

Particle reconstruction with Neural Networks for the IDEA Dual Readout Calorimeter

Federico Montereali,

Michela Biglietti, Biagio Di Micco, Roberto Di Nardo

Rome III INFN & University

o.b.o. the IDEA and HiDRa Collaborations

This work was supported by the
AIDAINNOVA HORIZON 2020 Project,

Grant Agreement ID:
101004761, WP 12, Task 12.3



中国科学技术大学
University of Science and Technology of China



INSTR 2026
Instrumentation for Colliding Beam Physics

October 12–16, 2026 Hefei | China
University of Science and Technology of China

FCC-ee and the IDEA Detector Concept

The **FCC programme** plans for a 90.7 km tunnel with **4 IPs** and **8 surface sites** hosting 2 stages:

- **Stage 1** → **FCC-ee: e^+e^- collider**, $\sqrt{s} = 88 \rightarrow 365$ GeV (*4 runs at Z pole, WW th., ZH, top-pair th.*)
- **Stage 2** → **FCC-hh (@ 100 TeV)**: exploration of the pp energy frontier

FCC-ee: beyond a *Higgs factory* !

Continuous beams, **extremely clean environment**, peak lumi $140 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ per IP at the Z pole

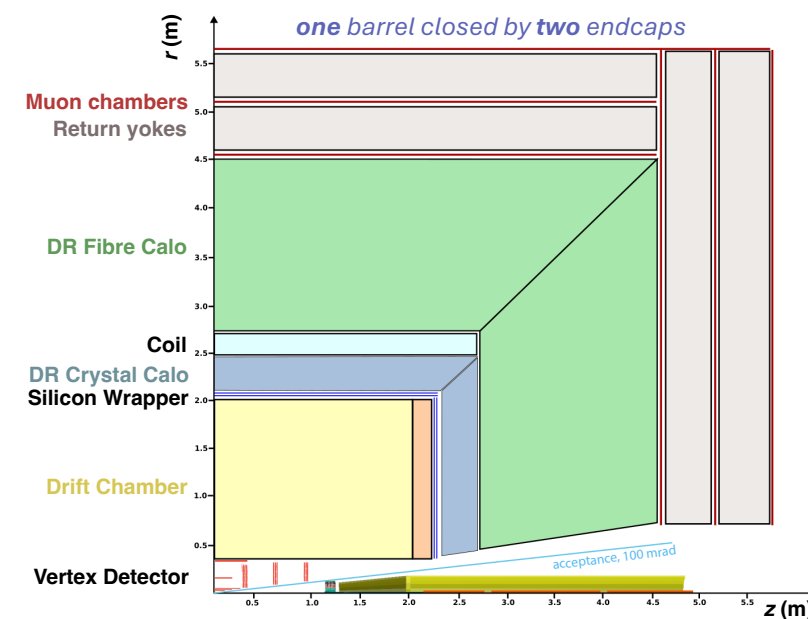
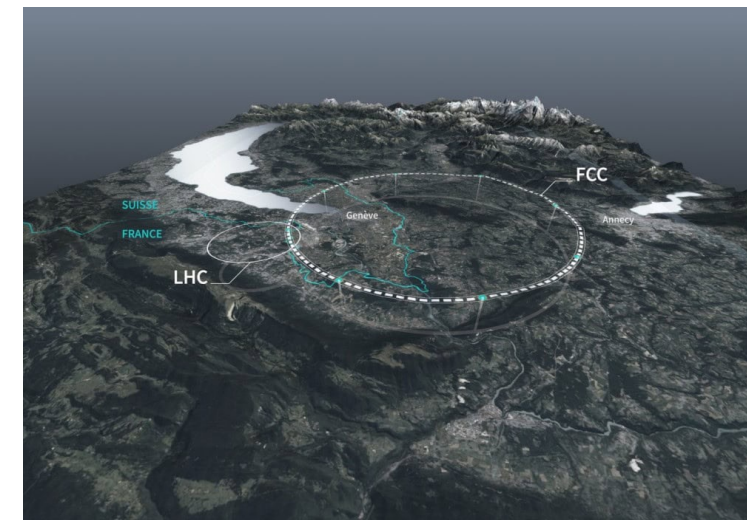
Precision exploration of the **electroweak sector**, **Higgs couplings**, **flavour physics**, and **BSM physics**

— The Innovative Detector E+e- Accelerator (IDEA) Concept —

➤ **First general-purpose detector** conceived specifically for **FCC-ee**

○ **Detector components** (moving outward from the IP):

- High-granularity MAPS **silicon pixel vertex detector**
- **Ultra-light Drift Chamber** + **silicon strip wrapper** tracking, PID, $O(100 \text{ ps})$ timing
- Fine-segmented **Crystal Dual-Readout EM calorimeter** $\sigma/E \approx 3\%/\sqrt{E}$
- **Superconducting solenoid magnet** up to 3 T
- **Fiber Dual-Readout Hadronic Calorimeter** $\sigma_{HAD}/E \approx 30\%/\sqrt{E}$
- **Three μ –RWELL muon stations** $O(100 \mu\text{m})$ spatial resolution



[NIM article on IDEA](#) to be published soon

Talk Overview

The idea is to exploit the **extremely high channel granularity** of the Dual-Readout fibre calorimeter (*HiDRa technology*) to reconstruct **energy and direction** of electrons and hadrons with **neural networks**,
directly from the raw fibre hits

➤ *how much can a network recover that the analytic dual-readout formula cannot?*

IDEA Full simulation → **hit-level input** → **DR reference** → **Deep Neural Network**
→ *Graph Neural Network*

- **Full Simulation:** new *IDEA* Calorimeter geometry, Scintillation and Cherenkov signal production (*slides 4-7*)
- **Input:** from the fired fibres to a flat network input (*slide 8*)
- **Reference:** standard DR reconstruction from input (*slide 9*)
- **DNN:** architecture, loss, and performance on energy and direction (*slides 10-14*)
- **Outlook:** First GNN results (*slide 15*)

IDEA Dual-Readout Calorimeter

Dual-readout calorimetry uses two readout channels, **Scintillation (S)** and **Cherenkov (C)**, to measure the **EM fraction f_{em}** event by event, enabling corrections for fluctuations in hadronic showers and improving the energy resolution

$S \propto$ deposited energy; C predominantly probes the relativistic e^+ of the EM component

INFN HiDRa project

a complete prototype qualified in beam at CERN in 2026

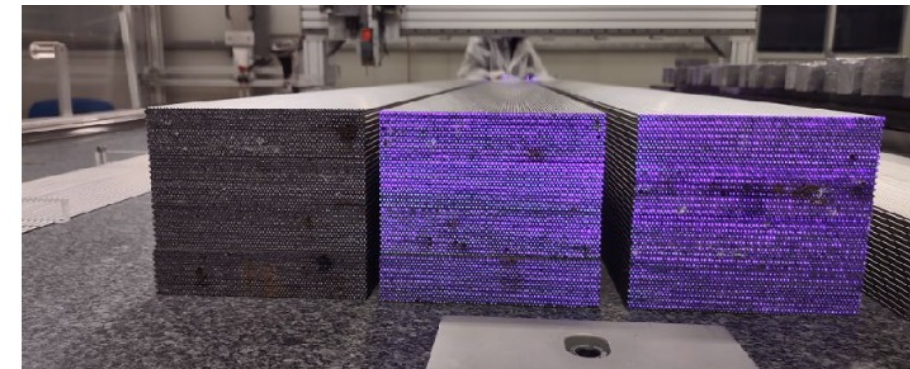
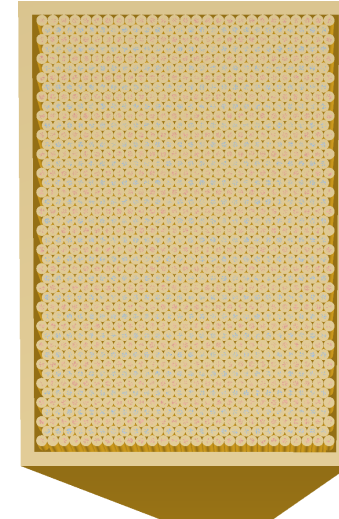
- **“Bucatini” fibre-sampling design**, alternating rows of 1 mm scintillating and clear Cherenkov fibres inside 2 mm stainless steel absorber capillary tubes

Polystyrene / PMMA cores

- **Geometry**, fibers projective to the IP
- **Readout**, one SiPM per fibre, summed analogically to the desired granularity (*1 Readout channel: 8 SiPM*)

Longitudinally unsegmented: depth information from Cherenkov time-of-arrival (< 100 ps under test), not exploited here

Granularity $O(10^3)$ independent channels per event, providing a high-dimensional, spatially structured input for deep neural networks

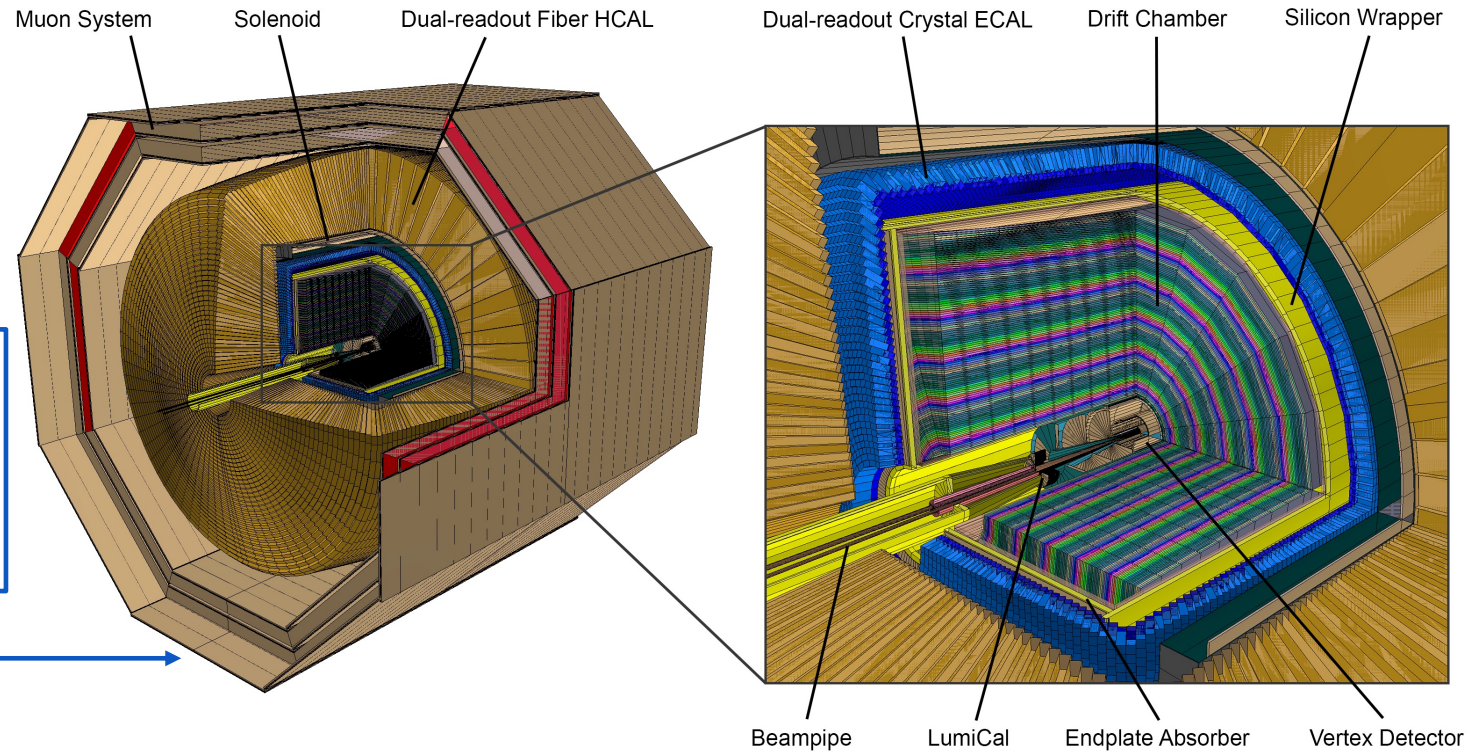


IDEA Full Simulation

There are two simulations for the IDEA Detector Concept in the FCC SW stack (**Key4hep**: Geant4, DD4hep, Gaudi, ...)

