Summary of Software development in AIDAinnova

Gérald Grenier
material taken from J. Back, G. Cibinetto, F. Gaede, T. Madlener, A. Zaborowska

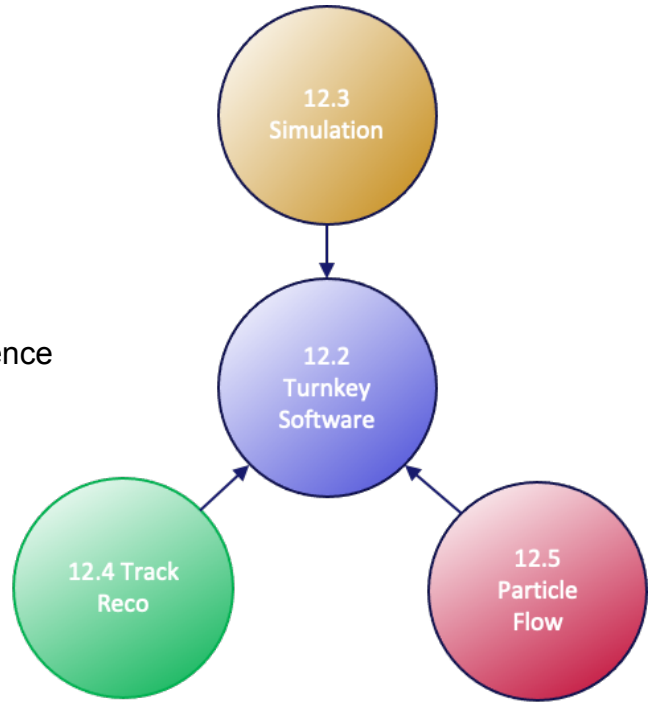
Institut de Physique des 2 Infinis de Lyon (IP2I)

4th FCC/DRD France Workshop, Strasbourg
Nov 22-24, 2023

département
Physique
Université Claude Bernard Lyon 1

WP12 Structure

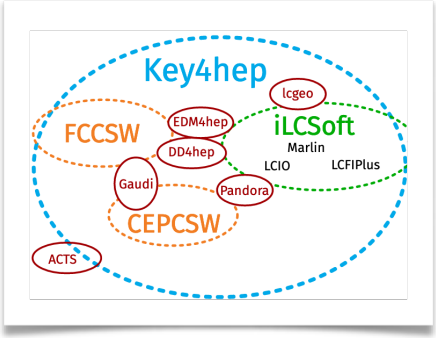
- **Task 12.2. Turnkey Software**
  - Turnkey Software Stack, for physics and performance studies, EDM4hep, PODIO and Digitisation toolkit
  - R&D study on frameworks to manage heterogeneous resources
- **Task 12.3. Simulation**
  - Fast simulation techniques integrated into Giant
  - Machine learning based calorimeter simulation toolkit for training and inference
- **Task 12.4. Track Reconstruction**
  - complete track reconstruction with ACTS composable algorithms and for heterogeneous computing
  - Machine learning reconstruction algorithm for MPGD detectors
- **Task 12.5. Particle Flow Reconstruction**
  - PFA algorithms for DUNE and dual-readout calorimeters, APRIL PFA for hadronic jets

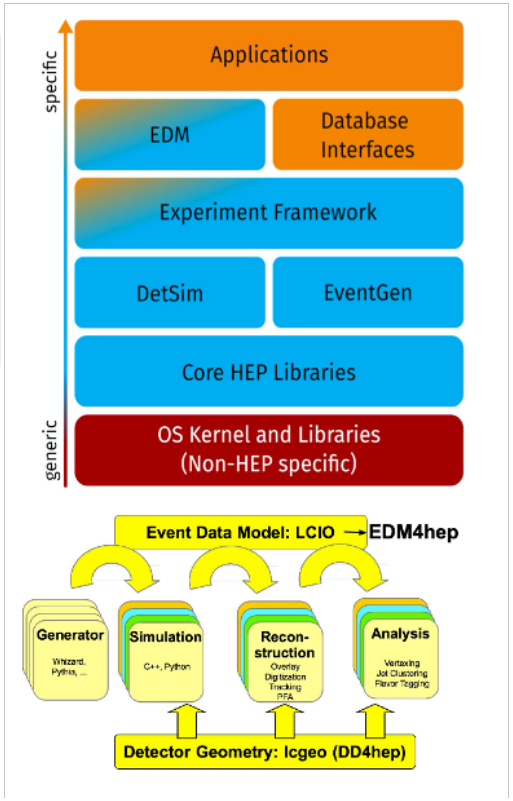


Task 12.2 Turnkey Software

turnkey software stack for all future collider projects

- take existing tools where possible - reuse existing software from the shared iLCSoft developed by ILC and CLIC
- all major players involved: CEPC, CLIC, FCC, ILC, EIC
- provide a complete data processing framework
  - shared components reduce overhead for all users
- make things as easy to use as possible for everybody (librarians, developers, users)
- supported by HSF, CERN EP R&D and **AIDAinnova**





