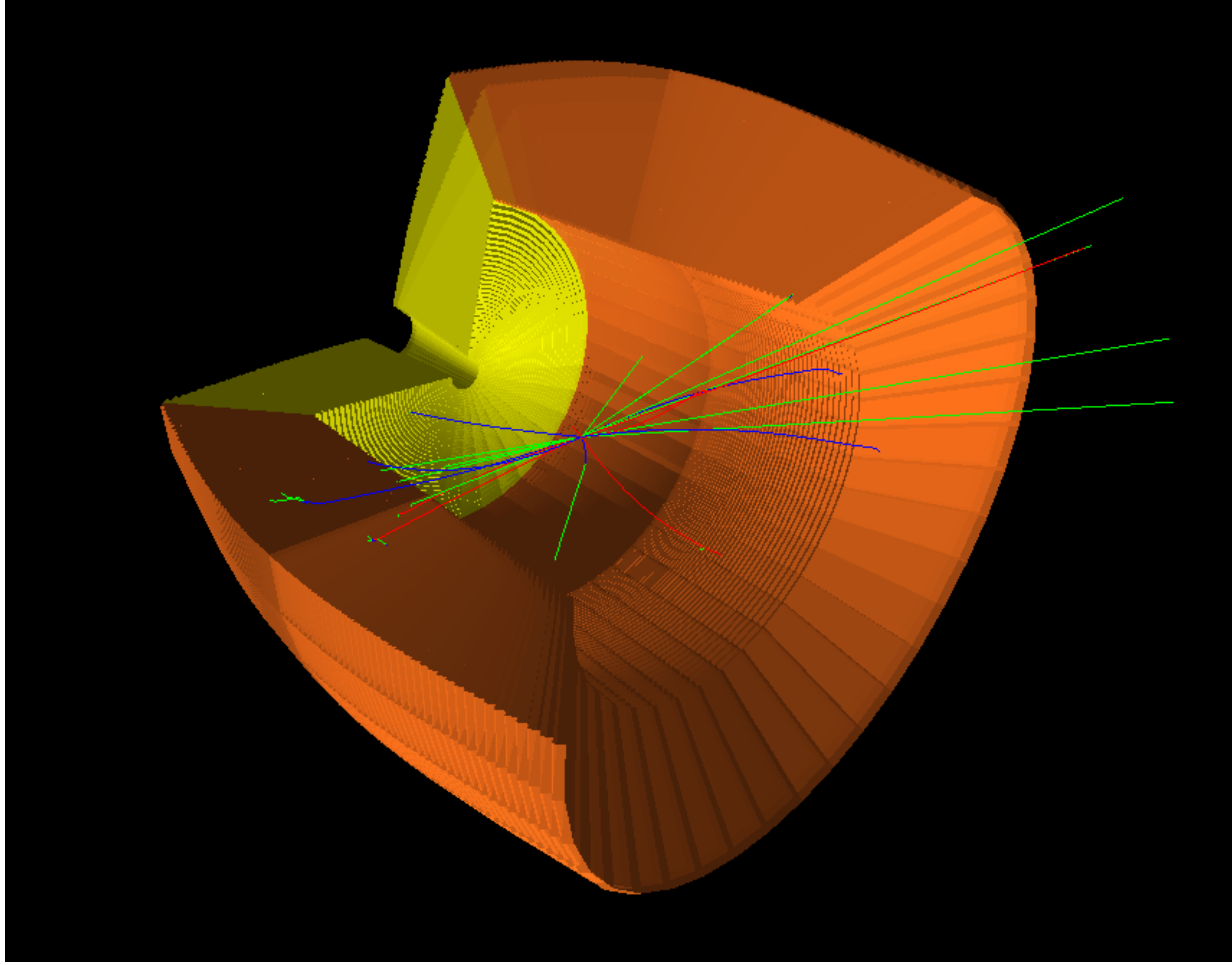
Dual-Readout Calorimeter Simulation



Lorenzo Pezzotti
CERN [EP-SFT]
previously University of Pavia & INFN

on behalf of the IDEA Calo Group

ECFA Higgs Factories: 1st Topical Meeting on Simulation
1-2 February 2022

Why dual-readout calorimetry?

A dual-readout fiber calorimeter at future Higgs factories:

- provides flexible compensation (i.e. regardless the absorber medium, sampling fraction and signal integration time)
- being an unsegmented calorimeter simplifies the calibration procedure (no need to correct for response variation according to the shower age)
- can be coupled to a dual-readout crystal section for extreme EM resolution
- is a granular calorimeter:
  - tunable 2D segmentation according to SiPM signal grouping
  - 3D segmentation could be restored through signal timing analysis
- is sensitive to different features for each event (S/C signals) thus provides richer inputs to any machine learning based estimation (*i.e.* possible to build 2-stacked convolutional neural networks)

Flexible compensation
↓
Calibration at EM scale
↓
Crystal friendly
↓
2D/3D granularity
↓
Gold mine for machine learning

Geant4 Sim, our starting point

To perform proof of concept studies we developed a Geant4-based simulation:

Pythia8

↓

HepMC

↓

GEANT4 → Qt - visualizer

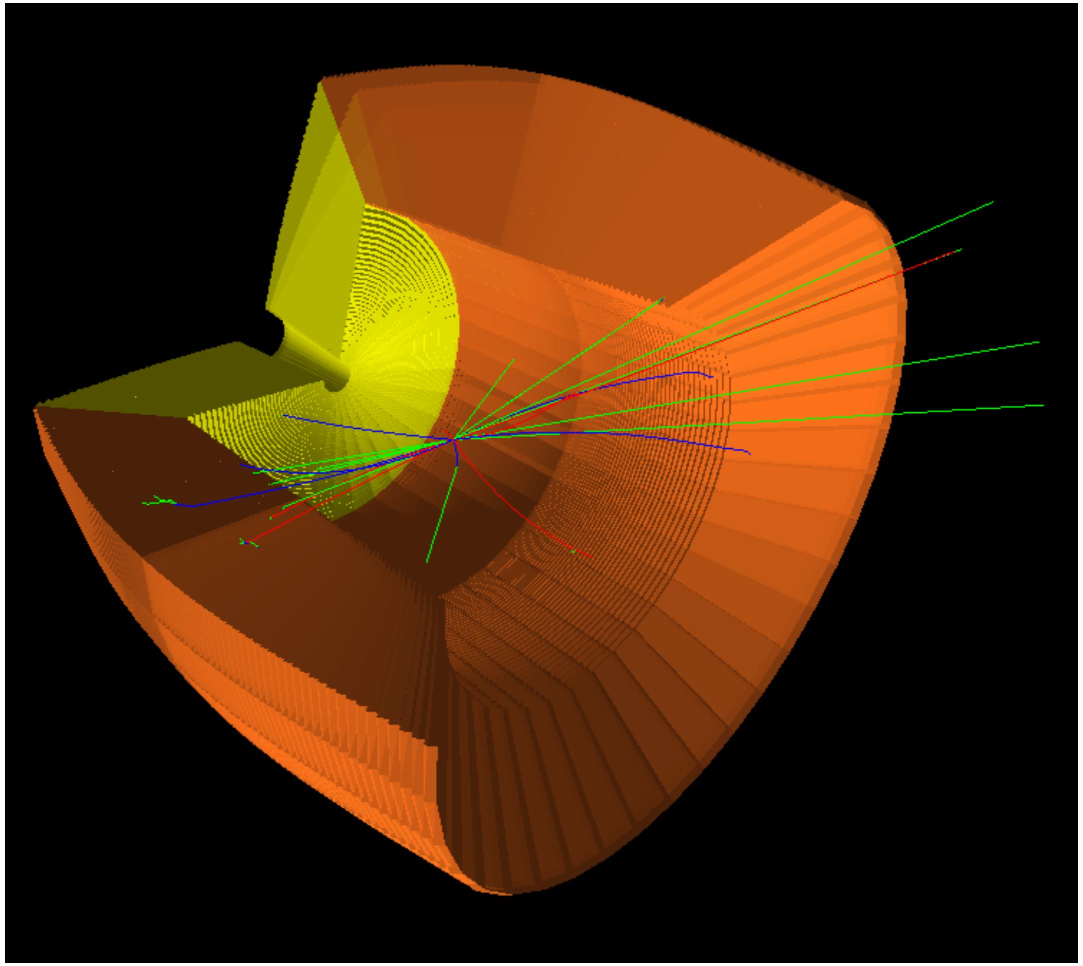
↓

ROOT
Event Data Model

↓

Data analysis

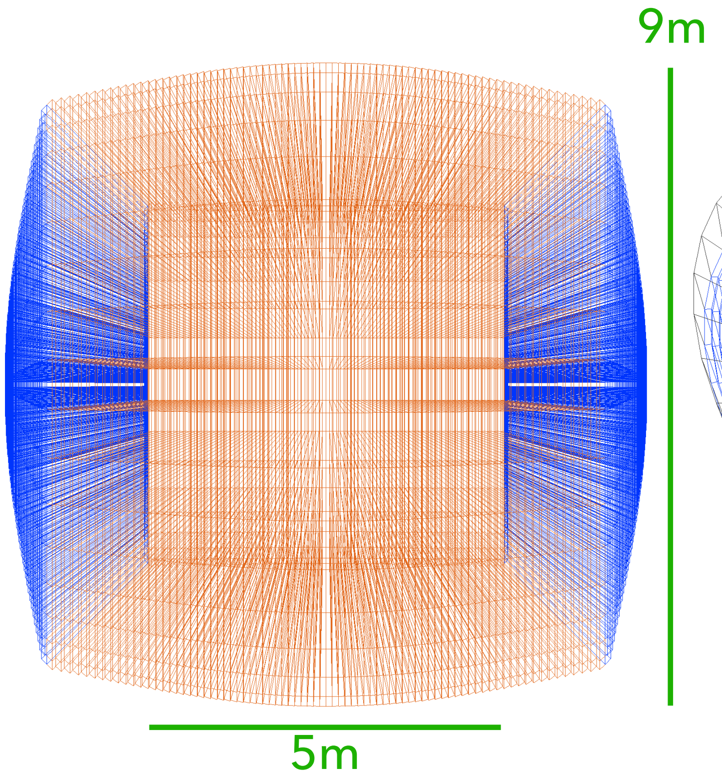
GEANT4 - Qt visualizer - IDEA / $e^+e^- \to jj$

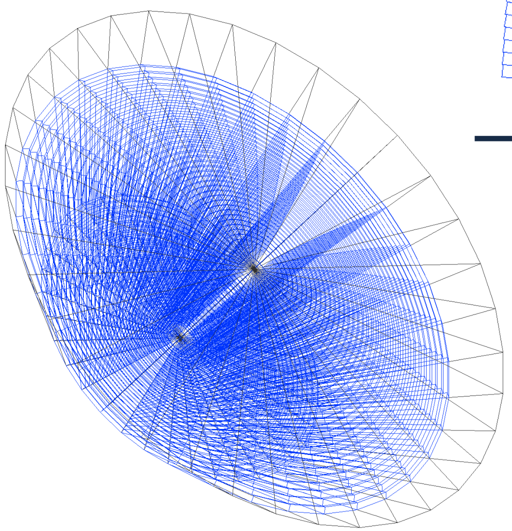


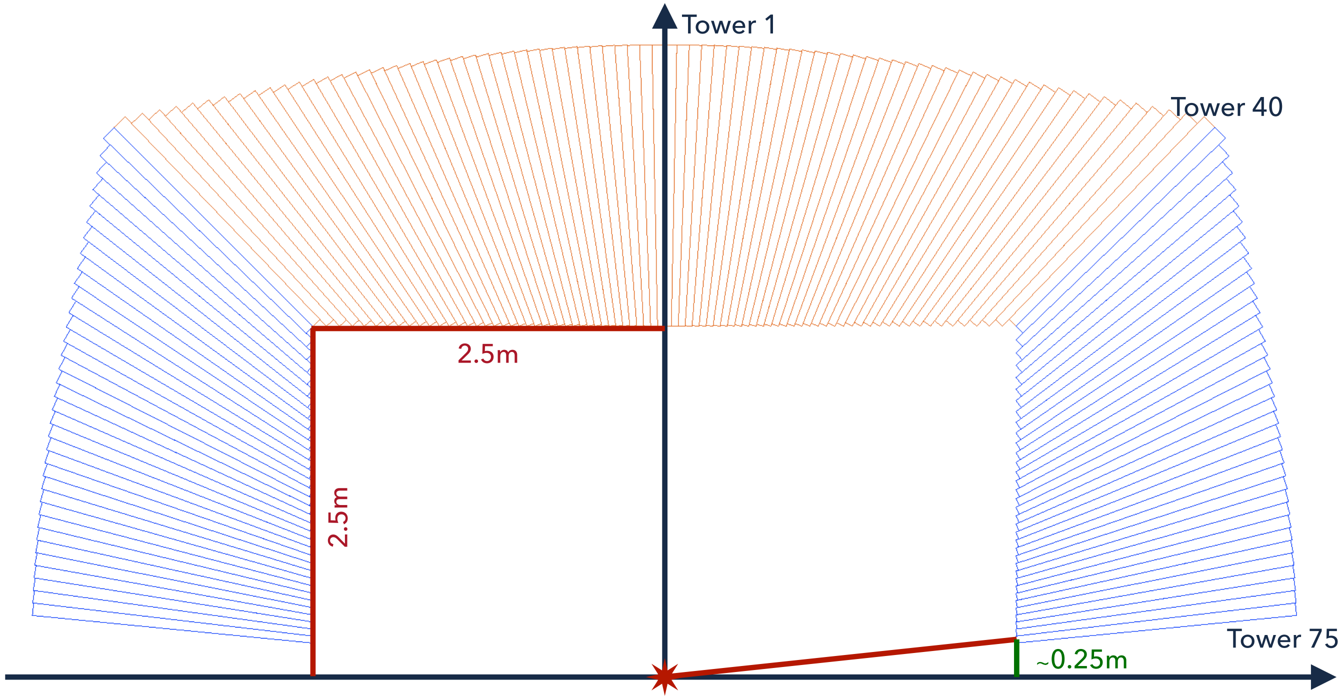
A benchmark geometry

A benchmark IDEA Calo implementation:

- Exploits 5400 towers: $\Delta\vartheta = 1.125°$, $\Delta\phi = 10.0°$
  Theta coverage up to ~0.100 rad







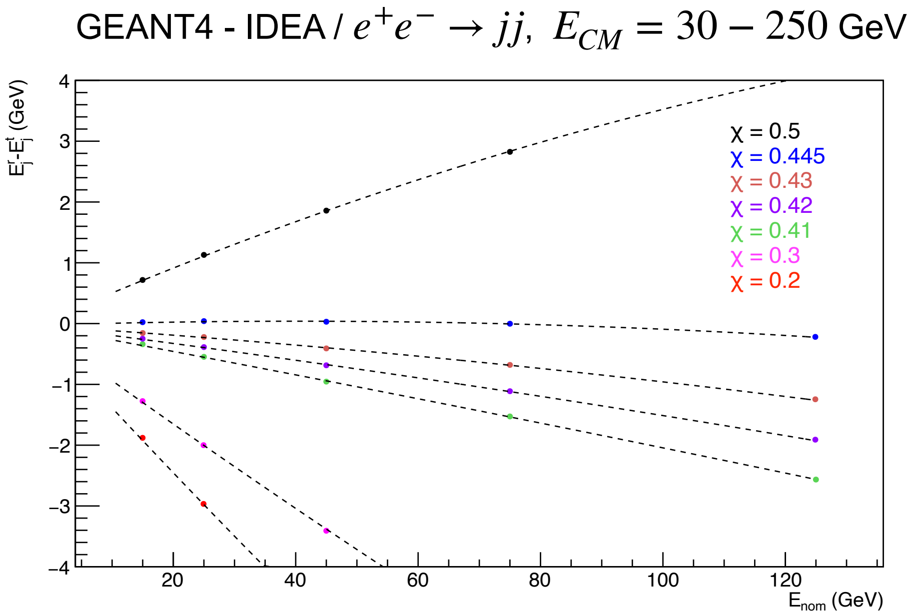
- 2 m long copper based towers
- 36 rotations around the beam axis
- Inner diameter: 5 m
  Outer diameter: 9 m @ 90°