- *Original design (IDEA_o1):*
monolithic calorimeter:
fibres embedded in a solid
absorber (*EM + hadronic device*)

- *New design (IDEA_o2):*
crystal-em calorimeter +
new *HiDRa*-like *HAD* calorimeter:
Fibres in individual capillary tubes



For the studies presented, the latest geometry (**IDEA_o2**) is used **without the crystal section**, allowing the response of the fibre calorimeter to be isolated and evaluated

IDEA Full Simulation – *geometry implementation*

➤ Tubes-based dual-readout calo @IDEA_o2

Four levels of nesting

the *capillary-tube technology* has no native "tower" object

→ a **trapezoidal tower** is built by *gluing many tubes together*

ϕ -slice (air) → θ -towers (air) → **capillary tubes** (metal) → **optical fibres** (sensitive)

- **Towers placed projectively inside the ϕ -slice**

dimensions adapted to θ to fit inside the slice

- **Uniform tower length**

same calorimeter depth at all θ

- **Stainless steel tubes**

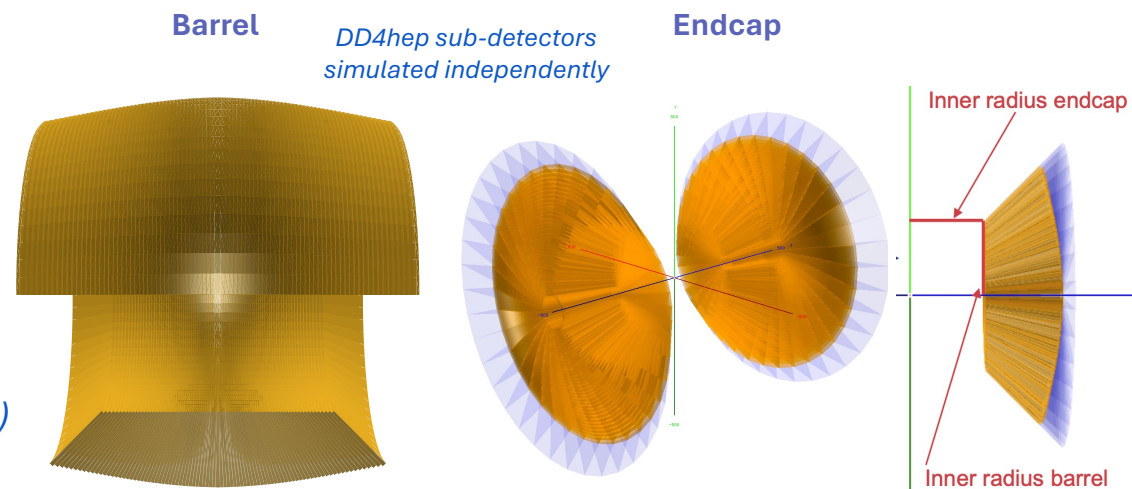
cylinders housing one S or C fibre each, S/C fibers alternate row by row

tubes cut at the first tower face → trapezoidal tower shape

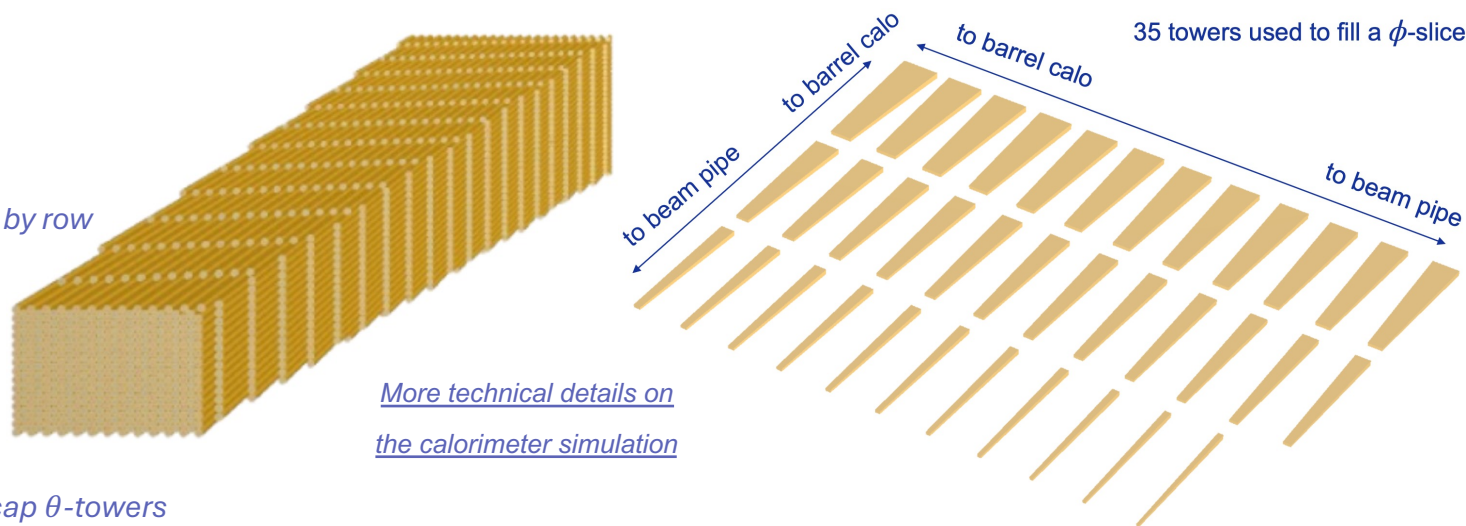
only tubes > 5 cm are retained

- ✓ **Modular geometry**

Inner 2.8 m · tower length 1.8 m · 72 ϕ -staves · 45 barrel / 35 endcap θ -towers



The full calorimeter (barrel + endcap) simulation corresponds to $\sim 8 \times 10^7$ tubes

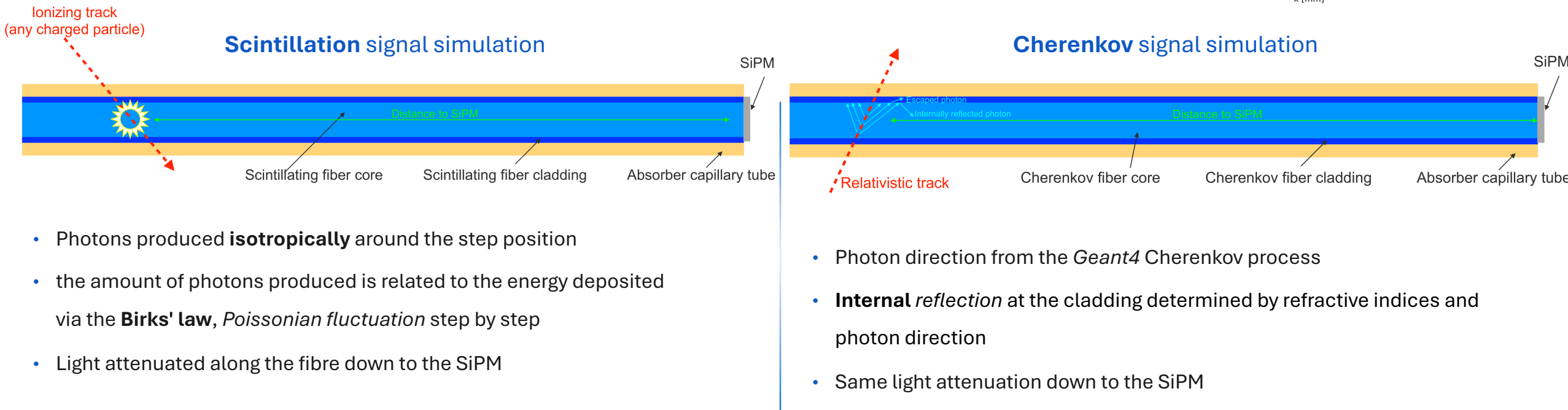
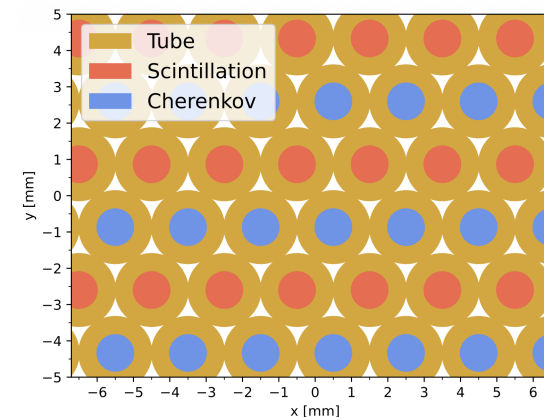


[More technical details on the calorimeter simulation](#)

IDEA Full Simulation – *signal simulation*

- **Two signals to simulate:** scintillation and Cherenkov light, including their production and propagation in the fibres
- **Parametrized light propagation:** based on experimental inputs to enable large-scale MC production,

with more detailed implementations possible for improved physics accuracy



✓ *Both approaches validated against 2021 and 2023 test-beam data, good MC-to-data agreement*

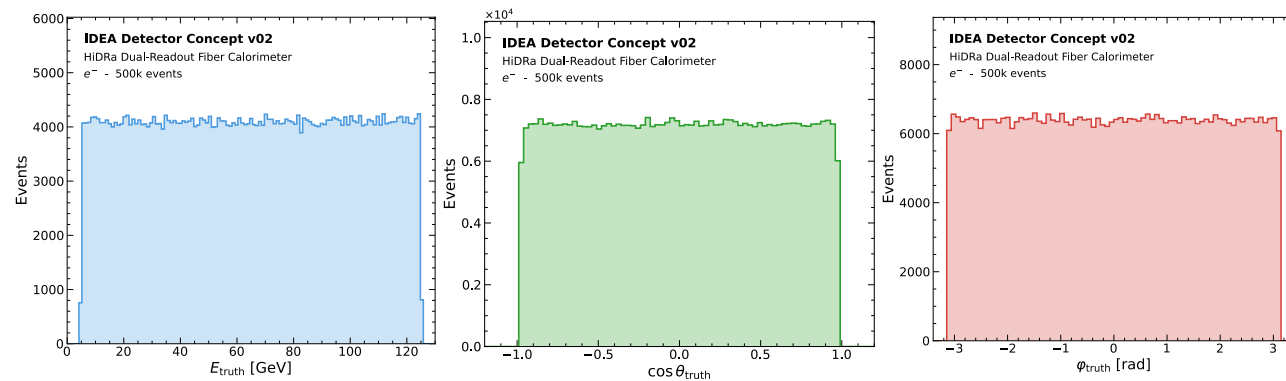
Selection and input distributions

- **Six single-particle samples** are produced using *IDEA_o2* full-sim geo

$$e^-, \pi^0, \pi^-, K^-, K_L^0, K_S^0$$

from the IP, uniform in E (0–125 GeV), $\cos\theta$ and ϕ

e^- : 500k events , hadrons: 250k events



- **Event Selection:** $10^\circ < \theta < 170^\circ$ and $E_{truth} > 5 \text{ GeV} \rightarrow > 94\%$ of events retained

removes inefficient forward/backward regions and

reduces ambiguity for the network

- **Input built directly from the hits**

each fired fibre $\rightarrow$ type, N_{pe} , (x, y, z) of its inner tip

$$x = (\text{type}_1, N_1^{pe}, x_1, y_1, z_1, \text{type}_2, N_2^{pe}, x_2, y_2, z_2, \dots)$$

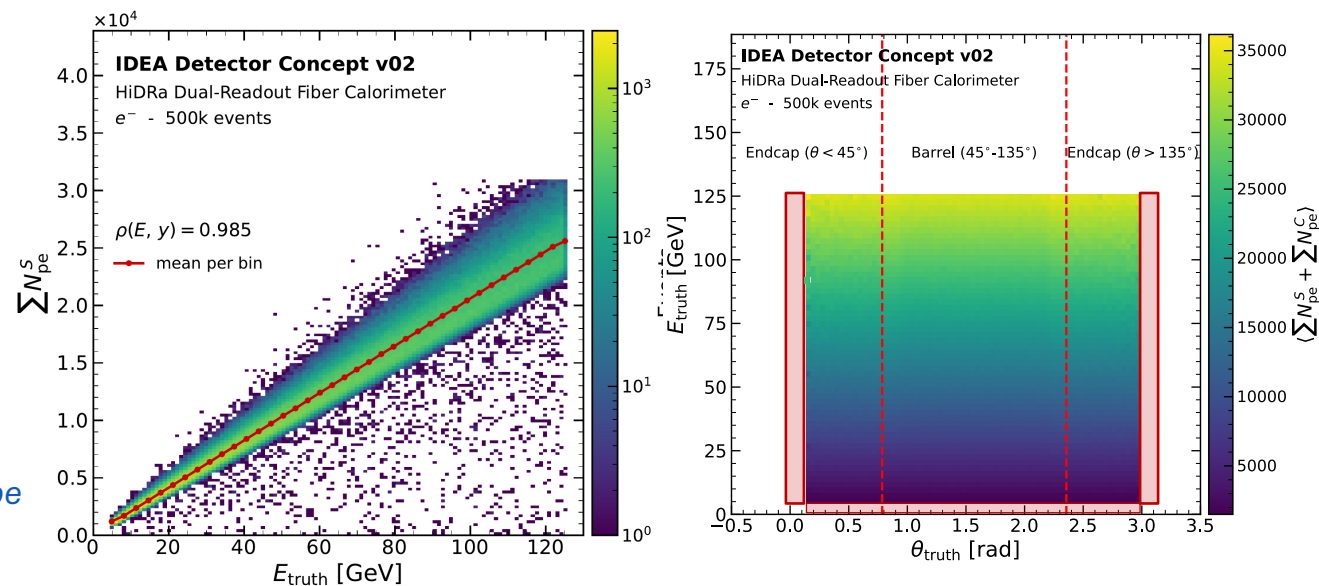
$\rightarrow$ event = flat vector, zero-padded (if needed) up to the fixed input length

