

CNN-Based Hadron Energy Reconstruction in an Imaging Calorimeter System

Machine-learning software compensation

Xin Xia IJCLab

With Roman Pöschl and the CALO5D collaboration

Why do we need precise energy measurement?



Physics analyses

Precise particle energies
for physics observables



Detector R&D

Optimize detector designs
Compare technologies with
a common benchmark

Performance measures in this study

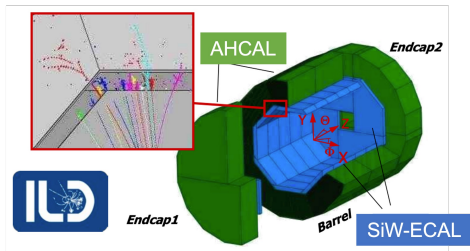
Linearity

Mean response vs. incident energy

Resolution

Event-to-event spread

A case study: the combined ILD ECAL and HCAL



Reconstruct the shower across the whole system.

Two complementary calorimeters

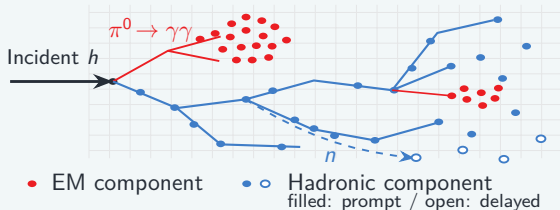
	ECAL	HCAL
Active material	Silicon	Scintillator
Absorber	Tungsten	Steel
Cell size	$5.5 \times 5.5 \text{ mm}^2$	$30 \times 30 \text{ mm}^2$
Layers	30	48
Interaction length	$\sim 1\lambda_I$	$\sim 5\lambda_I$
Response	$e/h \neq 1$	$e/h \neq 1$

More on the SiW-ECAL: Valeria's talk tomorrow.

Shower information enables software compensation

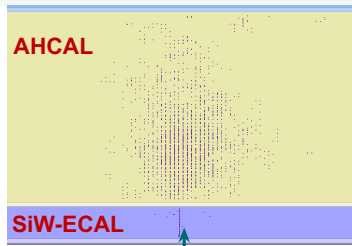
Why compensate? $e/h \neq 1$: changing EM fractions change the visible response.

Schematic hadronic shower in imaging calorimetry



EM: dense, compact, mainly prompt.

Hadronic: sparse / track-like, prompt + delayed.



Incident π^+

Simulated 50 GeV π^+ shower

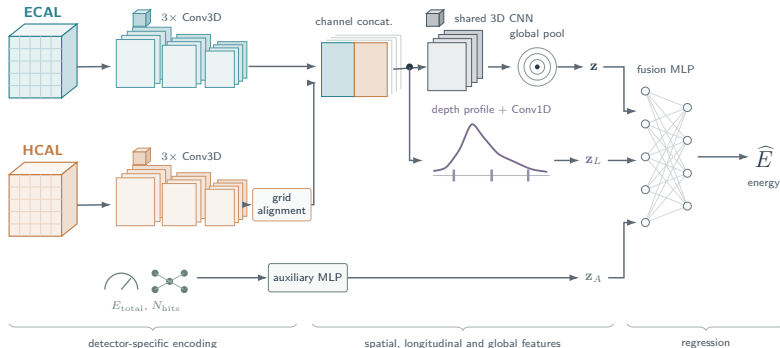
Rich shower information
from imaging calorimetry

CNN-based software compensation

Better energy
reconstruction

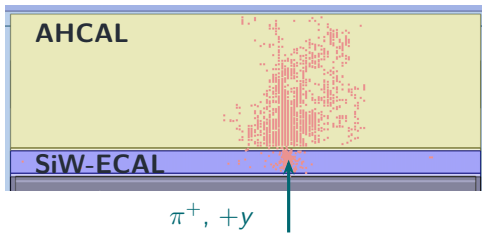
How the CNN estimates particle energy

Detector-aware voxelized representation of reconstructed hits



Separate ECAL/HCAL encoding combines spatial, depth and global features into one energy estimate.

Study configuration



Fixed-direction shower (simulation)

Particle configuration

Single π^+ at the ECAL front surface.
Fixed $+y$ direction; smeared impact position.

Training sample

About 1.3M events in total.
Train : validation : test $\approx 7 : 3 : 3$.
Training energy: 0.2–200 GeV.

Standard ILC software chain

simulation (DD4hep), digitization and reconstruction (Marlin).
Hit energythreshold: 0.5 MIP.

