Reconstruction for e+e- Higgs Top & Electroweak Factory Experiments

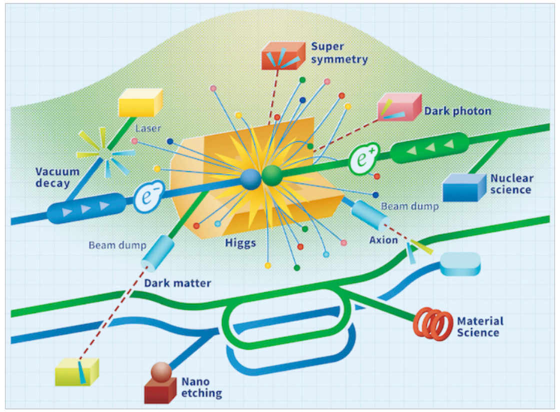
3rd ECFA workshop on Higgs, Top & Electroweak Factories

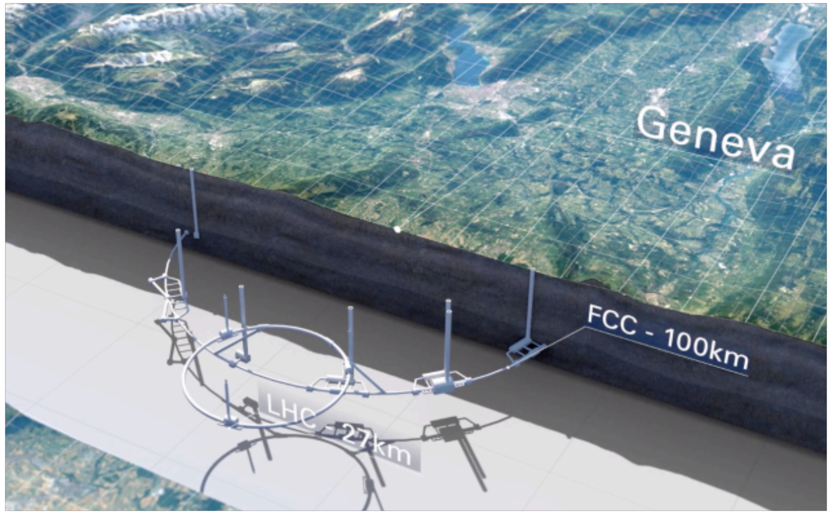
9-11 Oct, Paris

Frank Gaede, DESY

Outline

- Introduction
  - Key4hep, DD4hep and MarlinWrapper
- Existing reconstruction code
  - MarlinTrk, PandoraPFA,…
- Highlights of more recent developments:
  - Tracking, PFA, HF-Tagging, PID,…
  - with and without AI/ML
- Summary





ECFA Higgs Factories: 1st Topical Meeting on Reconstruction

4 May 2022, 09:00 → 5 May 2022, 18:00 Europe/Zurich

DESY

ECFA Higgs Factories: 2nd Topical Meeting on Reconstruction

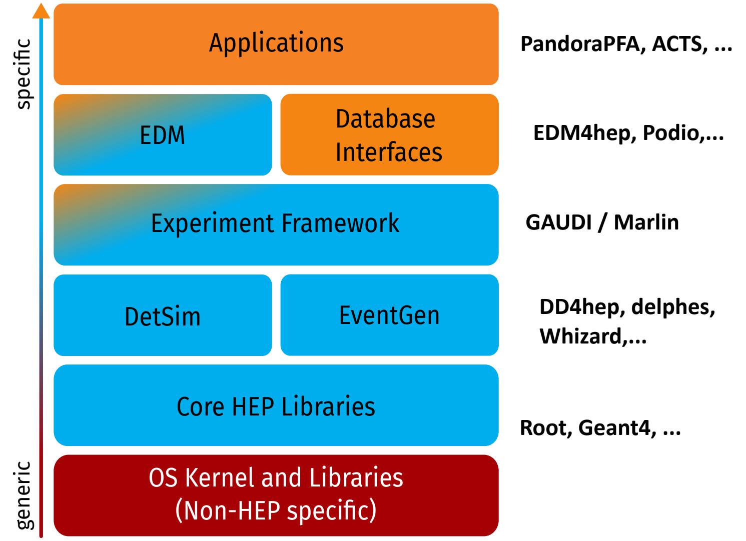
11 Jul 2023, 09:00 → 12 Jul 2023, 18:00 Europe/Zurich

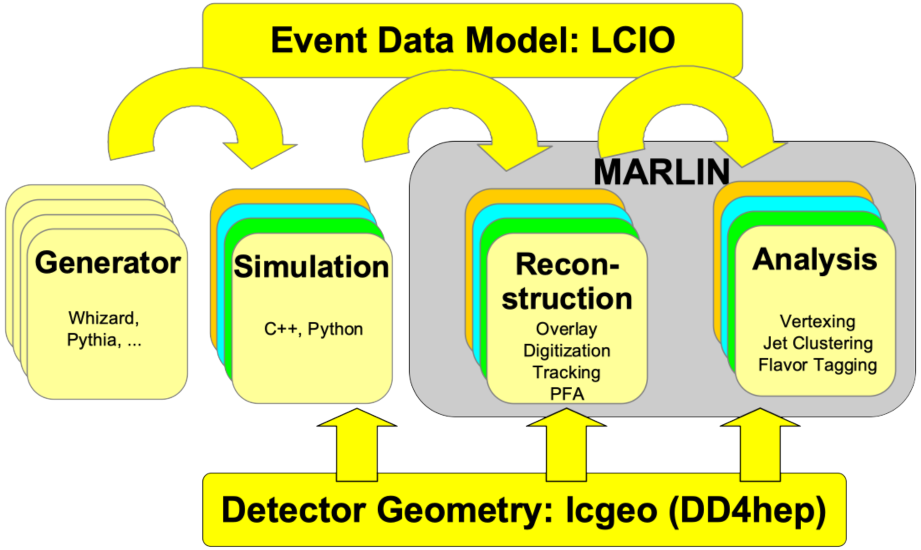
40/S2-C01 - Salle Curie (CERN)

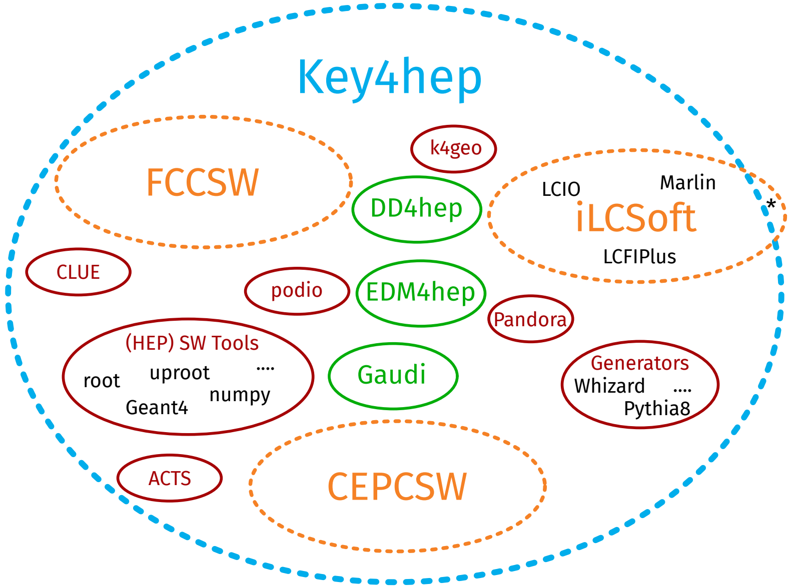
Key4hep

the turnkey software stack for ALL future colliders

- HEP community decided 5 years ago to develop a **common turnkey software stack** – for future collider studies
- create a software ecosystem integrating in an **optimal way the best software components** to provide a **ready-to-use full-fledged solution** for data processing of (future collider) **HEP** experiments
- involved communities/contributors: CEPC, CLIC, EIC, FCCee, FCChh, ILC, LUXE, Muon Collider …





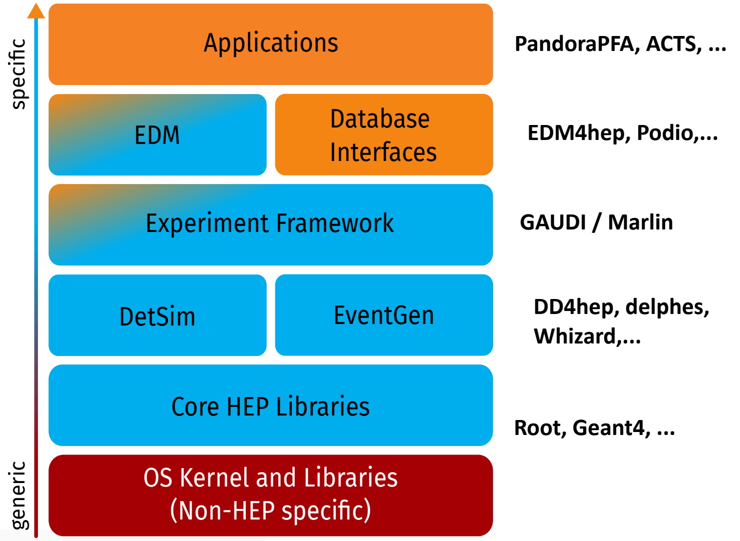


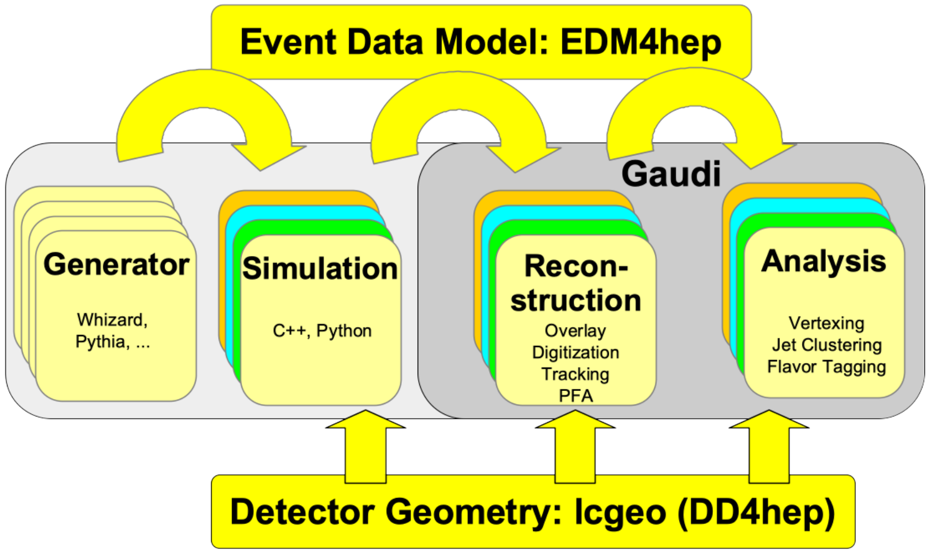
see talk by J.Smiesko, Wed.

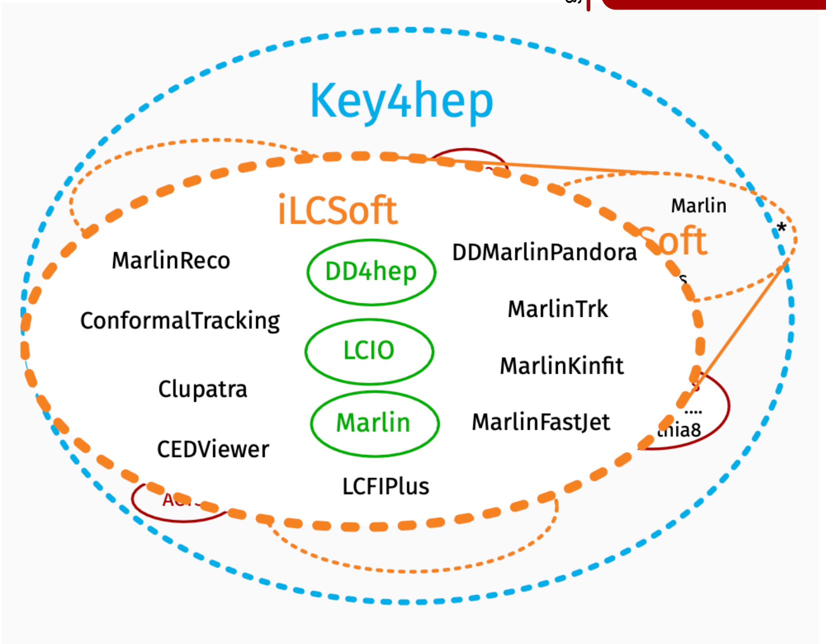
Key4hep

the turnkey software stack for ALL future colliders

- HEP community decided 5 years ago to develop a **common turnkey software stack** – for future collider studies
- create a software ecosystem integrating in an **optimal way the best software components** to provide a **ready-to-use full-fledged solution** for data processing of (future collider) **HEP** experiments
- involved communities/contributors: CEPC, CLIC, EIC, FCCee, FCChh, ILC, LUXE, Muon Collider …





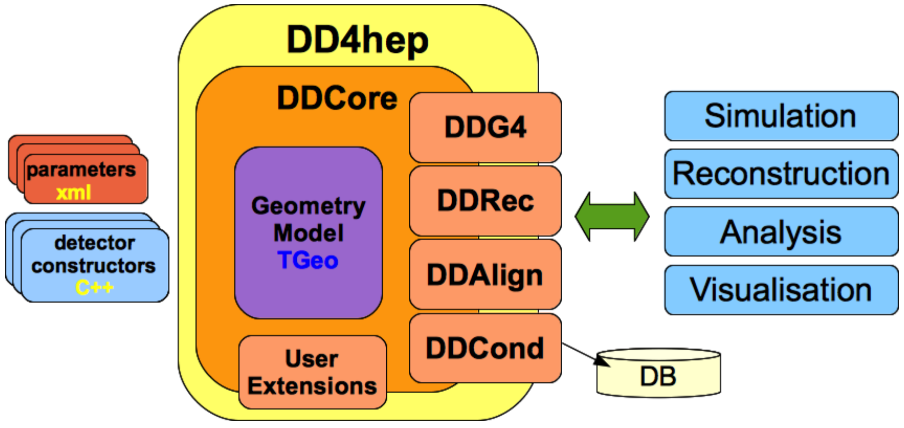


see talk by J.Smiesko, Wed.