- **Length capped at the 99.9th percentile of the multiplicity, per fibre type**

hits beyond the cap are merged into a single energy-conserving overflow hit

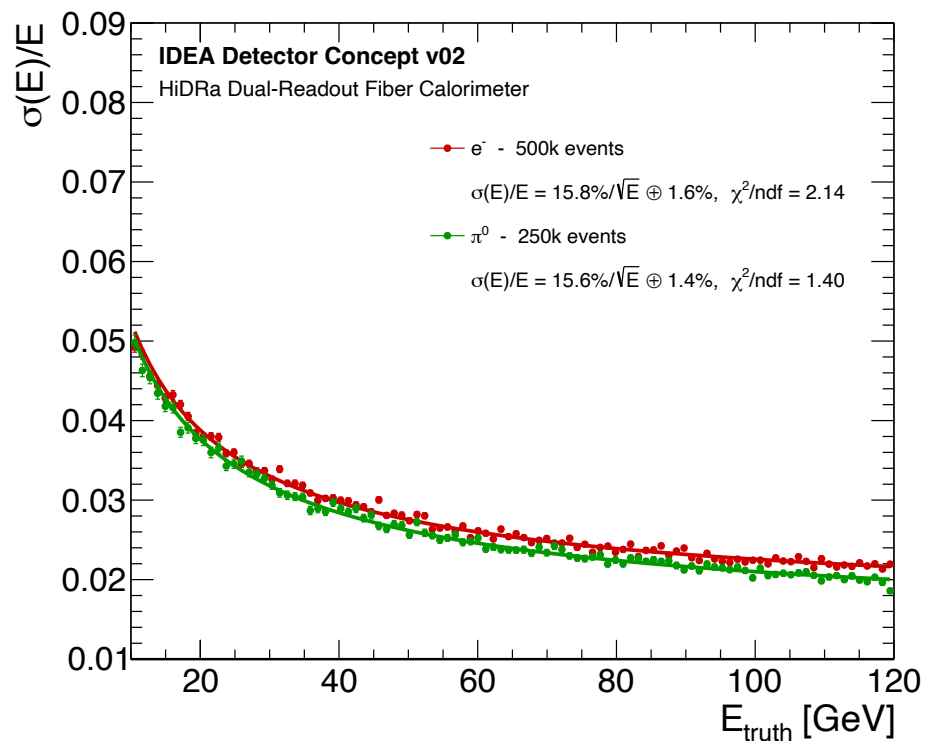
- **Two contiguous blocks (all S fibres by decreasing N_{pe} , then all C)**

in a fully connected network every input slot must keep a stable detector meaning across events



Reference curve reconstruction – *energy*

- A DR reconstruction is applied to the same selected hit-level information used by the DNN, *providing a direct baseline for the network performance*



- For a purely EM shower, $f_{em} = 1$, and both channels are calibrated to $\langle E_S \rangle = \langle E_C \rangle = E_{truth}$

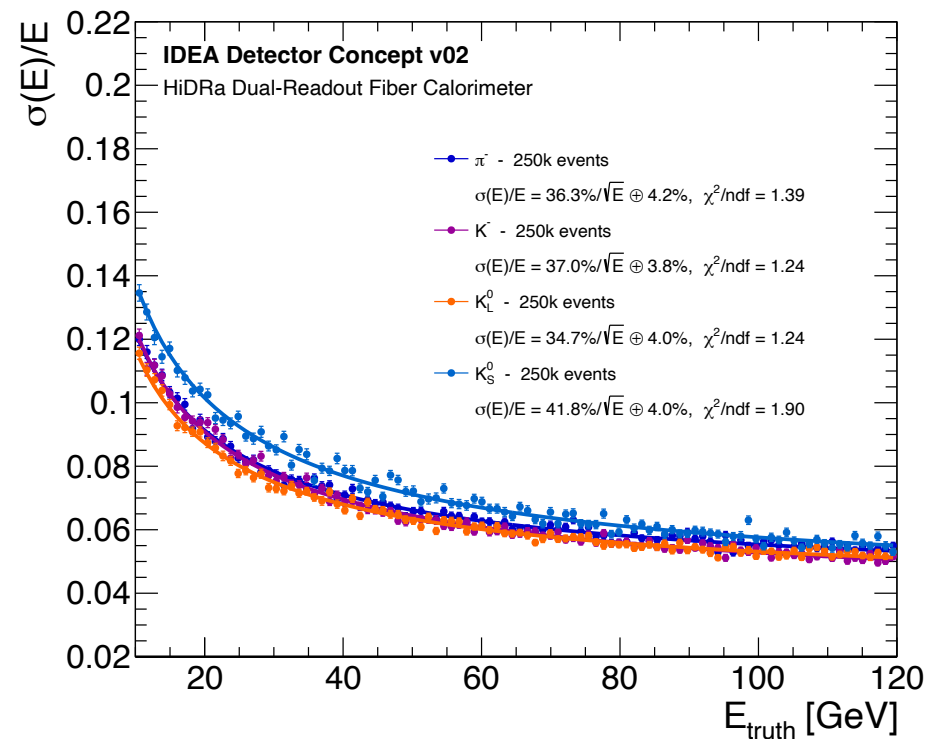
The e^- energy is reconstructed using the *inverse-variance weighted combination* of E_S and E_C :

$$E_{comb} = \frac{E_S/\sigma_S^2 + E_C/\sigma_C^2}{1/\sigma_S^2 + 1/\sigma_C^2}$$

- π^0 : $E = \frac{E_S + E_C}{2}$, giving a comparable stochastic term

$$E_S = N_{pe}^S / K_S$$

$$E_C = N_{pe}^C / K_C$$



- The different response to the *non-EM* component is corrected with

$$E = \frac{E_S - \chi E_C}{1 - \chi} \quad \chi = \frac{1 - (h/e)_S}{1 - (h/e)_C}$$

χ is extracted from a *bias-nulling scan* by matching the E_{reco} to E_{truth}

χ is stable across hadron species: ~ 0.38 , as expected (cal property)

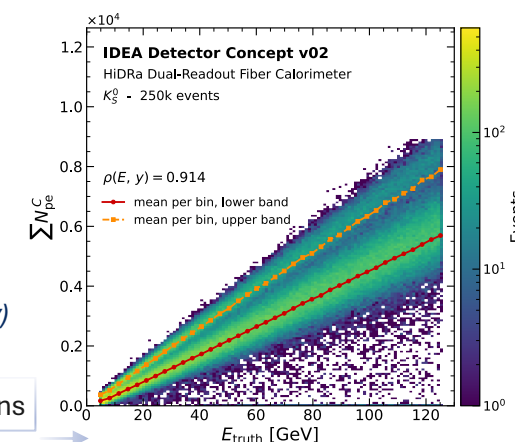
K_S^0 : 69% HAD ($\pi^+\pi^-$) + 31% EM ($\pi^0\pi^0$) → one χ cannot describe both populations

Resolution Fit Function

$$\frac{\sigma(E)}{E} = a/\sqrt{E} \oplus c$$

100 energy bins between 10-120 GeV

	Light yields		
	e^-	π^0	hadrons
K_S [pe/GeV]	206.7	205.6	~ 185
K_C [pe/GeV]	67.3	67.3	~ 46



The Deep Neural Network model

(1)

- The model is implemented in **TensorFlow** through the **Keras** interface

every training on a **single NVIDIA Tesla V100 GPU** (16 GB), **INFN Roma Tre** computing centre

Dataset split: Train set: 80% Test set: 20 % Validation set: 20% of train

- **Input Normalization:** (train-only stats)

- Mean/std computed on x_{train}
- Applied everywhere → **prevents data leakage**

- Target Normalization: Stabilizes training and speeds convergence

→ it allows the network to handle inputs with different scales

zero padded 2D matrix → rows = events, cols = features × hits

Rows

event 1

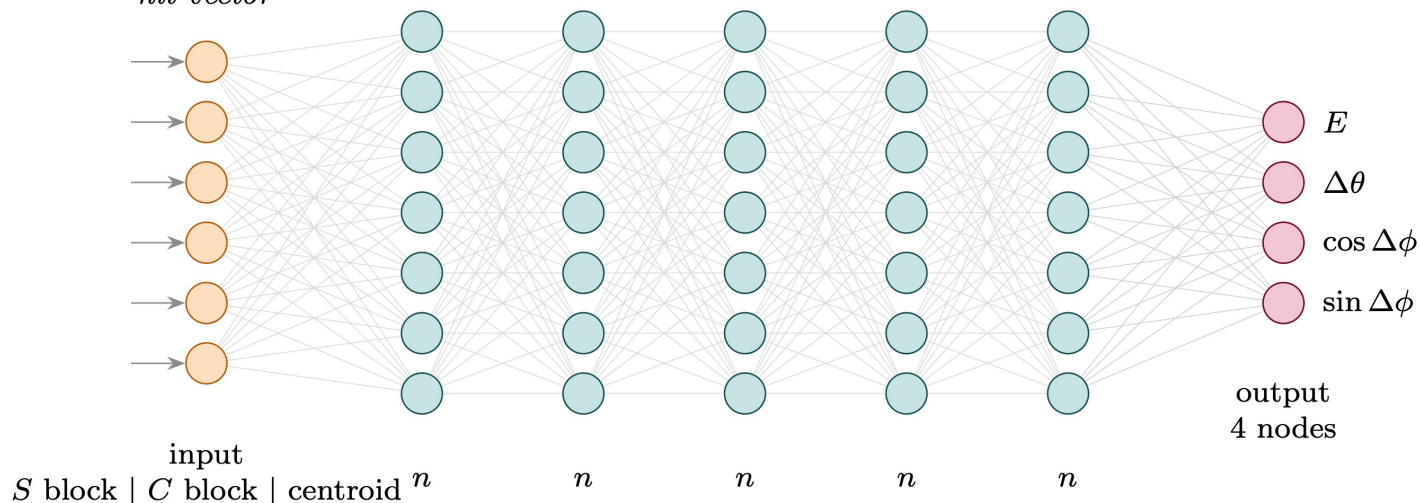
Fibre-type¹, N_{pe}^1 , x^1 , y^1 , z^1 , Fibre-type², N_{pe}^2 , x^2 , y^2 , z^2 , ...

Columns

...

zero-padded
hit vector

5 hidden layers, width n , ReLU



- **Multi-output regression**

Energy + angles predicted as a **correction to the centroid**

$\Delta\theta = \theta - \theta_c$, $\Delta\phi = \phi - \phi_c$ → outputs ($E, \Delta\theta, \cos \Delta\phi, \sin \Delta\phi$)

$$x_c^i = \frac{\sum_i^N x^i \cdot N_{pe}^i}{\sum_i^N N_{pe}^i}$$

→ preventing $-\pi/\pi$ ϕ problems

INFN Roma III Cluster

~75 servers (Blade), ~3200 VCPUs,

Infiniband (DDR 20Gbps / QDR 40Gbps)

128 logical cores per node

4 × NVIDIA Tesla V100 GPUs for ML workloads

1.5 PB storage via GPFS (IBM Parallel FS), 8/16 Gbps FC



The Deep Neural Network model

(2)

➤ **Architecture** → 5 dense hidden layers of equal width, linear output of 4 nodes

Batch size: 5120 + Learning Rate: 10^{-4} from a joint KerasTuner scan

Optimizer: Adam Layer Activation Function: ReLU Early Stopping after 15

➤ **Layer sizing driven by memory to maximize GPU utilization**

$$N_{par} \approx dn + (L - 1)n^2 + no, \text{ with } n \text{ solved at fixed budget}$$

