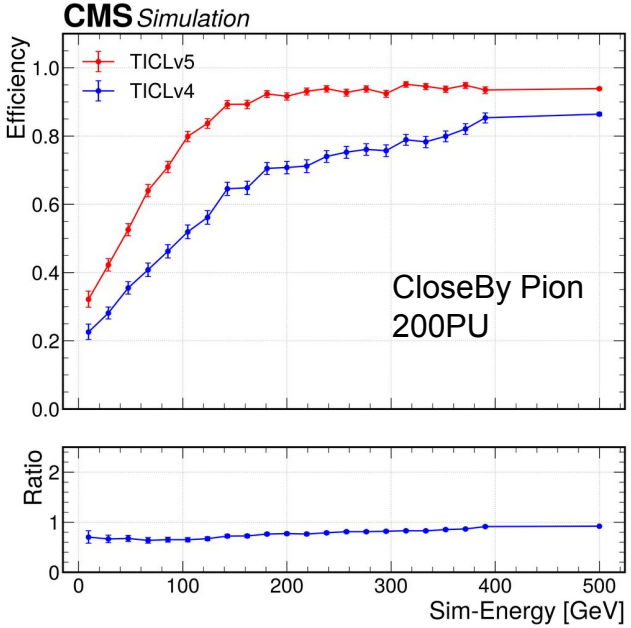
$$\text{Efficiency} = \frac{\text{simTracksters with a reco trackster that shares > 50\% of the energy}}{\text{simTracksters}}$$



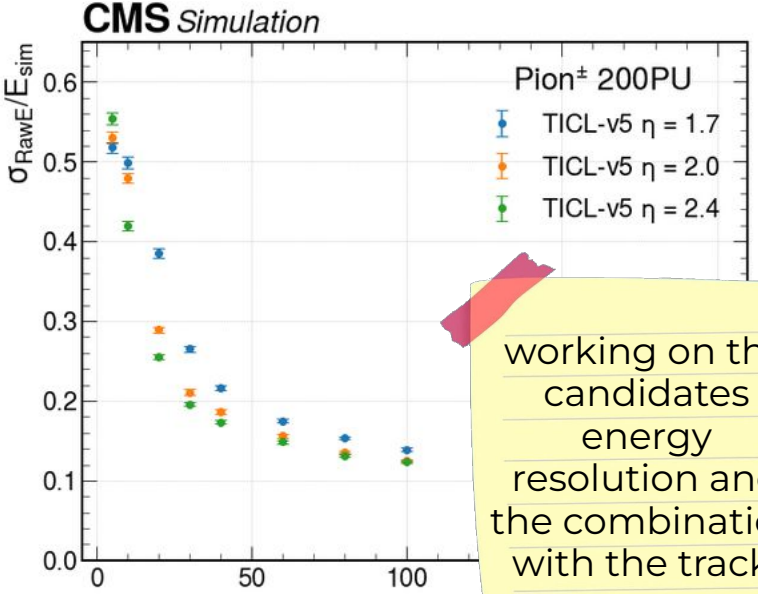
Energy resolution of the tracksters for different values of η: each point is the effective σ of the distribution of the best (= shares > 50% of the energy with a simTrackster) reco tracksters raw energy

TICLv5 first results

Wahid Redjeb



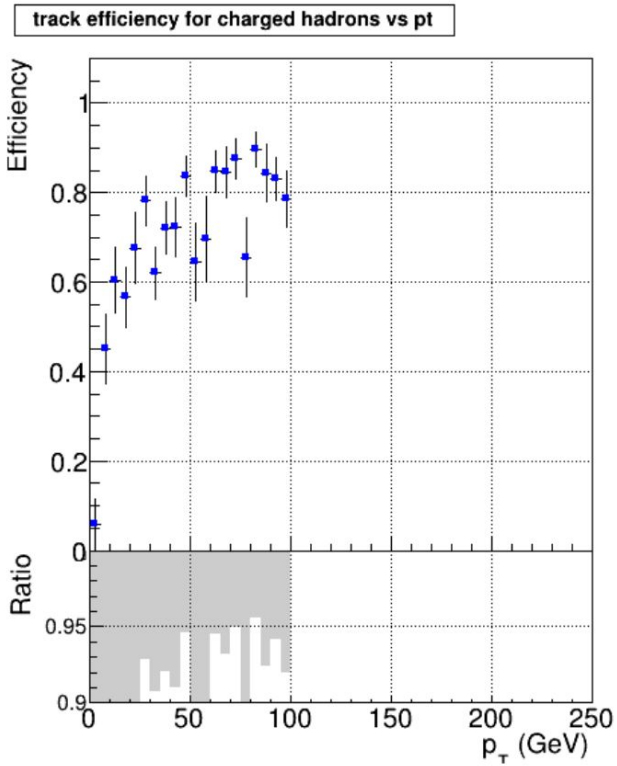
$$\text{Efficiency} = \frac{\text{simTracksters with a reco trackster that shares > 50\% of the energy}}{\text{simTracksters}}$$

