Co-funded by
the European Union

Transformer-based reconstruction of electromagnetic showers in CMS

Yuliia Maidannyk on behalf of the CMS Collaboration

IRFU, CEA Saclay

ML4Jets 2026, Vienna, 14^{th} – 18^{th} September

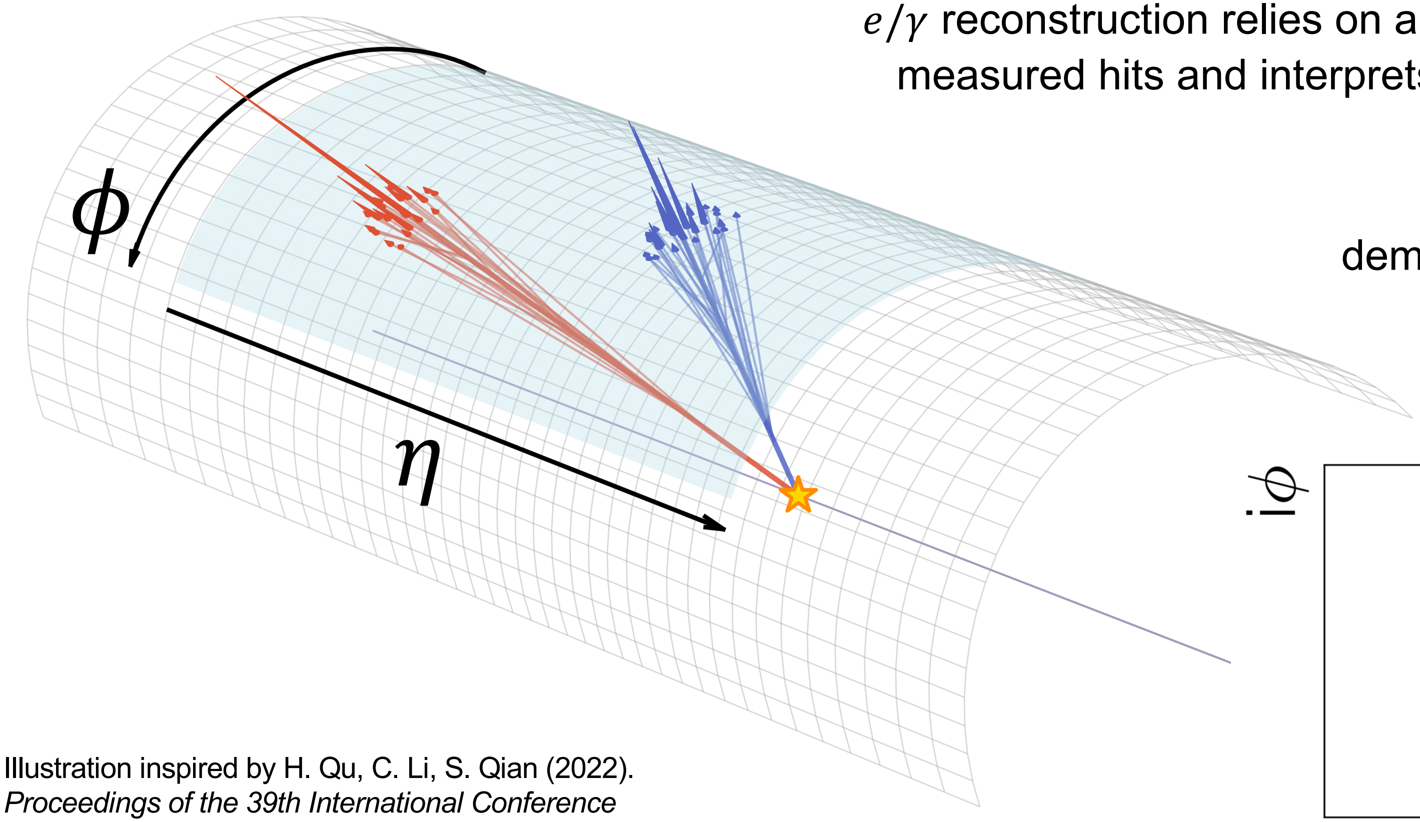
15^{th} September 2026

EM showers and the role of a clustering algorithm

An Electromagnetic CALorimeter (ECAL) detects photons and electrons in the form of **electromagnetic showers.** Looking at the **2D map of crystal coordinates** ($\eta - \phi$ plane) shows **clusters of detector hits** with energy deposits.