AIDAinnova data model

PODIO

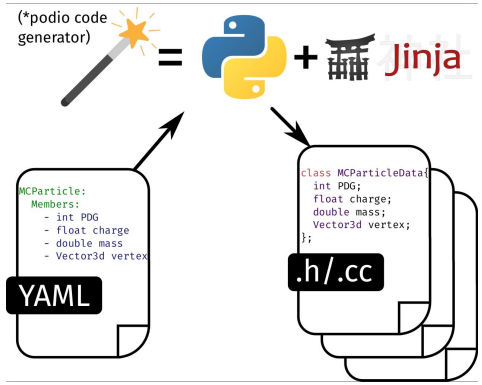
The event data model toolkit

github.com/AIDASoft/podio

- Generate code from simple yaml definition of EDM
- Based on using and storing POD (plain old data) structures
- Make it possible to target different I/O backends

PODIO → C++, pyROOT, Julia

EDM4hep → LCIO, JSON, ...

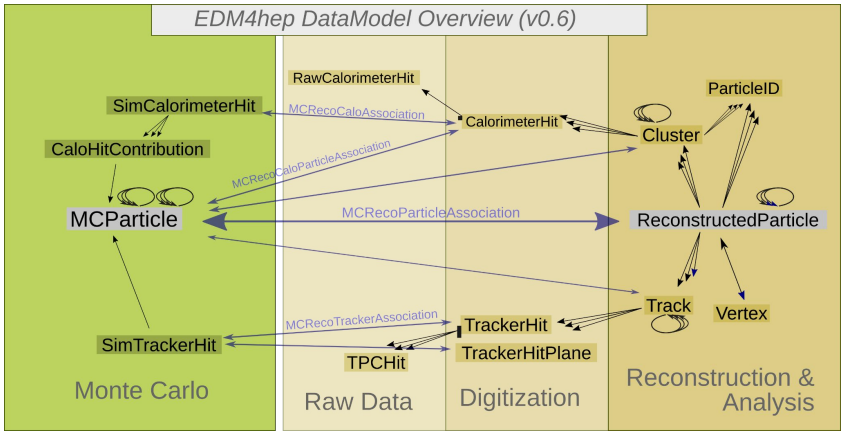


EDM4hep

The common event data model

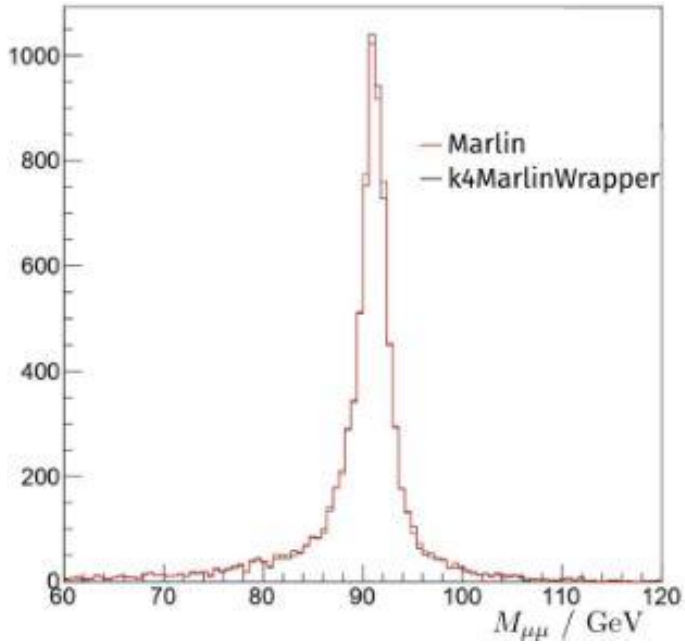
github.com/key4hep/EDM4hep
edm4hep.web.cern.ch

