未来正负电子对撞机实验中的粒子流算法

李波 （西南科技大学）

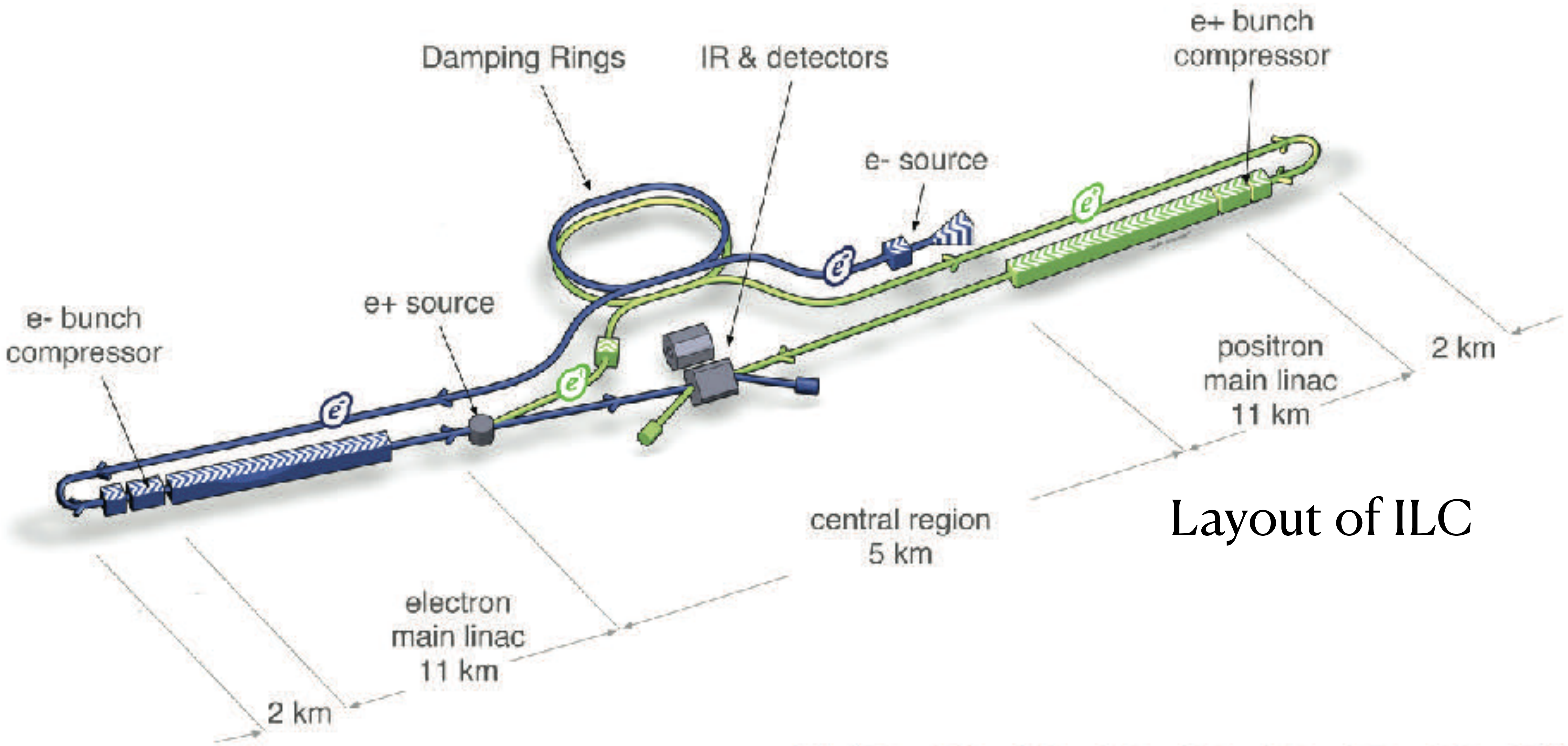
2020.7.1

Content

- Electron-positron collider experiments
- Particle flow calorimetry
  - High granularity calorimeter
  - Particle flow algorithm (PFA)
- The proposal for improving the PFA performance

International Linear Collider

- Future e^+e^- colliders: CEPC, FCC-*ee*, CLIC, and ILC
- "An electron-positron Higgs factory is the highest priority next collider." (from the future European Strategy for particle physics)



Layout of ILC

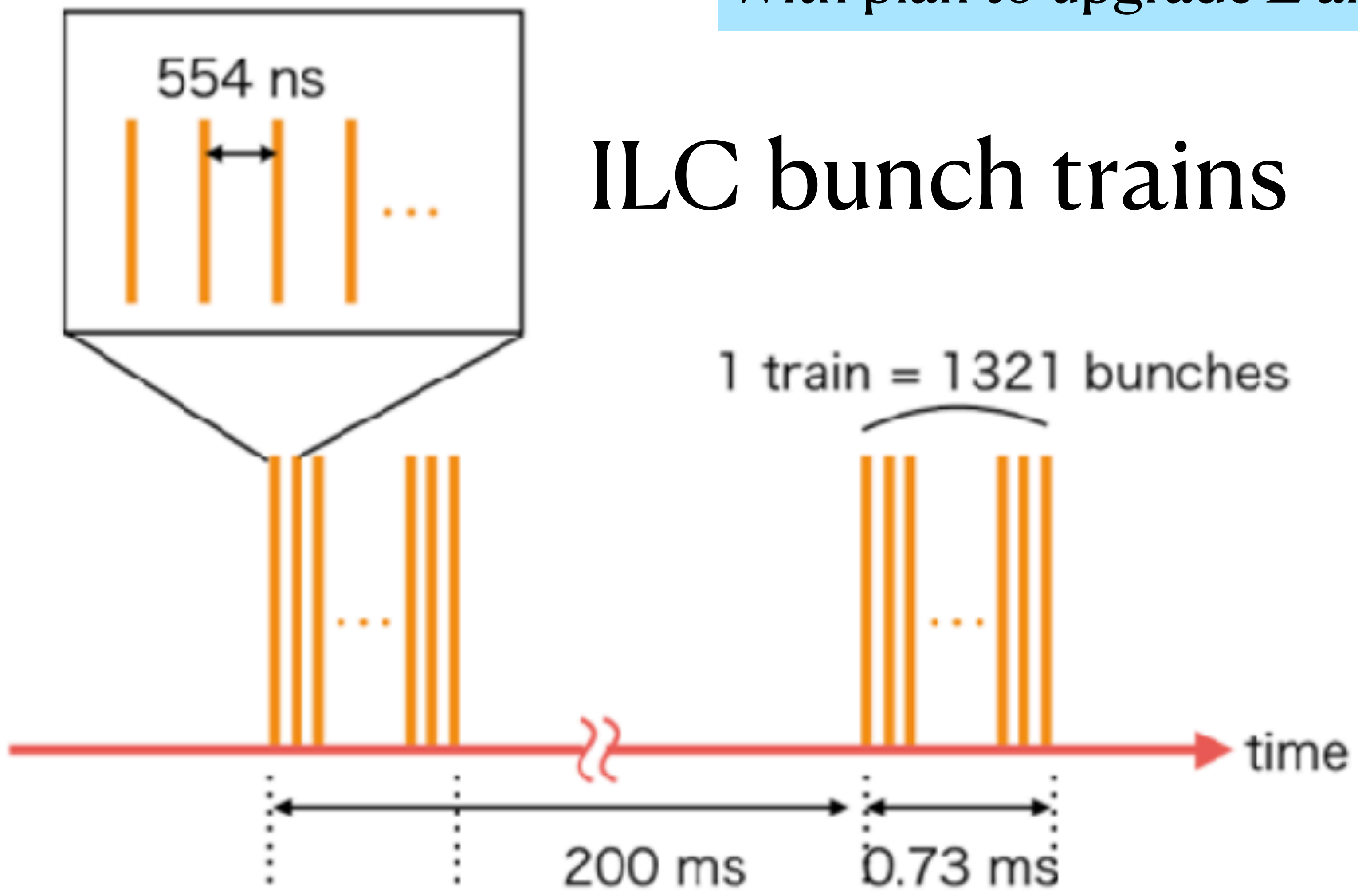
Superconducting RF cavity, average gradient 31.5 MV/m

ILC machine parameters

			Baseline 500 GeV Machine		
Centre-of-mass energy	E_{CM}	GeV	250	350	500
Collision rate	f_{rep}	Hz	5	5	5
Electron linac rate	f_{linac}	Hz	10	5	5
Number of bunches	n_b		1312	1312	1312
Bunch population	N	$\times 10^{10}$	2.0	2.0	2.0
Bunch separation	Δt_b	ns	554	554	554
Pulse current	I_{beam}	mA	5.8	5.8	5.8
Luminosity	L	$\times 10^{34}$ cm^{-2}s^{-1}	0.75	1.0	1.8
Fraction of luminosity in top 1%	$L_{0.01}/L$		87.1%	77.4%	58.3%
Average energy loss	δ_{BS}		0.97%	1.9%	4.5%
Number of pairs per bunch crossing	N_{pairs}	$\times 10^3$	62.4	93.6	139.0

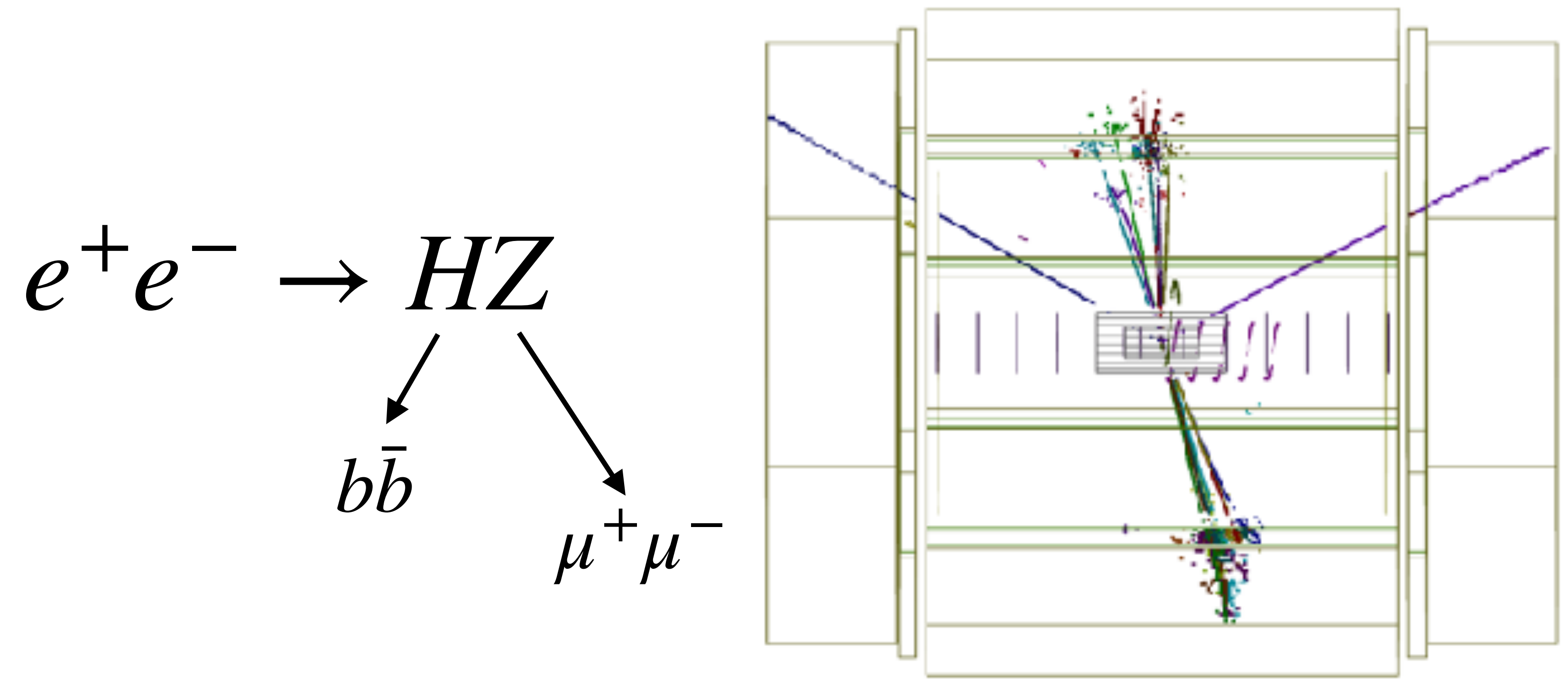
With plan to upgrade L and E_{cm}

ILC bunch trains

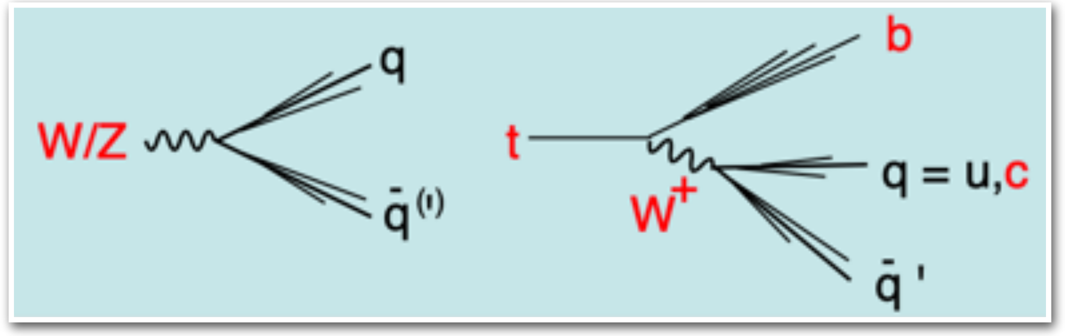


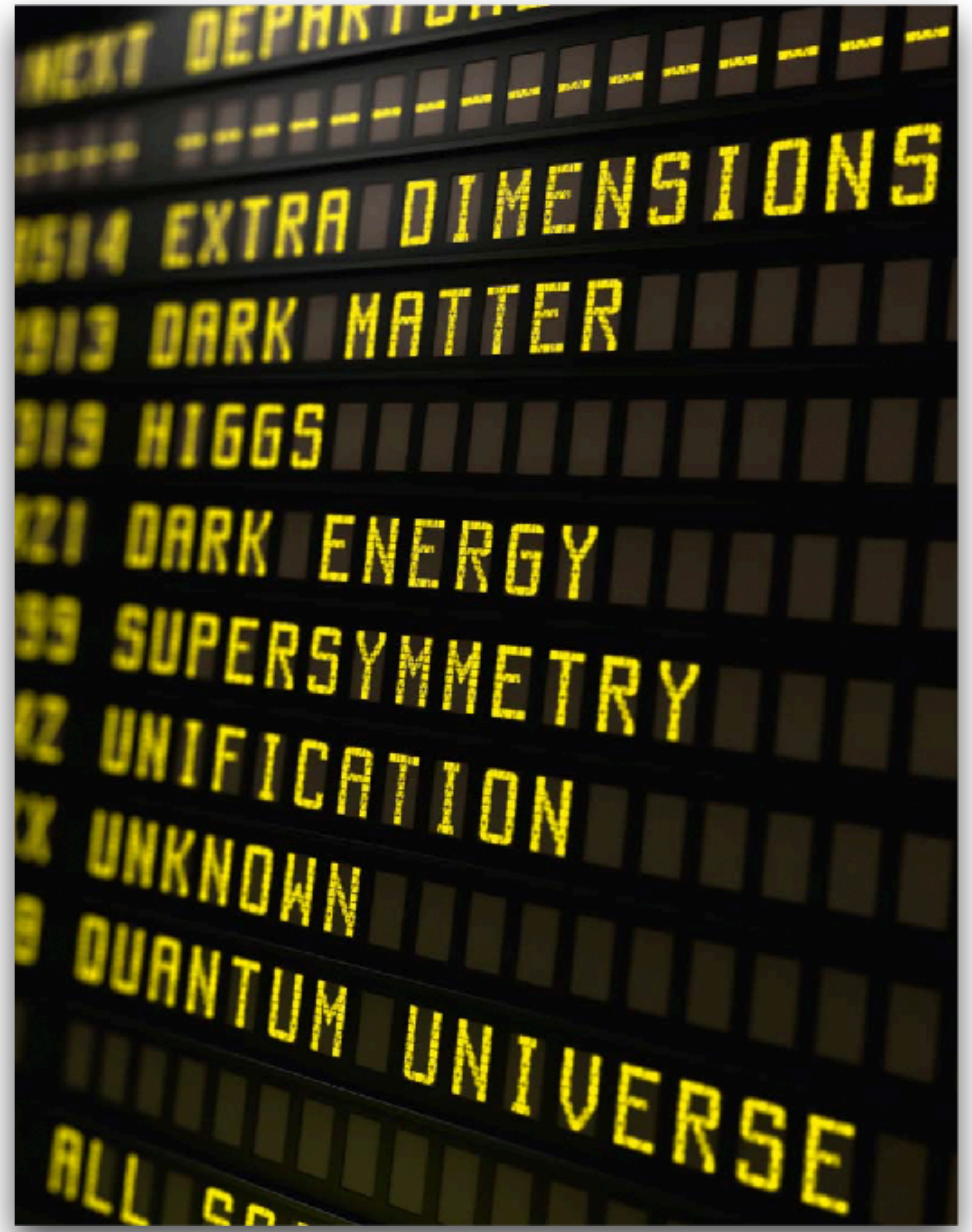
Physics and detector

- **ILC Physics goals:**
  - **Precision measurements**
    - H: $e^+e^- \to HZ$ (250 GeV), $e^+e^- \to t\bar{t}H$ (500 GeV) , $e^+e^- \to ZHH$ (500 GeV)
    - Z: $e^+e^- \to Z$ (91 GeV)
    - W: $e^+e^- \to WW$ (160 GeV, 350 GeV)
    - t: $e^+e^- \to t\bar{t}$ (350 GeV)
  - Search for: SUSY (if any) ...
  - Advantage: Cleanliness of signal

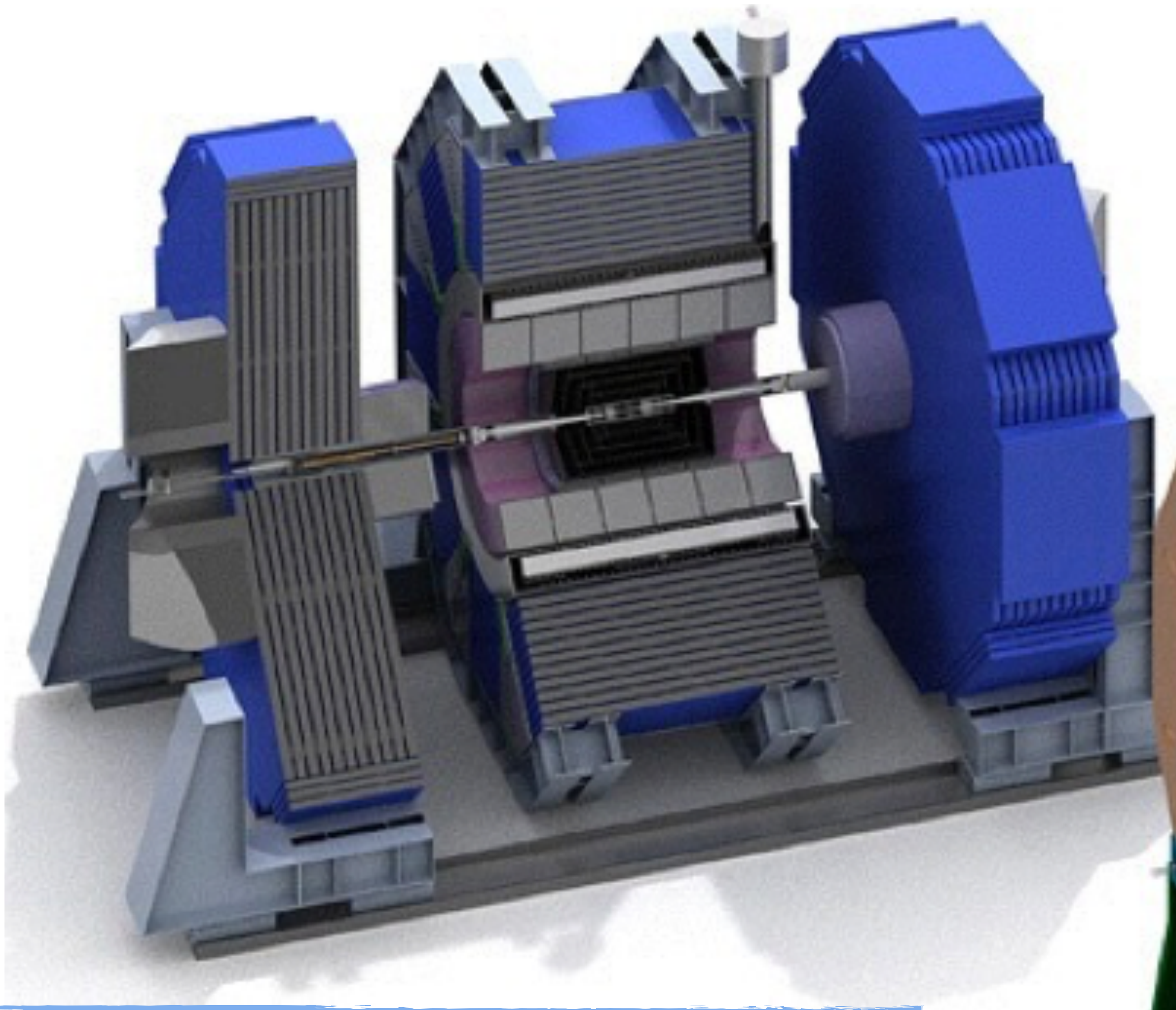


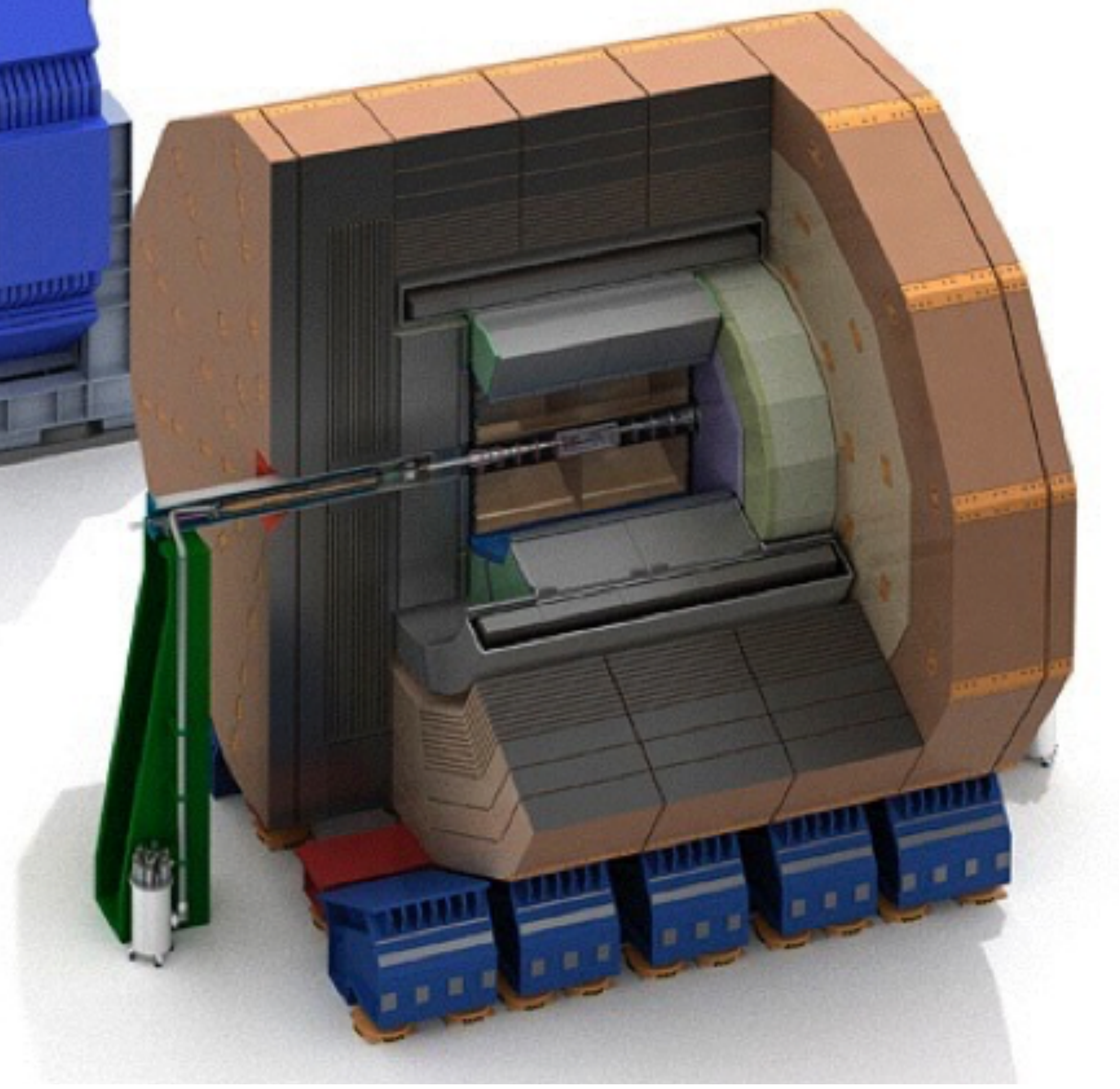
- **Detector**
  - The most precise detector for studying the interaction of high energy particles.
  - The particles must be reconstructed in multi-jet final states and identified with invariant mass.
    - High precision tracker
    - High granularity calorimeter





Two detectors





ILD

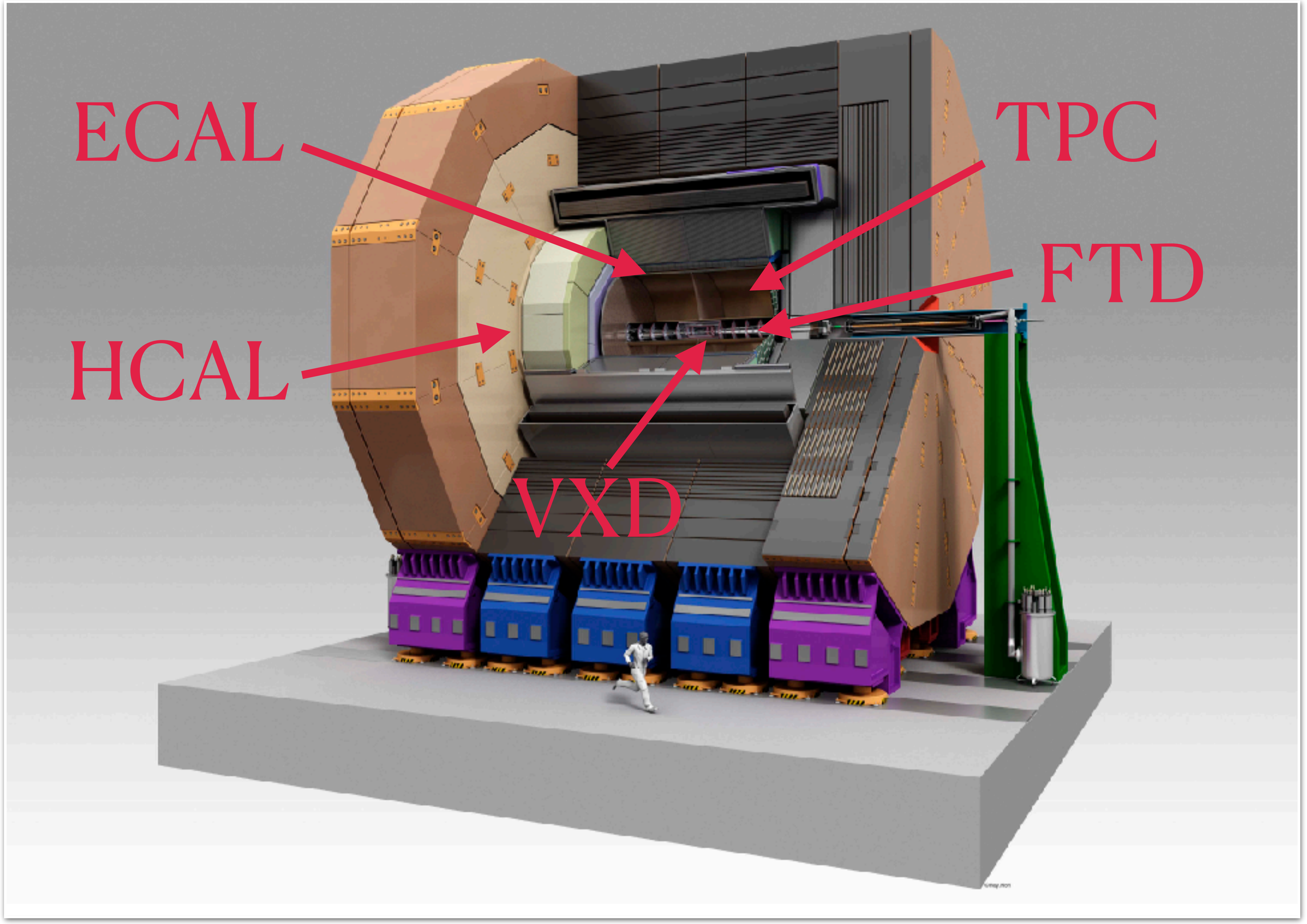
- Large R with TPC tracker
- 32 countries, 151 institutions, ~700 members
- B=3.5T, TPC + Si trackers
- ECAL: R=1.8m

The ILC collaboration (@LCWS2019)

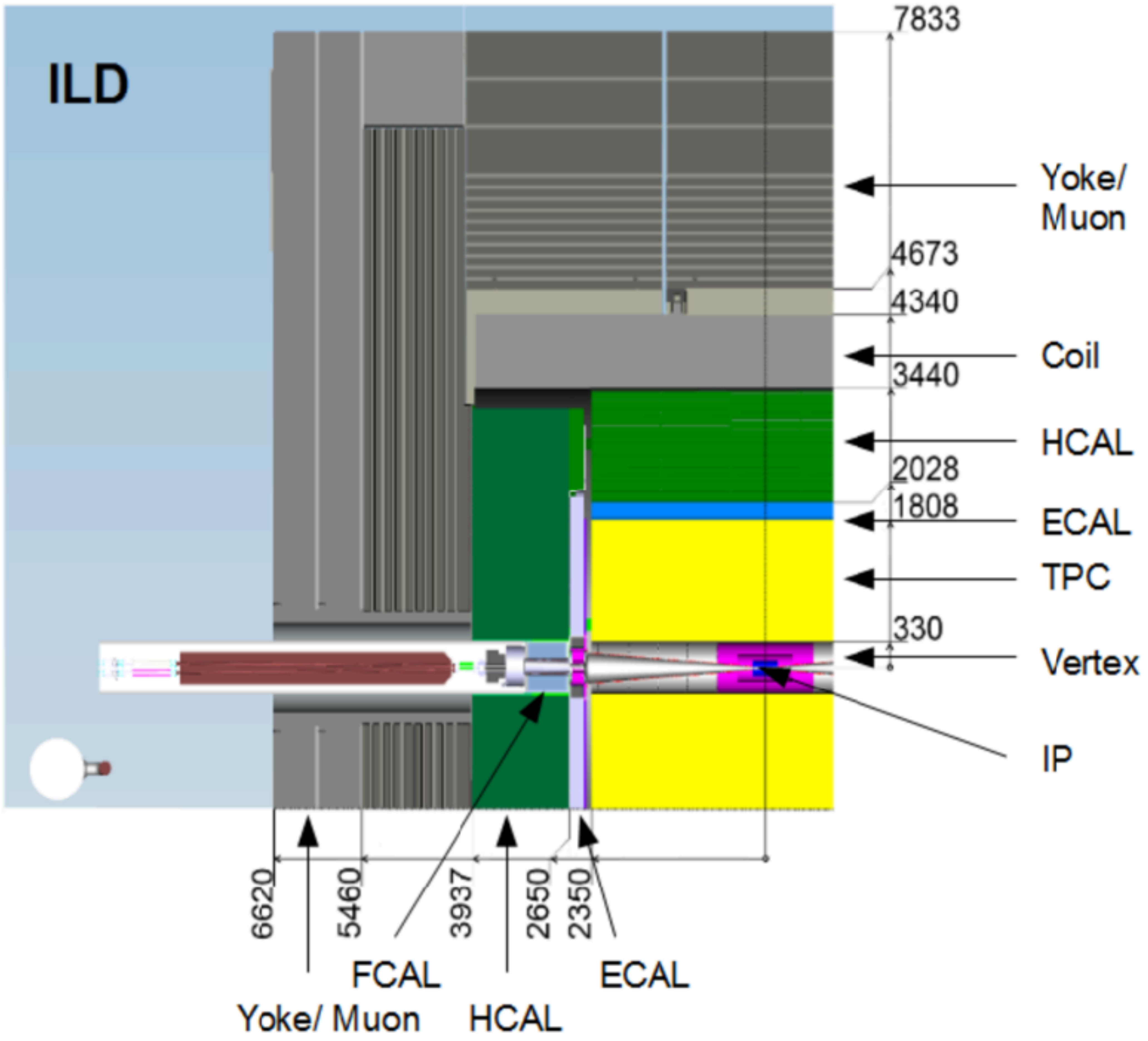
SiD

- High B with Si strip tracker
- 18 countries, 77 institutions, ~240 members
- B=5T, Si only tracker
- ECAL: R=1.27m

ILD

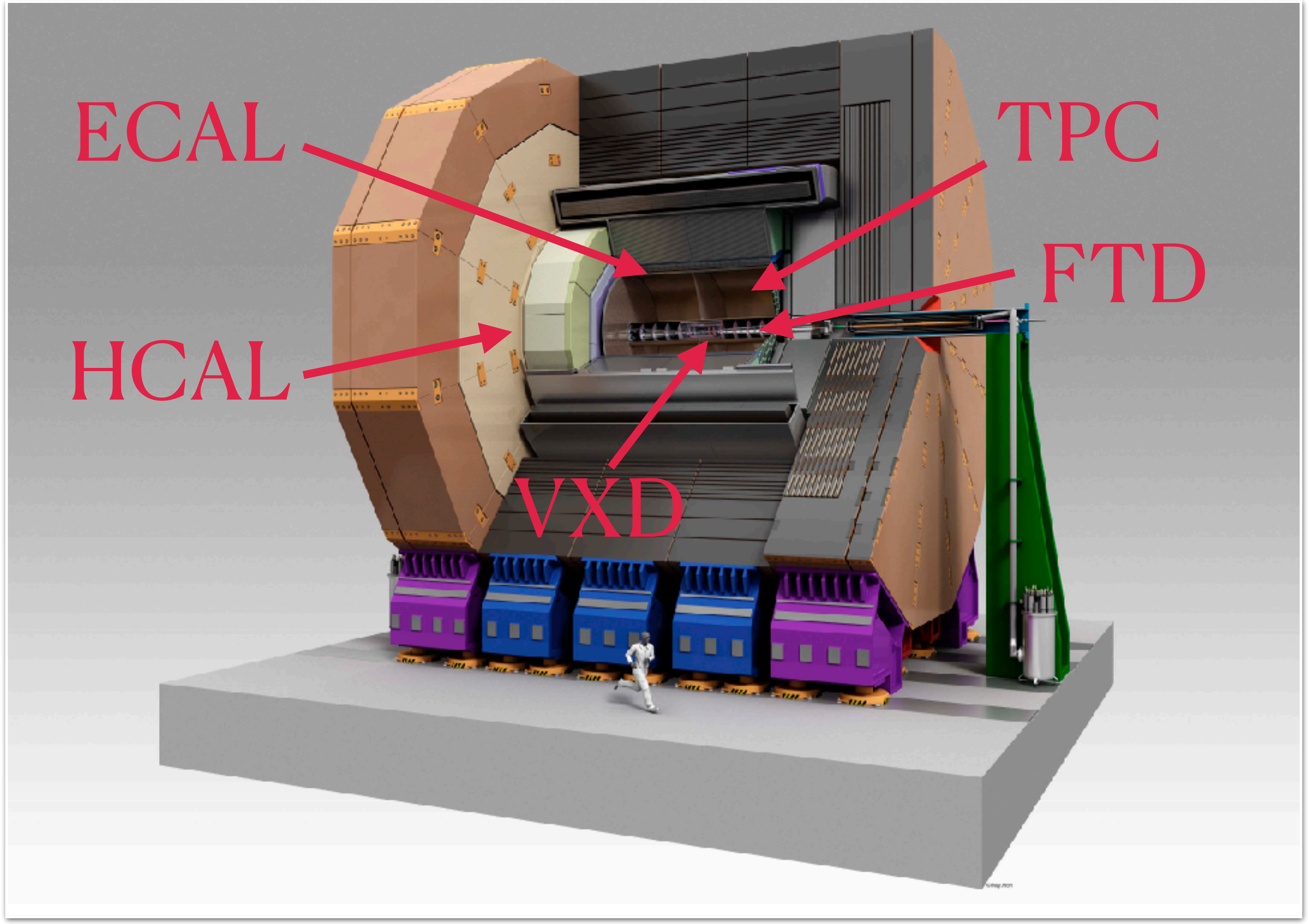


ECAL
TPC
FTD
HCAL
VXD

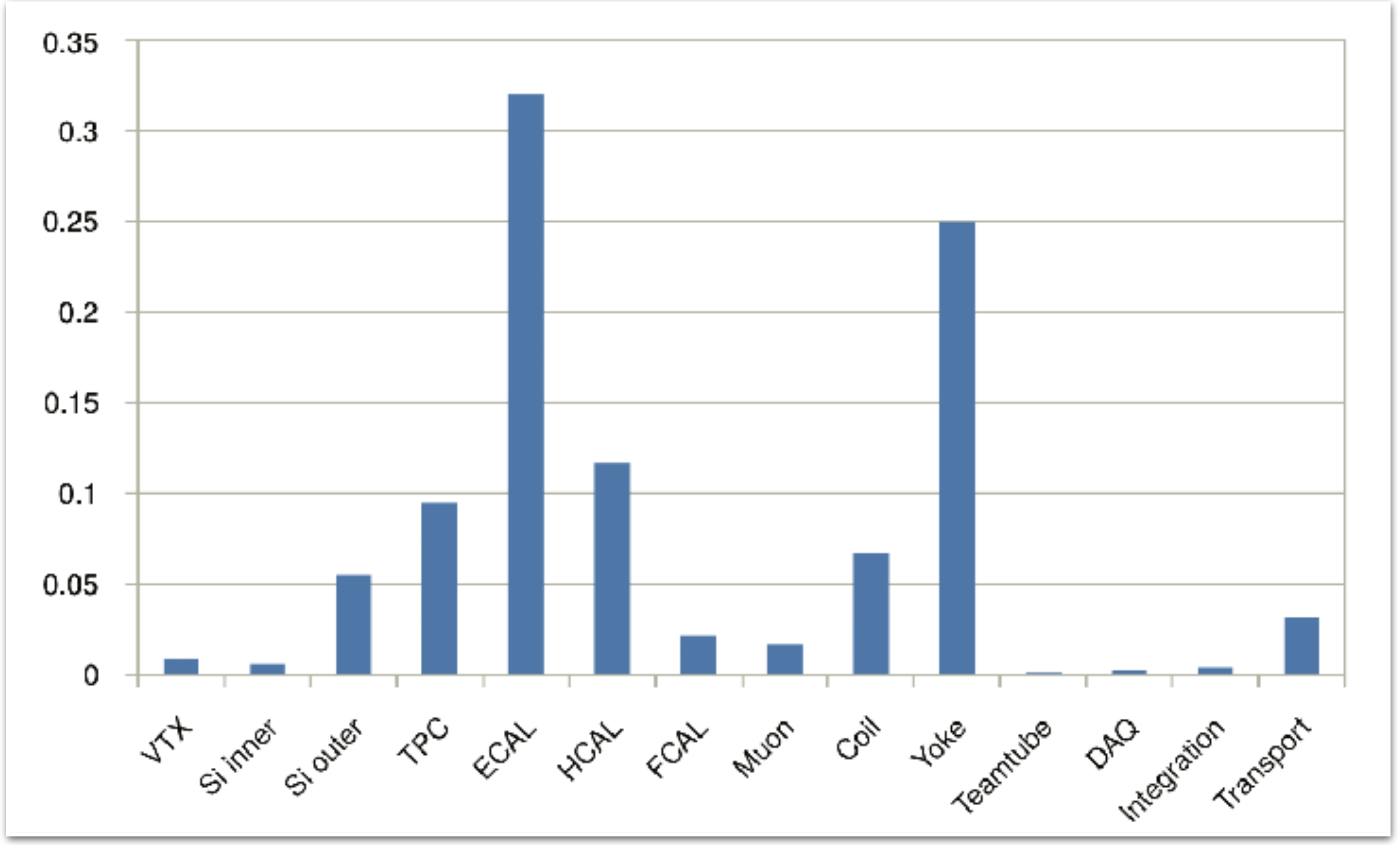


ILD
7833
Yoke/
Muon
4673
4340
Coil
3440
HCAL
2028
1808
ECAL
TPC
330
Vertex
IP
6620
5460
3937
2650
2350
FCAL
ECAL
Yoke/ Muon
HCAL