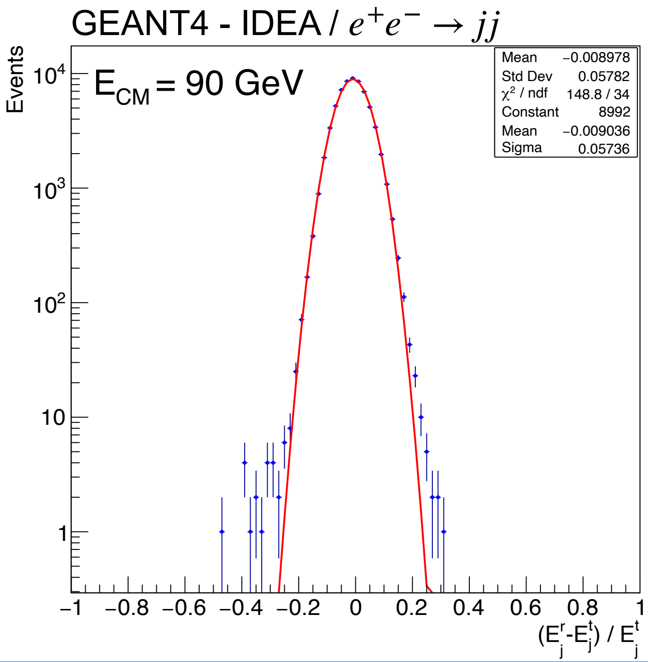
L. Pezzotti, G. Polesello, *et. al.*

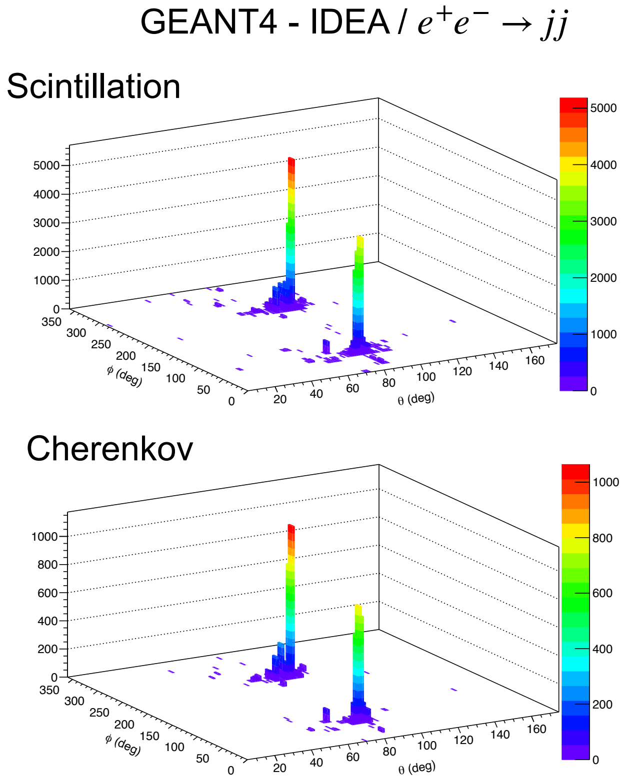
Proof of concept

Geant4 indications on the expected performance (selected results):

- $10\% - 15\,\%/\sqrt{E}$ EM energy resolution
- $25\% - 30\,\%/\sqrt{E}$ energy resolution for single hadrons (including neutral hadrons)
- $5\,\%$ energy resolution for jets at 50 GeV
- Sub-percent linearity in the FCCee energy ranges for e^-/γ, hadrons and jets.







M. Lucchini, *et. al.*

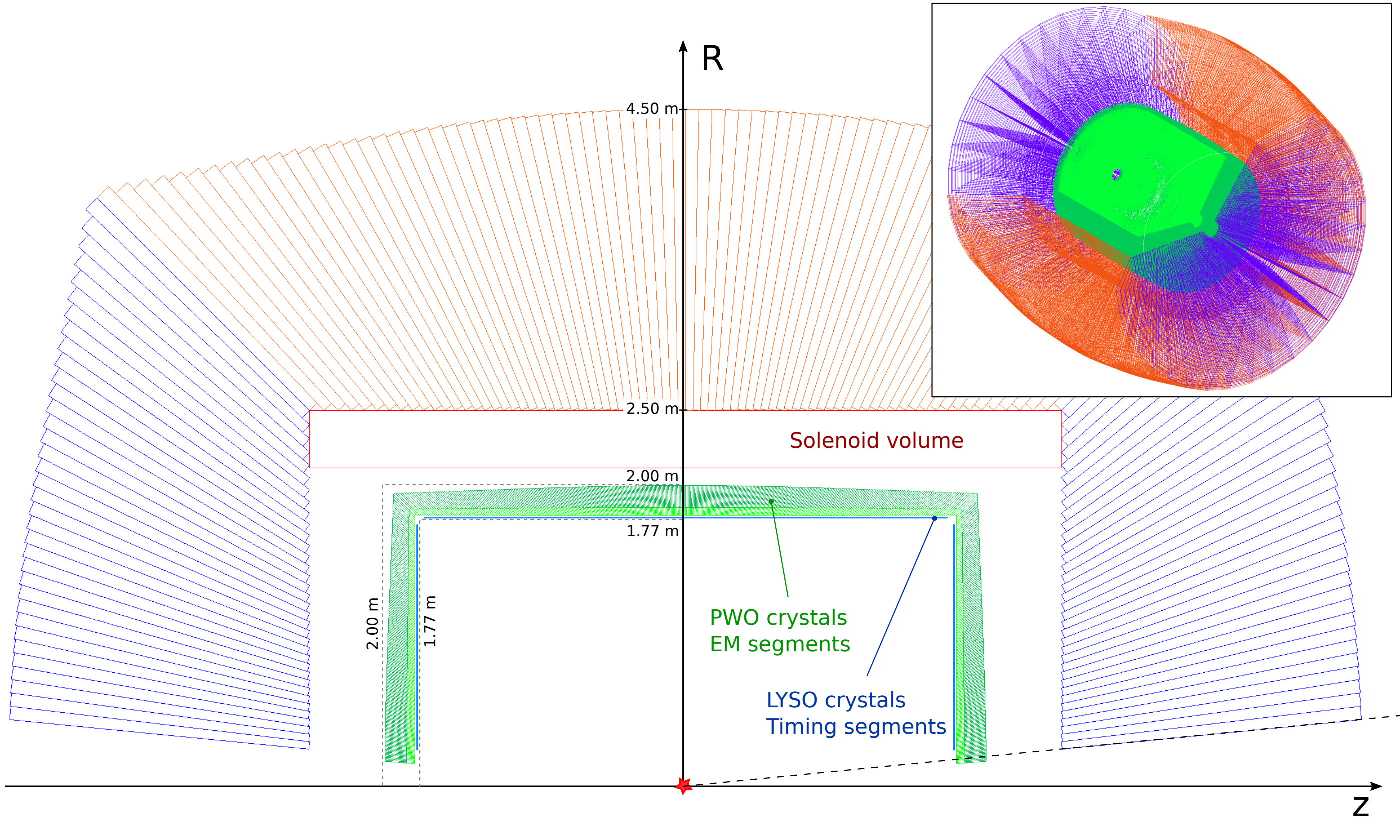
Combining DR crystals and fibers

Integration of a crystal calorimeter option in the 4π Geant4 IDEA simulation:

- Barrel crystal section inside solenoid volume
- 1x1 cm^2 PWO segmented crystals granularity
- Radial envelope: ~1.8-2.0 m

Standalone performance:

- Single hadrons: $25\% - 30\,\%/\sqrt{E}$ (including neutral hadrons)
- Single $e^{\pm}/\gamma$: $3\,\%/\sqrt{E}$
- Jets: $32\,\%/\sqrt{E} \oplus 3.4\,\%$ (from $e^+e^- \rightarrow jj$ events)



Standalone performance for hadrons and jets compatible with the IDEA baseline option, a proto-PF approach was tested to further improve these numbers

