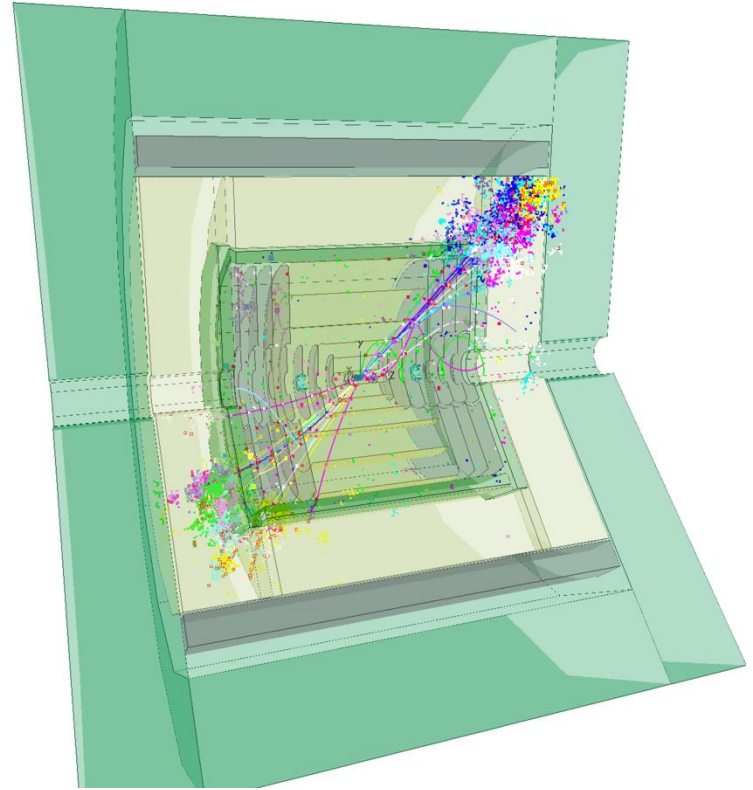


Particle Flow_s at FCC-ee

Dolores Garcia

Summary of the talk

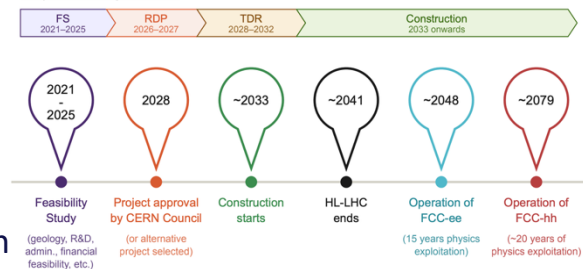
- Classical PF algorithms at FCC
- ML based PF algorithms at FCC
- Generalization to different detector concepts
- Robustness and systematic uncertainties of these algorithms



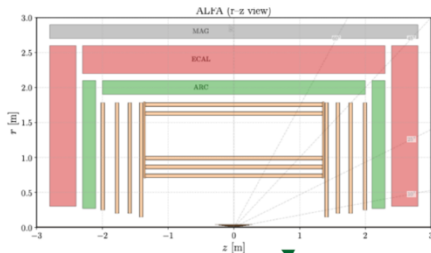
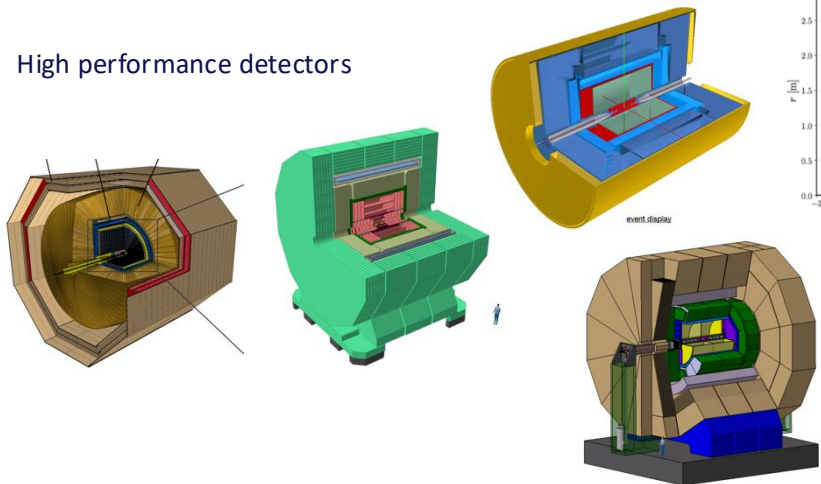
Event display in the CLD detector for a $Z/\gamma^* \rightarrow \bar{q}q$ event 365 GeV. From Bacchetta et al 2019.

Motivation

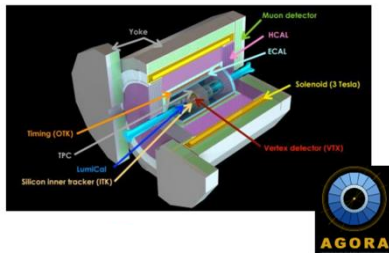
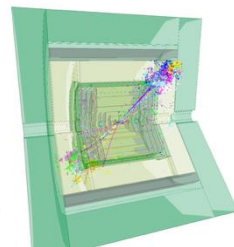
- ❑ Entering the Reference Design Phase era
 - ❑ Goal is to enable FCC-ee detector design optimization
 - ❑ Solidify the physics-reach and physics case with robust full-simulation implementations
- ❑ In order to **optimize detectors** and **compare designs** fairly we need **adaptable reconstruction algorithms**



High performance detectors



High quality reconstruction

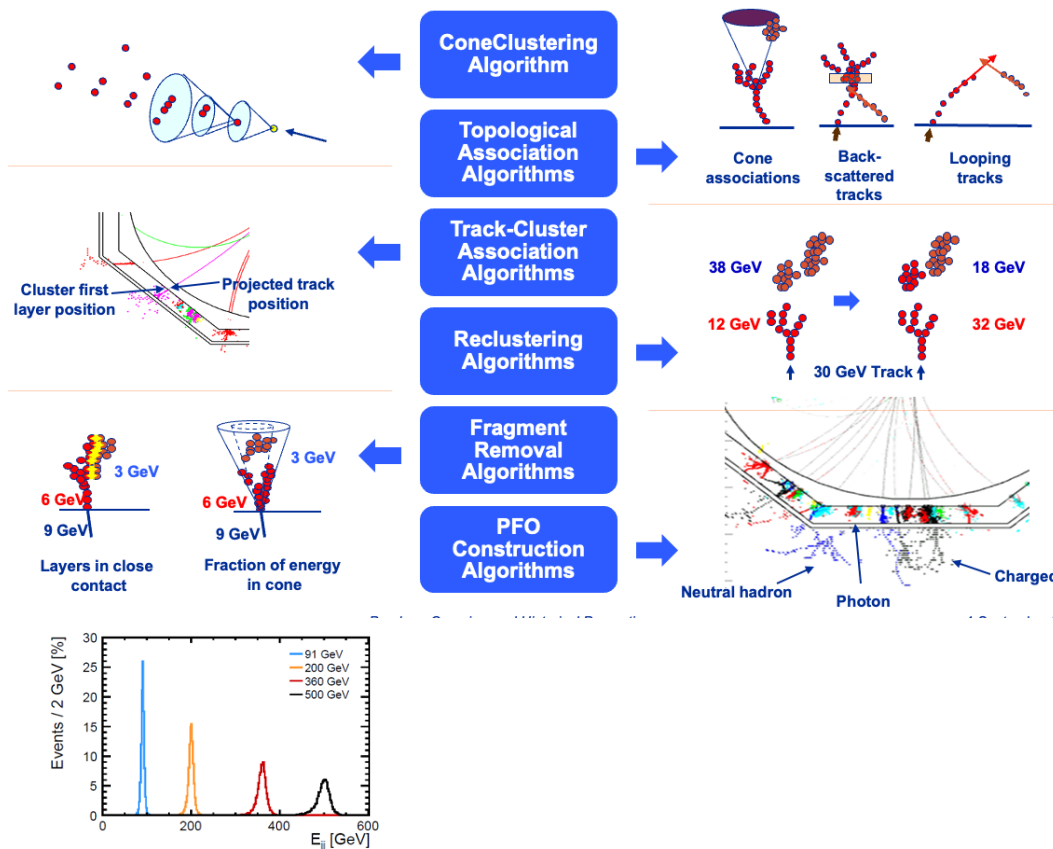


What is Particle Flow?

- ❑ **The goal:** Optimally combine measurements from all sub-detectors
 - Vertex, tracker, ECAL/HCAL, muon detector
 - From **hits** to **identified final state particles**
- ❑ For each **~stable particle**, this combination aims at returning a **reco particle**
 - With the best **direction** and **energy** estimate
 - With the most likely **type** (photon, electron, muon, charged hadron, neutral hadron)
 - With its origin vertex (primary, secondary, etc.)
- ❑ The list of individual particles can then be used
 - To determine the **total/missing energy** and momentum; build jets; tag **jet flavour**; identify and reconstruct **taus** from their decay products; select leptons or photons; etc.
 - With the best possible energy/angular **resolution**, efficiency, purity, ...
 - With **no** overlap or **double counting** - as with a generator-level analysis

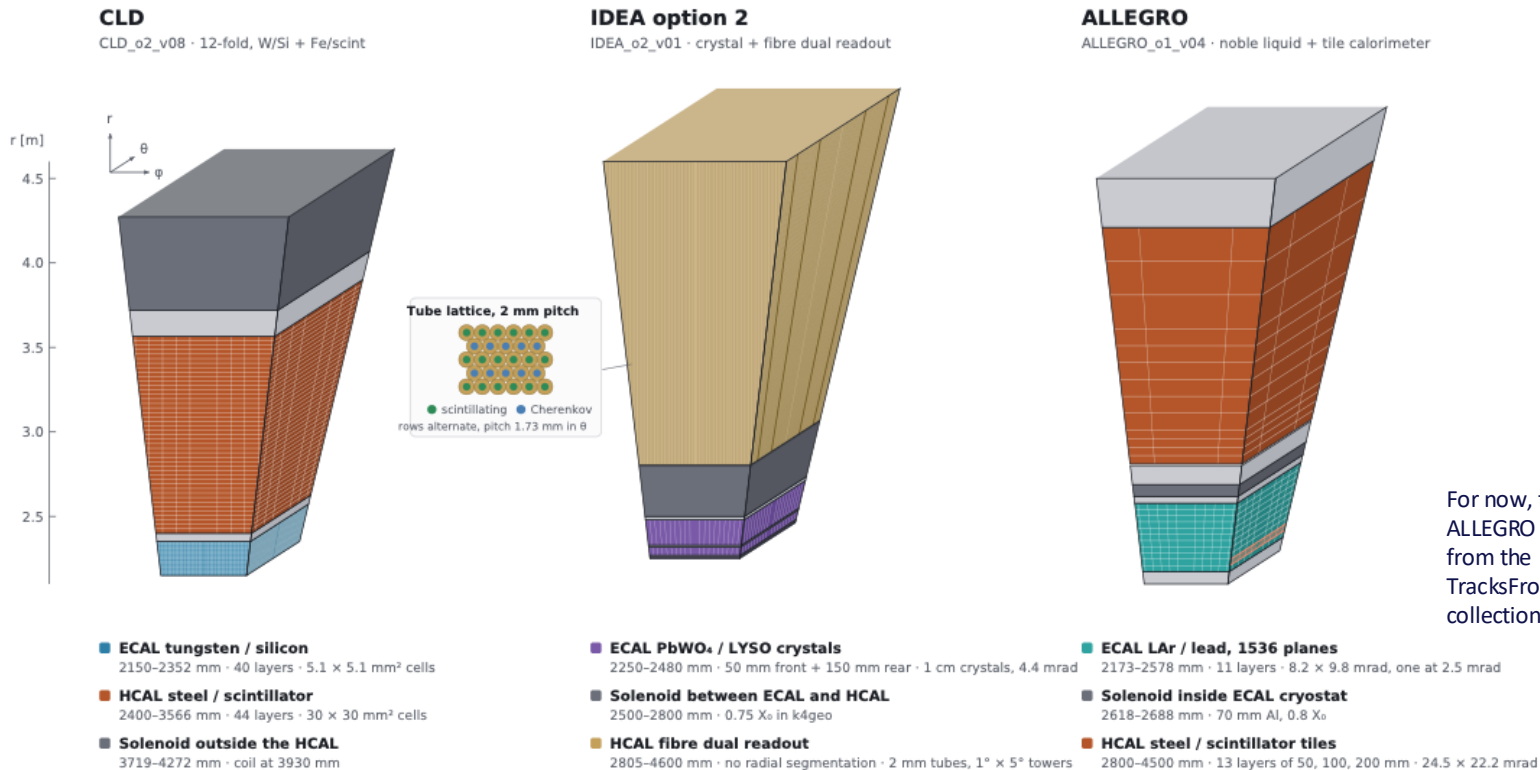
How is it tackled?

- ❑ Classically, multiple algorithms:
 - Tracking (one per sub-detector), clustering (ECAL/HCAL), and linking tracks and clusters
 - ALEPH, DELPHI, CMS, ATLAS, CLIC,...
- ❑ Pandora SDK (baseline FCC)
 - Series of distributed algorithms
 - Iterative clustering
 - Track cluster association and reclustering
 - Neutral identification
 - Relies on tuning, limiting flexibility during detector design phase (allows for geometry studies e.g transverse calo segmentation)



Example 1: Clustering block in CLD vs IDEA vs ALLEGRO

5°x5°
wedge



For now, for IDEA and ALLEGRO tracks are taken from the TracksFromGenParticles collection

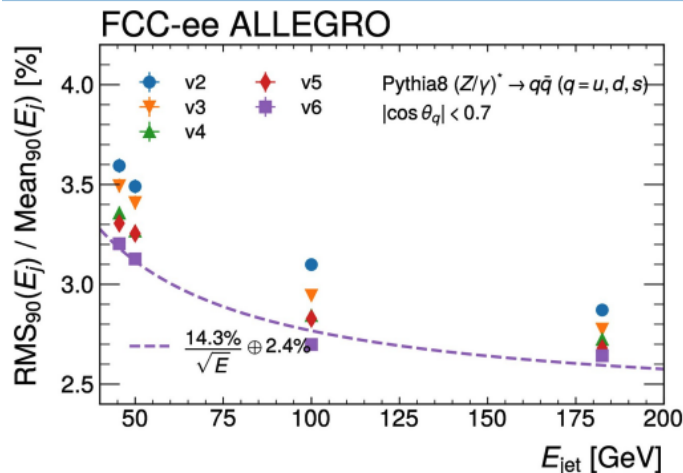
Example 1: Clustering block in CLD vs IDEA vs ALLEGRO

- Pandora clustering algorithm implemented in LCContent [1, 2]
- Cone clustering:
 - Flag hits as isolated or MIP (**IsolationMaxNearbyHits**, **IsolationNLayers**, **MipLikeMipCut**, **MipMaxNearbyHits**)
 - Isolated hits are then excluded from clustering
 - Cone clustering: forward pass on pseudo-layers
 - Seed clusters (with tracks) and attach cluster's direction (**NLayersSpannedForFit**)
 - Fit and test direction vs original
 - Find hits in previous and same layers to reattach previously left hits
 - A hit calculates distance to closest 3 clusters and associates itself with the closest one with a distance cut (**GenericDistanceCut**)
- After the first pass, topological clustering association (13 algorithms), reclustering with 12 different cluster configurations
- The full pandora algorithm has ~800 parameters, CLD varies ~60, a Pandora retuning is needed to augment performance (inherited from ILD/CLIC defaults)

Example 1: Clustering block in CLD vs IDEA vs ALLEGRO

@Archil Durglishvili

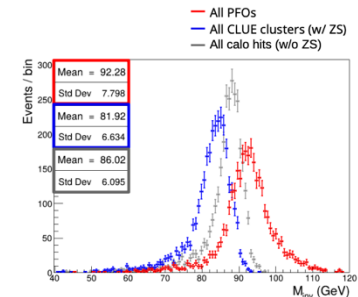
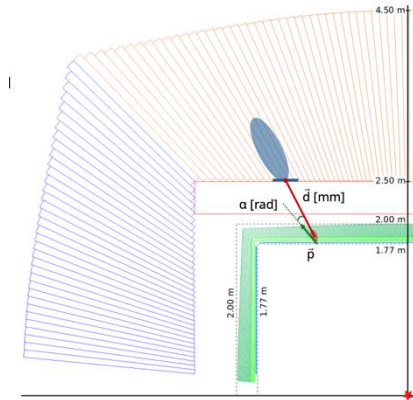
- Adapting Pandora's clustering by parameter tuning
- Hadronic scale calibration for charged hadronic cluster creation (uses E/p) (calibrated with 30 GeV K_{0L})
- Adaptation of the cone clustering algorithm parameters (distance between ECAL and HCAL layers impose a threshold $MaxClusterDirProjection$)
- v2 – PhotonRecovery algorithm is disabled and cone-clustering for primary clustering is tuned ($MaxClusterDirProjection = 300\text{mm}$)
- v3 – in addition to v2 changes, photon LH classifier is trained for ALLEGRO
- v4 – in addition to v3 changes, $MaxClusterDirProjection$ value for primary clustering is changed to 350mm
- v5 – in addition to v4 changes, $MaxClusterDirProjection$ value for reclustering algorithm is changed from 200mm to 350mm
- v6 – in addition to v5 changes, hadronic energy resolution value is changed from 60% to 40%



Example 1: Clustering block in CLD vs IDEA vs ALLEGRO

- CLUE clustering algorithm on ECAL and HCAL
 - Different set of parameters (4 each) tuned for $Z \rightarrow qq$
- Then ECAL/HCAL cluster association:
 - Associate tracks to ECAL clusters based on length from extrapolation
 - Round robin association based on cone-based matching: sort clusters in ascending order E/p , each cluster takes one HCAL cluster with the smallest angle
 - Repeat for neutrals
 - Cut based on angle