ILD



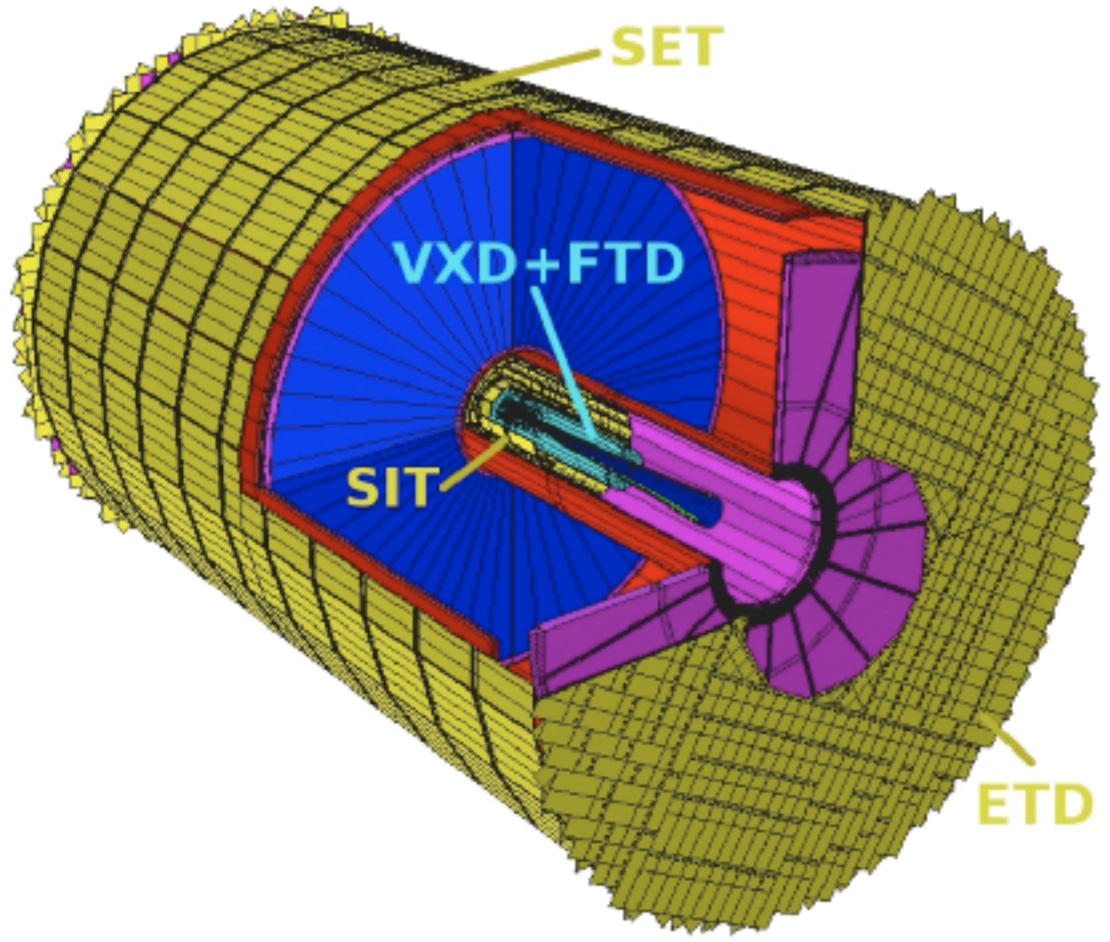
50% cost of ECAL is silicon



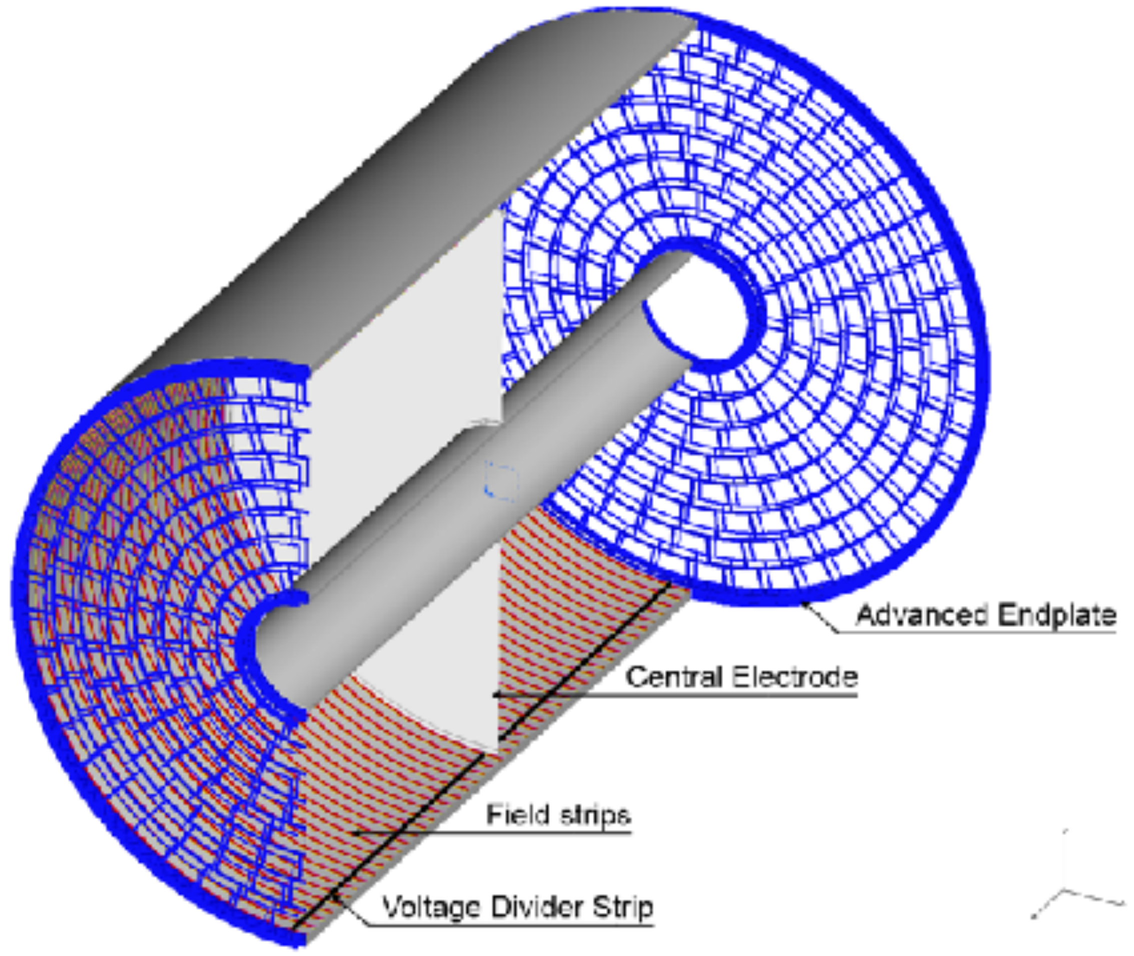
ILD cost fraction [total = 392 M ILCU (USD@2012)]

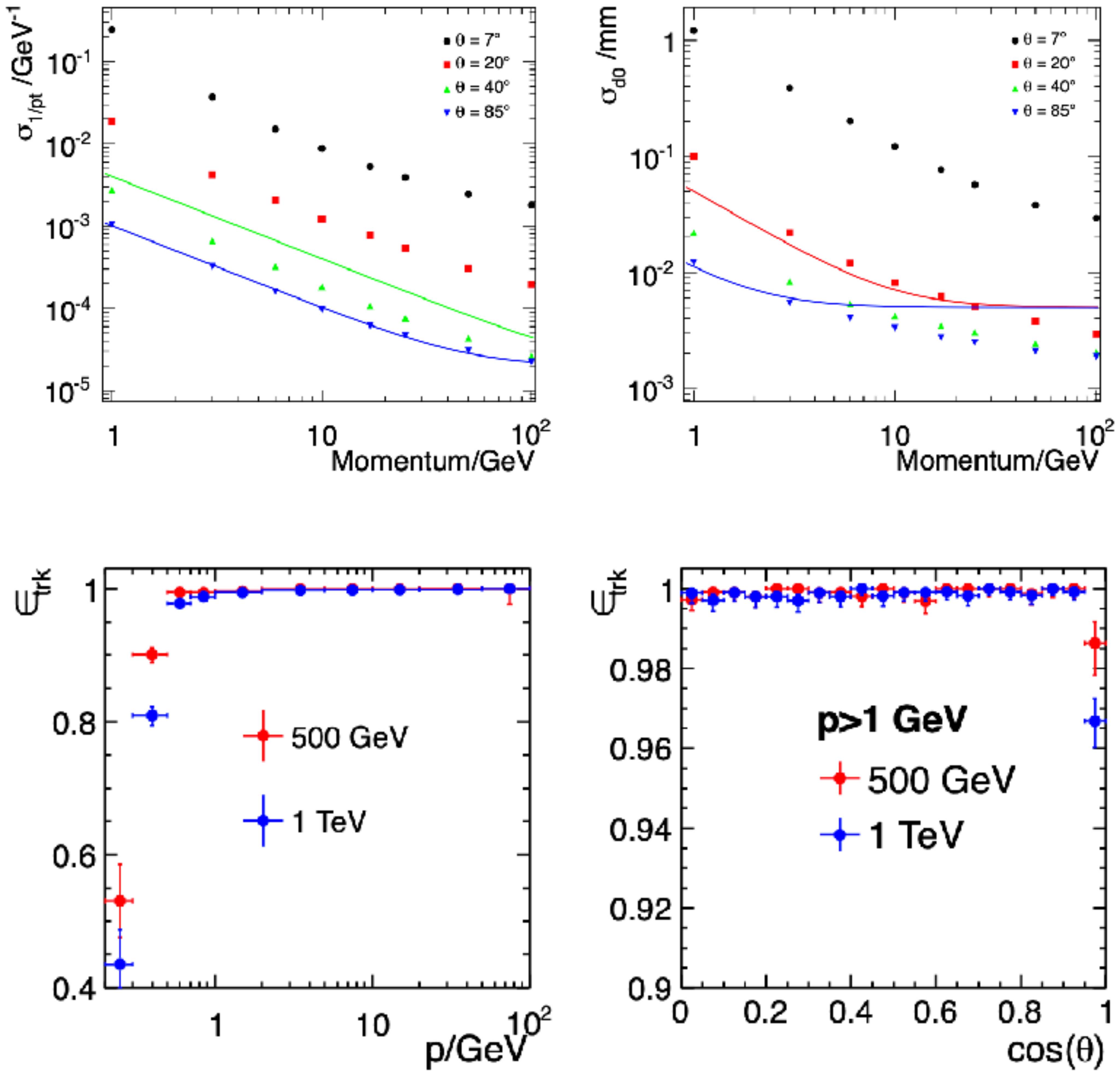
Trackers

Si tracker



TPC

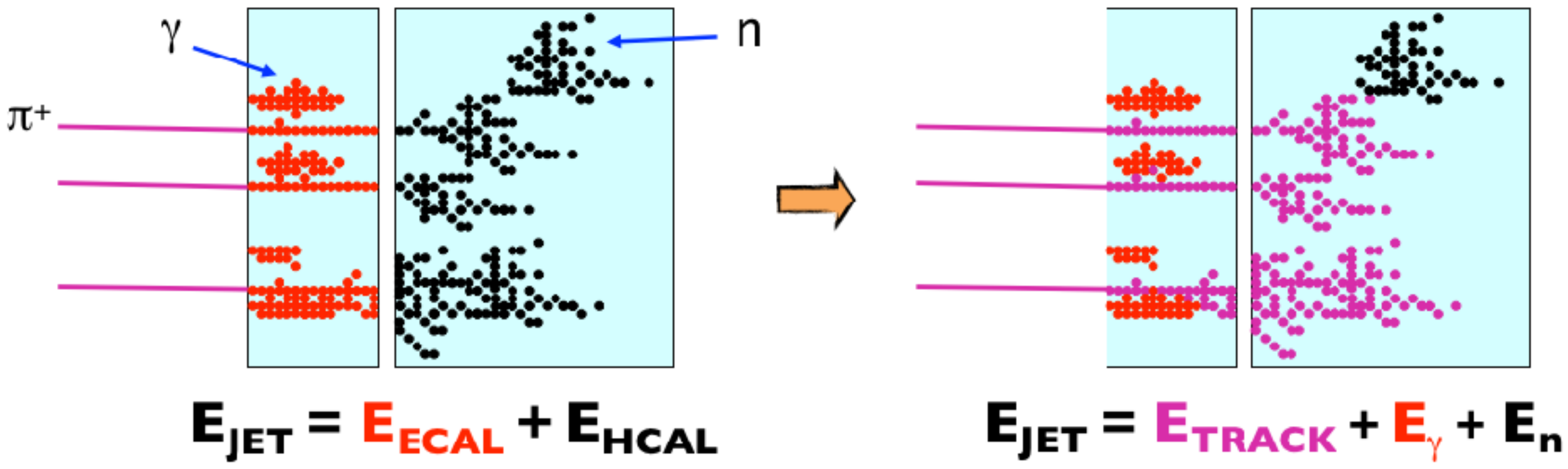




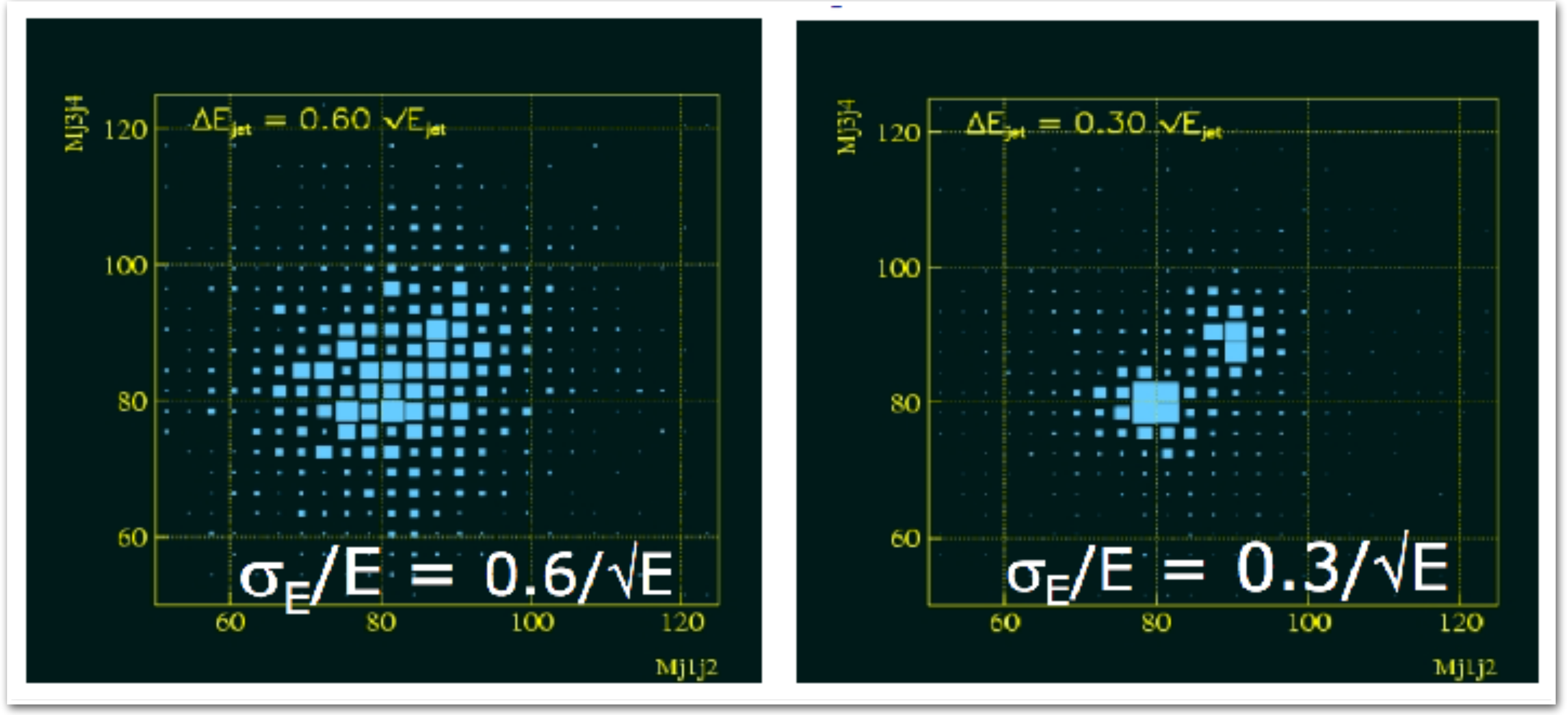
- Silicon tracker $\sigma_x \sim 5\ \mu$m
- TPC $\sigma_x \sim 100\ \mu$m, $N_{meas} = 220$, low material budget
- Momentum resolution at high momenta: $\sigma_{1/p_t} = 2 \times 10^{-5}\ \text{GeV}^{-1}$
- Good tracking efficiency for $p > 1$ GeV

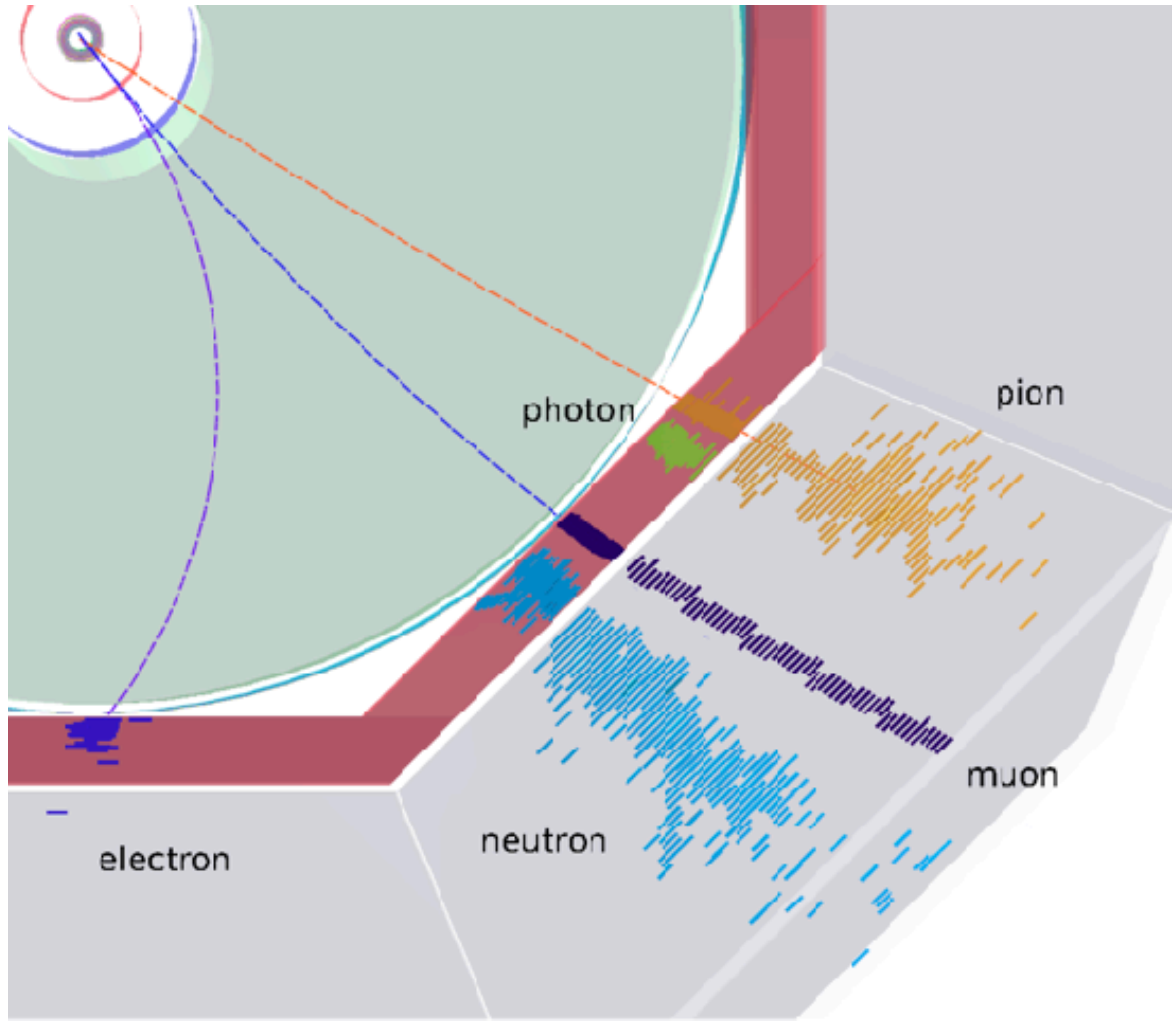
Particle flow calorimetry

- From energy flow to particle flow



- W and Z separation by di-jet invariant mass at two jet energy resolutions



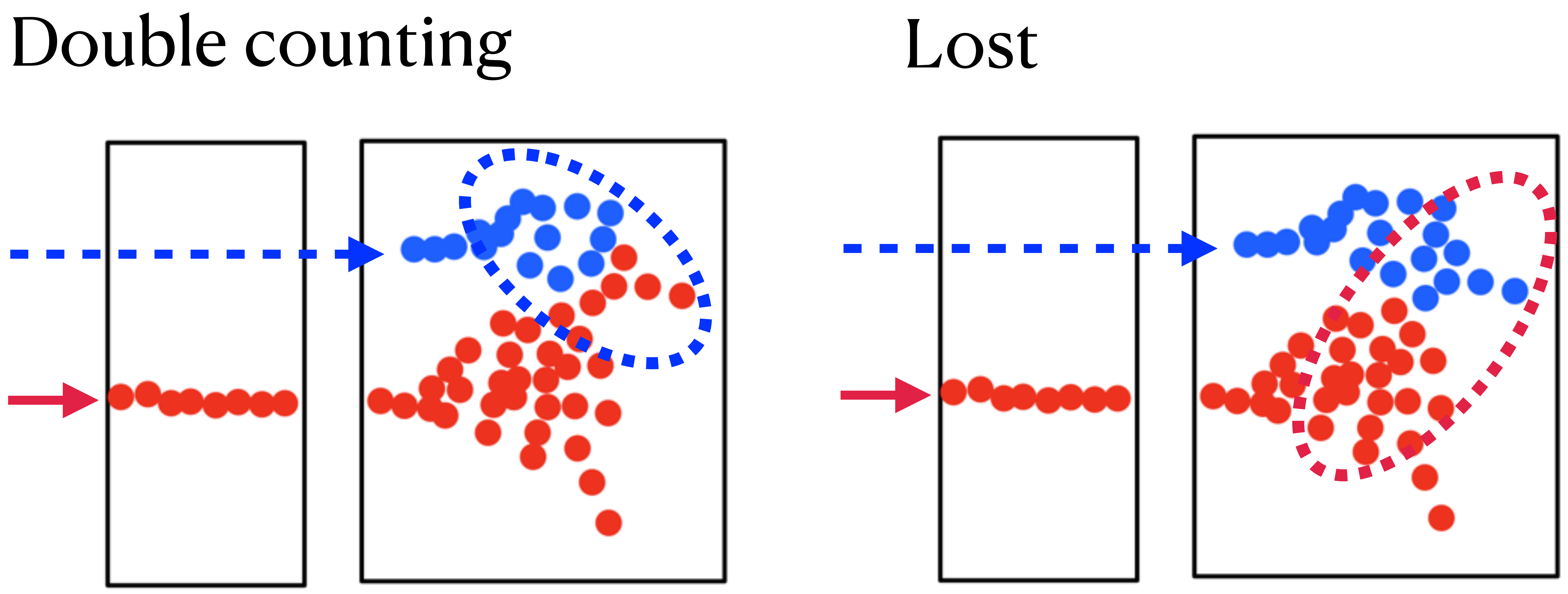


- Particle flow calorimetry aims to reconstruct visible final state particles from the information recorded by detector
- High precision tracker and high granularity calorimeter are mandatories
- Jet energy resolution requirement: $\sigma_E/E = 3 \sim 4\,\%$ for 50 to 500 GeV jets

Why particle flow calorimetry ?

Component	Detector	Energy fract.	Energy res.	Jet energy res.
Charged particles ($X^{\pm}$)	Tracker	$\sim 0.6Ej$	$10^{-4}E_{X^{\pm}}^{2}$	$< 3.6 \times 10^{-5}E_j^2$
Photons (γ)	ECAL	$\sim 0.3E_j$	$0.15\sqrt{E_\gamma}$	$0.08\sqrt{E_j}$
Neutral Hadrons (h^0)	HCAL	$\sim 0.1E_j$	$0.55\sqrt{E_{h^0}}$	$0.17\sqrt{E_j}$

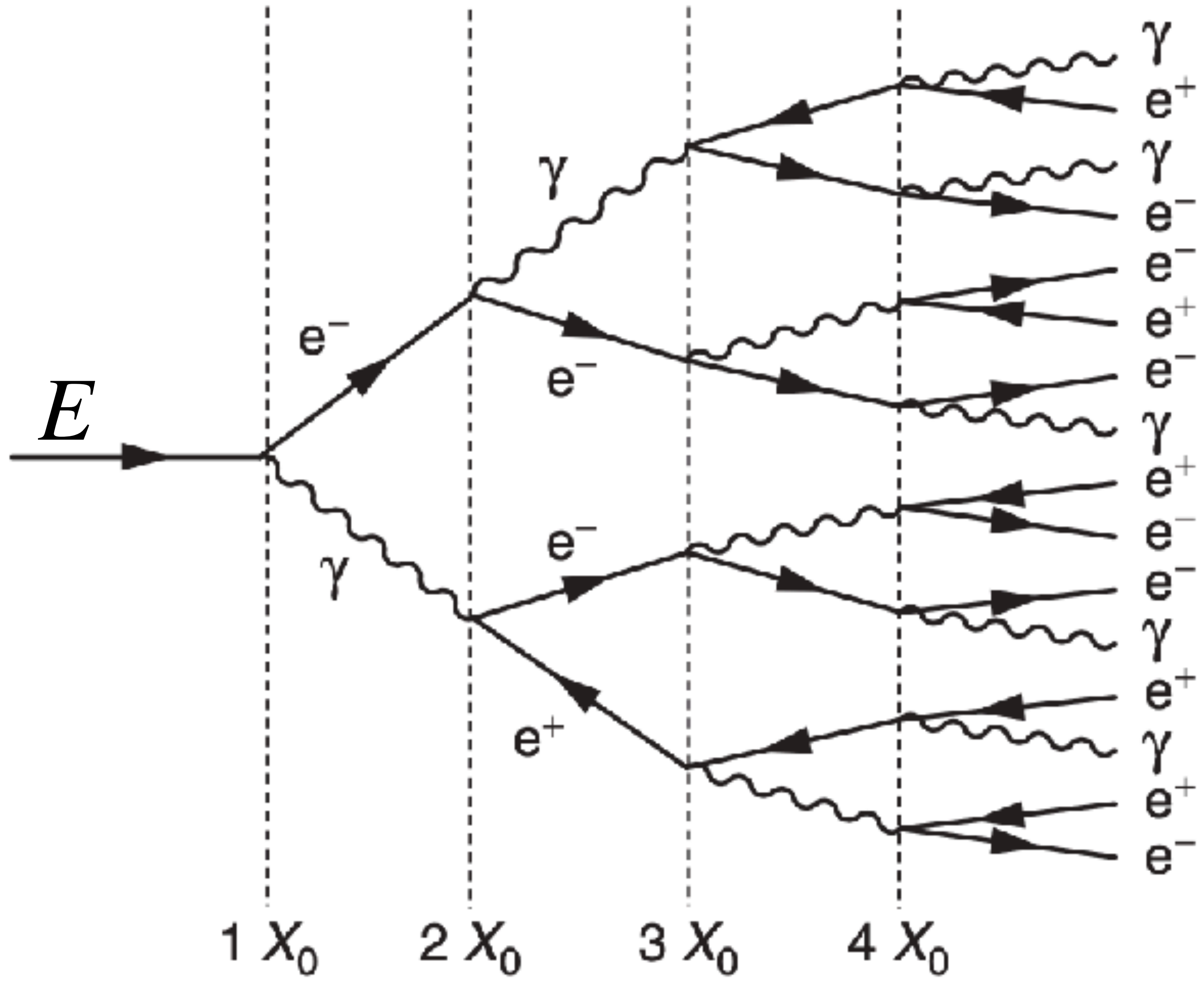
- 60% of jet energy is measured by tracker, which has a very good energy resolution.



- The cluster separation, especially the separation between charged cluster and neutral cluster, is the crucial aspect for particle flow calorimetry.
- The confusion (double counting and lost) should be made as small as possible.

Electromagnetic shower

The development of an electromagnetic shower:



Bremsstrahlung + e^+e^- pair production

Average particle energy after $nX_0 : \bar{E} = \frac{E}{2^n}$

Critical energy: $E_c = 800 \text{ MeV}/Z$

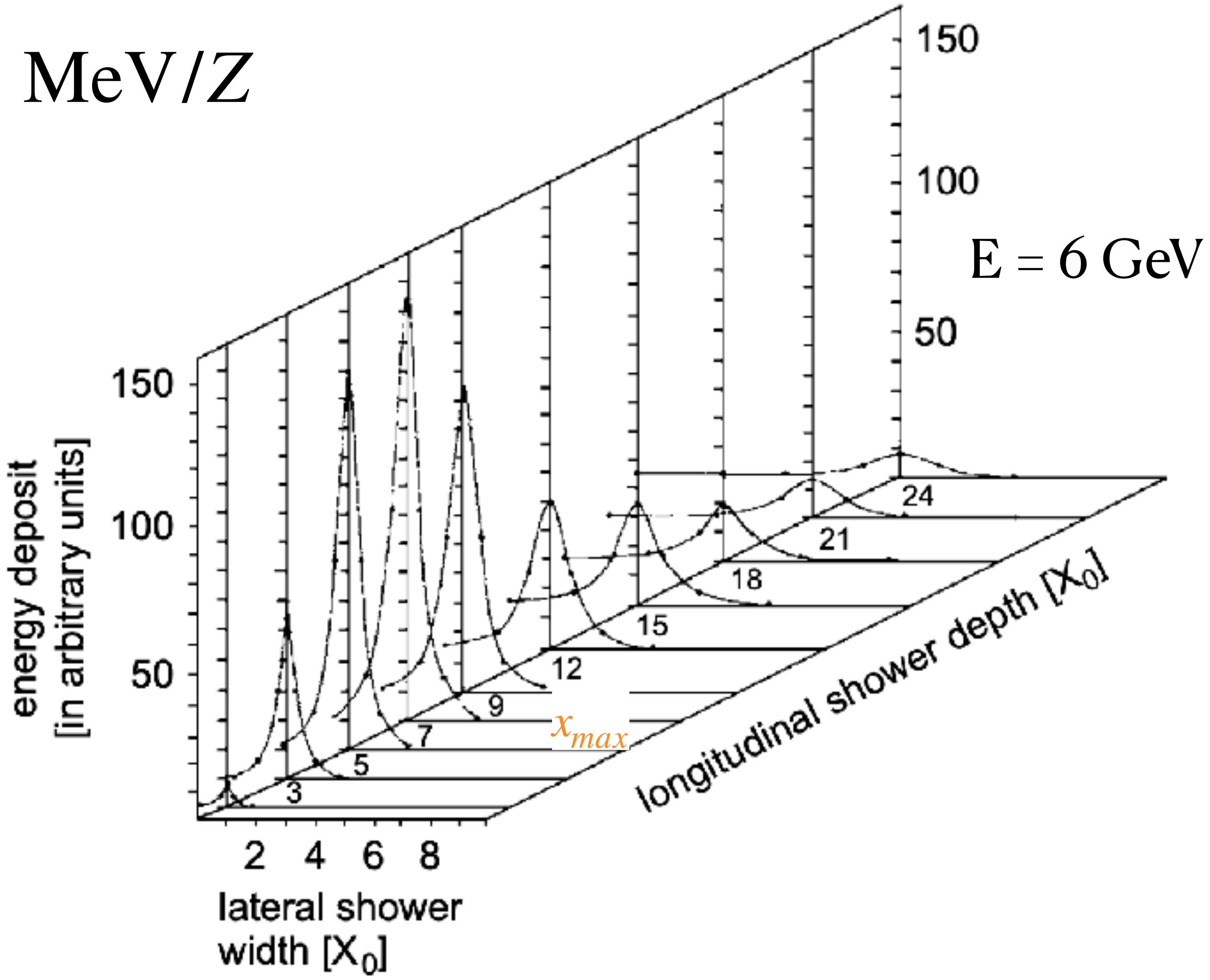
$$x_{max} = \frac{1}{\ln 2} \ln(E/E_c) \ [X_0]$$

The absorption of 95% photons needs additional $\sim 10 \ X_0$

Molière radius:

$$R_M = \frac{21\text{MeV}}{E_c} X_0$$

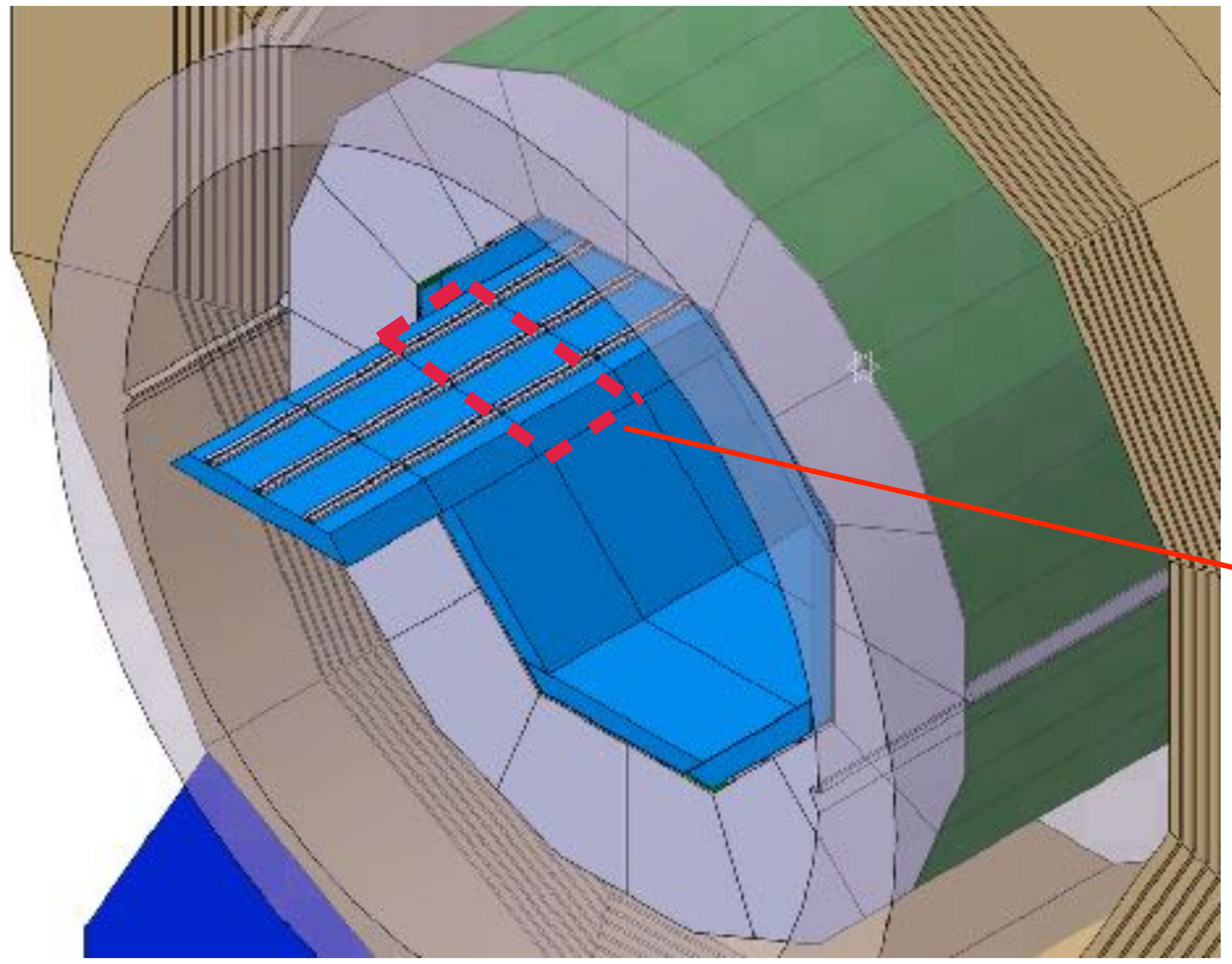
95% of the lateral shower is contained in $2R_M$