e/γ reconstruction relies on a **clustering algorithm** that collects the individual measured hits and interprets them a single particle with $(E_{\mathrm{reco}}, i\eta_{\mathrm{reco}}, i\phi_{\mathrm{reco}})$.

This task becomes increasingly challenging with demands for large datasets, rare decay observations and high-precision measurements.



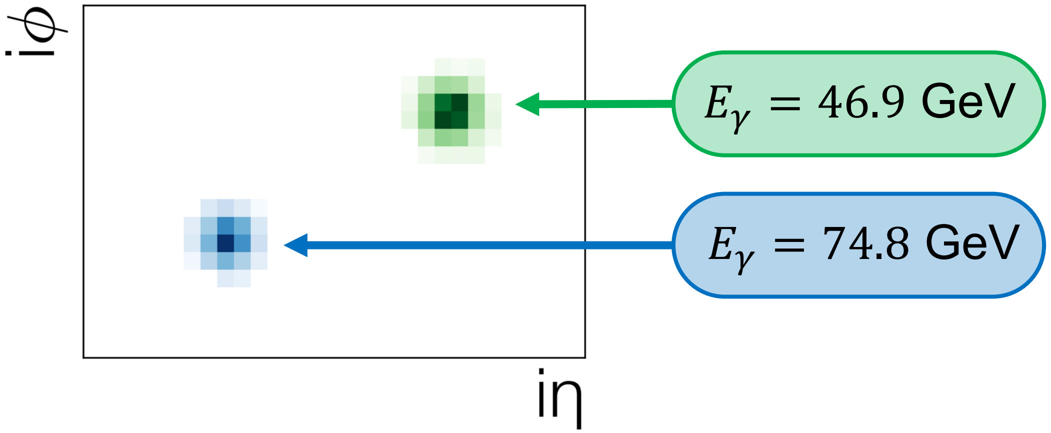
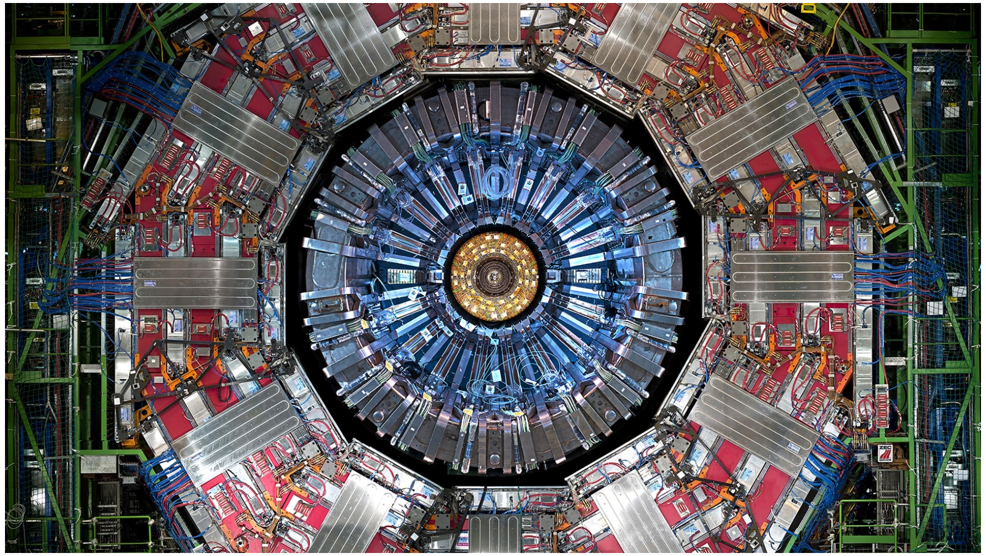


Illustration inspired by H. Qu, C. Li, S. Qian (2022). *Proceedings of the 39th International Conference on Machine Learning*, PMLR 162:18281-18292

The CMS experiment



CMS-PHO-GEN-2008-028

Compact Muon Solenoid (CMS) is a general-purpose detector at the Large Hadron Collider (LHC) at CERN

It collects data from proton-proton collisions at the centre-of-mass energy of 13.6 TeV

Ongoing upgrades to prepare for High Luminosity LHC (2030–), which will deliver a significant increase in the amount of data.