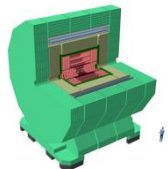
@Sanghyun Ko



The need for an end-to-end approach: Multiple detectors

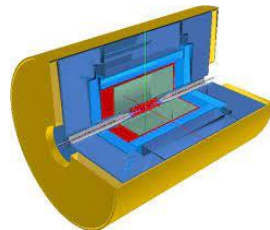
□ Adaptable reconstruction is crucial for a systematic **design optimization**:

- Parameter tuning is not straight forward; the detector will still change + different detectors
- Costly conventional implementation



CLD

- Silicon pixel + silicon tracker
- ECAL Granular Silicon-Tungsten 5x5 mm² cells (15-17 %)
- Scintillator-steel, 44 layers, 30x30 mm²
- Solenoid outside calo

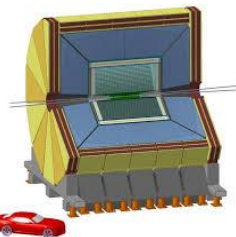


ALLEGRO

- Silicon pixel + drift chamber + silicon wrapper
- ECAL granular noble-liquid sampling LAR (8-10 %)
- Scintillating-tile HCAL
- Solenoid behind ECAL

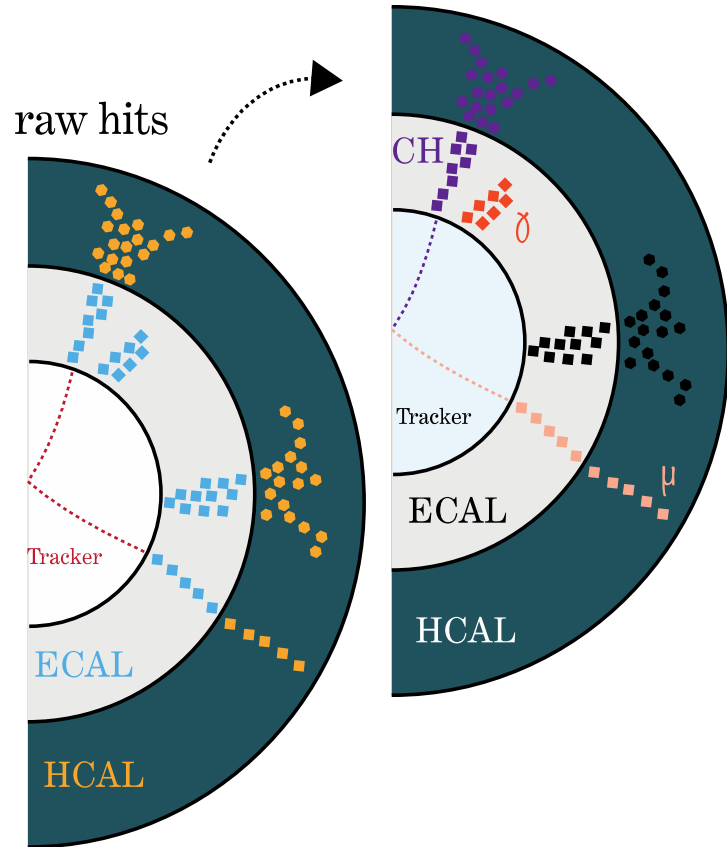
IDEA

- Silicon pixel + drift chamber + silicon wrapper
- Dual-readout crystal
- Dual-readout fiber scintillating and cherenkov
- Solenoid between ECAL/HCAL



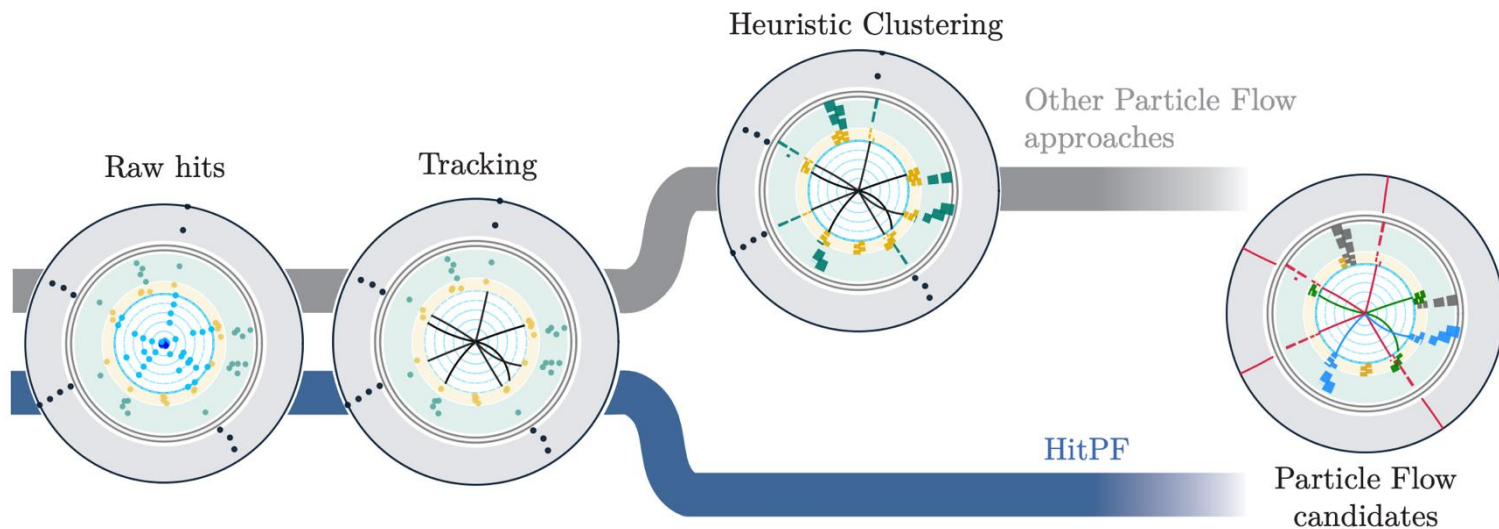
And others... ALFA, ILD

How to set it as an ML problem



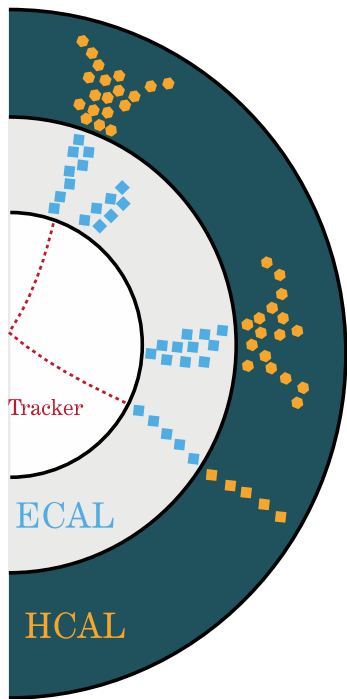
Hit-based approach to Particle Flow

Michele Selvaggi
Lena Herrmann
Gregor Krzmarc
Bohdan Dudar



The workflow: Inputs

raw hits



Track 'hits':
Reference point in calorimeter (position)
 $\mathbf{p}$ at vertex and calo
 χ from track fit
Sub-detector ID



ECAL hit
Position (x,y,z)
E energy
Sub-detector ID



HCAL hit
Position (x,y,z)
E energy
Sub-detector ID



Muon hit
Position (x,y,z)
E energy
Sub-detector ID

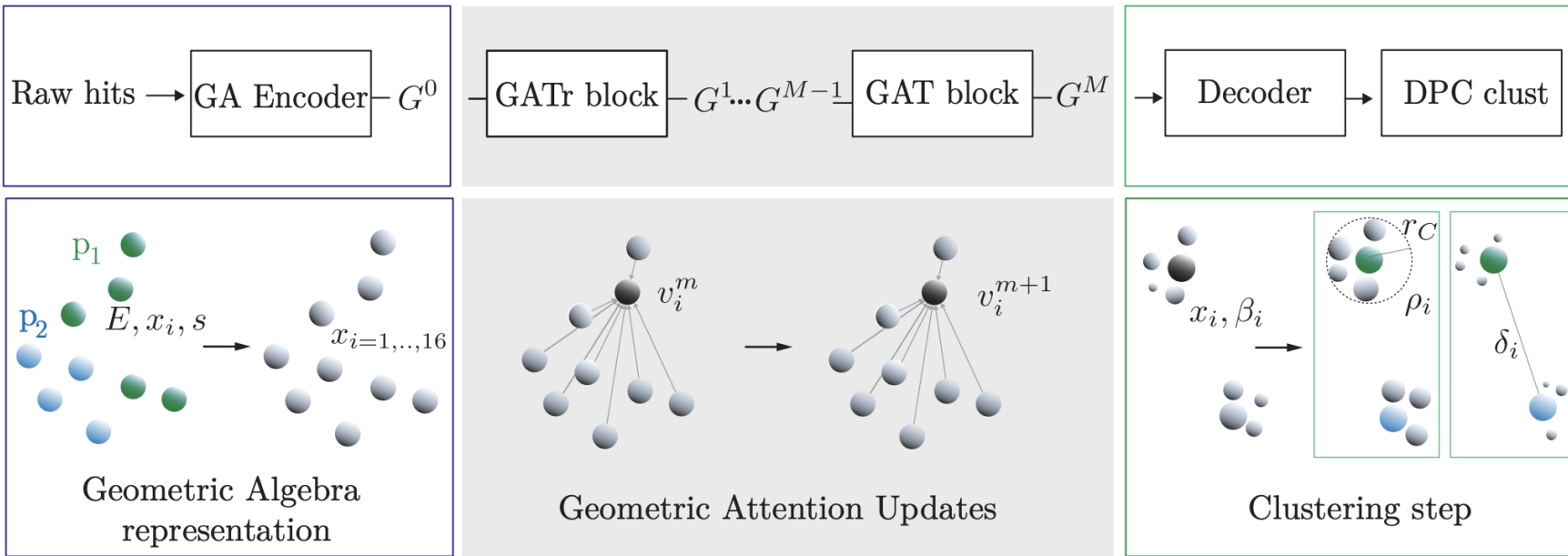


Other sub-detectors?
eg. ARC

pointcloud

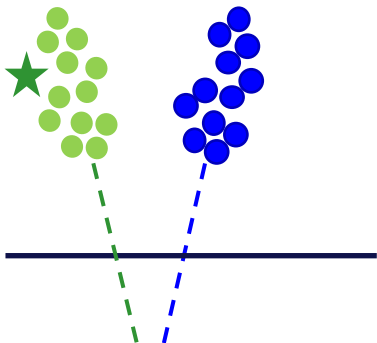
Geometric Algebra Transformer

v1, v2 coming out soon



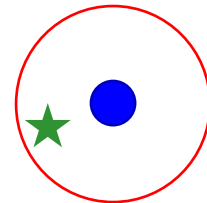
Loss: object condensation

Inputs (K=2)



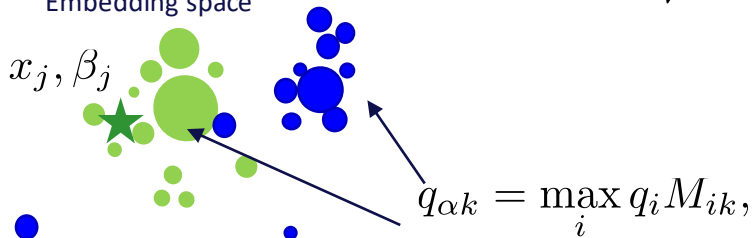
$$\star \bullet \quad \check{V}_k(x) = \|x - x_\alpha\|^2 q_{\alpha k},$$

$$\star \bullet \quad \hat{V}_k(x) = \max(0, \underline{1 - \|x - x_\alpha\|}) q_{\alpha k}$$



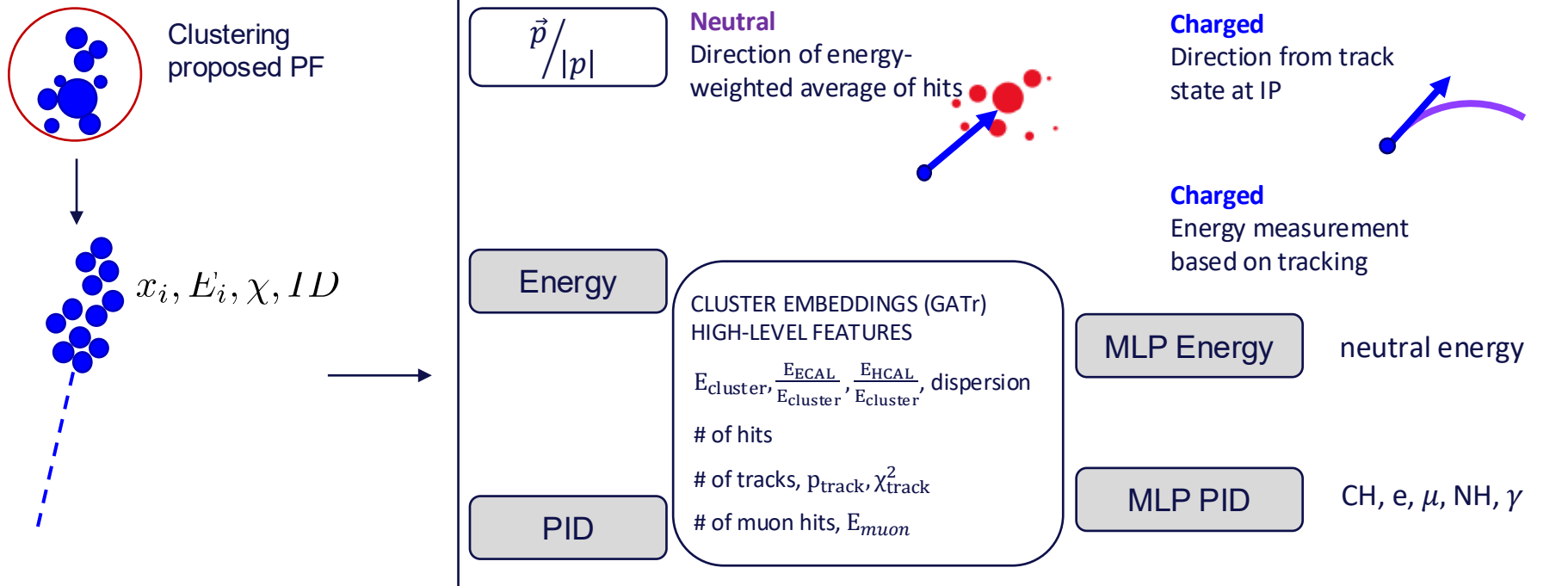
$$L_V = \frac{1}{N} \sum_{j=1}^N q_j \sum_{k=1}^K \left(M_{jk} \check{V}_k(x_j) + (1 - M_{jk}) \hat{V}_k(x_j) \right)$$

Embedding space



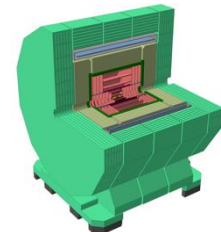
Regression & Classification

Per PF candidate property determination



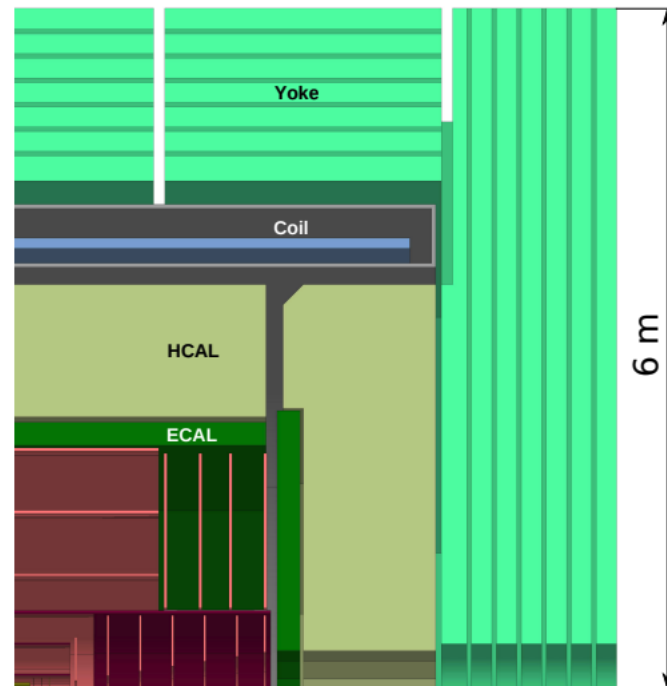
CLD a classic detector

Key4hep

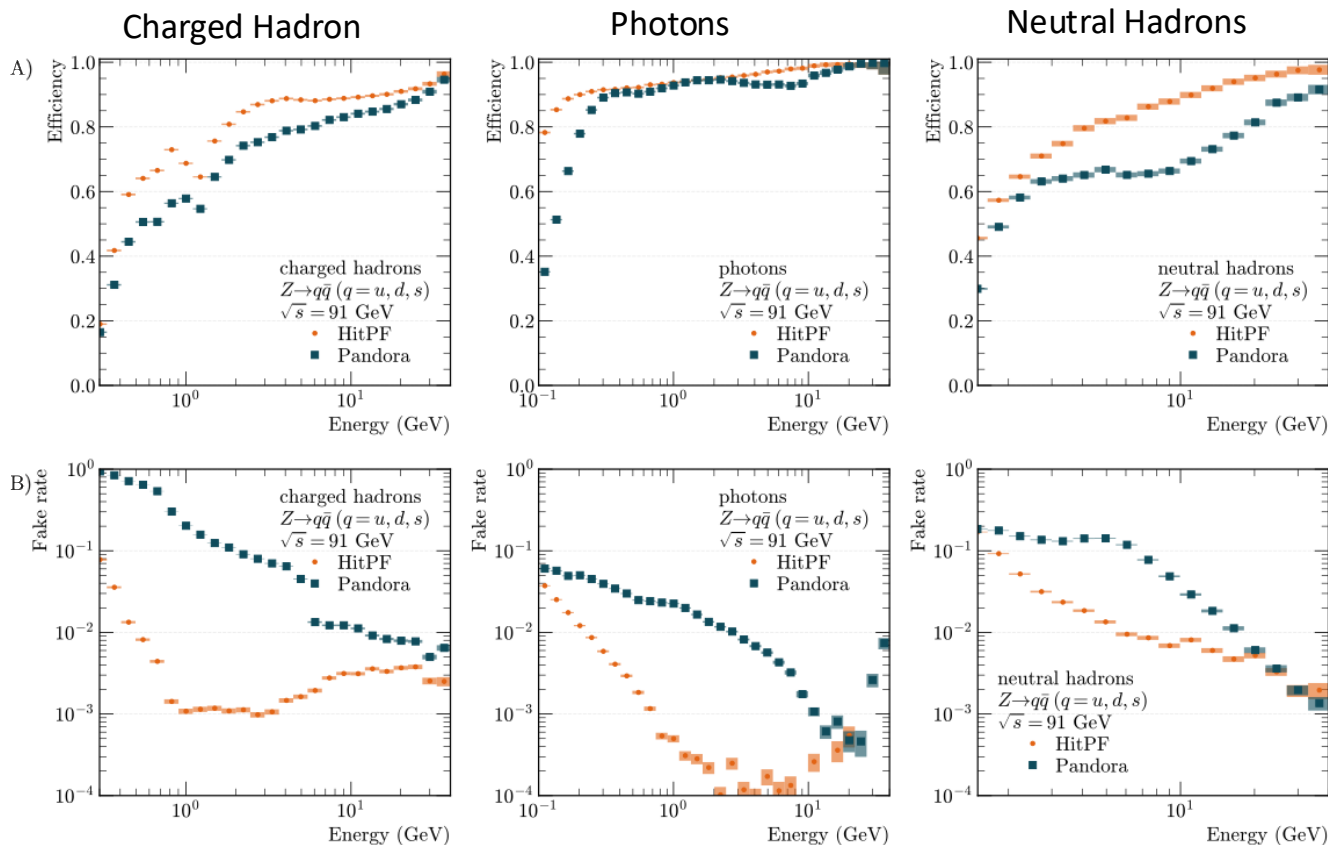


Full simulation

- ☐ Implemented in DD4hep
- ☐ Full simulation in Key4hep
- ☐ Comparison againsts Pandora tuned for CLD (CLD_o2_v05)
- ☐ $Z \rightarrow q\bar{q}, q = (u, d, s)$ 91 GeV
- ☐ Jet like events:
 - ☐ Close particles
 - ☐ High particle multiplicity



PID: Efficiency and fake rate



CH:

- Up to 20% higher eff.
- Fake rate $\mathcal{O}(2)$ smaller
- Eff. limited by track reco.