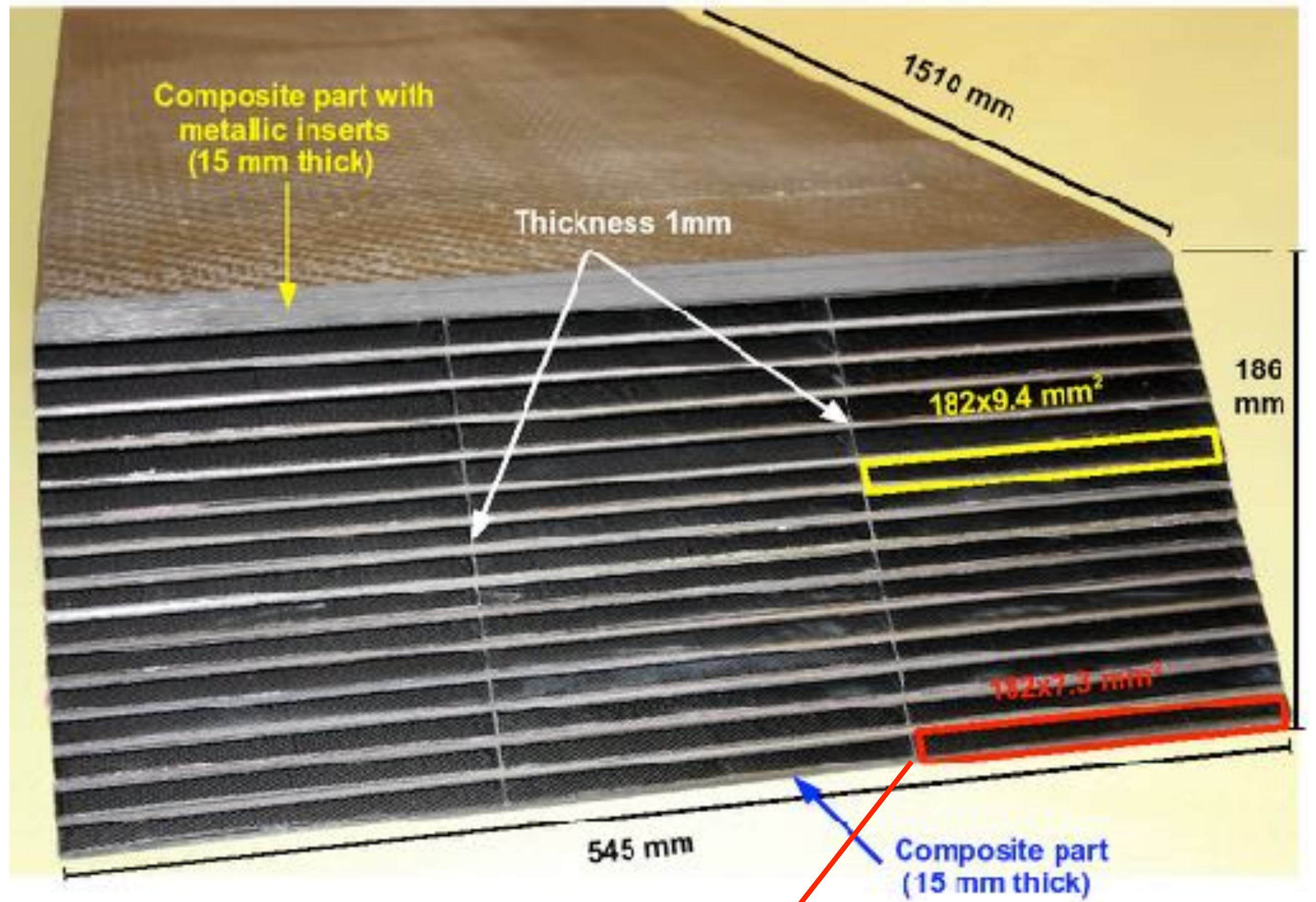


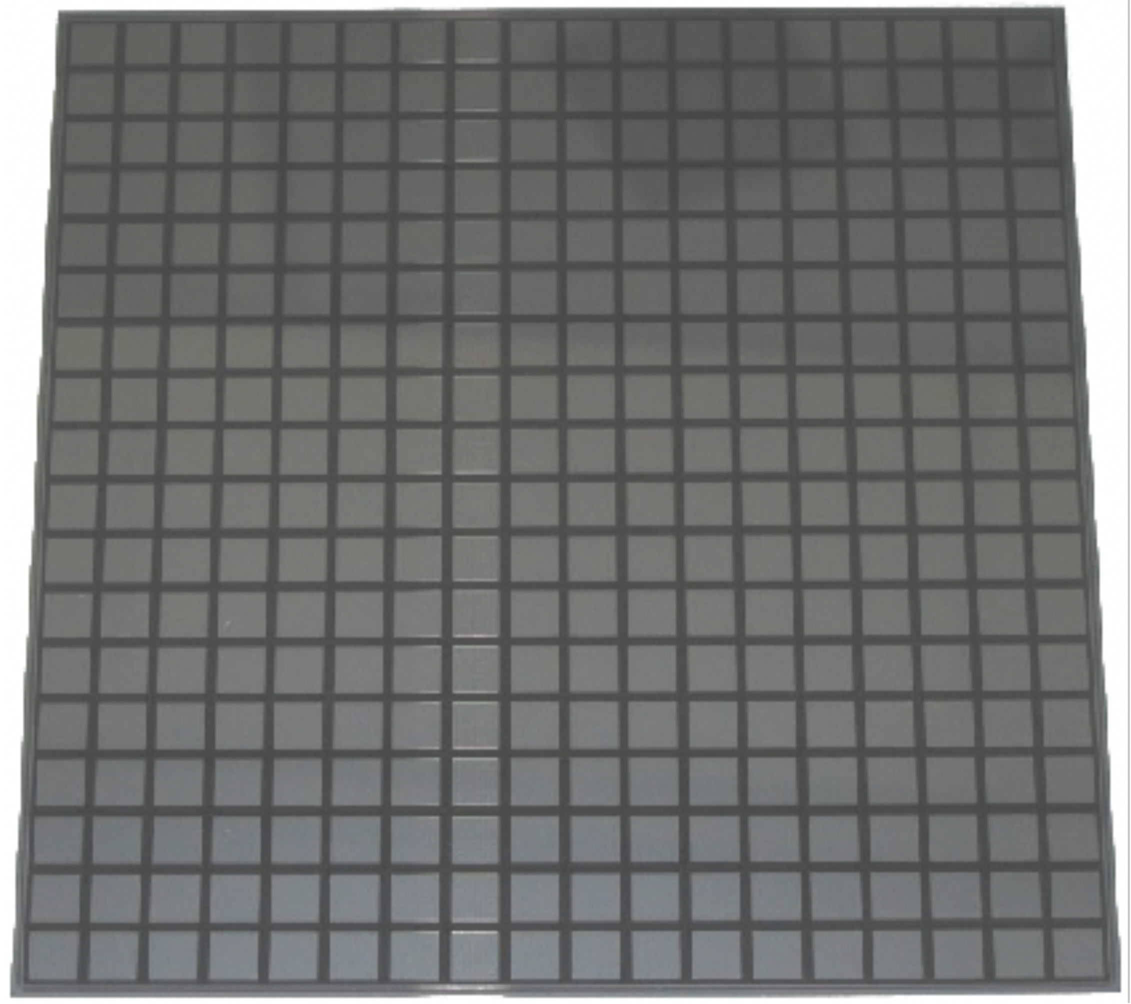
Material	Z	X_0 (cm)	E_c (MeV)
Iron	26	1.76	20.7
Tungsten	74	0.35	8.0
Lead	82	0.56	7.4

SiW ECAL

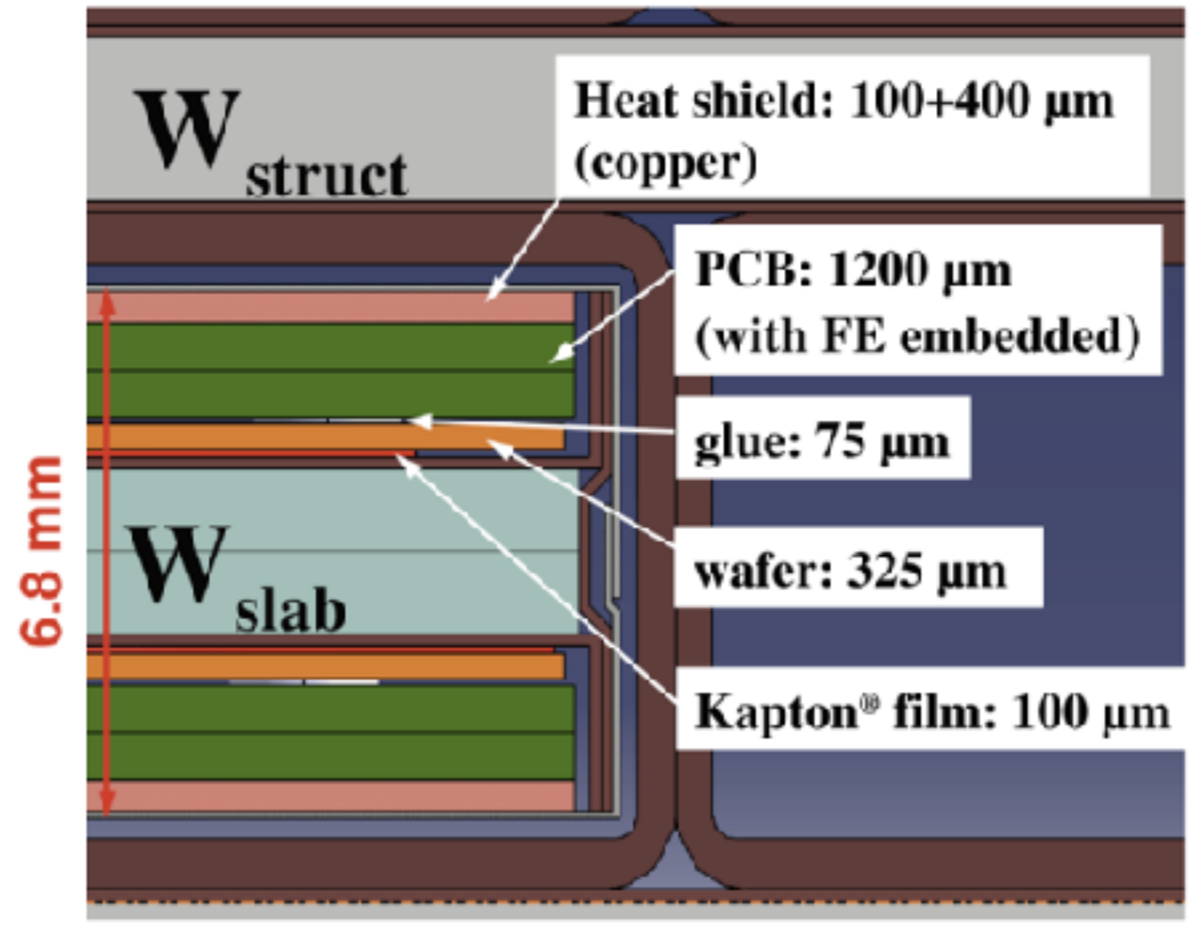
arXiv:1306.6329

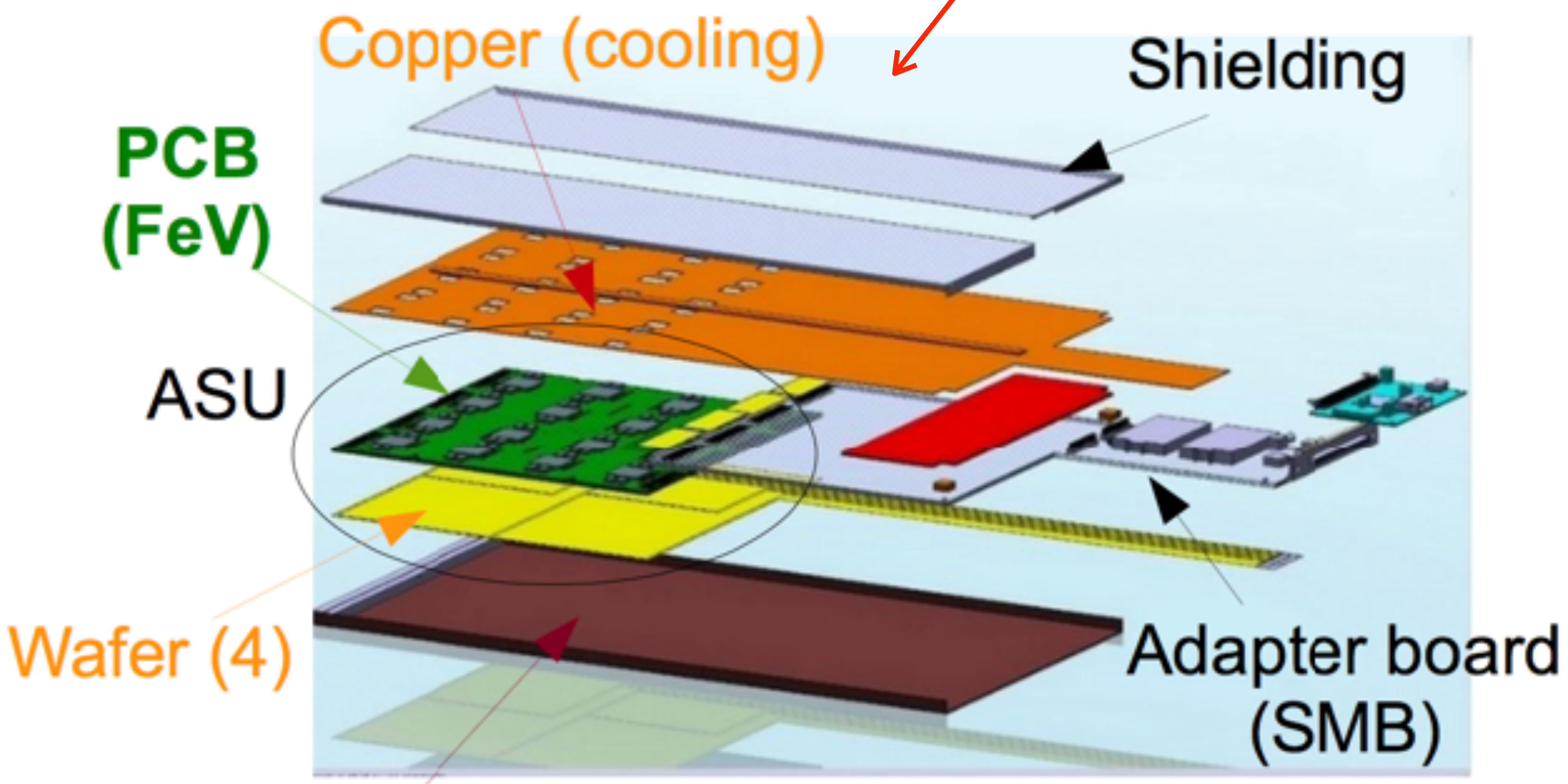






Silicon wafer matrix



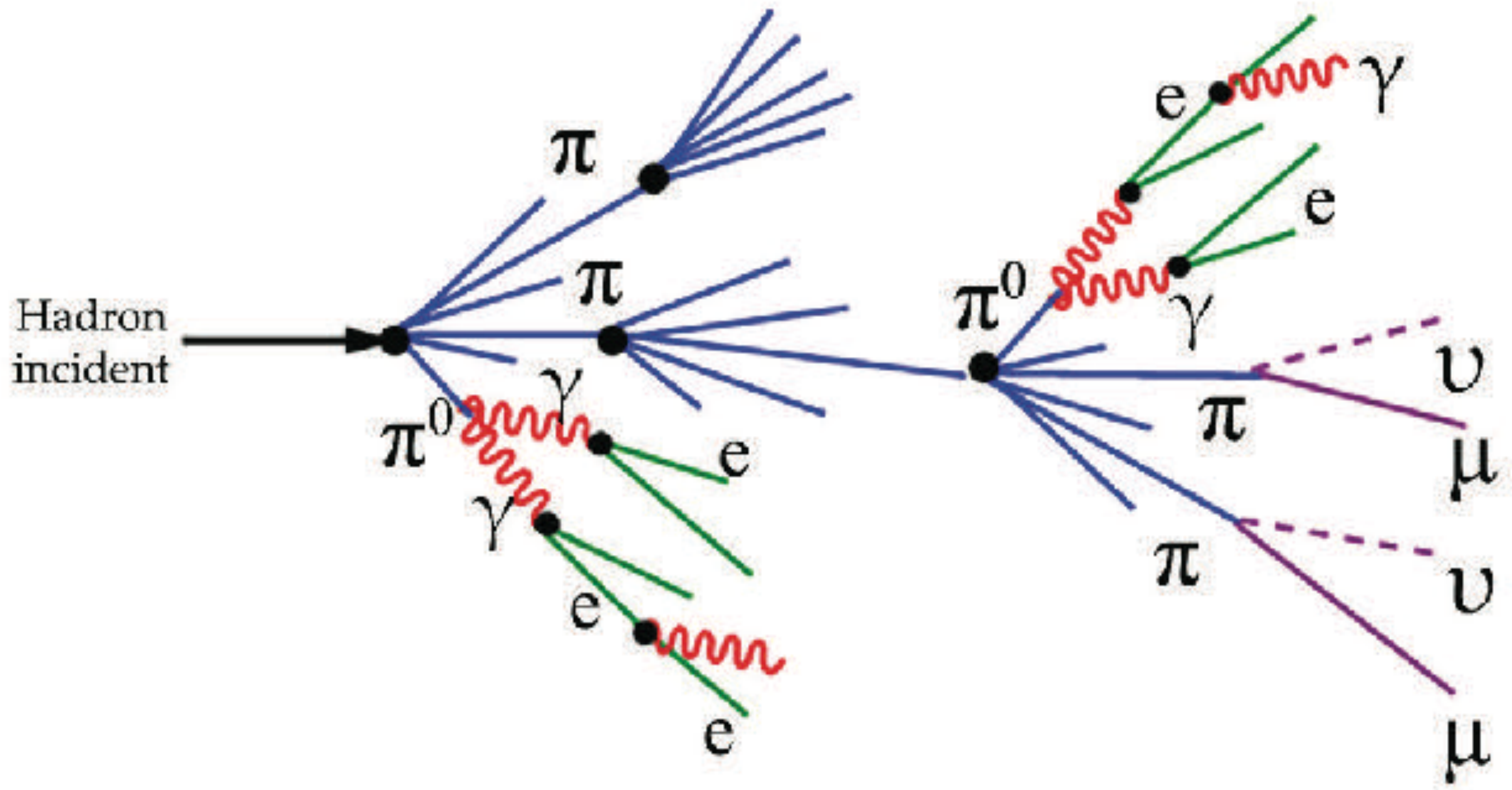


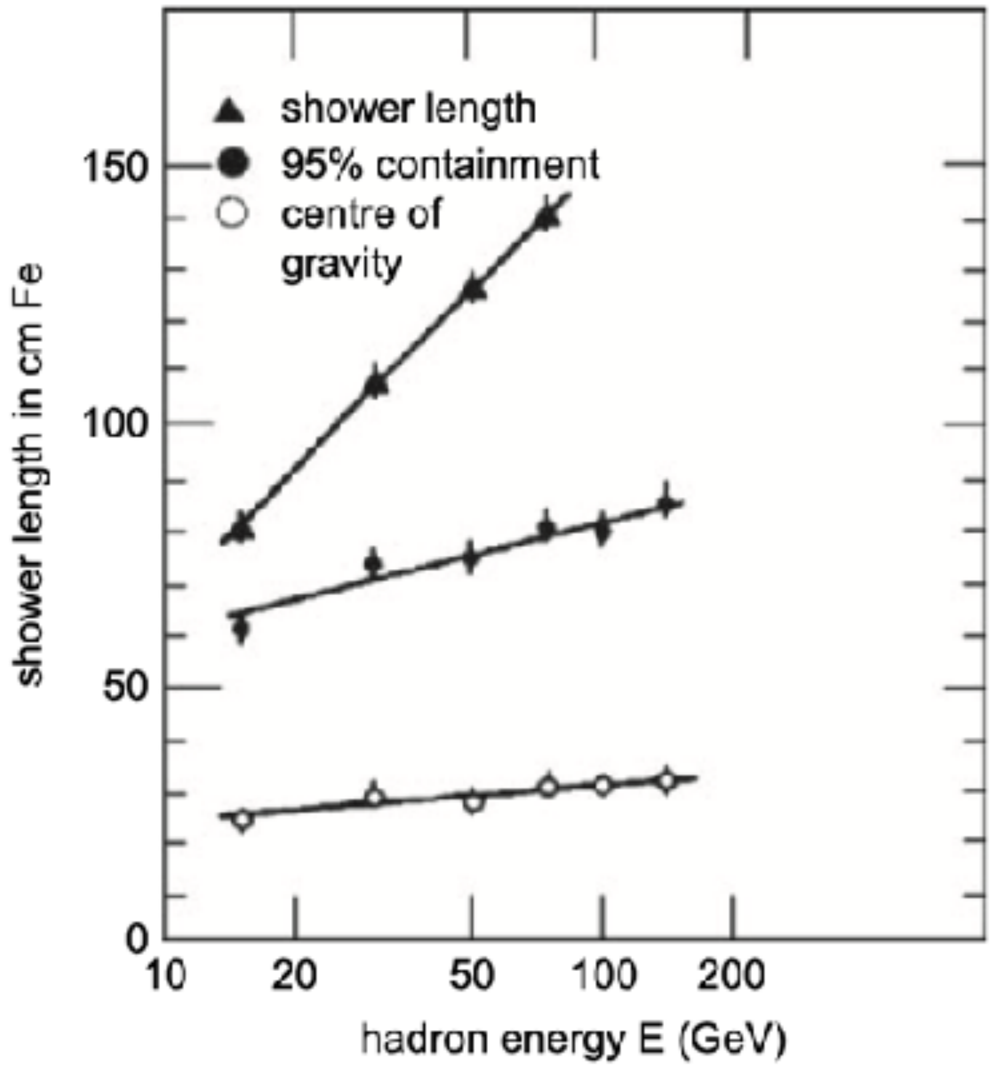
U layout of a short slab

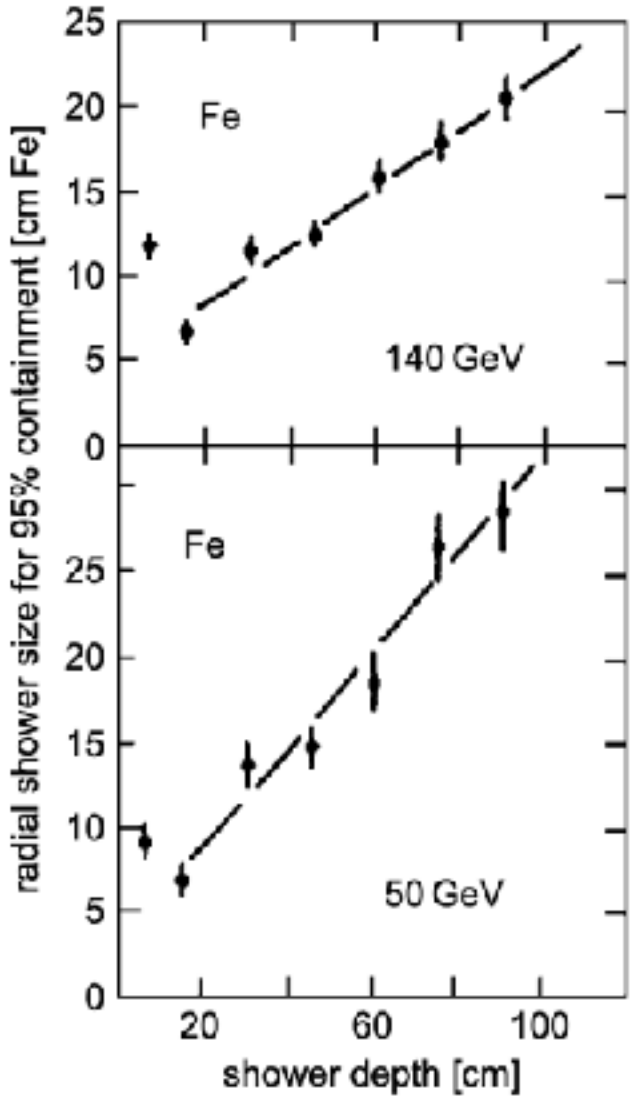
- Cell size : 5×5 mm²
- Thickness: 30 layers, 24 X_0
- Energy resolution: $0.15/\sqrt{E}$
- S/N ratio: 10:1 at 1 MIP level
- Total wafer surface: 3000 m²
- Leak-less water system for cooling

Hadronic shower

- Nuclear interaction length: $\lambda_I \approx 35\ \mathrm{g/cm^2} A^{1/3}$, usually much larger than X_0 (Iron: 16.8 cm)
- Leakage
- Invisible energy: ~ 30%
- The ratio of EM and hadronic components e/h > 1
- Hadronic calorimeter compensation; software compensation
- Dual readout: measure the e/h by Cherenkov light and scintillation

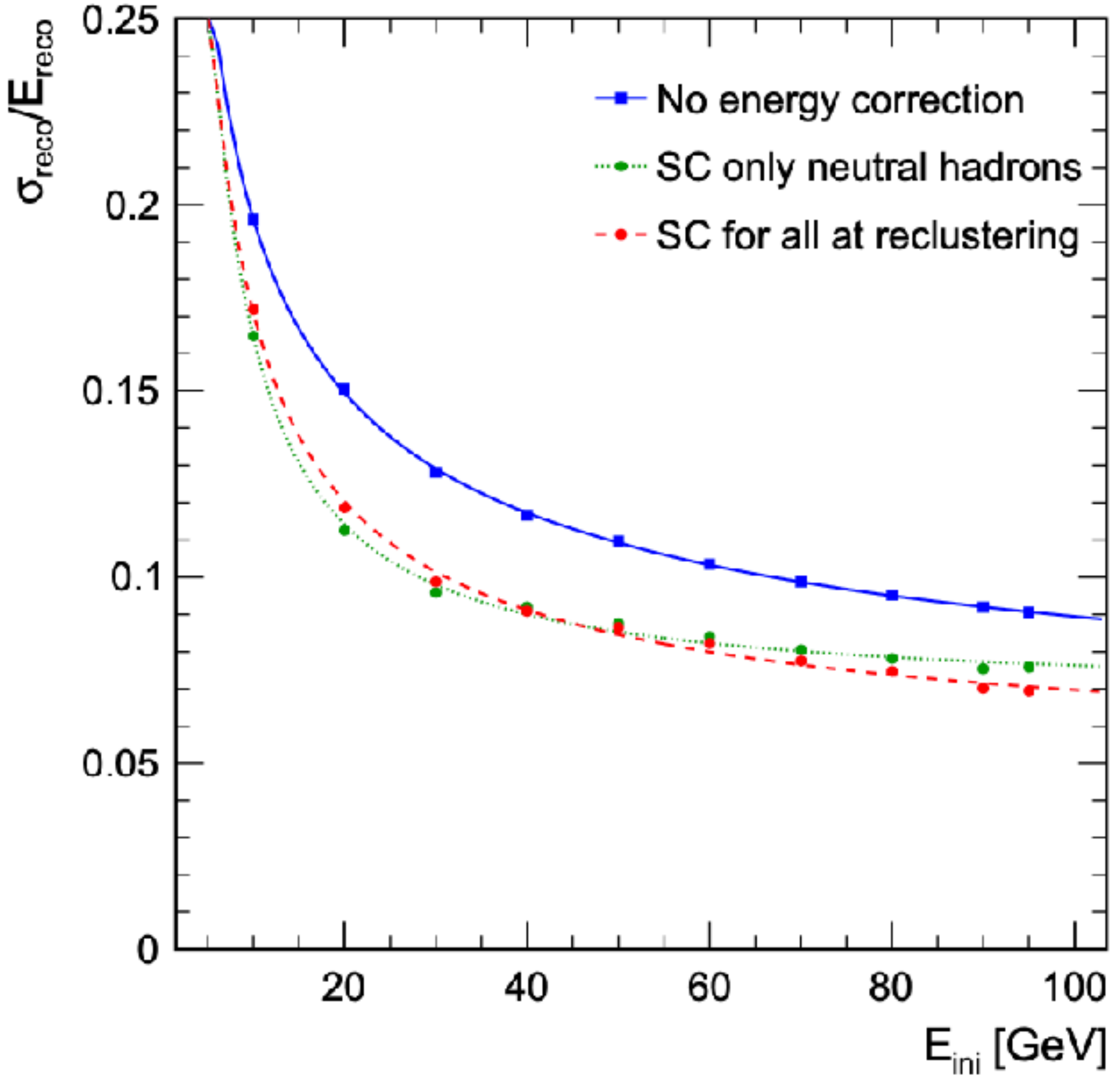




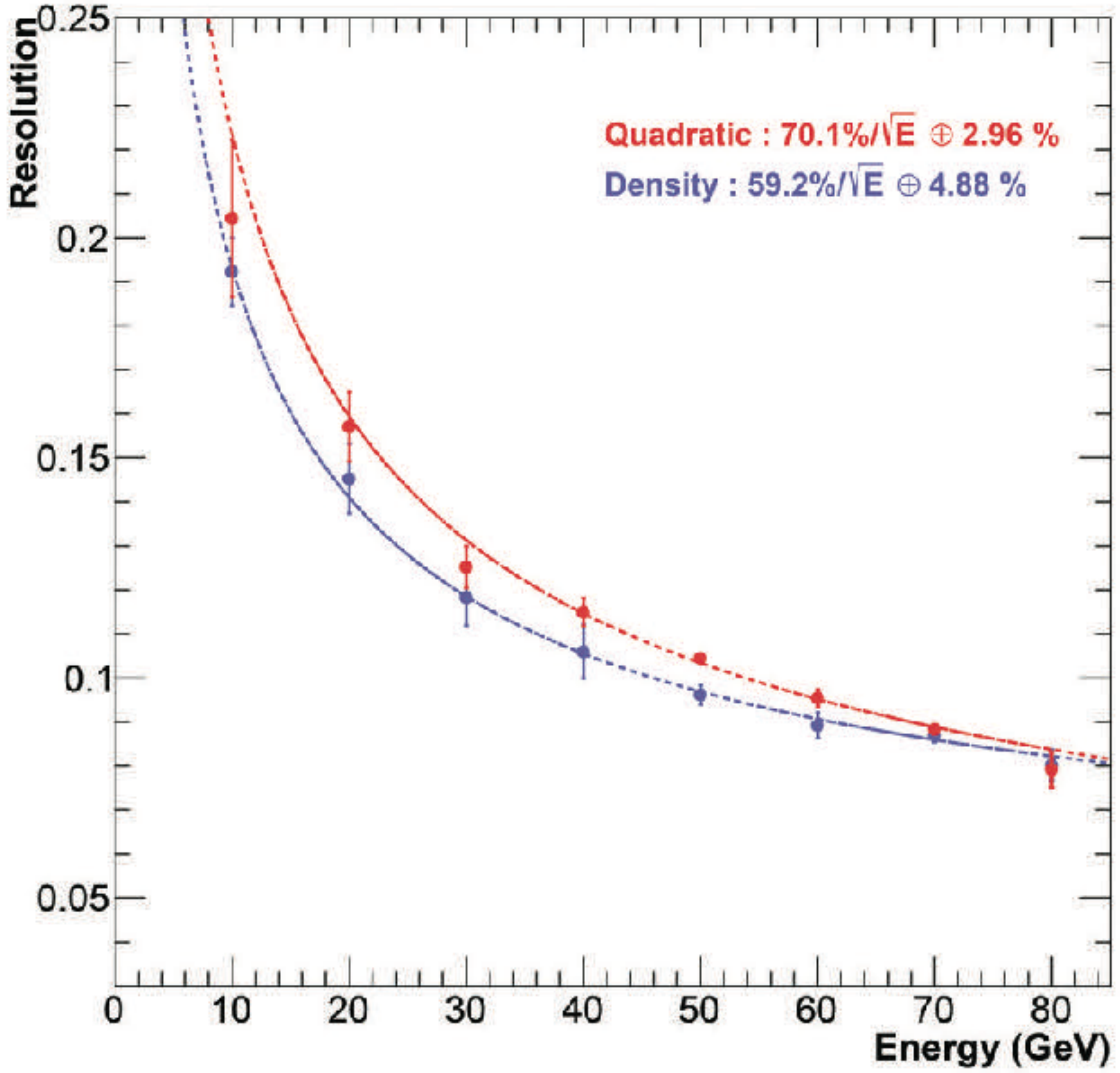


Software compensation

- Two HCAL options at ILD: Analog HCAL (AHCAL) and Semi-digital HCAL (SDHCAL)



AHCAL



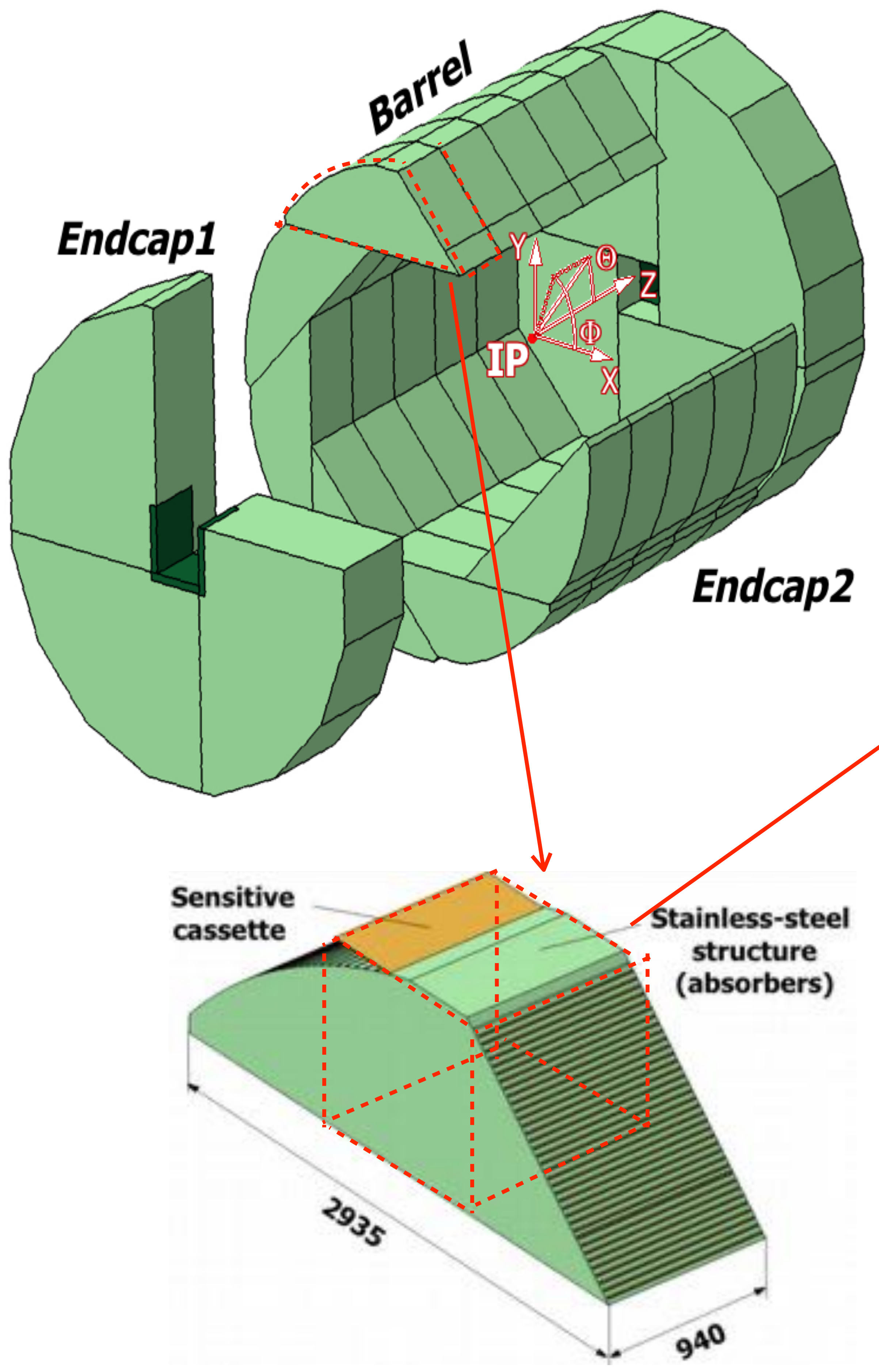
SDHCAL

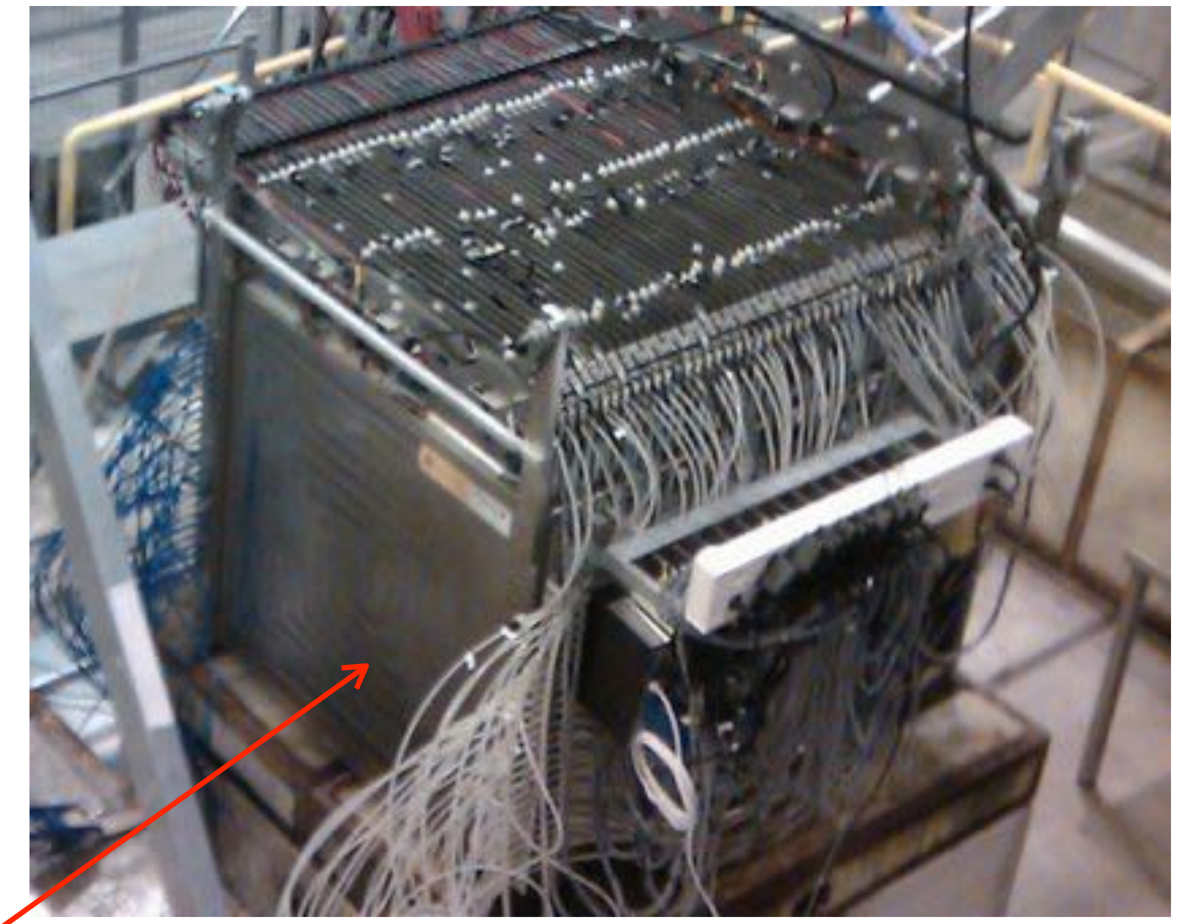
Energy resolution

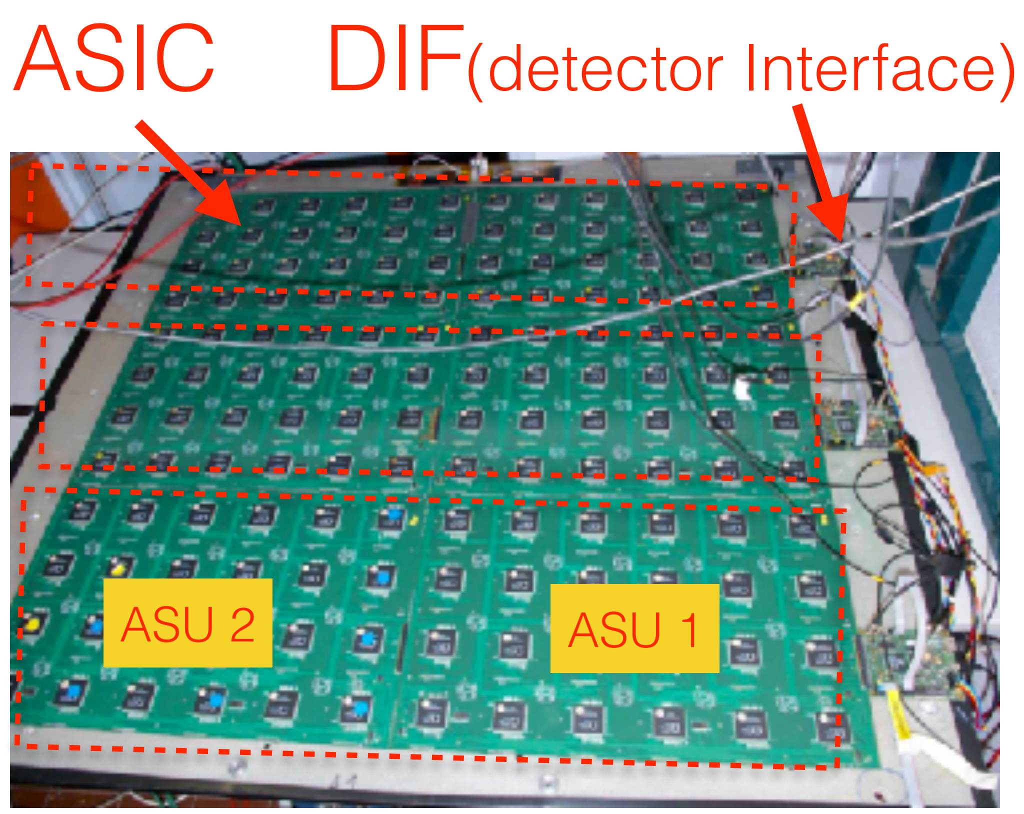
$$\frac{\sigma_E}{E} \approx 0.60/\sqrt{E}$$