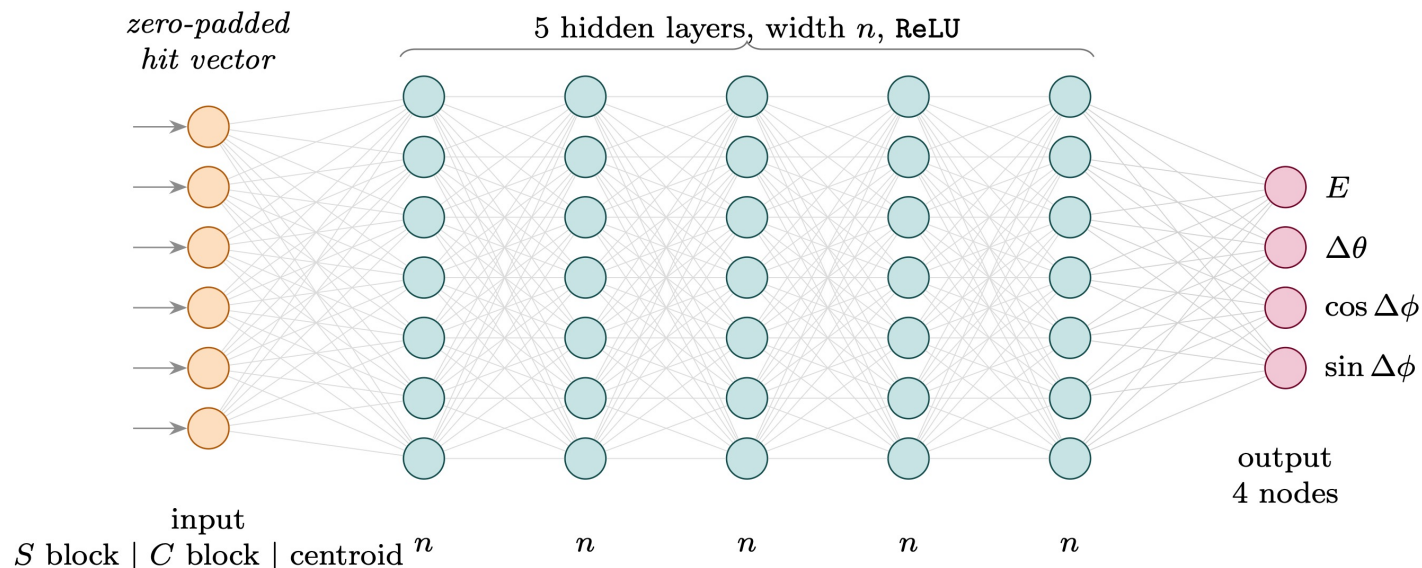
footprint constant across the six species, 12–13 GB at the peak

➤ **Loss Function**, **relative** error on the energy + quadratic terms on the centroid-relative angles

$$\mathcal{L} = E_w \cdot \frac{1}{N} \sum_{i=1}^N \left(\frac{E_{pred}^i - E_{true}^i}{E_{true}^i} \right)^2 + \frac{1}{N} \sum_{i=1}^N [(\Delta\theta_{pred}^i - \Delta\theta_{truth}^i)^2 + (\cos \Delta\phi_{pred}^i - \cos \Delta\phi_{truth}^i)^2 + (\sin \Delta\phi_{pred}^i - \sin \Delta\phi_{truth}^i)^2]$$

d input dimension
 o output dimension
 n number of nodes
 L hidden layers

$E_w = 5$ for π^- , K^- , K_L^0 ; $\times 10$ for e^- , π^0 , K_S^0 due to their EM component



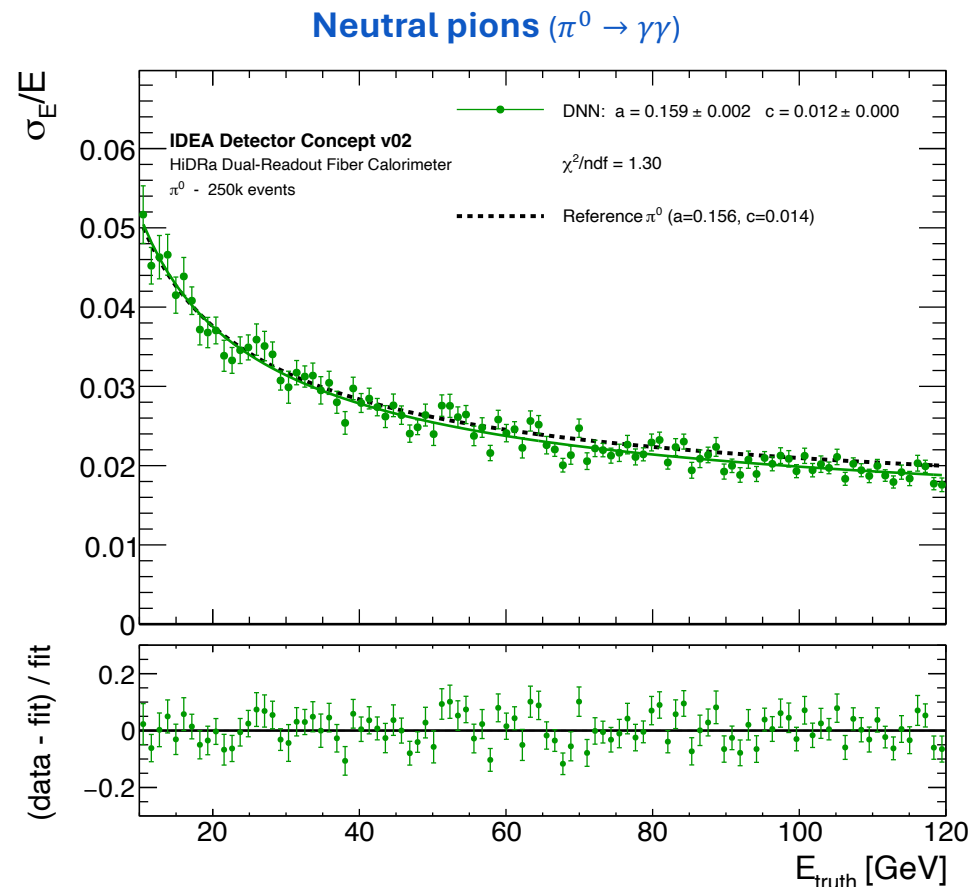
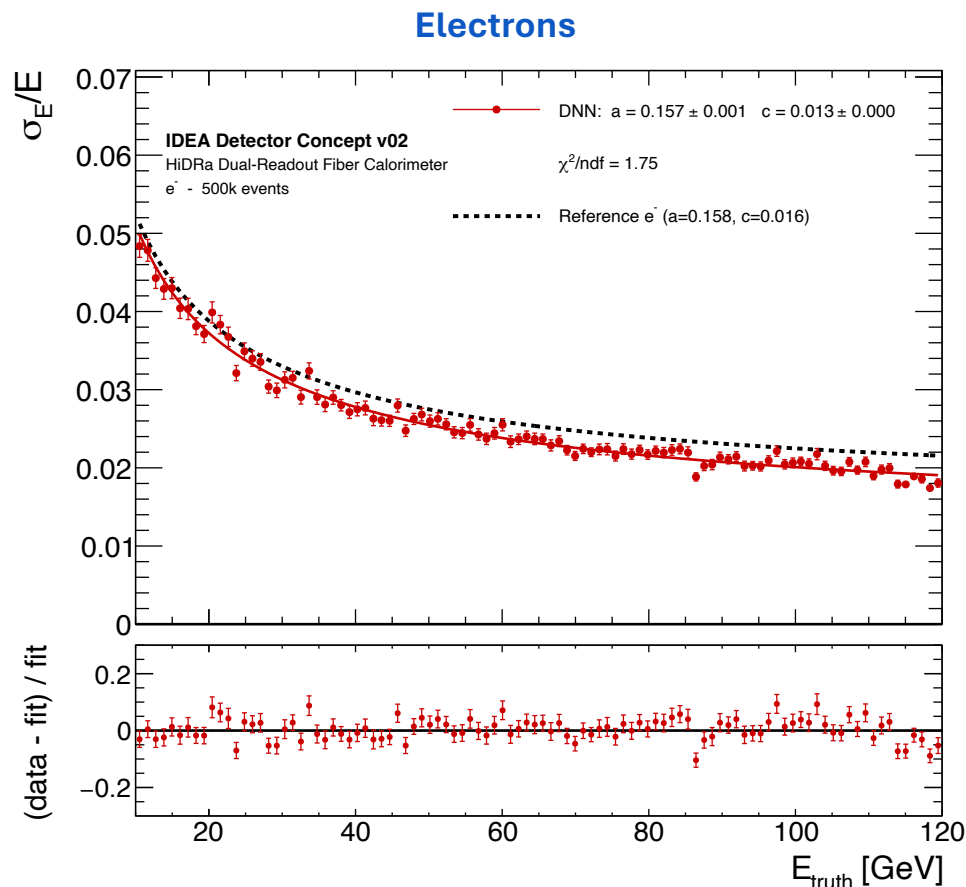
Target model size: **2.8 GB** (pars only) hidden layers: 5

Sample	d	n	Params	w_E
e^-	13 920	12 000	7.4×10^8	50
π^0	15 390	11 900	7.5×10^8	50
π^-	29 260	9 800	6.7×10^8	5
K^-	29 330	9 800	6.7×10^8	5
K_L^0	29 195	9 800	6.7×10^8	5
K_S^0	29 420	9 800	6.7×10^8	50

The Deep Neural Network model – Results

(1)

➤ Electromagnetic cases



The *network matches the analytic DR reconstruction*, already for purely EM showers

Resolution Fit Function

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus c$$

e^- **15.7%** (DNN) vs **15.8%** (Ref)

π^0 **15.9%** (DNN) vs **15.6%** (Ref)

constant term (c): **<1.3%**

stochastic term (a)

The Deep Neural Network model – Results

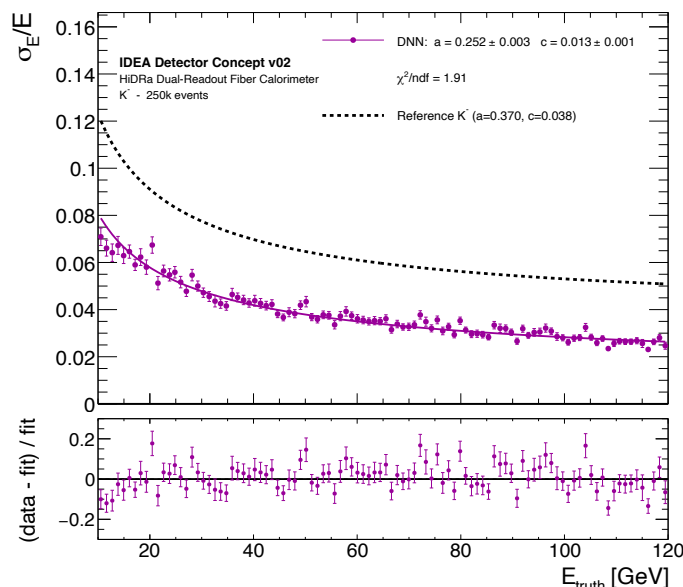
(2)

➤ Hadronic cases

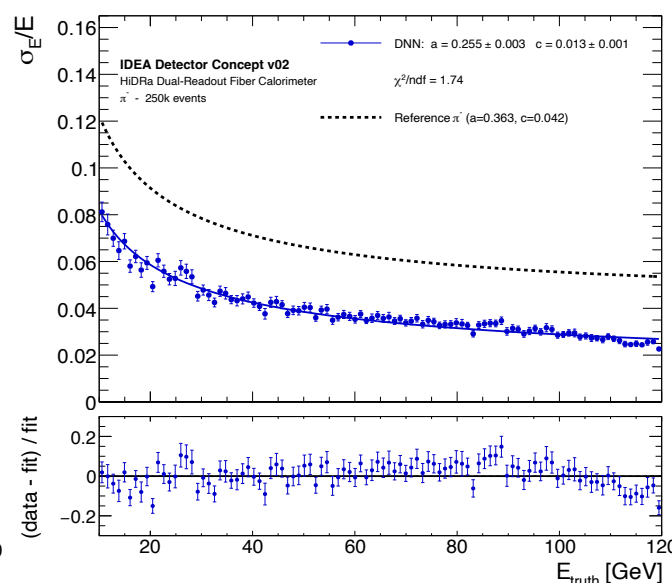
Resolution Fit Function

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus c$$

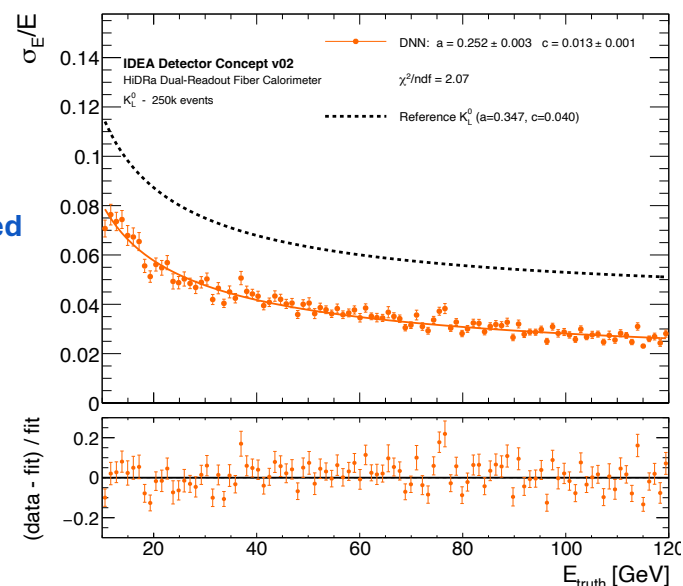
Negative
Kaons



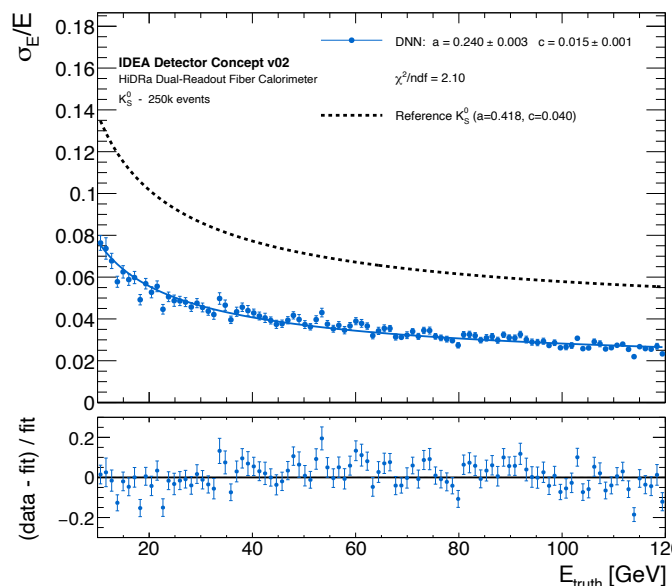
Negative
Pions



Long-lived
neutral
Kaons



Short-lived
neutral
Kaons



The **network outperforms** the
conventional energy-sum reconstruction:

π^- **-30%**

K^- **-32%**

K_L^0 **-27%**

K_S^0 **-43%**

stochastic term (a)

constant term (c): $\approx 4\% \rightarrow 1.3-1.5\%$

The conventional correction uses a **single global χ** to account for fluctuations in f_{em}

The DNN can additionally exploit
event-by-event correlations between the
S/C response and the **shower spatial pattern**

Angular reconstruction – *Definitions and corrections*

(a)

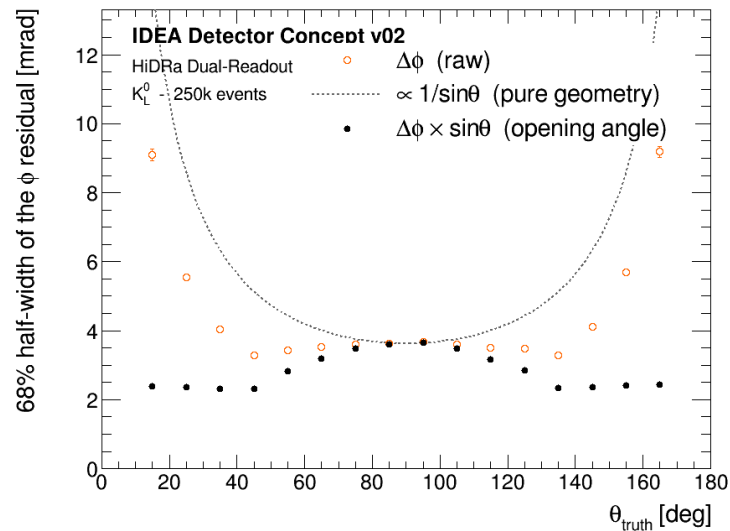
➤ **Reference direction:** N_{pe} -weighted centroid of the fired fibres, taken at the inner fibre tip and projected onto the front face

- The polar angle θ is straightforward
- The azimuthal angle ϕ is harder, for purely geometric reasons

- The azimuthal residual is quoted as an opening angle $\Delta\phi \times \sin\theta$
the angular distance between true and reconstructed direction

a transverse displacement d gives $\Delta\theta = d/r$ but $\Delta\phi = d/(r \sin\theta)$

→ the raw $\Delta\phi$ grows as $1/\sin\theta$ towards the endcaps



- S-shape correction

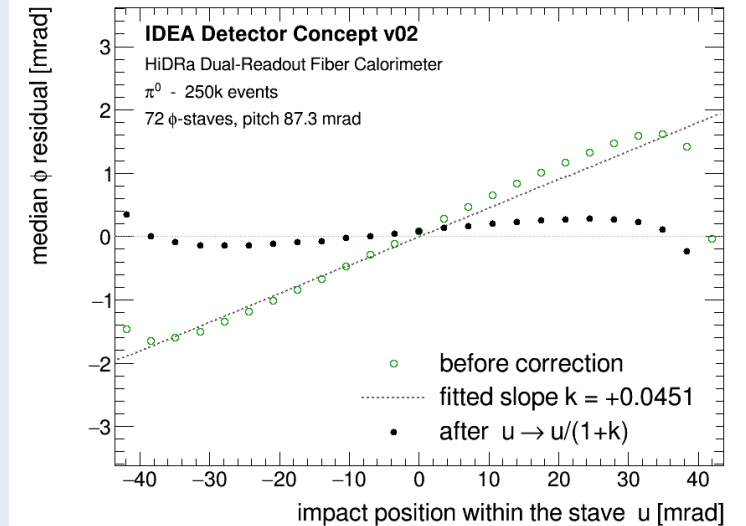
A shower landing near a stave edge has fibres on one side only, so the centroid is pulled towards the stave centre

folding ϕ_{reco} on the stave pitch gives u , the impact position inside the stave, zero at its centre

› the median ϕ residual versus u is a straight line:
 $u \rightarrow u/(1+k)$

→ map back to ϕ

similar behaviour already observed in previous analyses



Angular reconstruction – Definitions and corrections

(b)

- The effect of the two operations is *visible*

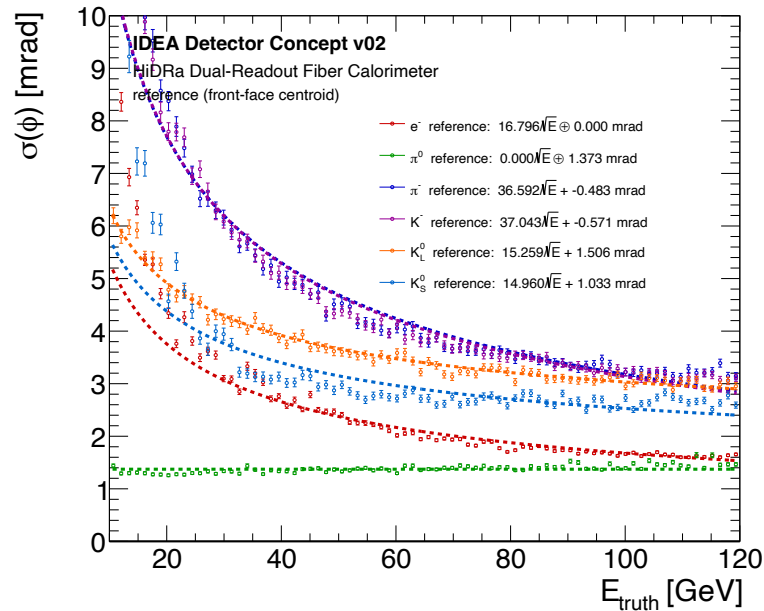
reference, 68% half-width of the $\Delta\phi$ residual

Resolution Fit Function

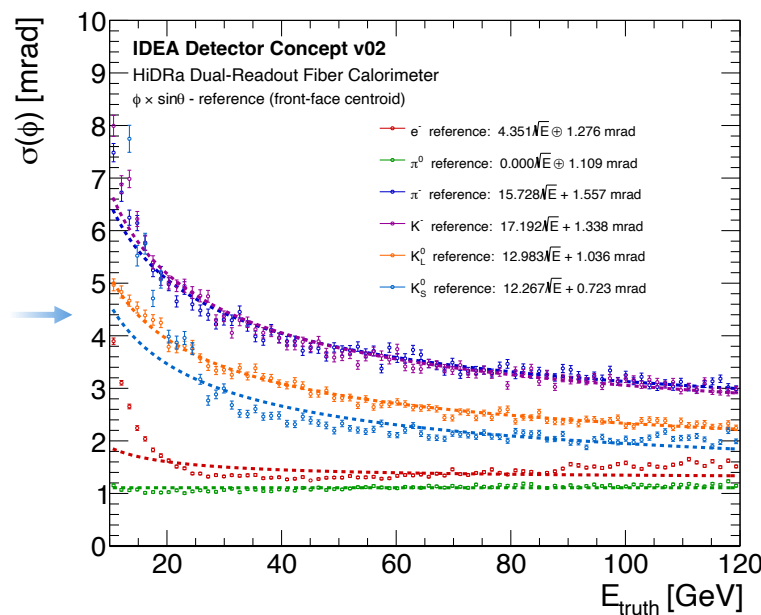
$$\sigma = \frac{a}{\sqrt{E}} \oplus c$$