The good

- Searches for rare decays and new physics
- Higher precision measurements

The challenges

- Pileup mitigation (60 interactions → 200)
- Radiation damage
- Computing demands
- Overlapping particle tracks

The CMS experiment: ECAL

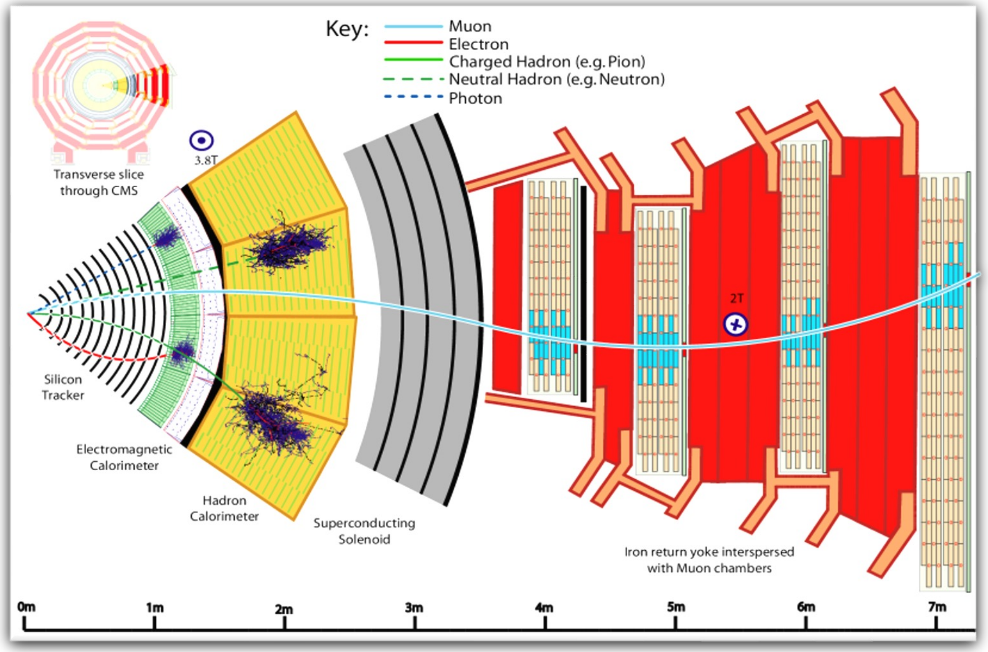
The CMS subdetectors

silicon tracker, electromagnetic and hadronic calorimeters, muon system

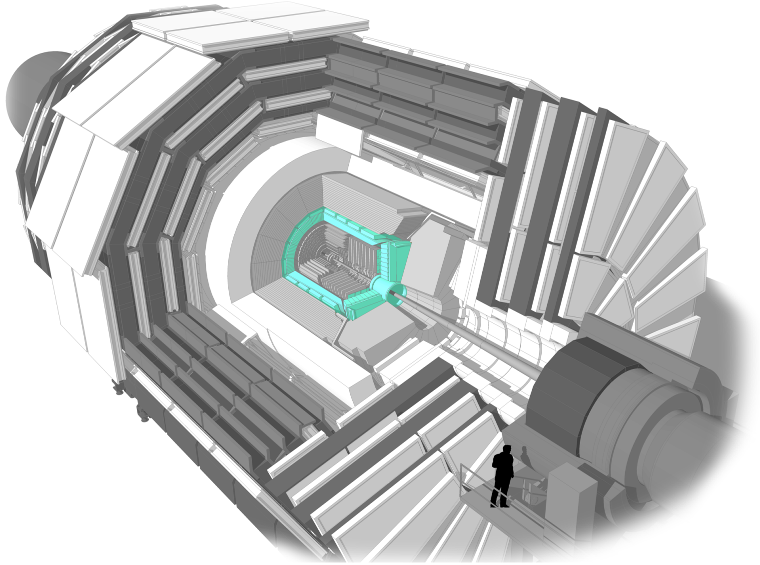
inner → outer

Each subdetector performs its local reconstruction

Particle flow framework combines information from every subdetector



JINST 12 (2017) P10003. 10.1088/1748-0221/12/10/P10003



Electromagnetic CALorimeter (ECAL)

Cylindrical **barrel region** ($|\eta| < 1.5$) + two endcap disks ($1.5 < |\eta| < 3.0$)