Semi-Digital HCAL

arXiv:1602.02276

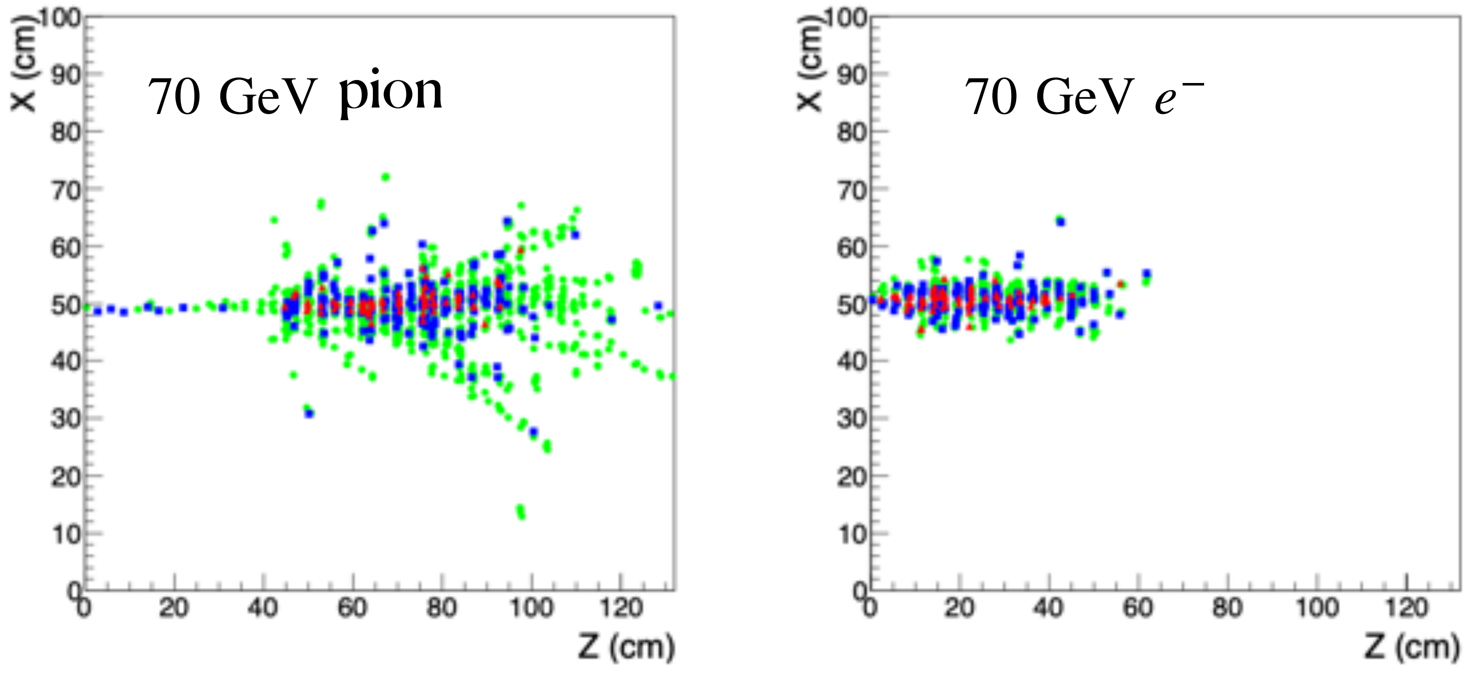


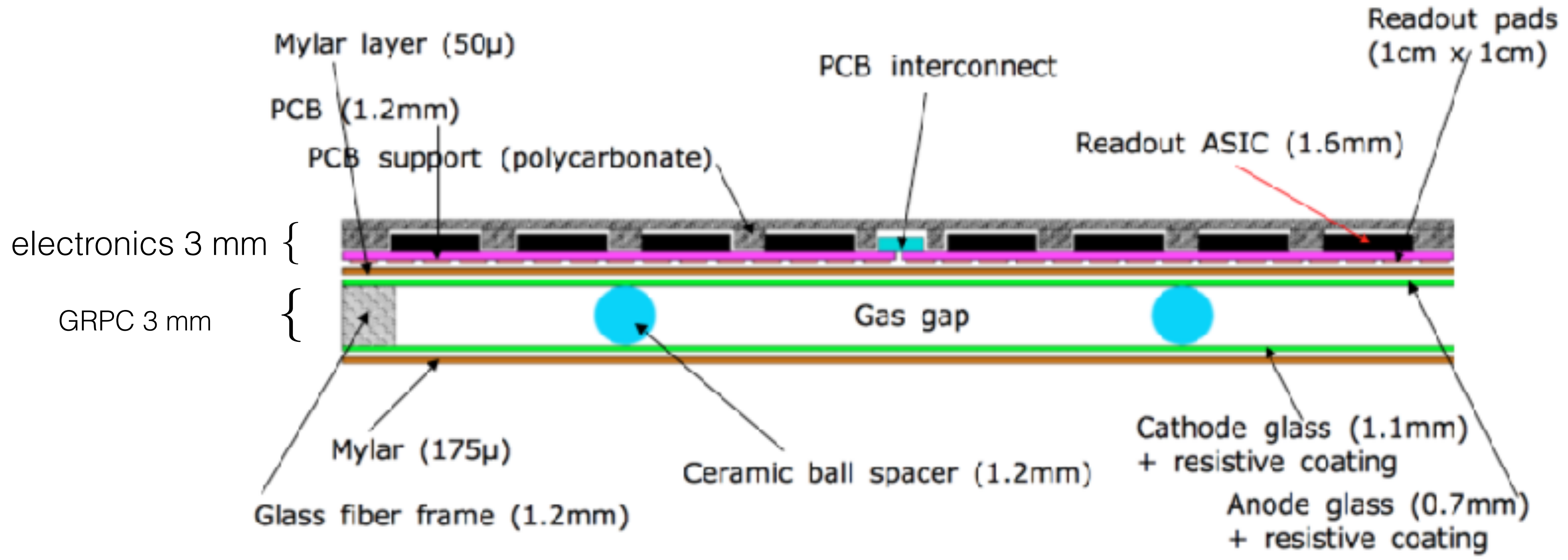




- 48 layers, 6 λ_I
- GRPC , cell size : 1×1 cm²
- One layer ASIC: 12 x 12
- Each ASIC: 64 ch. , 9612 ch. in total
- Three thresholds readout (2 bits): (0.11, 5, 15) pC
- Power-pulsing electronics, power reduction factor: ~100, 25μW/ch .

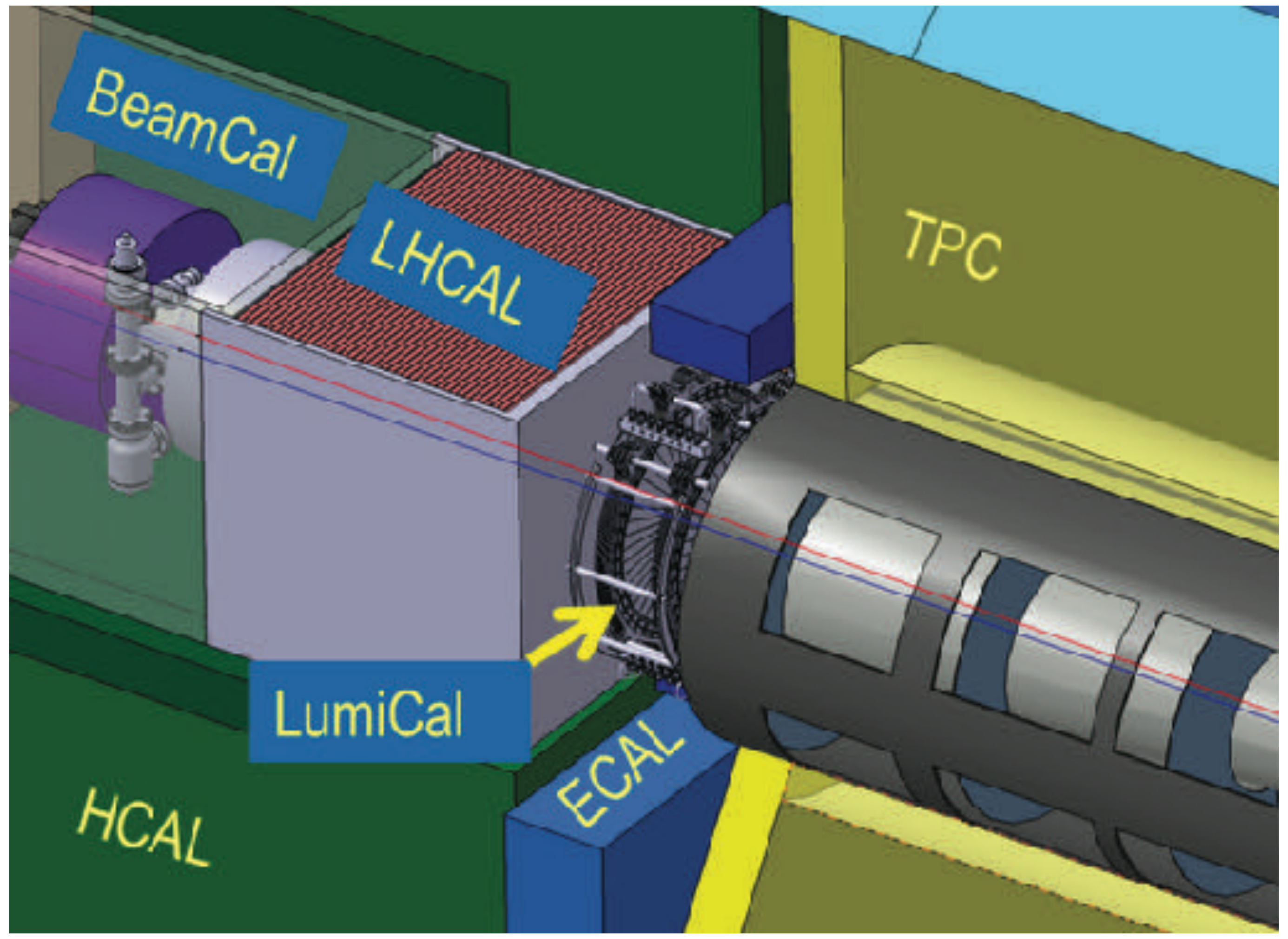
- Self-supporting mechanical structure as absorber as well
- Very compact, with negligible dead zones

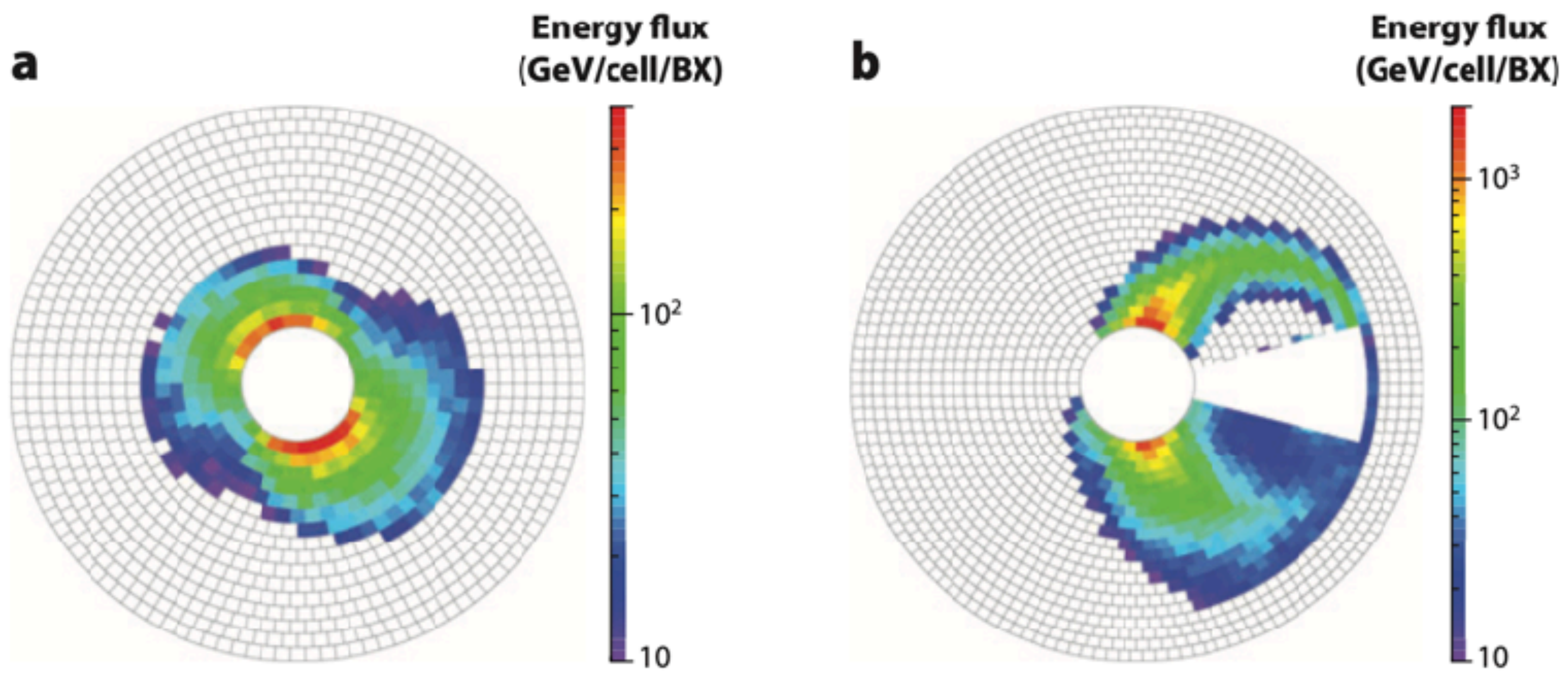




Forward calorimetry

- BeamCal and LumiCal provide fast and precise measurements of the beam properties and luminosity and to ensure detector hermeticity.
- The expected radiation dose is up to 10 MGy per year for the nominal accelerator parameters at a center-of-mass energy of 500 GeV
- CVD diamond sensor is the primary candidate
  - Radiation tolerance - no frequent replacements
  - Low dielectric constant - low capacitance
  - Low leakage current - low readout noise
  - Good insulating properties - large active area
  - Room temperature operation - no cooling necessary
  - Fast signal collection time – no ballistic deficit
  - *Smaller signal than silicon – larger energy to create eh-pair*
- Silicon sensors is also considered

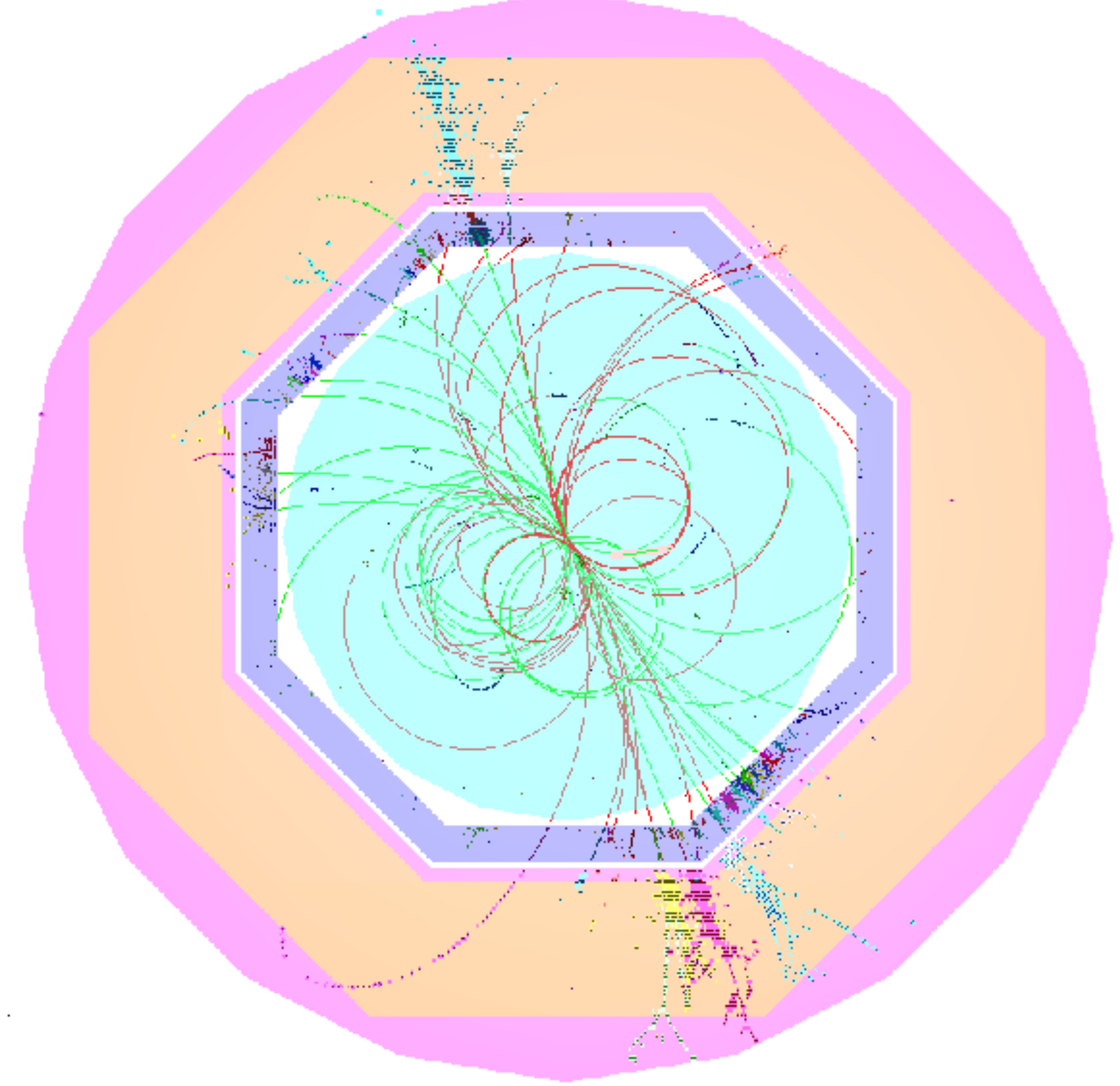


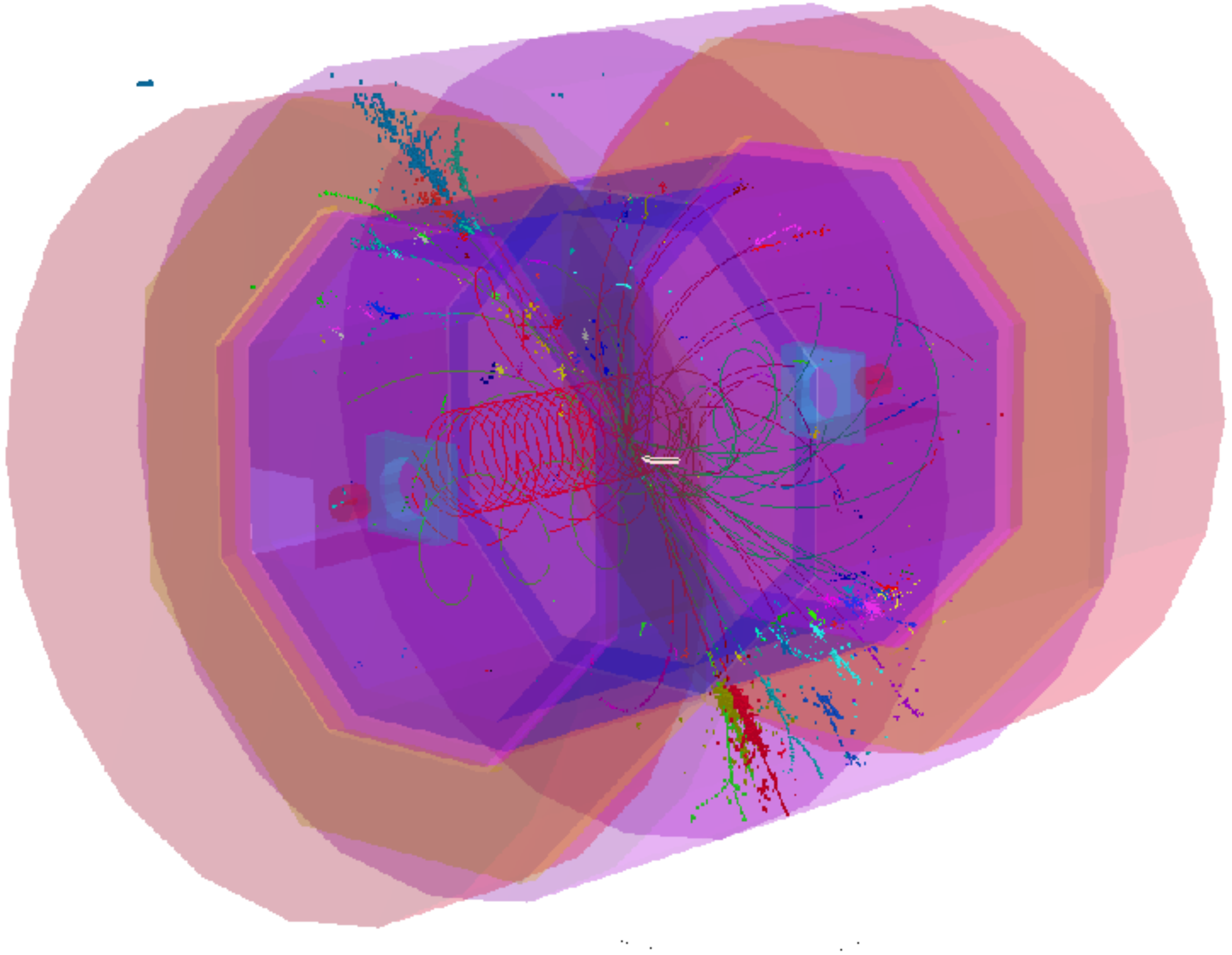


2-mrad crossing angle

20-mrad crossing angle

The necessity of Particle Flow Algorithm(PFA)





- Cluster building
- Track-cluster matching
- Energy correction
- PID

$$e^+e^- \to q\bar{q}, E_{\rm cm} = 500\ \text{GeV}$$

ILD_l5_o2_v02: SiW ECAL + SDHCAL

Particle flow calorimetry = High precision tracker + high granularity calorimeter + particle flow algorithm

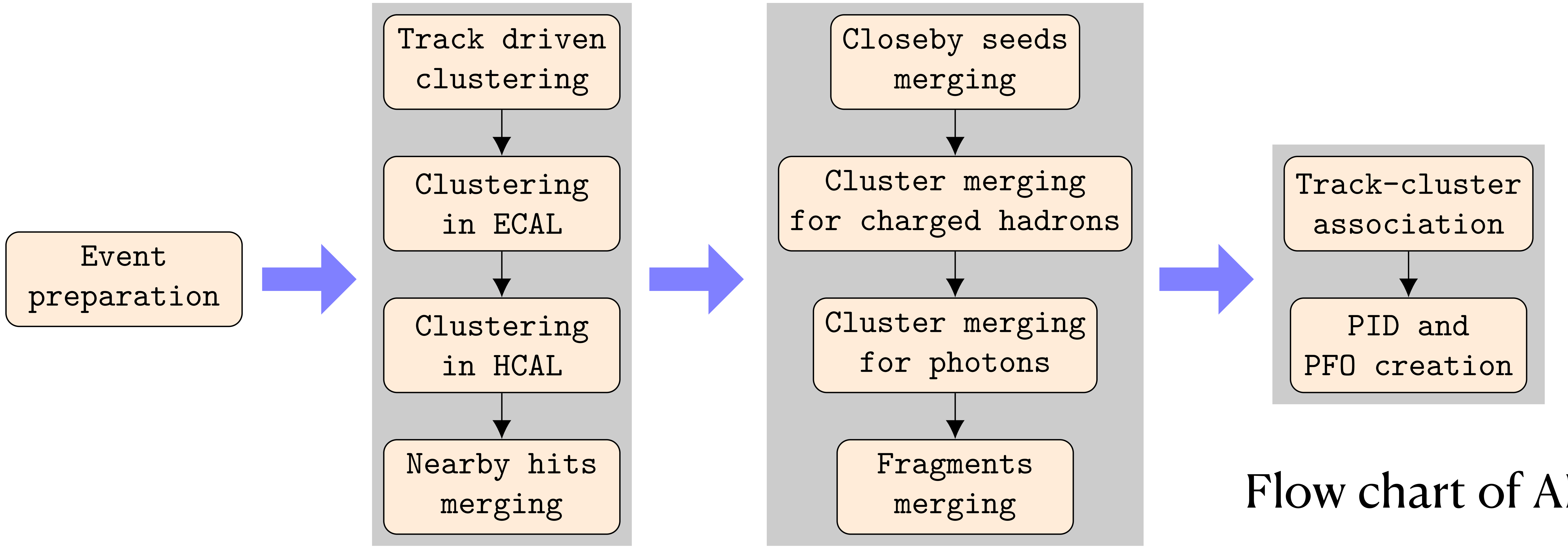
PFAs

The PFAs for e^+e^- collider experiments

- **PandoraPFA**: the standard PFA in ILCSoft. It derived a general pattern recognition package, **PandoraSDK**. NIMA 611 (2009) 25–40
- **ARBOR**, originated from LEP era; now used for the CEPC studies. arXiv:1403.4784, Eur. Phys. J. C78 (2018) no. 5, 426
- **GARLIC** (GAmma Reconstruction at LInear Collider experiment): MVA for PID. JINST 7 (2012) P06003
- **APRIL** (Algorithm of Particle Reconstruction for the ILC, Lyon): use PandoraSDK as the framework, and the basic idea of ARBOR for clustering. arXiv:2002.09678

Jet energy composition

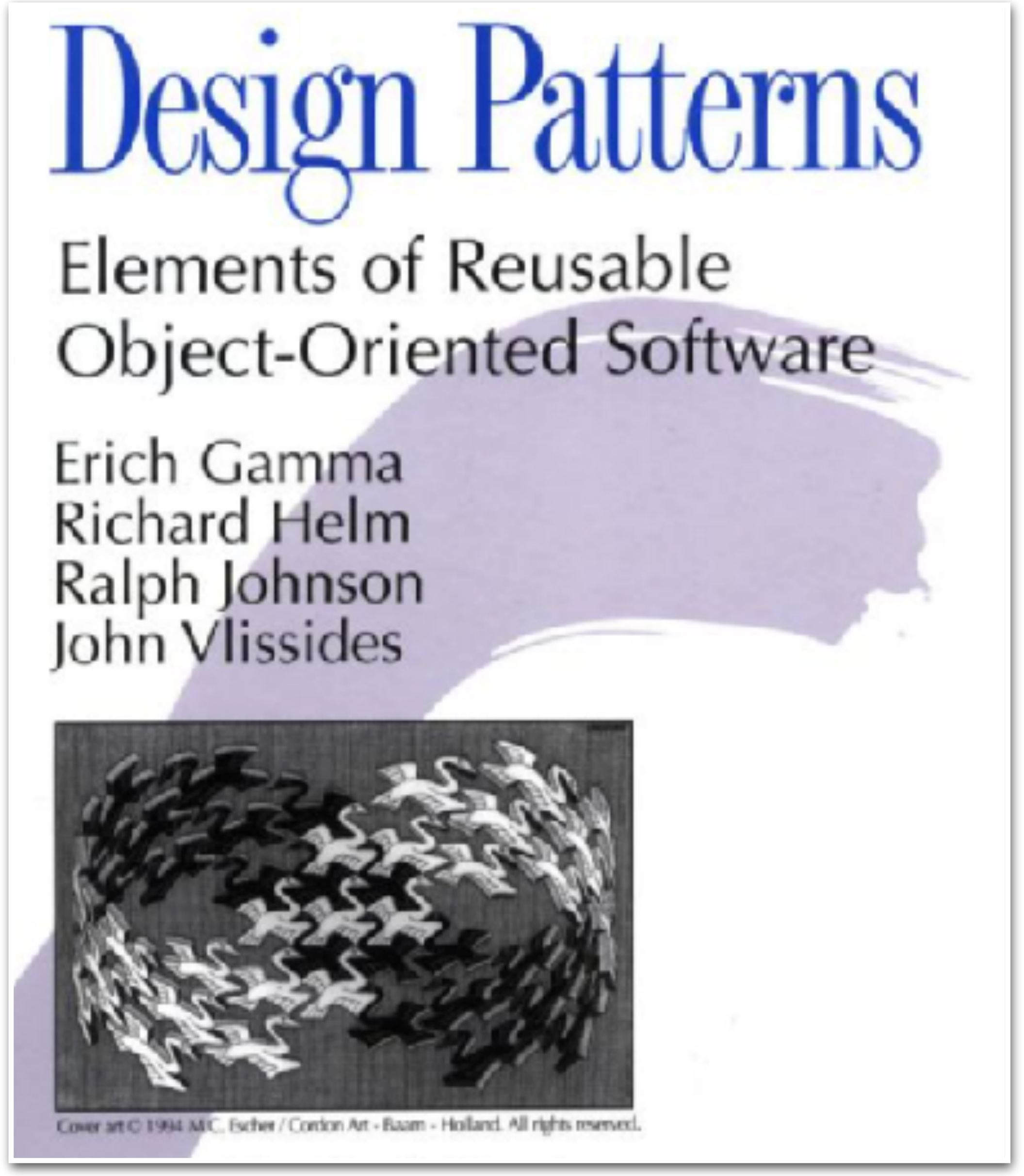
- Charged particles: ~ 60%
- Photons: ~ 30%
- Neutral hadrons: ~ 10%

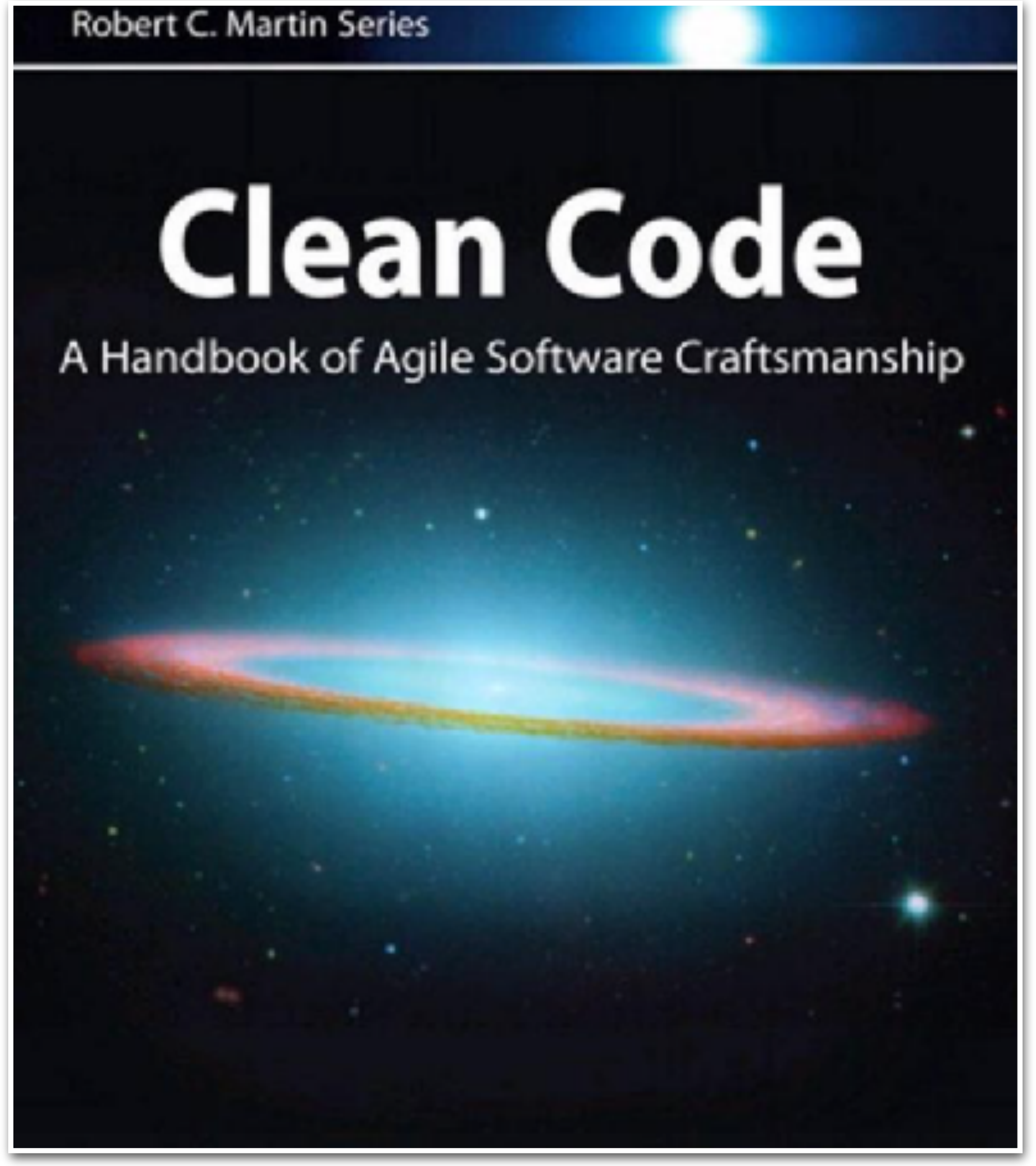


Flow chart of APRIL

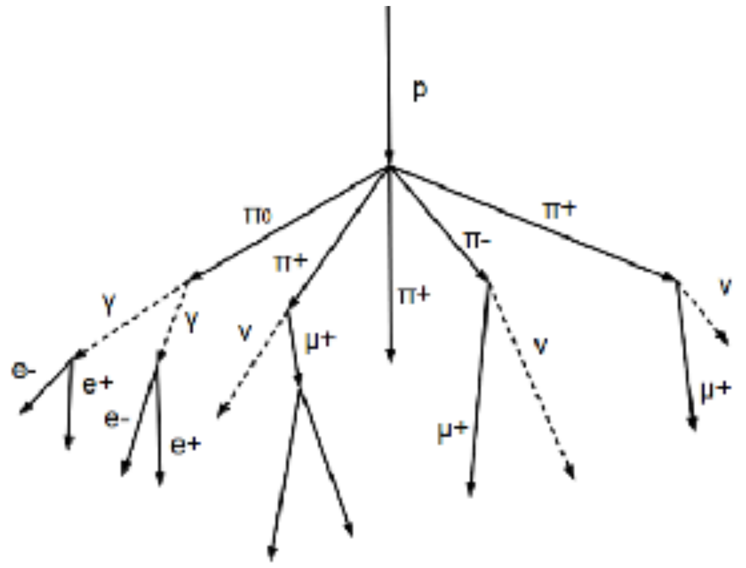
The code dependency of APRIL

- Algorithms developed by using the **PandoraSDK**
  - Multi-algorithm approach Eur. Phys. J. C75 (9) (2015) 439
  - Objects: track, hit, cluster, PFO
- **ILCSoft** (https://github.com/iLCSoft)
  - *Marlin*: the reconstruction framework
  - Tracking packages for all components of tracker
  - Calorimeter digitizers (*SimDigital* for SDHCAL)
  - *lcgeo*: the ILD detector model, which is based on DD4hep
  - *LCCalibration*: automated energy calibration for calorimeters at ILC
- **mlpack**: *NeighborSearch* for nearby hits search; *DBSCAN* for the clustering of unusual hits