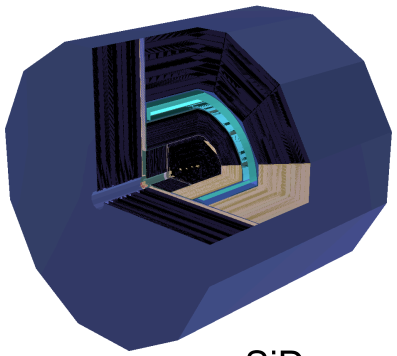
DD4hep geometry toolkit

defining the detector geometry and different views on it

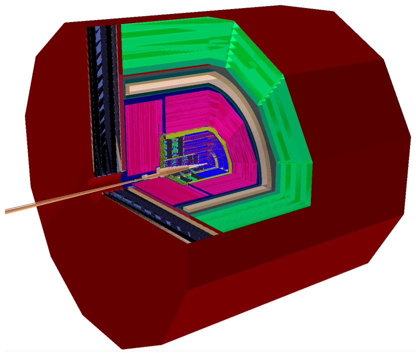
- supporting the full life cycle of the experiment
- **single source** of information for full **simulation, reconstruction, conditions, alignment, visualisation and analysis**
    - used by CEPC, CLIC, CMS, EIC, FCC, ILC, LHCb, …



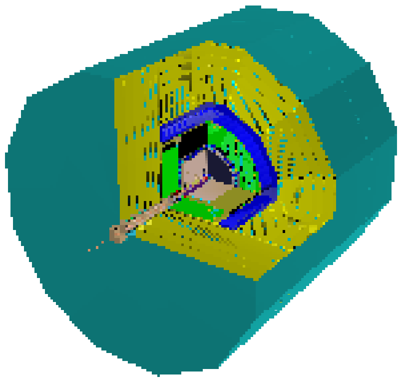
DD4hep: de facto industry standard



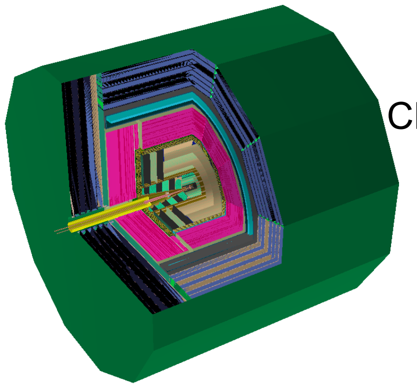
SiD



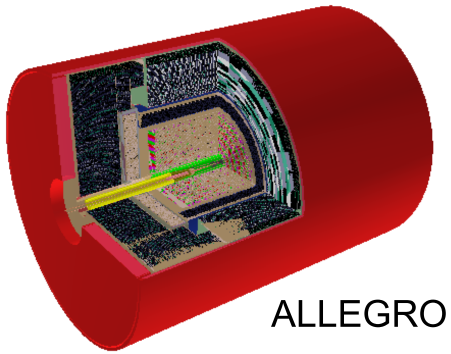
CLICdet



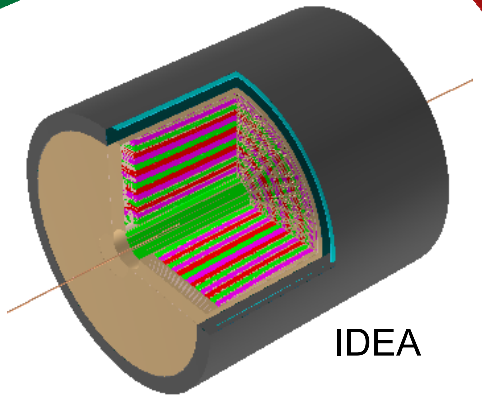
ILD



CLD



ALLEGRO



IDEA

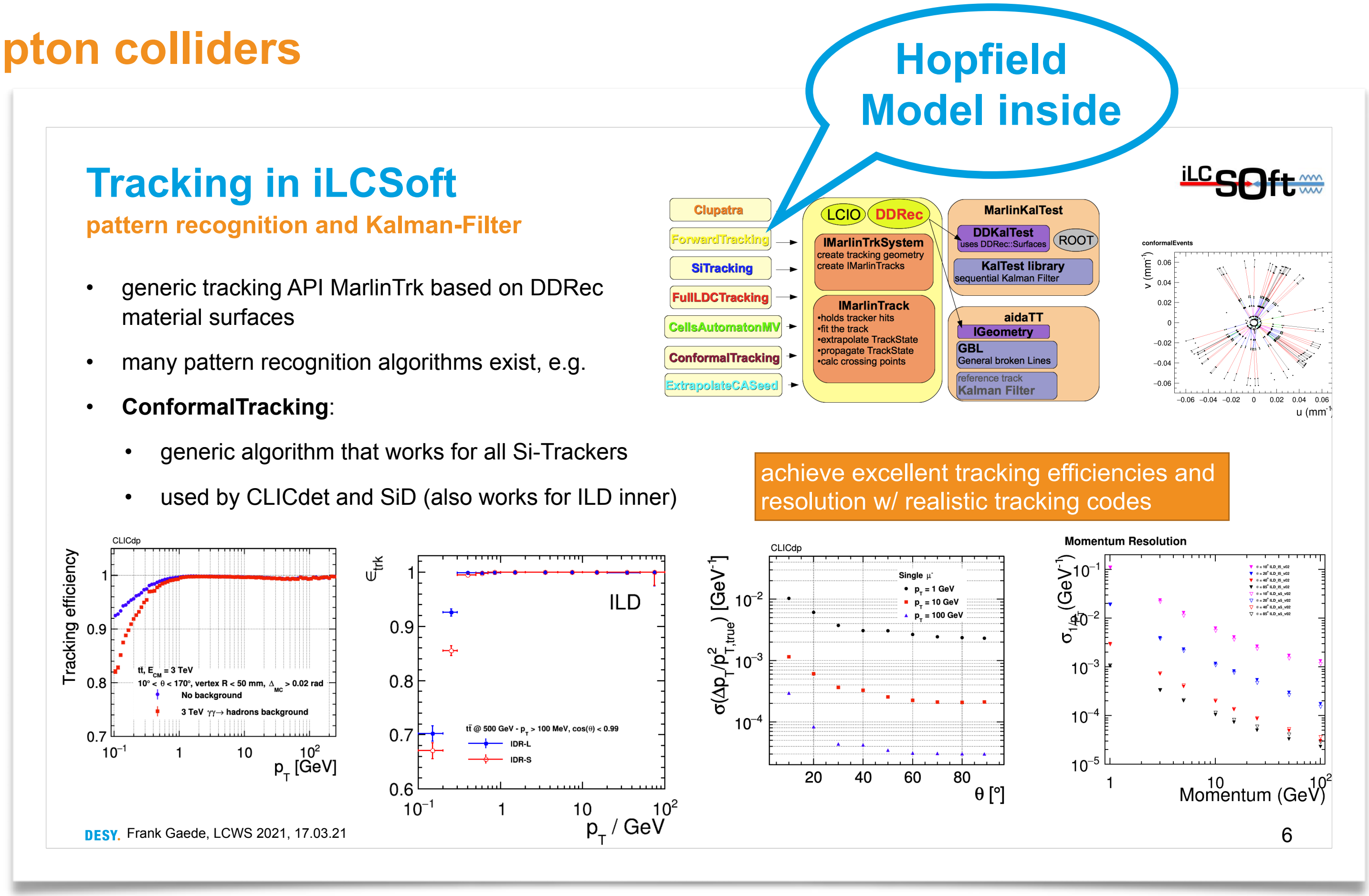
all future Higgs factory detector simulation models in **one package**
https://github.com/key4hep/k4geo
=> allows to **re-use reconstruction code** (for similar sub detectors) across detector concepts

see previous talk by A.Tolosa Delgado

large reconstruction code base in iLCSoft

Developed over >15 years for (linear) lepton colliders

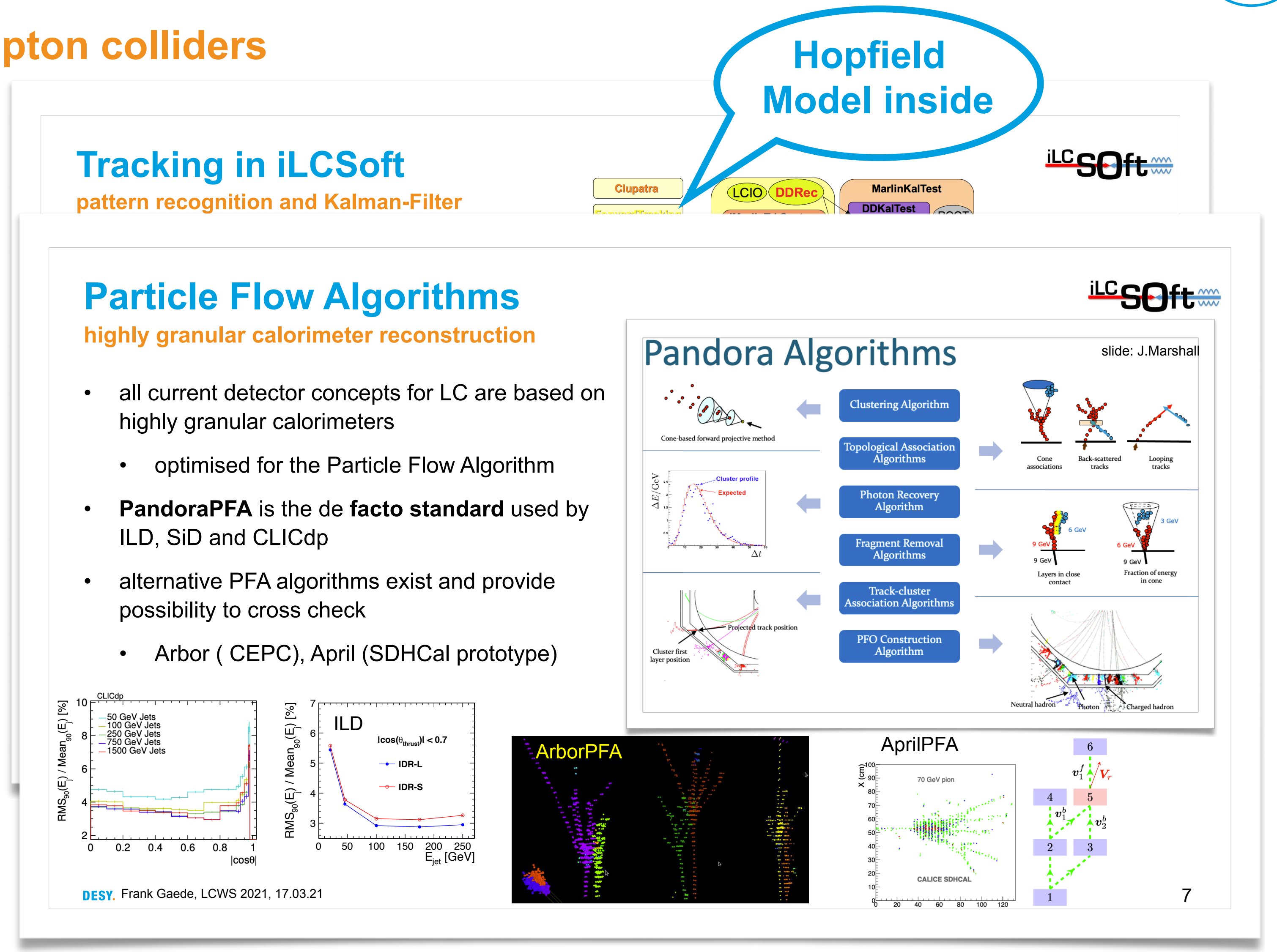
- realistic detector models for incl. tracking/reconstruction geometry
- track reconstruction
    - generic API for fitting algorithms
    - large number of pattern recognition algorithms



large reconstruction code base in iLCSoft

Developed over >15 years for (linear) lepton colliders

- realistic detector models for incl. tracking/reconstruction geometry
- track reconstruction
  - generic API for fitting algorithms
  - large number of pattern recognition algorithms
- particle flow algorithms
  - PandoraPFA ans Arbor, AprilPFA

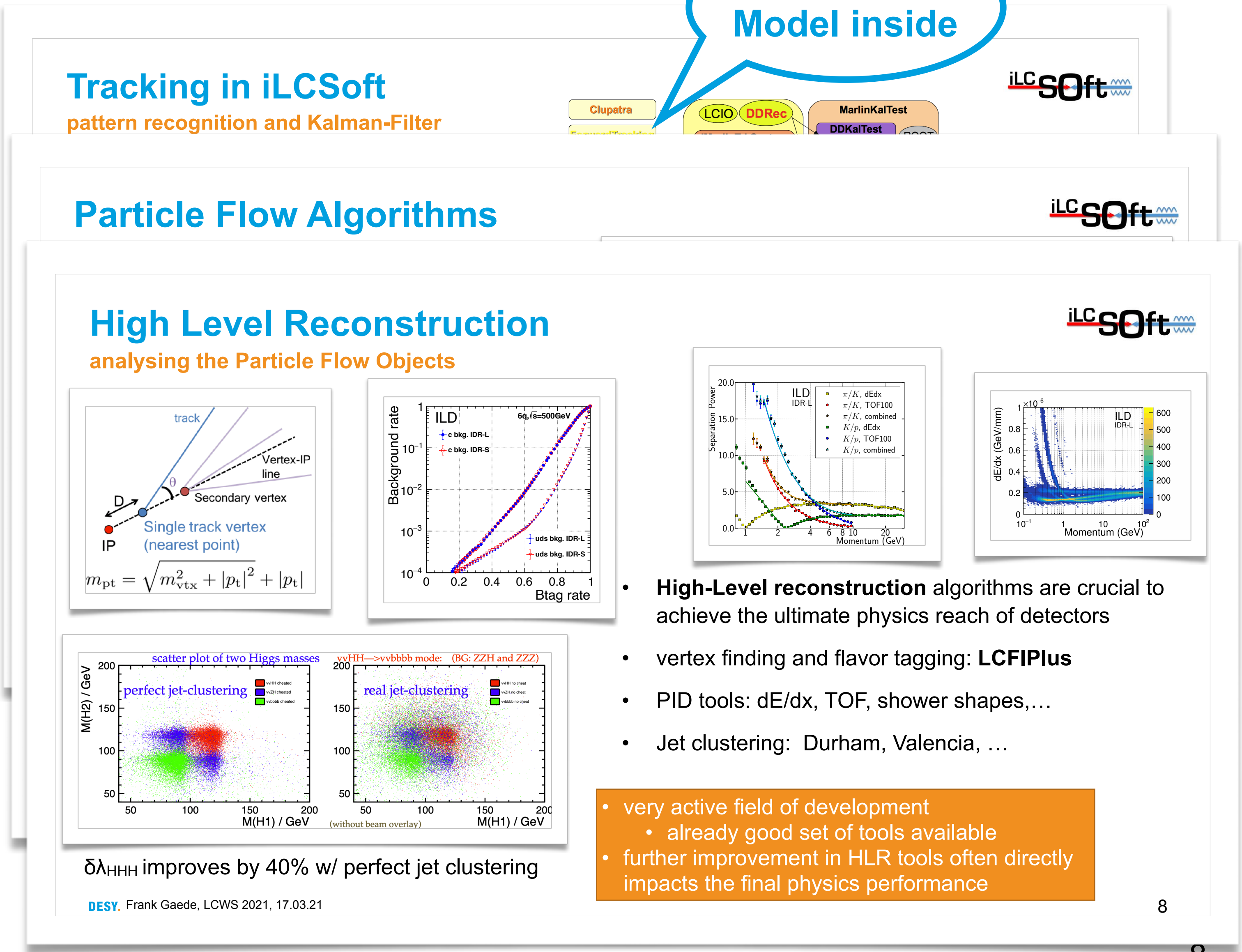


large reconstruction code base in iLCSoft

Developed over >15 years for (linear) lepton colliders

- realistic detector models for incl. tracking/reconstruction geometry
- track reconstruction
  - generic API for fitting algorithms
  - large number of pattern recognition algorithms
- particle flow algorithms
  - PandoraPFA ans Arbor, AprilPFA
- high level reconstruction
  - jet finding, flavor tagging, PID, TOF,…