Reconstruction baselines

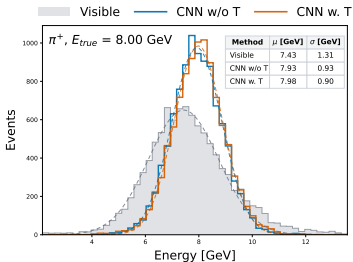
Estimator	Calorimeter information and reconstruction
Visible	Calibrated sum of reconstructed hit energies.
MaxPFO	Highest-energy Pandora particle-flow object, including local software compensation (without tracker information).
CNN w/o T	Detector-aware spatial-energy voxels from ECAL and HCAL.
CNN w. T	Detector-aware spatial-energy-timing voxels from ECAL and HCAL.

Same calorimeter sample and reconstructed hits, with no tracker information.

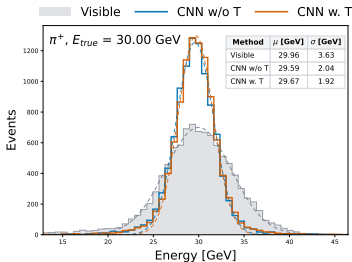
Baseline: reconstructed energy at selected energies

Fixed-direction π^+ at the ECAL surface

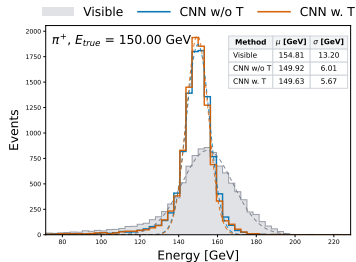
8 GeV



30 GeV



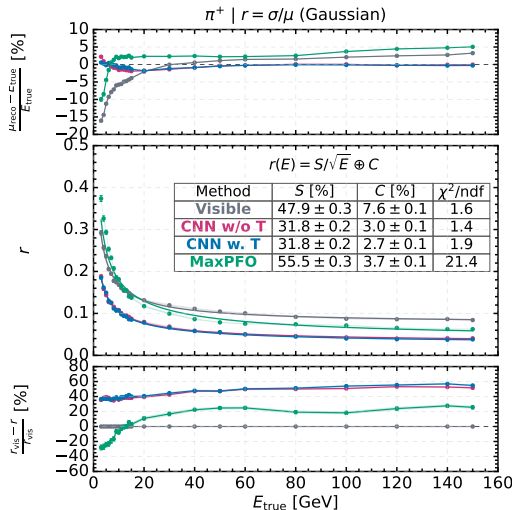
150 GeV



Visible-energy sum CNN without timing CNN with timing

CNN reconstruction narrows the energy distributions across the range.

Baseline: linearity and resolution across energy



Linearity

Nonlinearity within about 2%.

Resolution

About **35–60%** better than visible energy.

$S \simeq 31.8\%$, $C \simeq 2.7\%$ with timing.

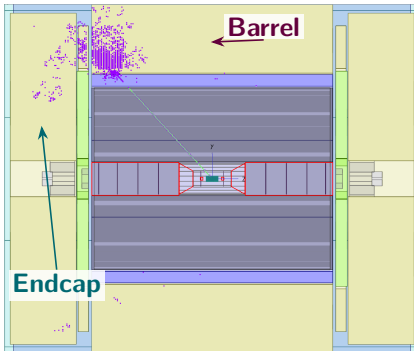
Which information contributes most?

Spatial-energy information provides most of the gain. Timing adds a modest benefit.

Extension: from fixed incidence to full angular coverage

Full-angle sample

Interaction point, isotropic directions

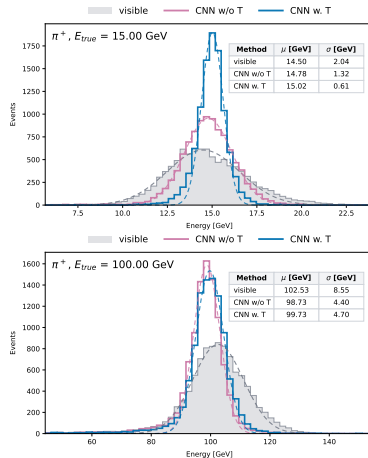
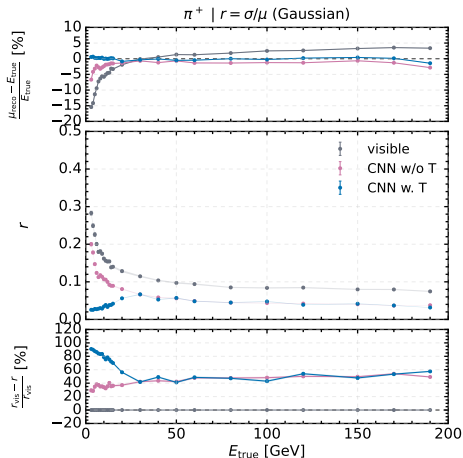


Beyond ECAL-surface, fixed-direction pions

- **Detector region:** barrel, endcap and transition regions.
- **Flight distance:** varying paths from the interaction point.
- **Shower inclination:** different orientations in the voxel grid.
- **Effective depth:** different traversed interaction lengths.