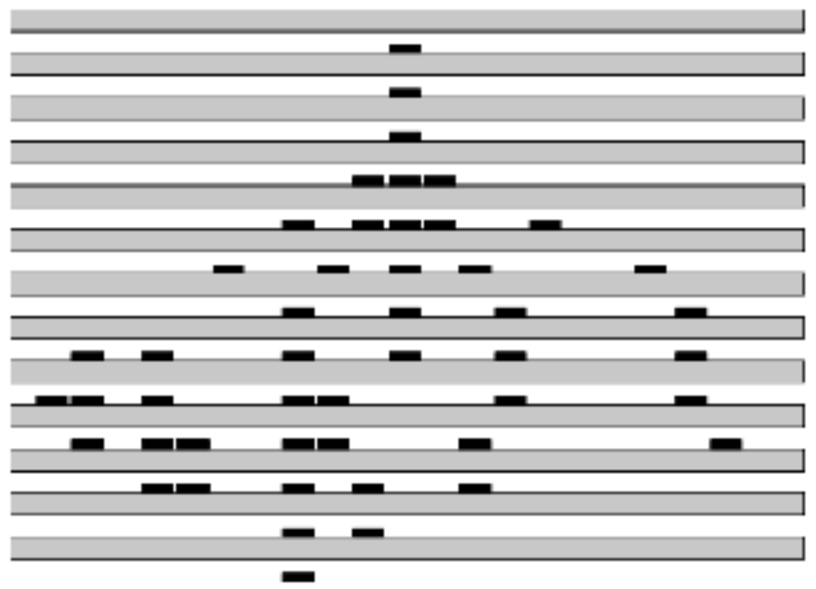




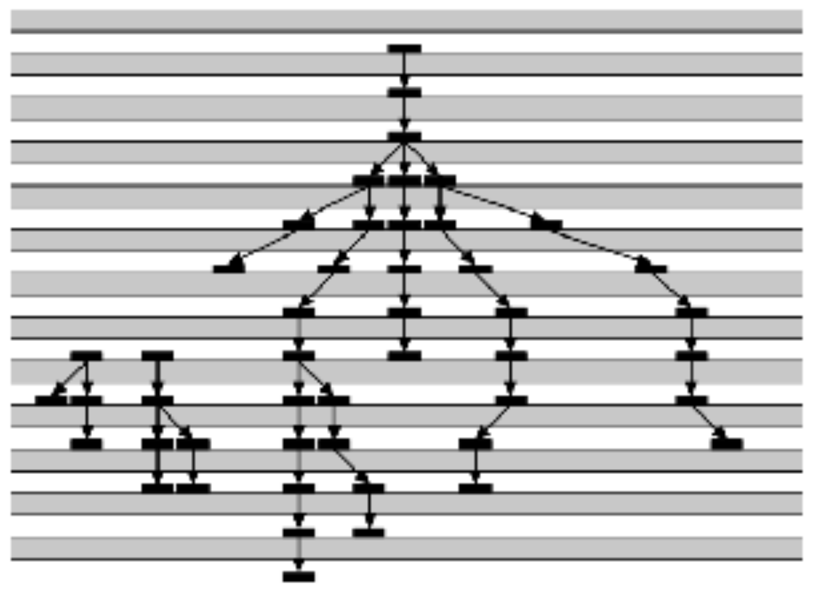
Clustering



a shower



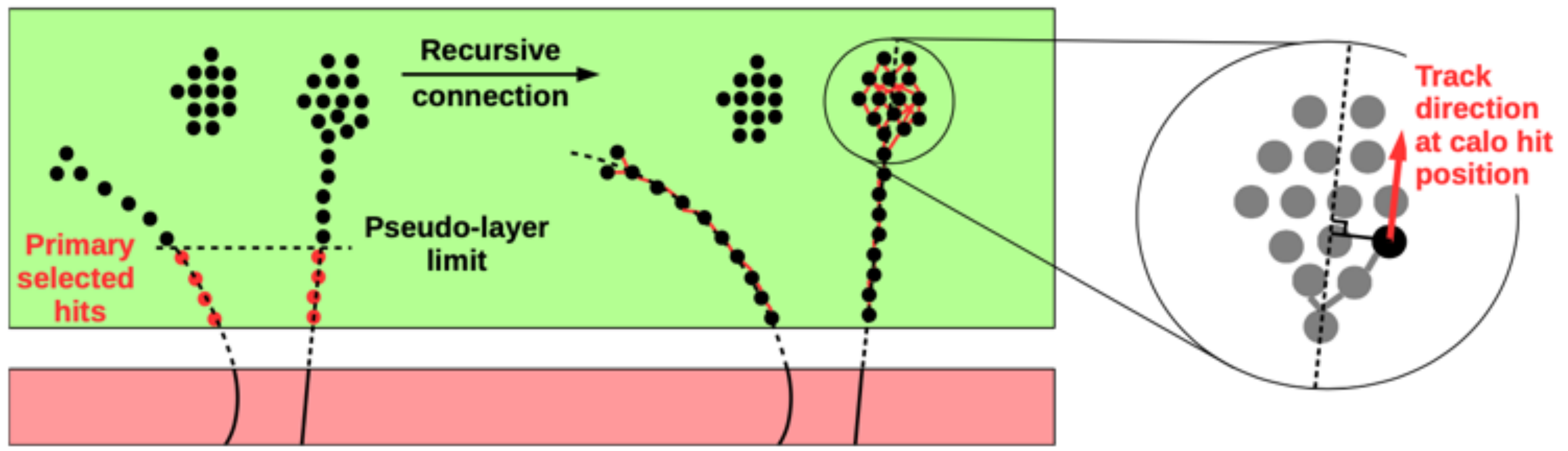
hits in calo



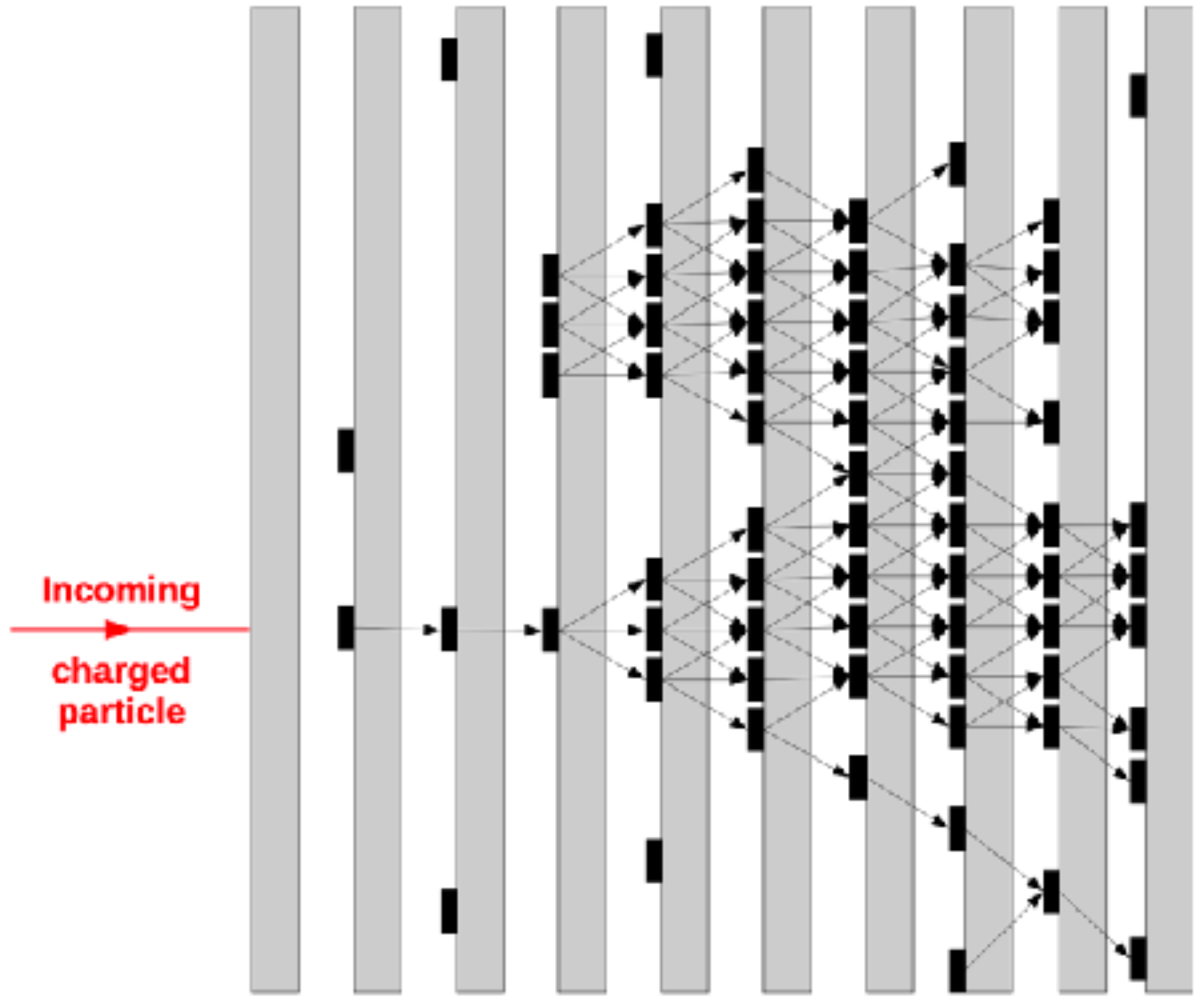
reconstruction as
tree topology

The basic clustering idea of ARBOR

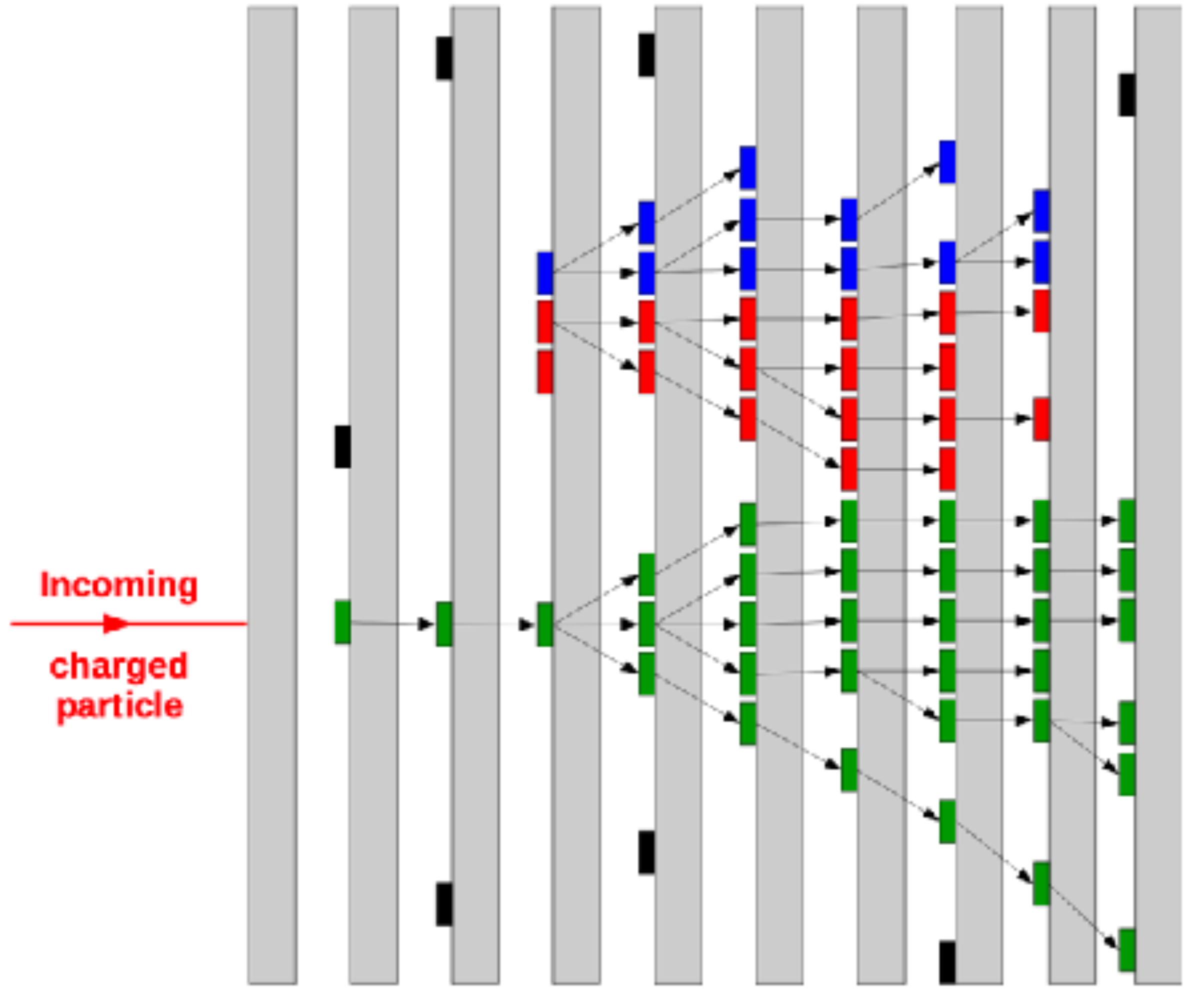
Track-driven clustering



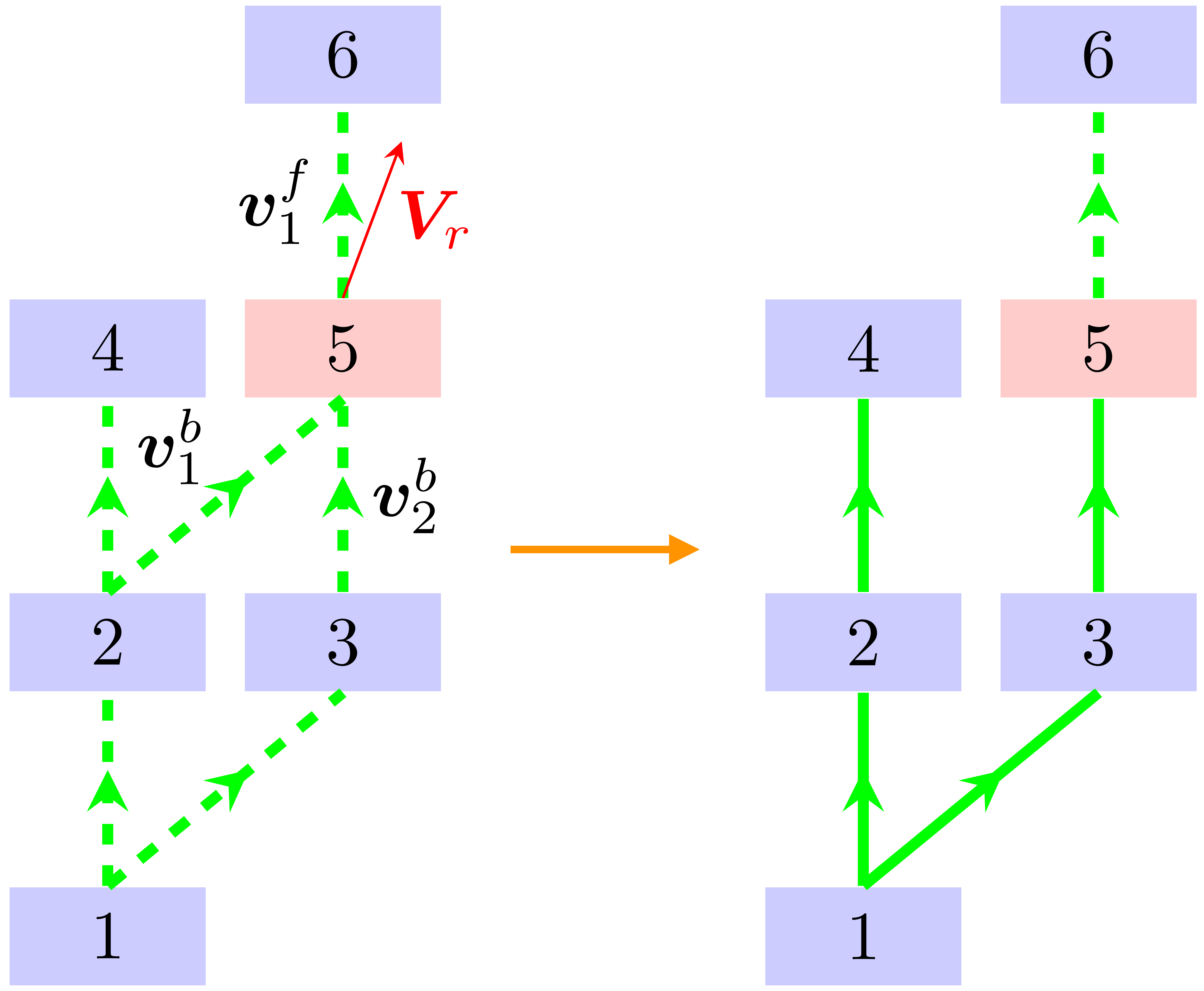
Tracking in the calorimeter !



clean connections



Hits connection



‘Small’ parameters are chosen, so the clustering at this stage has small error on merging two clusters into a single one

- Reference direction

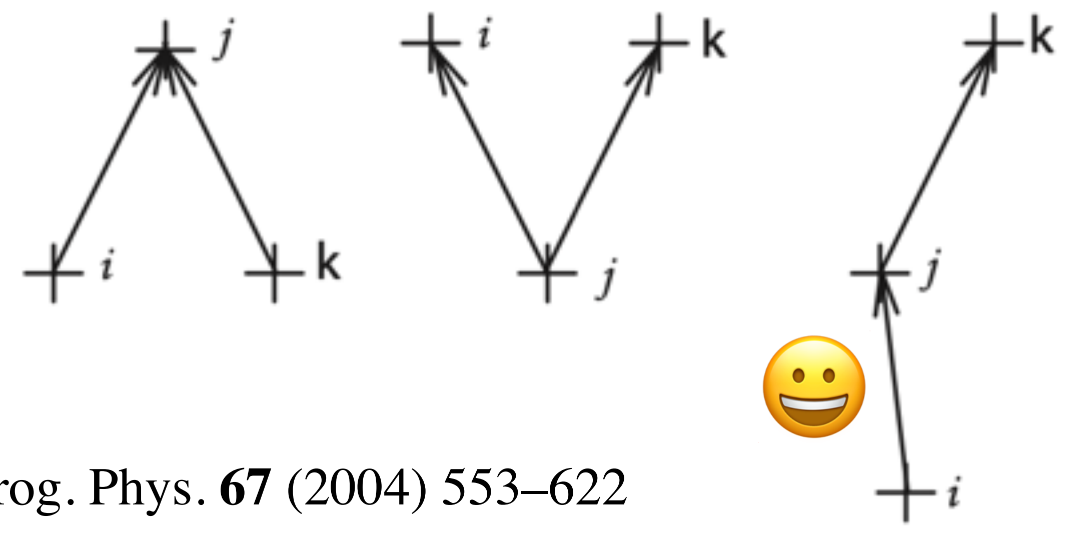
$$V_r = w_b \cdot \sum_i \boldsymbol{v}_i^b + w_f \cdot \sum_j \boldsymbol{v}_j^f$$

- Connection order

$$\kappa = \theta^{p_\theta} \cdot d^{p_d}$$

- θ: angle between connection and reference direction
- d: the connection length

- It is similar to the tracking method with neural network



Rep. Prog. Phys. **67** (2004) 553–622

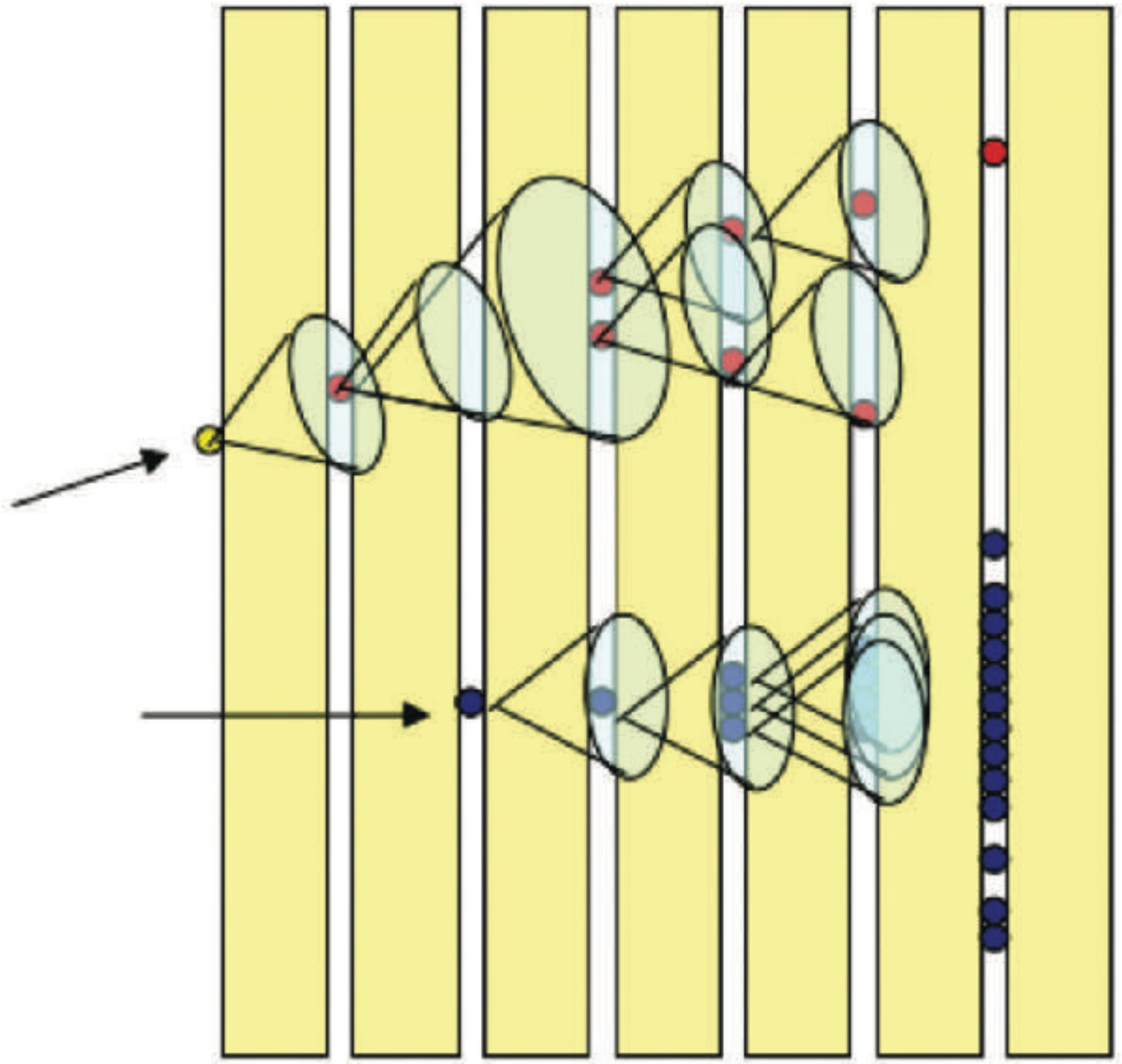
The system energy E is a function of

$$\frac{-\cos^m \theta_{ijl}}{d_{ij} + d_{jl}}$$

In principle, this clustering method of ARBOR can be updated to NN method

The clustering in PandoraPFA

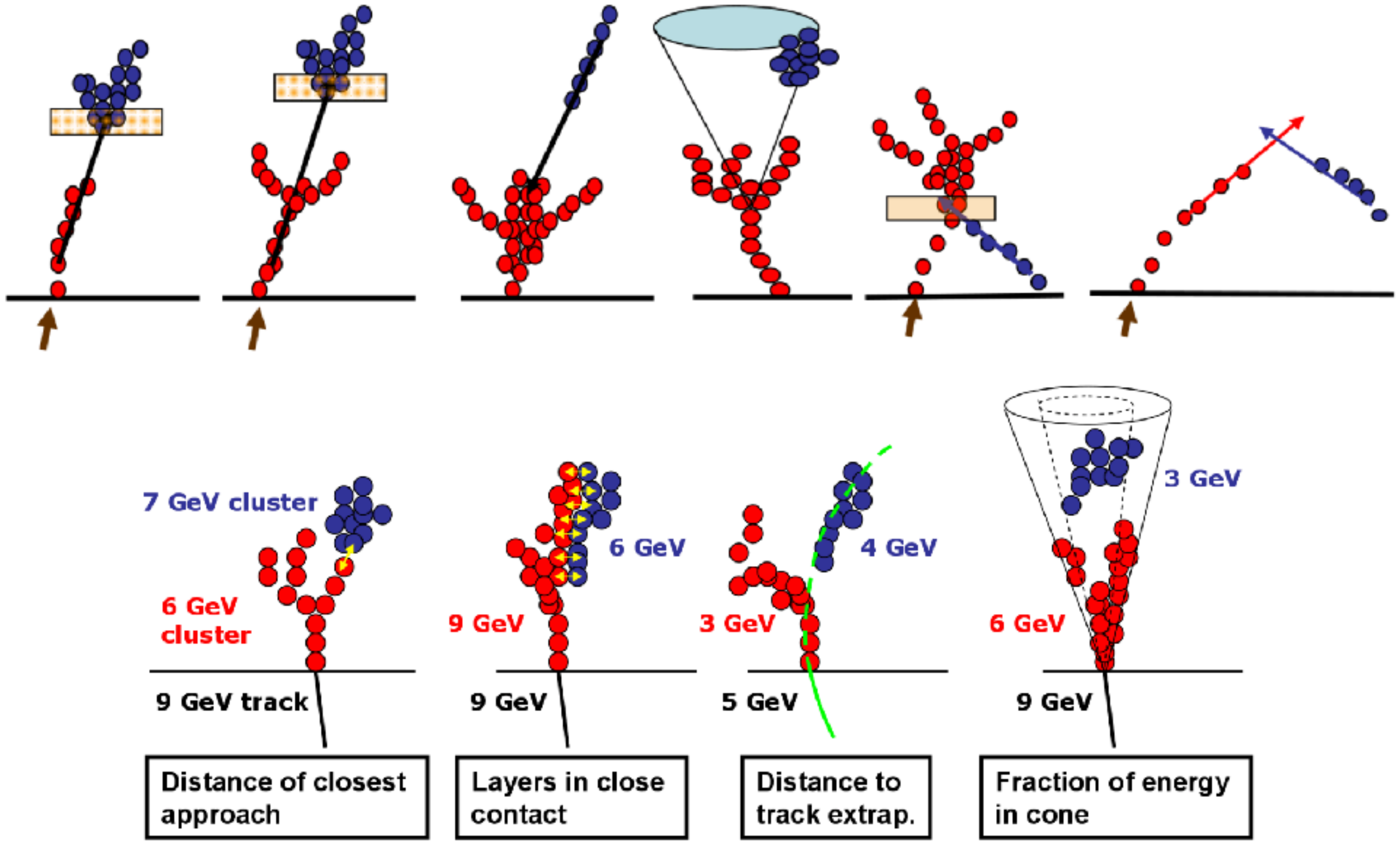
Nucl. Instrum. Meth. A611 (2009) 25–40



Cone-based clustering

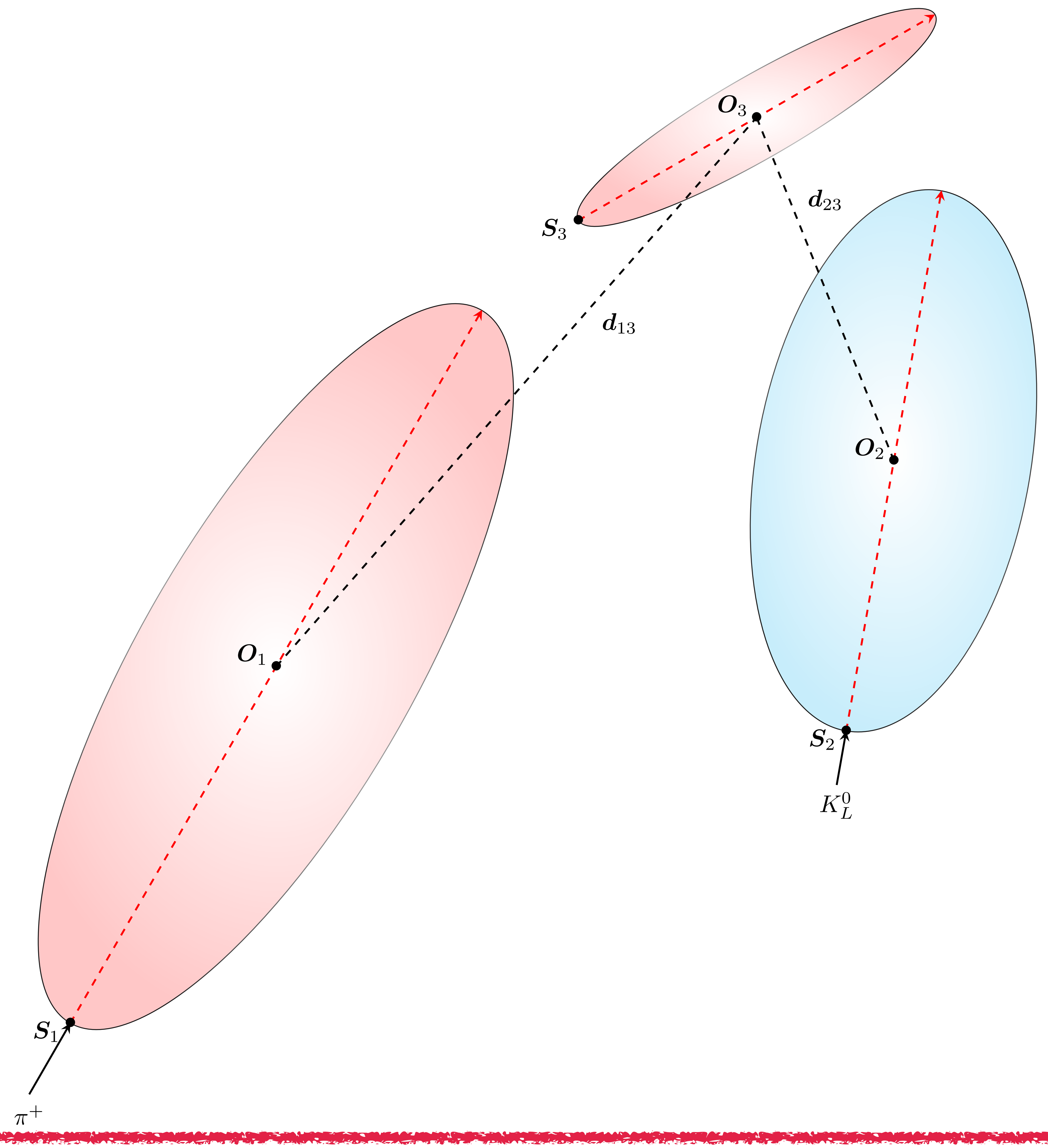
The cone parameters: cone angle and distance

A cluster merging procedure is needed after clustering. Without that, it may induce double counting.



Topological cluster merging in PandoraPFA

Cluster merging in APRIL



- The idea of ARBOR clustering for hits is extended to the cluster merging by using
  - Distance between two clusters: d
  - Cluster pointing angle: θ
- The cluster connection order is defined as that for hits

$$\kappa = \theta^{p_\theta} \cdot d^{p_d}$$

- The degree of clusters energy deviation from the track is evaluated before cluster merging

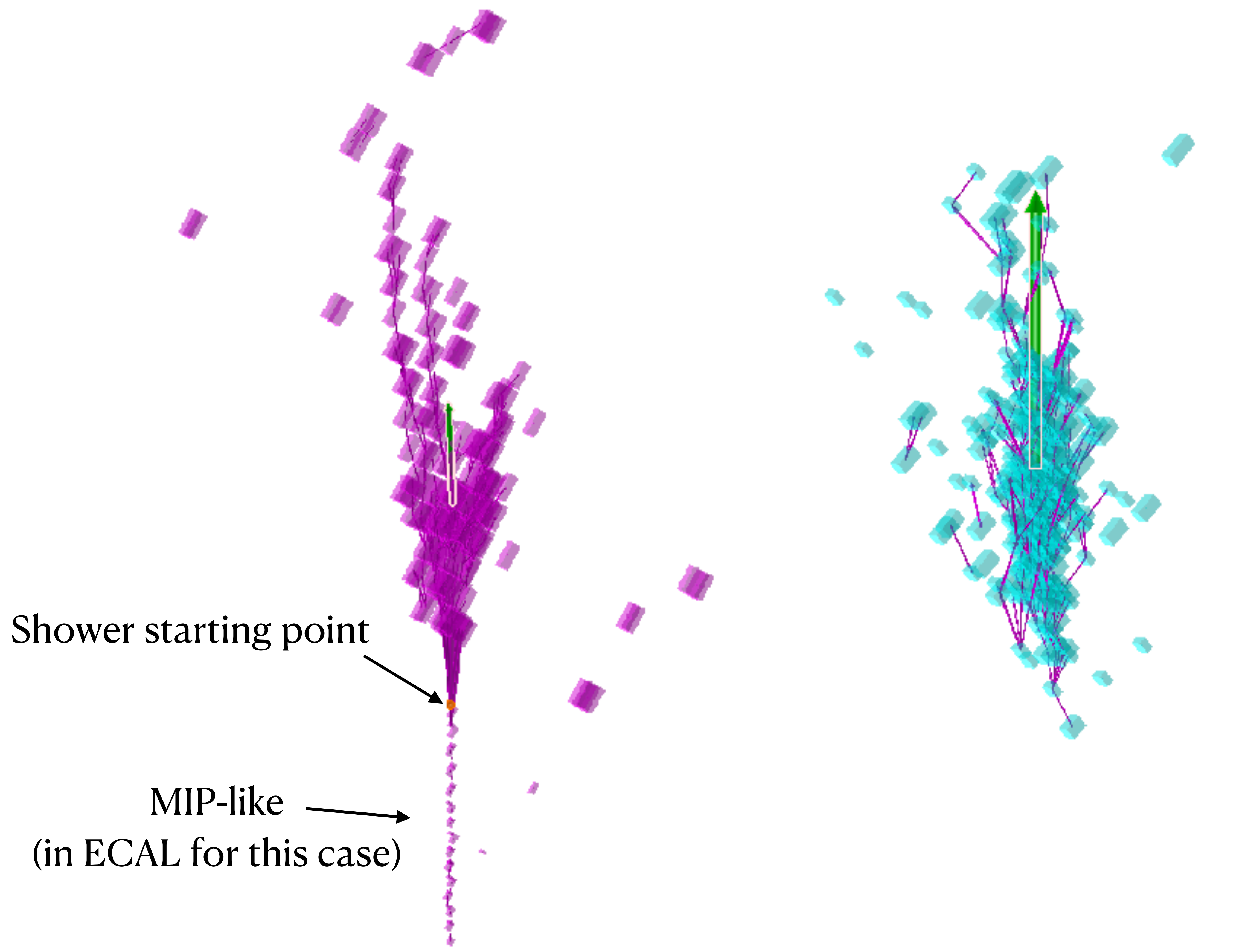
$$\chi = \frac{E_c - E_t}{\sigma_E}$$

E_t: track energy

E_c: sum of clusters energy

For the σ_E of shower, it is approximate to take $0.15/\sqrt{E_c}$ in ECAL, $0.55/\sqrt{E_c}$ in HCAL

The built clusters



10 GeV charged pion   10 GeV photon

- A cluster can be considered as a rigid body
- The **barycenter** for a cluster is defined by the energy weighted hit positions as

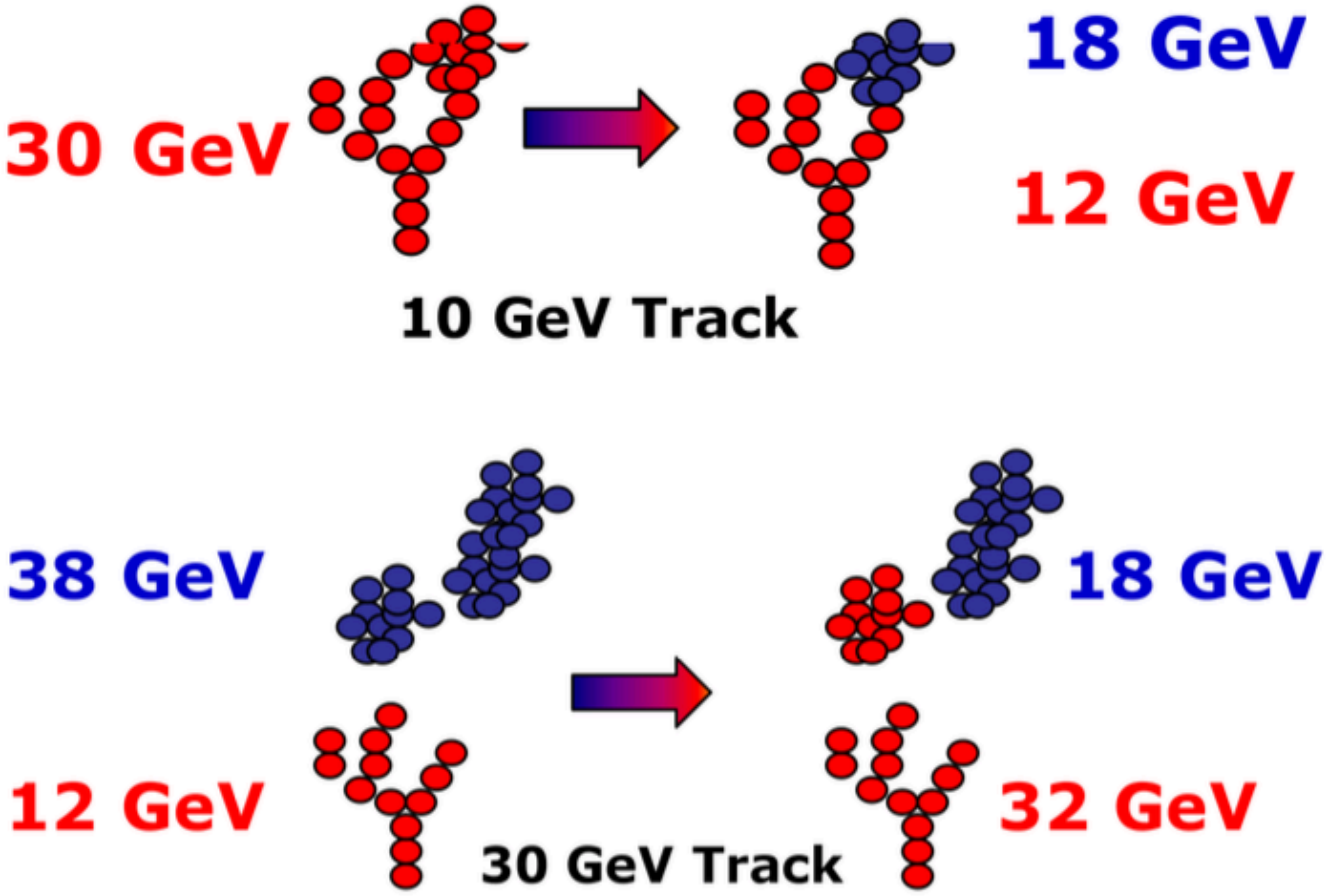
$$\boldsymbol{o} = \frac{1}{E}\sum_{k} e_k \boldsymbol{r}_k$$

- Cluster principle **axes** are obtained from the eigen vectors of cluster inertial tensor $\boldsymbol{I}$, which is defined by

$$I_{ij} = e_k \sum_{k} (\boldsymbol{r}_k^2 \delta_{ij} - \boldsymbol{r}_k^{(i)} \boldsymbol{r}_k^{(j)})$$

- The shower starting point is calculable

Reclustering in PandoraPFA



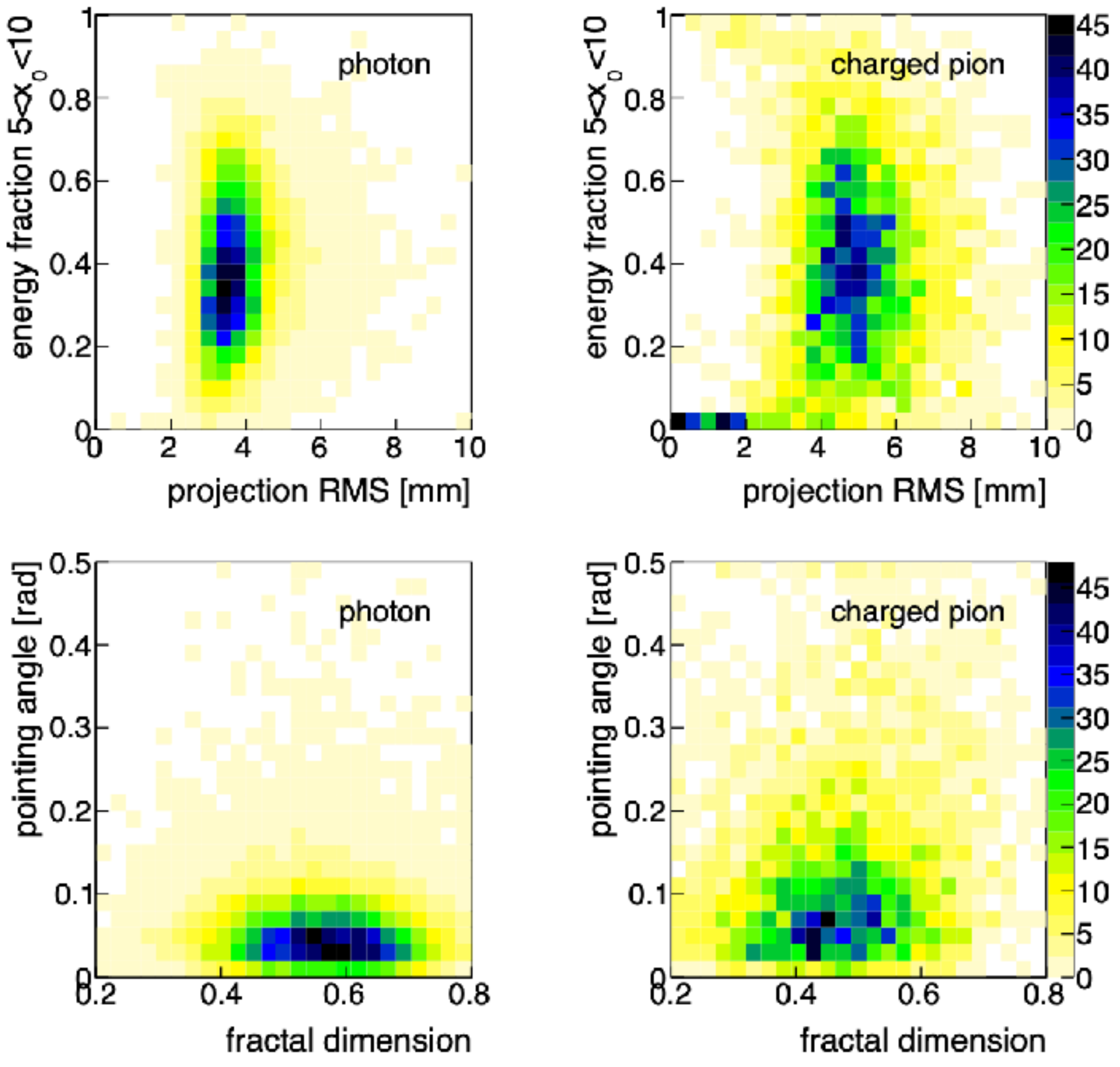
- For jet energy greater than 50 GeV, the reclustering procedure can improve the reconstruction (shower separation)
- However,
  - the performance depends on cluster energy. When the jet energy becomes high, the benefit form this seems trivial.
  - it scans 12 sets (10 parameters for each set) of clustering parameters, and take the minimal of