The barrel consists of a single layer of 61,200 **scintillating $PbWO_4$ crystals**

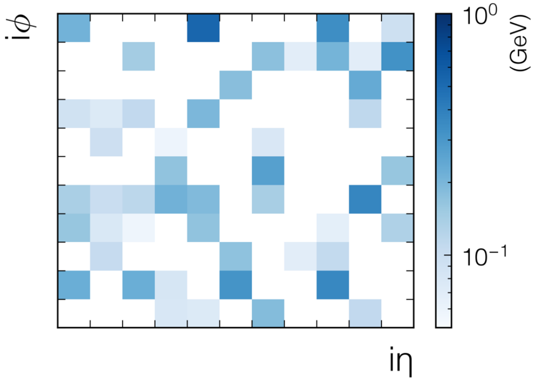
Crystal properties

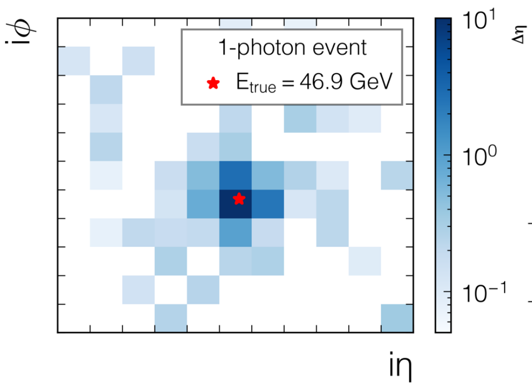
- 2.2x2.2 cm^2 front face and 23 cm depth
- Short radiation length ($X_0 = 0.89$ cm)
- Moliere radius ≈ crystal size

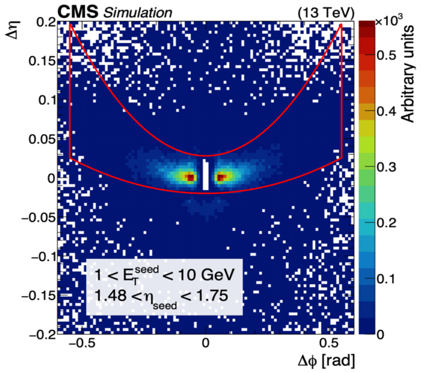
 CMS-OUTREACH-2018-005

The role of a clustering algorithm

PFRechits → PFClustering → PFClusters → SuperClusters







[2]

- **PFRechits:** Calibrated signal amplitudes recorded in individual ECAL crystals.

- **PFClusters**: Output of a **clustering algorithm** that gathers the signal from nearby crystals and interprets them as a single particle. The standard clustering algorithm is **PFClustering** [1], which builds topological clusters around local energy maxima.