$\oplus$ for EM, $+$ for hadrons

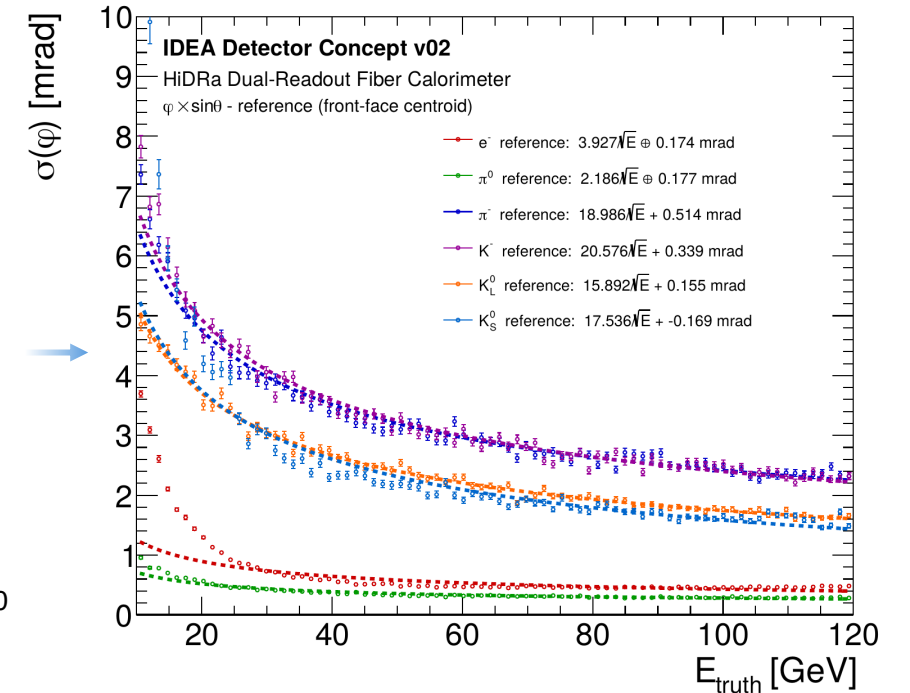
No $\sin\theta$, no S-shape



No S-shape



Including $\sin\theta$ + S-shape

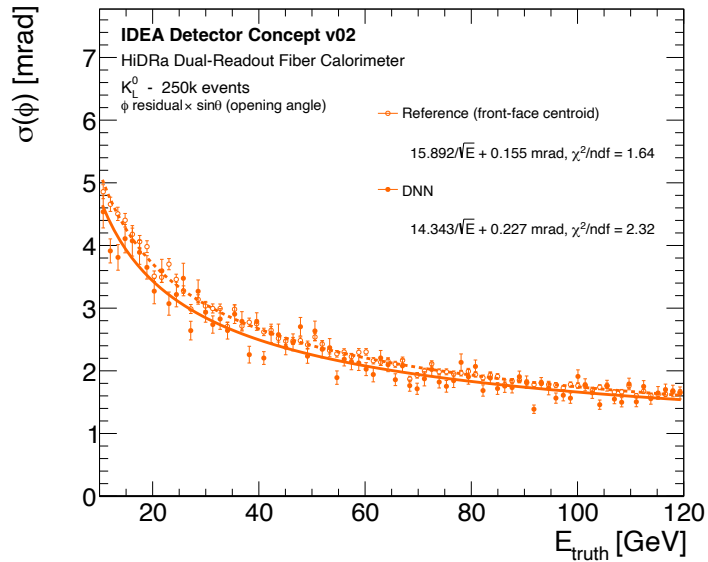
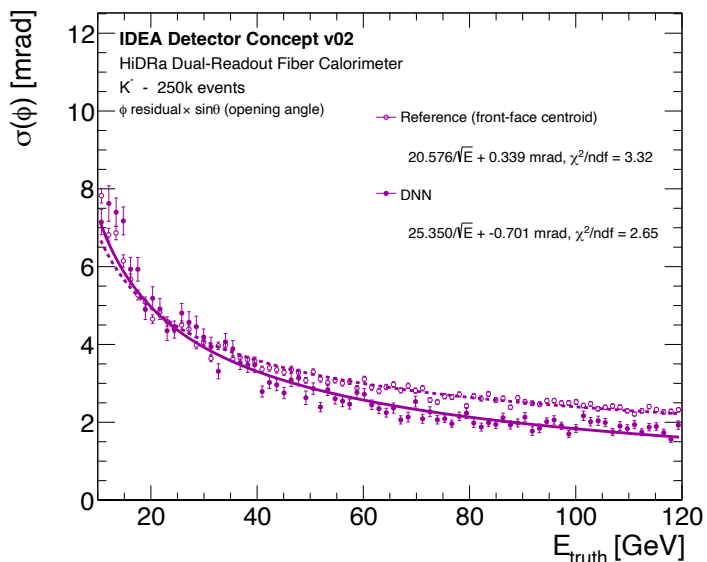
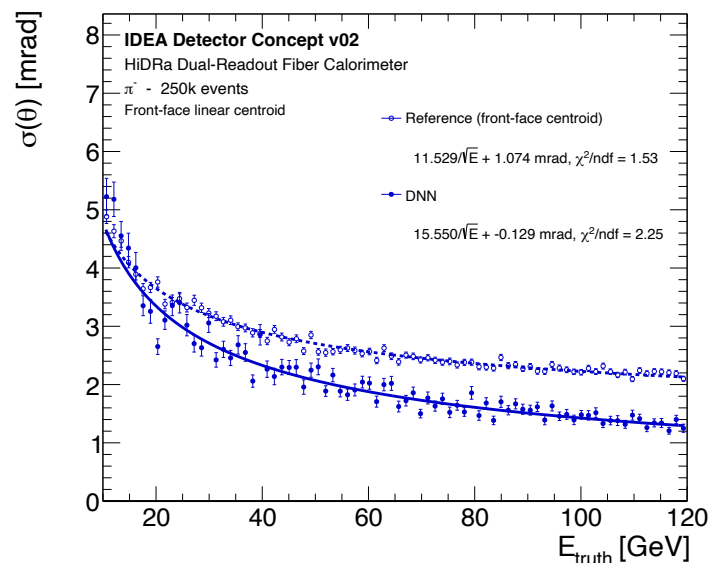
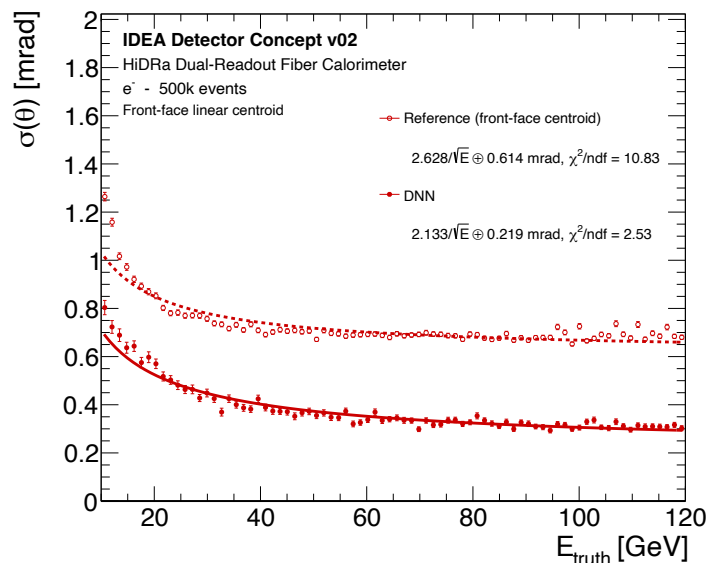


- **without them** the curves are dominated by geometry, not by resolution *e^- gives $16.8/\sqrt{E}$ and π^0 returns no stochastic term at all*
- **with $\sin\theta$** the resolution no longer degrades towards the endcaps *the stochastic term settles on the expected scale: $\mathcal{O}(1 \text{ mrad})/\sqrt{E}$ for e^- / π^0 , $\mathcal{O}(10 \text{ mrad})/\sqrt{E}$ for hadrons*
- **with the S-shape** the bias of the centroid is removed: the constant term drops by about 1 mrad for all samples *and π^0 recovers a sensible $2.19/\sqrt{E} \oplus 0.18$*

the same two operations are applied independently to the reference and DNN predictions

The Deep Neural Network model – Results

(3)



➤ Polar angle $\theta \rightarrow$ **Largest gain**

➔ DNN improves over essentially the full energy range

DNN constant term (c) compatible with zero for all samples, against 0.6–1.1 mrad for the reference

➤ $\Delta\phi \times \sin\theta$ opening angle

DNN comparable or *better* than the centroid estimator over essentially the full range (a/c improvements depend on the sample)

Note

A peculiar low-energy degradation, roughly $\propto 1/E$, affects the **samples with an EM component**

➤ Present in both the reference (see prev. slide) and the DNN

Not corrected for, but possibly consistent with:

- › e^- , magnetic bending
- › π^0, K_S^0 , upstream decays

Under investigation

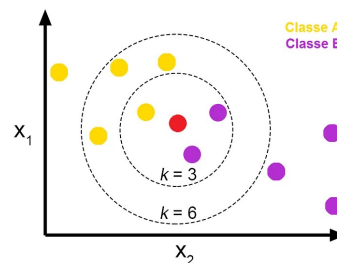
The Graph Neural Network model

implemented in **TensorFlow** through the **Keras** interface

While a flat network treats the event as a fixed-size vector,
a GNN exploits local spatial relationships between hits without requiring a regular grid

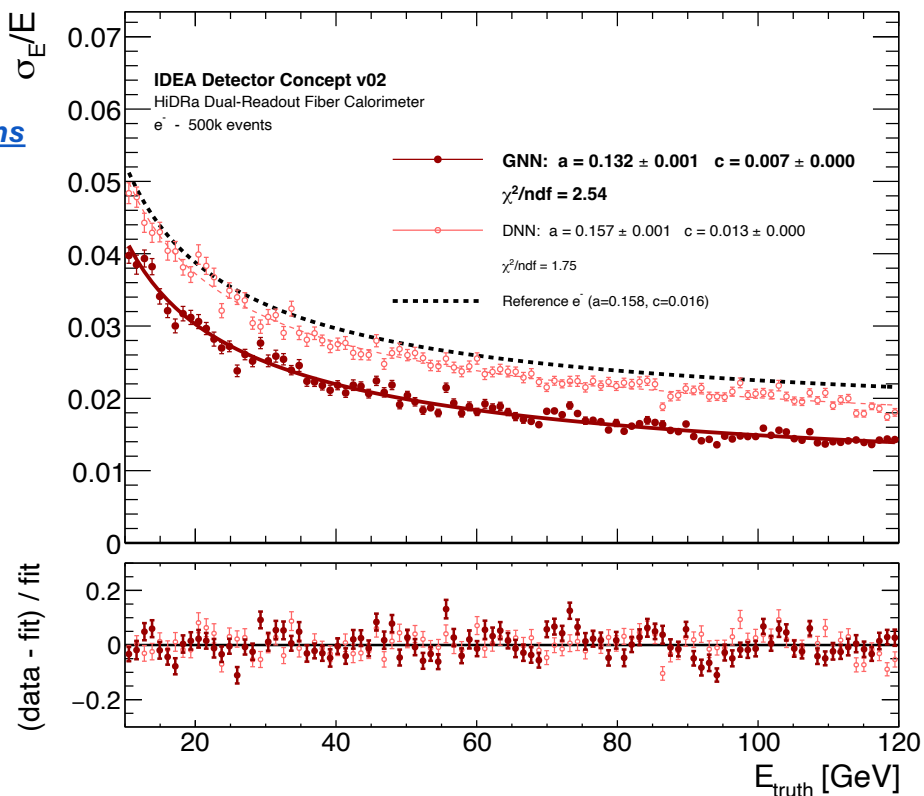
Graph construction

- **Node:** calorimeter hit
- **Edges:** kNN graph, built once offline before training
- **EdgeConv layers:** each hit learns from its local neighbourhood
- **Sum + Mean global pooling** → event-level representation → **MLP head**

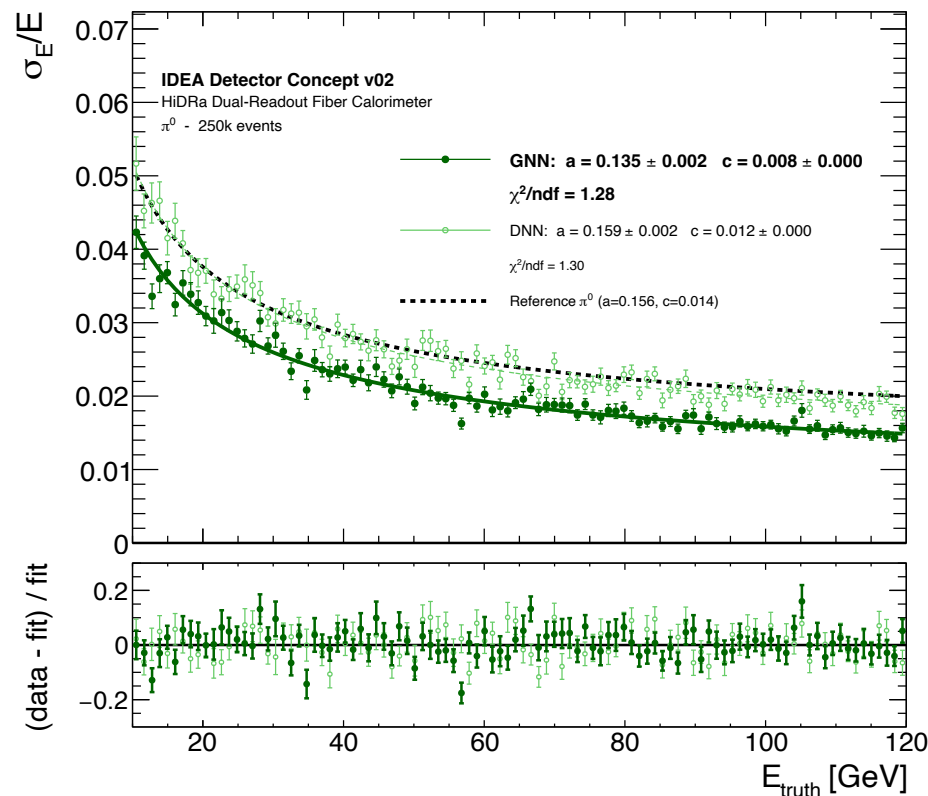


EdgeConv produces substantially larger activations
→ training needs **four GPUs**
→ hit multiplicity **truncated to a maximum number of hits per event**

Electrons



Neutral pions ($\pi^0 \rightarrow \gamma\gamma$)



Very promising results,
with further studies and
optimization ongoing

Summary and Conclusions

FCC-ee is a high-luminosity e^+e^- collider for precision studies from the Z pole to the top threshold, motivating high-performance and highly granular detector technologies

IDEA is a detector concept developed specifically for FCC-ee, combining precision tracking with crystal-EM and *fibre DR calorimetry*

- *The IDEA full simulation is continuously under development*
- *Implemented and optimised a **DNN** that **reconstructs energy and direction** of single particles in the **HiDRa calorimeter**, directly from the hit-level response of S and C fibres*

IDEA_o2 geometry · continuous energy spectrum · full angular acceptance · six EM and hadronic final states

- ✓ **Electromagnetic energy** → on par the analytic DR reconstruction **15.7%** vs 15.8% (e^-), **15.9%** vs 15.6% (π^0)
- ✓ **Hadronic energy** → stochastic term **-27** to **43%**
- ✓ **Direction** → competitive with a dedicated position estimator in θ over the full range and in ϕ

Next Steps

- ❖ **Graph Neural Network** studies already running (**-16.5%** (e^-), **-14.5%** (π^0)). The work now moves to optimisation
- ❖ **Include the crystal electromagnetic section**, improving the EM performance
- ❖ Use of **timing information**, which may provide effective longitudinal segmentation
- ❖ **Extend from single particles to jets**, calorimetric prediction with the tracking

Paper in preparation

Thanks for your attention !

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light to dark, creating a modern and dynamic visual effect. The shapes are primarily triangles and polygons, some of which are semi-transparent, allowing for layered visual depth. The overall composition is clean and professional.