Hopfield Model inside

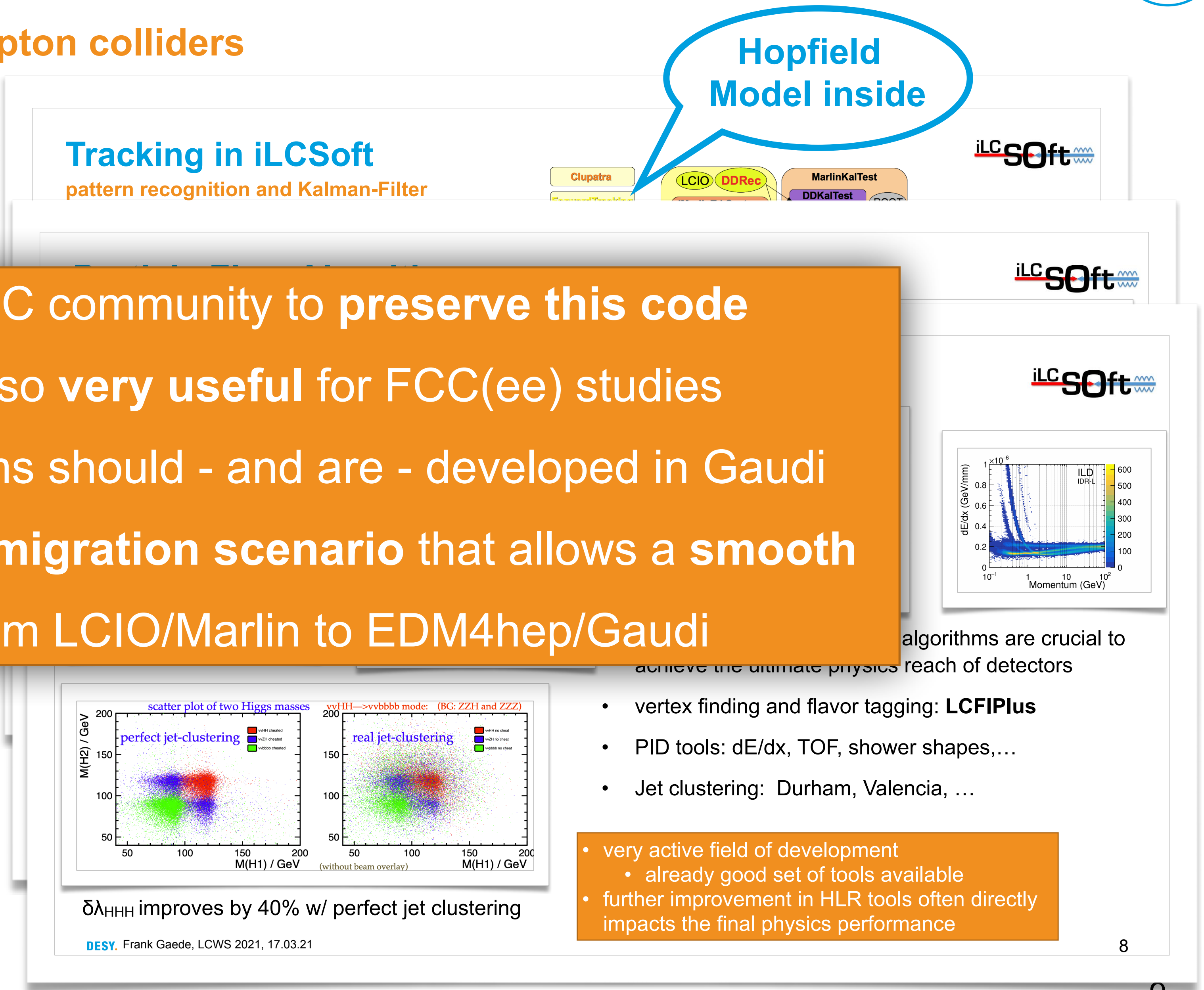


large reconstruction code base in iLCSoft

Developed over >15 years for (linear) lepton colliders

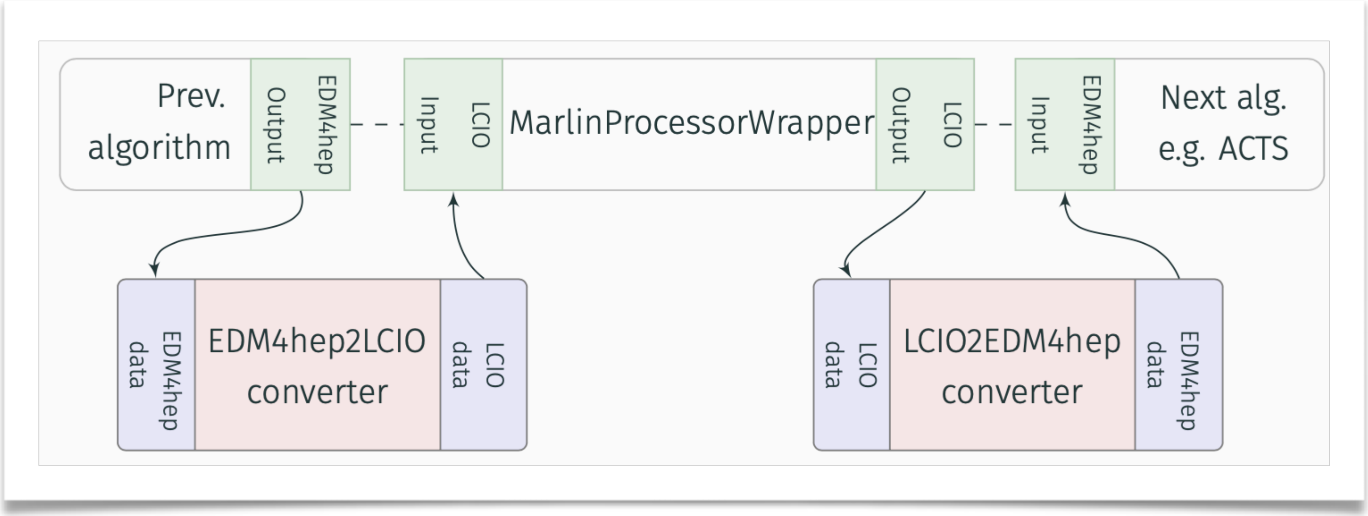
- realistic detector models for incl. tracking/reconstruction geometry
- track reconstruction
  - generic API for f
  - large number of recognition algo
- particle flow algorith
  - PandoraPFA ar
- high level reconstruction
  - jet finding, flavor tagging, PID, TOF,…

- **vital** for the FC community to **preserve this code**
- a lot of this also **very useful** for FCC(ee) studies
- new algorithms should - and are - developed in Gaudi
- developed a **migration scenario** that allows a **smooth transition** from LCIO/Marlin to EDM4hep/Gaudi



K4MarlinWrappper

the vision: mix and match Marlin and Gaudi algorithms



- in a transition phase algorithms **developed in the new EDM4hep/Gaudi** world can gradually replace older algorithms
  - e.g. eventually one might want to replace track fitting with **ACTS** also for LC detectors
- some technicalities are address *under-the-hood* via ***k4EDM4hep2LcioConv***

C.Hensel, ILD Meeting 2024

all existing (high level) reconstruction algorithms as Marlin processors fully available in Key4hep !

ILD Standard Reconstruction

now smoothly runs in in Key4hep

- translation of ***MarlinReco.xml*** to GAUDI python steering file ***ILDReconstruction.py***
- plan to use this as new standard for ILD soon

ILD standard reconstruction in Key4hep

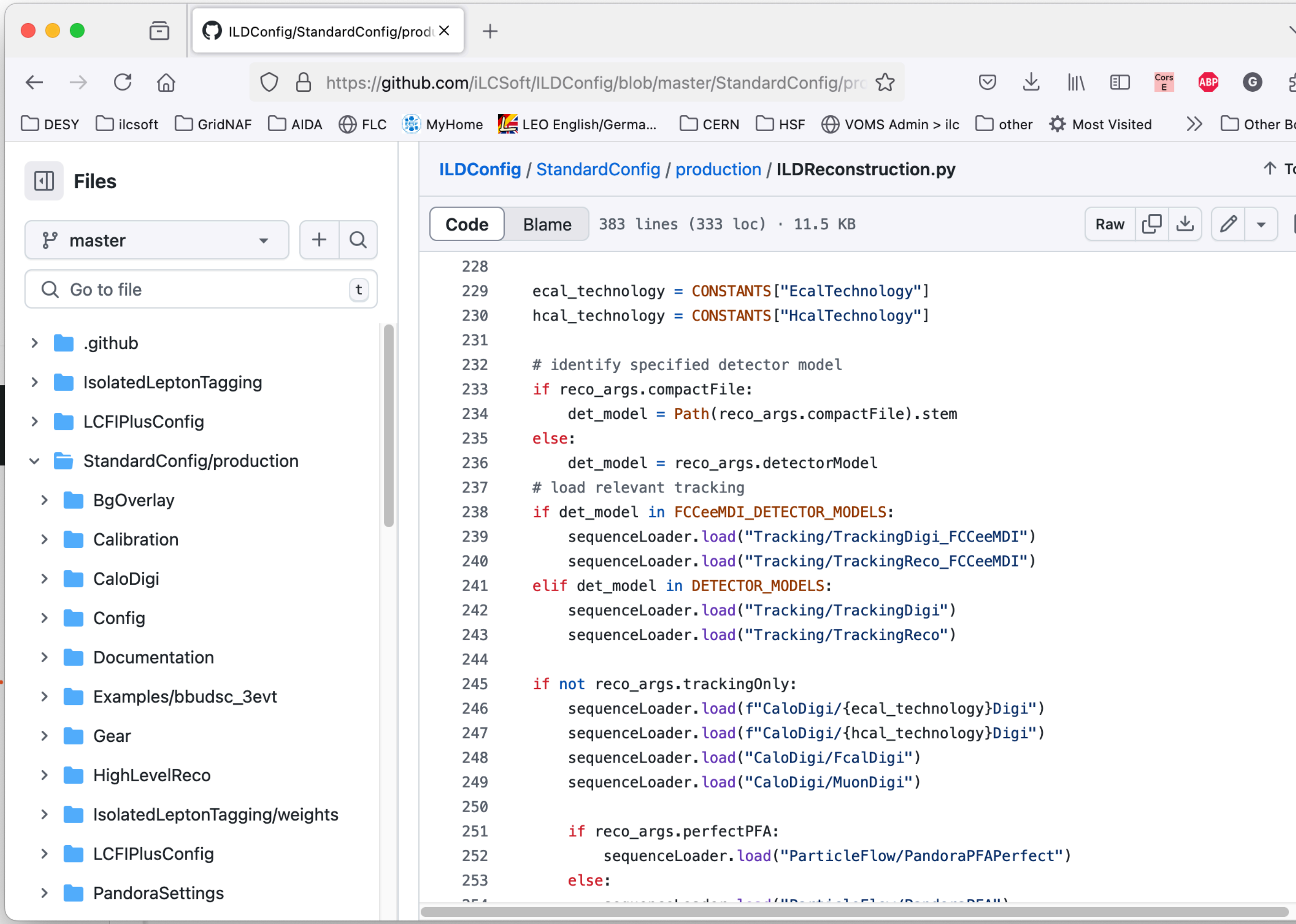
- All configuration available from iLCSoft/ILDConfig
- Everything that works in iLCSoft also works in Key4hep!

```
Marlin MarlinStdReco.xml --global.LCIOInputFiles=<input-file> [.
```

- Now also with Gaudi

```
k4run ILDReconstruction.py --inputFiles=<input-file> [...]
```

  - Works with EDM4hep and LCIO inputs
  - EDM4hep output by default, LCIO output via `--lcioOutput=[true|only]`
- Facilitates collaboration with other projects, e.g. CLD
- Full migration of all workflows will take some time but process started
  - Some new developments already done exclusively in Gaudi configuration