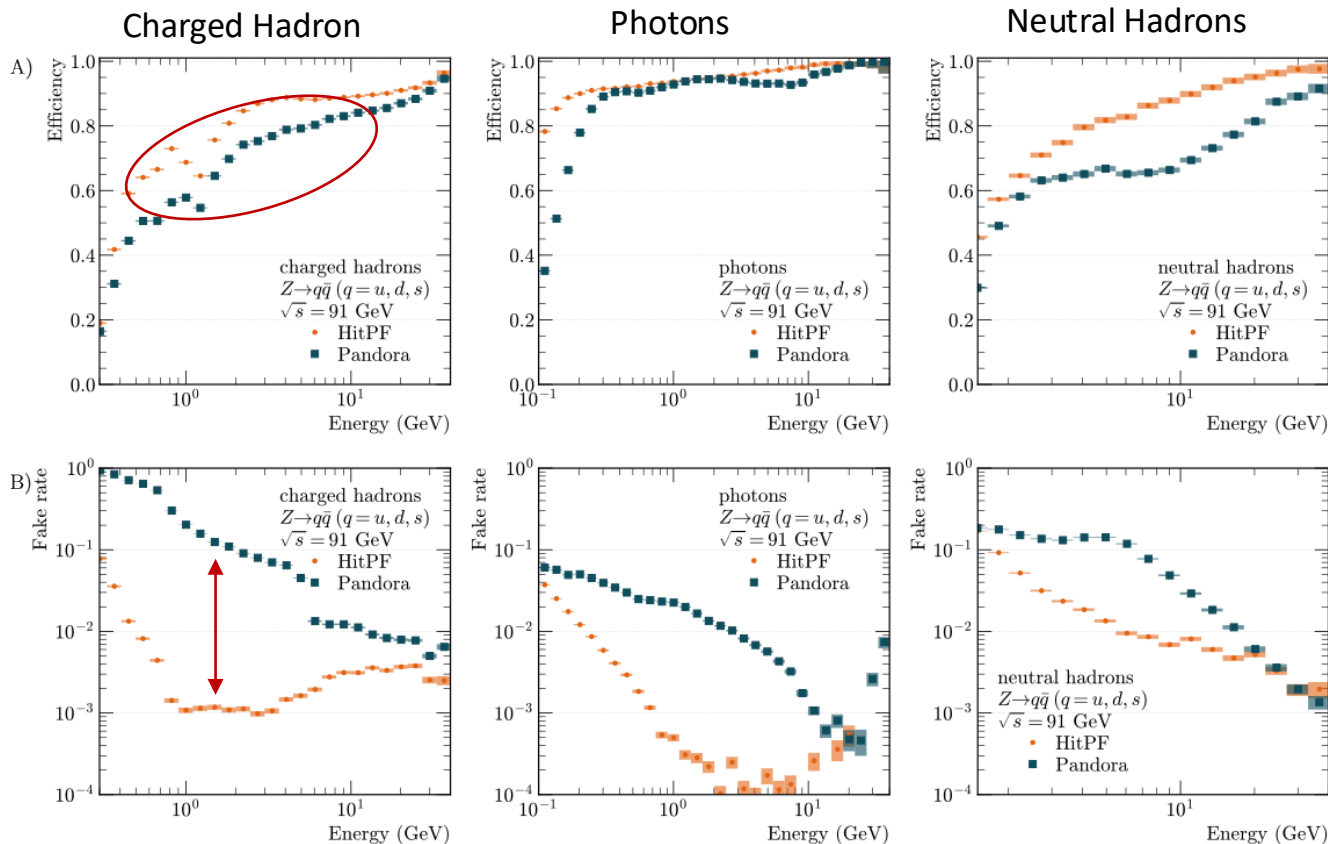
Photons:

- ~5% higher eff. at high E
- Improvement by up to factor 2 for low E
- Fake rate up to $\sim \mathcal{O}(1)$

NH:

- Up to 20% higher eff. with Fake rate reduced by almost factor 5 in this region
- Better clustering (no fake clusters from splitting)

PID: Efficiency and fake rate



CH:

- Up to 20% higher eff.
- Fake rate $\mathcal{O}(2)$ smaller
- Eff. limited by track reco.

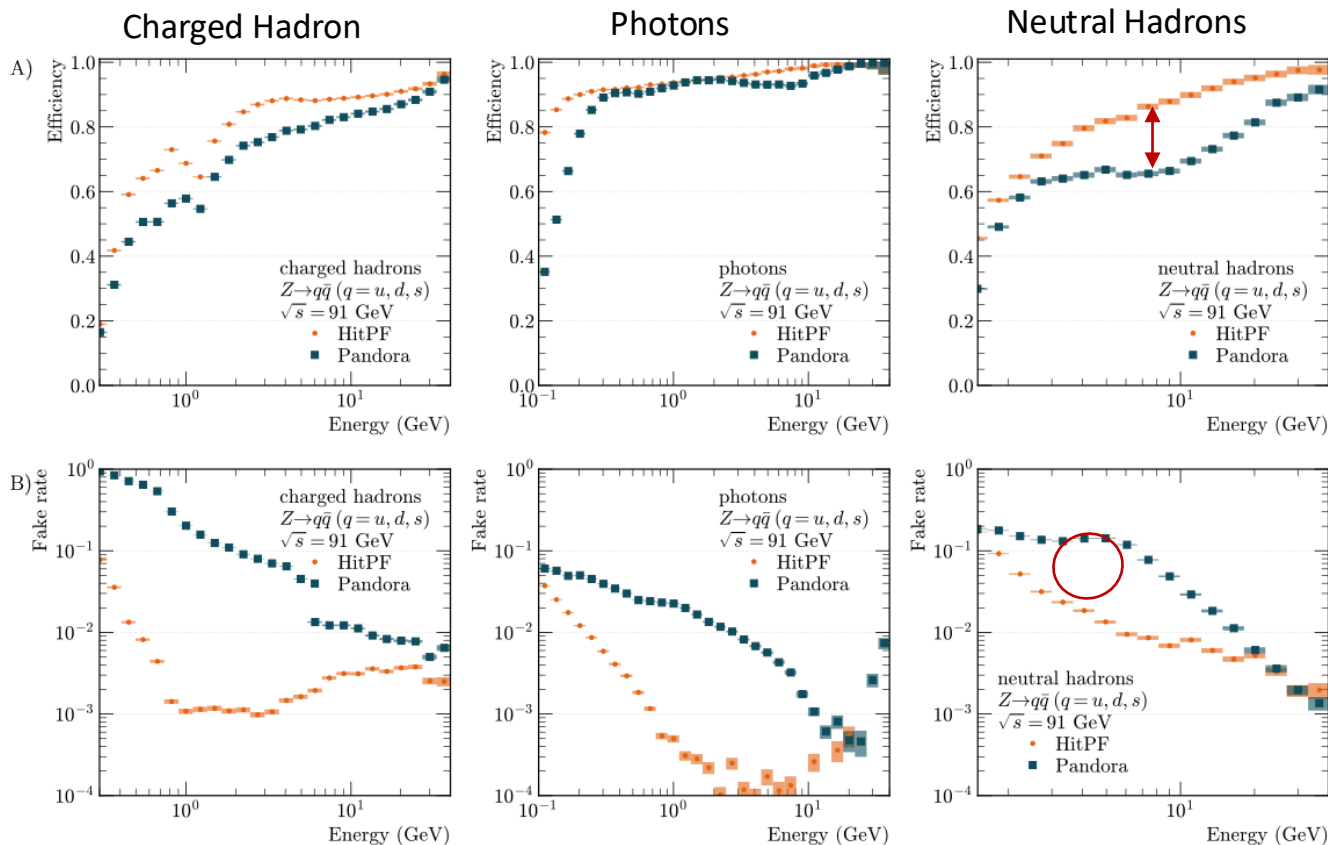
Photons:

- ~5% higher eff. at high E
- Improvement by up to factor 2 for low E
- Fake rate up to ~ $\mathcal{O}(1)$

NH:

- Up to 20% higher eff. with Fake rate reduced by almost factor 5 in this region
- Better clustering (no fake clusters from splitting)

PID: Efficiency and fake rate



CH:

- Up to 20% higher eff.
- Fake rate $\mathcal{O}(2)$ smaller
- Eff. limited by track reco.

Photons:

- ~5% higher eff. at high E
- Improvement by up to factor 2 for low E
- Fake rate up to $\sim \mathcal{O}(1)$

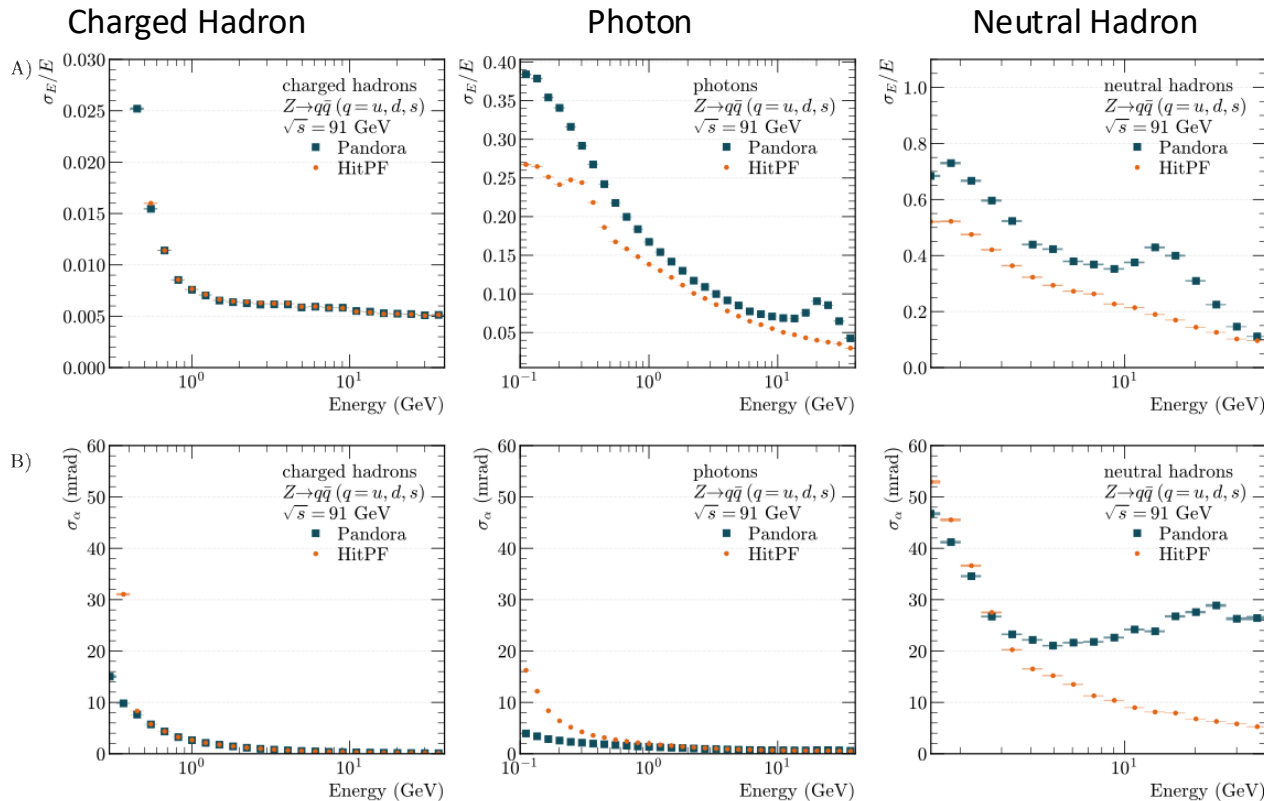
NH:

- Up to 20% higher eff. with Fake rate reduced by almost factor 5 in this region
- Better clustering (no fake clusters from splitting)

Energy and angular resolution

Energy Resolution: half-width of 16th-84th percentile normalized to median

Angular resolution: 68th percentile of the opening angle between rec and true



CH:

- Comparable by construction
- solely tracker information used

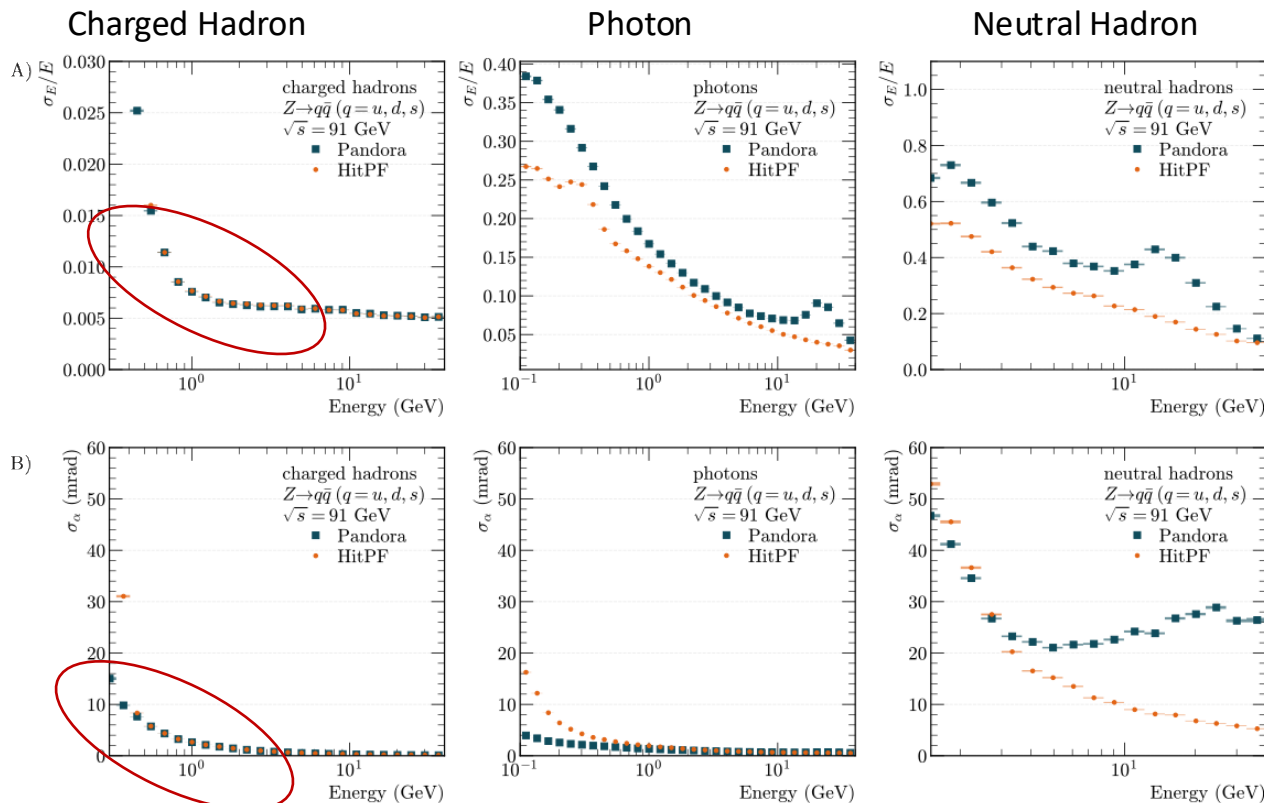
NH:

- Largest gain in energy/ angular resolution
- Less shower merging, better hit capture

Energy and angular resolution

Energy Resolution: half-width of 16th-84th percentile normalized to median

Angular resolution: 68th percentile of the opening angle between rec and true



CH:

- Comparable by construction
- solely tracker information used

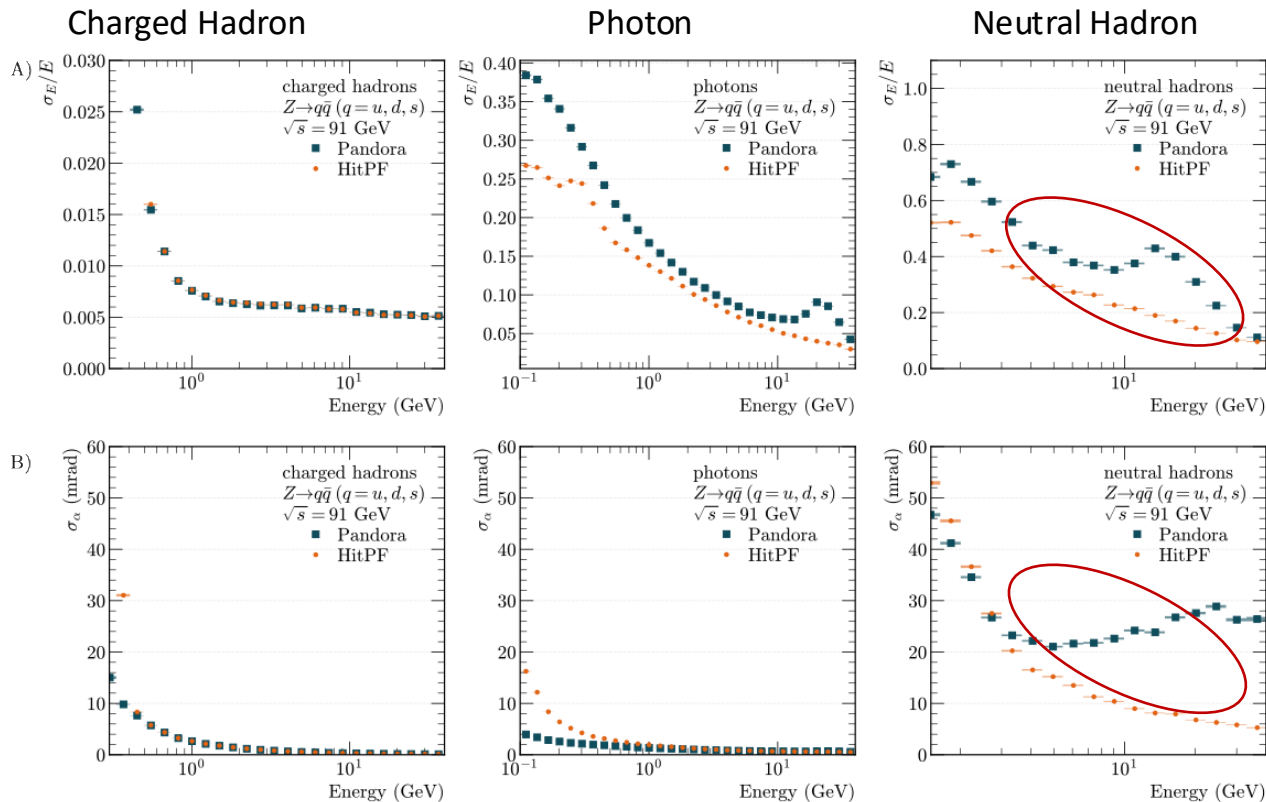
NH:

- Largest gain in energy/ angular resolution
- Less shower merging, better hit capture

Energy and angular resolution

Energy Resolution: half-width of 16th-84th percentile normalized to median

Angular resolution: 68th percentile of the opening angle between rec and true



CH:

- Comparable by construction
- solely tracker information used

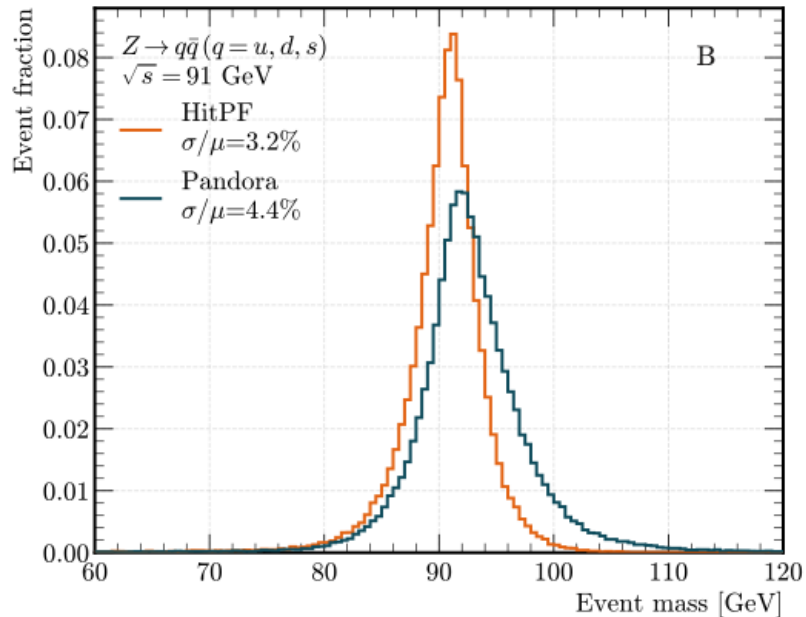
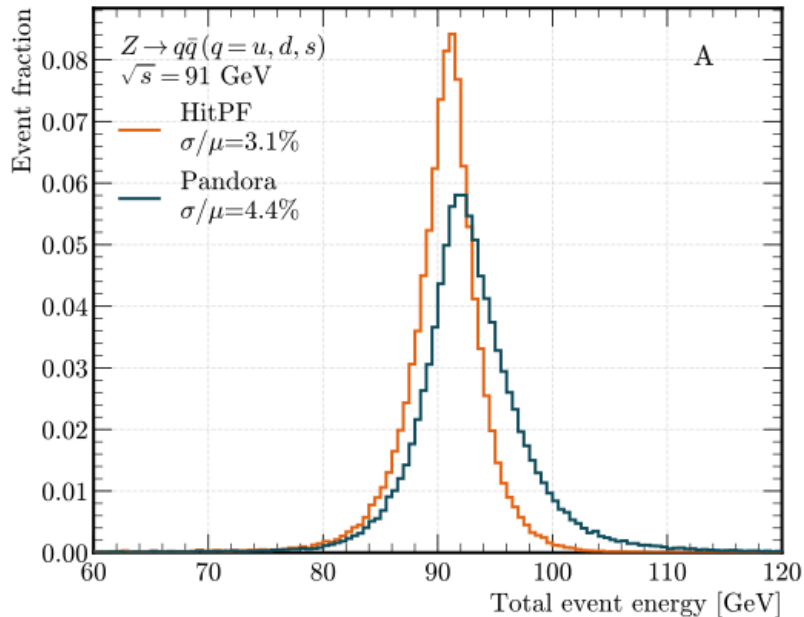
NH:

- **Largest gain in energy/ angular resolution**
- **Less shower merging, better hit capture**

Visible mass and energy

- ❑ Good complementary metric to per-particle metrics
- ❑ Improved mass resolution due to:
 - ❑ Higher efficiency in low energy neutrals
 - ❑ Improved energy resolution
 - ❑ Better track cluster-linking

27% relative improvement
of HitPF compared to
Pandora



ILD detector

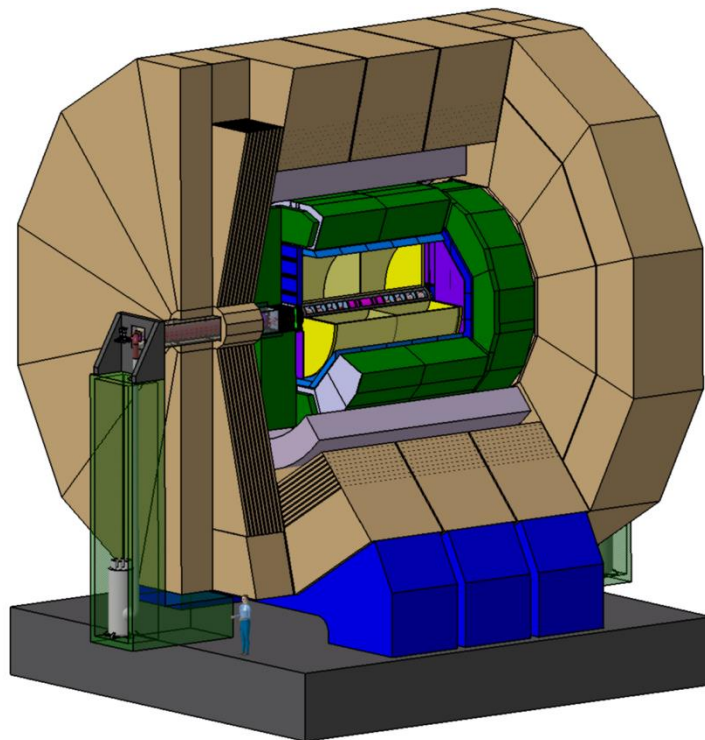


Full simulation (ILD_l5_o1_v02)

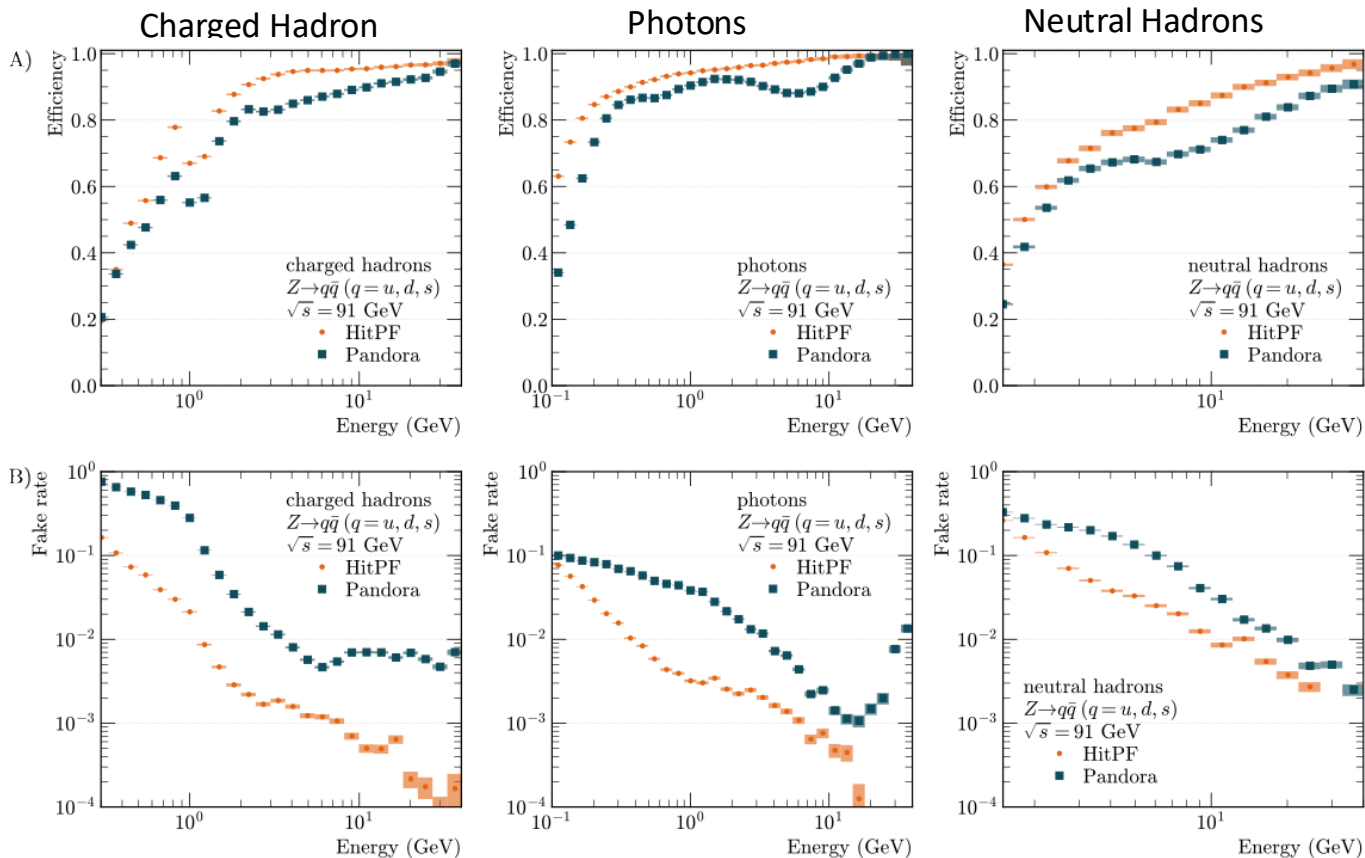
- ☐ Implemented in DD4hep
- ☐ Full simulation in Key4hep & Geant4
- ☐ Comparison againsts Pandora tuned for ILD

- ☐ $Z \rightarrow q\bar{q}, q = (u, d, s)$ 91 GeV
- ☐ Jet like events:
 - ☐ Close particles
 - ☐ High particle multiplicity

- ☐ Compared to CLD, higher field (3.5 T)
- ☐ Both have a highly granular silicon-tungsten ECAL and steel-scintillator HCAL



PID: Efficiency and fake rate (ILD)



CH:

- Up to 20% higher eff.
- Fake rate $\mathcal{O}(2)$ smaller
- Eff. limited by track reco.

Photons:

- ~5% higher eff. at high E
- Improvement by up to factor 2 for low E
- Fake rate up to $\sim \mathcal{O}(1)$

NH:

- Up to 20% higher eff. with Fake rate reduced by almost factor 5 in this region
- Better clustering (no fake clusters from splitting)

Energy and angular resolution (ILD)

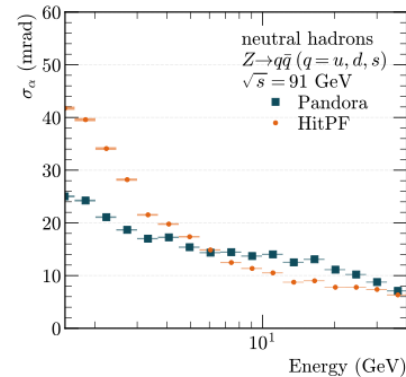
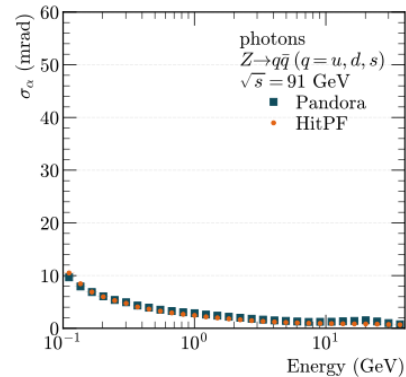
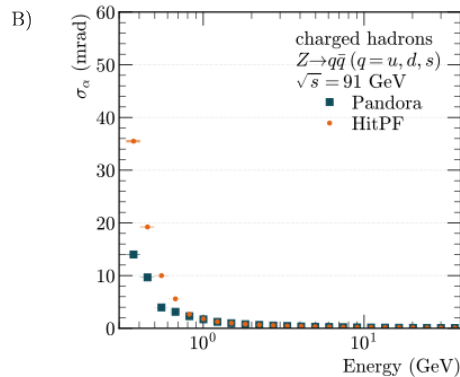
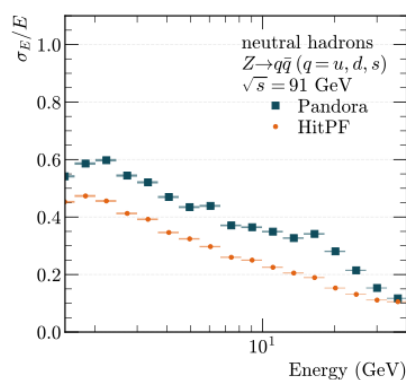
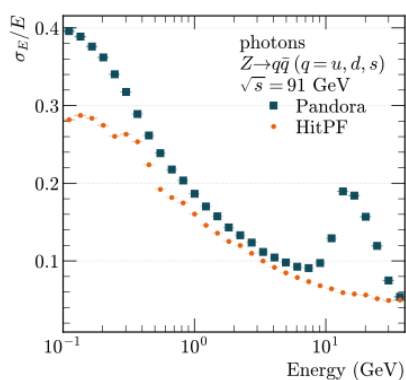
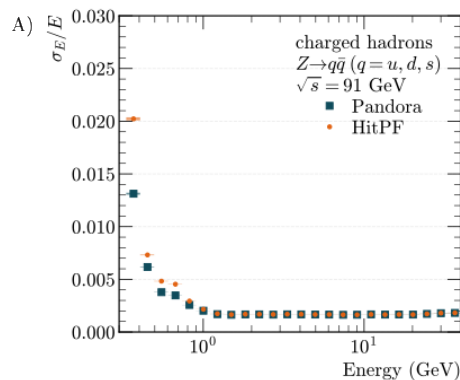
Energy Resolution: half-width of 16th-84th percentile normalized to median

Angular resolution: 68th percentile of the opening angle between rec and true

Charged Hadron

Photon

Neutral Hadron



CH:

- Comparable by construction
- solely tracker information used

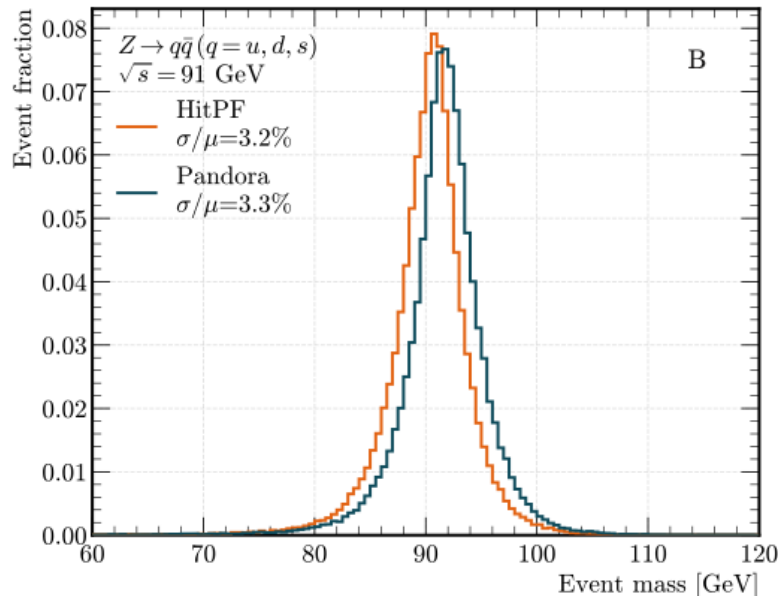
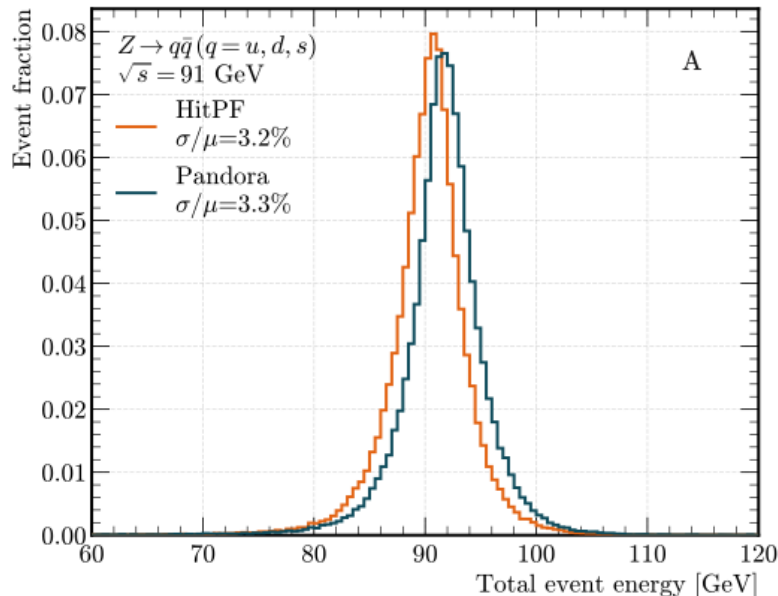
NH:

- Largest gain in energy/ angular resolution
- Less shower merging, better hit capture

Visible mass and energy (ILD)

- ❑ Good complementary metric to per-particle metrics
- ❑ Improved mass resolution due to:
 - ❑ Higher efficiency in low energy neutrals
 - ❑ Improved energy resolution
 - ❑ Better track cluster-linking

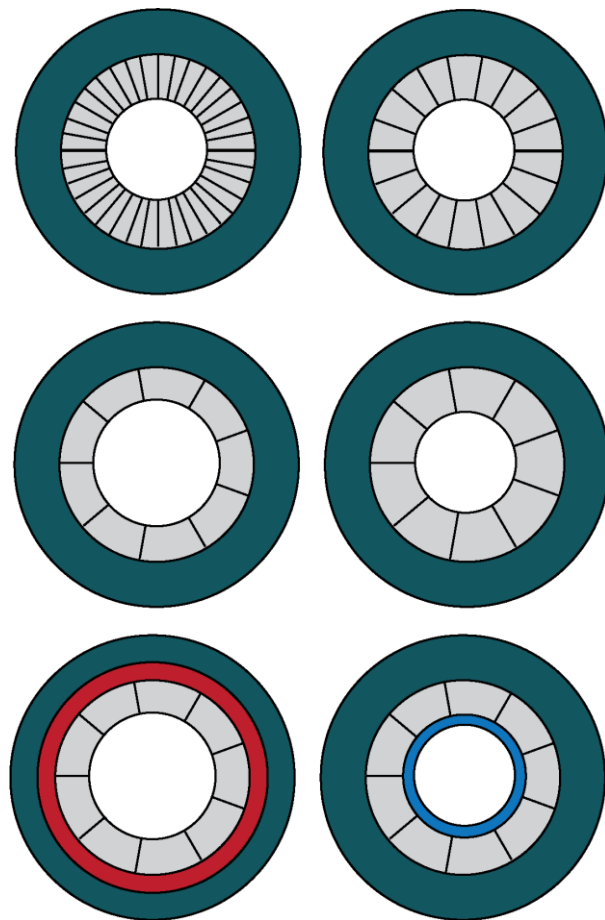
3% relative improvement of HitPF compared to Pandora



What does this give access to?

Full simulation physics studies

- ☐ Impact of beam background
- ☐ Effect of timing information in clustering
- ☐ Impact of low momentum tracks reco on PF
- ☐ Impact of acceptance
- ☐ Tau reconstruction
- ☐ Physics validation
 - Retrain jet taggers with HitPF input
 - Deploy for physics analysis in CLD



Systematic uncertainties and robustness

- ❑ Methods for calculating systematic uncertainties have been developed in HEP (tagging)
- ❑ Three uncertainty sources:
 - ❑ **Statistics** (from data and MC)
 - ❑ **Systematics** (at inputs) eg tracking efficiency, physics modeling..
 - ❑ Calculate systematic uncertainty of variations by varying the parameter and recalculating
 - ❑ Scale factor treated as a nuisance
 - ❑ **Systematics** (model) posterior of the weights distribution (dropout, ensembles)
- ❑ These methods extend to the networks applied in hit-based particle flow
- ❑ Studies are under way to determine the robustness of these algorithms

Extending hit-based models to data

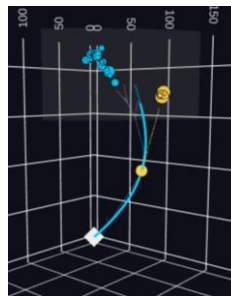
Florian Jonas
David D'Enterria
e+e- alliance

Vincent Richers

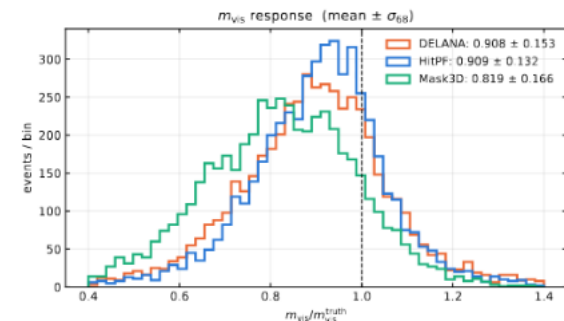
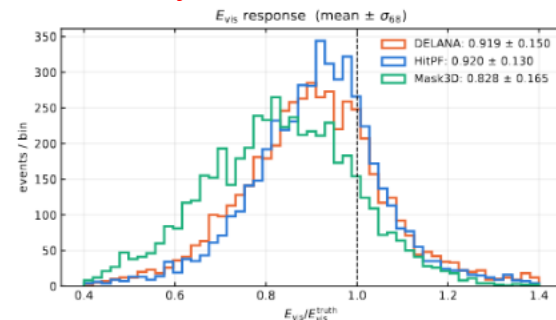
A proxy for FCC reconstruction

- ❑ Raw DELPHI data can test FCC high-level reconstruction algorithms on real detector data
- ❑ Using AI we developed a truth-linking (in FORTRAN!):
 - Required to dump cell recording blocks created during simulation in the DELPHI detector-response cards
 - Recording the parent label for each deposit
 - Created a family tree dictionary walker (for ZEBRA banks) equivalent to `getParent()` method

Truth labeling brems



Preliminary!



Summary



- ❑ Improved performance in PF compared to classical approaches across all metrics: efficiency in complex hadronic topologies, neutral hadron efficiency, fakes and visible mass resolution (improvements by 27-3%)
- ❑ A first version is already available in key4hep for CLD (talk by Lena HLR meeting)
- ❑ Ongoing work to deploy hit-based approaches for other detector concepts (eg. ALLEGRO)
- ❑ Ongoing work with LEP data to determine the MC to data transfer

Thank you to the key4hep team and all FCC collaborators (Michele Selvaggi, Lena Herrmann, Gregor Krzmarc, Bohdan Dudar, Brieuc Francois, Vincent Richers) Jan Kieseler, Alessandro Brusamolino, Katherina Schauble @ KIT & Justin Solomon @MIT



FCC Reconstruction Workshop

taking place from 26–29 October at CERN and online.

Topics:

- Vertexing
- Flavour tagging
- Particle Flow
- Occupancy and beam-induced -background studies

What is there, what is missing, what do physics groups need to move forward?

FCC Reconstruction Workshop

26–29 de octubre de 2026
Europe/Zurich zona horaria

Introduzca su término de búsqueda

Vista general

Horario

Lista de contribuciones

Inscripción

Lista de participantes

This year's FCC Reconstruction Workshop focusses on high-level reco, namely vertexing, flavour tagging and Particle Flow, as well as occupancy and beam-induced -background studies.

The aim is to provide an overview what reco tools are there, what is still missing and discuss what the options are, in particular for common developments across detector concepts. The target are fully implemented reco chains.

On the last day we get together in smaller teams to work hands-on on reco code to address identified gaps.



Empieza 26 oct 2026, 9:00
Finaliza 29 oct 2026, 17:30
Europe/Zurich



CERN



Inscripción

La inscripción de este evento está actualmente abierta.

[Inscribirse ahora](#)



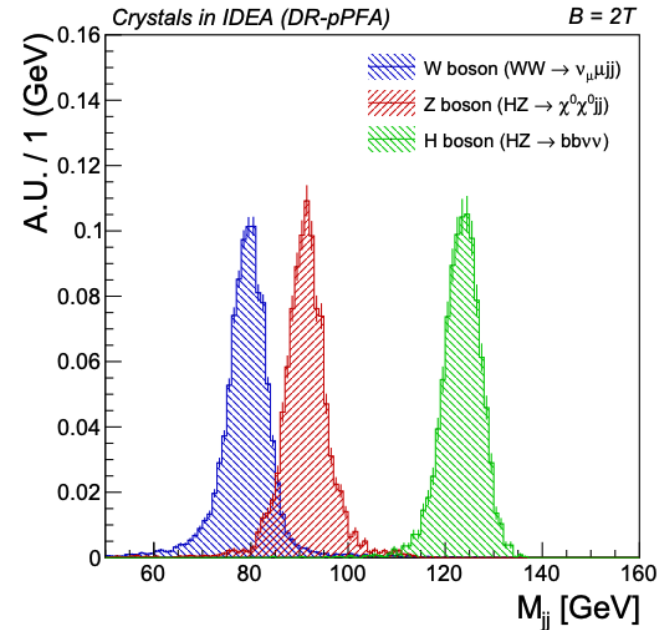
Thank you
for your attention.



Why Particle Flow reconstruction?

From raw data to physics observables

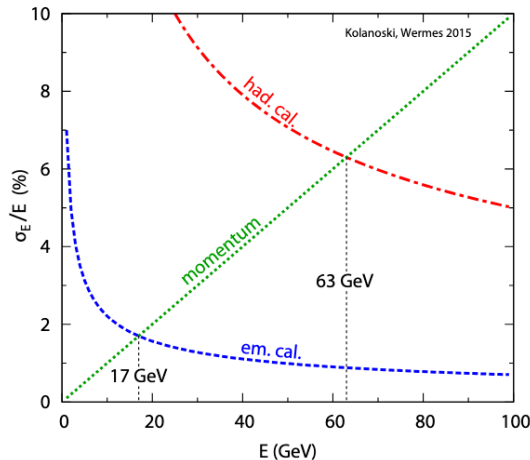
- ❑ Reconstructing individual particles from raw signals connects experimental measurements with physics observables
- ❑ Achievable sensitivity on key SM measurements is driven by the resolution with which the visible final state particles and their invariant masses can be reconstructed
- ❑ Rare higgs decays ($H \rightarrow cc$, $H \rightarrow ss$, relative uncertainty scale is sensitive to visible mass resolution)
- ❑ Hadronic final states large particle multiplicity and geometrical overlap



The need for an end-to-end approach

Increased performance?

- ❑ Asymptotic resolution is 1-2% @100 GeV. Current performance 3-4% @100 GeV [arXiv:0907.3577v1](https://arxiv.org/abs/0907.3577v1)
- ❑ Key drivers of performance:
 - Tracking & calo cluster efficiency/purity
 - → Cluster merging: if two clusters are merged, with a track we lose resolution because we use the total calorimeter energy
 - → Shower splitting: double counting energy
 - Track cluster matching
 - Arbitration of track vs cluster energy



Average Jet Composition

- 60% charged particles
- 30% photons
- 10% neutral particles

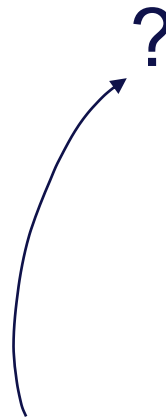
Superior System

- tracker
- ECAL
- HCAL

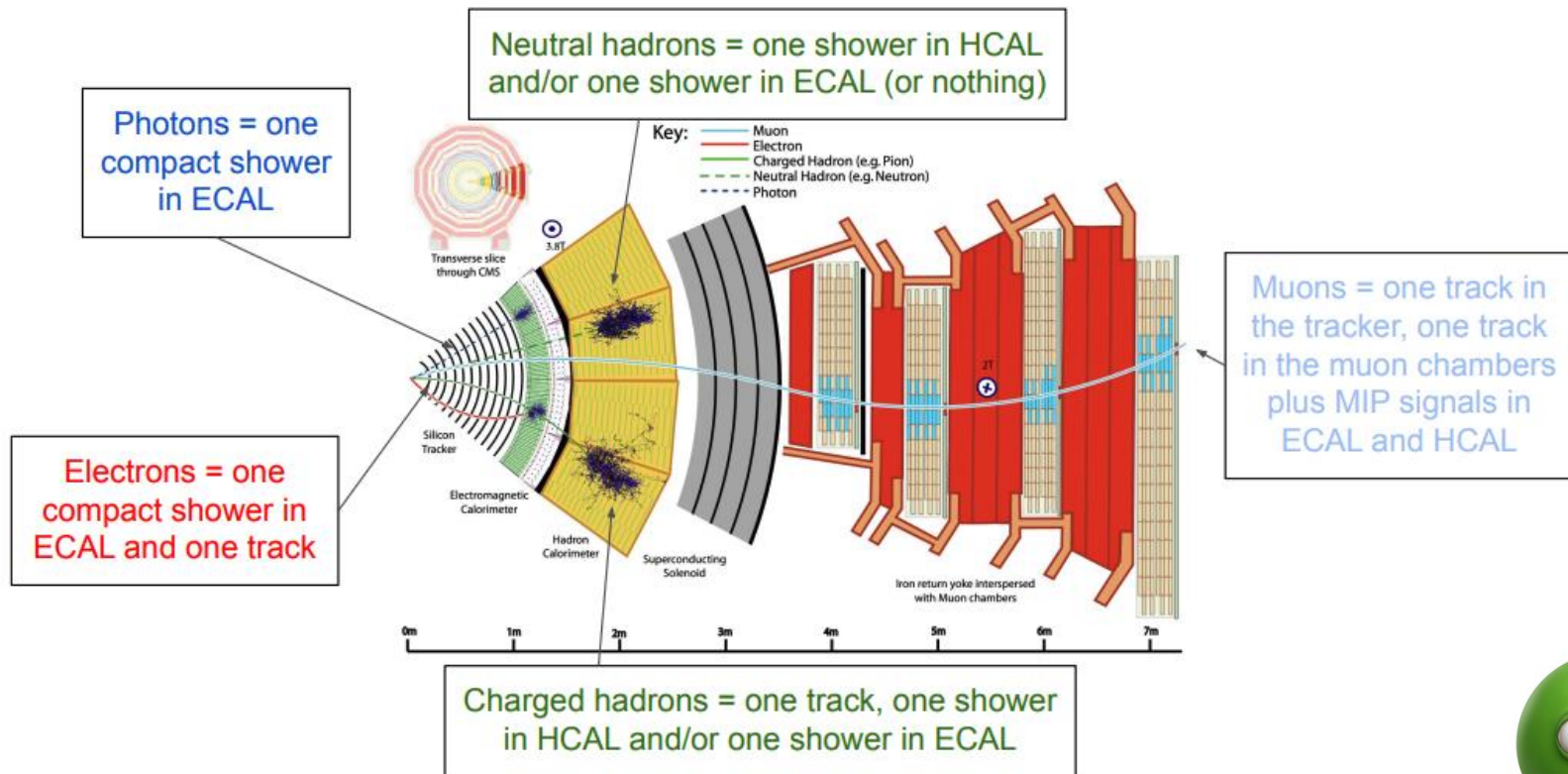
$$\sigma_{jet}^2(E_{vis}) = \sum_{i \in tr} \sigma_{tr}^2(E_{tr}^{(i)}) + \sum_{i \in \gamma} \sigma_{ECAL}^2(E_{ECAL}^{(i)}) + \sum_{i \in h^0} \sigma_{HCAL}^2(E_{h^0}^{(i)})$$

Why is it hard?

