IDEA Simulation status and ongoing Software Activities

P. Azzi (INFN PD), **N. De Filippis (INFN BA+POLIBA)**
Responsibles of "Physics and Software" Group activities in RD-FCC (Italy)

DISCLAIMER
Summary report of many ongoing activities, no time to go in details. Set of links to talks and presentations given.
Tried to give latest updates but lots of work ongoing: please comment if your latest plot is not included!
Big thanks to all the authors we stole our slides from!

Meeting December 17 2024

Full Simulation for FCC:

The FCC Software fully adopts Key4hep;

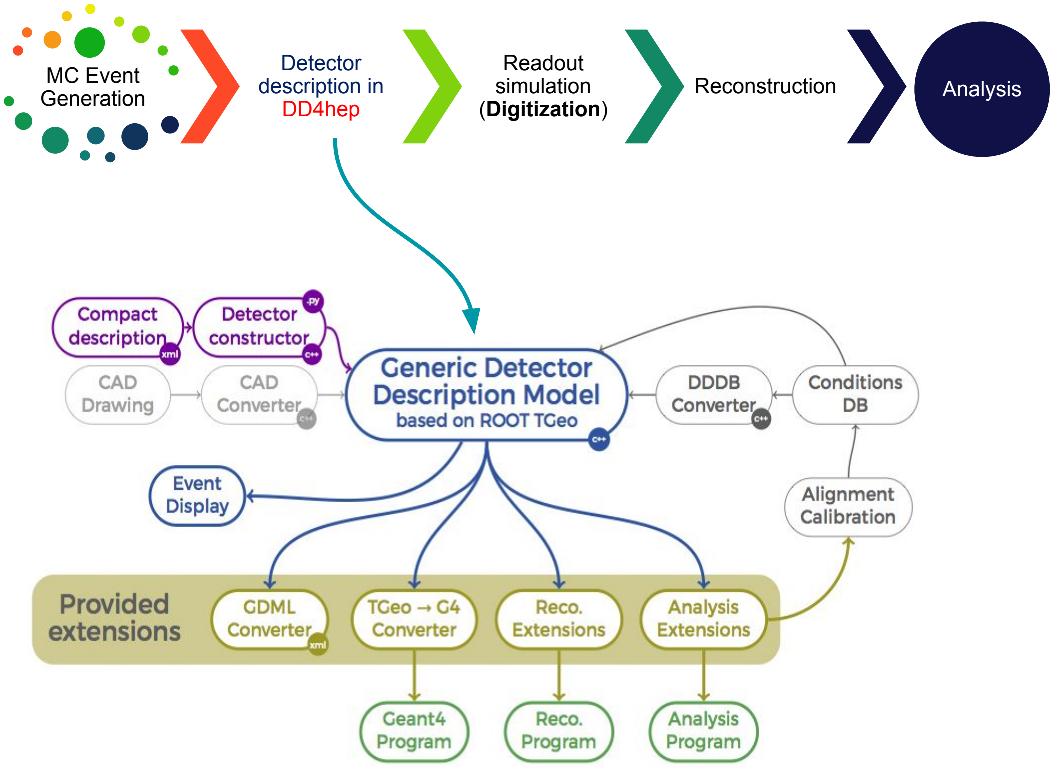
Key4hep: Complete data processing framework, from generation to data analysis

- Data format is EDM4hep.
- Detector description is DD4hep.
- Algorithm orchestration done by Gaudi.

We are going to discuss:

1. Detector description.
2. Digitization.
3. Reconstruction (if available).

For each IDEA sub-detector.



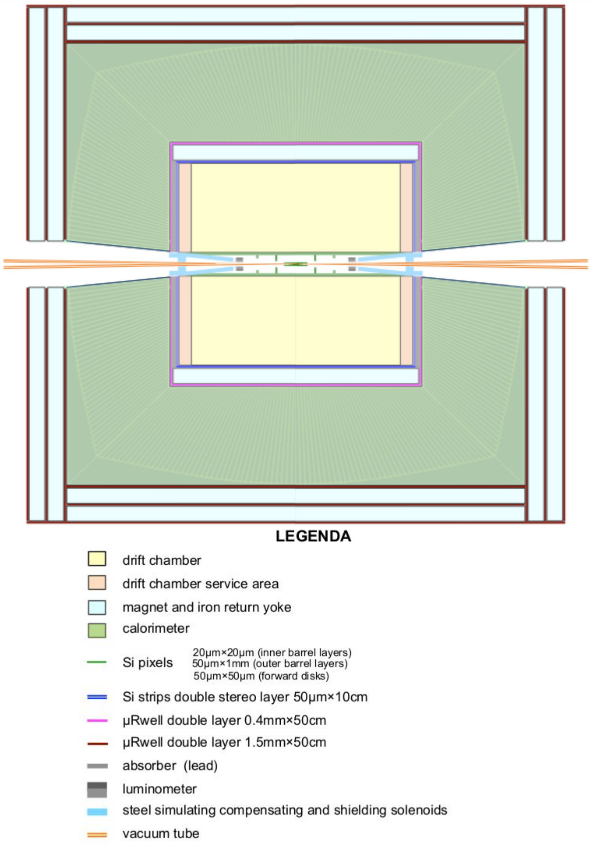
Innovative Detector for *e*+ *e*− Accelerator (IDEA)

IDEA detector concept consists of:
- Silicon pixel vertex detector
- Large-volume extremely light drift wire chamber
- "Wrapper" of Silicon micro-strip detectors (maybe TOF?)
- Dual readout crystal calorimeter.
- Thin low-mass superconducting solenoid coil.
- (Pre-shower detector based on *μRWELL* technology)
- Dual readout fiber calorimeter.
- Muon chambers based on *μRWELL* technology inside the magnet return yoke.

Lots of work done for a coherent description without overlaps

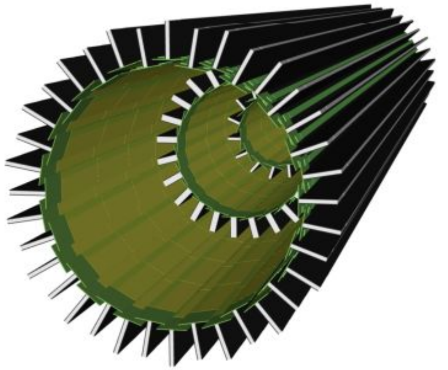
Mahmoud Ali, Brieuc Francois

https://indico.cern.ch/event/1471173/contributions/6201740/attachments/2957269/5200434/IDEA%20Sub-Detector%20Dimensions.pdf