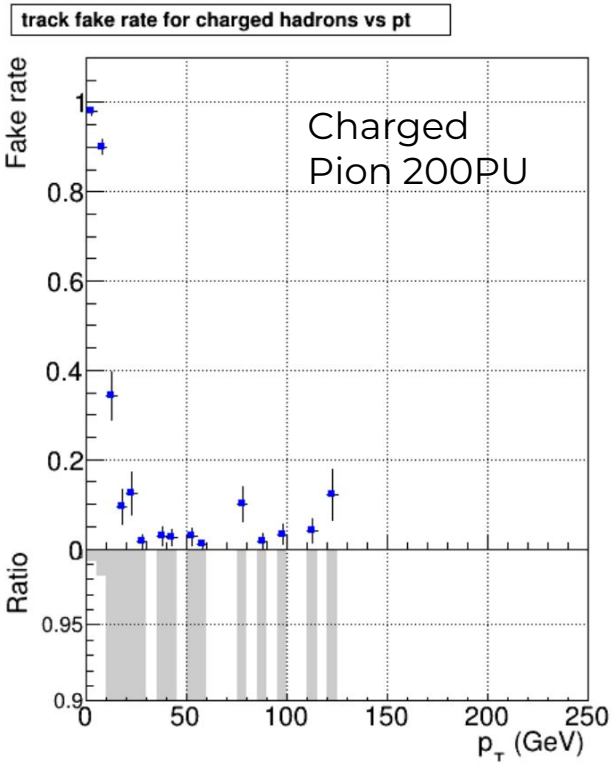


working on the candidates energy resolution and the combination with the track!

Energy resolution of the tracksters for different values of η: each point is the effective σ of the distribution of the best (= shares > 50% of the energy with a simTrackster) reco tracksters raw energy

TICLv5 first results





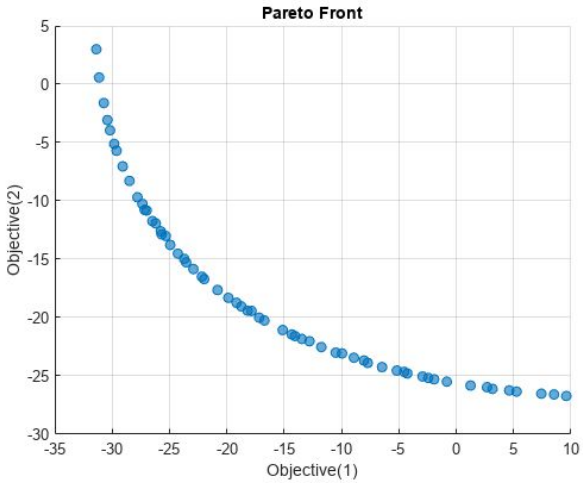
TICLCandidates **validation**

- integrated in CMSSW
- set of plots to **monitor** the **performance**
- web visualization
- entire set of plots here

Multi Objective Particle Swarm Optimization

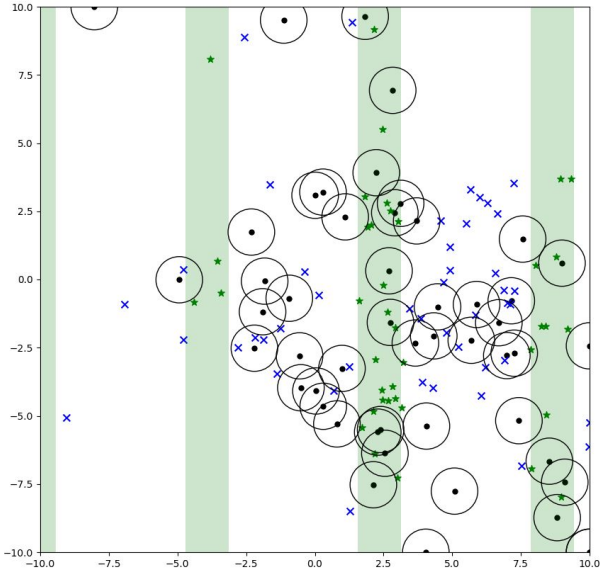
- computational method employed to **optimize** a **problem** defined by multiple objective functions (metrics)
- **population** of *particles* explores the phase space of the solution to find the minimum
  - each particle updates its state at every iteration and knows the global best

- the result is a *pareto front* of solutions
- has been used in CMS to **optimize** some of the pixel reconstruction parameters and in **TICL** to tune the pattern recognition and linking **parameters**
  - successful, but requires a lot of time
  - idea: skip unnecessary evaluations