- ❑ Classically, multiple algorithms, many moving pieces:
 - Tracking (one per sub-detector), clustering (ECAL/HCAL), and linking tracks and clusters
- ❑ We want every particle in the event, but not fakes, hard to tune
- ❑ **Fake tracks**
- ❑ **Lonely tracks** (tracks that are real but left no clusters)
- ❑ **Displaced tracks** (less hits, tracking classically not tuned for that)
- ❑ Physics:
 - Photons conversion in the tracker (detached vertex + e^+e^- , maybe two tracks maybe none)
 - Electrons radiation ('double counting')
 - Hadrons interaction (detached vertex + more particles)
 - Neutral hadrons decay (e.g., K_S^0 or Λ)
 - Many edge cases that are rare (hard to take all into account)



The 'game'



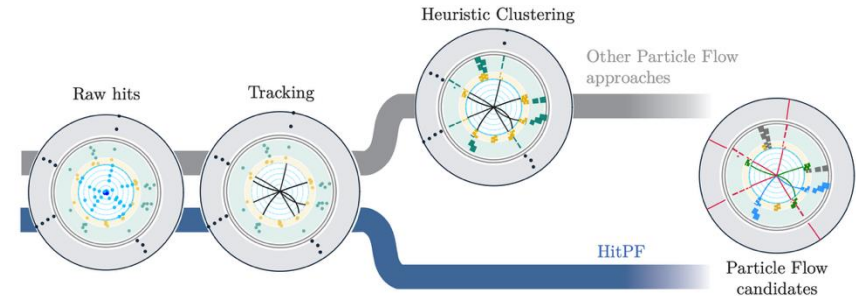
Comparison with other methods

Other PF approaches

- ☐ Require heuristic clustering step + cluster track matching
- ☐ Complex parameter tuning for new detector
- ☐ Requires geometry navigation

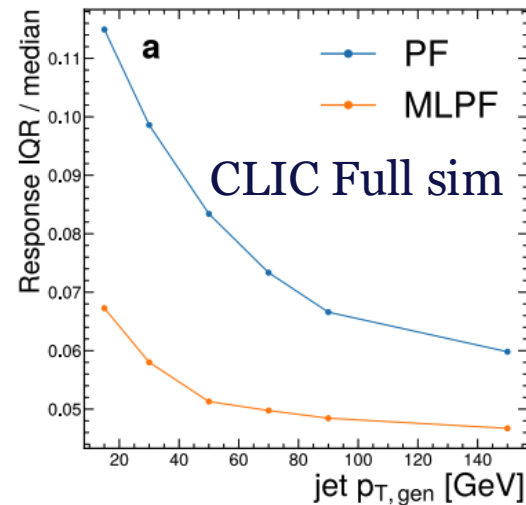
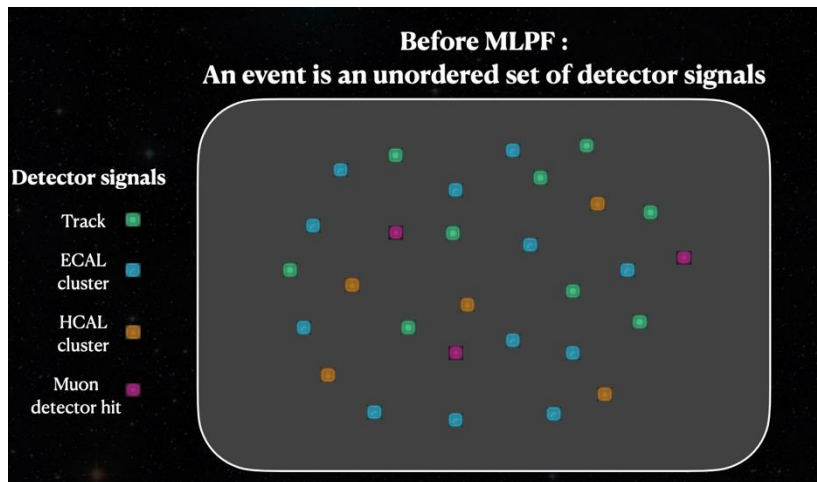
HitPF

- ☐ ML-based Clustering + track matching
- ☐ Out of the box with multiple detector and hit geometries
- ☐ Does not require the explicit detector geometry (navigation)



MLPF

- ❑ Cluster based approach:
 - Starts from pandora clusters and reconstructed tracks
 - In practice, very different problem
 - Improves the track-cluster linking and properties regression
 - 1 node in MLPF ~ 600 nodes in HitPF
 - Does not solve the geometry adaptation problem
 - New results for CMS



J.Pata et al 2024

Know problem in ML

Classification



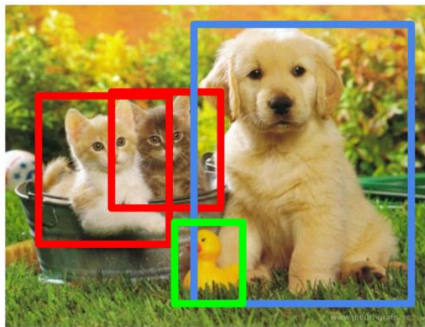
CAT

Classification + Localization



CAT

Object Detection



CAT, DOG, DUCK

Instance Segmentation



CAT, DOG, DUCK

Single object

Multiple objects

Example 1: Clustering block in CLD vs **IDEA** vs ALLEGRO

DR fiber calo

- Capillary tube w/ 2 mm diameter
- $\lambda_{\text{int}} \sim 25 \text{ cm}$

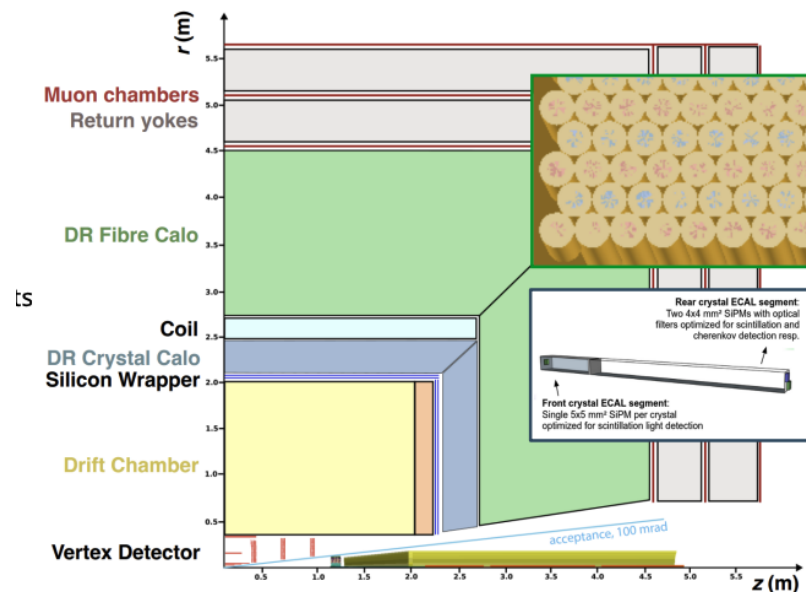
DR crystal calo

- PbWO4 crystals (baseline) w/ Cherenkov light detection using UV filters and/or timing differentiation
- 1x1 cm² for each front (6X0) and rear (16X0) end segments
- **Moliere radius** $\sim 2 \text{ cm}$

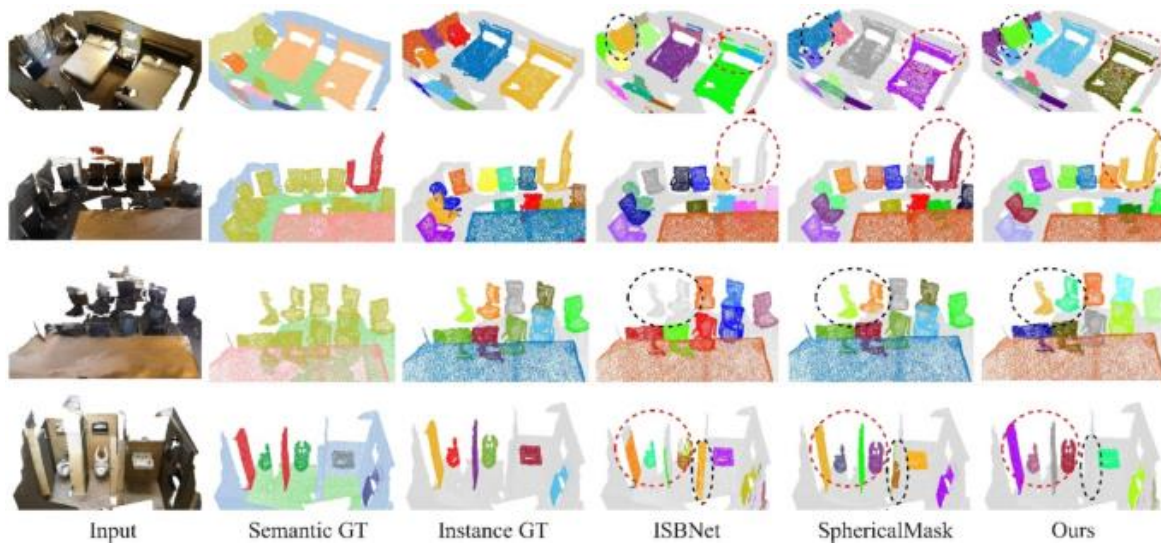
Solenoid

- 1.5X0 transparency based on arXiv:2505.21223 (however, the actual implementation in k4geo is a thin Aluminum can with 0.75X0)
- 30 cm gap is allocated between the ECAL & HCAL
- For now, tracks are taken from the TracksFromGenParticles collection

@Sanghyun Ko



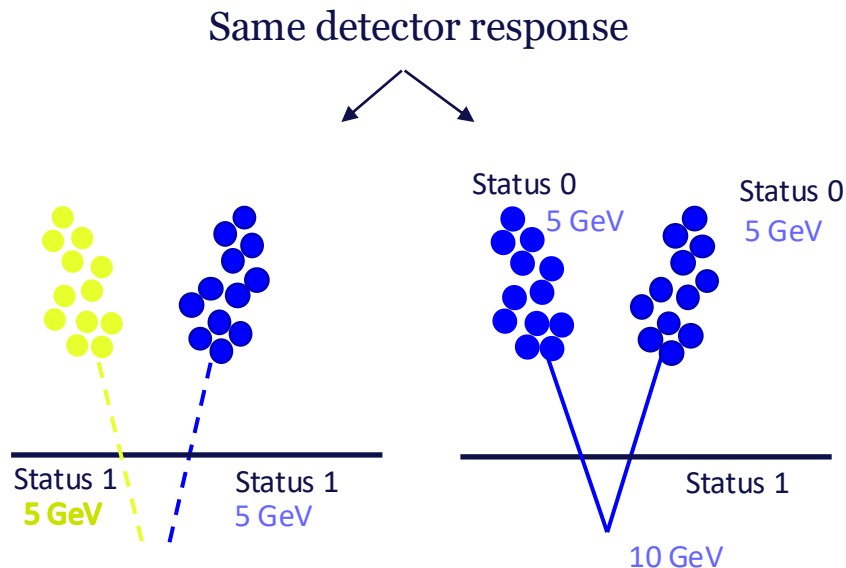
Know problem in ML



Supervision: Target definition

What is the best reconstruction target?

- ❑ Ideally, we would recover the stable particles that entered the detector (status 1) (previous studies)
- ❑ BUT this gives the algorithm conflicting training examples (why this is hard slide)
 - Conversions
 - Backscattering (clusters appear far in the detector)
- ❑ The **target** is defined as:
 - Resolvable detector objects
 - Ongoing work to add detector resolution
 - Defined in practice by the calorimeter boundary
 - This is what it is supervisable
 - And means that we are solving the L1 problem



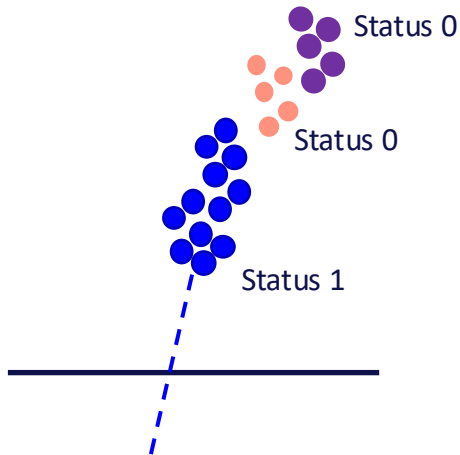
Different label?

Target definition

What is the best reconstruction target?

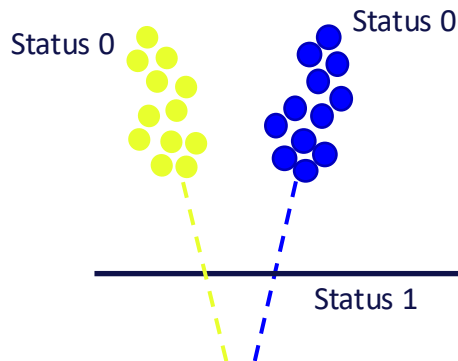
L0

- all Status 0 and status 1 particles eg. Hadronic showers resulting from nuclear interactions



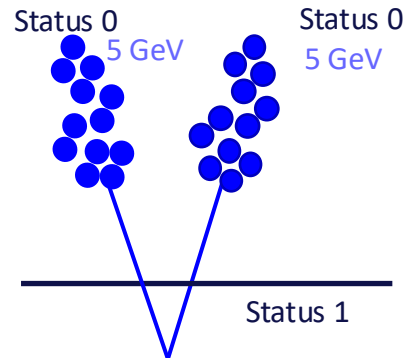
L1

- Calorimeter boundary abstraction
- Reconstruct two clusters for conversions, etc



L2

- Status 1
- Complex topologies
- Brems, kinks, K_S^0



The role of AI

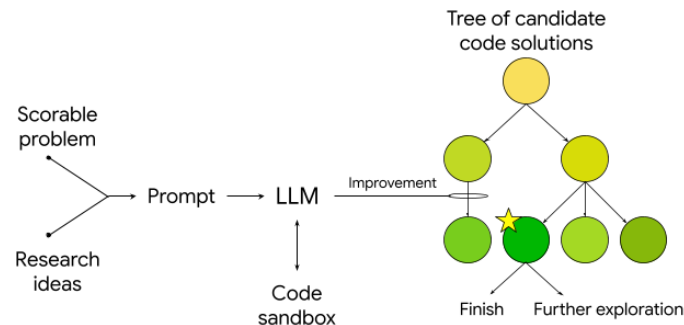
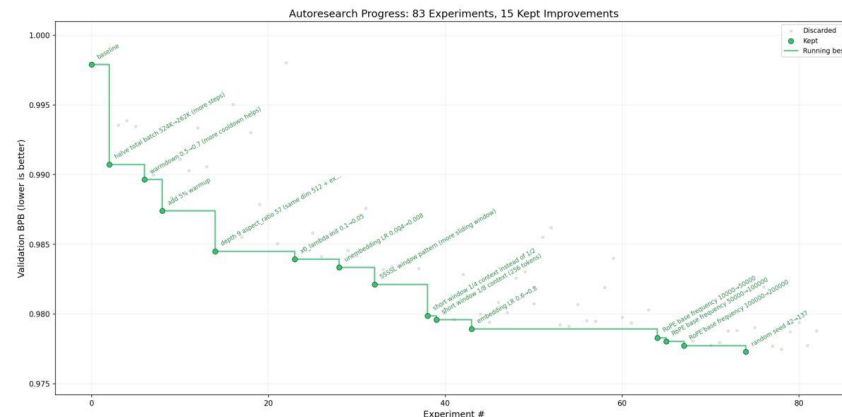
AI is being used to further improve scientific codebases

❑ autoresearch (Andrej Karpathy):

- Agents modify architecture, optimizer, hyperparameters to improve a loss function
- Fast, fixed-budget trial-and-error loop experiments result in improved loss

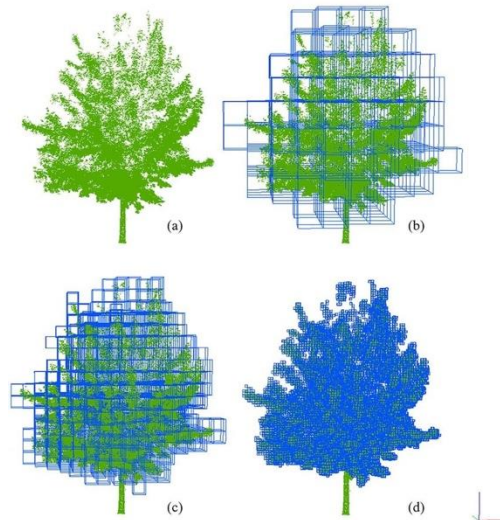
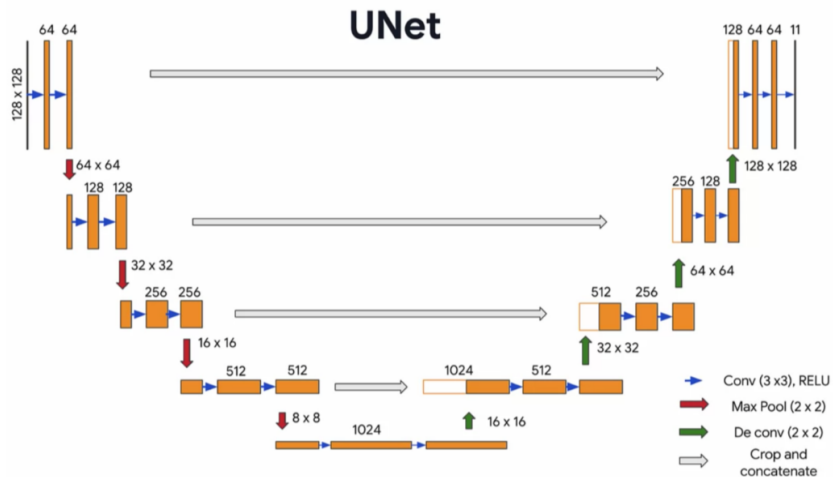
❑ LLMs + Tree search for scientific software (Google, Nature):

- Automatically write and iteratively improve scientific code that maximizes a quality metric
- Applied to multiple fields (Covid prediction to geospatial segmentation) obtains improved results over baselines
- A large share of the top-performing results came from recombination



The 3D ML architectures

UNet backbone + with voxelization



Inductive bias

Assumptions about the relationship between the input and output of a learning problem that are built into the model

Can improve generalization

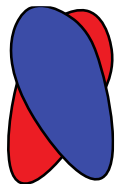
Can reduce required dataset size

Adapting ML methods to PF

Inductive bias

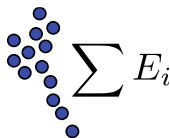
Self- intersections

No equivalent in 3D ML dataset



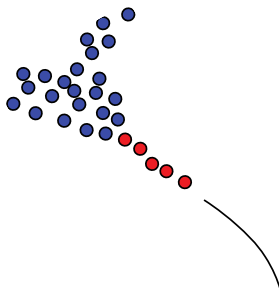
Multiple measurements in different subdetectors that 'add up' --> Optimal transport

No equivalent in 3D ML datasets (LiDAR scans of chairs do not add up to their mass)



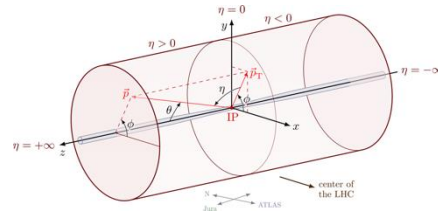
Need to preserve 'fine' structures in ECAL/tracker