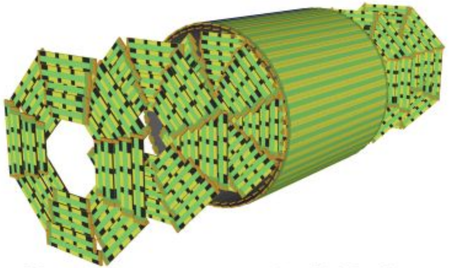


IDEA detector layout

IDEA vertex detector in DD4hep



Inner vertex barrel in DD4hep



Complete vertex in DD4hep

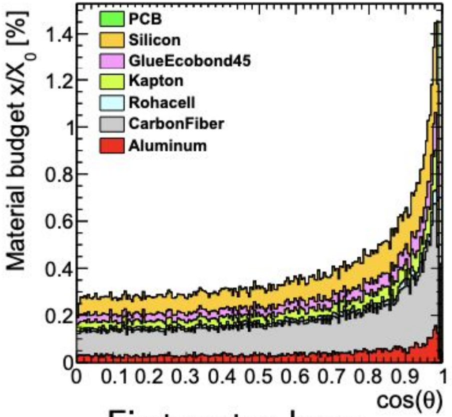
- Accurate description of engineered vertex detector
- Taking into account on-detector services and supports
- Realistic material budget evaluation

→ Compatible with CDR assumption

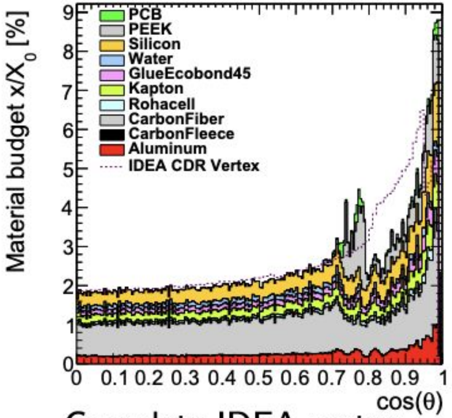
→ $\approx 0.25\%\ x/X_0$ at $\cos(\theta) = 90\,\text{deg}$ for first layer

- Correct description of sensor peripheries

→ Allows for more realistic vertex performance estimation than CLD vertex or previous fast simulation studies (Delphes)



First vertex layer

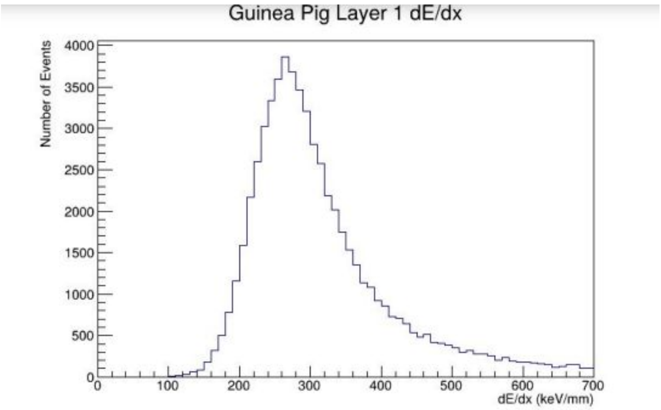


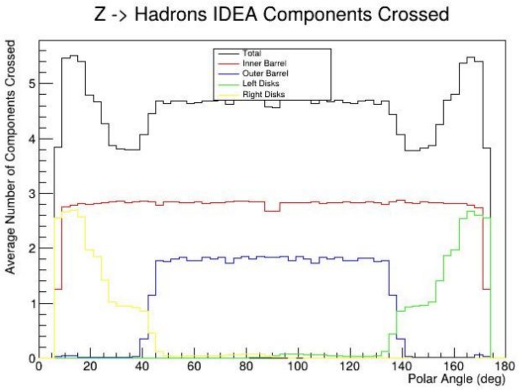
Complete IDEA vertex

Vertex detector DIGITIZATION & Background studies

N. Martinez + MIT group

- A simple digitization of Si-hits is ready in **k4RecTracker** repository as a Gaudi algorithm.
- Applying Gaussian smearing of the hits.
- The same algorithm is applicable for Si-Wrapper and CLD silicon layers.
- The implementation of a detailed digitizer (including charge sharing) is ongoing.
- Very important studies have started**
  - on the effect of background from incoherent pairs from GUINEAPIG
  - on acceptance uniformity
  - hit distribution and occupancy
- *ESSENTIAL FOR DETECTOR PERFORMANCE*





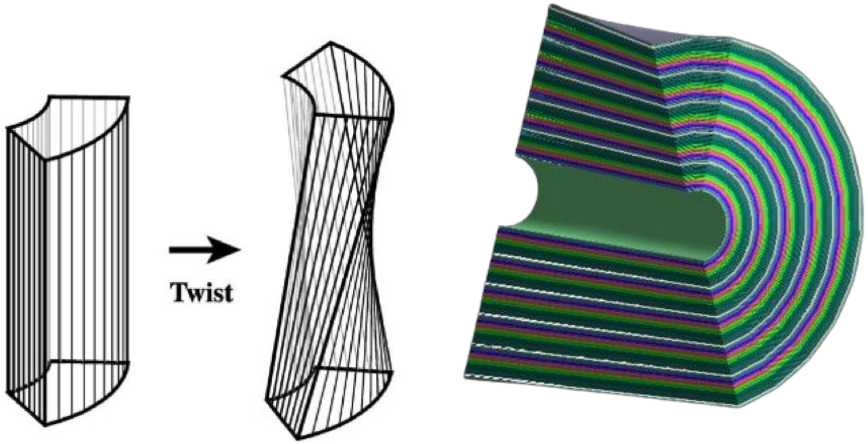
**https://indico.cern.ch/event/1463349/contributions/6161464/attachments/2949268/5183648/Nate_oct16.pdf

Drift Chamber New Geometry description

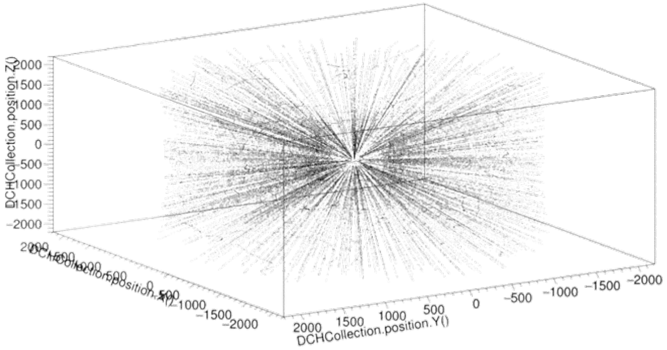
Alvaro Tolosa Delgado

Large-volume extremely light drift wire chamber Evolving from the detectors built for KLOE and MEG2 experiments: is a full-stereo unique volume, co-axial with the 2T solenoid field, with high granularity, low mass and short drift path.