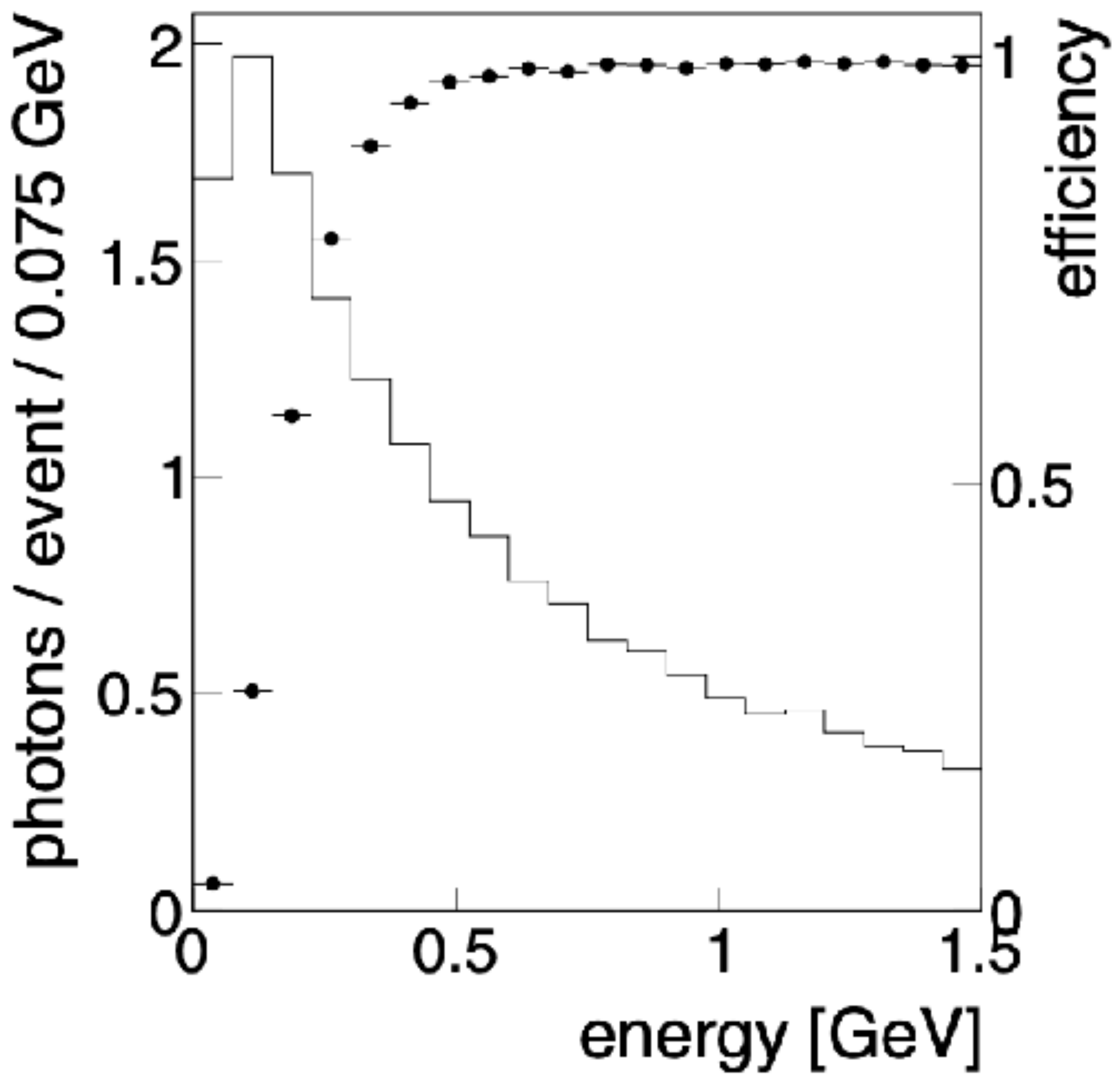
$$\chi^2 = (\frac{E_c - E_t}{\sigma_E})^2$$

The PID with MVA in GARLIC

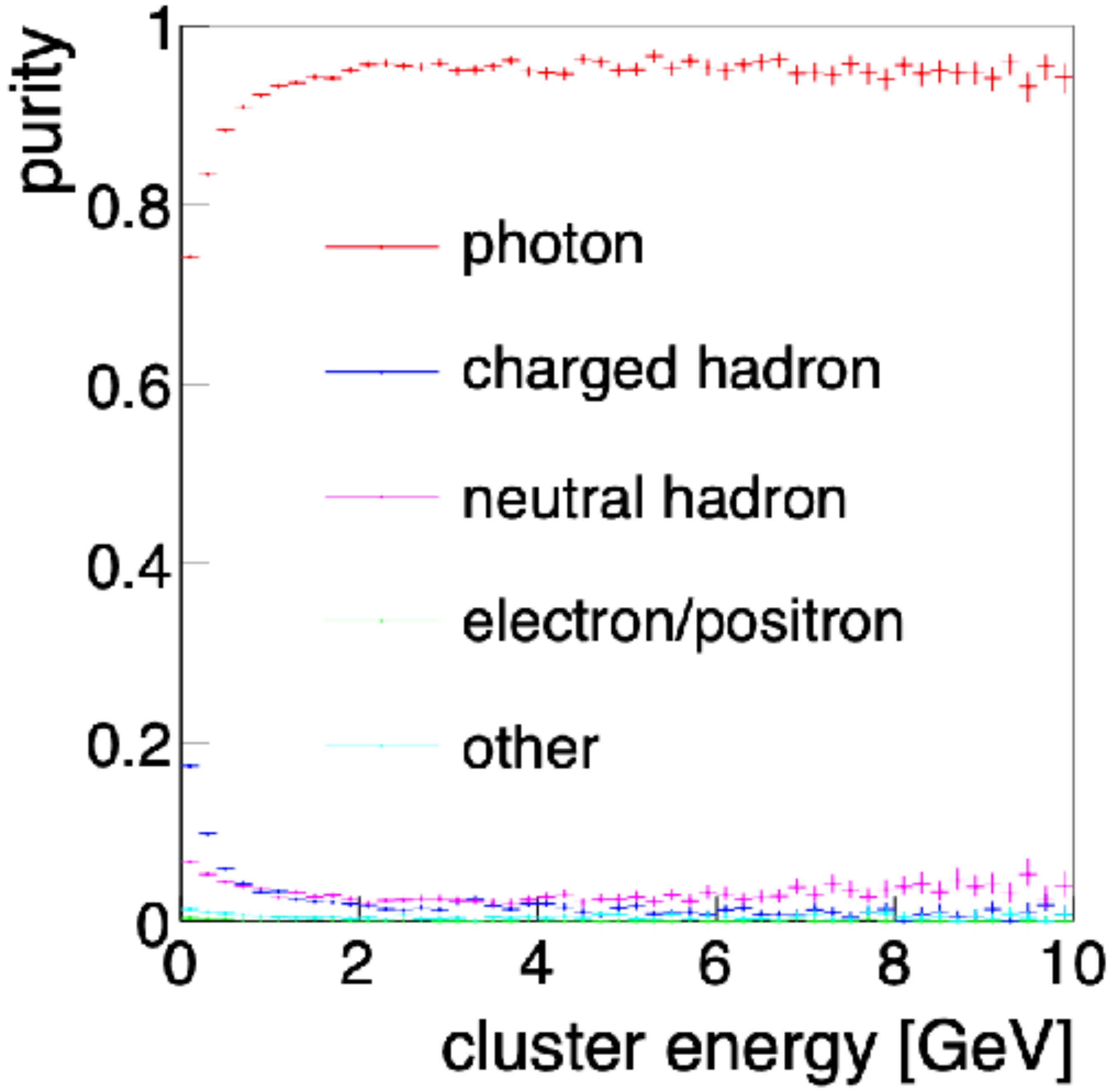
arXiv: 1203.0774

Some of input variables of MVA





Photon reconstruction efficiency

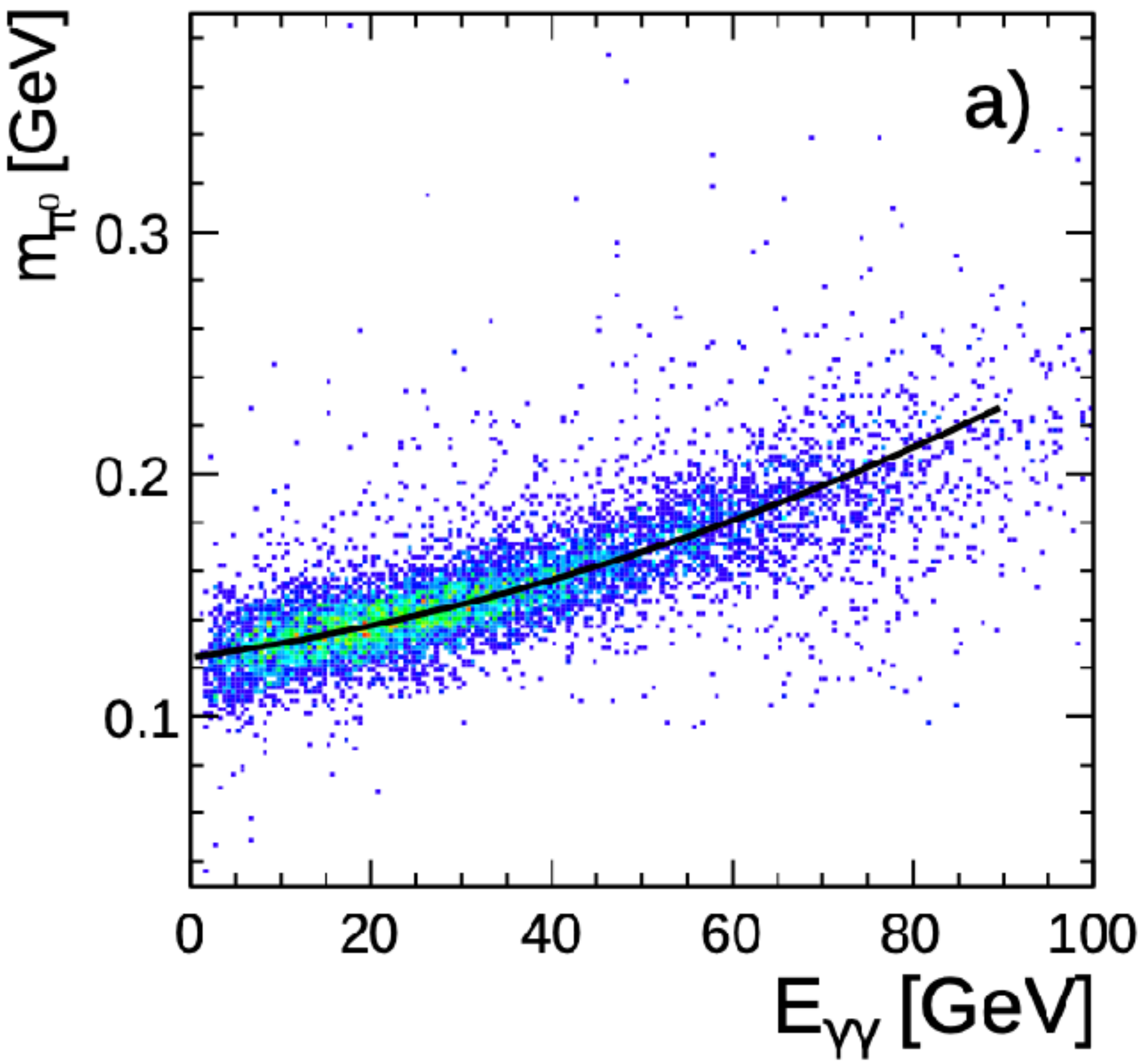


Photon purity

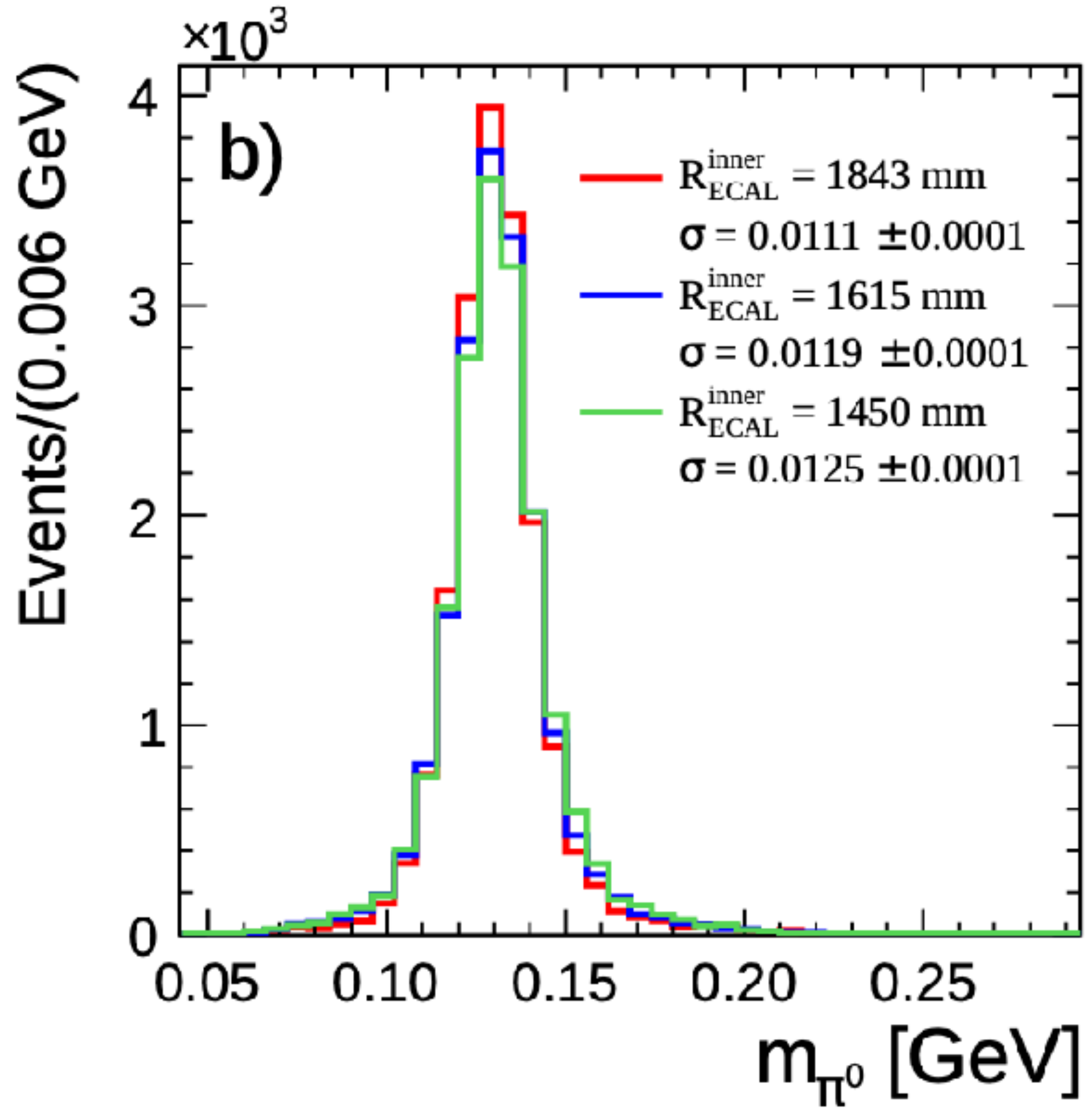
- Projection RMS: the RMS of hit distribution around the cluster axis
- Pointing angle: the angle between cluster axis and the direction of cluster barycenter
- Fractal dimension: related to the hit density

π^0 reconstruction with GARLIC

arXiv:1510.05224v2



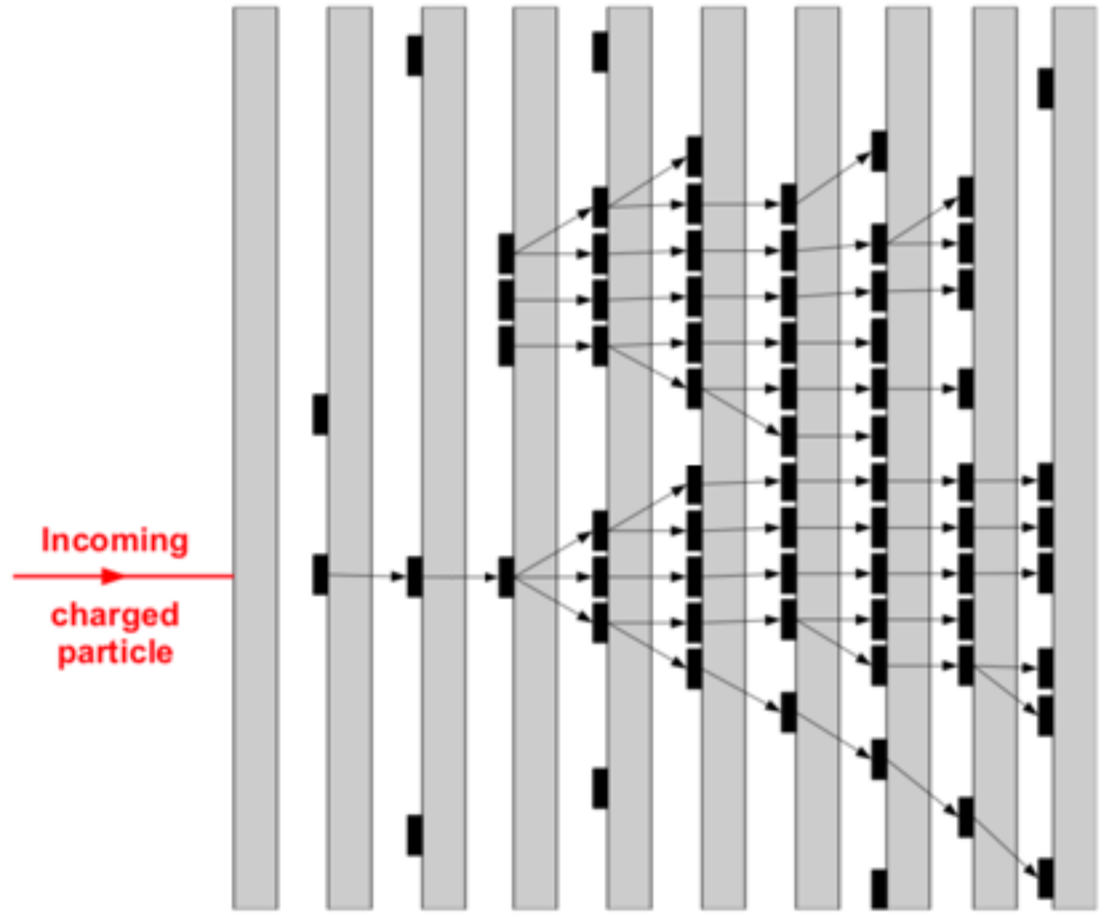
Affected by the clusters overlap !

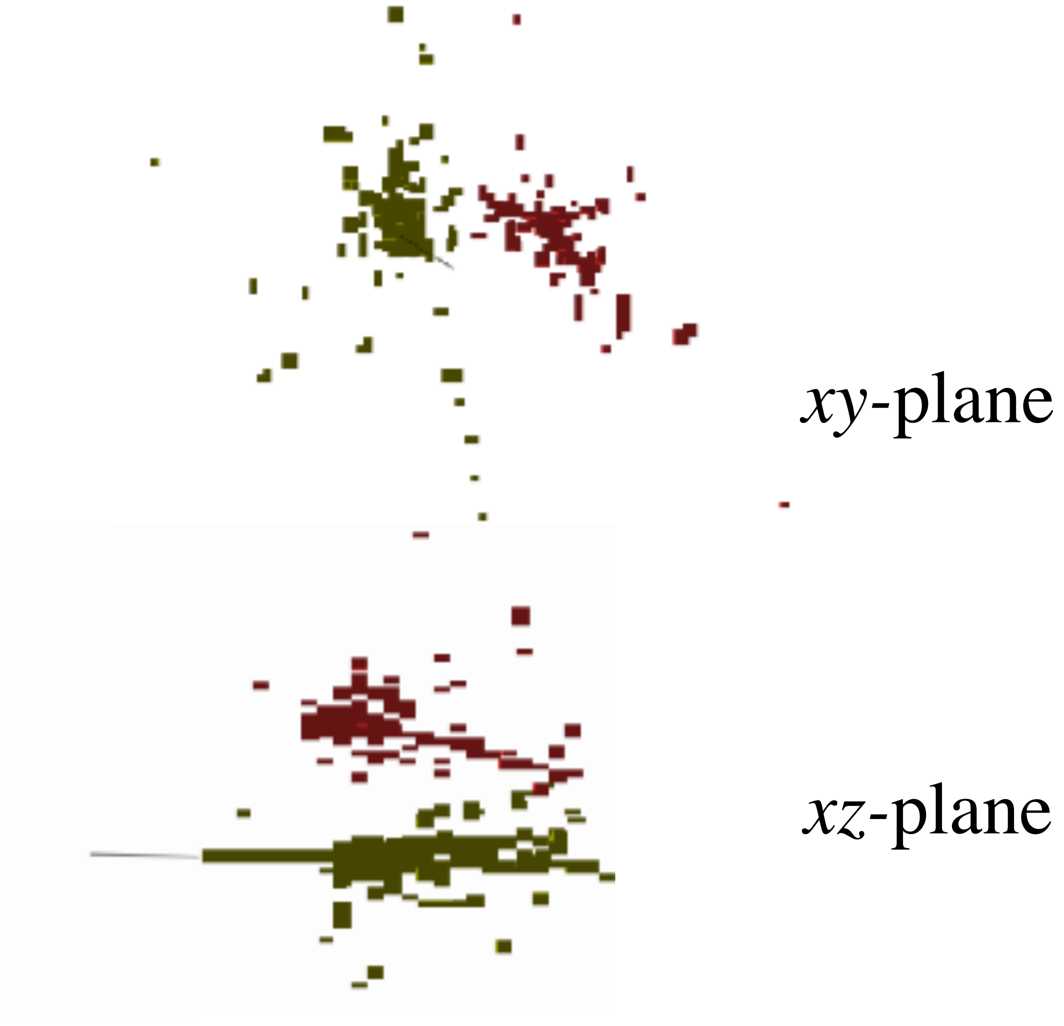


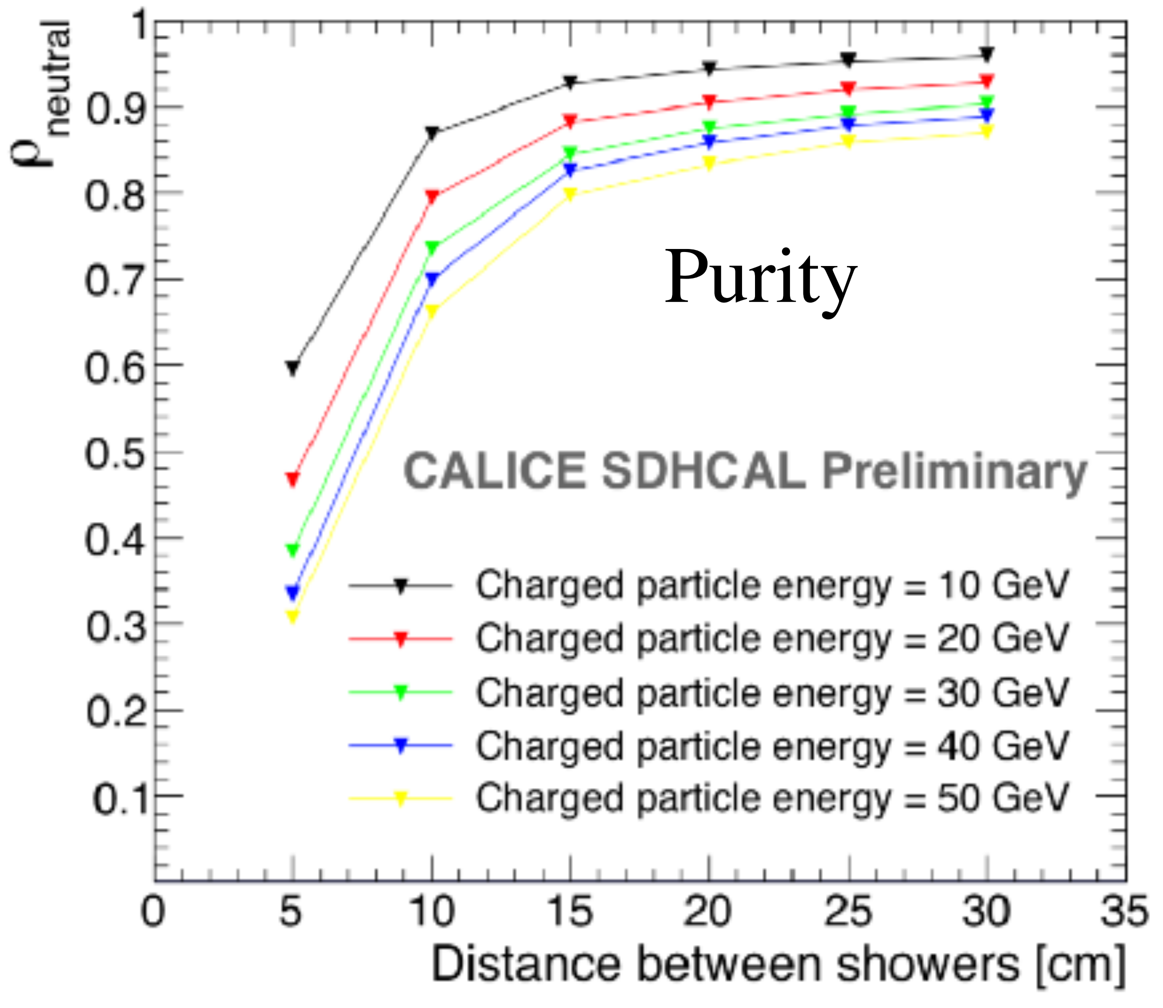
Result with energy correction

Cluster separation

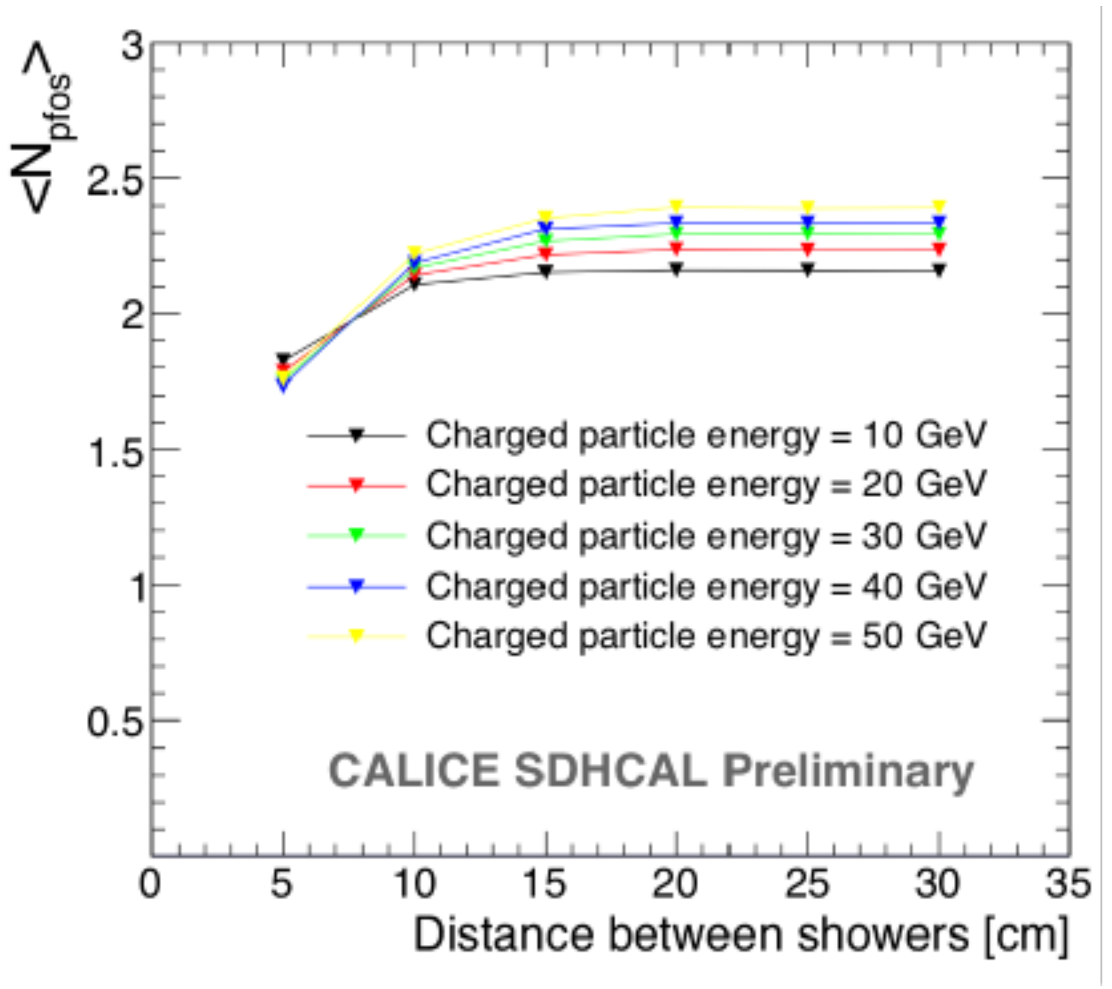
CALICE note CAN054

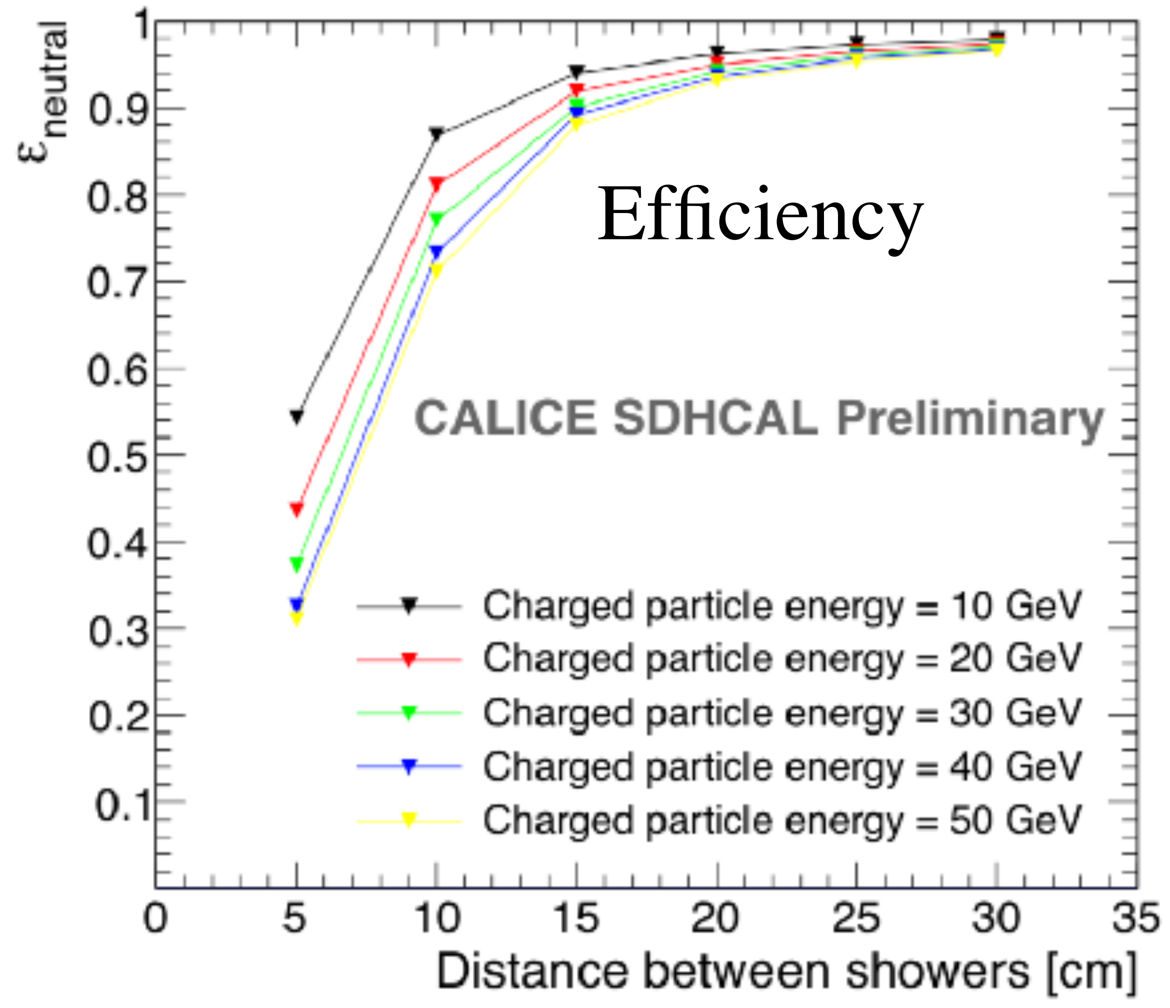






10 GeV neutral kaon + charged pion

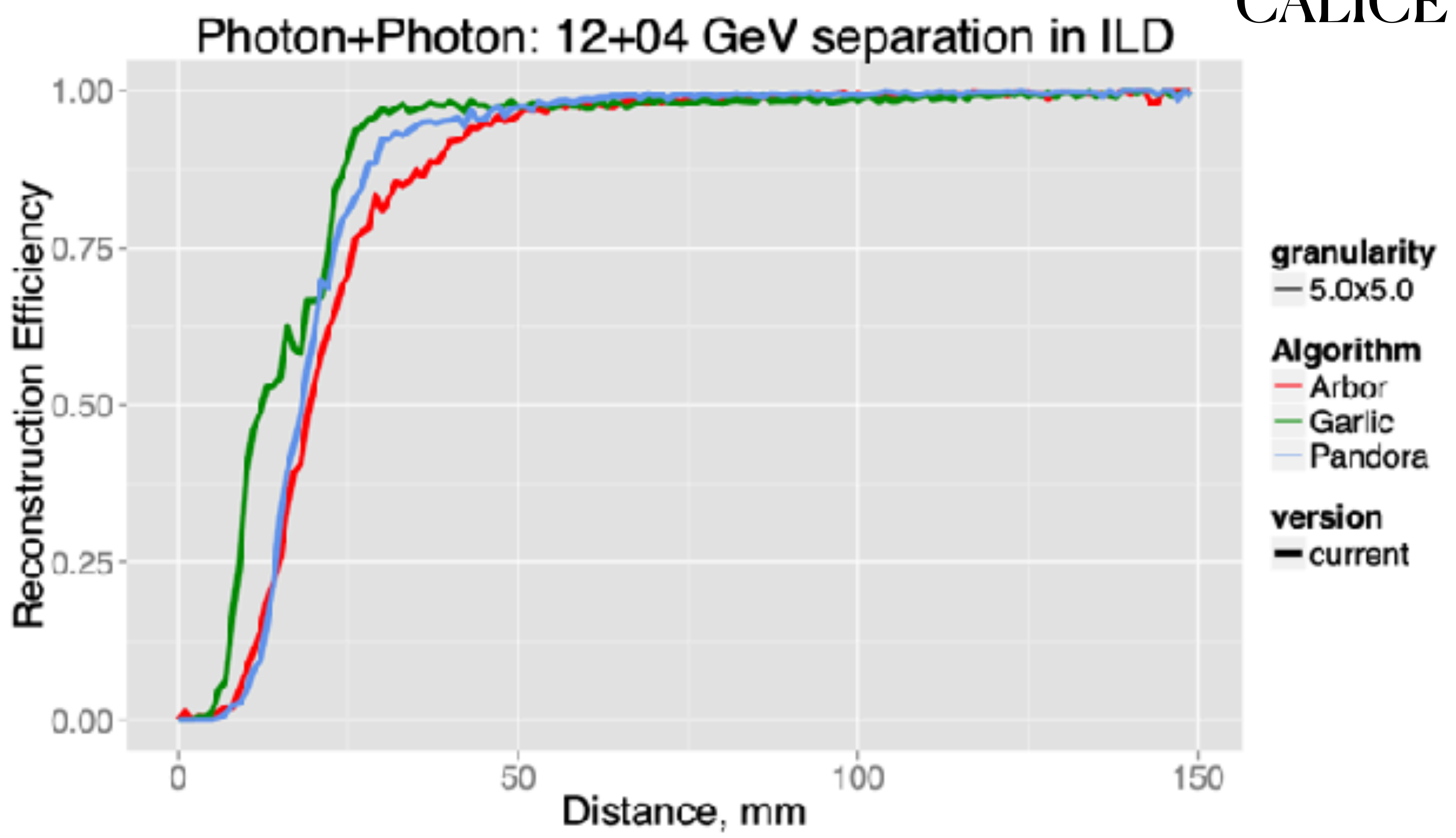


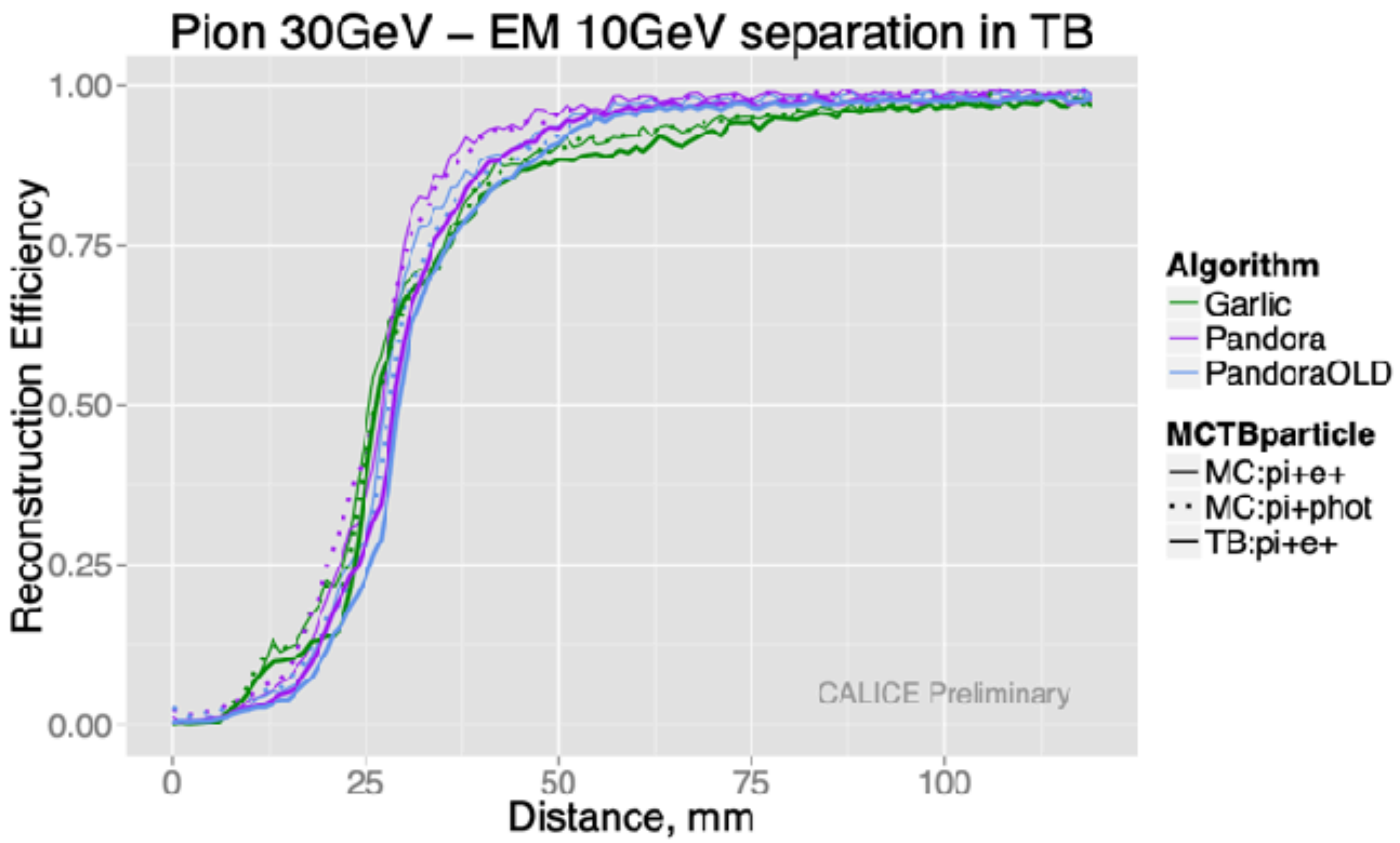


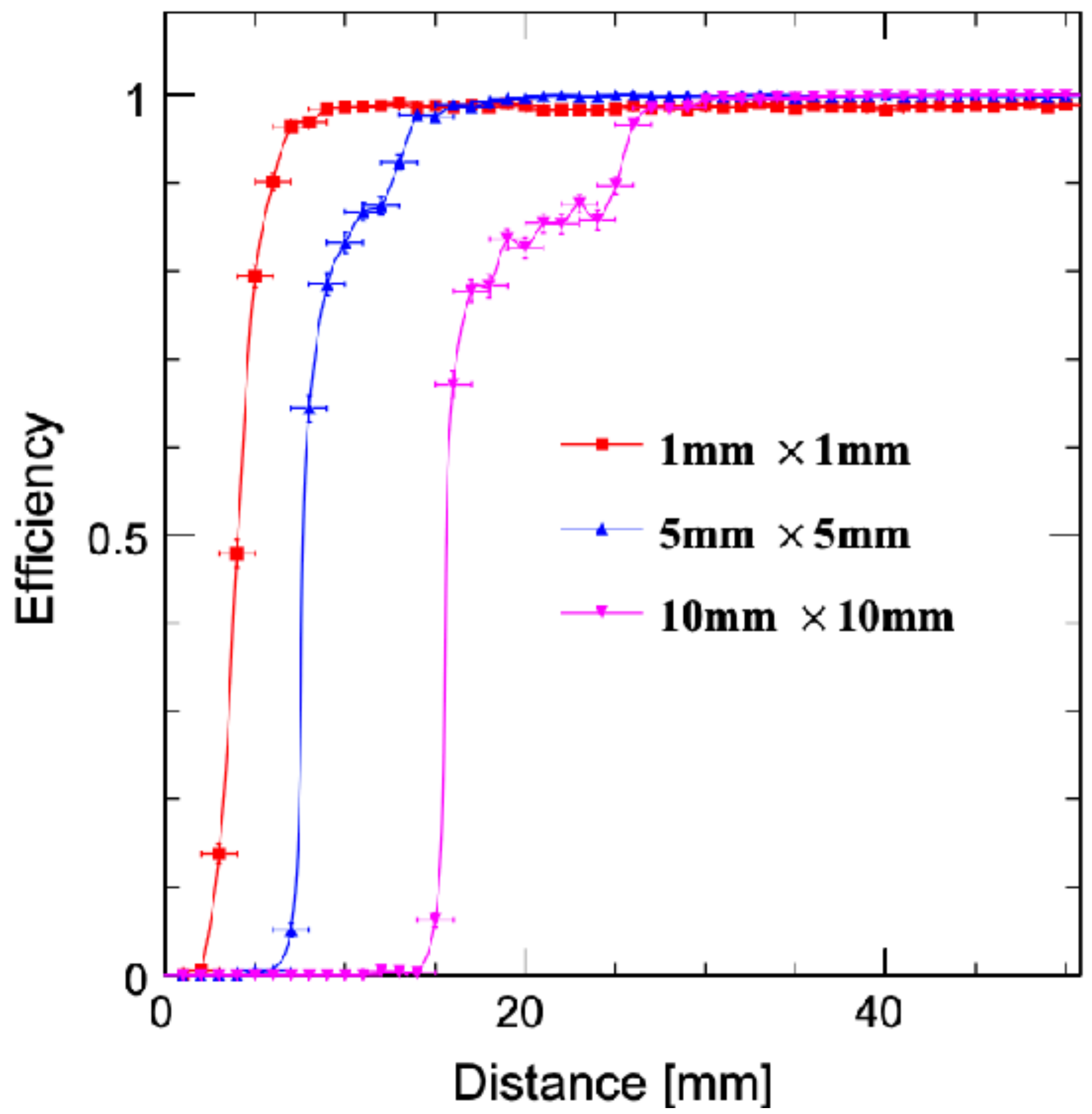
- The reconstruction of tree topology may help the power of separation.
- The high granularity calorimeter support this idea at the hardware side.
- The separation depends on cluster energy