Back-up

Dual-Readout Calorimetry

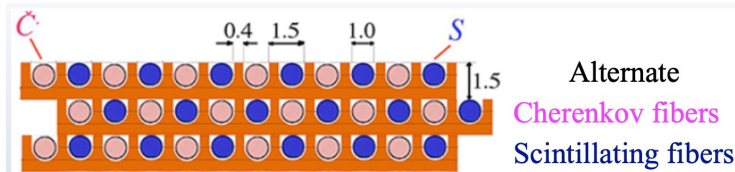
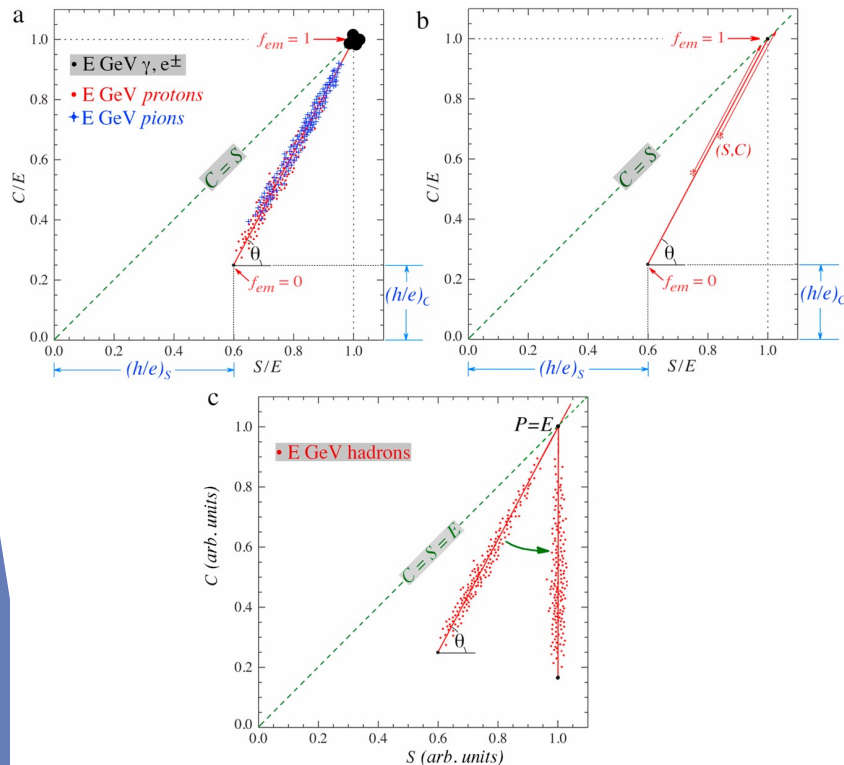
A variable fraction f_{em} of a hadronic shower goes into an electromagnetic component, mostly through π^0 decays, and fluctuates strongly event by event

If $h/e \neq 1$ the energy resolution is **limited by the fluctuations of f_{em}**

Dual readout uses **two channels with different h/e** to measure that fraction event by event and correct for it

$S \propto$ deposited energy; C almost only from the relativistic e^+ of the electromagnetic component

Combining the two turns a non-compensating device into one with a **35–37%/√E** hadronic stochastic term



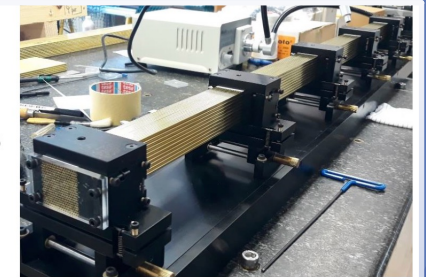
- ❖ Measure simultaneously:
 - > Scintillation signal (S)
 - > Cherenkov signal (Q)
- ❖ Calibrate both signals with e^-
- ❖ Unfold event by event f_{em} to obtain corrected energy

$$S = E[f_{em} + (h/e)_S(1 - f_{em})]$$

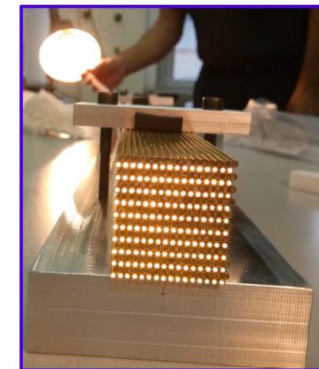
$$C = E[f_{em} + (h/e)_C(1 - f_{em})]$$

$$E = \frac{S - \chi C}{1 - \chi} \quad \text{with: } \chi = \frac{1 - (h/e)_S}{1 - (h/e)_C}$$

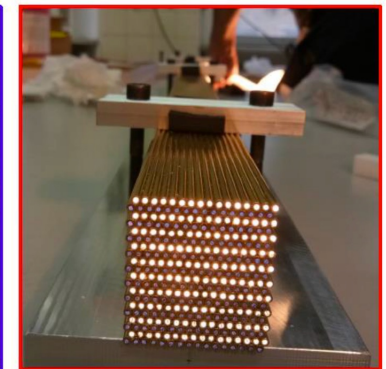
~2m long capillaries



Newer DR calorimeter
(bucatini calorimeter)



Scintillation fibers



Cherenkov fibers

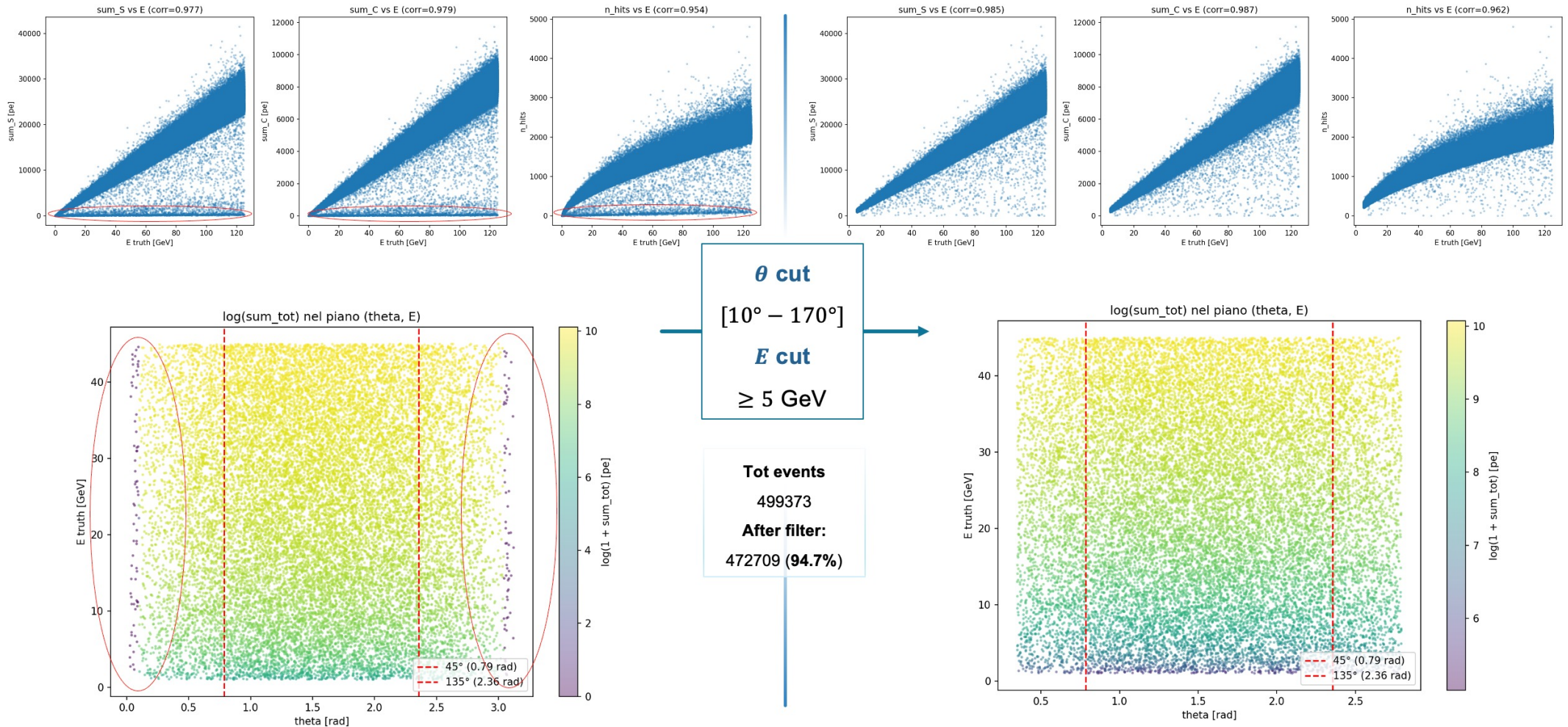
Selection and input distributions

- Six single-particle samples are produced using *IDEA_o2* full-sim geo

$$e^-, \pi^0, \pi^-, K^-, K_L^0, K_S^0$$

from the IP, uniform in E (0–125 GeV), $\cos\theta$ and ϕ

Selection: $10^\circ < \theta < 170^\circ$ and $E_{truth} > 5$ GeV $\rightarrow$ **94.6% of events retained**



Selection and input distributions

- Six single-particle samples are produced using *IDEA_o2* full-sim geo

$$e^-, \pi^0, \pi^-, K^-, K_L^0, K_S^0$$

from the IP, uniform in E (0–125 GeV), $\cos\theta$ and ϕ

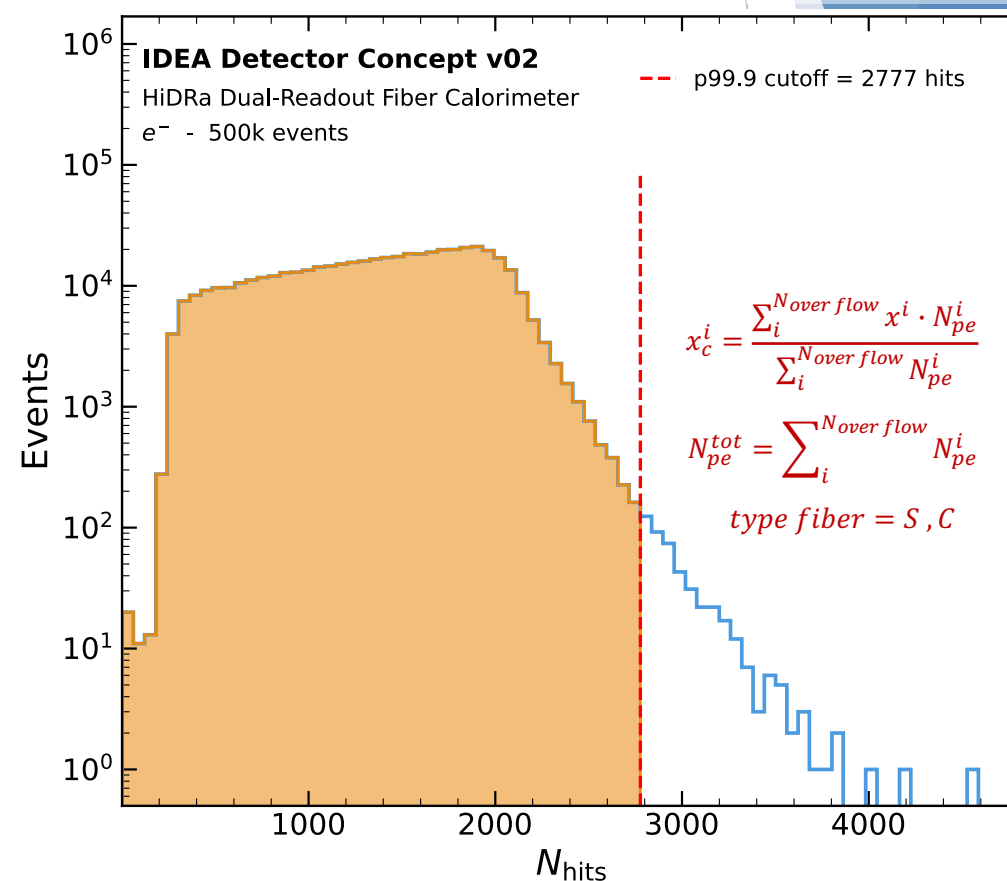
Selection: $10^\circ < \theta < 170^\circ$ and $E_{truth} > 5 \text{ GeV}$

Sample	Generated	Selected	$\epsilon_{\text{sel}} [\%]$	Input feat.
e^-	499 373	472 709	94.7	13 920
π^0	250 000	236 357	94.5	15 390
π^-	249 697	236 260	94.6	29 260
K^-	249 750	236 396	94.7	29 330
K_L^0	249 691	236 318	94.6	29 195
K_S^0	249 841	236 249	94.6	29 420

- A small fraction of events contains extremely large hit multiplicities !

...preserving the relevant physical info...

Length capped at the 99.9th percentile of the multiplicity



- Drastically reduction of CPU(-GPU) request memory

Reference curve reconstruction

Table 2: Reference reconstruction: light yields obtained from each sample, value of χ from the bias-nulling procedure (hadronic showers only), and parameters of the fit of Eq. (1) between 10 and 120GeV.