- Cylindrical wall made of carbon fiber
- Cylindrical volume filled of gas mixture.
- 112 hyperboloidal layers filled with gas mix.
- Cells are twisted tubes (twisted tube results from layer segmentation in phi, keeping the twist angle), made of gas mix. These cells are the sensitive volumes!
- Field (x5) and sense (x1) wires inside each cell.
- The new version (v02) is in [k4geo](k4geo).



DriftChamber o1_v02 Twisted tubes.



Drift Chamber New DIGITIZER

Alvaro Tolosa Delgado

- New Drift chamber digitizer, PR27 in k4Rectracker [link], which handles:
  - Hit position projection and smearing.
    - Each simulated hit is transformed into a digitized hit. The digitized hit position is the projection of the simulated hit position onto the sense wire (at the center of the cell)
    - Smearing of the digitized hit position along the wire and radially is done according to the input parameter values
    - Debug histograms are created if `create_debug_histograms` option is enabled
    - It requires that the cellID contain the layer and number of cell within the layer (nphi). It does not matter if the segmentation comes from geometrical segmentation by using twisted tubes and hyperboloids (and the cellID is created out of volume IDs), or the segmentation is virtual DD4hep segmentation
    - Data extension has now the functionalities to calculate the required quantities
  - Adds dN/dx information: number of clusters and their size, which are derived from precalculated distributions contained in an input file specified by the parameter `fileDataAlg`. The method and distributions corresponds to the option 3 described in F. Cuna et al, arXiv:2105.07064
- New digitized hit class is used as an EDM4hep data extension, to be integrated into EDM4hep
- Random number generator uses the seeds calculated on an event basis by the UID service, from the podio header information (run/event number)

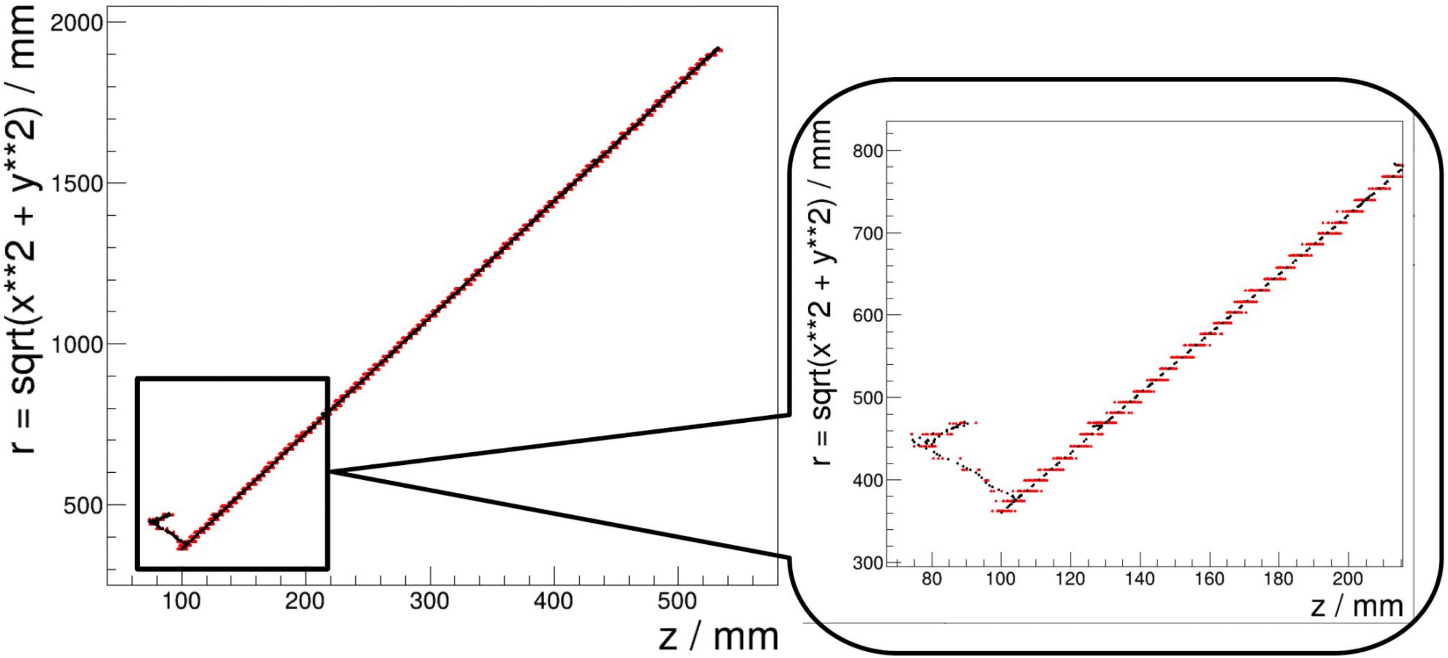
Significant developments also extending/modifying the Edm4HEP DATA Member.
VALIDATION NEEDED & checks with real data.
In particular for the dN/dx debugging.

https://indico.cern.ch/event/1481286/contributions/6247930/attachments/2975421/5237707/dch_digi_atd.pdf

An updated digitizer will be also discussed by N. De Filippis at the FCC Physics week on January 16

DCH DIGITIZER: Position projection and smearing

- Example of proton at 10 GeV, direction (3,2,1), 10 events, seconday production range cut 0mm, maxstep 1cm
- Black dots correspond to Geant4 energy deposit positions, red dots to the same positions projected onto the closest wire and smeared along and perpendicularly to it



Alvaro Tolosa Delgado

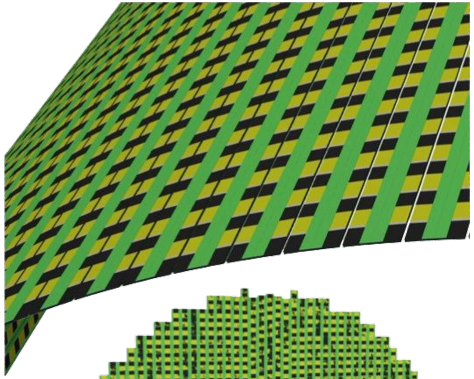
Armin Ilg

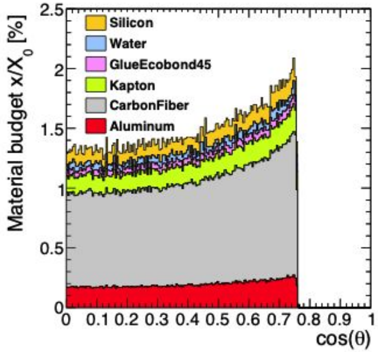
Silicon Wrapper

- Being further away from the interaction point, the level of details needed to get accurate simulations is probably lower than for the vertex detector where we need the great details.
- By using the same detector builder and digitizer as for the vertex detector, a first version od Si-Wrapper is ready.
- Large surface (112 m^2), tiled with ~ 4x4 cm^2 modules.
- A huge number of modules □ Slow and large memory consumption.
- A second version which is lighter (memory consumption-wise) is implemented. For the moment a single-layer of 0.050 x 1 mm strips.