M. Lucchini, *et. al.*

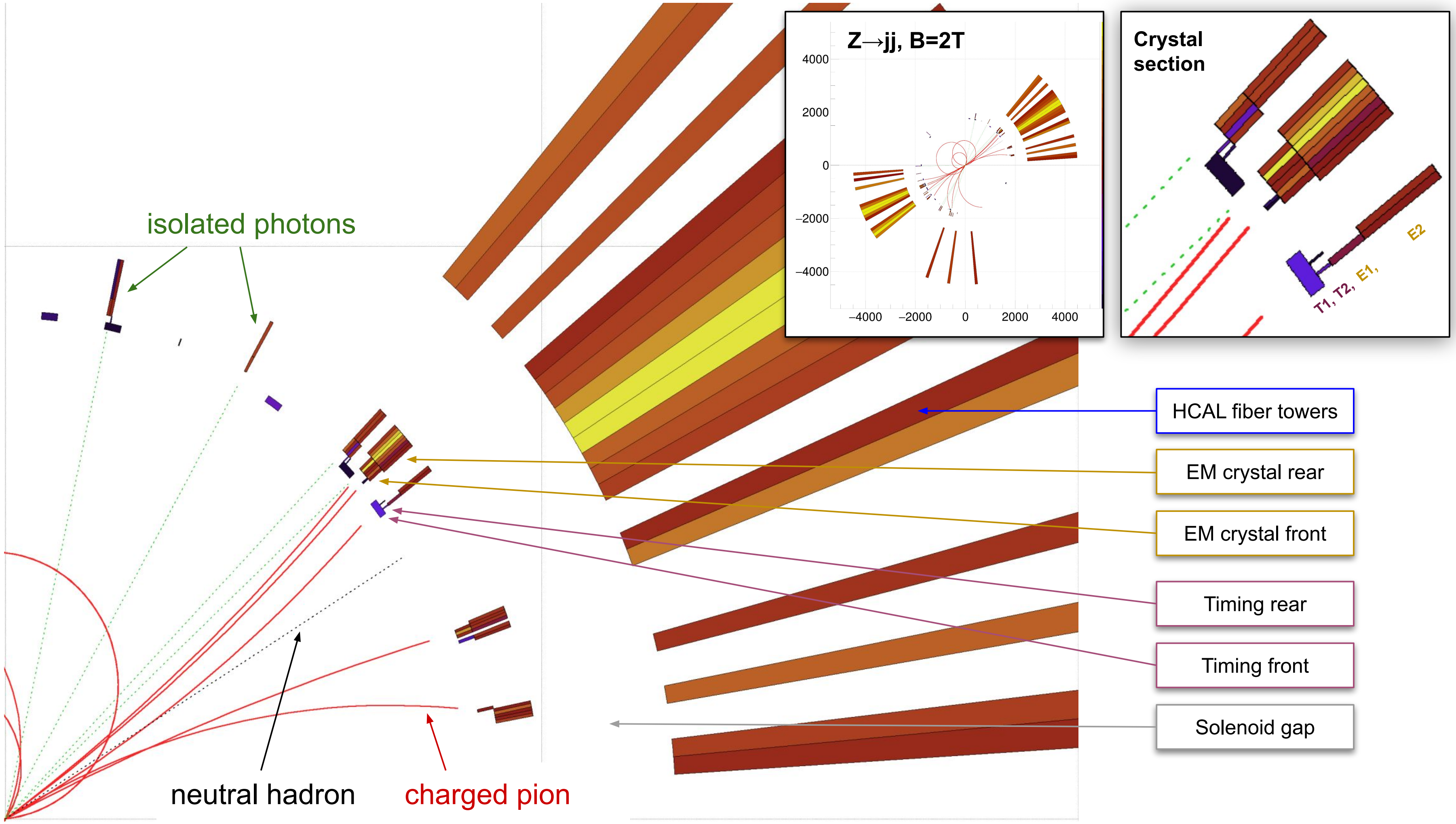
Combining calo hits and tracks

1) Hits around *crystal neutral seeds* are clustered as *photon hits* and are not associated to tracks.

2) Calorimeter hits are associated to tracks based on their track-distance.

↓

If the sum of the energy associated to a track is within 1σ from the expected energy the calorimeter hits are replaced with the track momentum.



Combining calo hits and tracks

1) Hits around *crystal neutral seeds* are clustered as *photon hits* and are not associated to tracks.

2) Calorimeter hits are associated to tracks based on their track-distance.

↓

If the sum of the energy associated to a track is within 1σ from the expected energy the calorimeter hits are replaced with the track momentum.

3) The Durham algorithm is fed with the collection of

- all photon hits from 1)
- a collection of tracks
  - charged particles not reaching the calo
  - tracks that were swapped with calorimeter hits from 2)
- All the other calo hits (ECAL+HCAL) not swapped out
- The algorithm (FASTJET) clusters the 4-momenta into *jets*
- The jet energy *(non-swapped component)* is corrected with DRO equation

$$E_{jet} = C_{PFA} \cdot \left[\sum E_{hits,\gamma} + \sum E_{tracks} + \sum E_{hits,leftover,DRO} \right]$$

M. Lucchini, *et. al.*

Combining calo hits and tracks

1) Hits around *crystal neutral seeds* are clustered as *photon hits* and are not associated to tracks.

Crystal section is crucial in 1) ↑

2) Calorimeter hits are associated to tracks based on their track-distance.

↓

If the sum of the energy associated to a track is within 1σ from the expected energy the calorimeter hits are replaced with the track momentum.

3) The Durham algorithm is fed with the collection of

- all photon hits from 1)
- a collection of tracks
  - charged particles not reaching the calo
  - tracks that were swapped with calorimeter hits from 2)
- All the other calo hits (ECAL+HCAL) not swapped out
- The algorithm (FASTJET) clusters the 4-momenta into *jets*
- The jet energy *(non-swapped component)* is corrected with DRO equation

$$E_{jet} = C_{PFA} \cdot \left[\sum E_{hits,\gamma} + \sum E_{tracks} + \sum E_{hits,leftover,DRO} \right]$$

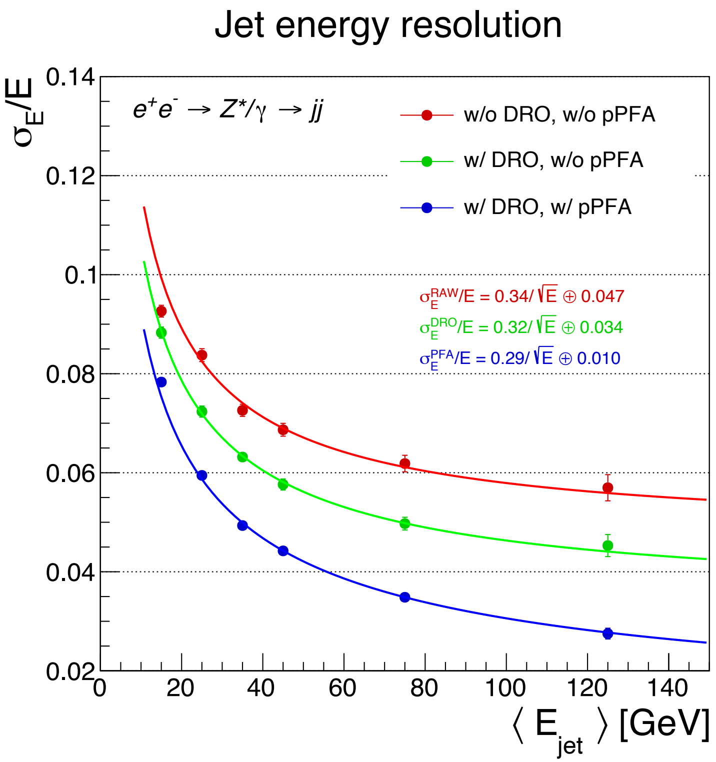
DRO energy correction is helpful in both 2) and 3)

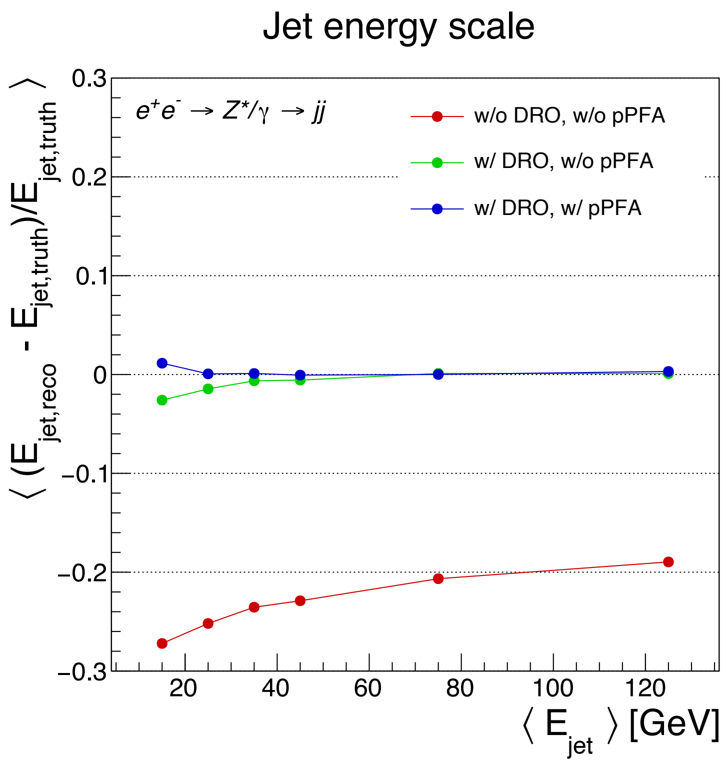
L.Pezzotti, G.Polesello, M. Lucchini, *et. al.*