- **SuperClusters**: A superclustering algorithm gathers PFClusters coming from the same primary particle. See backup

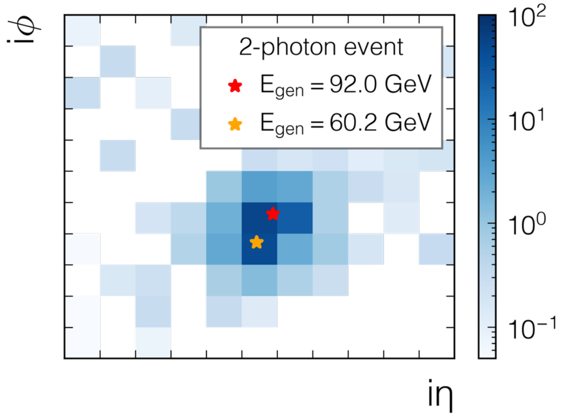
PFClusters are not the final physics objects

[1] *JINST* 12 (2017) P10003. 10.1088/1748-0221/12/10/P10003
[2] *JINST* 16 (2021) P05014. 10.1088/1748-0221/16/05/P05014

Motivation for ML-based clustering

PFClustering advantages

- Simple
- Good energy resolution after correction
- High signal efficiency for isolated particles, even at low energies

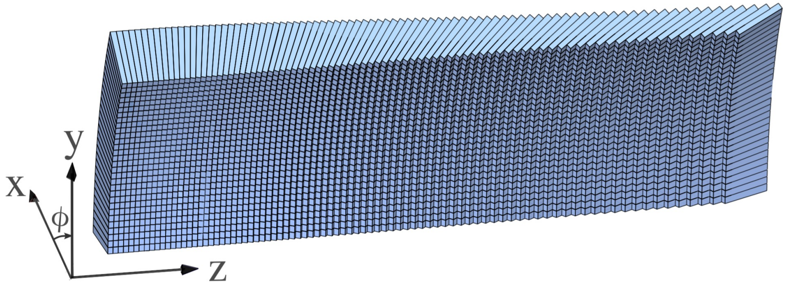


Shortcomings	Problematic for ...
overlapping showers	HL-LHC, $\pi^0 \to \gamma\gamma$, $H \to AA \to 4\gamma$ search, converted γ, $\gamma + \text{jet}$
high background rate at low E	HL-LHC, an aging detector
requires an energy correction	shower leakage, dead crystals

Goal: use advanced ML techniques to reconstruct e/γ from calorimeter hits and improve upon PFClustering