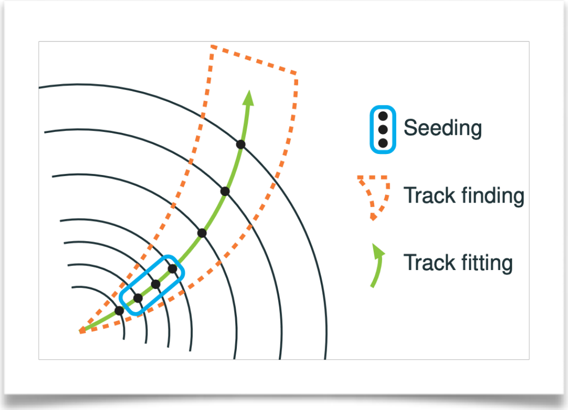
T.Madlener

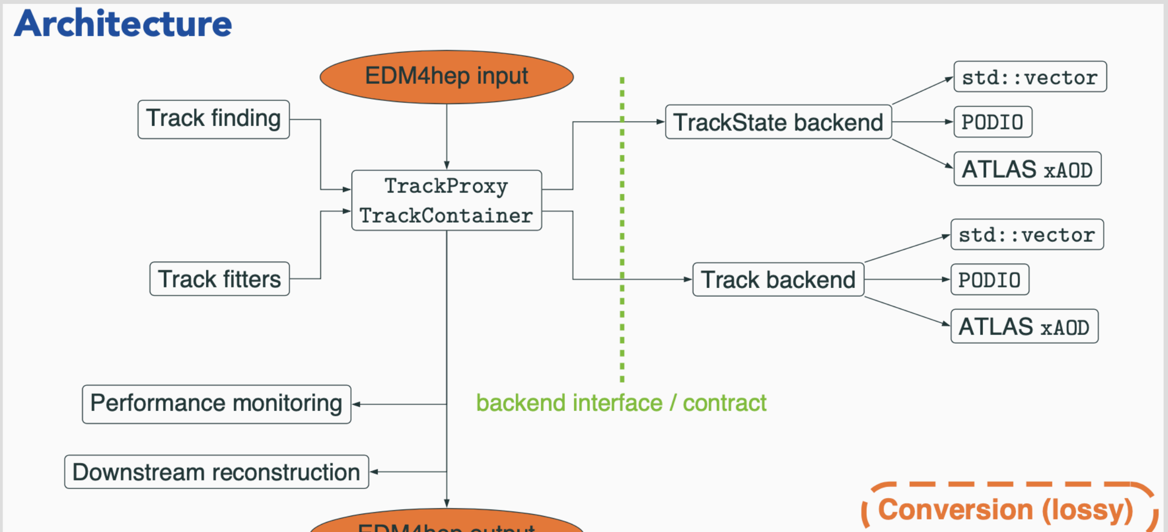
ACTS - A Common Tracking Toolkit

used in Key4hep

- **ACTS** tracking toolkit is the the current choice for track fitting (and finding) in Key4hep
- new package **k4ACTSTracking** provides ACTS fitter
  - implemented as GAUDI algorithm
  - uses DD4hep geometry
- first implementation of TrueTrackfinder for CERN OpenDetector
- successfully used for MuonCollider w/ handcrafted tracking geometry
- ongoing work on **automatic** and **transparent** construction of **tracking geometry** for all detectors in k4geo (Key4hep)
- not clear how soon this will be resolved…

L.Reichenbach



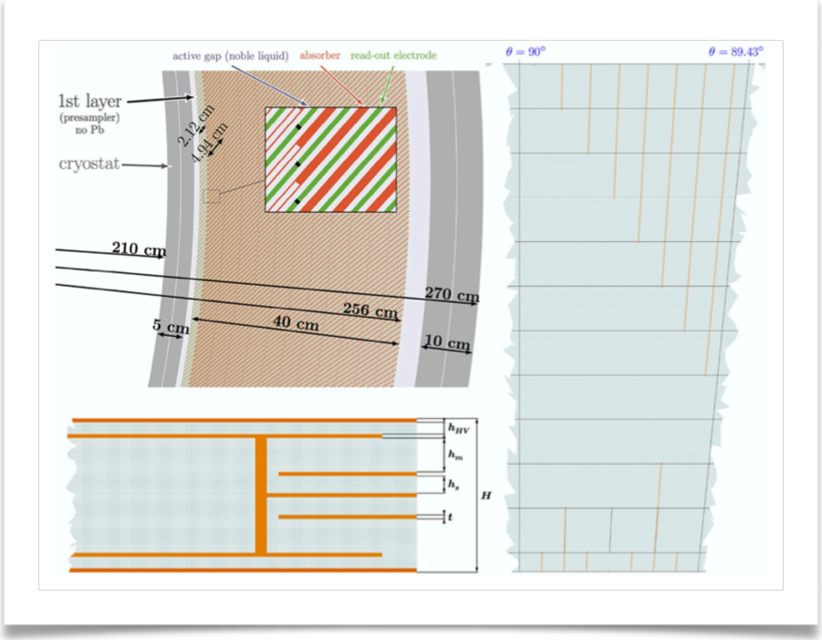


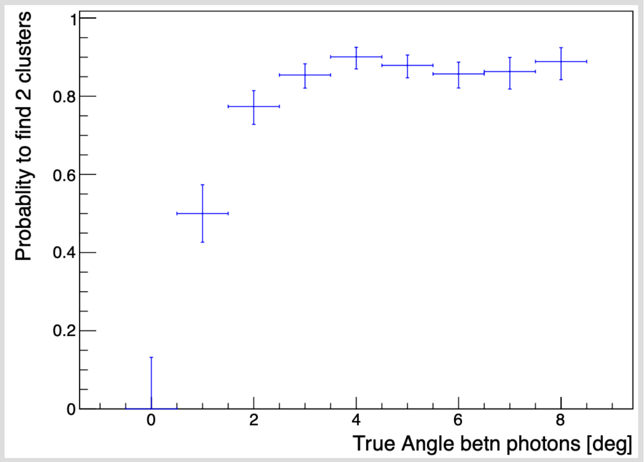
P.Gessinger, CHEP 2023

PFA for LAr Calorimeters

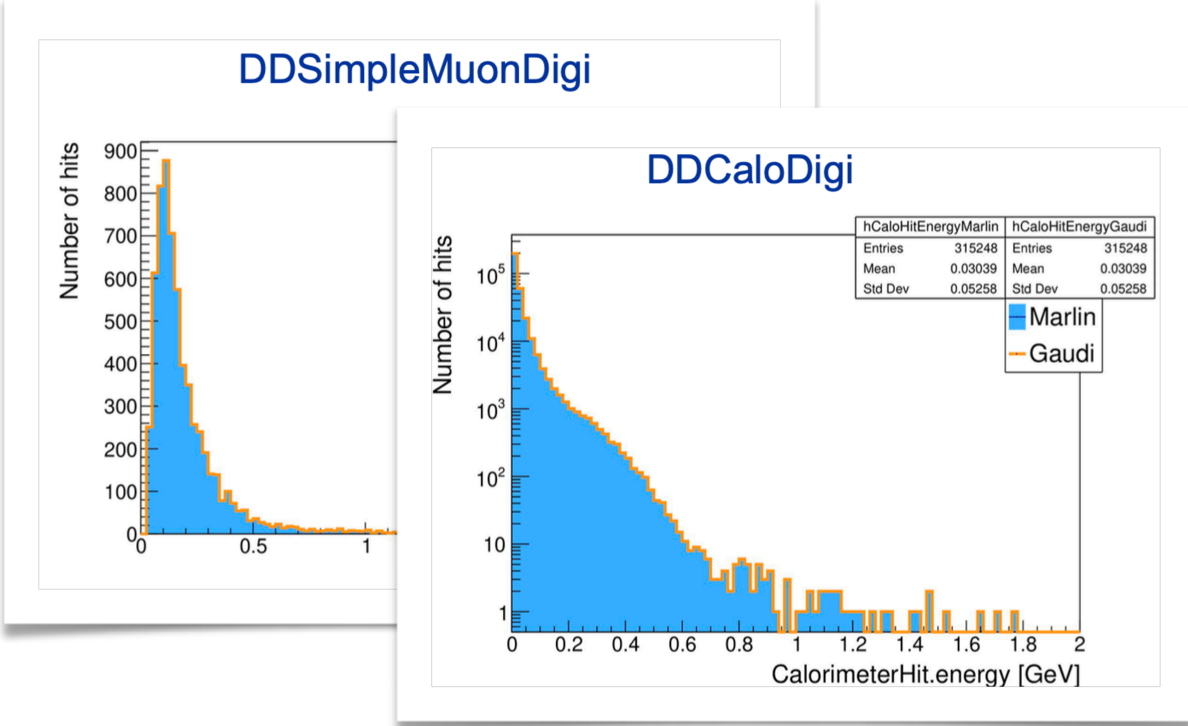
k4GaudiPandora: interfacing PandoraPFA to Key4hep

- goal to have general interface to PandoraPFA for all future collider detectors that have a DD4hep geometry model
- first implementation adapt use of **PandoraPFA for LAr calorimeter** reconstruction
  - needed ‘re-interpretation’ of calorimeter layers as used in highly granular calorimeters
  - adressed in DD4hep::MaterialManager
- used special version of CLD detector with a LAr calorimeter
- started porting DDMarlinPandora to **k4GaudiPandora**





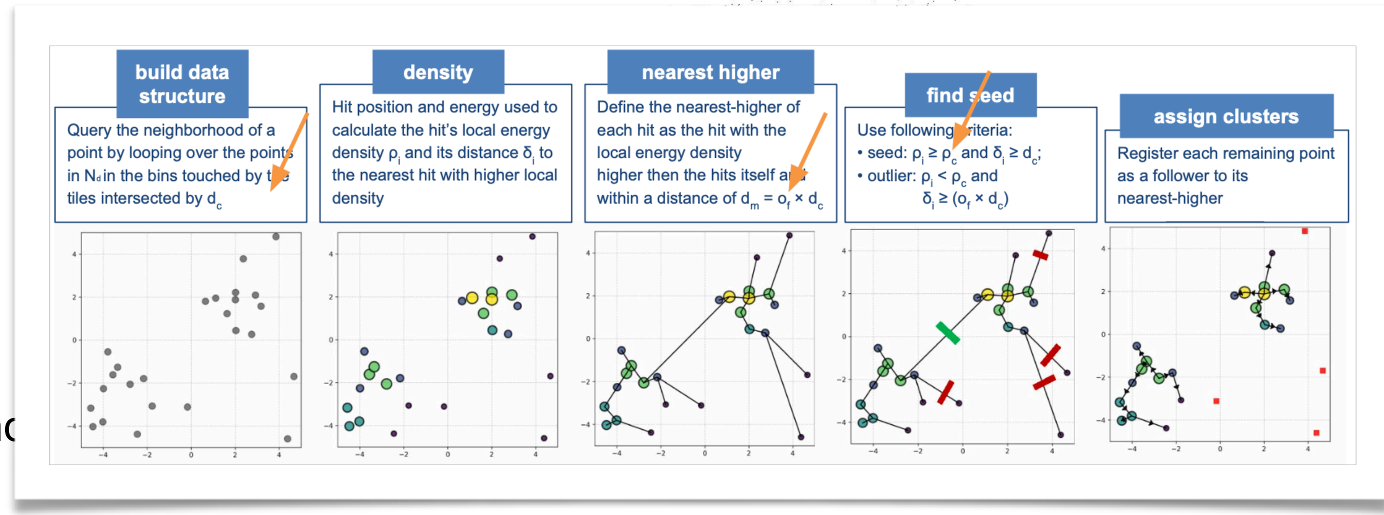
S.Sassikumar



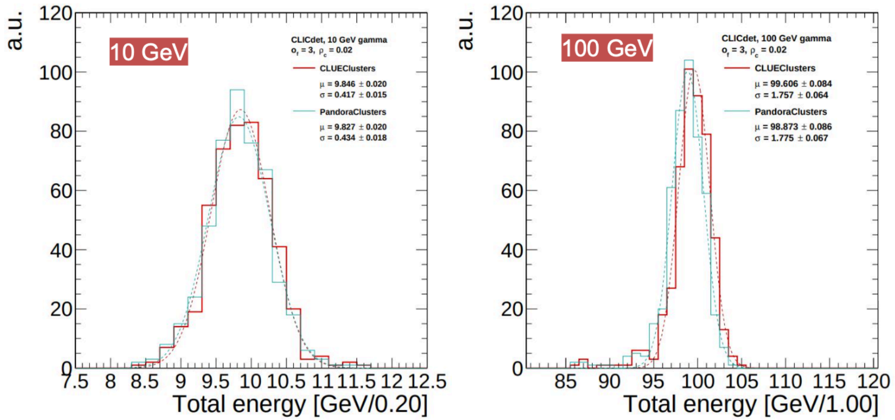
k4Clue

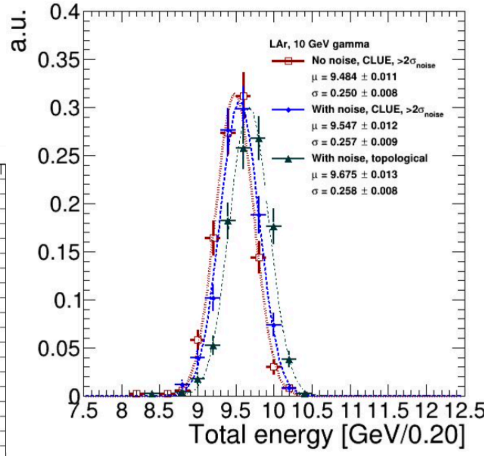
CLUStering Energy

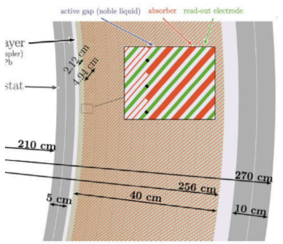
- originally developed for CMS HGCal:
  - fast clustering for high granular calorimeters - based on local energy density
- ported to Key4hep and extended to 4pi geometry
- dedicated tuning of clustering parameters for CLD and LAr ECal
  - shows performance comparable to pre-existing dedicated algorithms
- versatile clustering algorithm for a variety of different calorimeter technologies