Prevents voxelization and standard pulling operations



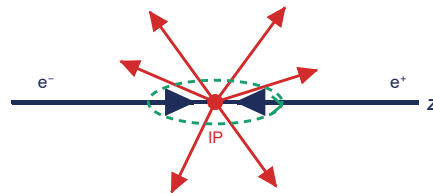
SO(2) + z mirror (endcaps)

~E(3) rotations, translations, and reflections



Beam direction

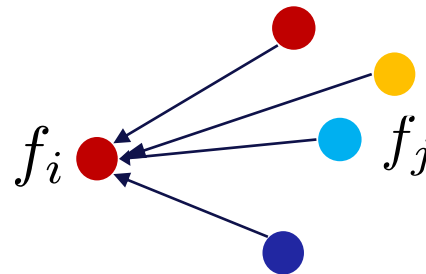
Preferred direction set by gravity



ML 3D data

Detector data

Model GNNs, transformers



$$f'_i = \square_{j:(i,j) \in \mathcal{E}} h_{\Theta}(\mathbf{f}_i, \mathbf{f}_j).$$

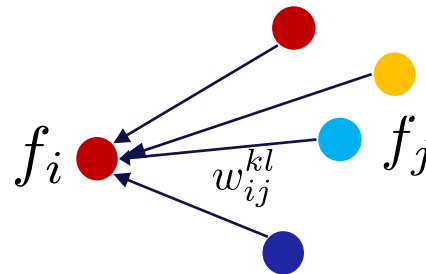
Building block 1

$\square_{j:(i,j) \in \mathcal{E}}$ Aggregation operator

h_{Θ} Non-linear function with learnable parameters

Model GNNs, transformers

$$f'_i = \square_{j:(i,j) \in \mathcal{E}} h_{\Theta}(\mathbf{f}_i, \mathbf{f}_j).$$



Aggregation operator

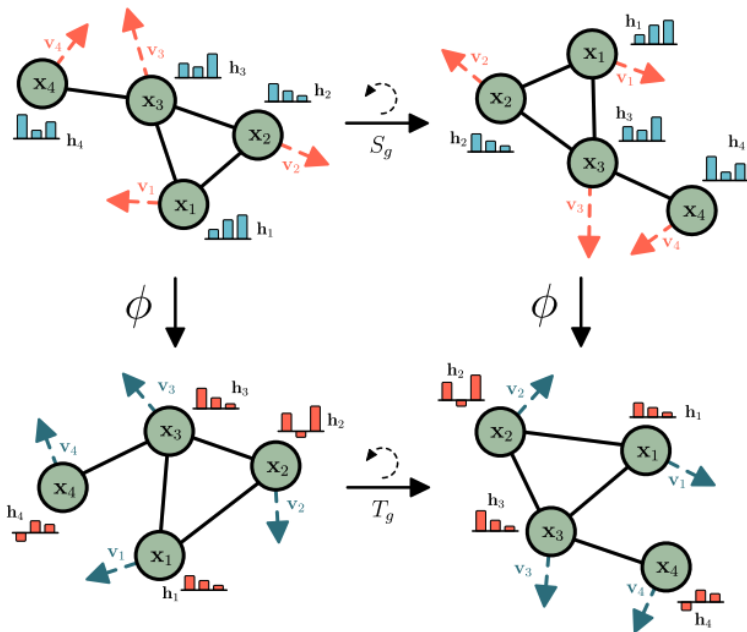
$$\square_{j:(i,j) \in \mathcal{E}} \longrightarrow f'_i = \sum_{j:(i,j) \in \mathcal{E}}$$

Non-linear function with learnable parameters

$$h_{\Theta} \longrightarrow h_{\Theta}(\mathbf{f}_i^l, \mathbf{f}_j^l) = w_{ij}^{kl} \mathbf{V}^{kl} \mathbf{f}_j^l,$$

$$w_{ij}^{kl} = \text{softmax}_j \left(\frac{(\mathbf{Q}_q^{kl} \mathbf{f}_i^l)^{\top} (\mathbf{K}_k \mathbf{f}_j^l)}{\sqrt{d}} \right)$$

Equivariant networks



Building block 2

A symmetry group $\mathfrak{G}$
 With a group representation

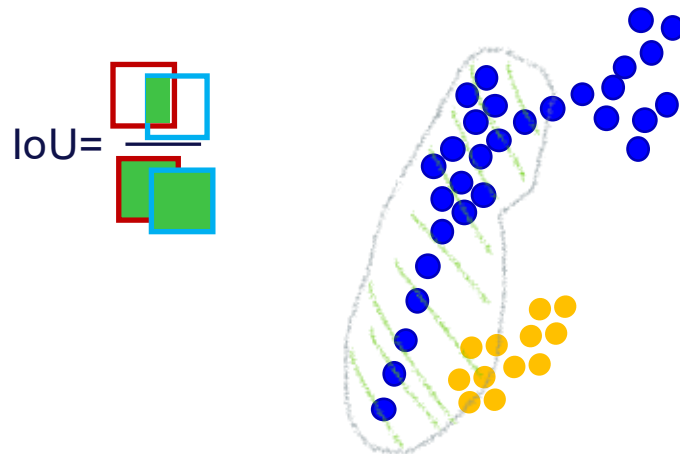
$$\rho : \mathfrak{G} \rightarrow GL(\mathcal{X}(\Omega))$$

Equivariant if

$$f(\rho(g)x) = \rho(g)f(x)$$

Matching

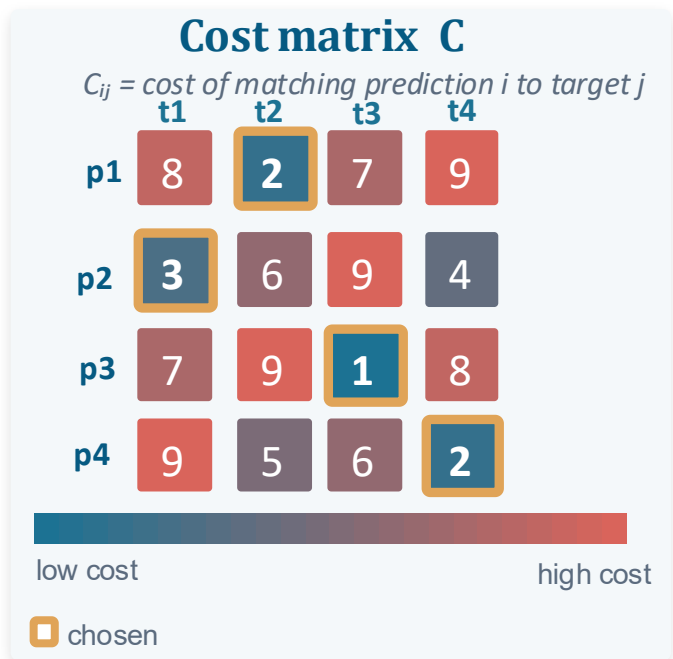
- We use a **Hungarian matching** to match a reco particle to a target particle with the intersection over union cost
- Builds a cost matrix
- One-to-one constraint
- Cost minimization



Matching

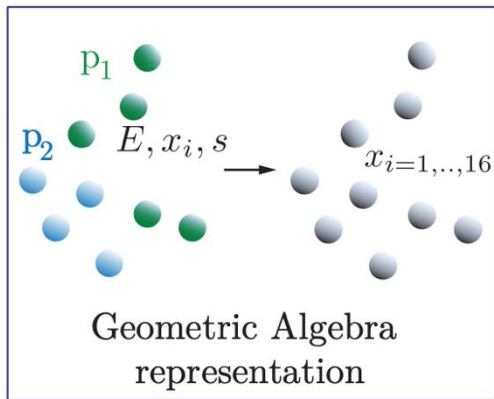
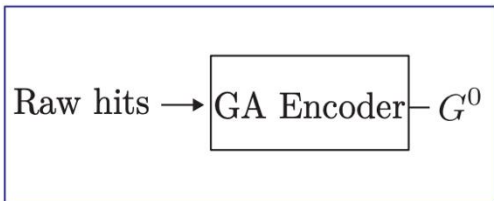
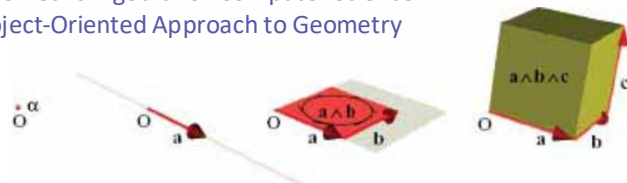
- We use a **Hungarian matching** to match a reco particle to a target particle with the intersection over union cost
- Builds a cost matrix
- One-to-one constraint
- Cost minimization

$$\min_{X \in \{0,1\}^{N \times N}} \sum_{i=1}^N \sum_{j=1}^N C_{ij} X_{ij} \quad \text{s.t.}$$
$$\sum_j X_{ij} = 1 \quad \forall i, \quad \sum_i X_{ij} = 1 \quad \forall j.$$



Multivectors

Geometric Algebra for Computer Science : An Object-Oriented Approach to Geometry



- Representation of multiple geometric objects of different grades
- Outer product $a \wedge b$ (oriented span of a and b)
- Projective algebra (adds displacement from the origin)
- 3D object representation:
 - Planes are represented as vectors (normal + displacement)
 - Lines are intersection of two planes (bivectors)
 - Points are intersections of three planes (trivectors)
- Multivectors:
 - $x = x_s + x_1 e_1 + x_2 e_2 + x_3 e_3 + x_{12} e_1 e_2 + x_{13} e_1 e_3 + x_{23} e_2 e_3 + x_{123} e_1 e_2 e_3$, x_i real coefficients
 - Built with the geometric product (metric)

Object / operator	Scalar 1	Vector $e_0 \ e_i$		Bivector $e_{0i} \ e_{ij}$		Trivector $e_{0ij} \ e_{123}$		PS e_{0123}
Scalar $\lambda \in \mathbb{R}$	λ	0	0	0	0	0	0	0
Plane w/ normal $n \in \mathbb{R}^3$, origin shift $d \in \mathbb{R}$	0	d	n	0	0	0	0	0
Line w/ direction $n \in \mathbb{R}^3$, orthogonal shift $s \in \mathbb{R}^3$	0	0	0	s	n	0	0	0
Point $p \in \mathbb{R}^3$	0	0	0	0	0	p	1	0
Pseudoscalar $\mu \in \mathbb{R}$	0	0	0	0	0	0	0	μ
Reflection through plane w/ normal $n \in \mathbb{R}^3$, origin shift $d \in \mathbb{R}$	0	d	n	0	0	0	0	0
Translation $t \in \mathbb{R}^3$	1	0	0	$\frac{1}{2}t$	0	0	0	0
Rotation expressed as quaternion $q \in \mathbb{R}^4$	q_0	0	0	0	q_i	0	0	0
Point reflection through $p \in \mathbb{R}^3$	0	0	0	0	0	p	1	0

Geometric Algebra Transformer

Equi
linear

K-blade projections, no grade mixing

$$\phi(x) = \sum_{k=0}^{d+1} w_k \langle x \rangle_k + \sum_{k=0}^d v_k e_0 \langle x \rangle_k$$

Geo
bilinear

Geometric product $x, y \rightarrow xy = \langle x, y \rangle + x \wedge y \rightarrow$ grade mixing

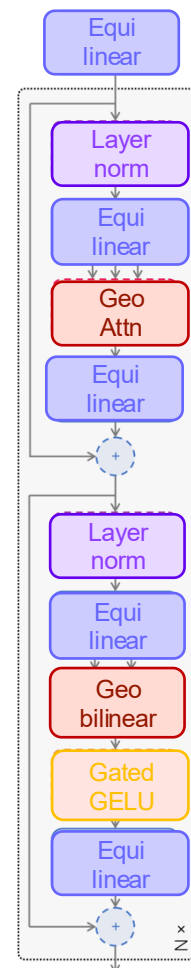
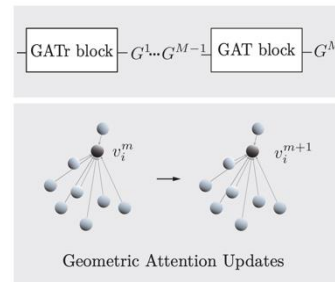
Join product $x, y \rightarrow (x^* \wedge y^*)^*$

$$\text{Geometric}(x, y; z) = \text{Concatenate}_{\text{channels}}(xy, \text{EquiJoin}(x, y; z))$$

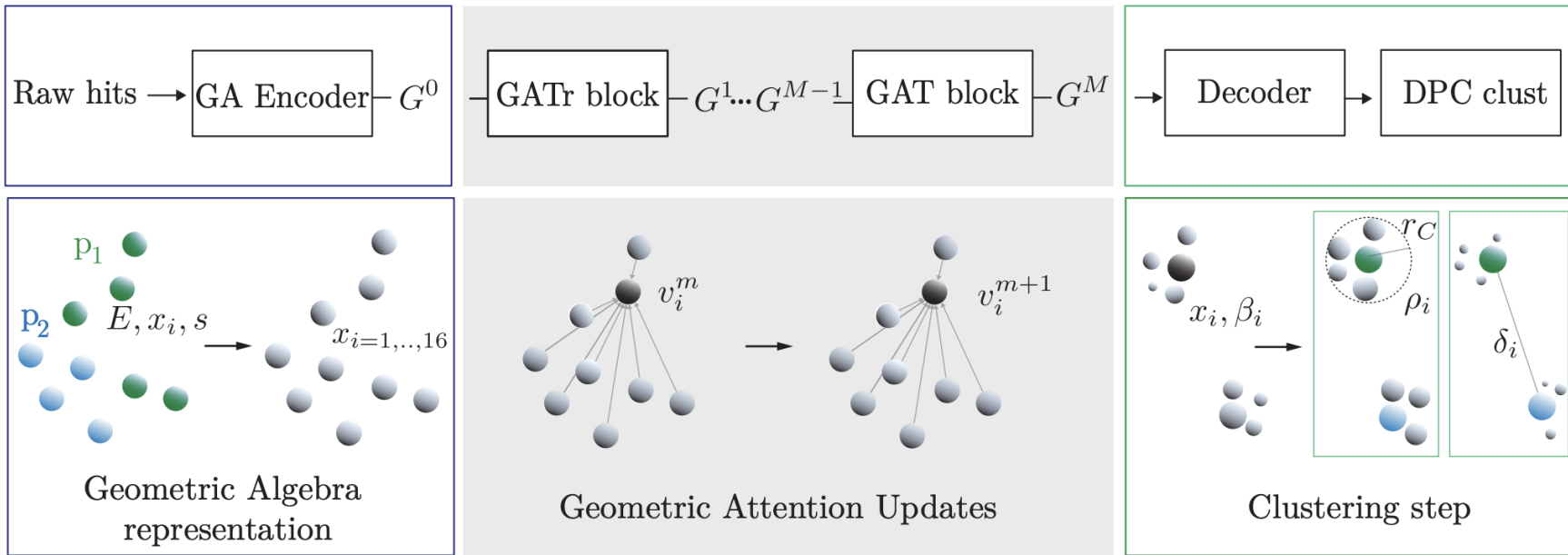
Geo
Attn

Attention mechanism as described previously but with inner product of the algebra

$$\text{Attention}(q, k, v)_{i'c'} = \sum_i \text{Softmax}_i \left(\frac{\sum_c \langle q_{i'c}, k_{ic} \rangle}{\sqrt{8n_c}} \right) v_{ic'}$$



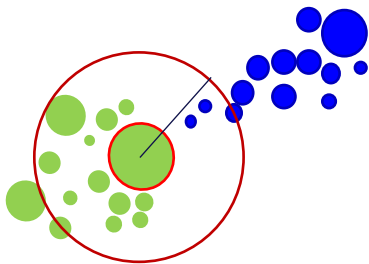
Geometric Algebra Transformer



Clustering in embedding space

Two methods

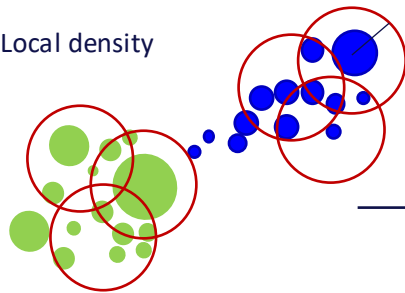
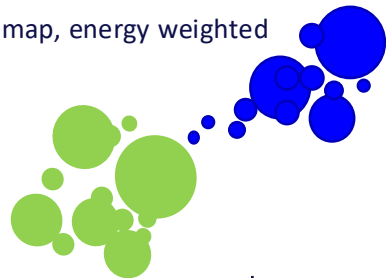
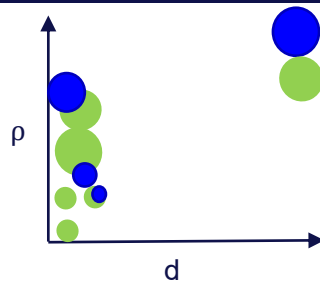
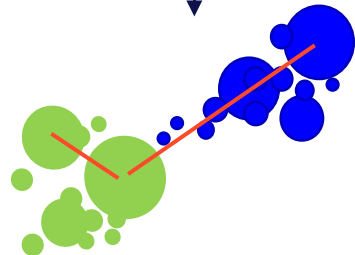
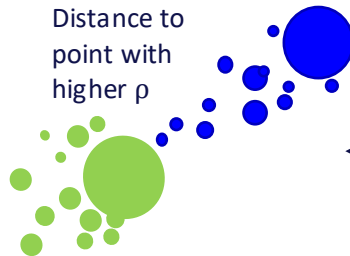
Beta based clustering [1]

 x_j, β_j


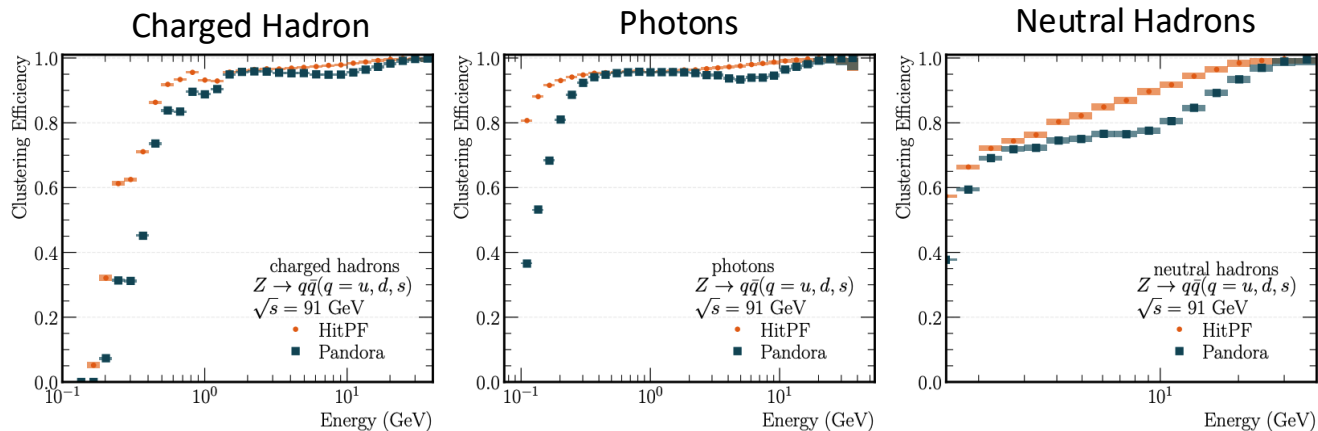
- Two parameters, $\beta_{\min}$, distance threshold
- Depends on good training of β distribution