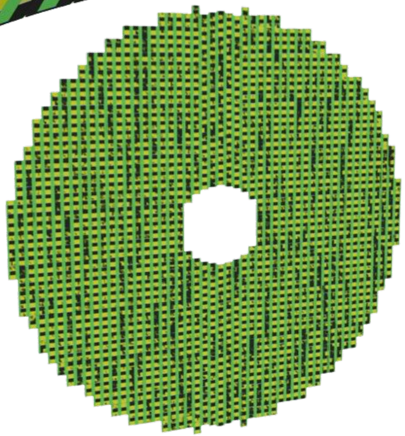
Two layers of silicon micro-strip detectors or One hermetic layer of pixels ?

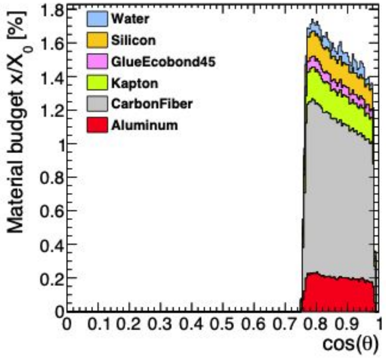
Barrel





Endcap





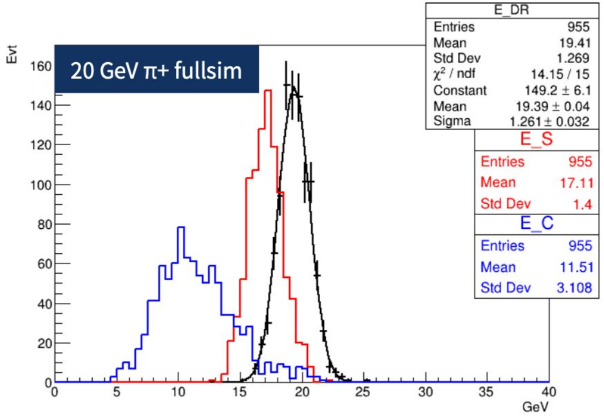
Fiber sampling dual-readout calorimeter

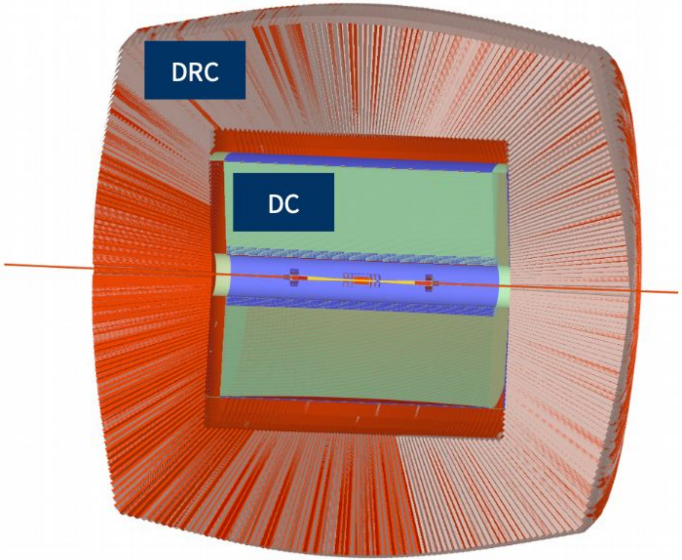
Version 1: SangWhun Ko, SungWon Kim

The full simulation of fiber sampling DR-Calo has been implemented, and a PR has been opened for k4geo:

Successfully demonstrated the principle of DR-Calo with full simulation

- A dedicated SiPM emulation library ("SimSiPM") has been developed:
  - Able to simulate the output waveform of SiPM based on parameterized inputs from the datasheet (dark counts, crosstalk, afterpulse, saturation, noise, ...)
- Using fast sim for photon transportation in fiber.
- Some events produced to study clustering



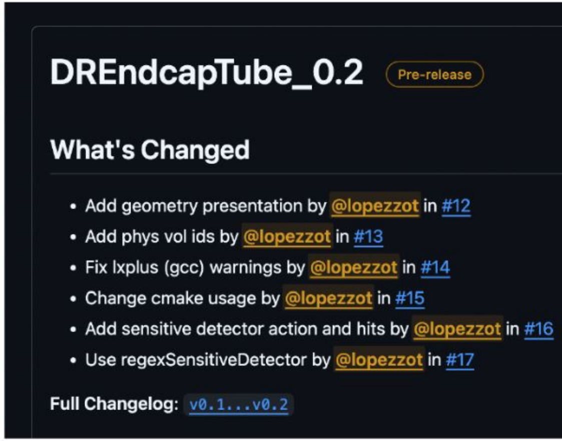


Fiber sampling dual-readout calorimeter :

Version 2:

L. Pezzotti

- The next goal was to develop a DD4hep sensitive (detector) action (`Geant4SensitiveAction<>`) for this detector which:
  - Allows to simulate events at a single-threaded rate of $O(1) - O(10)$ s/evt at all the FCCee energy poles
  - Reduces the (too) large memory footprint problem
  - Is validated against experimental test-beam data
- A solution is proposed for the endcap simulation
  - available as v0.2 of DREndcapTubes
  - the same solution is applicable to be the barrel simulation by Andreas



https://indico.cern.ch/event/1457476/contributions/6136107/attachments/2929865/5144647/lopezzot_fccsim_1892024.pdf

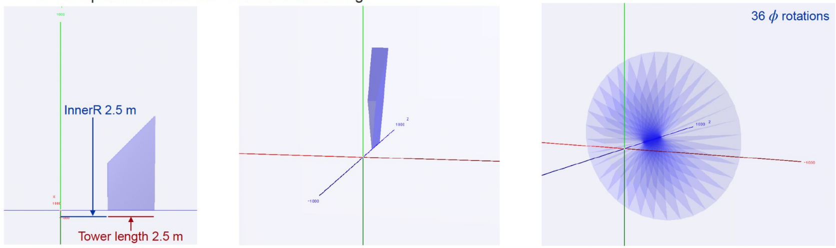
Fiber sampling dual-readout calorimeter :

L. Pezzotti

Version 2:

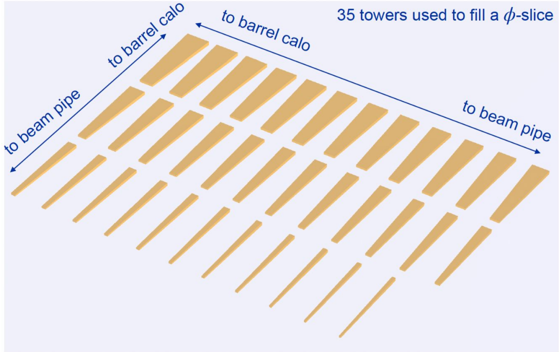
(Recap) ϕ-slice

- A single ϕ-slice is a DD4hep EightPointSolid (equivalent to root Arb8, or Geant4 G4GenericTrap)
  - It is build according to the calorimeter inner radius (2.5 m), tower length (2.5 m), and ϕ-unit (2*π/36)
  - Assumes that the barrel and endcap region will touch at π/4
  - Is replicated around the z-axis to full coverage

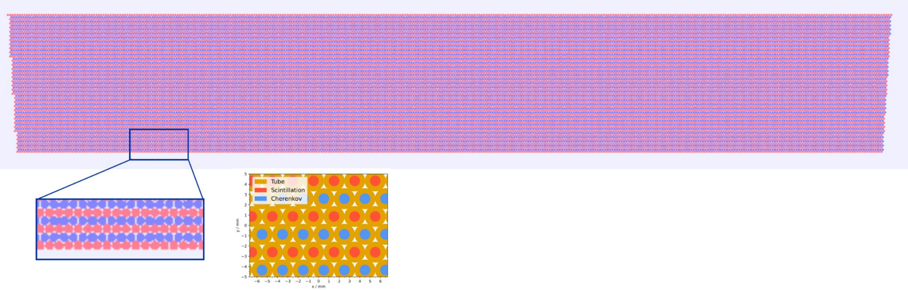


(Recap) Towers

- Towers are Trap (equivalent to Geant4 G4Trap) that define a θ-region inside the ϕ-slice. The center of each tower points to the IP.
- 40 towers are considered, each covering a region of $\Delta\theta = \dfrac{\pi/4}{40}$
- As towers must fit inside the ϕ-slice their dimension changes with θ

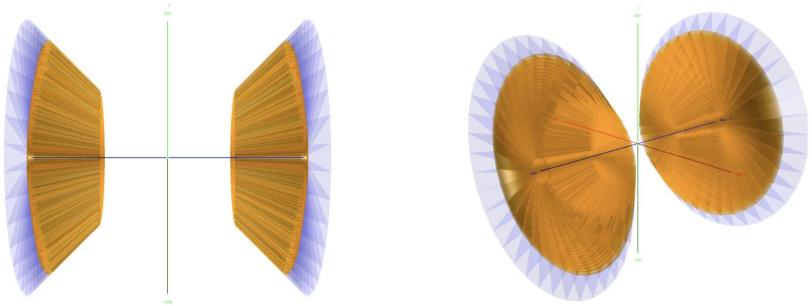


(Recap) Filling the towers



- Tubes are 1-mm-thick radius Tubs (equivalent to G4Tubs) and house Scintillating or Cherenkov (clear) optical fibers. In the following optical fibers inside tubes are not displayed to aid visibility.
- Tubes z-axis is always parallel to the tower axis, i.e. they are projective pointing to the IP
- Tubes are placed starting from the back face of the tower (the biggest one)

- The final geometry is obtained with a simple repetition of the ϕ-slice around the z-axis (right endcap, $z > 0$), and a reflection + translation for the left endcap ($z < 0$)



Fiber sampling dual-readout calorimeter :

L. Pezzotti

Version 2:

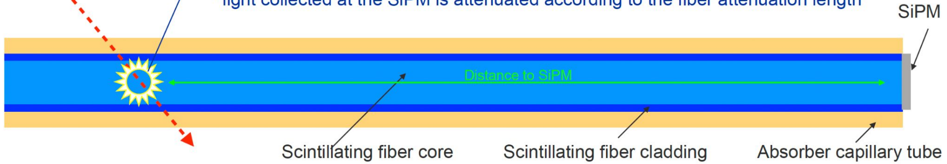
Signal treatment

- This calorimeter simulation must include two signals:
  Scintillation light in scintillator-doped optical fibers and *Cherenkov light* in clear optical fibers

- Simulating light in calorimeters has always been a problem:
  as of today no LHC Experiment simulation include light propagation in calorimeters, instead it is parametrized based on experimental inputs

Scintillation signal simulation

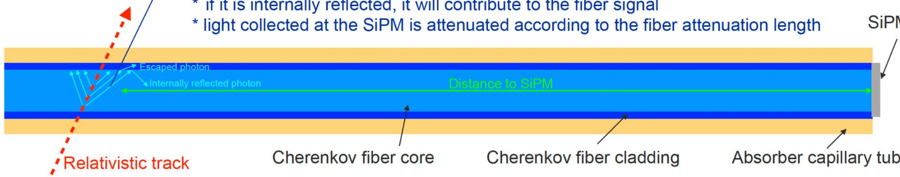
* optical photons are produced *isotropically* around the step position → no need to take into consideration their direction in the Monte Carlo
* the amount of photons produced is related to the energy deposited via the Birks Law
* on a step-by-step basis it fluctuates according to Poissonian statistics
* light collected at the SiPM is attenuated according to the fiber attenuation length



- This approach was validated against test-beam data from 2021 finding a good MC-to-data agreement [Article]
  - It will be refined using new results from 2023 and 2024 test beams which are being analyzed

Cherenkov signal simulation

* optical photons direction is calculated by the Geant4 Cherenkov process
→ the emission of the optical photon must be included in the Monte Carlo
* once reached the fiber cladding, photons can be internally reflected or may escape the fibe
according to the material refractive index and the photon direction
→ we check with Geant4 if the photon undergoes an internal reflection
* if it is internally reflected, it will contribute to the fiber signal
* light collected at the SiPM is attenuated according to the fiber attenuation length



- This approach was validated against test-beam data from 2021 finding a good MC-to-data agreement [Article]
  - It will be refined using new results from 2023 and 2024 test beams which are being analyzed

Fiber sampling dual-readout calo. VERSION 2: Digitization

L. Pezzotti

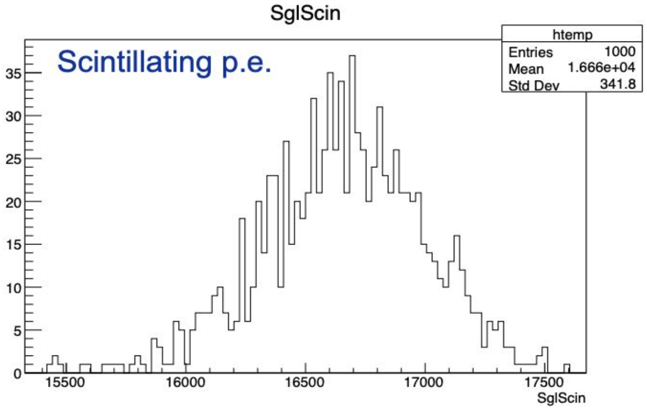
- An edm4hep hit collection is created per each endcap (right and left) and per each fiber type (Scintillation and Cherenkov)
- A hit is created per each fiber with a signal (photo-electrons) above 0, i.e. applying a zero-suppression → the hit collection size represents the number of fibers with a signal in the event