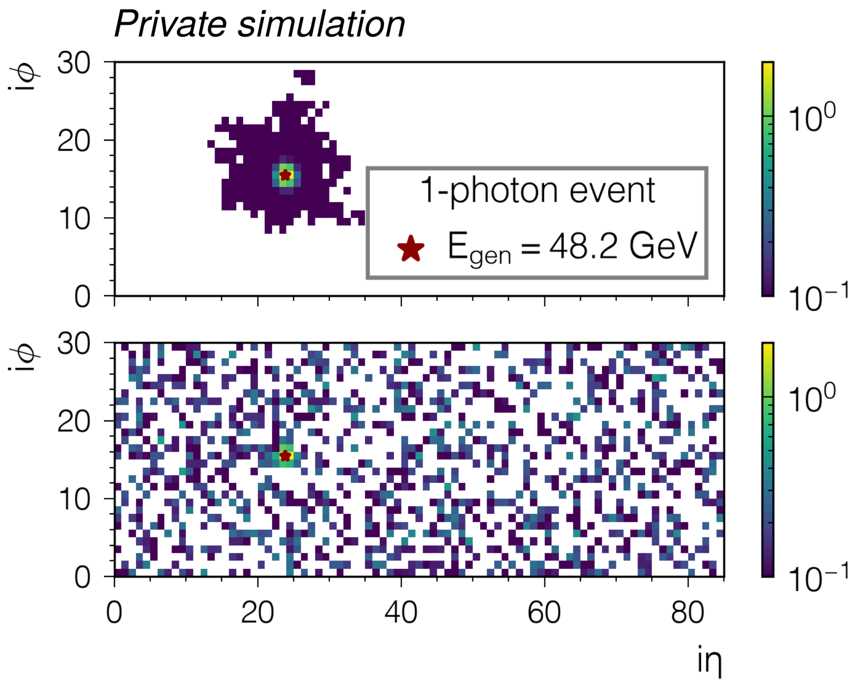
Data generation

30 x 85 x 1 crystals
Size: 2.2 x 2.2 x 23 cm
Material: $PbWO_4$



Toy calorimeter – Proof of concept before the full detector simulation

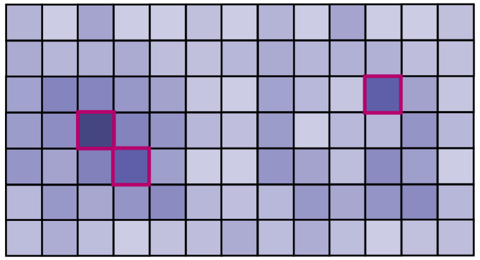
- GEANT4 simulation
- Quasi-projective geometry causes explicit (ϕ, η) dependence
- Particle gun at interaction point
- 1 layer of 85x30 crystals $(0 < \eta < 1.479,\ -15° < \phi < 15°)$
- 33 mm of upstream silicon (tracker material budget)



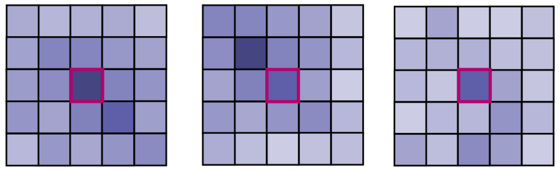
- Photon gun with uniform $(E_{gen}, \eta_{gen}, \phi_{gen})$ covering 1-100 GeV range → sample of **isolated showers** (1-photon)
- **Calorimeter windows**: records of energy deposited into each cell E_{xtal}
- Build a sample of **overlapping showers** (2-photon) requiring $\Delta R < 3$ crystals
- Add **Gaussian noise** to the readout

$$\sigma^2 = (0.03)^2 E_{xtal} + (0.0035)^2 (E_{xtal})^2 + (0.167)^2$$

Toy calorimeter: 2-step approach



Identify $E_{xtal} > 0.5$ GeV



NxN candidate seed windows

1. Classifier

2. Regressor

Goal: simulated calorimeter windows → $(\boldsymbol{E_{\mathrm{reco}}}, \boldsymbol{i\phi_{\mathrm{reco}}}, \boldsymbol{i\eta_{\mathrm{reco}}})$ of each cluster

The network should be scalable to **arbitrarily large detectors** and **arbitrarily large number of seeds** (noise-dependent)

Showers are localised so we can create **7x7 seed windows** around each crystal with $E_{\mathrm{xtal}} > 0.5$ GeV (3σ of the noise level)

Each window is a candidate that may contain either a true cluster or background noise

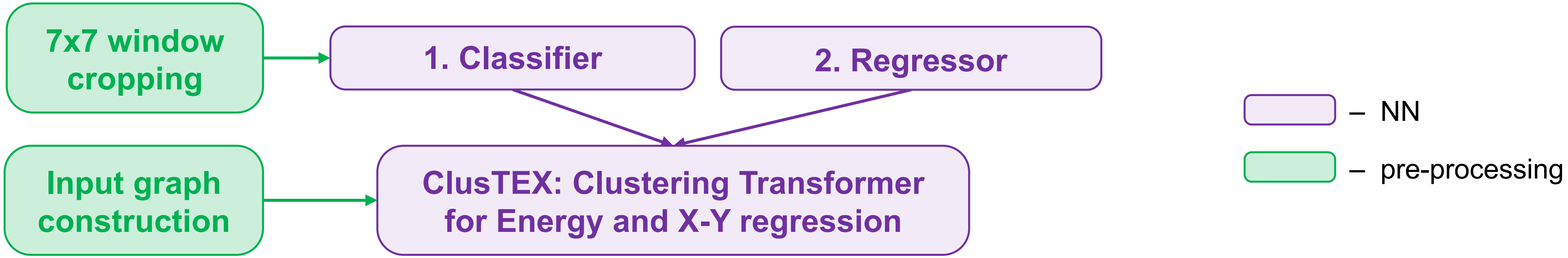
→ Common approach: dedicated classifier (signal/background) + regression network $(E_{\mathrm{reco}}, i\phi_{\mathrm{reco}}, i\eta_{\mathrm{reco}})$