Broader angular coverage changes both the calorimeter response and the shower patterns seen by the CNN.

Extension: results of full angular coverage

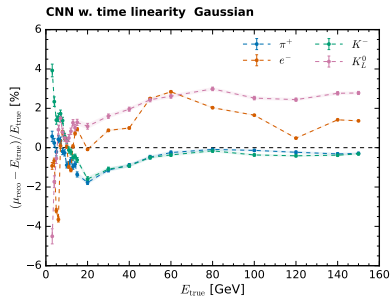


CNNs still outperform the visible-energy baseline.
TOF-like information likely helps, especially at low energies.

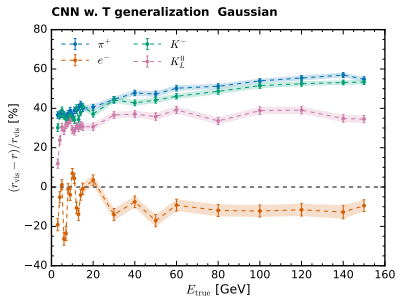
Extension: cross-particle generalization

Train on π^+ ; apply to K^- , K_L^0 and e^- without retraining.

Linearity



Resolution improvement

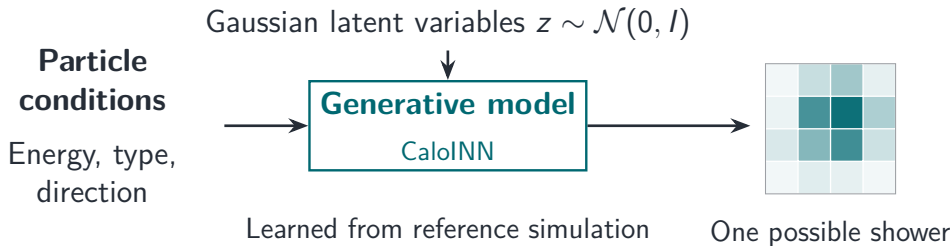


Nonlinearity within 5% for the tested species. Resolution gains persist for kaons; electrons slightly degrade.

Learned shower features transfer across hadrons, with a limit for electromagnetic showers.

Outlook I — Fast simulation: motivation

Generate showers faster while preserving Geant4 fidelity.



Fast simulation

Generate large shower samples faster.

Shower fidelity

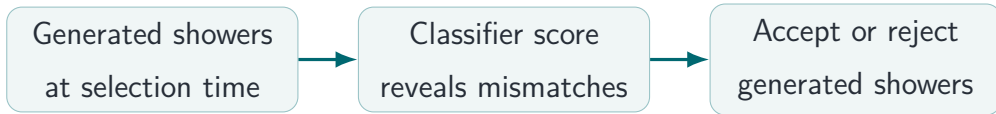
Preserve energy deposits, spatial structure and correlations.

Fast generation must preserve shower fluctuations and correlations.

Outlook I — Fast simulation: resampling

Use a classifier to improve an existing generator

Classifier training: generated showers vs. Geant4 reference.



Probabilistic selection

Use the score to set an acceptance probability for each shower.

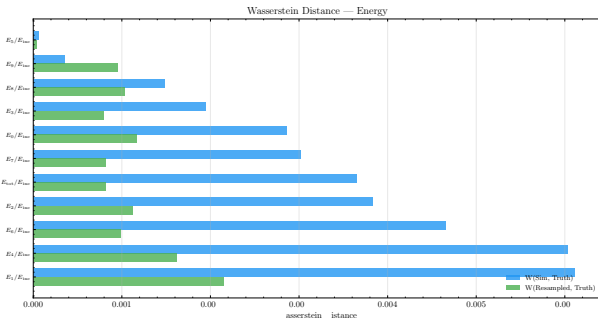
Showers stay unchanged

Select which showers to retain; do not modify their hit energies.

Select a shower sample that better matches Geant4.

Related IJCLab work [Minh: Improving generative models with a classifier](#)

Outlook I — Fast simulation: validation



Energy-distribution agreement

Total and per-layer energy distributions generally move closer to Geant4.