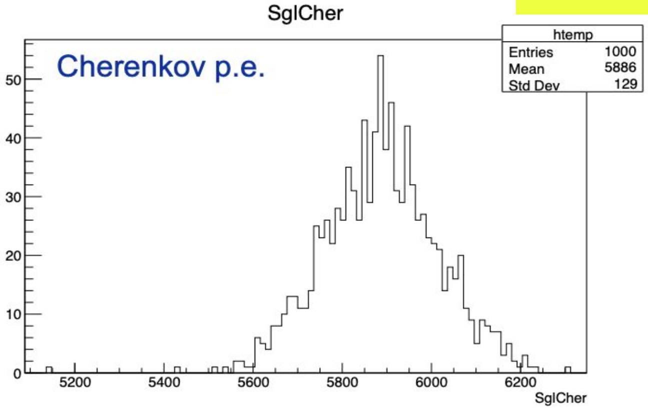
Number of fibers readout in each event @90 GeV e^-
~1300 Scintillating fibers and ~800 Cherenkov fibers

- The simulated signal in terms of photo-electrons is set to ~180 Scintillating p.e./GeV and ~66 Cherenkov p.e./GeV
- It will be re-tuned according to new experimental findings as they come from test-beams

Can study event rates and tune with TB data

Total signal in photo-electrons @90 GeV e^-
~180 Scintillating p.e./GeV and ~66 Cherenkov p.e./GeV



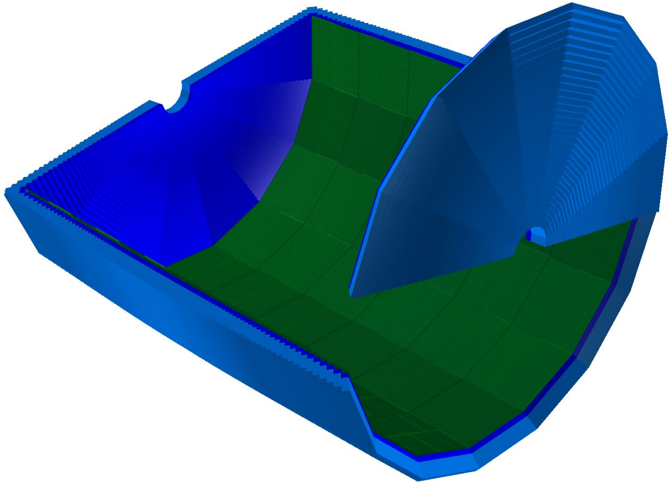


https://indico.cern.ch/event/1457476/contributions/6136107/attachments/2929865/5144647/lopezzot_fccsim_1892024.pdf

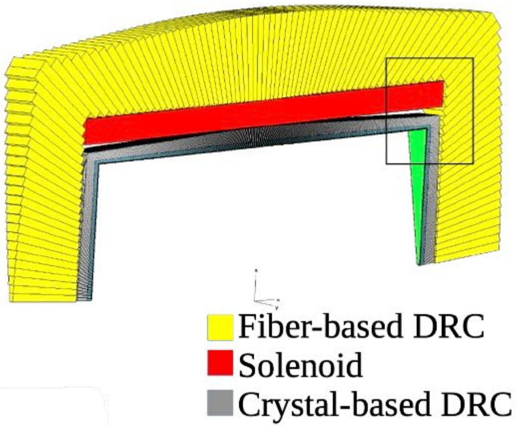
M. Lucchini, W. Chung et al.

Crystal-based dual-readout calorimeter

- Baseline IDEA now contains dual readout crystal EM.
- $PbWO_4$ crystals + LYSO timing layer.
- With 1x1cm crystal faces/thickness:
  - ~1.12 million barrel crystals
  - ~400,000 endcap crystals
  - ~30,000 timing crystals



DD4hep implementation of Crystal DR-Calo

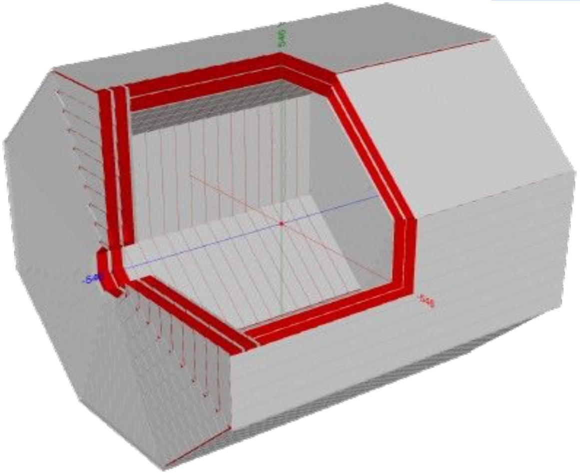


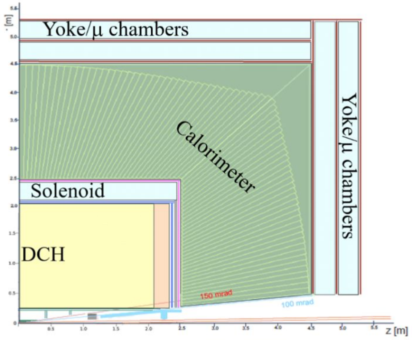
Muon system based on *μRWELL* technology

M. Ali

IDEA muon system primarily composed of 3 sensitive stations. Each station will consist of a large mosaic of μRWELL detectors.

- The basic μRWELL ”tile” will have an active area of 50 × 50 cm^2.
- The layers are placed between layers of the iron yoke that closes the magnetic field.
- A strip pitch ~ 1.2 mm and 500 mm length.
- A 2D readout system for each individual chamber.





- A simple digitization algorithm is ready, which smears the hit position in the local μRWELL chamber plan in 2D, with the space resolution of the chamber ~ 400 μm, and more features to be added (simulates the efficiency, fake rate (noise)).
- Currently working on reconstruction (Standalone muon system alg.).