Cluster separation

CALICE Note CAN-057



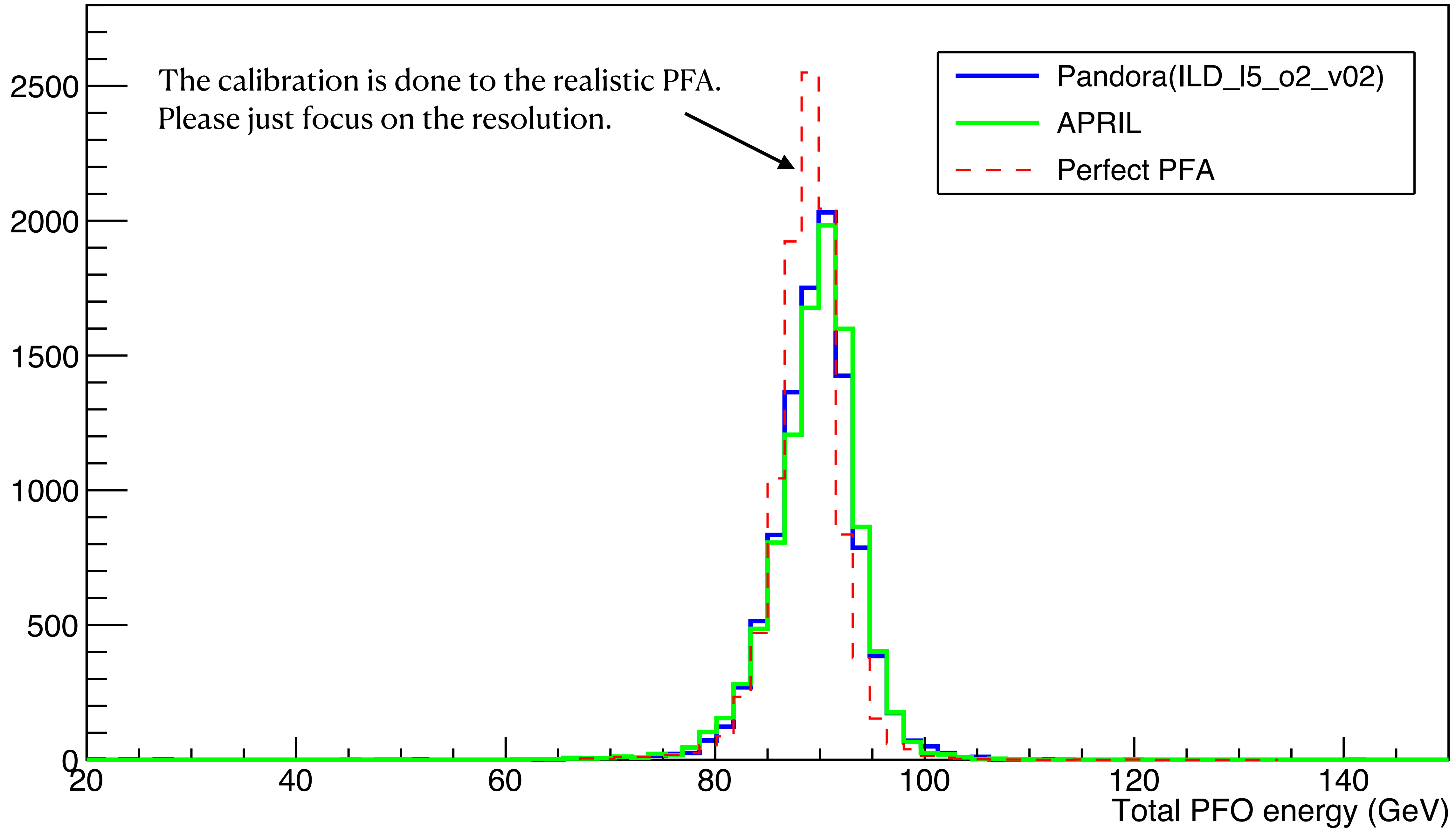


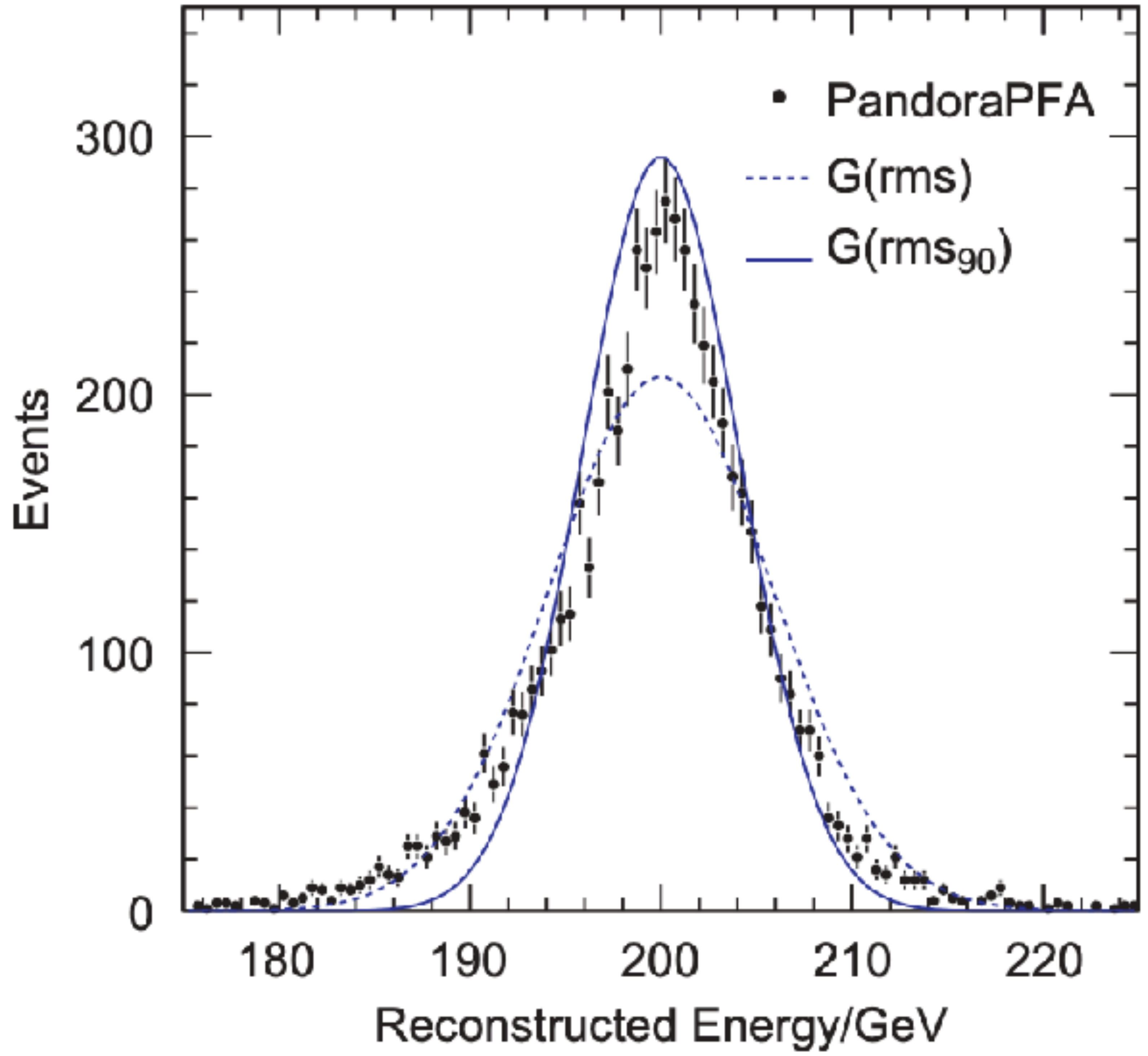


Eur. Phys. J. C (2018) 78:426

- These algorithms have consistent performance on cluster separation;
- And Arbor has further updated its EM shower separation power.

Jet energy resolution - 1





The event samples are from the processes $e^+e^- \to q\bar{q}$,
where $q = u, d, s$ $\quad |\cos\theta_{q\bar{q}}| < 0.7$

RMS90

Jet energy resolution $\dfrac{\mathrm{RMS}_{90}(\mathrm{E}_j)}{\mathrm{mean}_{90}(\mathrm{E}_j)} = \sqrt{2}\cdot\dfrac{\mathrm{RMS}_{90}(\mathrm{E}_{jj})}{\mathrm{mean}_{90}(\mathrm{E}_{jj})}$

ARPIL follows the procedures of JER calculation in Ref. NIMA 611 (2009) 25–40

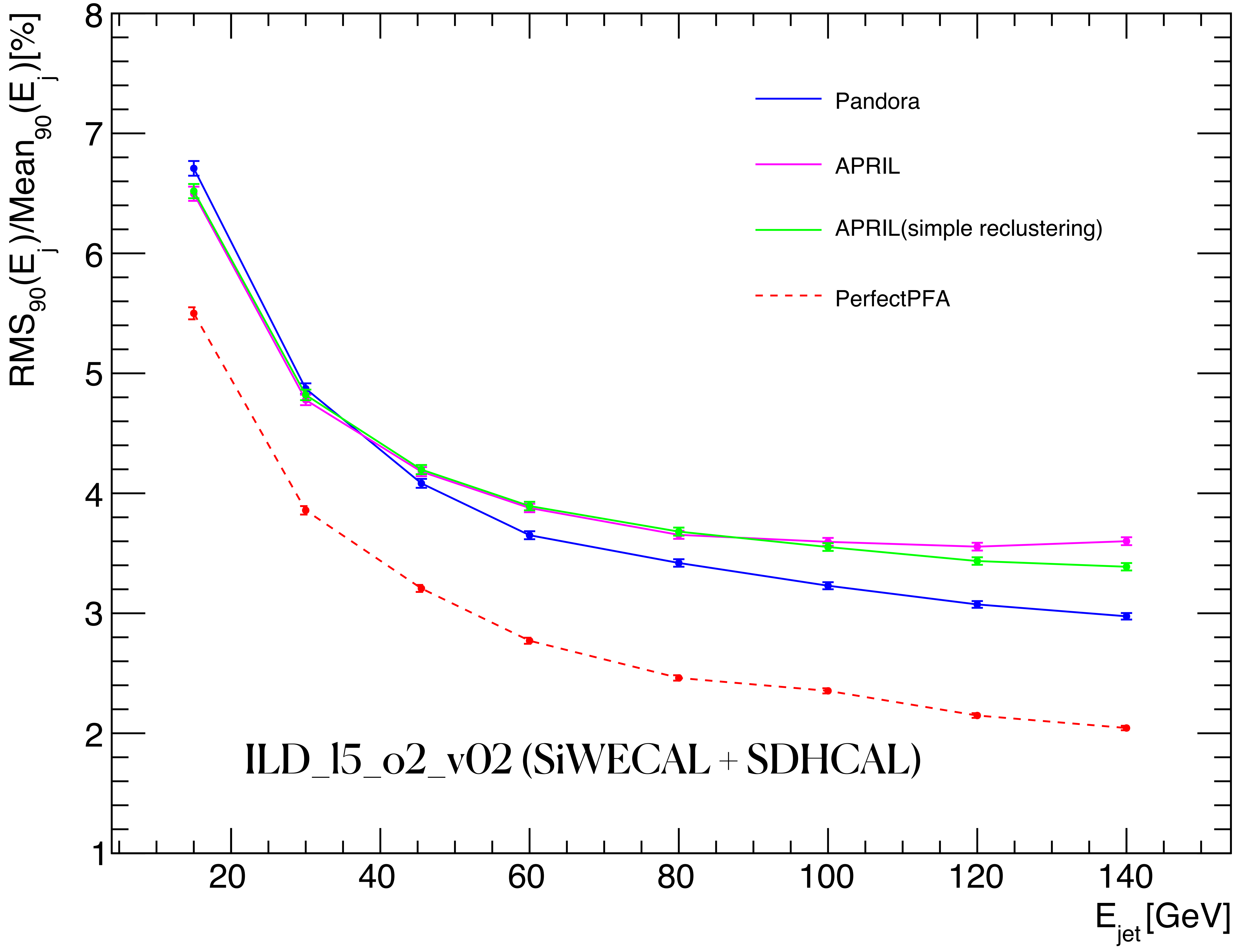
Errors

- The error of each reconstruction stage is estimated according to MC truth (APRIL, 91.2 GeV)

Stage	Error
Clustering	0.05%
Nearby hits merging	0.15%
Cluster merging	0.30%
Track-cluster association and PFO creation	0.30%

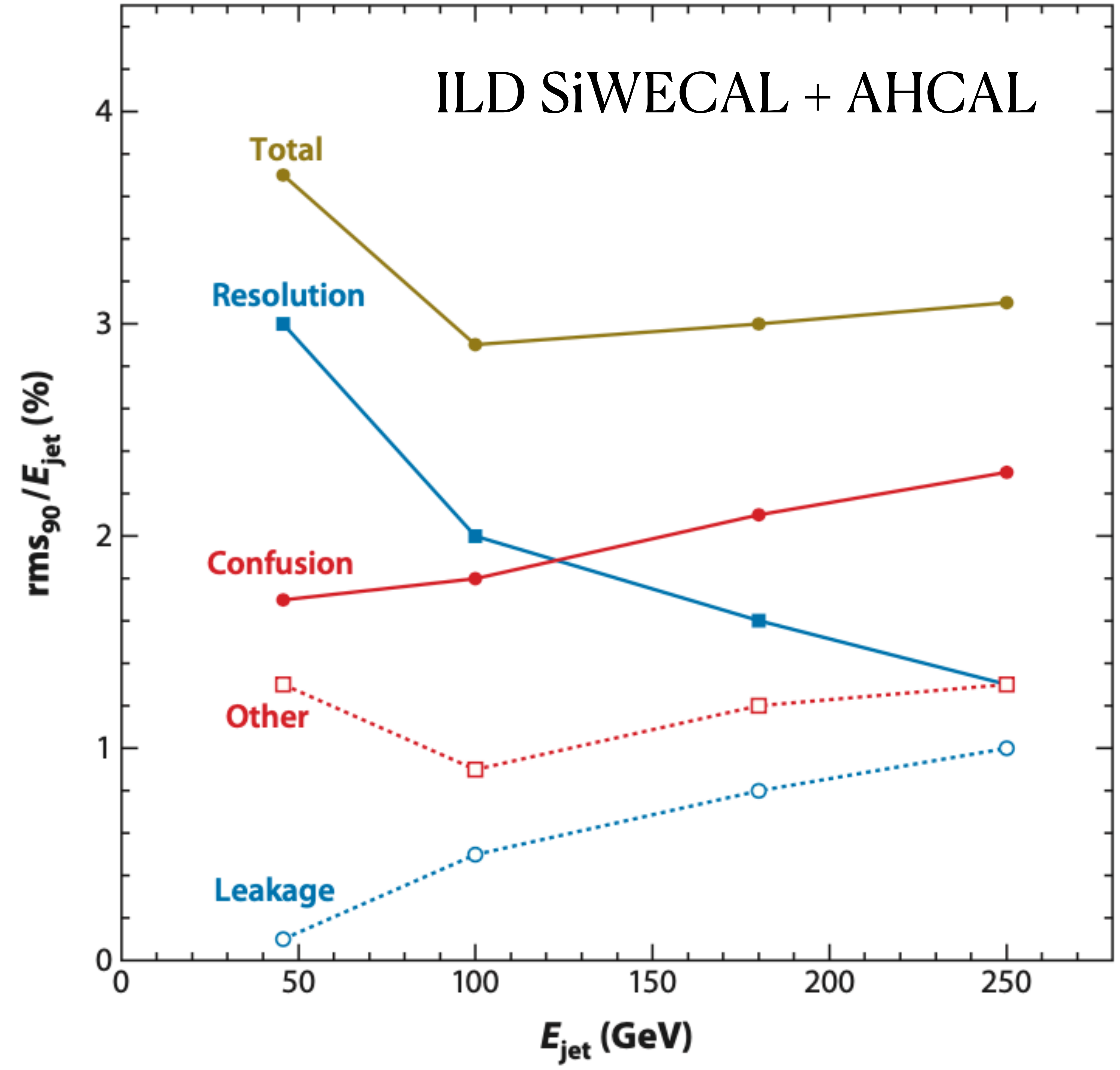
- The initial clustering has relatively small error, as designed
- Further optimisation to the algorithm and parameters

Jet energy resolution - 2

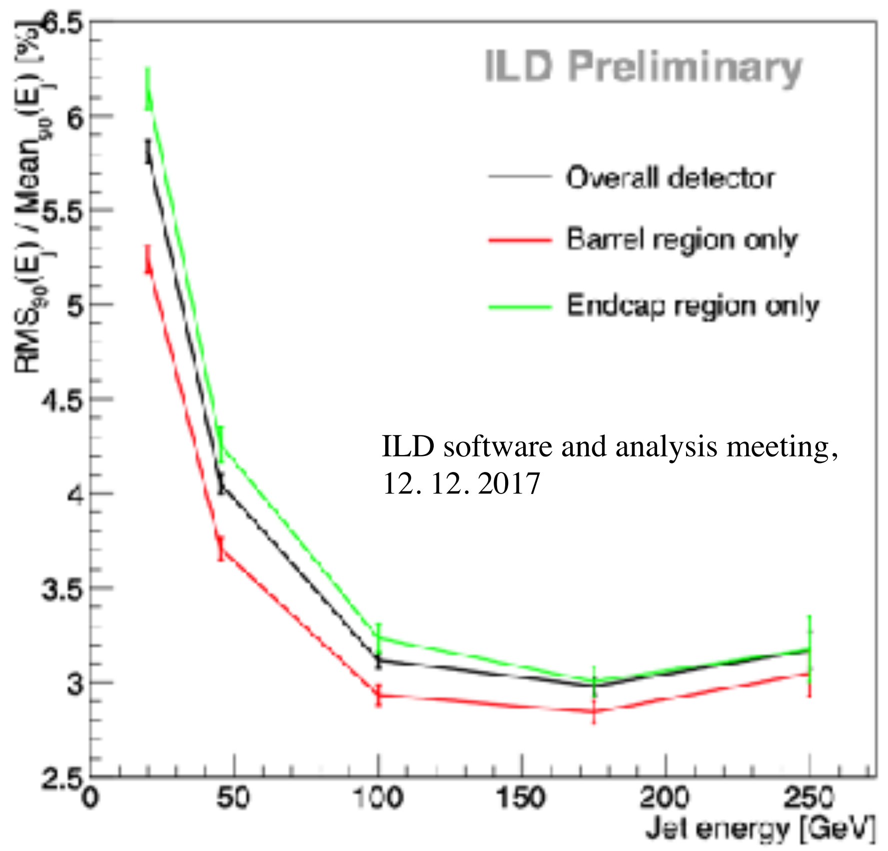


- Without the reclustering procedure, the jet energy resolution by APRIL is not bad (3 - 4% for jet energy 40 - 140 GeV).
- A new cluster separation algorithm for APRIL is needed.

Jet energy resolution - 3



- The jet energy resolution is 3 - 4%
- 'Other' term : track-cluster matching, PID
- The jet energy suffers from the the 'confusion' term, which is due to the shower overlap
  - The 'confusion' term dominates for jet energy greater than 100 GeV
  - Based on the current PFAs, the high granularity of calorimeter is limited at high jet energy
- Keep in mind that
  - No detector noise is considered
  - Using jets on the barrel
  - Detector optimisation



- So, new ideas to improve the PFA performance ?

Can we get better cluster separation?

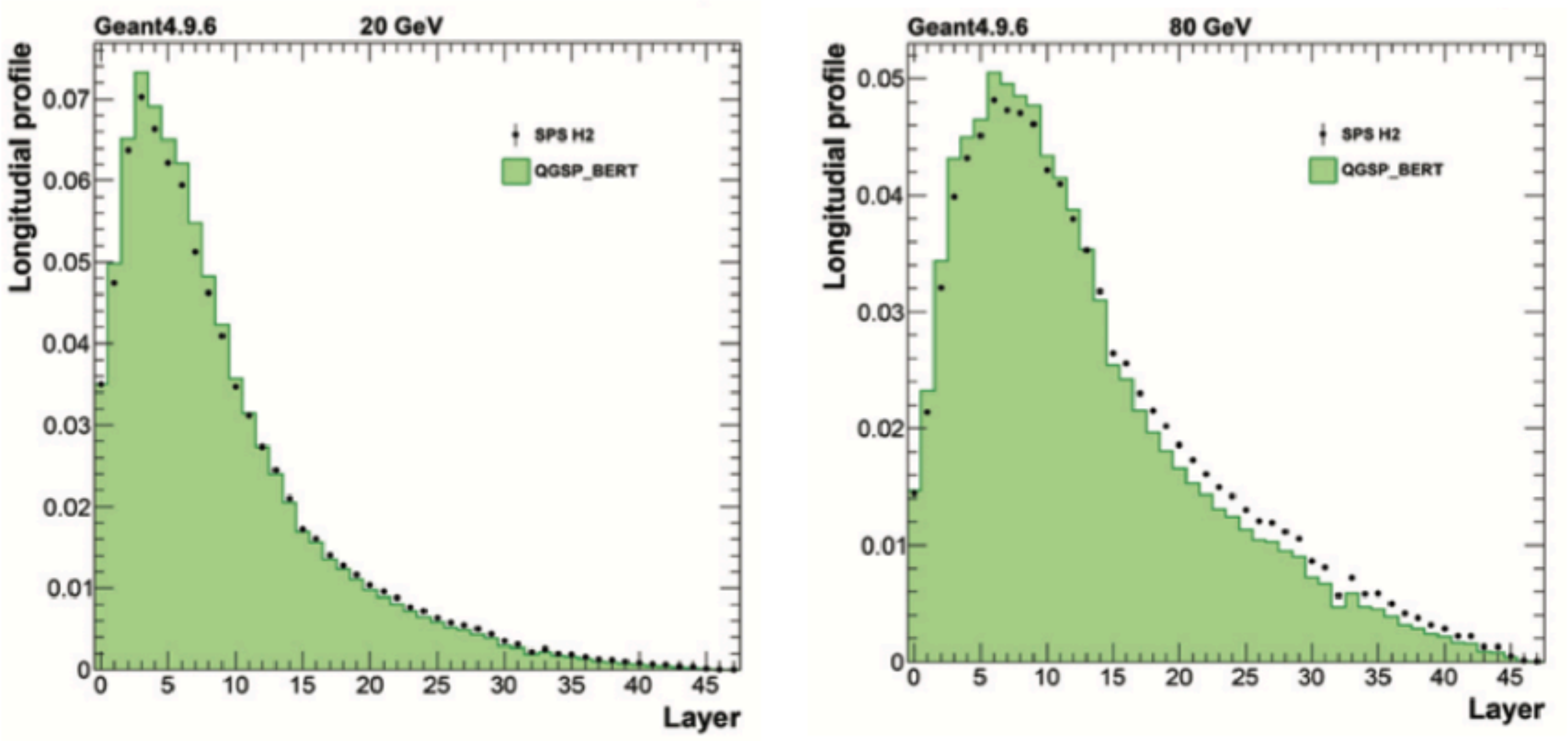
- Although the calorimeter can be 5-d, the clustering methods in the PFAs are still 3-d (hit position)
- Only local information is taken into account in the clustering
- The clustering parameters are tuned manually

The typical clustering result we now get from two overlapping clusters

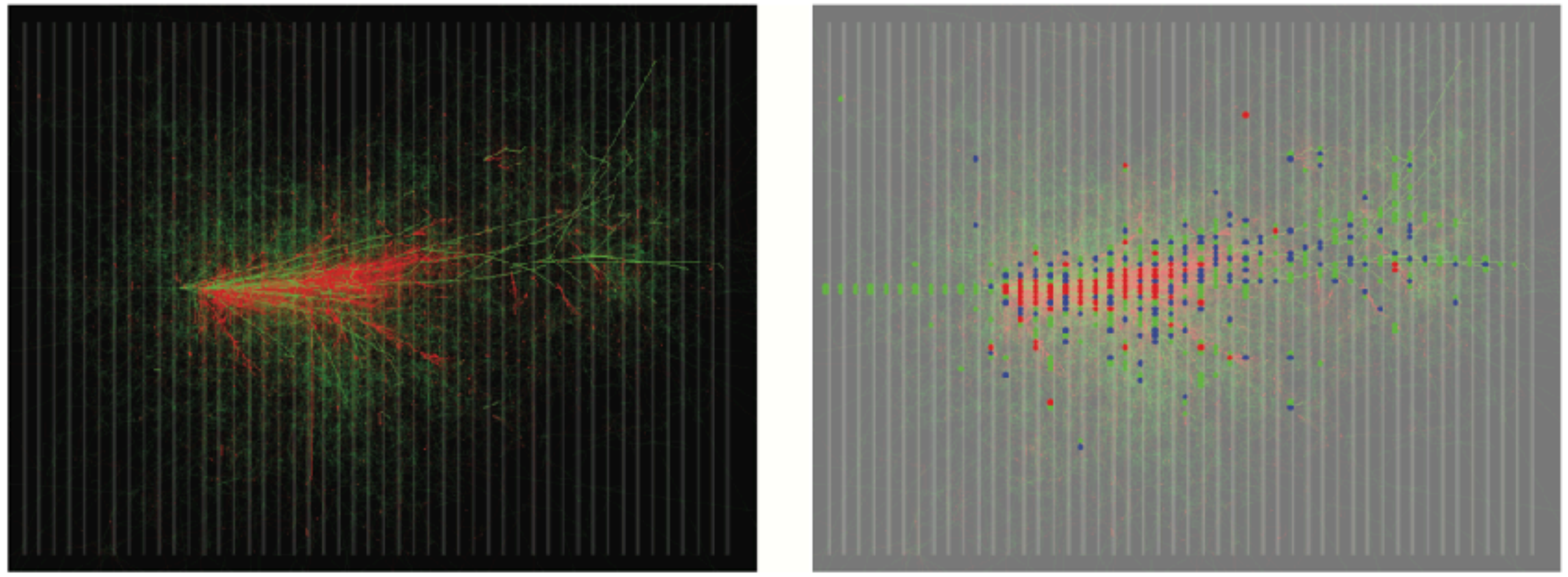
This seems more reasonable.

How can we make it true?

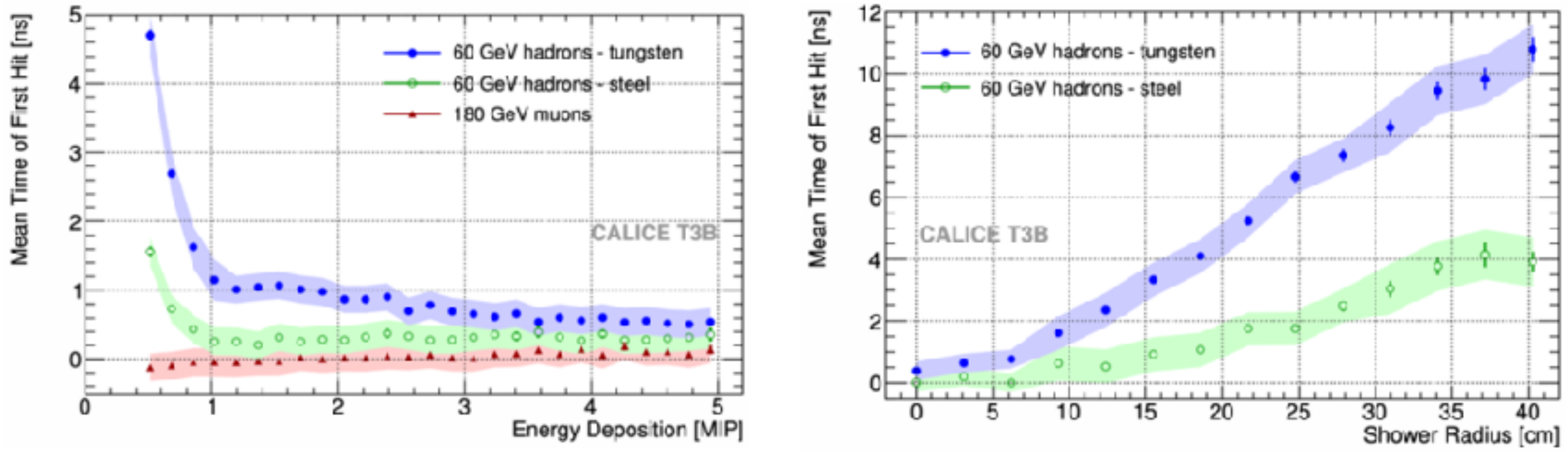
Proposal



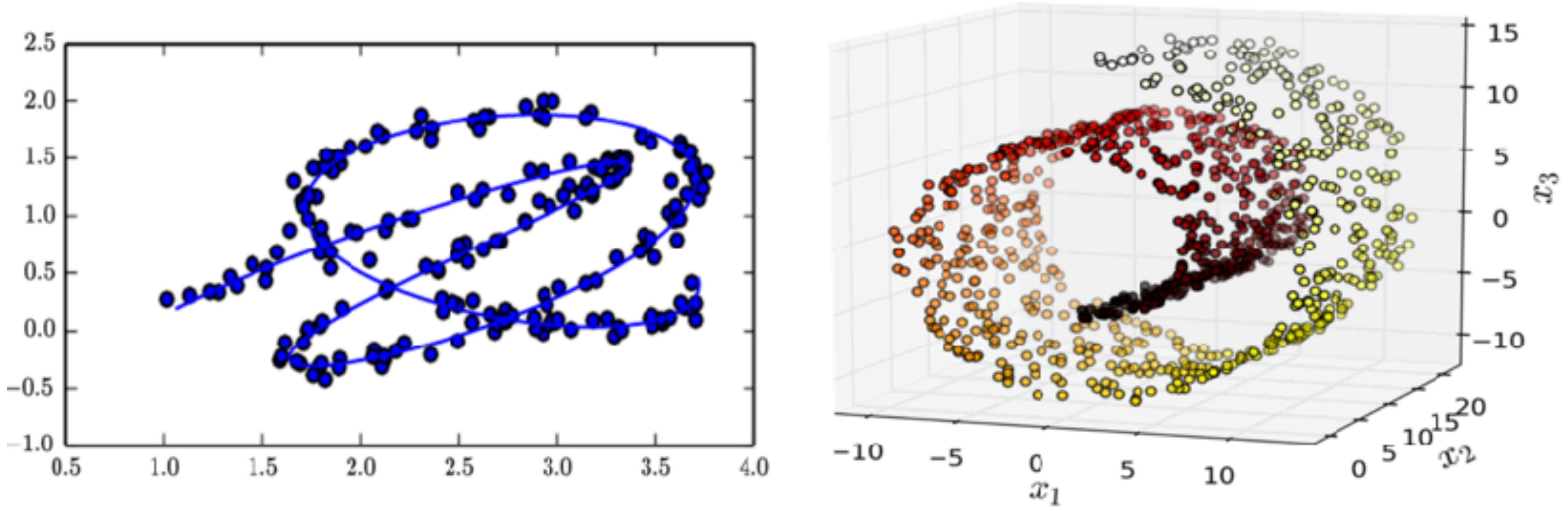
Energy density



A kind of hidden pattern may exist in some space ?