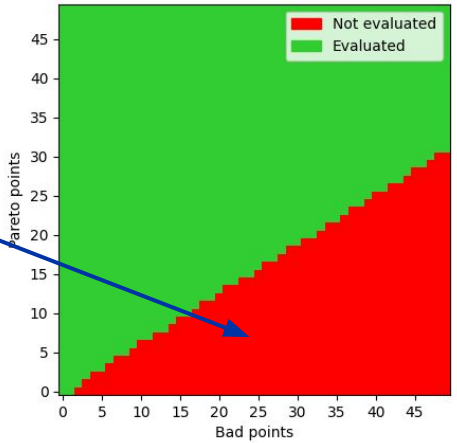
MOPSO with RL

Marco Picione



- strategy: use reinforcement learning to skip unnecessary evaluations
  - better than skipping them randomly
  - converges faster than standard MOPSO
- a point is bad if there are no pareto points in its surroundings
  - RL decides how many points to evaluate

evaluations that will be skipped

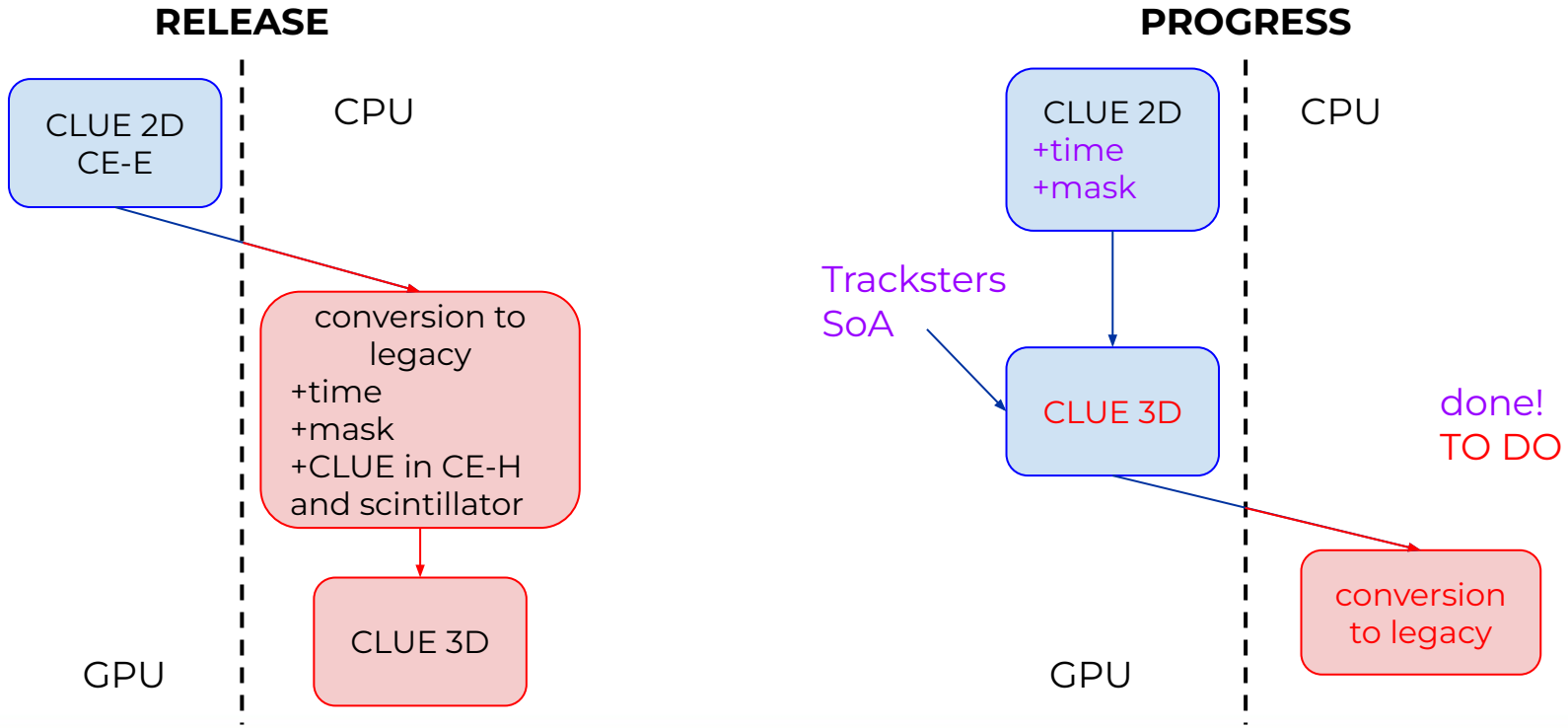


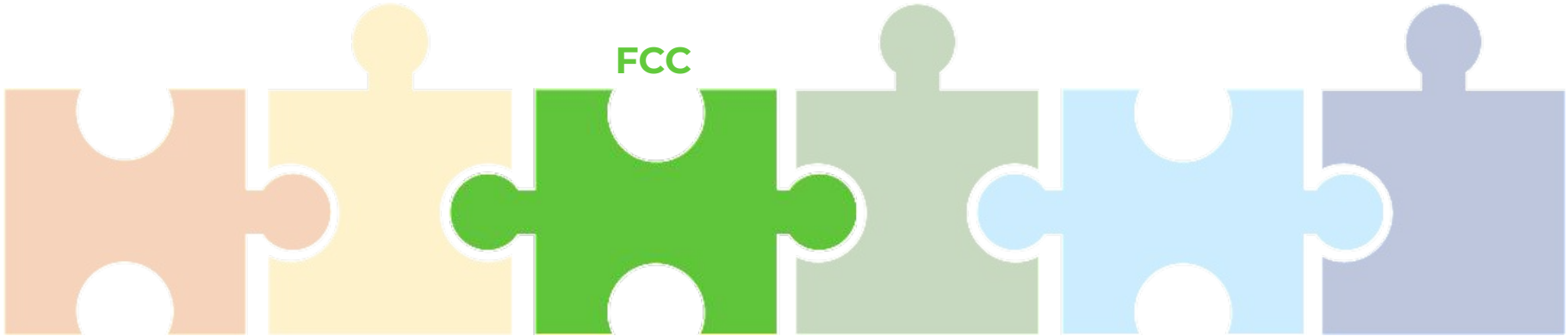
- Marco will test it on TICL for his thesis

CLUE 3D @ alpaka

Muhammad Arham

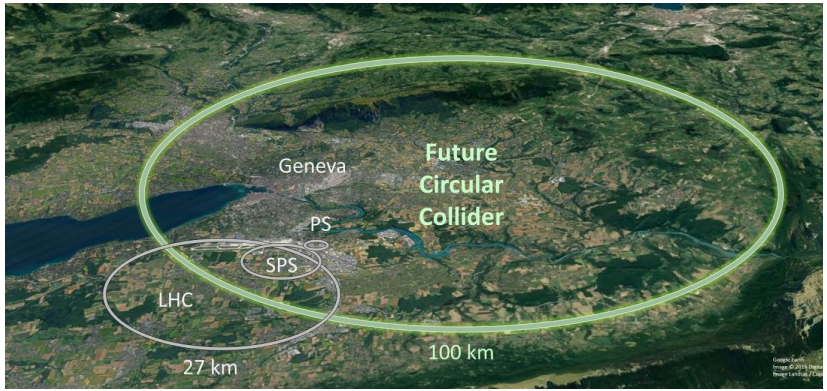