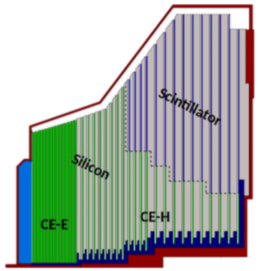


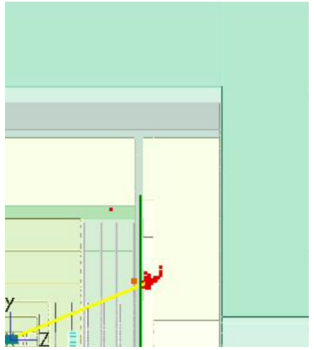
E.Brondolin









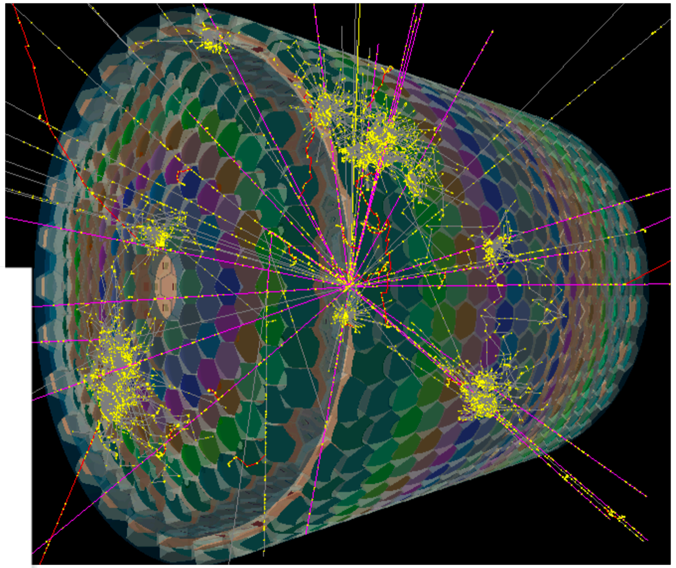


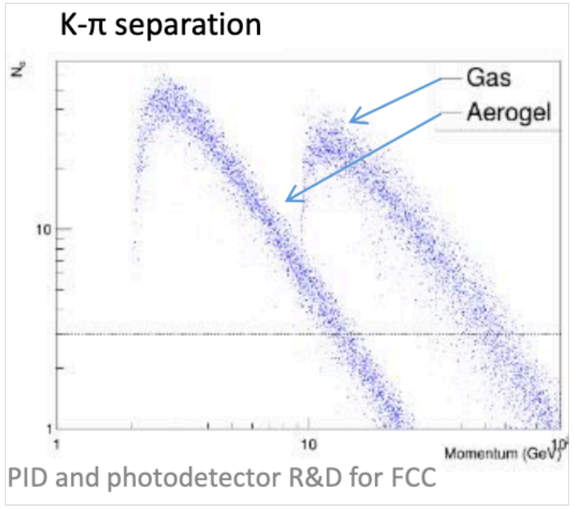
ParticleID performance with the ARC

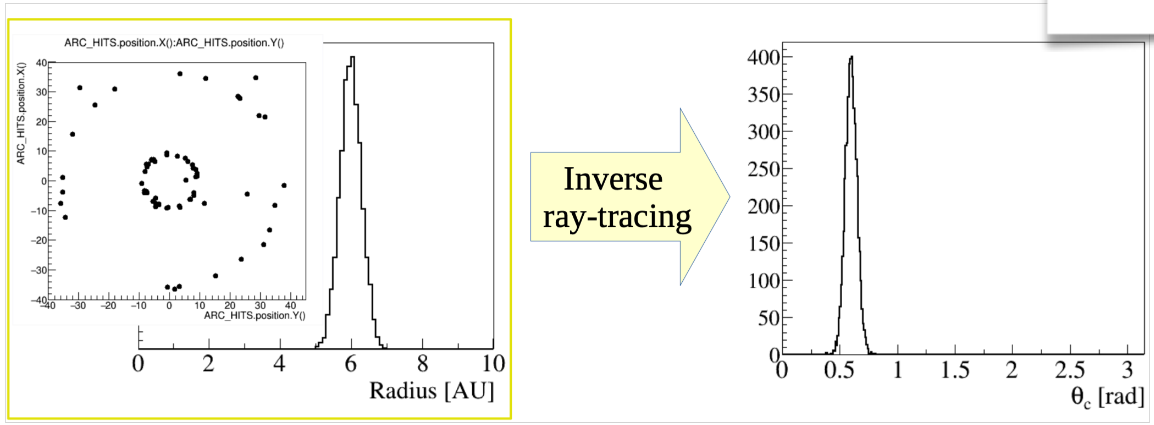
a novel GAUDI algorithm in Key4hep

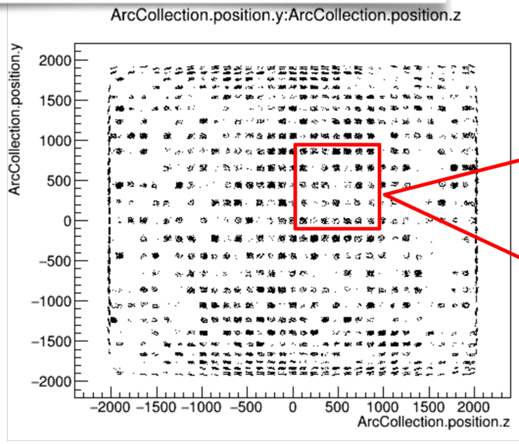
A. Tolosa Delgado

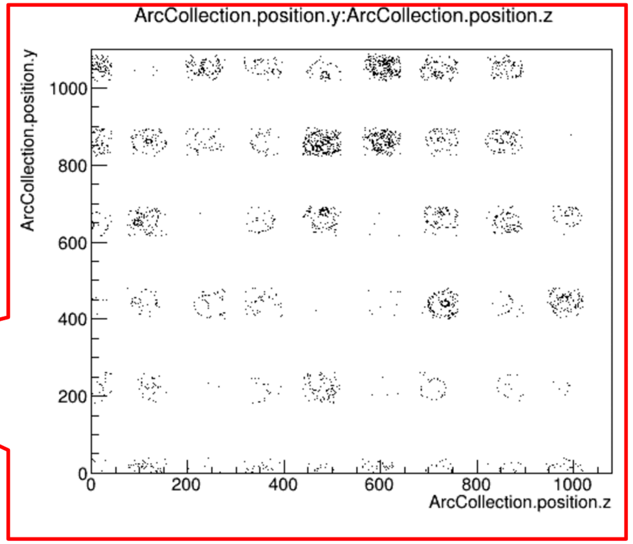
- can simulate full events in **CLD w/ ARC** with dddim (DD4hep)
- standalone reconstruction w/ inverse ray-tracing exists for single cell
  - should provide excellent K-pi separation from 2-50 GeV
- ongoing work: full ARC reconstruction in Gaudi - aim for end of summer











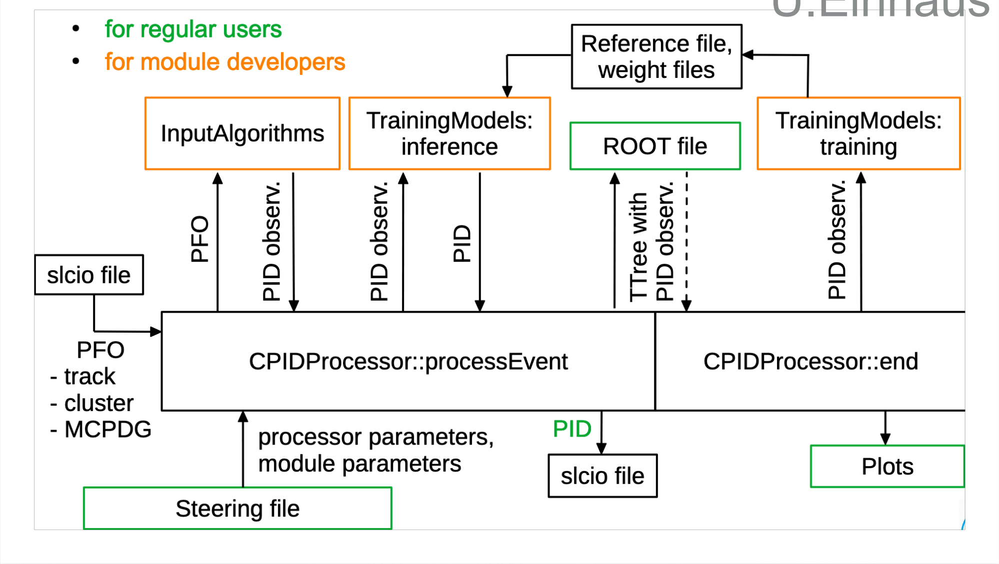
Sagital cut of the detector (events on the endcaps are compressed in Z, not visible in this projection)

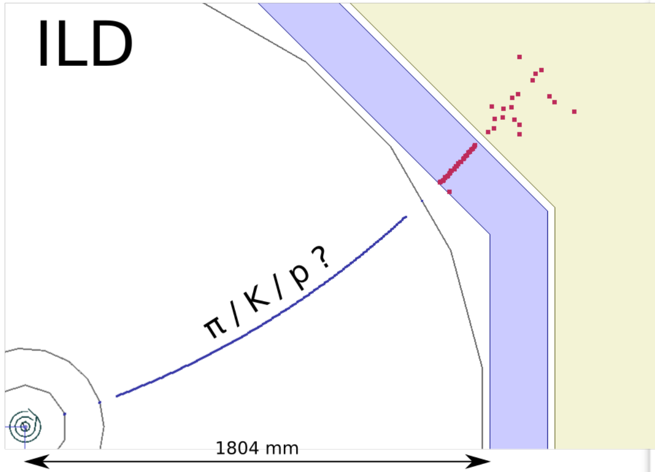
Particle identification

Using dE/dx, TOF and AI/ML

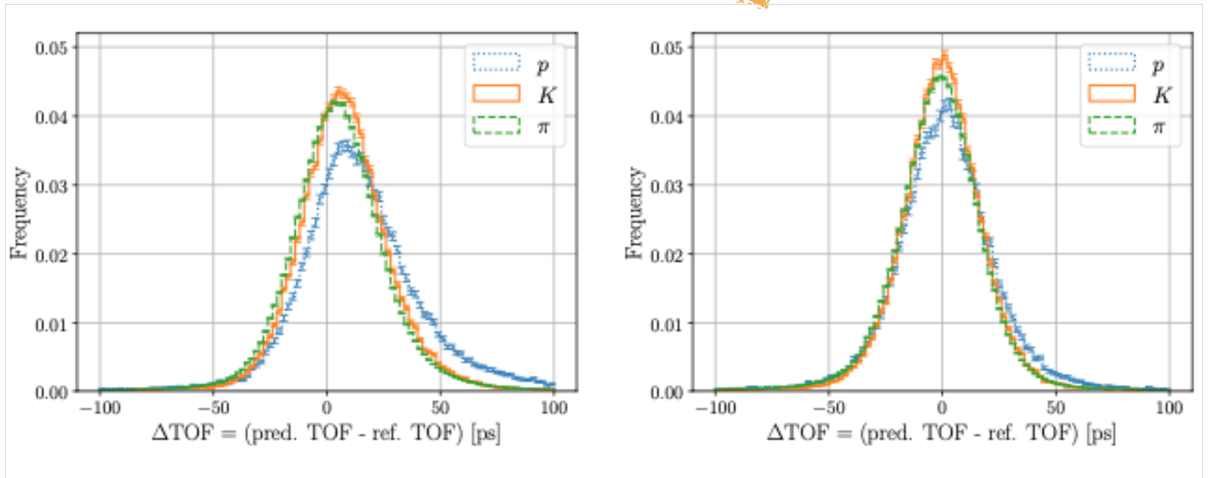
- comprehensive particle identification **CPID:**
  - toolkit for ML based PID (in ILD) using: dE/dx, **TOF**, PandoraPFAID, dedicated leptonID
  - **applicable to other (all) HET detectors**
- dedicated study of extracting **TOF from the calorimeter in ILD**
  - w/ classical and AI/ML estimator

U.Einhaus



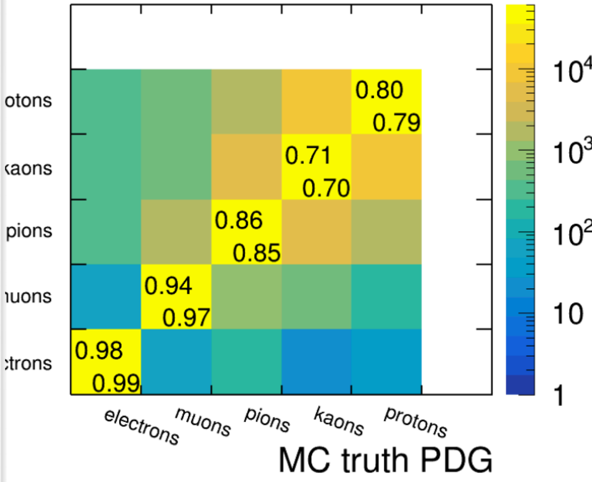


$$L = \sum_i^{N_{\text{hits}}} L_i = \sum_i^{N_{\text{hits}}} \frac{|z_{i+1} - z_i|}{|\tan \lambda_i|} \sqrt{1 + \tan^2 \lambda_i}$$



(a) Cylinder algorithm. (b) TOF transformer.