- EDM4hep defines the common *language* for all Key4hep components to communicate
- Heavily inspired by LCIO that has been successfully shared by ILC and CLIC
  - Additional novel ideas from fcc-edm
- Generated by the PODIO EDM toolkit
  - EDM4hep and EICD main customers of PODIO

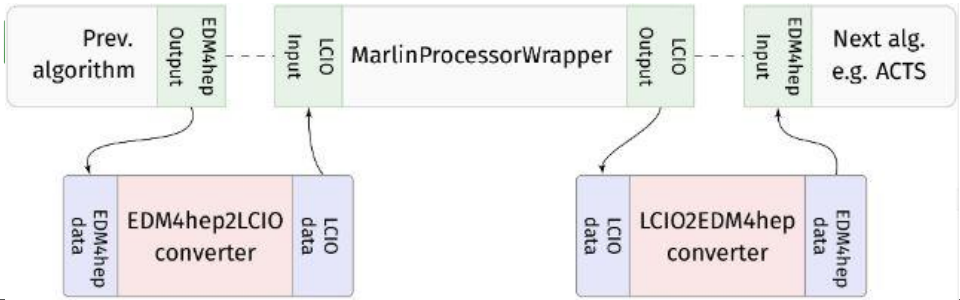


Key4hep Framework

- Gaudi based core framework
  - **k4FWCore** provides I/O for PODIO based EDMs
  - k4SimDelphes for Delphes integration
  - **k4MarlinWrapper** for calling Marlin processors
  - **k4geo** for detector models (rebranded from lcgeo)
  - k4SimGeant4 for Geant4 based simulation
  - k4Gen for generator integration
  - …



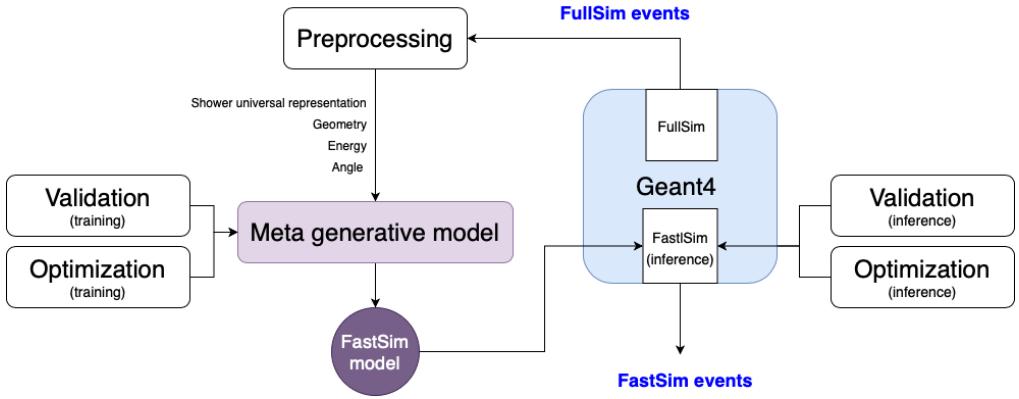
✅ **MS47: Comparable performance to iLCSoft via k4MarlinWrapper**



Integration of ML models

Integration of Machine Learning (ML) models into standard simulation toolkit (GEANT4)

- Demonstration of ML inference in C++ framework
- available in GEANT4 11.0 release, but can be also used with 10.7
- Incorporation of few libraries: ONNX Runtime, LWTNN, Torch
  - Torch was integrated during the last AIDAinnova hackathon, thanks to everyone involved! (CERN, DESY, UniMan)
  - available in GEANT4 11.1 release
- Implemented as a Geant4 example Par04, includes a trained model: Variational Autoencoder (VAE)
- Described in AIDAinnova milestone report



1/7

Various ML models are tested.

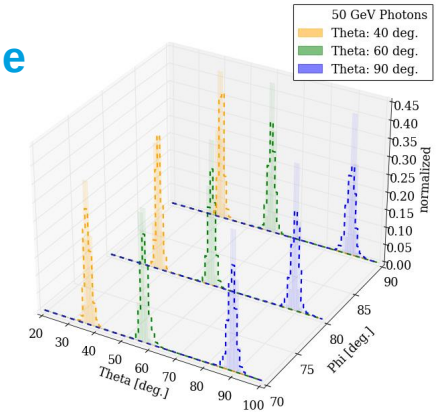
fastsim for Highly Granular Calorimeter

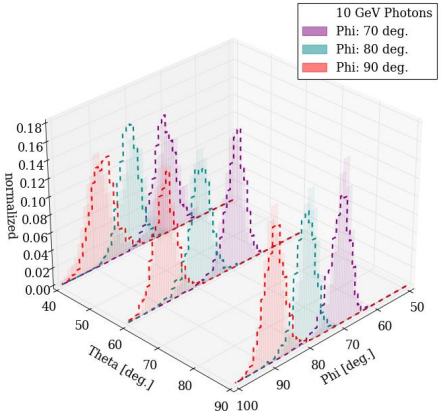
- LHCb Lamarr Gaussino → key4hep.
- Test G4 photon fast sim on ILD with Bounded-Information-Bottleneck Auto Encoder.