This summer I supervised a summer student who started the effort of writing the pattern recognition algorithm CLUE 3D with alpaka



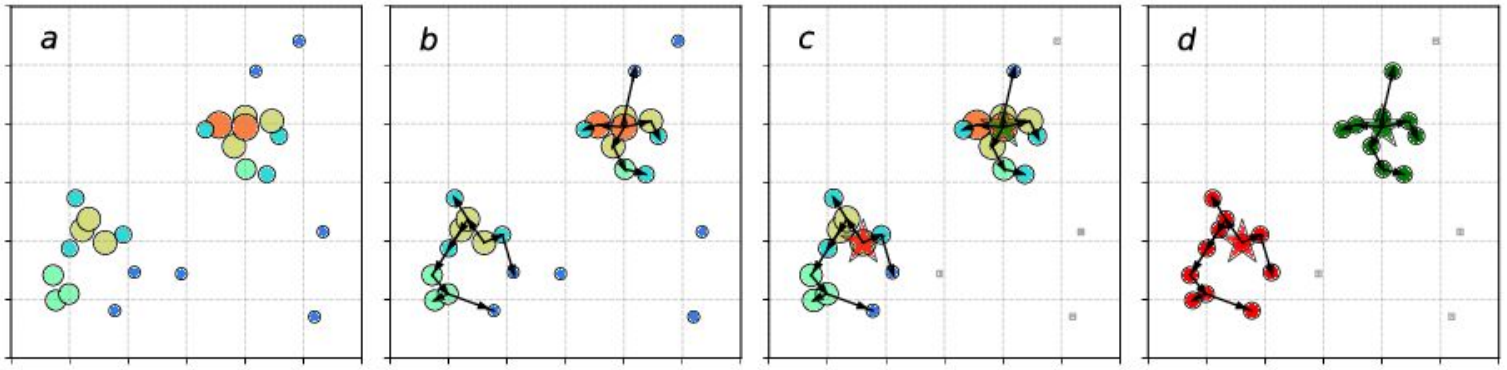

FCC

plans for future colliders

- Part of my PhD is focused on **future colliders**
- 5 years plan:
  - explore the time information available in the FCC detectors
  - integrate **time** information in **clustering** and **pattern recognition** algorithms
  - use the timing of Particle Flow Candidates in Particle Flow algorithms
  - adapt the algorithms for execution in a **heterogeneous framework**
  - combination with the tracking reconstruction