Jet resolution in the IDEA Crystal option

Jet energy resolution and linearity as a function of the jet energy (from $e^+e^- \to jj$ events at different center-of-mass-energies) for:

- Crystals + IDEA Calo w/o DRO
- Crystals + IDEA Calo w/ DRO
- Crystals + IDEA Calo w/ DRO + pPFA





pPFA leads to a sensible improvement in jet resolution using dual-readout information from crystals and fibers → 3%-4% for jet energies above 50 GeV, within most physics requirements at Higgs factories

S. Giagu, L. Torresi, M. Di Filippo *et. al.*

Machine Learning & DR Calorimetry

The same GEANT4 simulation was used to pioneer the usage of machine learning solutions to maximize the physics potential of future collider experiments.

- Case study: $\tau^{\pm}$ lepton decay mode ID to
  - leverage modern machine learning methods based on differentiable deep neural networks
  - study performance using only standalone DRC information
  - optimizing the detector and design of the readout electronics
- Tasks:
  - classification of τ decays and separation from QCD jets based on Dynamic Graph Neural Networks (DGCNN)
  - DGCNN-based object detection (e.g. identification of γ and n inside hadronic τ decays)

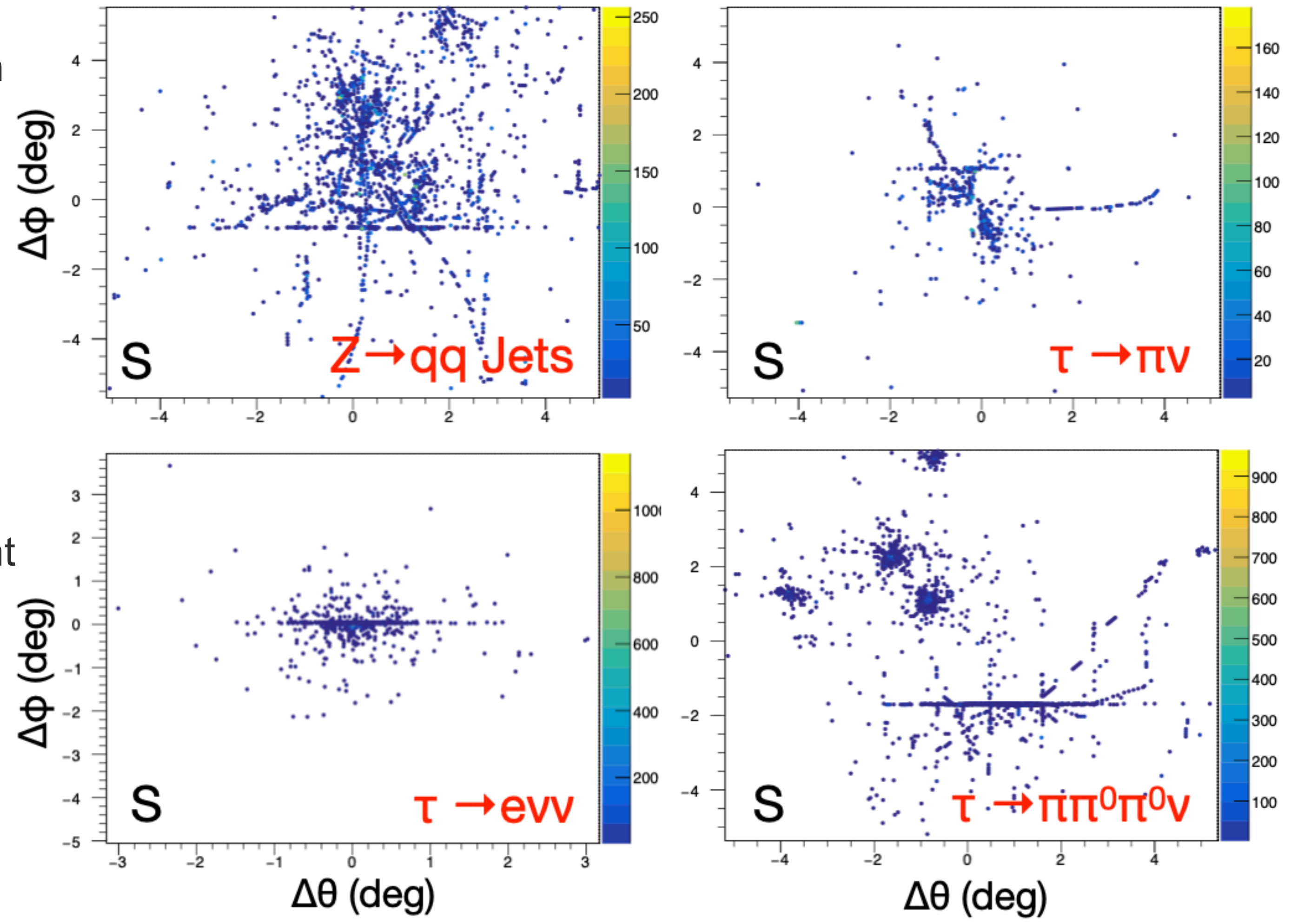
Decay	Label 8-class
$\tau^- \to e^- \nu_e \nu_\tau$	0
$\tau^- \to \pi^- \nu_\tau$	1
$\tau^- \to \pi^0 \pi^- \nu_\tau$	2
$\tau^- \to \pi^0 \pi^0 \pi^- \nu_\tau$	3
$\tau^- \to \pi^- \pi^- \pi^+ \nu_\tau$	4
$\tau^- \to \pi^0 \pi^- \pi^- \pi^+ \nu_\tau$	5
$\tau^- \to \mu^- \nu_\mu \nu_\tau$	6
$Z \to q\bar{q} \to jet\, jet$	7

S. Giagu, L. Torresi, M. Di Filippo *et. al.*

Events display and DGCNN inputs

Data representation

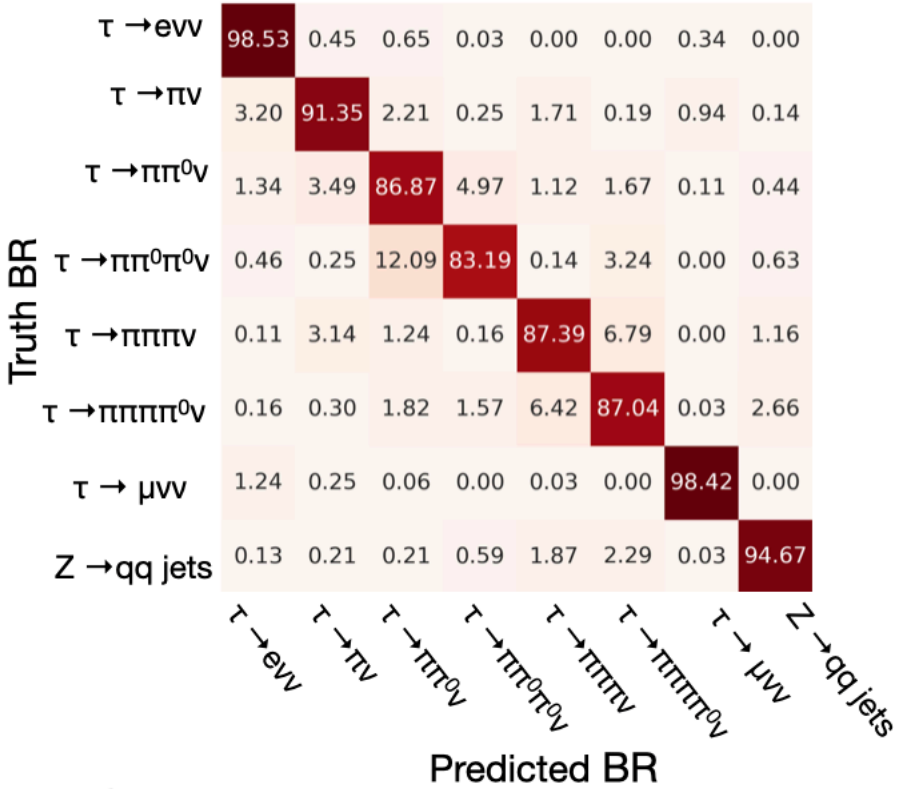
- Image-based: treating the energy deposition on each fiber as the pixel intensity creates an image of the event in fixed-shape mesh
  - CNN standard representation
  - unclear how to incorporate additional information of the fibers
  - very sparse and inefficient representation
- Point-cloud-based: unordered sets of entities distributed irregularly in space, analogous to the point cloud representation of 3D shapes
  - easy to incorporate additional information of the fibers (fibre type, energy, time information, …)
  - the architecture of the neural network has to be carefully designed to fully exploit the potential of this representation → Dynamic Graph CNN