Density Peak clustering [2]

Local density


 ρ map, energy weighted

Distance to point with higher ρ


Clustering performance

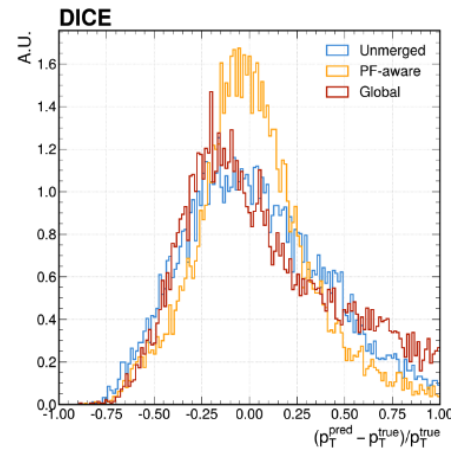
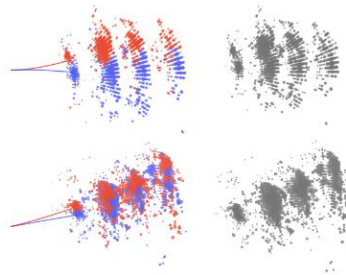


- It is efficient if for a given target particle it finds a reco particle (no PID), purely a measure of clustering performance
- vs the energy of the deposited hits (reco)
- Higher efficiency for all momentum range, substantial at energies below 1 GeV

The impact of the truth definition

Detector aware-target

- ❑ Define a shower merging algorithm based on a resolvability score and a connection score
 - Resolvability: what fraction of shower is not shared
 - Connection: show strongly shower B overlaps with A
- ❑ Graph with sim showers as nodes and possible absorptions as directed edges
- ❑ Algorithm to determine merging steps

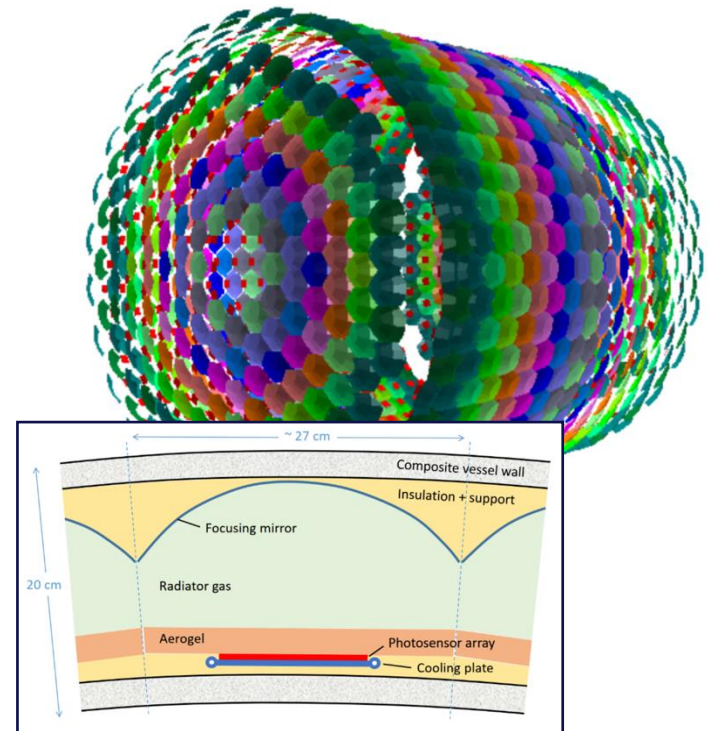


Transverse momentum response
in tau jet sample for models
trained with different target
definition

Impact of the ARC's material

- ❑ ARC adds charged-hadron PID capability to CLD, but its integration places material upstream of the ECAL and compresses the tracker
- ❑ Compact RICH-based PID concept for charged-hadron separation ($\pi/K/p$) via Cherenkov radiation.
- ❑ Uses gas and aerogel radiators with spherical focusing mirrors and SiPM-based photon detection.
- ❑ Tracker is compressed to accommodate ARC: outer radius reduced from 2145 to 1900 mm. Studies have assessed impact on track resolution but not effect on PF

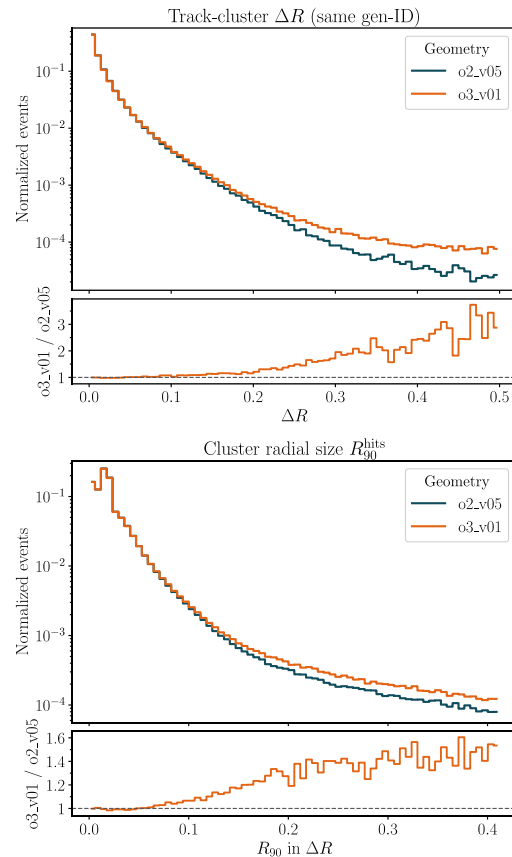
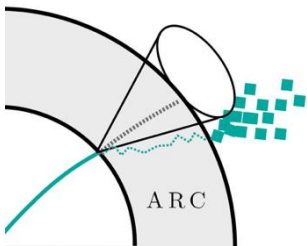
Vincent Richers



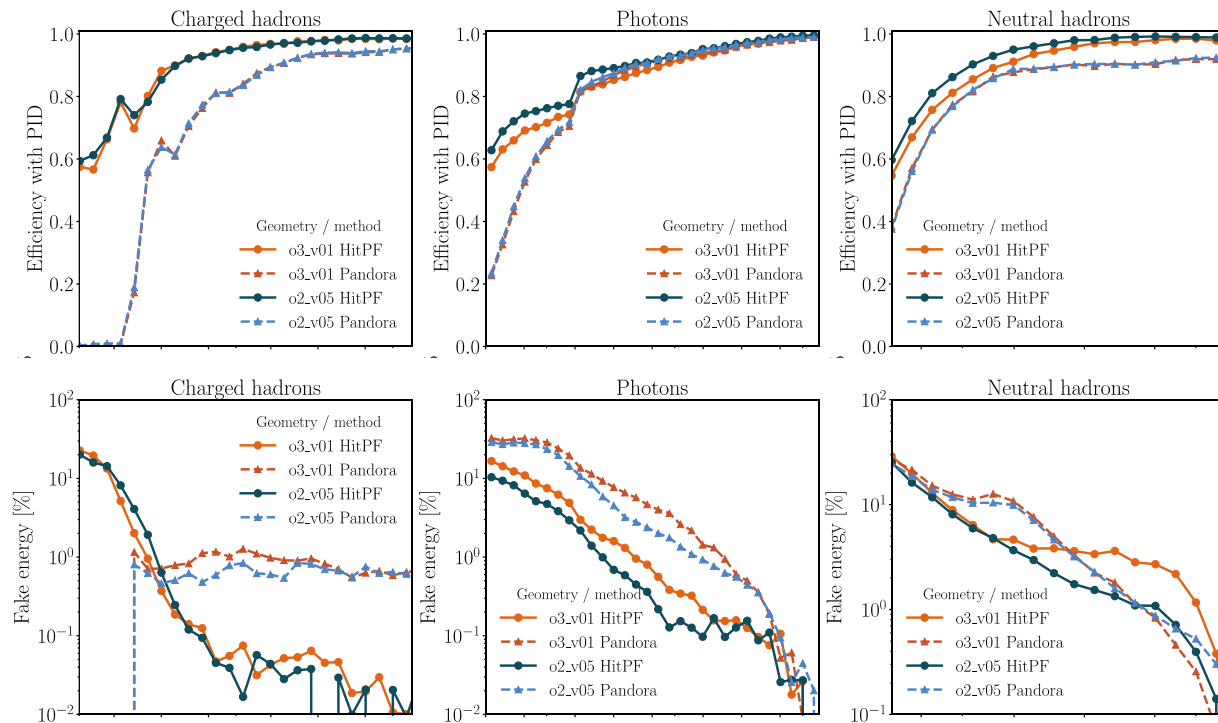
The ARC

Impact of the ARC's material

- ❑ Tracker is shrunk resulting in impact on momentum resolution [1] (track lever arm) (~15%) not necessarily a big impact on PF in itself
- ❑ But tracking-cluster linking might be impacted (multiple scattering)
- ❑ Tracking efficiency is also impacted
- ❑ R90 is impacted by earlier showering



Impact of the ARC's material

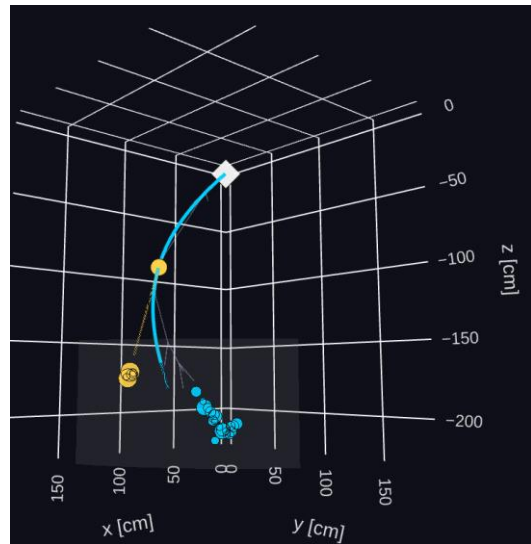


- ❑ Charged-particle reconstruction remains stable once tracking effects are controlled.
- ❑ The visible ARC impact is concentrated in photons and neutral hadrons.
- ❑ Photons and neutral hadrons show reduced PID efficiency, increased fakes, and broader response

Raw data from DELPHI

Working on legacy code (hard to build expertise)

- ❑ Cell recording blocks per sub-detector
 - Written during simulation in the DELPHI detector-response cards
 - Dumped at the end per event (addition to .fadsim)
- ❑ Parent label
 - Record a label for each deposit
 - New block added to delsim36.car an array running parallel to the DELSIM stack
- ❑ Family tree to dictionary walker
 - Walks the ZEBRA bank simulation tree and prints a plain-text dictionary
 - Original Pythia particle link
 - Added the equivalent to getParent() method and store vertex



MC particles (68)

Hits coloured by: calo-mother (ygen hit: PYTHIA primary that entered the calo)

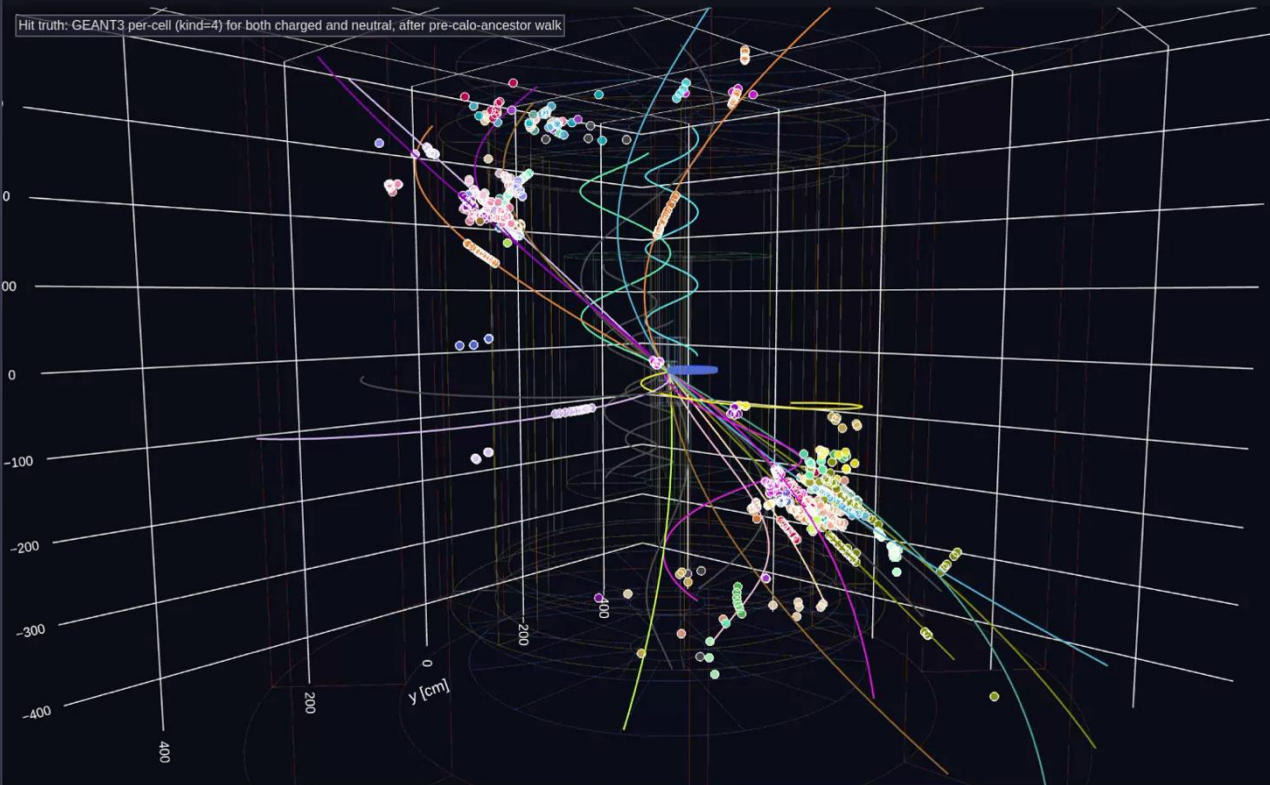
show all ✓ hide all ✕

Click a particle row → solo view (only that particle's hits + track).
Shift+click → toggle one particle on/off without affecting others.
Use the buttons above for bulk show/hide.

$\pi^\pm$ #35	p=18.15	h:35 t:1
$\pi^\pm$ #37	p=6.77	h:40 t:0
$\pi^\pm$ #39	p=6.64	h:66 t:0
K $^\pm$ #38	p=4.97	h:28 t:1
$\pi^\pm$ #15	p=4.12	h:0 t:1
γ #41	p=3.36	h:20 t:0
γ #16	p=2.81	h:28 t:0
n #43	p=2.70	h:50 t:0
$\pi^\pm$ #7	p=2.43	h:36 t:1
$\pi^\pm$ #47	p=2.21	h:18 t:1
$\pi^\pm$ #53	p=2.16	h:15 t:1
γ #18	p=1.94	h:13 t:0
$\pi^\pm$ #14	p=1.59	h:27 t:1
$\pi^\pm$ #48	p=1.58	h:12 t:0
n #42	p=1.53	h:2 t:0
$\mu^\pm$ #52	p=1.43	h:30 t:1
γ #62	p=1.29	h:20 t:0
$\pi^\pm$ #12	p=1.28	h:0 t:1
$\pi^\pm$ #10	p=1.28	h:9 t:1

pf_mlpf.parquet | event 2 — labels: calo-mother

Hit truth: GEANT3 per-cell (kind=4) for both charged and neutral, after pre-calo-ancestor walk



- γ #60 p=0.10 GeV
- γ #61 p=0.05 GeV
- γ #62 p=1.29 GeV
- γ #63 p=0.60 GeV
- e $^\pm$ #64 p=0.60 GeV
- γ #65 p=0.41 GeV
- γ #66 p=0.38 GeV
- e $^\pm$ #67 p=1.05 GeV
- e $^\pm$ #68 p=0.52 GeV
- e $^\pm$ #69 p=0.30 GeV
- γ #70 p=0.23 GeV
- e $^\pm$ #71 p=0.49 GeV
- e $^\pm$ #72 p=0.49 GeV
- e $^\pm$ #73 p=0.19 GeV
- γ #74 p=0.36 GeV
- γ #75 p=0.24 GeV
- γ #76 p=0.39 GeV

MC truth (calo-face dots)

- MC truth - other (2)
- MC truth - $\pi^\pm$ (28)
- MC truth - γ (28)
- MC truth - K $^\pm$ (1)
- MC truth - n (2)
- MC truth - $\mu^\pm$ (3)
- MC truth - p (1)
- MC truth - e $^\pm$ (12)

Tracks cut off in tracker (no calo entry)

- Decayed-in-tracker (1)