n-dimensional CLUE



evaluation of the density around each point

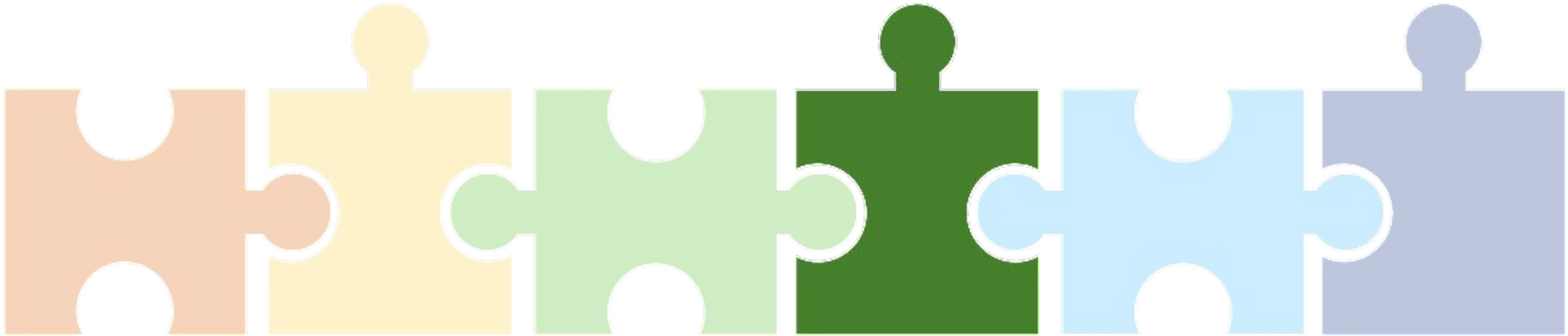
connection of each point with the nearest with higher energy

points promotion to seeds or outliers based on density and distance

clusters definition and outlier removal

A more generic version of CLUE has been developed (already on GPU), able to do the clustering in N dimensions, where N can be:

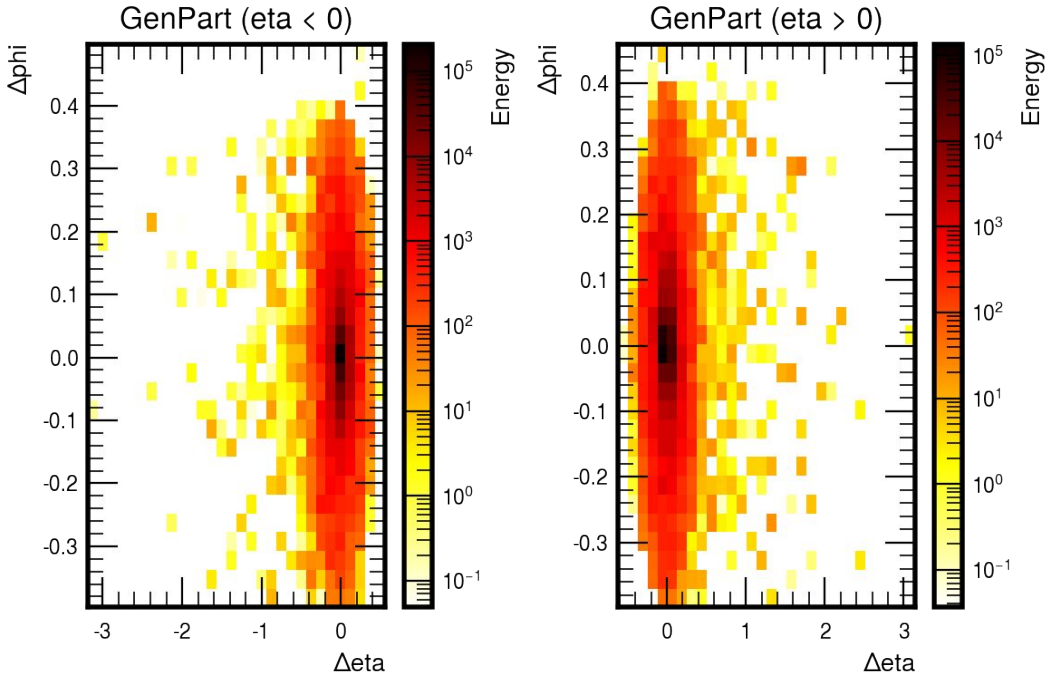
- 2: the current one
- 3: implementation ongoing, clustering directly the hits
- 4: with time (?)