B.Dudar, K.Helms

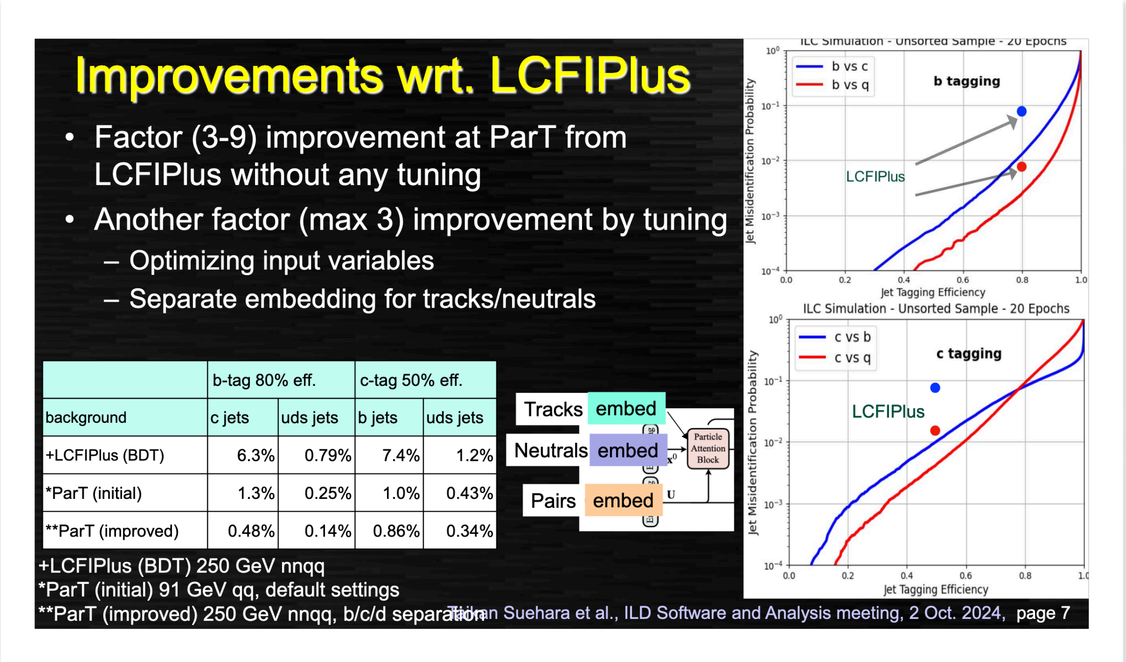


flavour tagging with deep learning methods

AI/ML in Key4hep

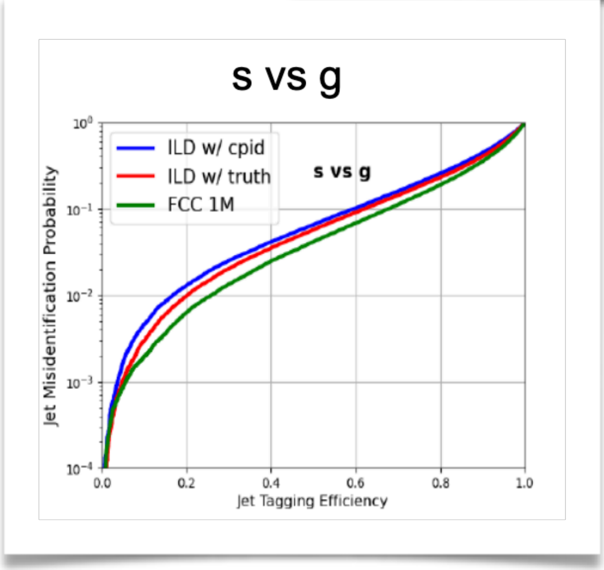
T.Suehara et al

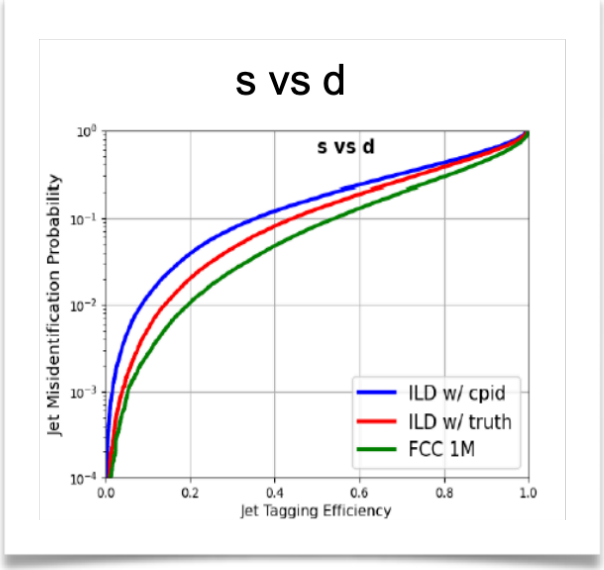
- implemented ParticleTransformer flavour tagging for ILD
  - originally developed for CMS and adapted to FCCee
  - achieve **dramatically better** results than LCFIPlus
- massive improvement w/ more training data at FCC -> to be studied !
- framework (Marlin/Gaudi?) inference work in progress
- started to look into **strange tagging**



Sample / sample size	b-tag 80% eff.		c-tag 50% eff.	
background	c jets	uds jets	b jets	uds jets
ILD full-sim 1M (optimized)	0.48%	0.14%	0.86%	0.34%
FCCee Delphes 1M (reduced)	0.47%	0.12%	0.64%	0.10%
FCCee Delphes 1M (full)	0.21%	0.054%	0.36%	0.059%
FCCee Delphes 4M	0.045%	0.025%	0.20%	0.033%
FCCee Delphes 6M	0.014%	0.010%	0.13%	0.022%
FCCee Delphes 8M	0.007%	0.006%	0.076%	0.021%

We see mild consistency between ILD and FCC!



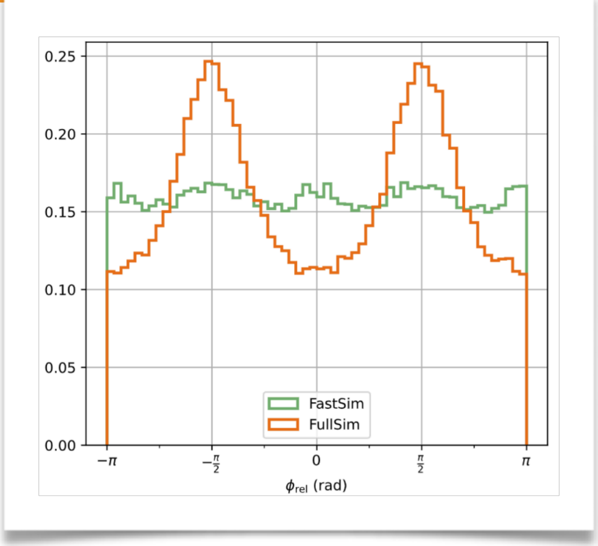


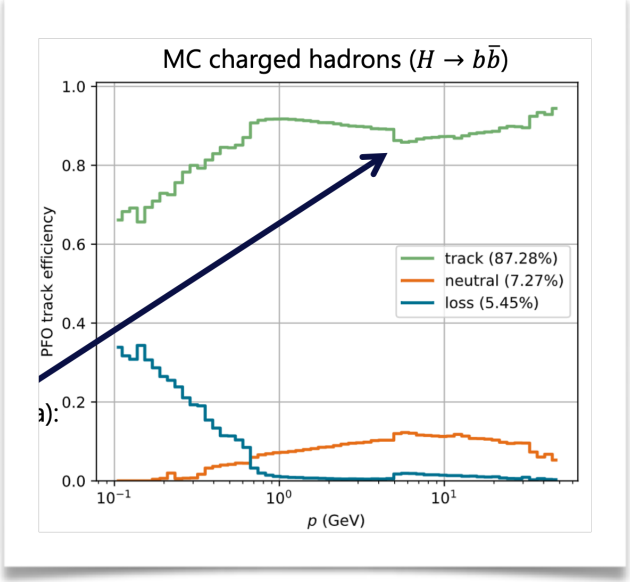
flavour tagging with deep learning methods

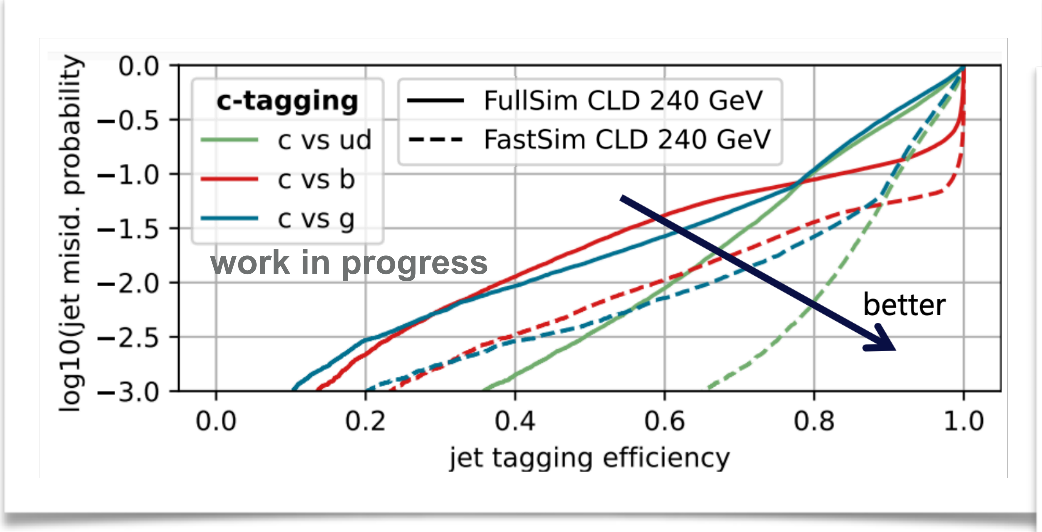
comparing fast sim and full sim based flavour tagging with CLD

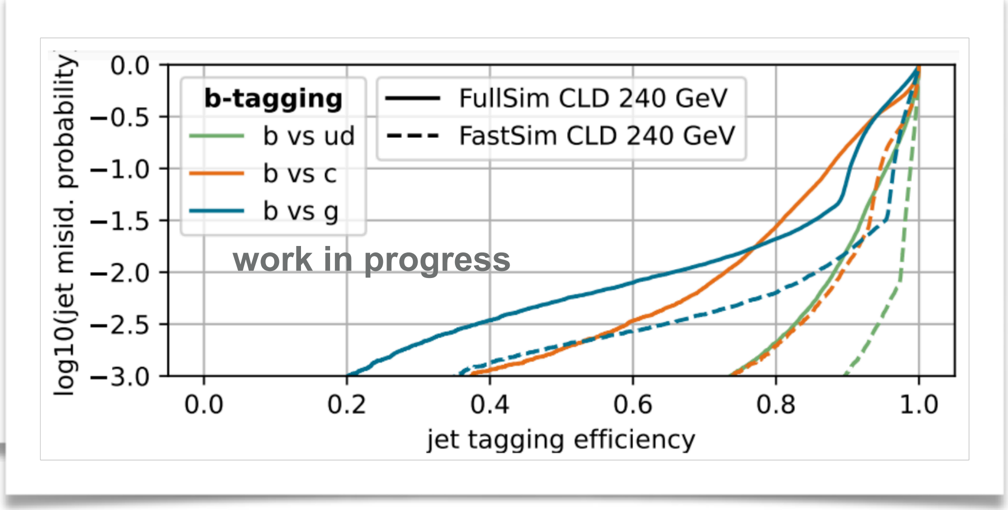
S.Aumiller

- training a PartTransformer for CLD w/
  - Delphes and Fullsim samples
- observe differences in extra neutrals due to
  - split clusters in real algorithm and due to tracks intentionally dropped as no cluster found
- leads to significant differences in tagging performance







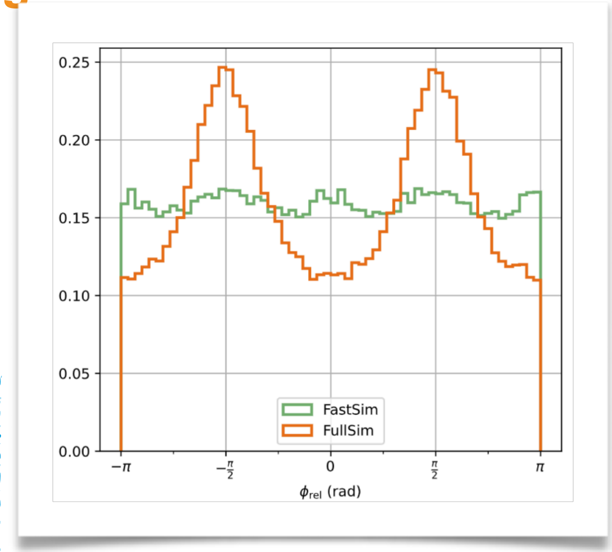


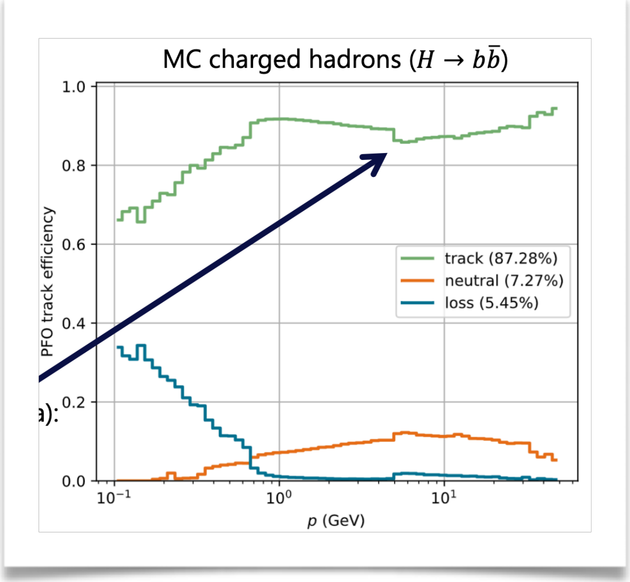
flavour tagging with deep learning methods

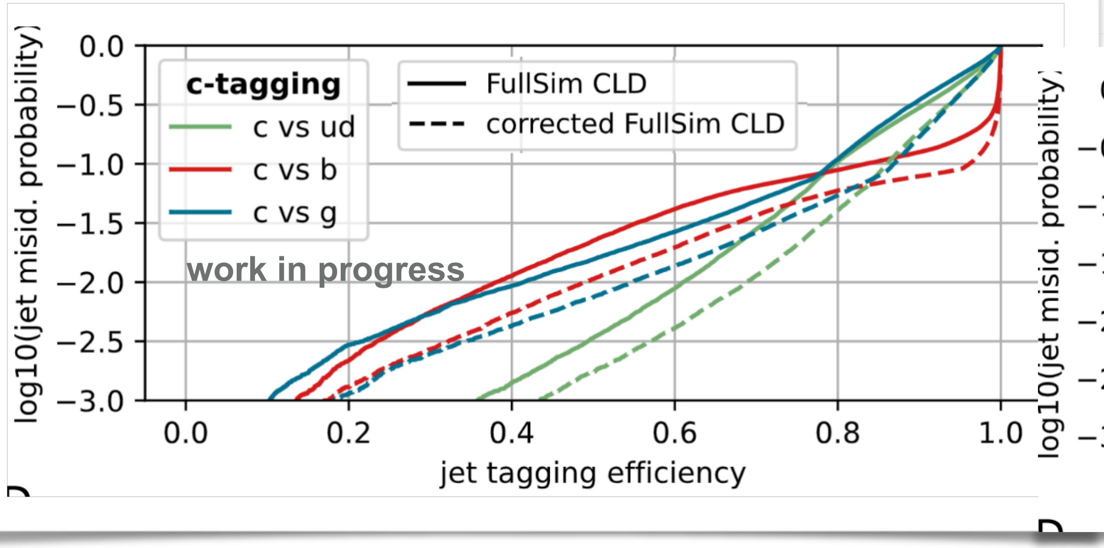
comparing fast sim and full sim based flavour tagging with CLD