Toy calorimeter: 1-step approach

Classification + regression in a single network

Propose an alternative approach [3] leveraging the power of transformers:

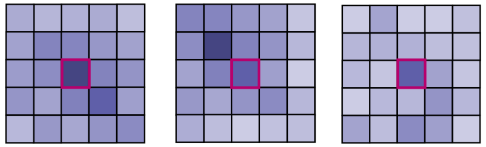
Collapse the classification and regression stages into a single inference step



See backup

ClusTEX is a graph-attention transformer: each 7x7 seed window is a graph node

Neighbouring windows show the same cluster: attention is a form of **message passing** that ensures information exchange between them



NxN candidate seed windows

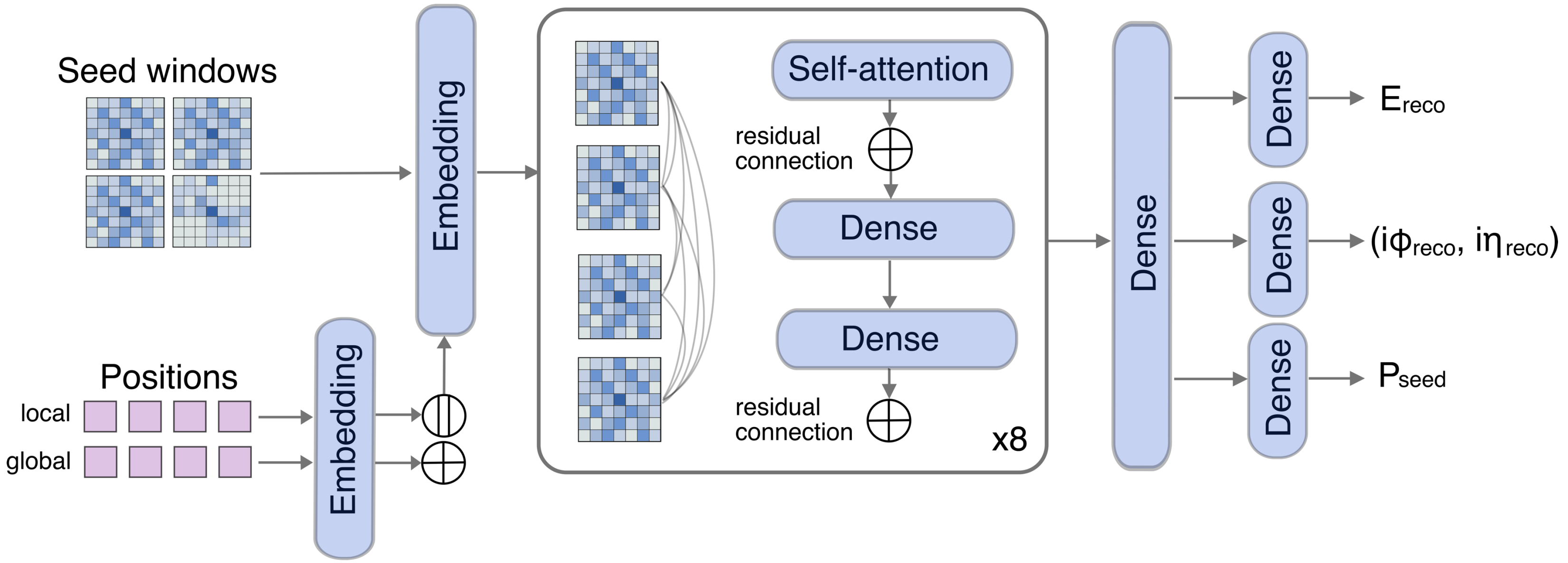
[3] *EPJC* 86, 873 (2026). 10.1140/epjc/s10052-026-16097-x

1-step approach: ClusTEX architecture

Transformer takes **graphs of 7x7 seed windows** as input and predicts the **energy and position of the cluster**

Architecture building blocks: encoding / attention / decoding

Joint reasoning among nodes to discriminate between **signal / background** and perform **kinematic reconstruction**