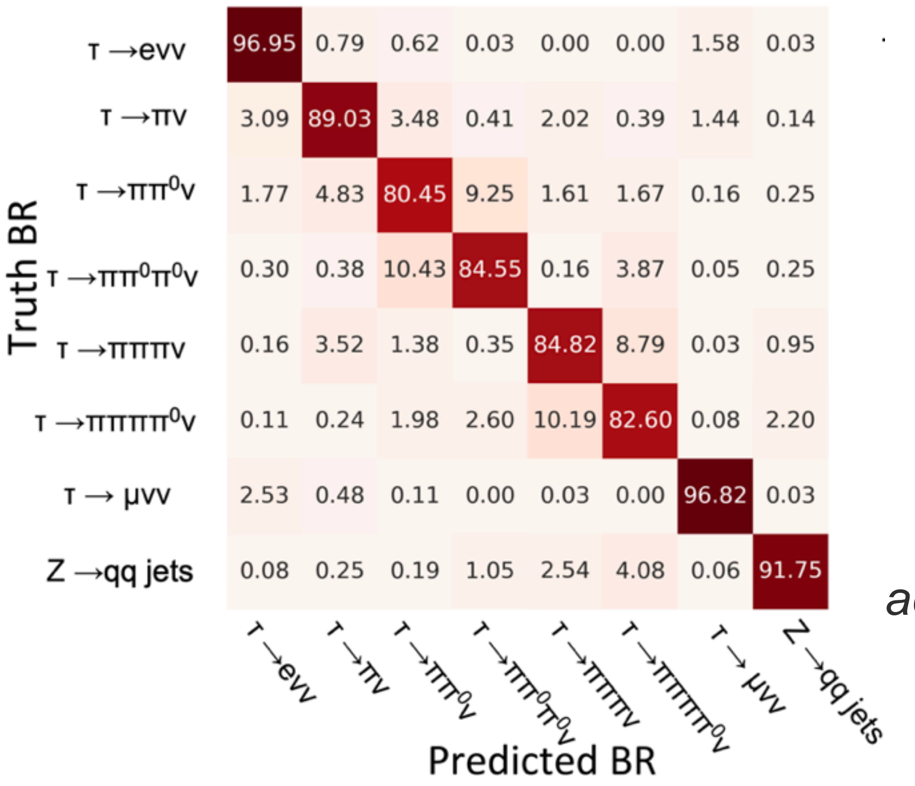
Results on $\tau^{\pm}$ ID with DGCNN

Average accuracy 90.8%

Using coordinates, fiber type (S/C) and p.e. in each fiber



Using only coordinates and fiber type (S/C)



Average accuracy 88.3%

Uncertainty on accuracies ~3-5%

- The calorimeter geometry alone allows excellent τ ID
- Results performed including SiPM emulation do not show significant performance reduction

S. Giagu, L. Torresi, M. Di Filippo *et. al.*

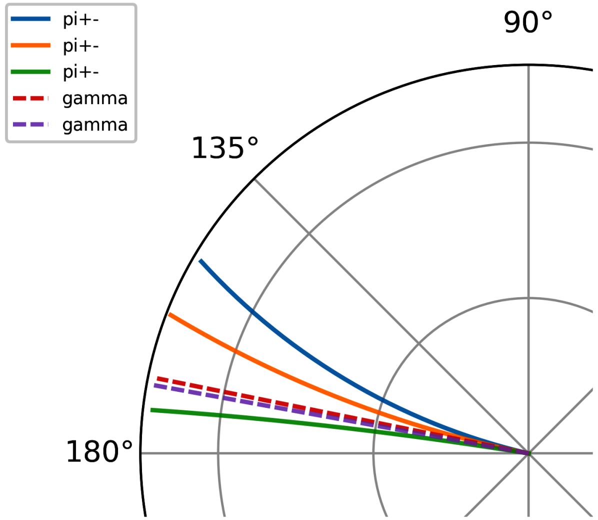
Segmentation, finding objects in calo-jets

DGCNN coupled to a high granularity dual-readout calorimeter can also be exploited for object (particle) detection inside τ and hadronic jets:

- label each fibre by extrapolating Monte Carlo truth particles from production to the DRC into the IDEA magnetic field
- train the DGCNN to predict the label associated to each fibre
- identify the particle associated to the larger energy deposit in each fibre

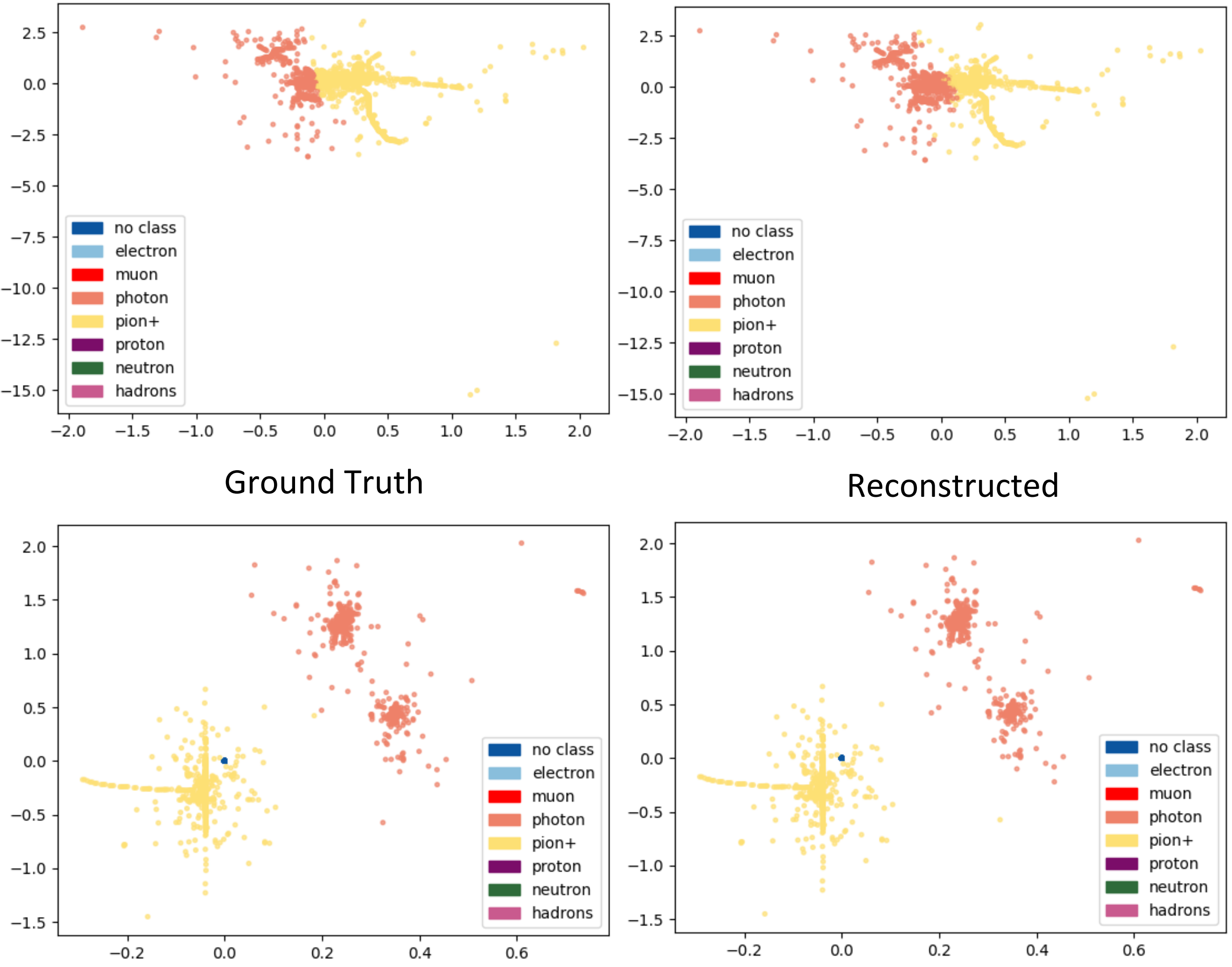
First study ongoing:

- initial tests only on photons/neutrons *VS* other particles identification in τ decays
- γ ID in jets is an open point in unsegmented calorimeters!
- Straightforward application in any Particle-Flow like algorithm



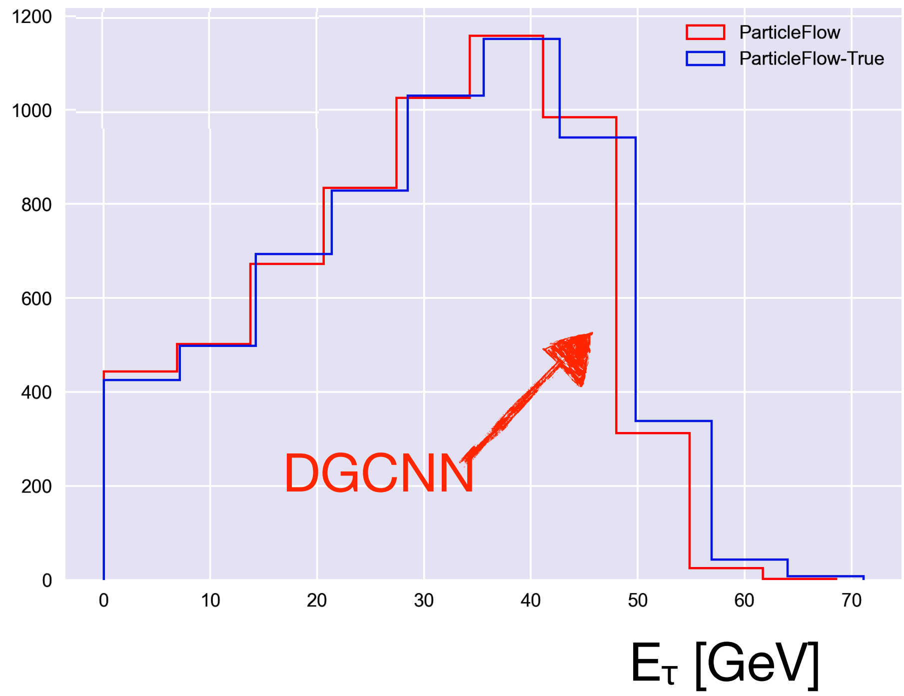
Results on segmentation

Example: segmentation of two $\tau \rightarrow \pi\pi^0\nu_\tau$ events



- τ visible energy reconstructed using
  - DRC for γ
  - MC truth for other particles

comparison of the distributions obtained when photons are identified by the DGCNN and when using the MC truth

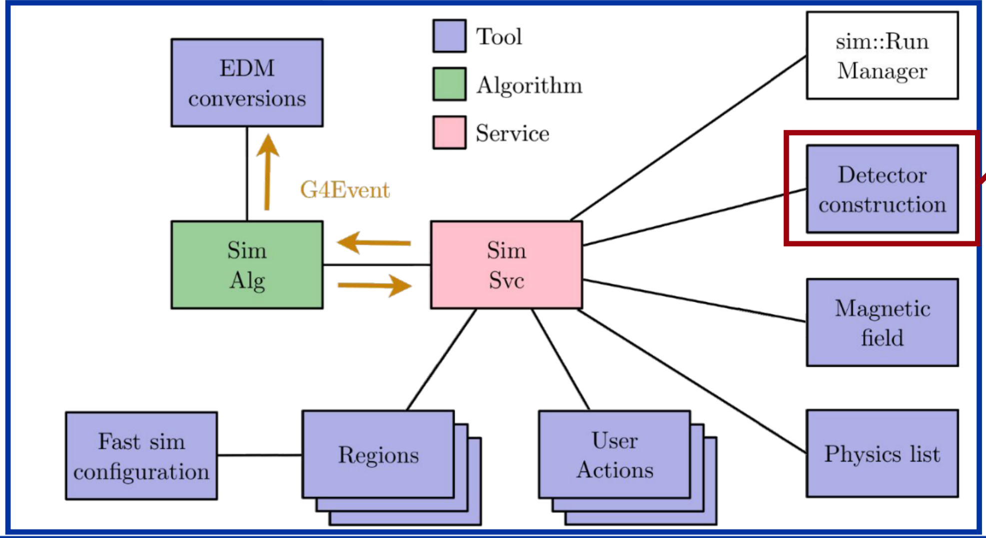


Sang Hyun Ko *et. al.*

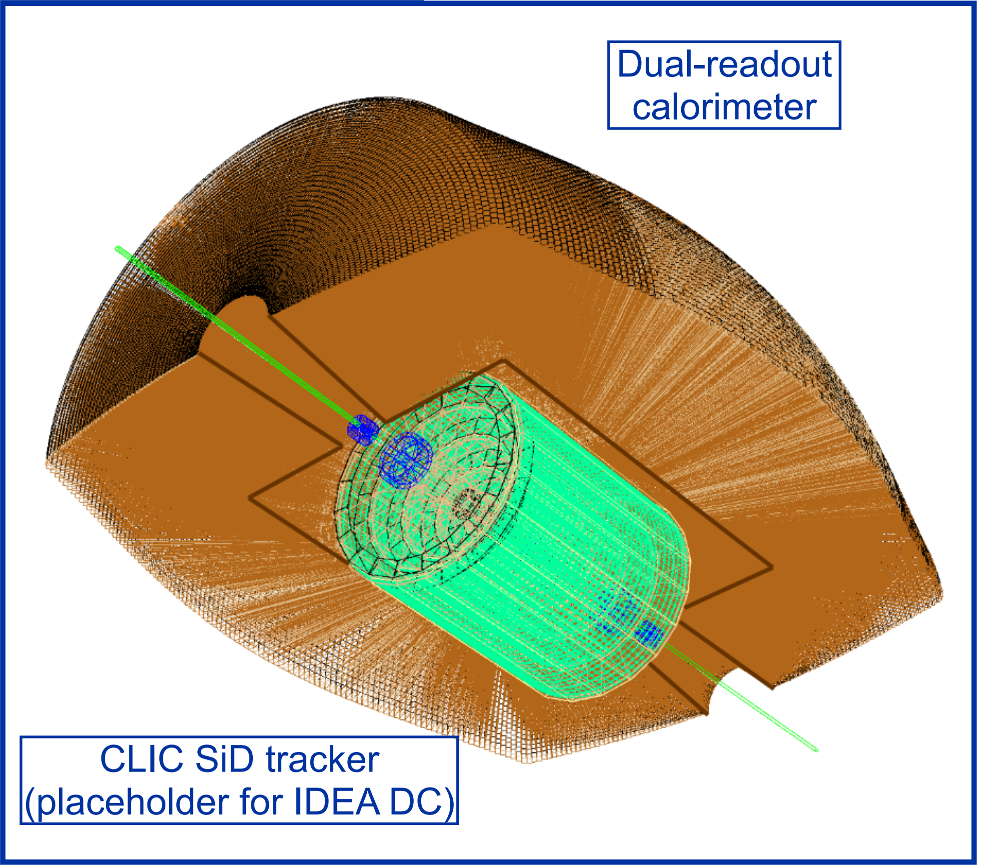
DD4hep geometry migration

- DD4hep is a main framework for detector description
- It is a first step to migrate to key4hep, common SW stack for FCC, ILC, CLIC, CEPC
- An IDEA DR-Calo description was implemented in DD4hep [git]
- To be coupled with a DD4hep description of the IDEA Drift Chamber