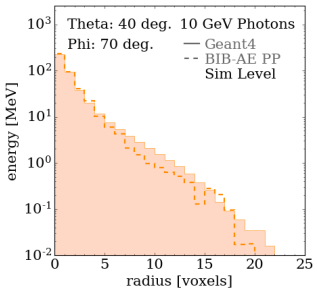
DESY

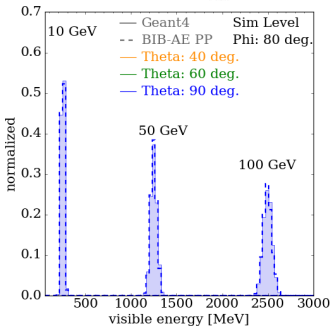
Adding Another Angle

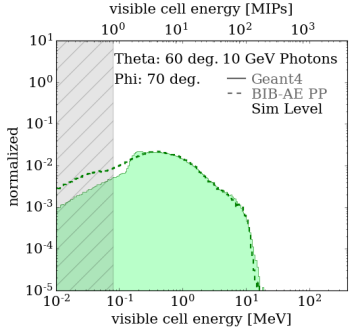
- Need to condition on **energy, theta and phi** for full application
- **Extending phase space** can be challenging









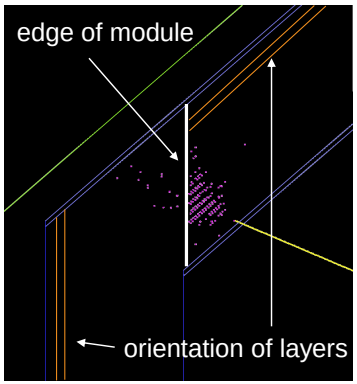


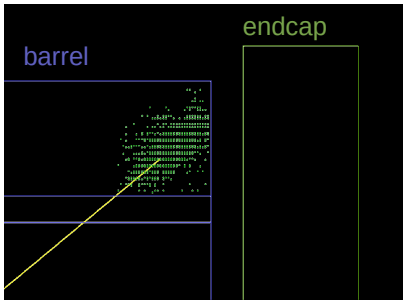
fastsim for Highly Granular Calorimeter (II)

DESY

BIB-AE Integration Into Realistic Geometry

- **BIB-AE** model with full conditioning now **integrated** into ILD detector simulation chain
- **Exclude regions** of detector where model cannot be applied to geometry
  - **Corners** of octagonal barrel
  - **Transition** between barrel and endcap





fastsim for Highly Granular Calorimeter (III)