- **Six single-particle samples** are produced using *IDEA_o2* full-sim geo

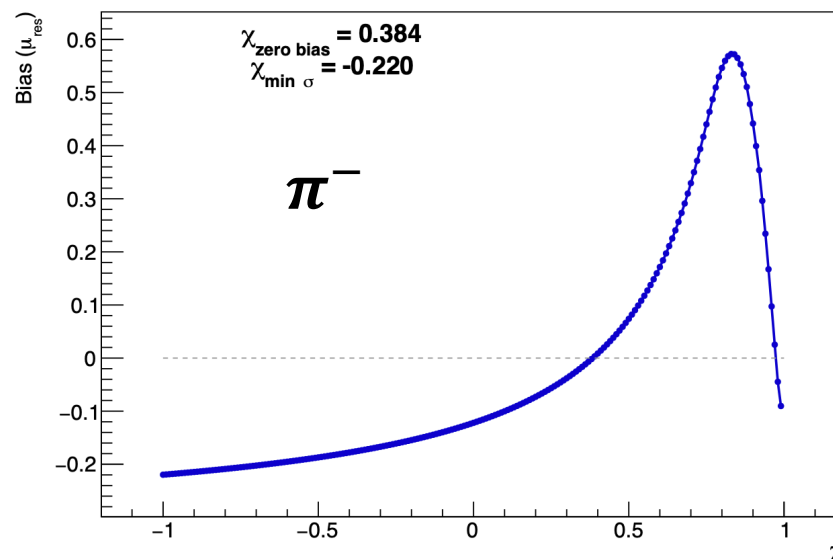
$e^-, \pi^0, \pi^-, K^-, K_L^0, K_S^0$

uniform in E (0–125 GeV), $\cos\theta$ and ϕ

Sample	K_S [pe/GeV]	K_C [pe/GeV]	χ	a [%]	c [%]
e^-	206.7	67.3	—	15.8	1.6
π^0	205.6	67.3	—	15.6	1.4
π^-	185.2	46.6	0.384	36.3	4.2
K^-	184.4	45.7	0.381	37.0	3.8
K_L^0	184.3	45.6	0.379	34.7	4.0
K_S^0	186.9	45.9	0.374	41.8	4.0

χ extrapolation

- Scan χ over 200 steps in the range $[-1, 1]$
- For each χ , reconstruct: $E_{reco} = \frac{S - \chi C}{1 - \chi}$
- Compute the residual $r_i = \frac{E_{reco} - E_{truth}}{E_{truth}}$
- Evaluate the mean bias over all events $bias(\chi) = \frac{1}{N} \sum_i r_i$
- Find the value of χ for which the mean bias is zero, using linear interpolation between two points with opposite-sign biases



GPU usage considerations

○ GPU Usage

CPU → data handling & preprocessing

GPU → feature computation & training

❖ **Model size:** $O(n_{\text{features}} \times n_{\text{neurons}})$ *observed max 1*

– 2 GB

→ *weights + gradients + optimizer states + activations ...*

→ **Goal:** Balance CPU / GPU memory to avoid bottlenecks

Multi-GPU scaling study: 1 → 3 GPUs evaluated

- **"1 GPU"**

no replication, minimal overhead, ...

→ **adopted**

- **"Multi-GPU (2–3, MirroredStrategy)"**

NCCL + replicas + extra threads, ...

→ *not beneficial*

Architecture Optimization

➤ The final architecture resulted from extensive configuration tests and iterative optimisation

- **Initial hyperparameter scan** performed entirely with Keras Tuner.

➤ Tests

- *Hyperparameters' 2D scan*: Batch size x LR scan, hidden layers x nodes
- *Activation Functions*: ELU, LeakyReLU
- *Network topology*: Hs with decreasing nodes, early compression, "hourglass" architecture
- *Energy-Loss weight*: α scan
- *Regularization*: dropout, batch normalisation

➤ **Final configuration**

→ 5 hidden layers, with an equal number of nodes per layer (*maximising GPU utilisation*)

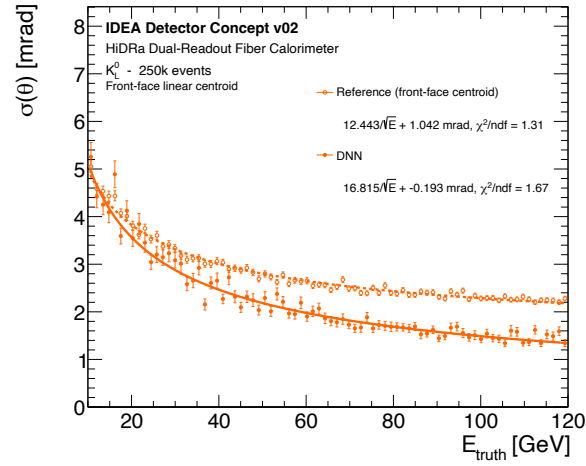
further manual scans, varying one parameter at a time around promising configurations

(e.g. 5 → 3 or 8 hidden layers), to assess its impact

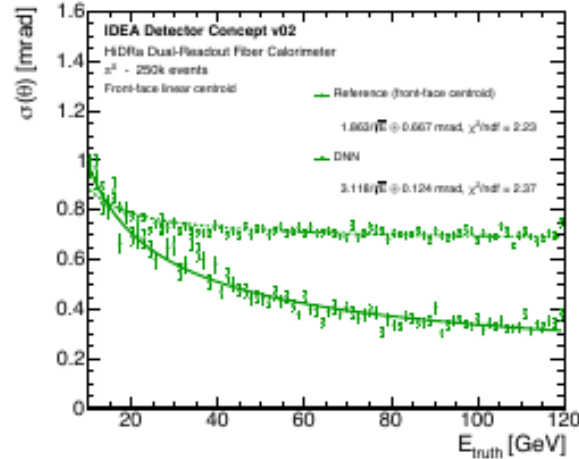
Batch size x LR scan

a larger batch produces fewer and less noisy gradient updates per epoch, so the step size has to grow with it to keep the same effective progress

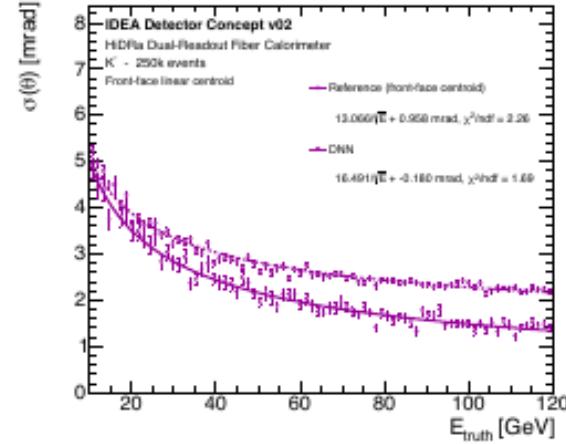
Angular resolution, remaining samples



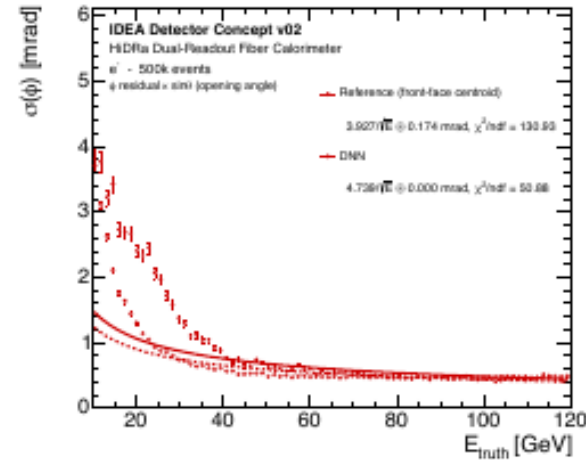
(a) θ, π^0



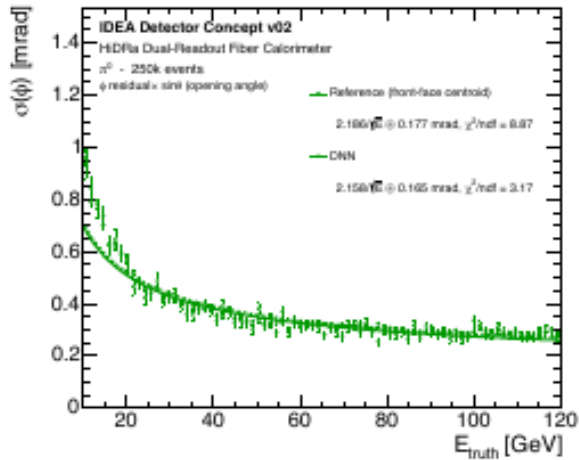
(b) θ, K^-



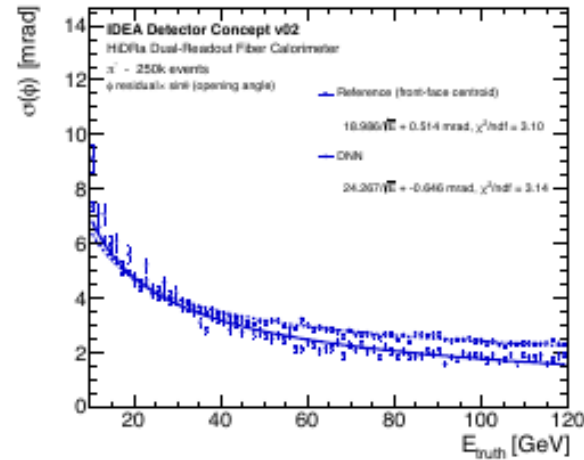
(c) θ, K_S^0



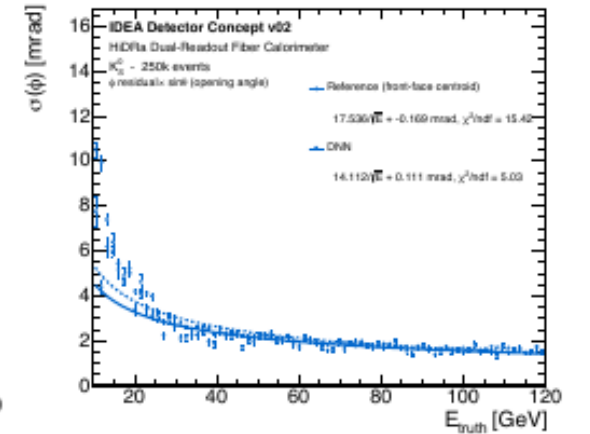
(d) ϕ, e^-



(e) ϕ, π^0



(f) ϕ, π^-



(h) ϕ, K_S^0

Angular resolution, all samples

Polar θ	a ref	c ref	a DNN	c DNN
e^-	2.63	0.61	2.13	0.22
π^0	1.86	0.67	3.12	0.12
π^-	11.53	1.07	15.55	-0.13
K^-	13.07	0.96	16.49	-0.18
K^0_L	12.44	1.04	16.82	-0.19
K^0_S	11.51	0.81	14.12	-0.20

Azimuthal ϕ	a ref	c ref	a DNN	c DNN
e^-	3.93	0.17	4.74	0.00
π^0	2.19	0.18	2.16	0.17
π^-	18.99	0.51	24.27	-0.65
K^-	20.58	0.34	25.35	-0.70
K^0_L	15.89	0.16	14.34	0.23
K^0_S	17.54	-0.17	14.11	0.11

a and c in mrad, 10–120 GeV. **In quadrature** ($\sigma = a/\sqrt{E} \oplus c$) for the EM samples, **linear** ($\sigma = a/\sqrt{E} + c$) for the hadronic ones

same convention as the existing single-particle studies of this calorimeter

In quadrature σ depends only on c^2 — the fit is degenerate in the sign and only $|c|$ is meaningful

in the linear form the sign counts: $c < 0$ means the points fall slightly faster than $1/\sqrt{E}$, and all the negative c sit in hadronic rows

Graph Neural Network

General Pipeline

CSV inputs



Preprocessing

No max_hits cut, Normalization (train) $x_{norm} = (x - \mu)/\sigma, \rightarrow *.npy$



Graph construction (kNN)

K-nearest neighbors, currently only coordinates are used (x,y,z), $k = 4$ (mem impact !)



EdgeConv GNN

Node features: (B, N, F) , Edge index: (B, N, k) , Mask: (B, N)

Layer: **EdgeConv** ($e_{ij} = \phi([h_i || (h_j - h_i)])$)

MLP ϕ learns which local differences are important

MLP in EdgeConv: Dense – BatchNorm – ReLU – Dense – BatchNorm – ReLU

Local Max aggregation $h'_i = \max_j e_{ij}$



Global Pooling

Sum (preserving total energy) + Mean (preserving mean density) [concatenation]{448+448=896}



MLP (Multi Layer Perceptron) head

Dense (256) – BatchNorm – ReLU – Dropout – Dense (128) – BatchNorm – ReLU – Dropout - ...



$E, \sin/\cos\phi, \theta$

4 Gpu needed

$(B, N, k, 2F)$: huge tensors

{393 millions $\rightarrow$ 4byte x float32 $\rightarrow$ 1.57 GB for a single tensor}

mem grows with $N \times k \times F$ with every EdgeConv

$\rightarrow$ Small number of parameters but huge activations

After building the edge features

$(B, N, k, 2F) \rightarrow$ for each event-hit-neighbor

$\rightarrow$ storing $[h_i || (h_j - h_i)]$

$(256, 3000, 4, 128)$ is already huge $\rightarrow$ mem consumption

$\rightarrow$ Transforming local differences into meaningful messages

Multi-Scale EdgeConv

EdgeConv 64



EdgeConv 128



EdgeConv 256

The features are concatenated

$64+128+256=448$