k4SimGeant4



DD4hep geometry

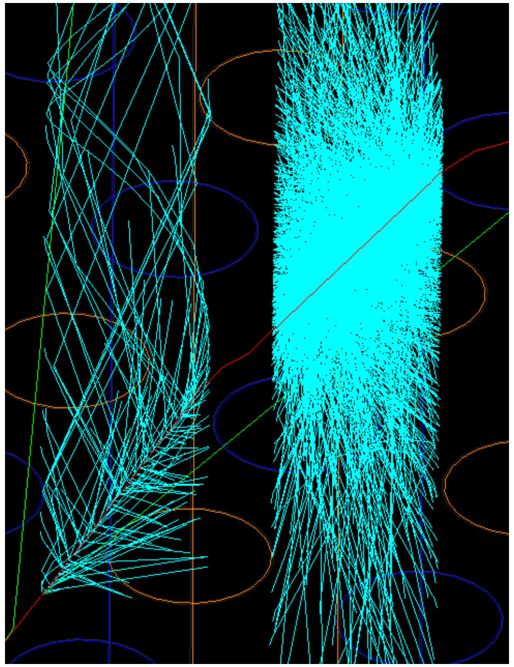


Optical physics simulation

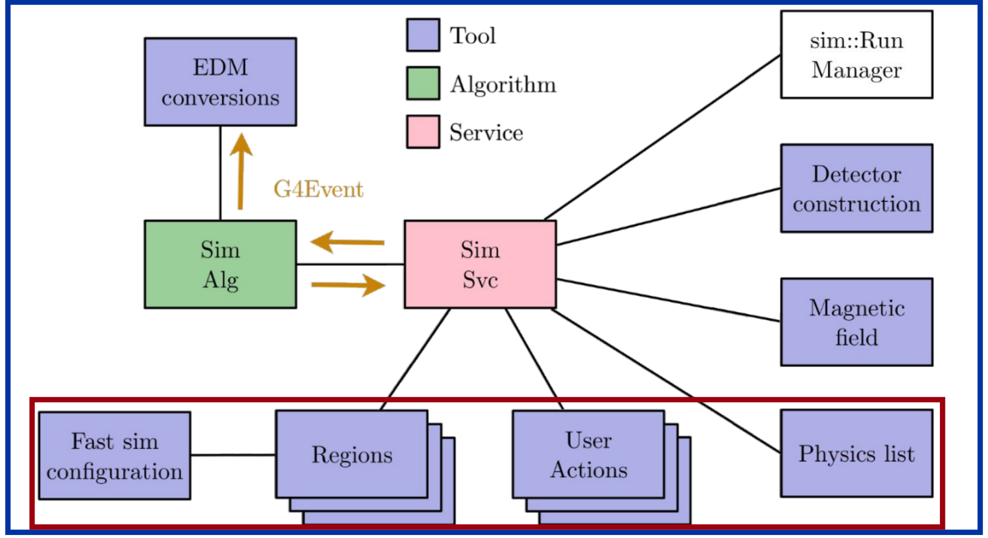
- Precise timing measurement is crucial to recover longitudinal segmentation in fiber calorimeters [link]
- Optical physics provides detailed timing information but at a high-cost of CPU (open problem in Geant4)
- Incorporating modularized G4 Physics List using FTFP_BERT, Geant4 optical physics and a Fastsim module applied to optical photons [link]

Cherenkov Scintillation

S. H. Ko *et. al.*



k4SimGeant4



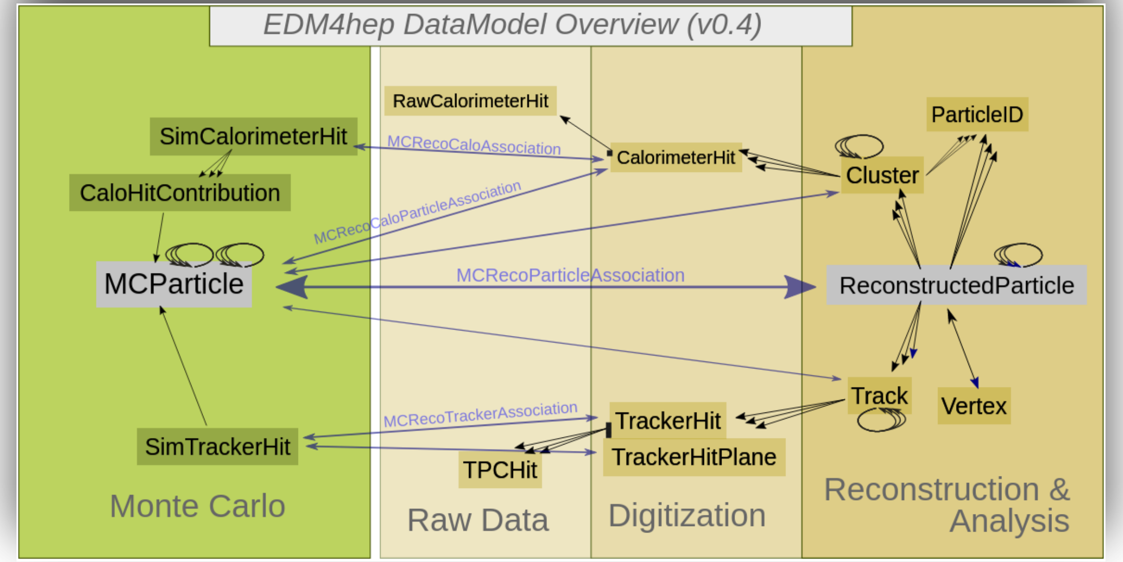
k4run configurator

```
regionTool = SimG4FastSimOpFiberRegion("fastfiber")
opticalPhysicsTool = SimG4OpticalPhysicsList("opticalPhysics", fullphysics="SimG4FtfpBert")
physicslistTool = SimG4FastSimPhysicsList("Physics", fullphysics=opticalPhysicsTool)

from Configurables import SimG4DRcaloActions
actionTool = SimG4DRcaloActions("SimG4DRcaloActions")

# Name of the tool in GAUDI is "XX/YY" where XX is the tool class name and YY is the given name
geantservice = SimG4Svc("SimG4Svc",
  physicslist = physicslistTool,
  regions = ["SimG4FastSimOpFiberRegion/fastfiber"],
  actions = actionTool
)
```

I. Vivarelli, Sang Hyun Ko *et. al.*

EDM4hep

- EDM4hep is a common EDM that can be used by all communities in the key4hep project
- Second step to migrate to the common SW stack
- Needed to interface G4Event/G4VHit of the DRC simulation to EDM4hep calorimeter hits.



- Migrated information:
  - MC truth (edep) → edm4hep::SimCalorimeterHits
  - Readout (#p.e.) → edm4hep::RawCalorimeterHits
  - Digitization (#ADC) → edm4hep::RawCalorimeterHits
  - Reco (2D/3D) → edm4hep::CalorimeterHits

```
from Configurables import SimG4SaveDRcaloHits, SimG4SaveDRcaloMCTruth
saveDRcaloTool = SimG4SaveDRcaloHits("saveDRcaloTool", readoutNames = ["DRcaloSiPMreadout"])
saveMCTruthTool = SimG4SaveDRcaloMCTruth("saveMCTruthTool") # need SimG4DRcaloActions

geantsim = SimG4Alg("SimG4Alg",
  outputs = [
    "SimG4SaveDRcaloHits/saveDRcaloTool",
    "SimG4SaveDRcaloMCTruth/saveMCTruthTool"
  ],
  eventProvider = edmConverter
)
```

Conclusions & Take Home

- Geant4 has been used to simulate a dual-readout calorimeter concept for IDEA providing good indications on the possibility of:
  - reconstructing $e^{\pm}$, hadrons and jets with superior HAD resolution and linearity
  - combining DR fibers and crystals (in a fully compensating segmented calorimeter)
  - using proto-PF approach improving the jet energy measurements
- CNN can be used to extract features from an unprecedented amount of calorimetric information ($\simeq 100$ million fibers). Case studies:
  - $\tau^{\pm}$ ID
  - Finding particles (γ, n) into jets
- Migration of the DR Calo simulation to the key4hep SW stack is at an advanced level (DD4hep and EDM4hep)

Geant4

↓

DR and PF proof of concept

↓

Machine learning (superior reconstruction)

↓

key4hep