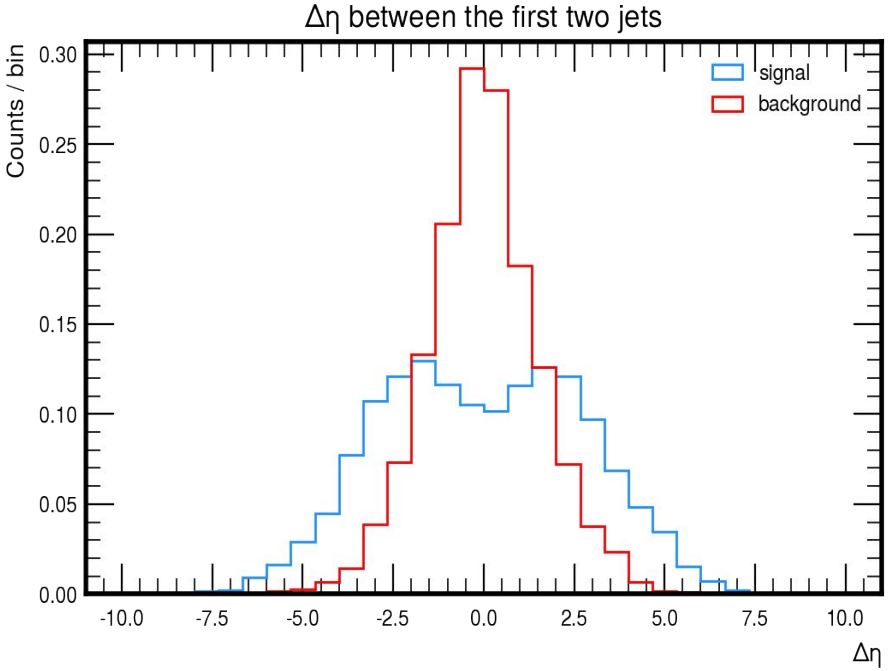
VBF / VBS tagger

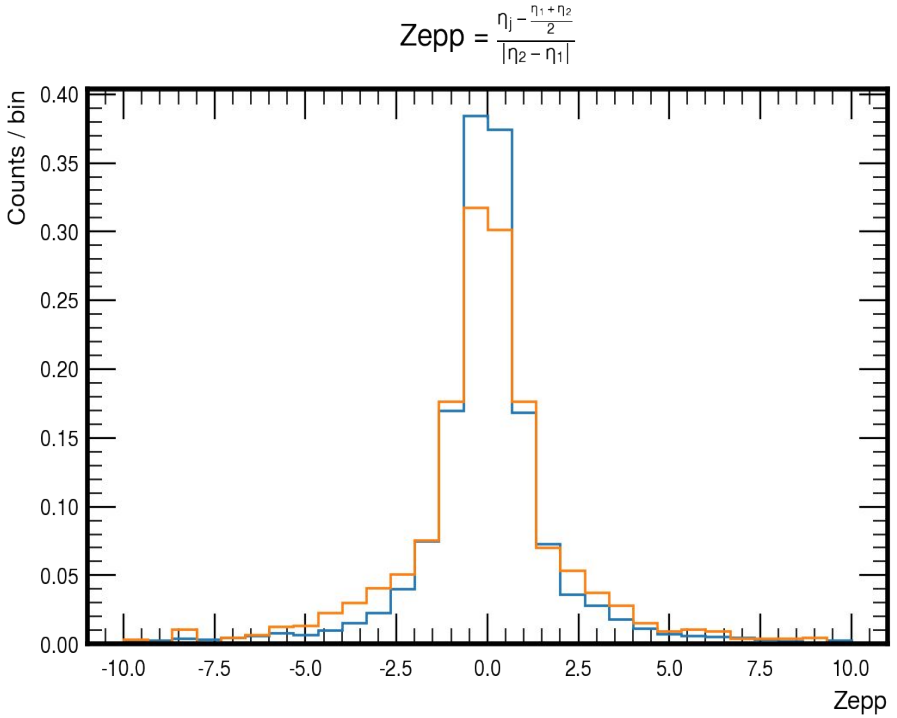
physics motivation behind the tagger

- based on the idea that the QCD activity between the tag jets and the beam is larger than in the space between jets
  - causes asymmetry in jets
  - from preliminary plots the asymmetry exists and is slightly more evident in the signal
- signal: VBF
  background: Drell Yan
- studies ongoing to find interesting variables for the tagger
  - jet pull
  - jet colour ring
  - Lund plane