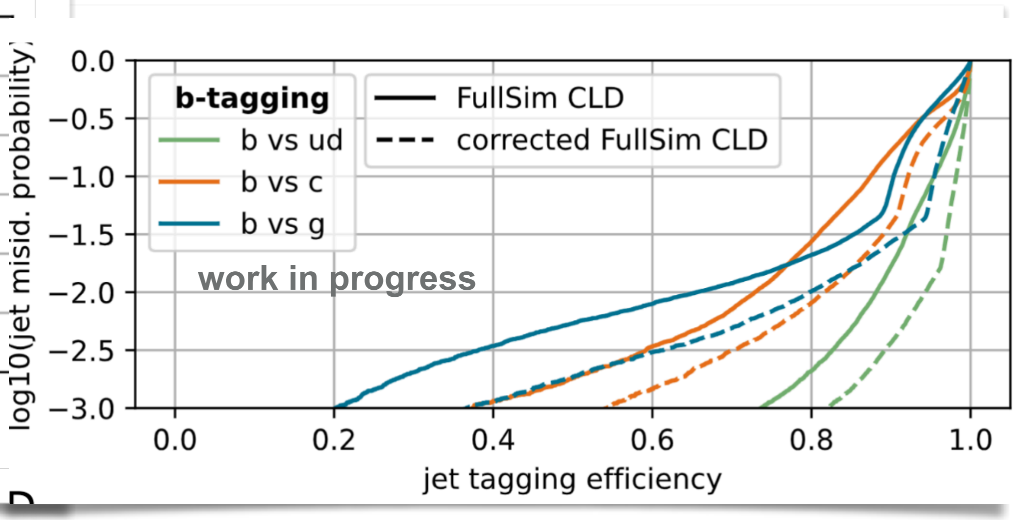
S.Aumiller

- training a PartTransformer for CLD w/
  - Delphes and Fullsim samples
- observe differences in extra neutrals due to
  - can recuperate some of the flavour tag performance with better reconstruction
    - here cheating w/ MCTruth for demonstration





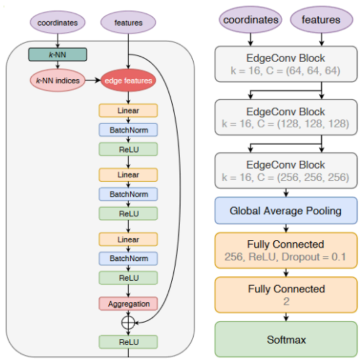




flavour tagging with deep learning methods

Flavour Taggers

- Long-time (> 10 a) standard tool: LCFIPlus, now out-of-date with BDT and b/c/light tag
- A number of new flavour taggers have been developed, making use of neural networks and advanced machine learning
  - ParticleNet (ILD/CLD) [Meyer]
  - ParticleNet (IDEA) [Garcia e.a.]
  - ParticleNet (CEPC baseline) [Ruan e.a.]
  - Particle Transformer (ILD, IDEA) [Suehara]
  - Particle Transformer (CLD, Si-IDEA) [Aumiller e.a.]
- Adding strange-tag via PID (kaonID), possibly individual u- and d-tag via FSR
- Systematic uncertainties on flavour taggers?
  → calibrate at Z-pole with double-tag method and 10^{12} events (FCC-ee) or some 10^{9} (LC)
  → limits to detector adaptions to center-of-mass energy?



Karlsruher Institut für Technologie

Ulrich Einhaus |

Jet Flavour Tagging at FCC-ee with a Transformer-based Neural Network: DeepJetTransformer

Freya Blekman [1,2,4] Florencia Canelli [3] Alexandre De Moor [1] Kunal Gautam [1,3] Armin Ilg [3] Anna Macchiolo [3] Eduardo Ploerer [1,3]

fast simulation (Delphes) of IDEA concept

D.Jeans

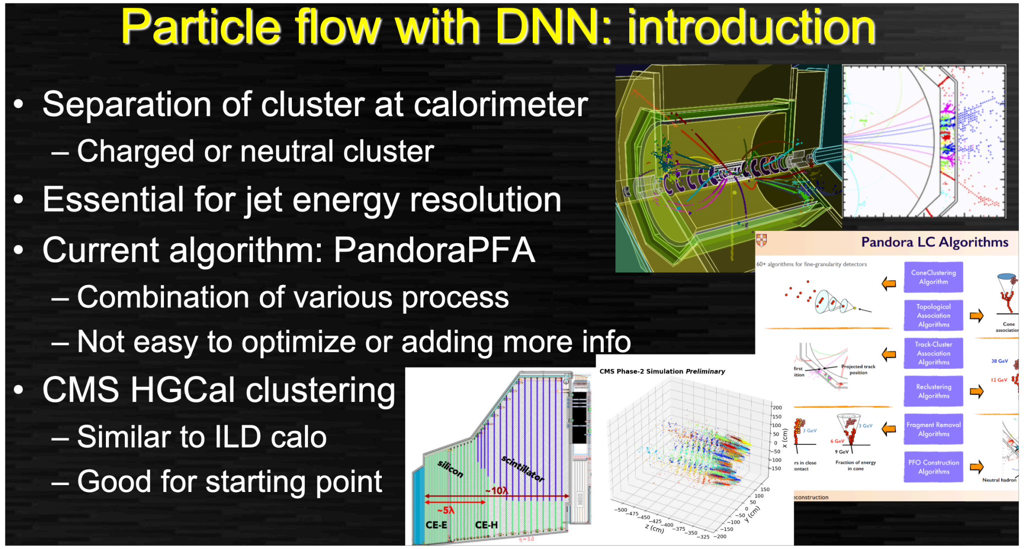
- large number of ML flavour taggers exist for HET detectors
- need to systematically benchmark and understand the performance of each algorithm (on each detector)

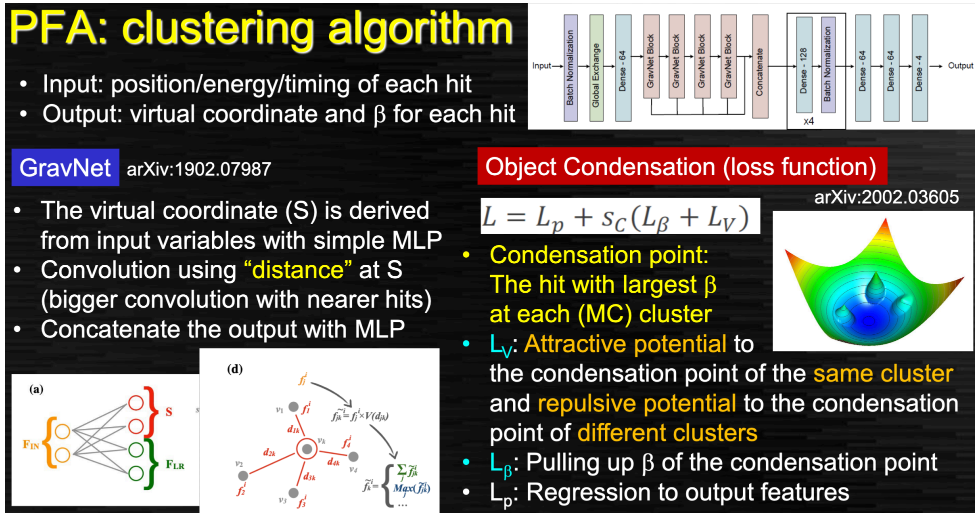
Particle Flow with Machine Learning

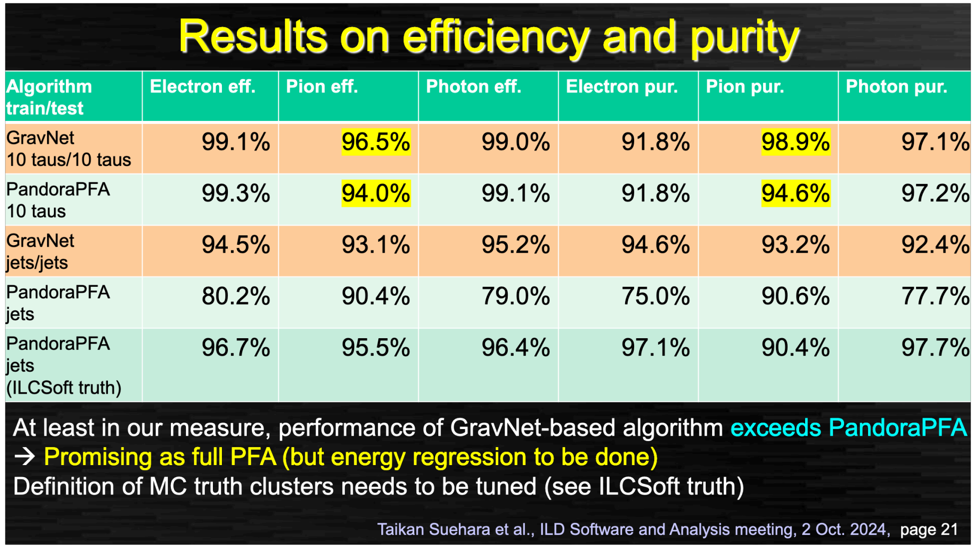
T.Suehara

AI/ML in Key4hep

- started to develop deep neural networks for particle flow in the ILD detector
    - using GravNet and Object Condensation
- after early, promising results **made significant progress**
    - **better** than **PandoraPFA** in one selected metric







Results on efficiency and purity

Algorithm train/test	Electron eff.	Pion eff.	Photon eff.	Electron pur.	Pion pur.	Photon pur.
GravNet 10 taus/10 taus	99.1%	96.5%	99.0%	91.8%	98.9%	97.1%
PandoraPFA 10 taus	99.3%	94.0%	99.1%	91.8%	94.6%	97.2%
GravNet jets/jets	94.5%	93.1%	95.2%	94.6%	93.2%	92.4%
PandoraPFA jets	80.2%	90.4%	79.0%	75.0%	90.6%	77.7%
PandoraPFA jets (ILCSoft truth)	96.7%	95.5%	96.4%	97.1%	90.4%	97.7%

At least in our measure, performance of GravNet-based algorithm exceeds PandoraPFA
→ Promising as full PFA (but energy regression to be done)
Definition of MC truth clusters needs to be tuned (see ILCSoft truth)

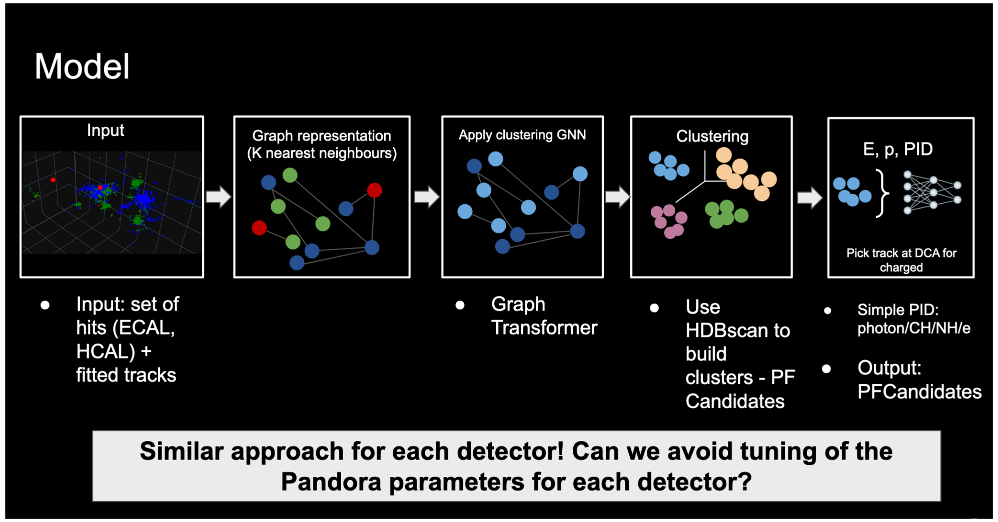
Taikan Suehara et al., ILD Software and Analysis meeting, 2 Oct. 2024, page 21

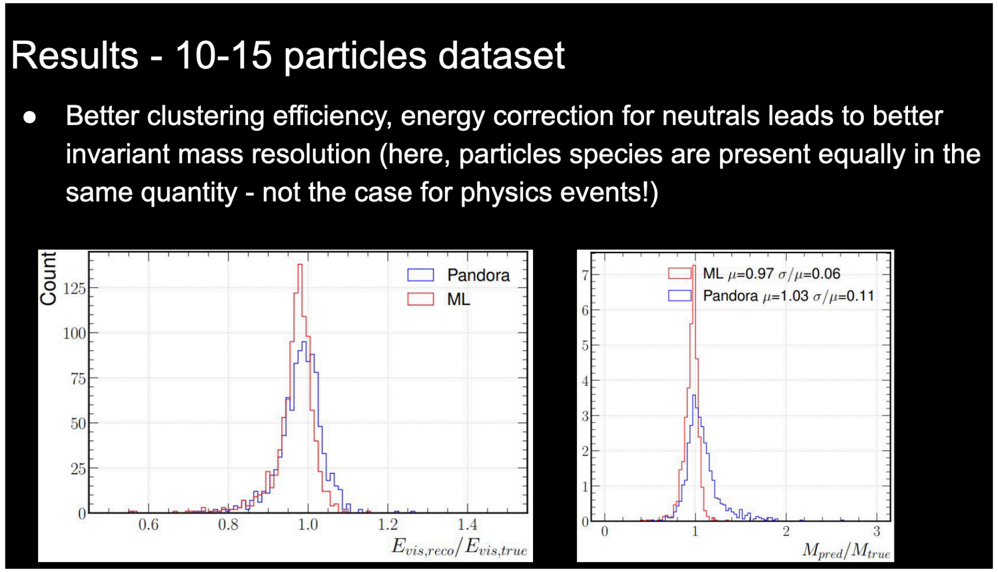
Particle Flow with Machine Learning

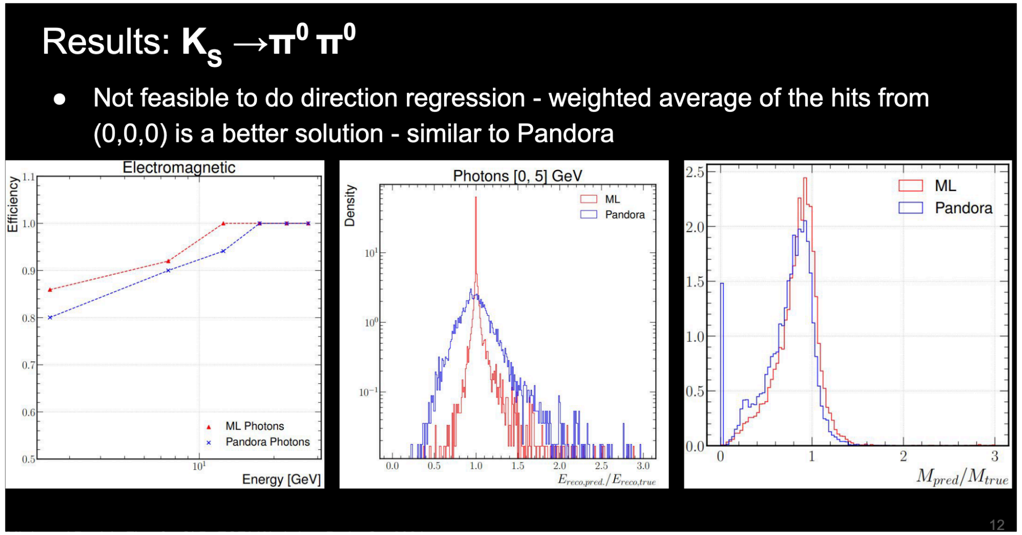
G.Krzmanc

AI/ML in Key4hep

- alternative approach ML-PFA w/ GNN using full simulation for CLD
- also observe very **promising results** - **better** than the well established **PandoraPFA** on simplified events
  - would be interesting to see in on more realistic collision events (ee->uds ?)