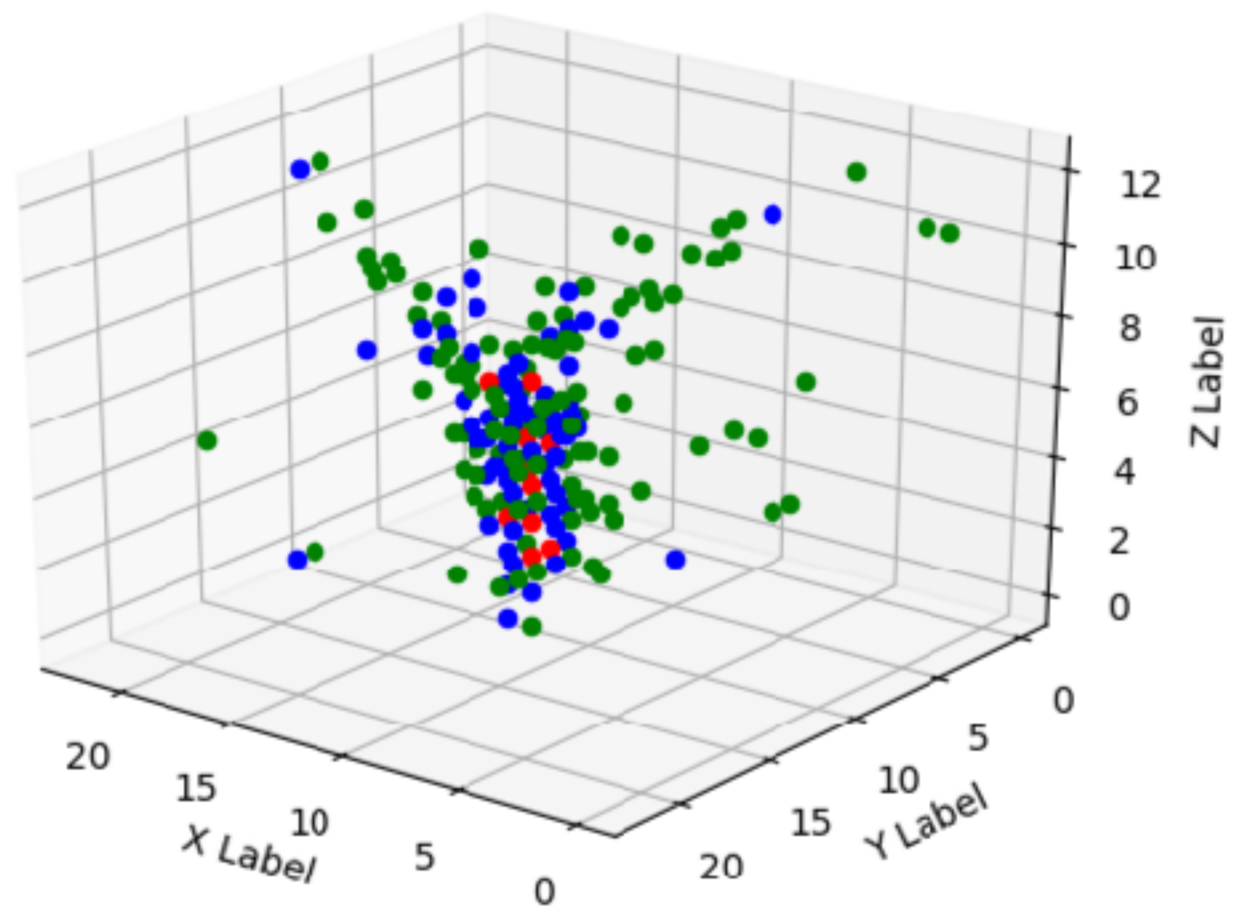
Hit time



Energy reconstruction with Convnet

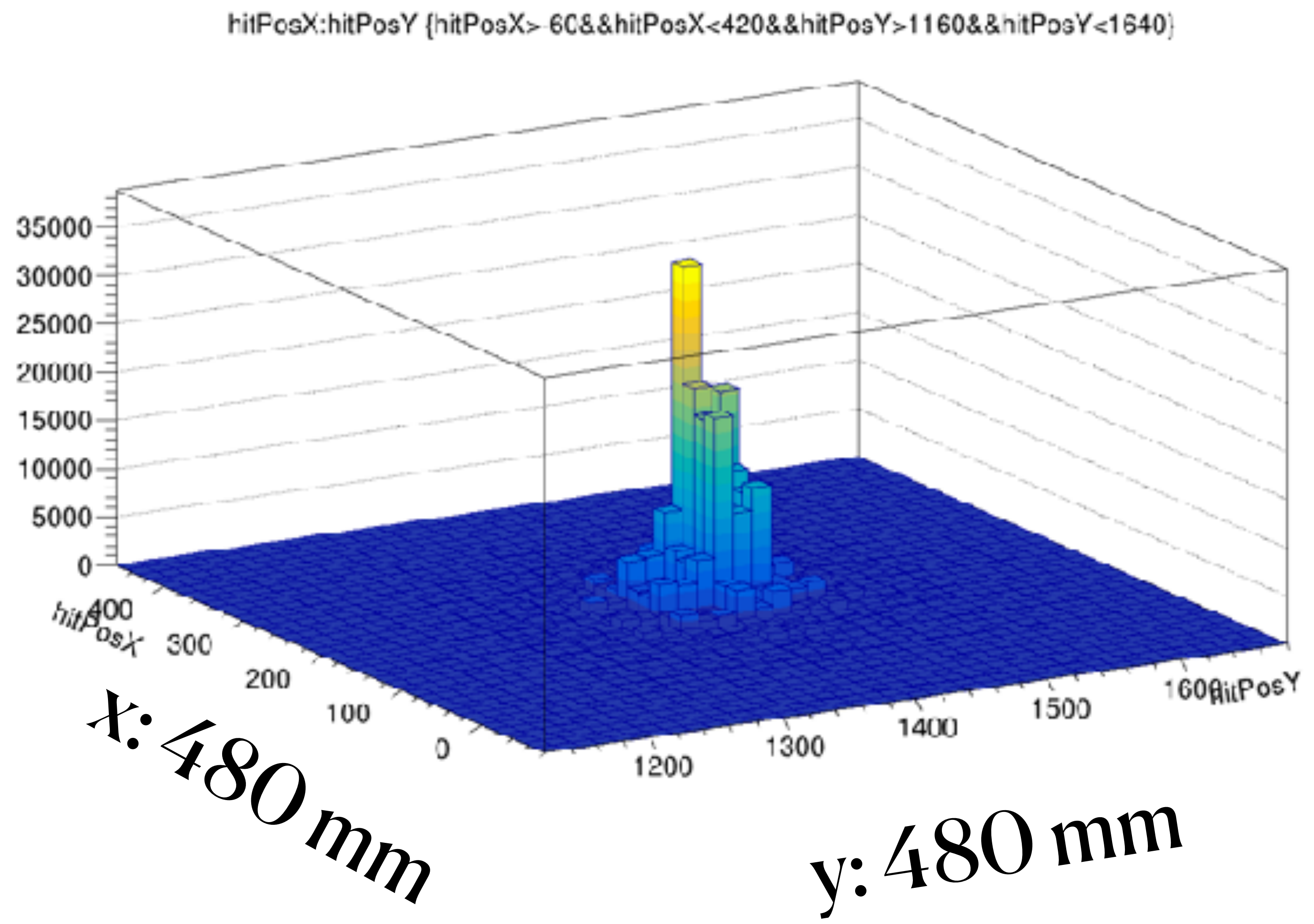
Simulation:

- training samples: K^0_L, 2 - 100 GeV
- test samples: K^0_L, 10, 20, … , 80 GeV
- ILD_l5_o2_v02 (SDHCAL), endcap
- ILCSoft v01-19-05



29 GeV

Hit map of 1k events



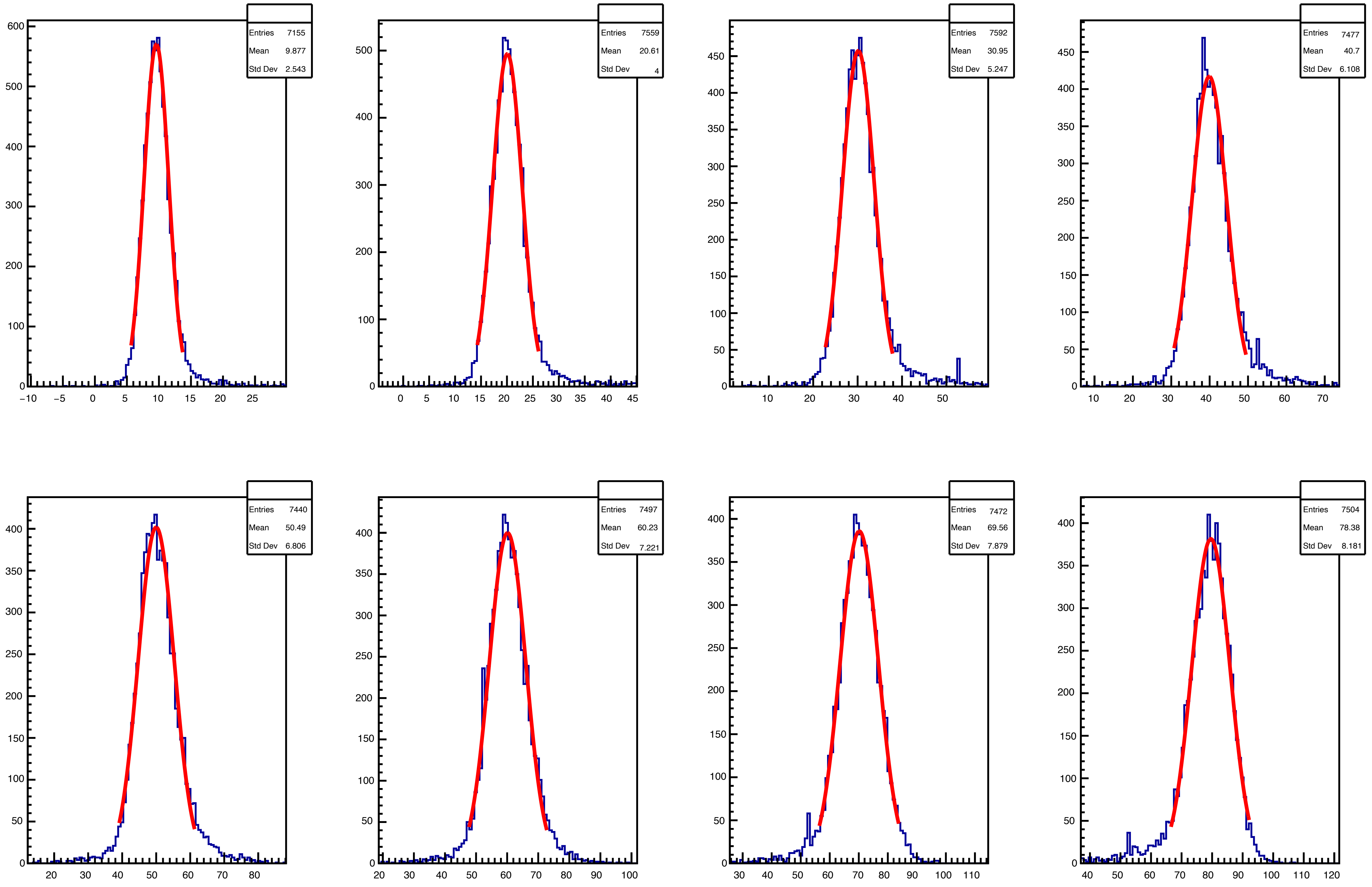
96% hits in this region

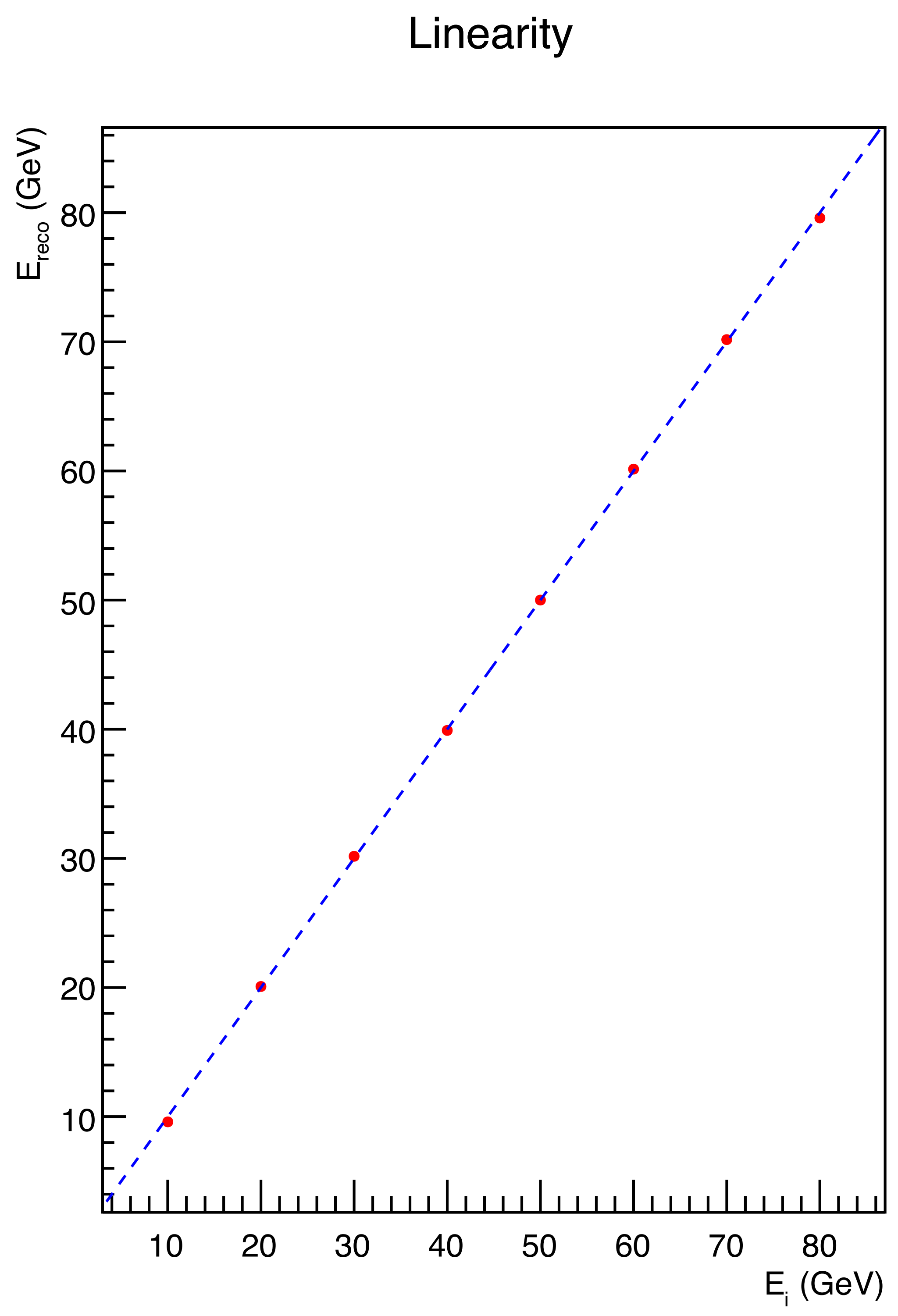
- Keras 2.2.0 (with tensorflow-1.8.0)
- GPU: GeForce GTX 1080 Ti
- Convnet model

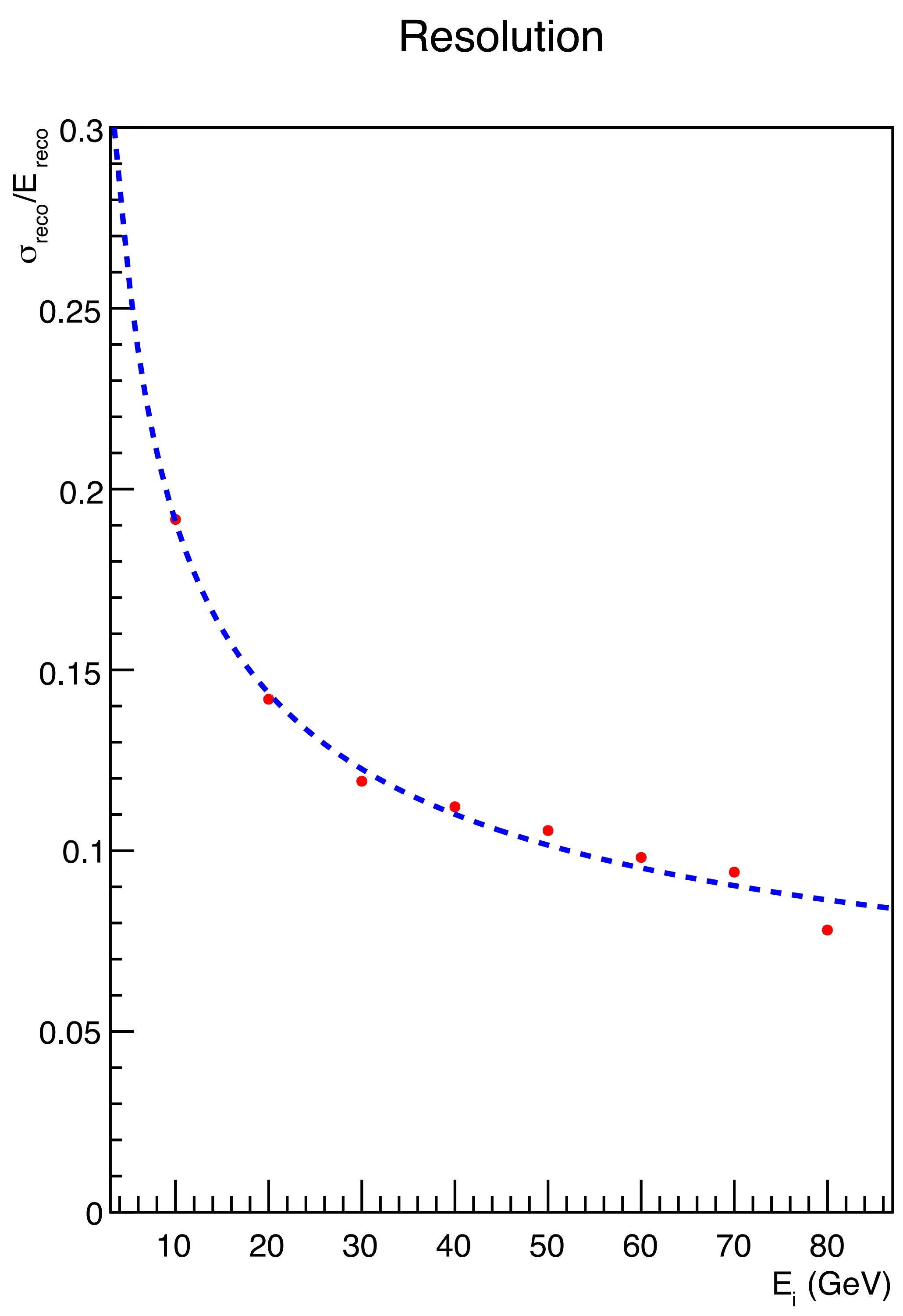
Layer (type)	Output Shape	Param #
conv3d_1 (Conv3D)	(None, 23, 23, 23, 16)	144
average_pooling3d_1 (Average	(None, 11, 11, 11, 16)	0
conv3d_2 (Conv3D)	(None, 10, 10, 10, 32)	4128
average_pooling3d_2 (Average	(None, 5, 5, 5, 32)	0
conv3d_3 (Conv3D)	(None, 4, 4, 4, 32)	8224
flatten_1 (Flatten)	(None, 2048)	0
dense_1 (Dense)	(None, 32)	65568
dense_2 (Dense)	(None, 1)	33

Total params: 78,097
Trainable params: 78,097
Non-trainable params: 0

Result







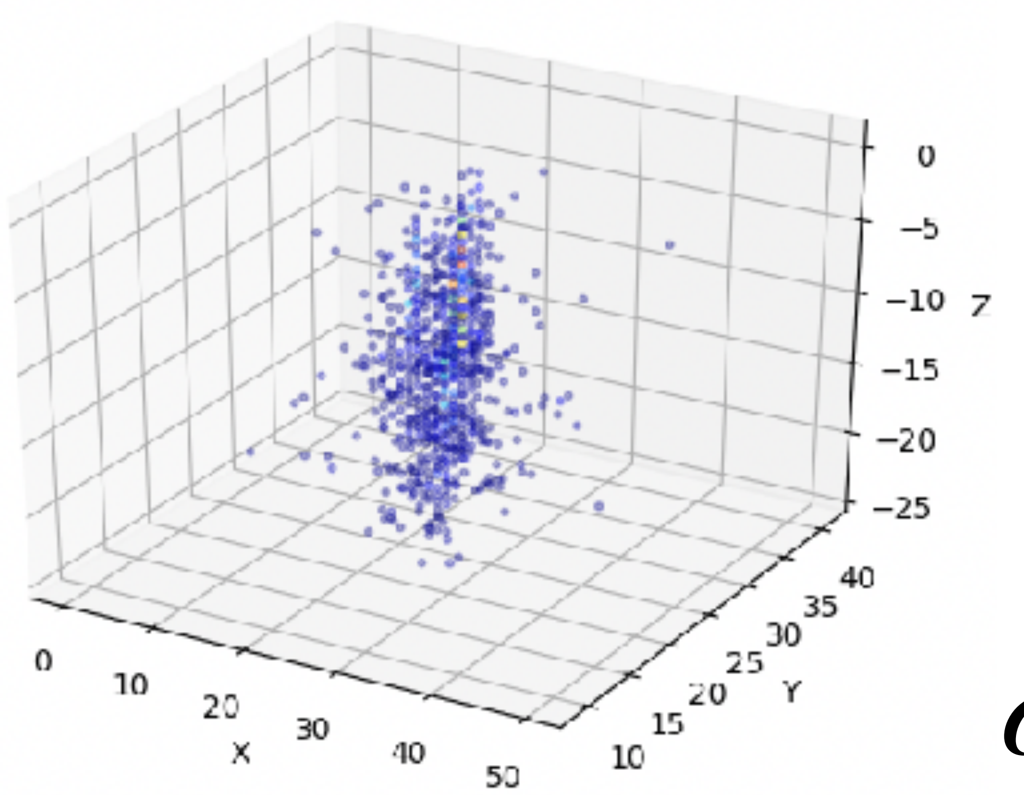
The linearity and resolution of energy are the same with that from software compensation in SDHCAL

New algorithms for LCD at CLIC

arXiv:1912.06794v3

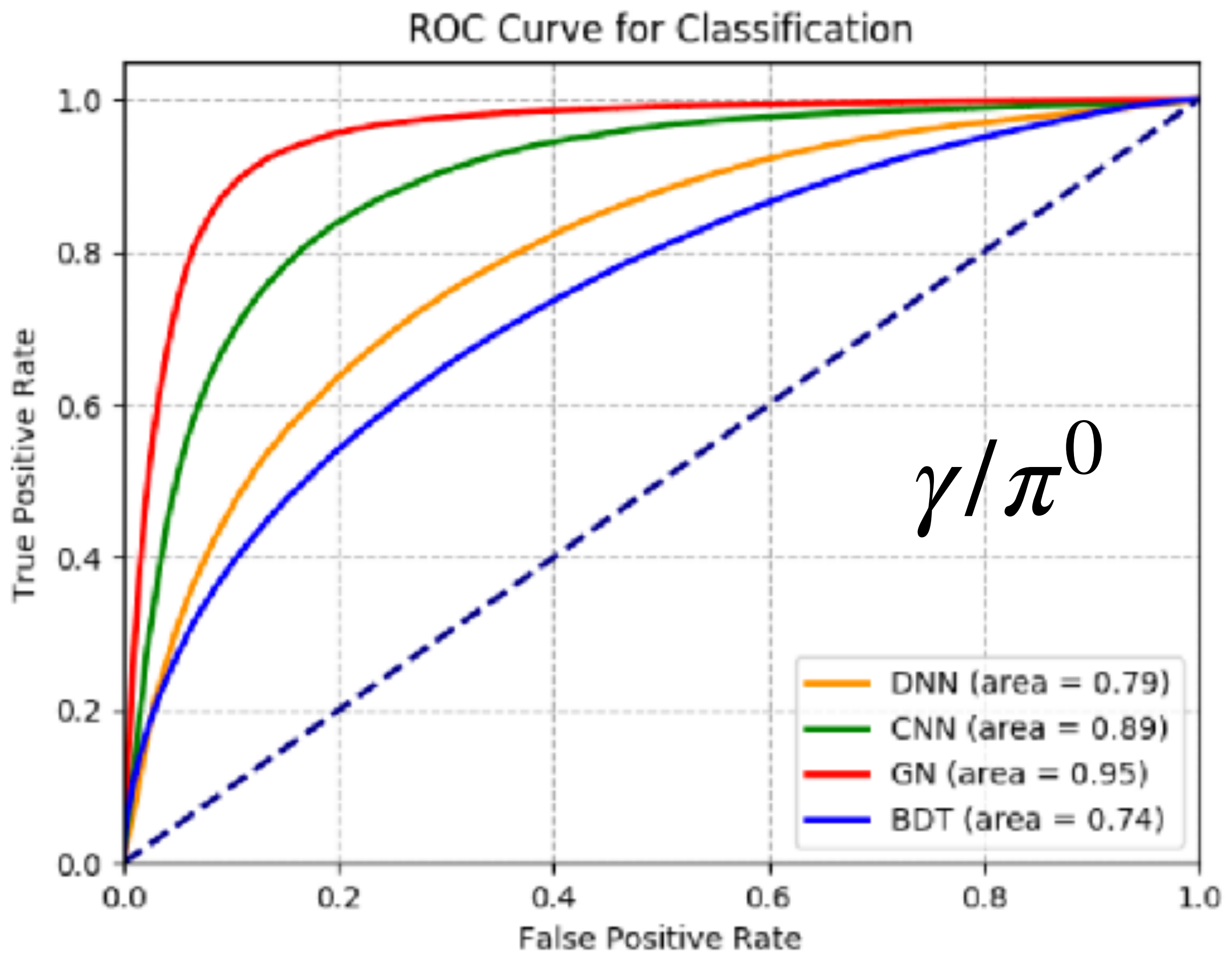
- Detector model: Linear Collider Detector (LCD) at CLIC
- Several kinds of networks are tested:
  - Dense Neural Network (DNN)
  - 3D convolutional network (CNN)
  - GoogLeNet (GN)

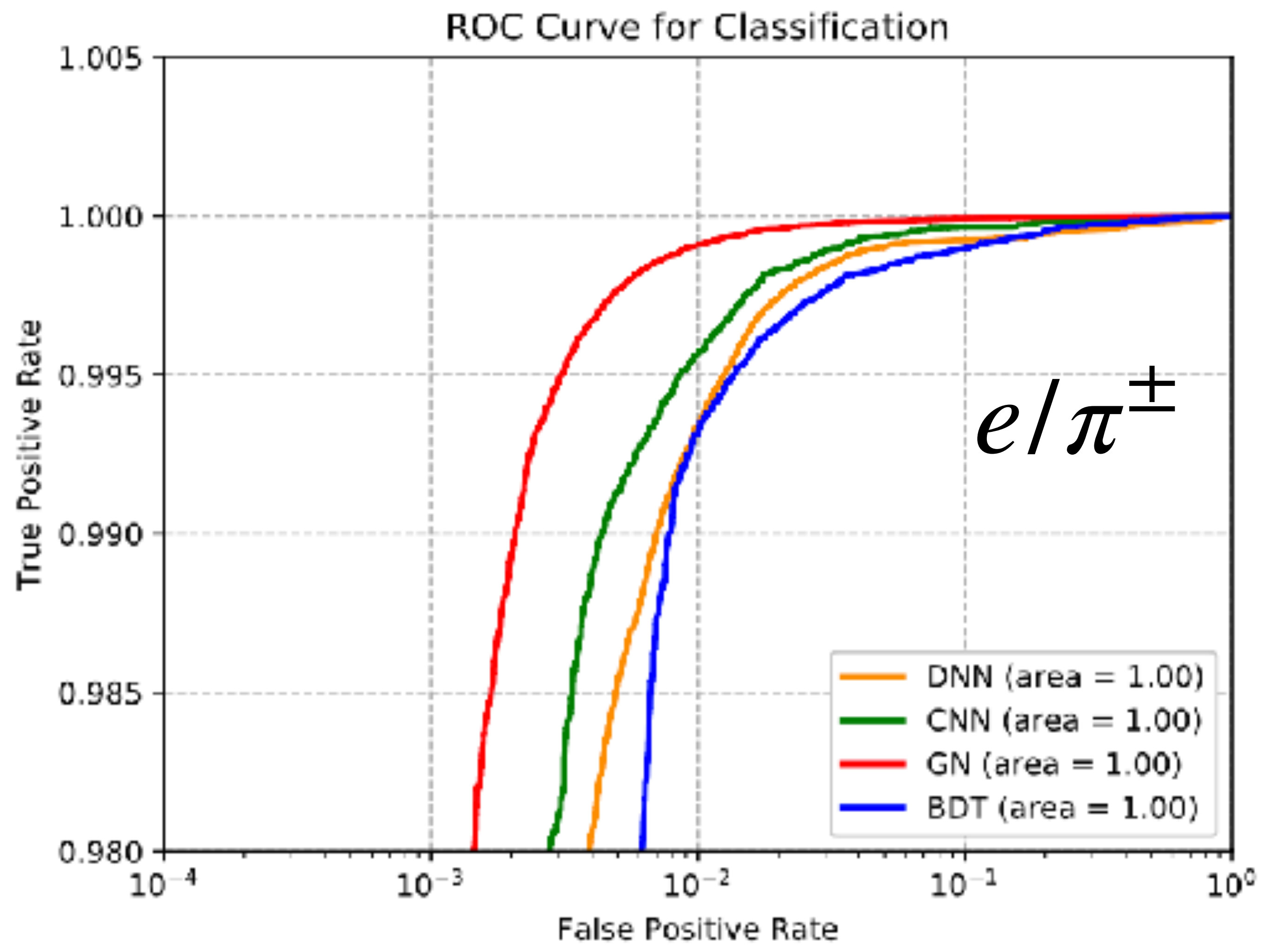
- ECAL: 25 layers, cell: 5.1×5.1 mm^2
- HCAL: 60 layers , cell: 3×3 cm^2



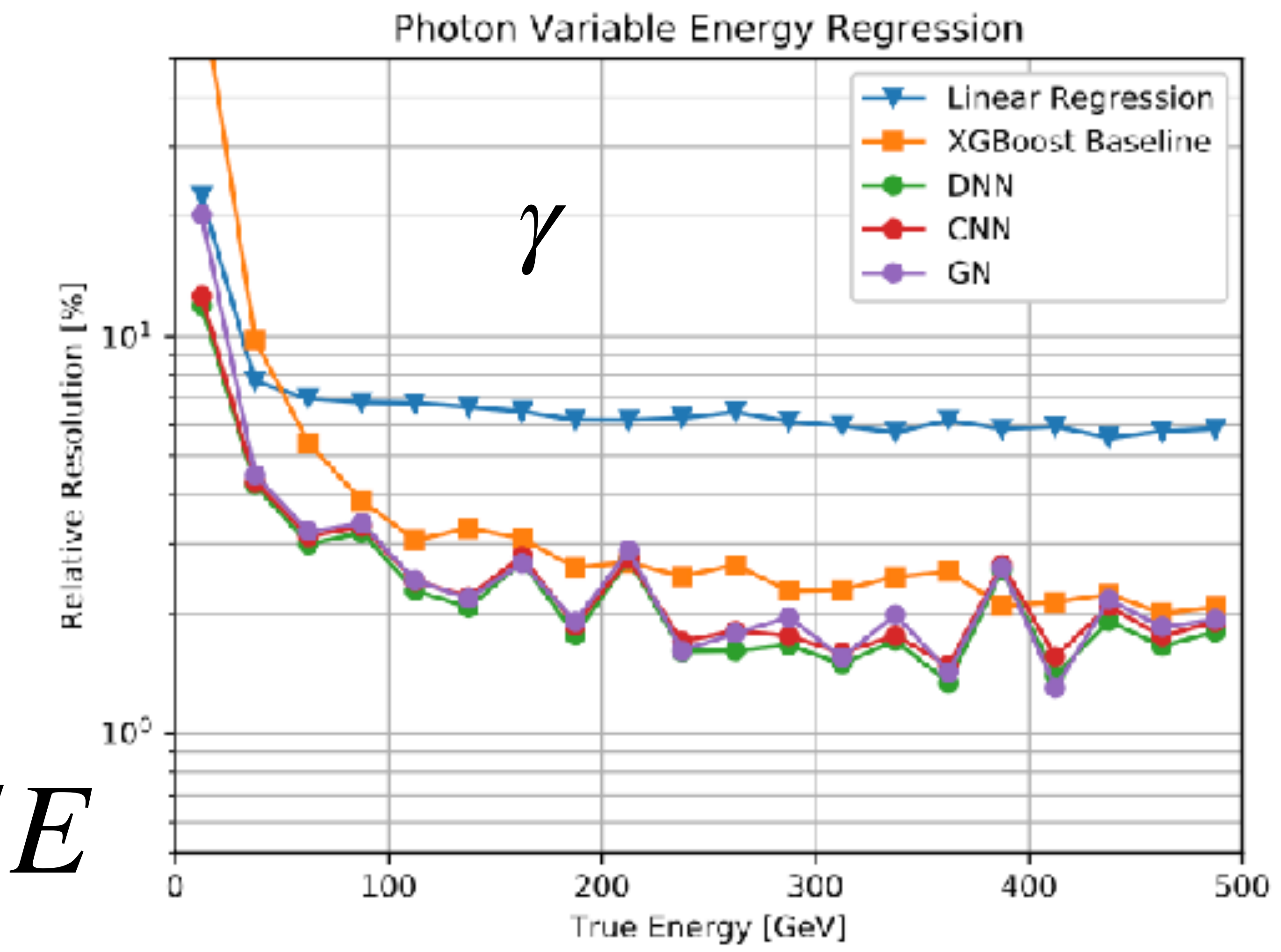
3D image of an EM shower

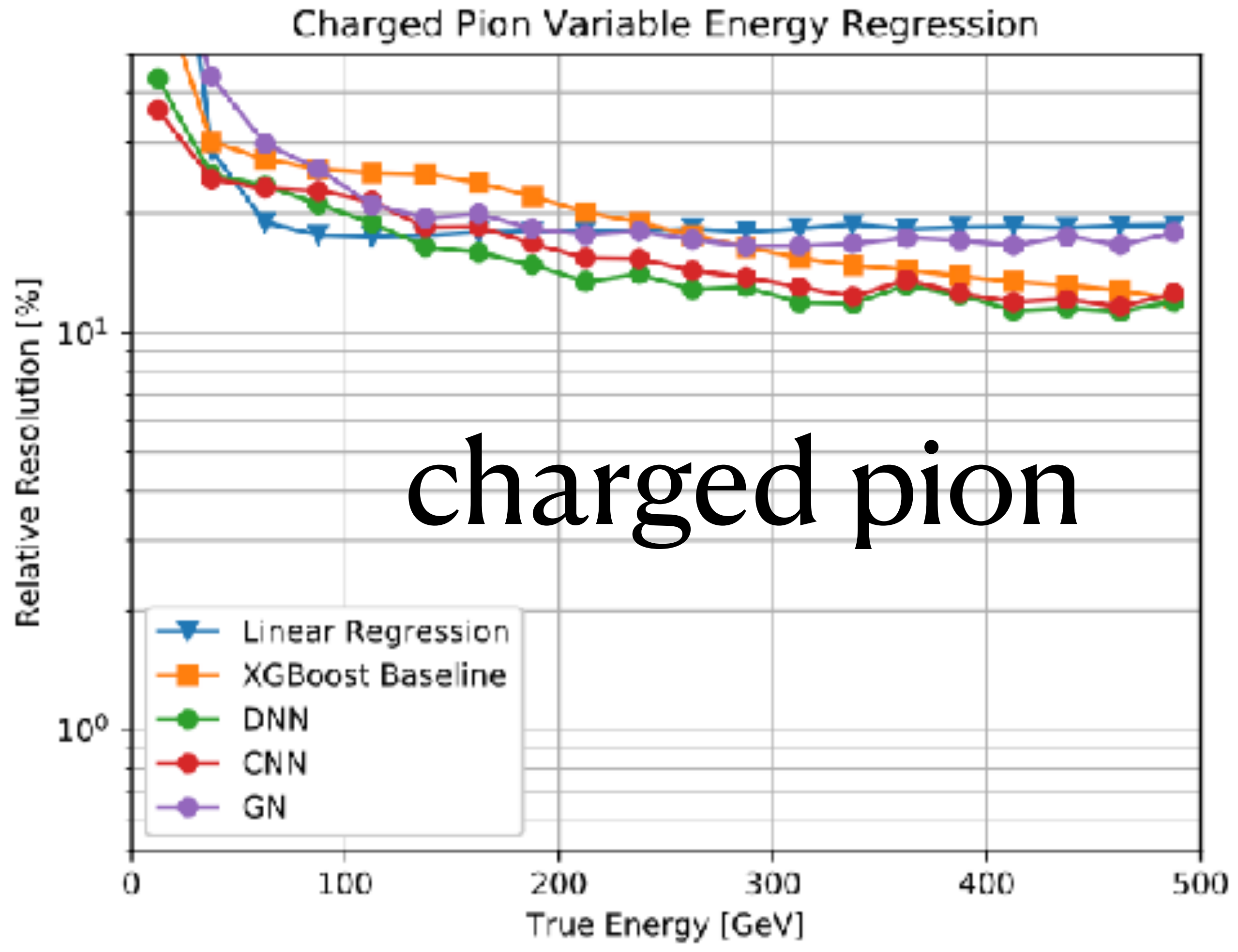
PID





σ_E/E



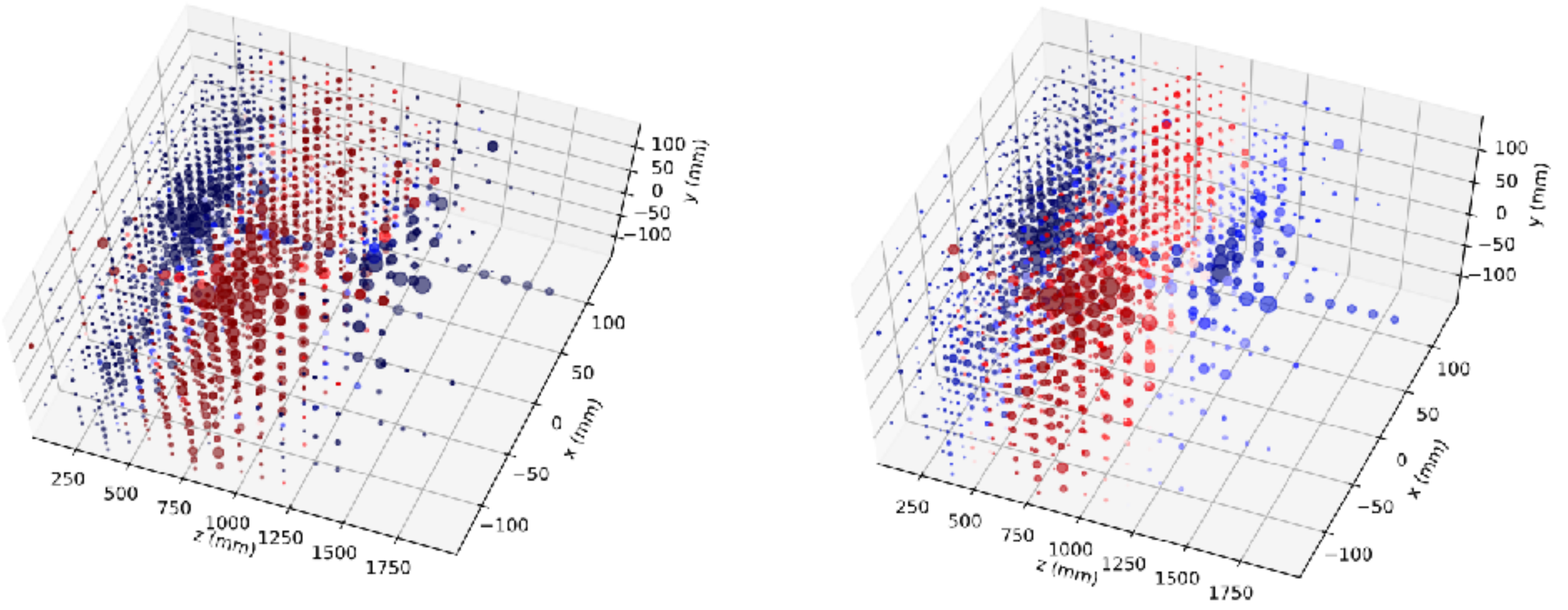


Neural network for shower separation

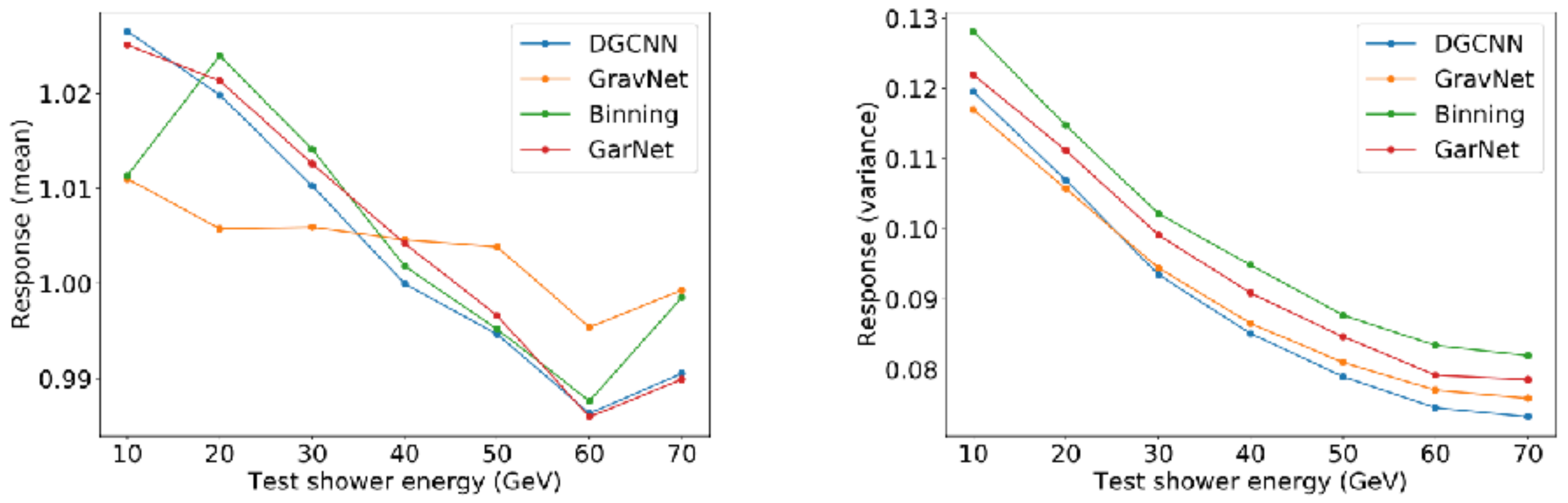
arXiv:1902.07987v2

- A so-called Gravitational Network (GravNets), with distance-weighted graph network architecture, is introduced
- Gaussian potential $V(d_{jk}) = \exp(-d_{jk}^2)$ for the connection

$$R_k = \frac{\sum_i E_i p_{ik}}{\sum_i E_i t_{ik}}$$



(a) Truth Two 50 GeV charged pions (b) Reconstructed



(c) Mean (d) Variance

Summary

- Particle Flow Calorimetry = high granularity calorimeter (together with high precision tracker) + Particle Flow Algorithm.
- Particle Flow Calorimetry is important for the future electron-positron collider experiments.
- The study based on simulation shows that jet energy resolution is 3 ~ 4%. It needs the validation by experiment.
- The new ideas for improving the PFA.