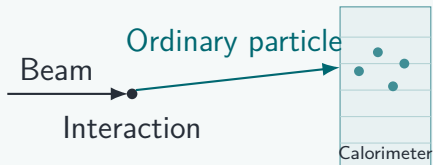
Additional test: reconstruction

Apply one fixed CNN to Geant4 and selected showers. Compare energy response, resolution and tails.

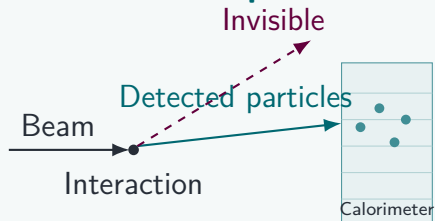
Reconstructed energy provides an additional test of shower fidelity.

Outlook II — Dark-sector searches

Accurate energy measurement



Invisible particles



A genuine energy deficit

Possible use: control hadronic backgrounds

Test whether shower-based reconstruction reduces hadronic events that mimic missing energy.

Summary and manuscript status

Main result

CNN-based software compensation substantially improves energy reconstruction in the combined ILD ECAL + HCAL system.

Resolution improves by $\sim 35\text{--}60\%$ relative to the visible-energy sum; nonlinearity is within about 2% in the original benchmark.

Manuscript status

DRD Calo collaboration review is ongoing
– Nicolas's team (RPTU) has approved

Thank you!

Backup: constructing the voxelized representation

1. Center each subsystem on its shower

$$(x_c, z_c) = \frac{\sum_j E_j(x_j, z_j)}{\sum_j E_j}$$

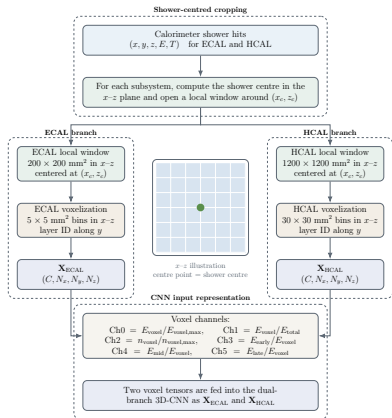
2. Accumulate hits in local spatial grids

	ECAL	HCAL
Window [mm ²]	200 × 200	1200 × 1200
Voxel bin [mm ²]	5 × 5	30 × 30
Grid (N_x, N_y, N_z)	(40, 30, 40)	(40, 48, 40)

3. Preserve detector depth and channels

Layer ID gives depth. Voxels encode energy and hit occupancy, with optional early/mid/late energy fractions.

4. Global inputs: E_{total} and N_{hits} .



Backup: spatial-energy input channels

Channel	Definition	Information retained
0	$E_v/E_{v,\max}$	Local energy concentration
1	$E_v/E_{\text{subsystem}}$	Share of subsystem energy
2	$n_v/n_{v,\max}$	Local hit occupancy

Each ECAL and HCAL branch receives its own normalized voxel grid.

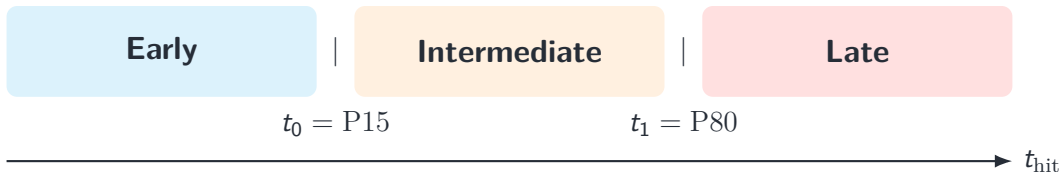
E_v : voxel energy; n_v : voxel hit count. Maxima are evaluated within the corresponding subsystem grid. Empty voxels are zero.

Global energy and hit count retain the event scale.

Shower shape and energy scale enter together.

Backup: how hit times become CNN inputs

Original benchmark: energy fractions in three time windows



Within each voxel, sum hit energies in each window:

$$f_{v,k} = \frac{\sum_{i \in v, t_i \in k} E_i}{\sum_{i \in v} E_i}, \quad k \in \{\text{early, mid, late}\}.$$

Three timing fractions + three spatial-energy channels $\rightarrow$ six input channels.

Time reference: generator production time. Earliest hit contribution; no TOF correction or added smearing.

Backup: timing representation and TOF-like information

Alternative representations

Propagation correction

$$t_i^{\text{corr}} = t_i^{\text{rec}} - \frac{|\mathbf{x}_i - \mathbf{v}|}{c}$$

Subtract a straight-line, light-speed reference.

Tested inputs

Five corrected-time energy bins, or voxel time moments (mean and variance).