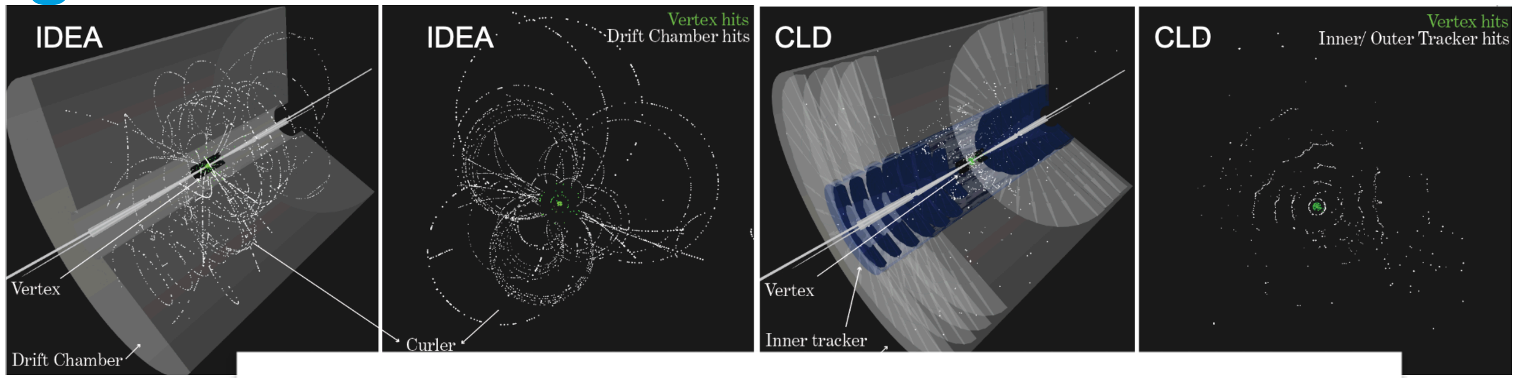


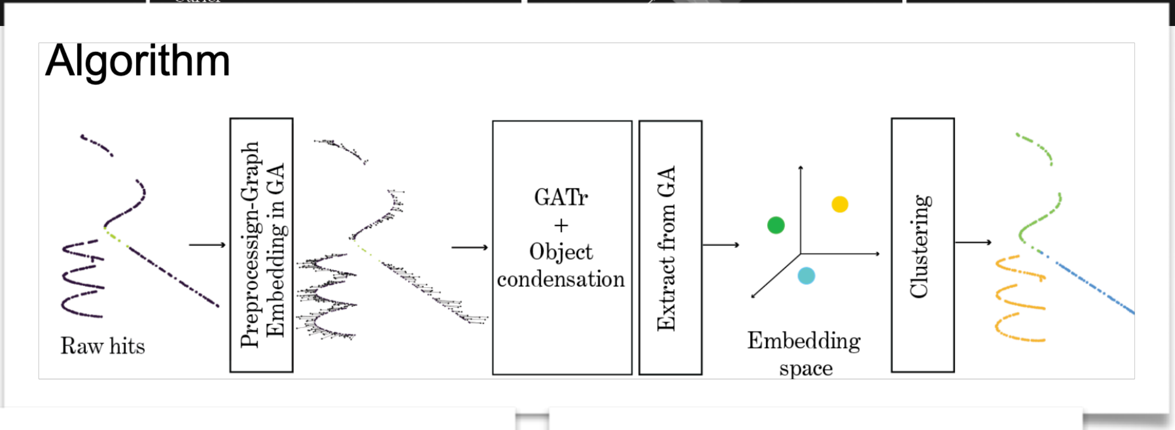
AI/ML based Pattern Recognition

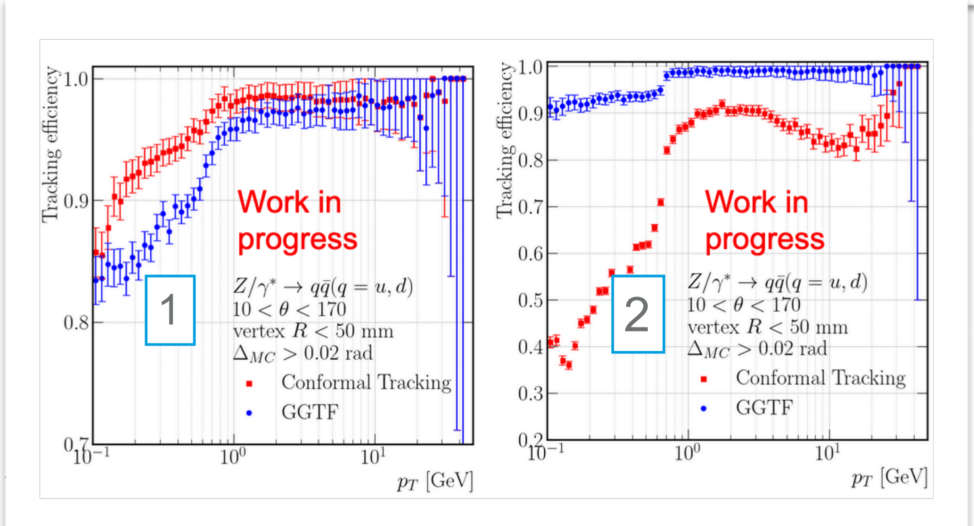
D.Garcia

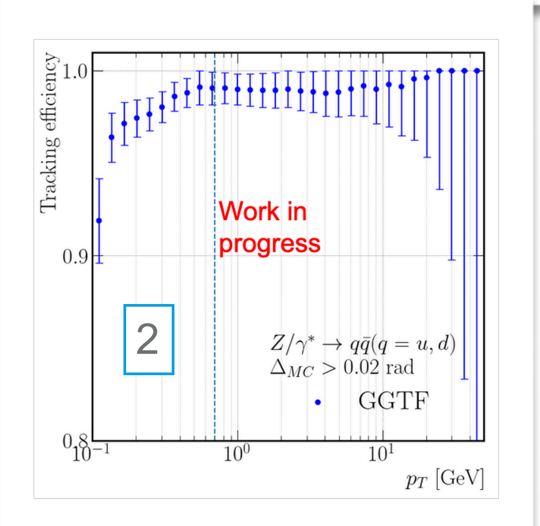
for CLD and IDEA

- goal: ML based track finding algorithms that is largely geometry and detector independent
- developed GGTF for IDEA and CLD !
- for CLD observe higher efficiency but lower purity compared to ConformalTracking
- works also nicely for IDEA w/ drift chamber
- implemented in Key4hep-GAUDI







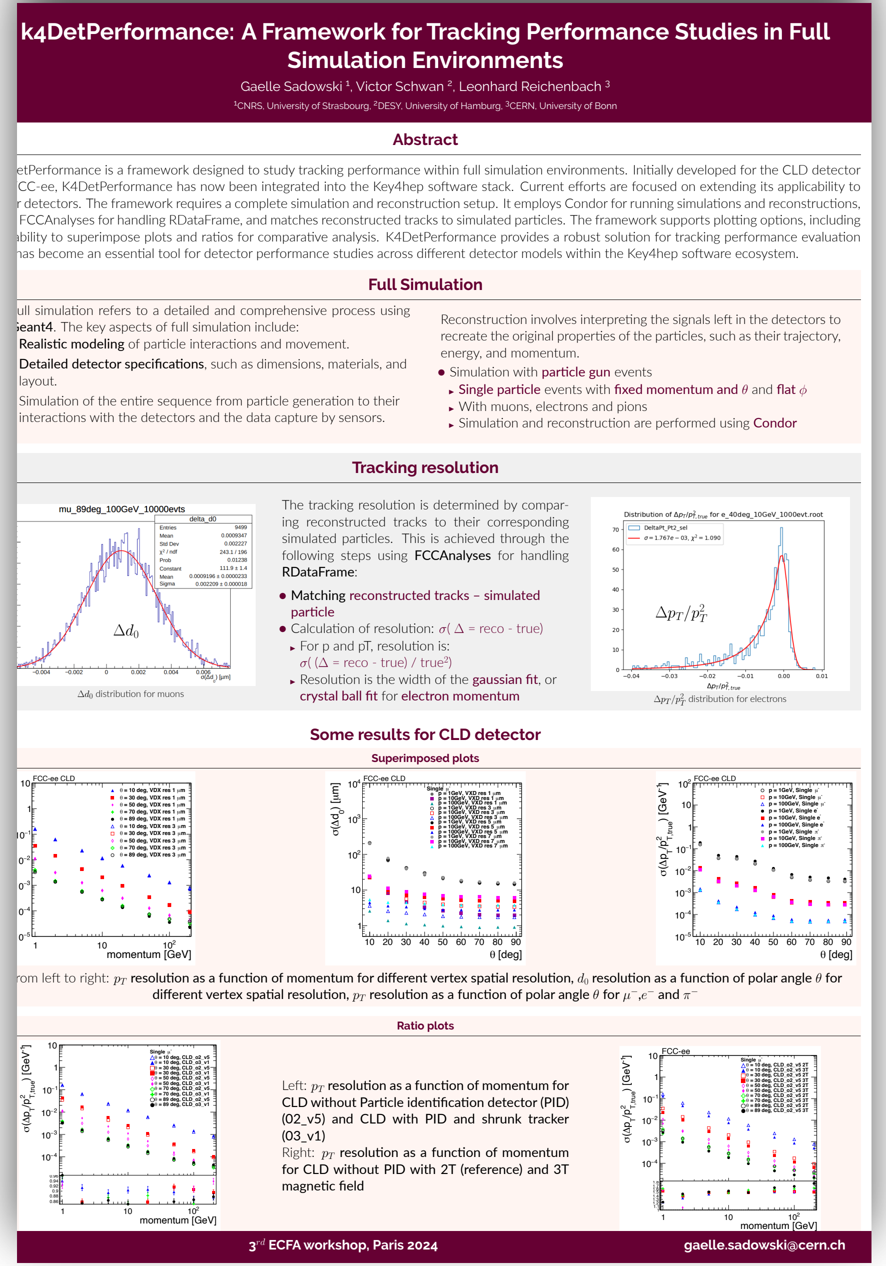


1: hit purity > 75%
2: hit purity & eff. > 50%

Benchmarking

Algorithms and eventually detector concepts

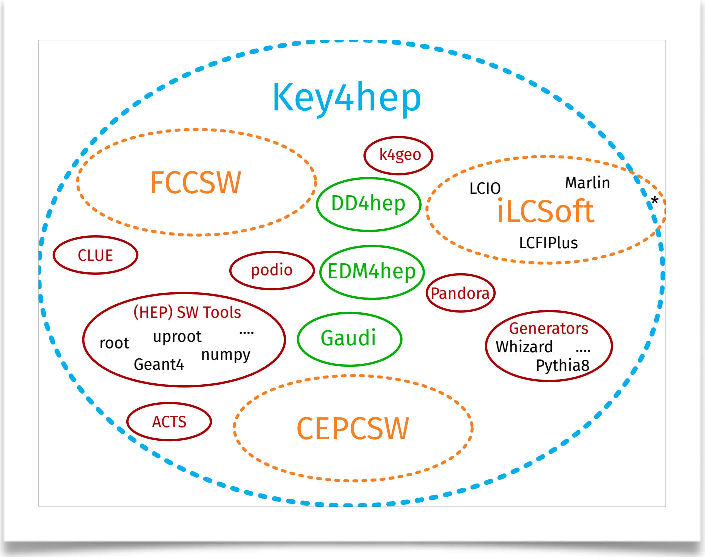
- ultimate physics performance is a combination of detector and reconstruction performance
- for developing optimised detectors benchmarking is absolutely indispensable
- started a new package *k4DetPerformance* w/
  - tracking performance benchmarks
    - residuals, pulls, resolutions, efficiencies
    - …
- in Key4hep-Gaudi that works for all HET detectors
- should be extended to PFA, favour tagging, PID …



G.Sadowski

Summary and Outlook

- **Key4hep** started as a new future collider community wide effort in 2020 to put together a modern turnkey software stack for all future colliders
  - contributors: CEPC, CLIC, FCC, EIC, ILC, LUXE, Muon Collider …
- existing **standard reconstruction** (linear colliders) can be run in Key4hep w/ **MarlinWrapper** as before - or w/ EDM4hep output
- first **genuine Key4hep**/EDM4hep/Gaudi **reconstruction algorithms** start to become available (k4Clue, k4ACTS, k4GaudiPandora,….)
- many new developments in (high level) reconstruction tools
  - more and more **ML/AI** based !
- **full simulation and reconstruction** is important for detector optimisation - as well as systematic **benchmarking**



ECFA Report: Reconstruction editors Ulrich Einhaus, Loukas Gouskos, Taikan Suehara

- development of (reconstruction) tools is mostly manpower limited
- vital for future collider community that Key4hep continues to receive funding