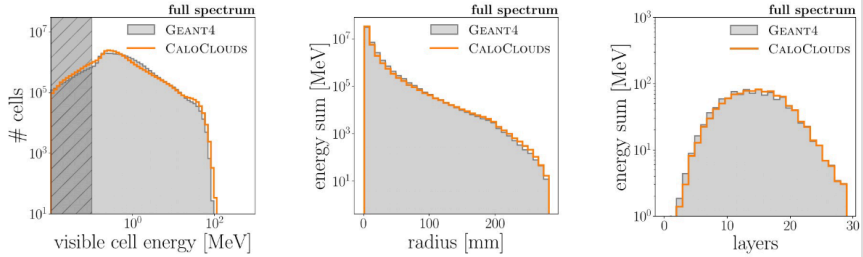
Simulating photons in ILD

DESY

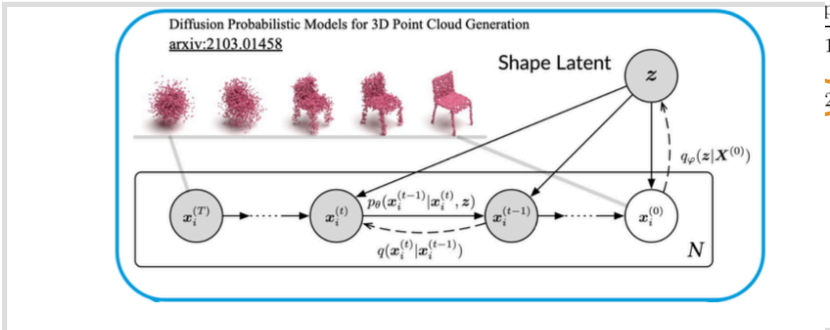
P.McKeown

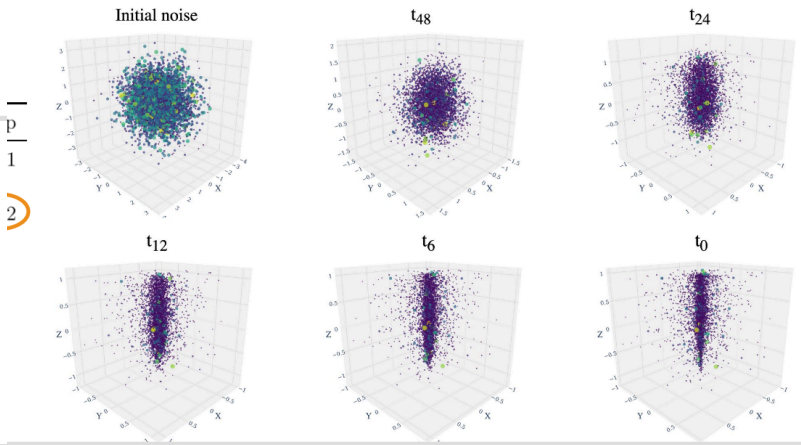
Update 3: CaloClouds 2

- Previously developed point cloud diffusion model on clustered Geant4 steps for photons
  - More granular than readout geometry
  - Achieve high degree of geometry independence



Difusion Point Clouds





Simulating photons in ILD

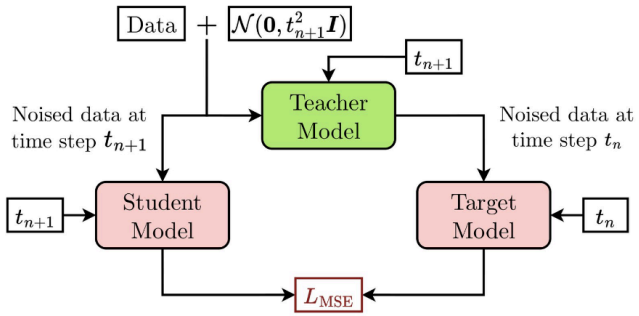
DESY

AIDA innova

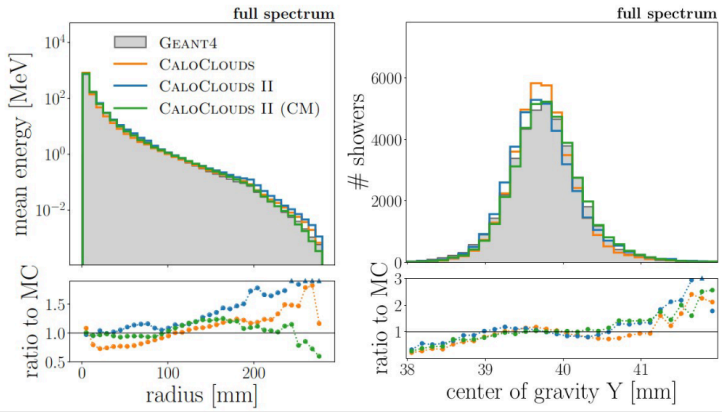
P.McKeown

Update 3 (continued)

- Now successfully applied consistency distillation for significant speed up
- Additional calibrations for energy per layer and CoG in X and Y
- Paper out soon!



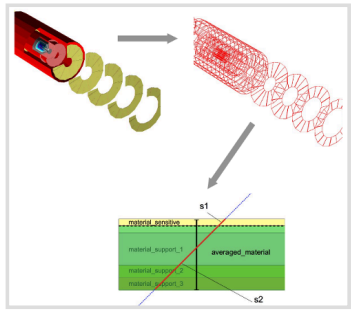
Hardware	Simulator	NFE	Batch Size	Time / Shower [ms]	Speed-up
CPU	GEANT4			3914.80 ± 74.09	×1
	CALOCLOUDS	100	1	3146.71 ± 31.66	×1.2
	CALOCLOUDS II	25	1	651.68 ± 4.21	×6.0
	CALOCLOUDS II (CM)	1	1	84.35 ± 0.22	×46
GPU	CALOCLOUDS	100	64	24.91 ± 0.72	×157
	CALOCLOUDS II	25	64	6.12 ± 0.13	×640
	CALOCLOUDS II (CM)	1	64	2.09 ± 0.13	×1873



Task 12.4 Track Reconstruction

Main work in track reconstruction task is framed inside the ACTS project

- A Common Tracking Software
- Project was spawned from ATLAS's tracking code
- Make this state-of-the-art track reconstruction experiment independent
  - Significant technical challenges!