Timing encodes shower development and upstream flight information.

Separate the contributions with controlled ablations and realistic timing resolution.

What can remain?

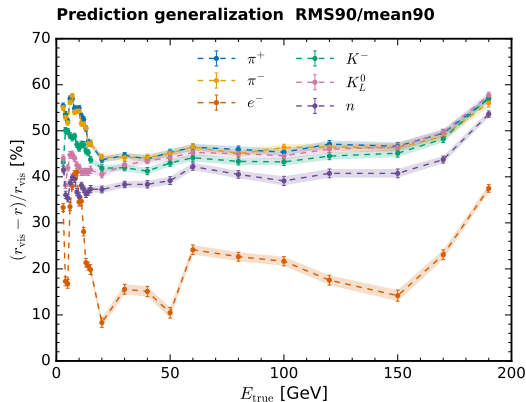
A slower incident particle retains a delay:

$$\Delta t_{\text{TOF}} \simeq \frac{L}{c} \left(\frac{1}{\beta} - 1 \right).$$

Path-length differences and shower delays also contribute.

Relative hit times describe the internal pattern; an event-level reference can retain the overall arrival-time offset.

Backup: one model trained on six particle types



Shared training

$\pi^+, \pi^-, e^-, K^-, K_L^0, n$.

3M isolated events: 500k per type.

Result

All six species improve relative to their own visible-energy baseline, including electrons.

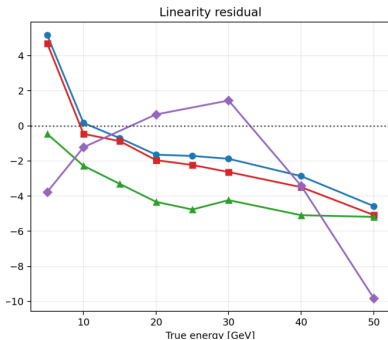
Resolution: $\text{RMS}_{90}/\text{mean}_{90}$.

Broader training extends coverage to both EM and hadronic showers.

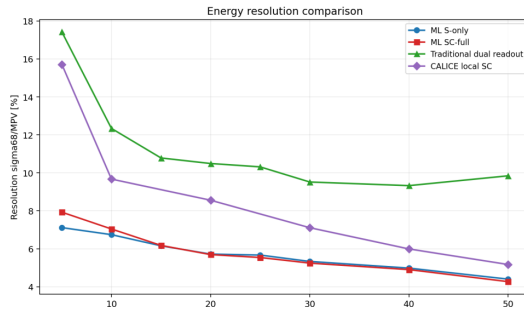
Backup: adaptation to a different calorimeter

Tile dual-readout HCAL: adapted and retrained on π^- , 5–60 GeV.

Linearity residual



Energy resolution



CNNs outperform local software compensation in resolution.

F. Guo, DRD Calo update, 2 October 2026.

Backup: from classifier score to acceptance probability

1. Score generated showers

$D(x)$: classifier probability for the Geant4 class.

With equal class priors and an ideal classifier:

$$\hat{w}(x) = \frac{D(x)}{1 - D(x)} \simeq \frac{p_{G4}(x)}{p_{\text{gen}}(x)}.$$

Larger weights favor underrepresented regions of the generated sample.

2. Accept or reject

Choose a weight scale $Z > 0$:

$$a(x) = \min\left(\frac{\hat{w}(x)}{Z}, 1\right).$$

Draw $u \sim \mathcal{U}(0, 1)$.

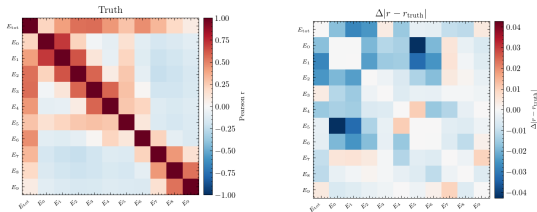
Accept the shower if $u < a(x)$; otherwise reject it.

Each accepted shower stays unchanged.

Only generated showers are scored during selection.

Finite classifier accuracy, support coverage and clipping limit agreement; rejection reduces effective throughput.

Backup: does resampling improve energy correlations?



(Left) PCC of total and per-layer energy from GEANT4.

(Right) $|r_{res} - r_{truth}| - |r_{sim} - r_{truth}|$ (blue is better).

Left: Geant4 reference

Pearson correlations between total and per-layer deposited energies.

Right: change in correlation error

$$|r_{res} - r_{G4}| - |r_{gen} - r_{G4}|$$

Blue: improvement. Red: degradation.

Most tested energy correlations move closer to Geant4 after resampling.