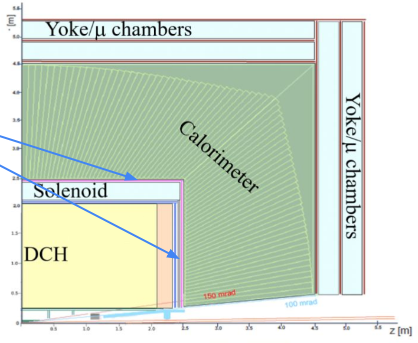
M. Ali

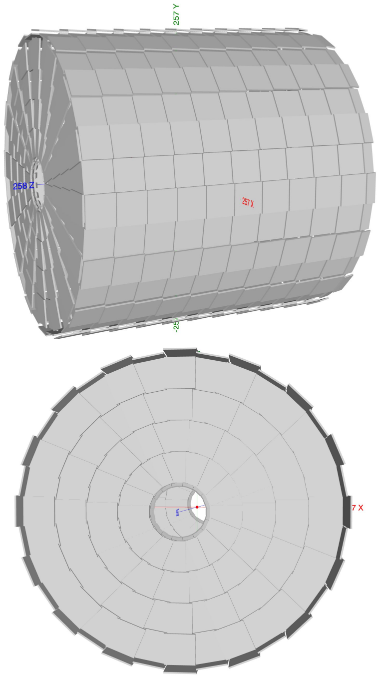
Preshower based on *μRWELL* technology

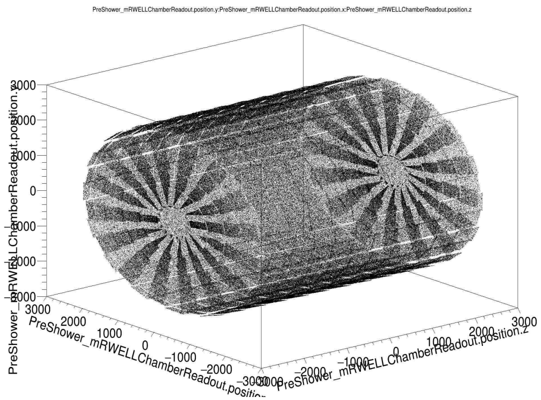
IDEA detector envisages one layer preshower system utilizing μRWELL technology. Both the preshower and muon systems have a modular design, with both sharing the **same builder** file.

- Pitch between readout strips: 400 μm
- A 2D readout system for each individual chamber.
- The implementation is ready in Dd4hep, and a PR is opened.
- It uses the same digitization of the muon system.
- A reconstruction alg. To be implemented.

One layer encapsulate the solenoid



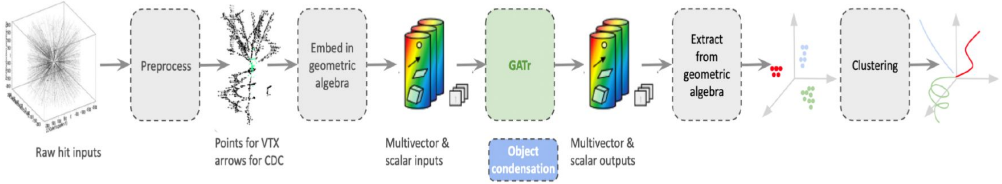




Tracking: Pattern recognition with GNN

D. Garcia, A. De Vita et al.

Generalised Geometric Track Finding algorithm



An end-to-end pipeline can be employed to process hits from all tracking systems, producing a complete set of tracks and **evaluating tracking performance**.

The analyses focus on the IDEA detector, which utilises a **vertex detector and a drift chamber**.

ML step: The Track-finding approach is based on a graph structure of the inputs, where geometric algebra transformations are applied. The result is a set of pairs (β, coordinates) in the embedding space.

Clustering step: Tracks can be identified in the embedding space by applying a clustering algorithm, establishing a one-to-one correspondence between clusters and tracks.

Track Finder - ML step

1. **Preprocessing** → Each set of hits X = {X_{vxt},X_{cdh}} has to be converted into a **graph of multivectors**, where each node corresponds to an hit.
2. **Geometric Algebra Transformer (GATr)** → Each multivector graph is transformed with **GATr layers**.
3. **Object Condensation** → Each node of the transformed graph is mapped into an **embedding space coordinate with a corresponding scalar** (β).

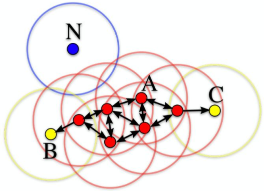
Track Finder - Clustering

Eleobert, DBSCAN - GitHub Repository

The clustering algorithm is implemented with **DBSCAN**.

DBSCAN uses a definition of clusters based on the notion of **density**:

- if a point has a minimum number of points within a certain epsilon distance (ϵ), it is classified as a **core point**;
- If a point is not a core point and it is not close to a core point, then it is classified as **noise**;



Starting with the core points, the clusters are expanded until all points are classified as noise or belonging to a cluster.

Tracking: Pattern recognition with GNN

D. Garcia, A. De Vita et al.

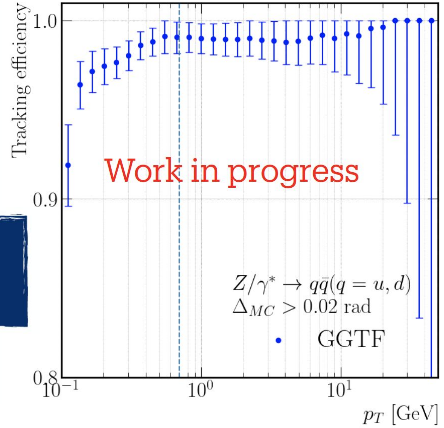
A particle is defined as **reconstructable** if it satisfies the following conditions:

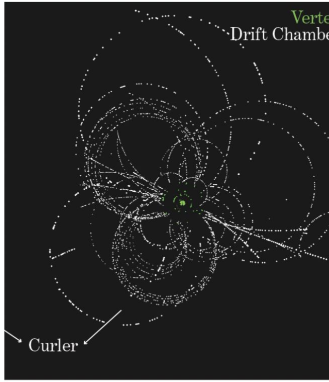
1. $p_T > 100$ MeV
2. $\cos(\theta) < 0.99$
3. Number of unique hits (Drift Chamber + Vertex) > 15
4. Number of Drift Chamber hits > 4
5. Generator Status == 1
6. Vertex < 50 mm

$$\text{Tracking efficiency} = \frac{\text{number of reconstructable and assigned particles}}{\text{number of reconstructable particles}}$$

Results

Performance for complex events - $Z/\gamma^* \to q\bar{q}\ (q = u, d)$





What's next

- **Optimization of the GGTF architecture** aims to enhance performance while reducing computational load.
- A pull request (PR) will be submitted to the central Key4Hep repository, implementing a **complete end-to-end pipeline for the GGTF**.
- Additionally, developing a **physics-based track-fitting** algorithm will be valuable as both an alternative and complementary tool to the ML-based track-finding algorithm.

https://indico.cern.ch/event/1457081/contributions/6166117/attachments/2961403/5208718/Andrea_DeVita_FCCVenice_6thNovember.pdf