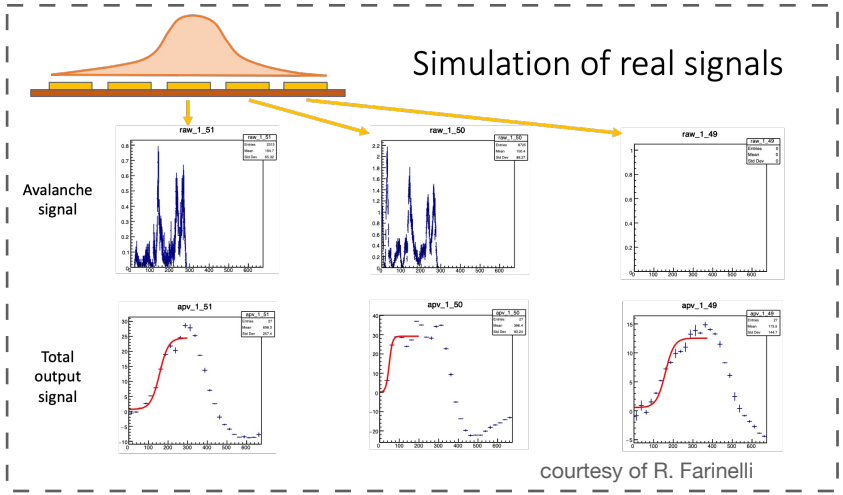
AIDA innova

Developed at IJClab.

tracking with MGPD

Aidainnova 4-year program

1. simulation of the μ-RWELL resistive layer
2. use of Machine Learning for cluster selection and track finding
3. track cleaning and refinement
4. application to IDEA framework



Task 2 started.

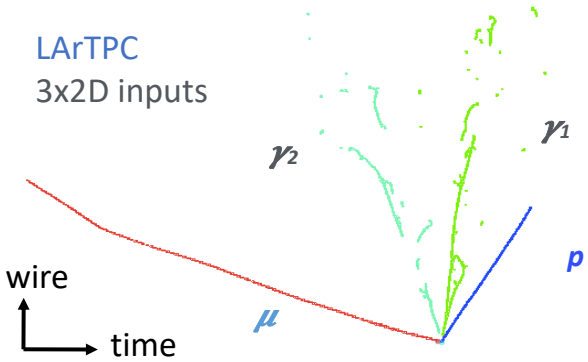
Pandora Software Development Kit

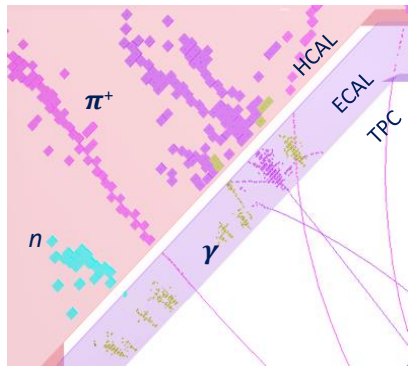
https://github.com/PandoraPFA

A single clustering approach is unlikely to work for complex event topologies:

- Mix of track-like & shower-like clusters

- Use **multi-algorithm** approach using the **Pandora SDK** to **build up events** gradually:
  - Each step is **incremental** - aim not to make mistakes (undoing mistakes is hard)
  - Deploy **more** sophisticated **algorithms** as picture of **event develops**
  - Algorithms: can use machine-learning methods & detector physics knowledge





ILC/CLIC
3D inputs

3

- PandoraSDK used for PFA for ILD, SDHCAL, DUNE LArTPC, IDEA, LAr ECAL (ALLEGRO).
- ML in PFA : used in LArTPC. DeepNN for PFA in IDEA not enough, need physics guidance.

Outlook

AIDAinnova software developement :

- a key item for future collider physics.
- covers :
  - experimental data format,
  - analysis framework,
  - (fast) simulations,
  - track recontruction,
  - calorimeter reconstruction (PFA).
- Deliverables and milestones are on time.