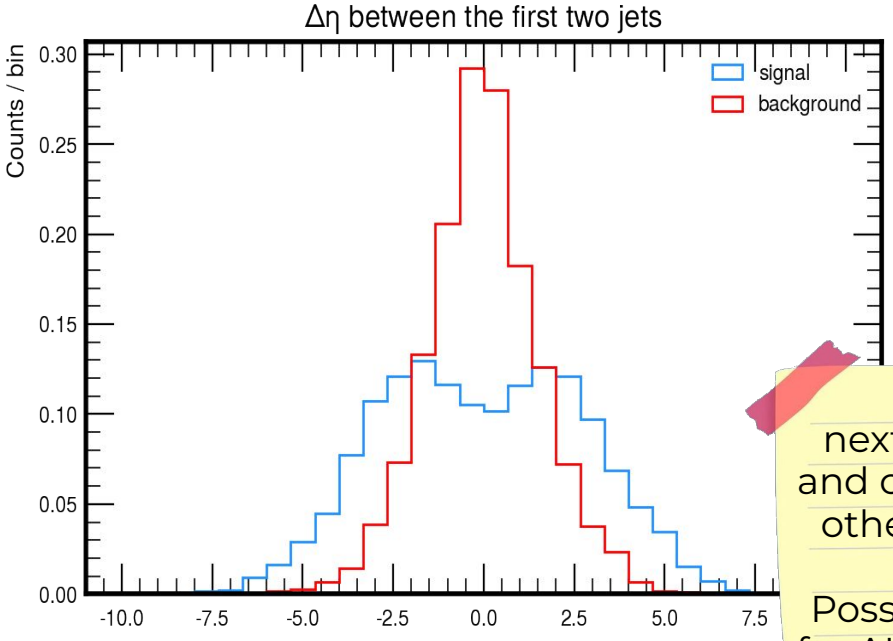


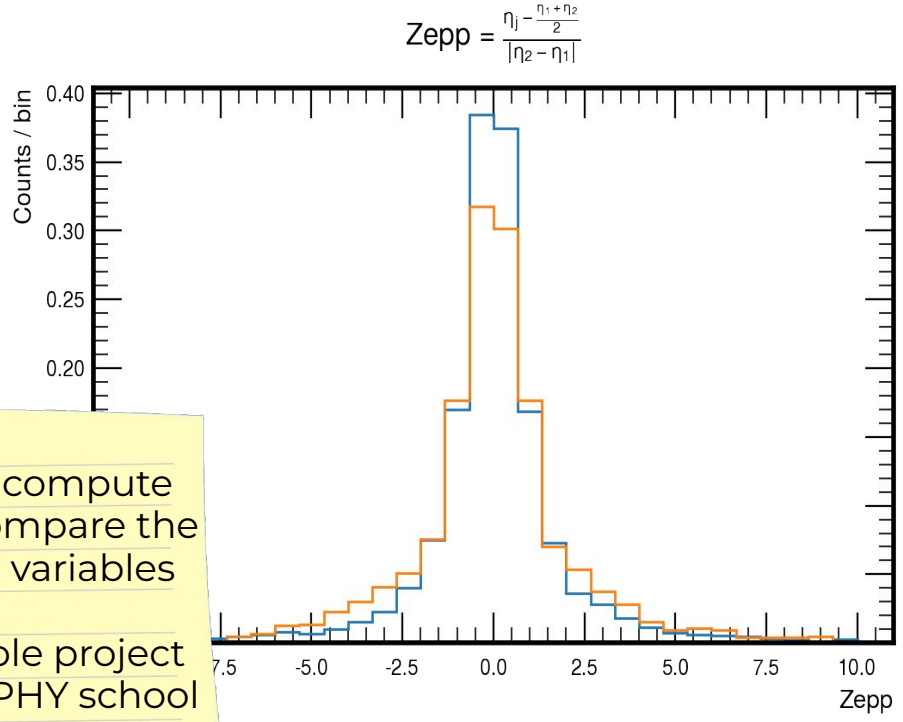
(preliminary) comparisons





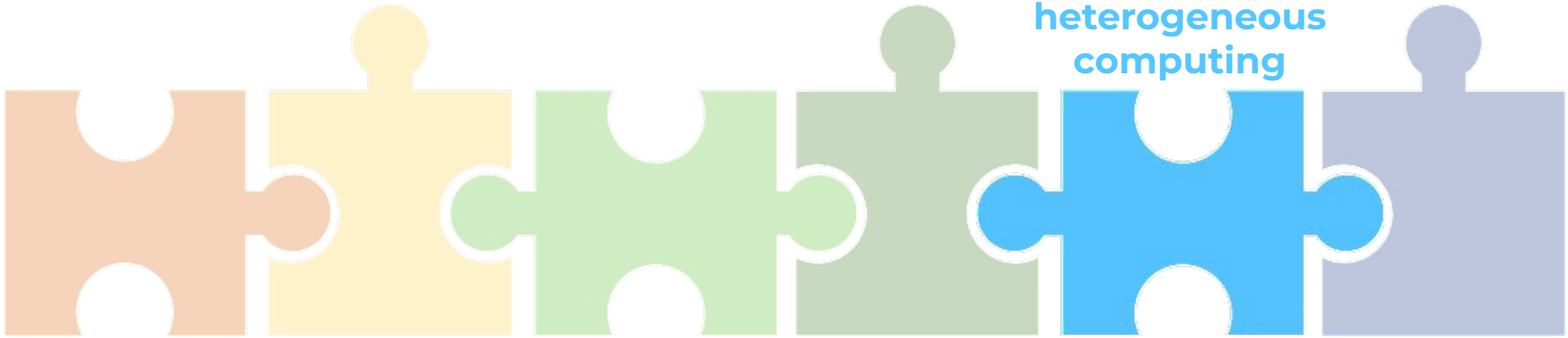
(preliminary) comparisons





next: compute and compare the other variables

Possible project for AI-PHY school


heterogeneous
computing

the alpaka library

- is a header-only C++17 abstraction library for accelerator development
- aims to provide performance portability across accelerators through the abstraction of the underlying levels of parallelism
- has several backends:
  - serial and parallel for CPUs
  - CUDA for NVIDIA GPUs
  - HIP for AMD GPUs
  - oneAPI for CPUs, Intel GPUs and FPGAs

latest updates:

v1.0.0: rewritten part of the oneAPI backend (USM, SYCL 2020)
v1.1.0: warp-level primitives
v1.2.0 (soon released): device global variables

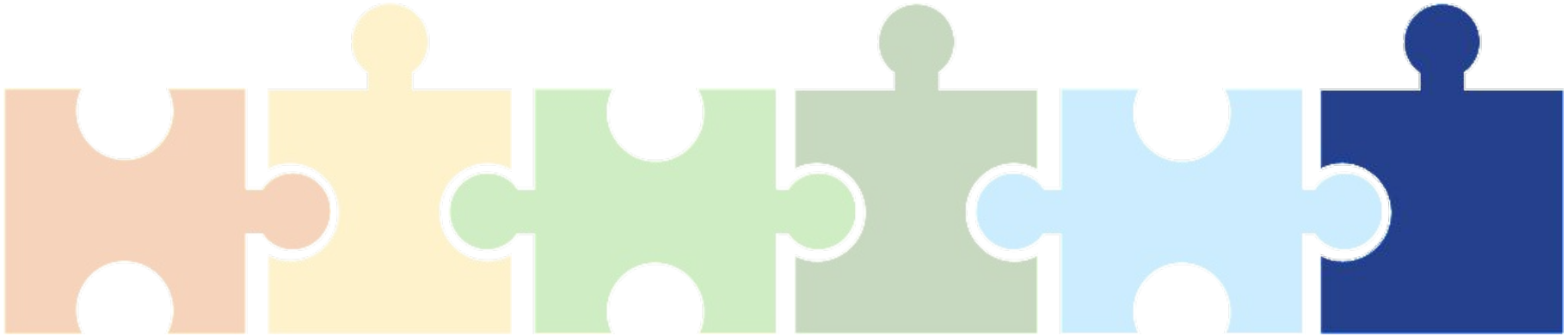
future plans:

- add SYCL CUDA / HIP backend
- rewrite the shared memory implementation, currently very slow
- test it on LUMI supercomputer
- (maybe) try the FPGA backend

the oneAPI backend now is fully supported!

heterogeneous computing in CMSSW

- the **CMSSW** framework already uses **alpaka** to achieve **performance portability** across CPUs and NVIDIA / AMD GPUs
  - the High Level Trigger farm is equipped with NVIDIA T4 and L4 GPUs
  - the fallback on CPU is automatic and transparent
  - the HLT reconstruction has been successfully tested on AMD GPUs
- with the complete **support** for **Intel GPUs** in alpaka, the plan is to extend their support also in CMSSW
  - ✓ write some simple tests in CMSSW with oneAPI and adapt the existing alpaka tests to use the oneAPI backend
  - (ask to ✓) implement in SCRAM the rules to compile for the oneAPI backend
  - add the support in CMSSW framework
  - test it!



other activities

Other activities

central DAQ shifts

for CMS data taking

PhD course Scientific computing with Python

Tutor of “Laboratorio di Calcolo e Statistica” of the second year of bachelor

Conferences

Upgrade week CALOR