Toward Particle Flow Reconstruction

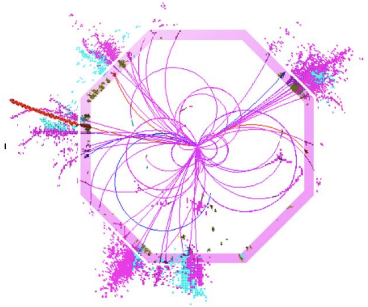
B. Francois, D. Garcia + all

PandoraPFA originally developed for the application for ILCSoft

Used in FCC-ee CLD detector simulation

Multi-algorithm pattern recognition

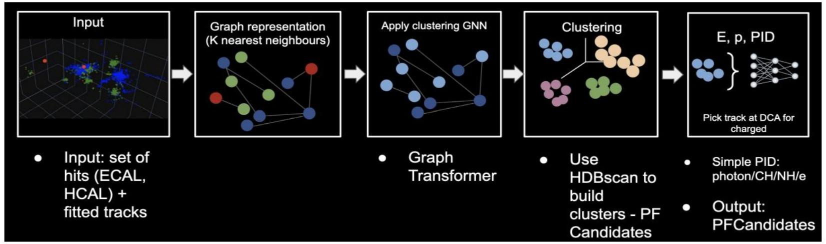
PandoraPFA



https://github.com/PandoraPFA
https://arxiv.org/abs/0907.3577
https://arxiv.org/abs/1506.05348
PandoraPFA algorithm overview

- Figures of merit to optimize detectors should be as close as possible to 'ultimate' solutions
  - Otherwise we will find 'wrong' minima
- Best performance expected from Particle Flow (PFlow) → **we need PFlow for all FCC detector implementations**
- Status: **PandoraPFA is integrated in Key4hep but little local expertise available**
  - Implemented for CLD baseline and used for CLD with LAr ECAL (need extra tuning)
- Possible (non mutually exclusive) paths:

  1) Revive PandoraPFA expertise locally: tune the CLD implementation, apply to other concepts
  2) Implement other classical approaches 'from scratch' (not relying on PandoraPFA)
  3) Pursue the Machine Learning based approach (longer term but promising)



Recent implementation of the PandoraPFA for ALLEGRO

Implementation of PandoraPFA in the ALLEGRO detector simulation has started and progressing well

- Currently focusing on the ECAL and HCAL barrels
- PandoraPFA can identify photons in the ALLEGRO detector very well
- Managed to reconstruct hadronic showers, however, too many clusters are created → needs further investigation
- Tracks for MCParticles are produced to use in the PandoraPFA algorithm
- Managed to reconstruct electrons from tracks and showers in the ECAL

Validation Suite for FCCSW

E. Lupi

https://indico.cern.ch/event/1457081/contributions/6166118/attachments/2961254/5208753/LupiEnrico_FCCSWValidation_FccVenezia.pdf

The main pupose of this project is developing new tools for the validation of the **FCC software**.

What is validation?

Fundamental step in the software development lifecycle

Ensures the final product meets the **user's need** and fulfills its **intended purposes**

And for physics?

Making sure that the physics **results** obtained from the simulations are **compatible with our expectations**

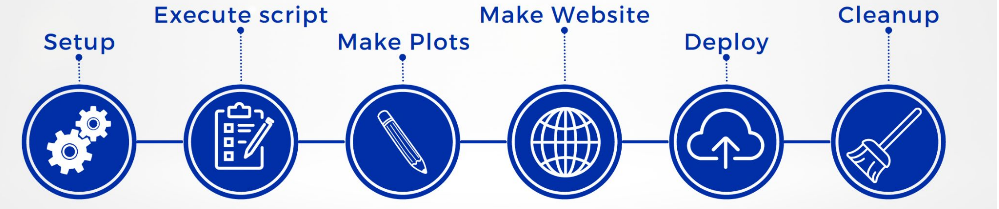
How to achieve it?

GitLab CI/CD pipeline to run daily automated tests

Results comparing the output of new versions of the stack with reference stable ones can be checked on **website**

WARNING!

This pipeline only validates the physics!
The software itself is tested by cron jobs in other repositories



Detector Version	Subsystem	Histograms

ALLEGRO_o1_v03	• Electromagnetic Calorimeter - Barrel	• CaloCluster Energy
		• CaloTopoCluster Energy
		• ECalBarrelModuleThetaMergedPosition:
		  ◦ total Energy per evt
		  ◦ X, Y and Z position

IDEA_o1_v03	• Drift Chamber • Vertex Detector • Vertex Inner Barrel • Vertex Outer Barrel	Number of Hits
CLD_o3_v01	• Standalone ARC Detector	• Photon counts per event • Photon counts vs. θ • Photon counts vs. θ of incoming particle

FUTURE DEVELOPMENTS

ERROR HANDLING SYSTEM:
- Implement fully functional check to handle software failure
- Avoid failure for one detector to affect the validation of others

MORE VALIDATION!
- The pipeline is ready, but there is still a lot of work to do...
- Add more tests and populate the validation website

Conclusions

Lots of activity in the Software:

- basic geometry description of IDEA in k4geo
- digitization work in progress: possible now to start studying more in details, compare with Test Beam data, validate with background, occupancies etc.
- reconstruction work started: pattern recognition with GNN and clustering. Lots to try, also traditional methods to compare.
- Validation suite in progress as well, will become more and more useful.
- Effort ongoing on distributed computing: essential for larger production
- Effort ongoing on the translation of the LEP data: essential for validation and development

LINKS TO RECENT UPDATES ON SIMULATION AND SOFTWARE STUDIES FOR IDEA

- New IDEA GEOMETRY WITH CRYSTAL ECAL

https://indico.cern.ch/event/1471173/contributions/6201740/attachments/2957269/5200434/IDEA%20Sub-Detector%20Dimensions.pdf

- NEW DIGI DRIFT CHAMBER

https://indico.cern.ch/event/1481286/contributions/6247930/attachments/2975421/5237707/dch_digi_atd.pdf

- NEW STUDIES WITH GUINEA PIG

https://indico.cern.ch/event/1463349/contributions/6161464/attachments/2949268/5183648/Nate_oct16.pdf

- NEW SIMULATION OF DR CALO

https://indico.cern.ch/event/1457476/contributions/6136107/attachments/2929865/5144647/lopezzot_fccsim_1892024.pdf

- GNN TRACKING IN IDEA

https://indico.cern.ch/event/1457081/contributions/6166117/attachments/2961403/5208718/Andrea_DeVita_FCCVenice_6thNovember.pdf

- Validation Package E Lupi here → Camila Paris Stagiaire in PAdova adding Tracker/Tracking Histograms from IDEA

https://indico.cern.ch/event/1457081/contributions/6166118/attachments/2961254/5208753/LupiEnrico_FCCSWValidation_FccVenezia.pdf

- Discussion about future needs on reco and computing:

https://indico.cern.ch/event/1457081/contributions/6166120/attachments/2961586/5209075/20241106_Discussion_Next_Step_FullSim_Brieuc_Francois_FCC_Italy_France.pdf