7x7 calorimeter windows
+ positional information

Input features

encoding / attention / decoding

$(E_{reco}, i\phi_{reco}, i\eta_{reco})$ of a cluster
+ signal/background discrimination

Output

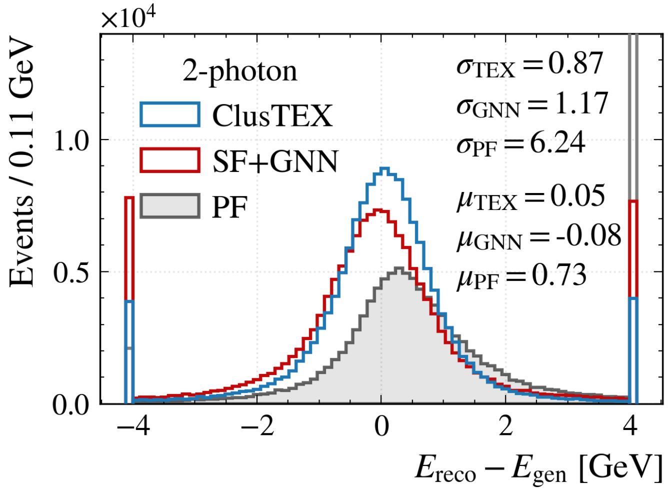
Toy calorimeter: results

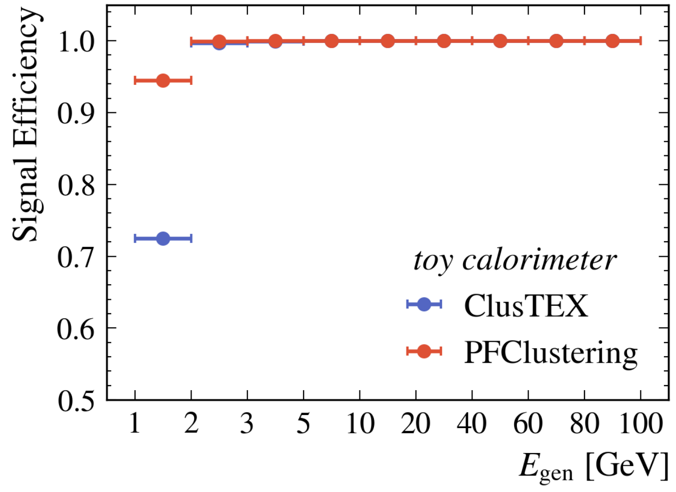
The single-step transformer (ClusTEX)

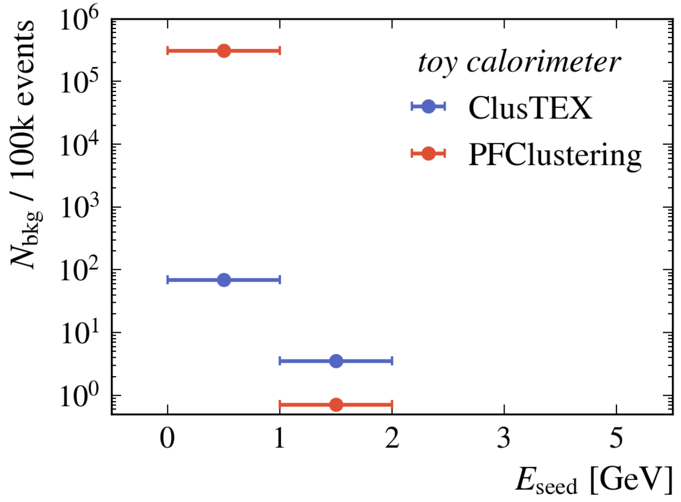
- provides a good performance on isolated showers
- outperforms PFClustering and other tested ML approaches* on the sample of overlapping showers
- has a low background yield

All while removing a dedicated classifier step

	1-photon sample		2-photon sample	
Metric	ClusTEX	PFClustering	ClusTEX	PFClustering
Efficiency (%)	99.7	99.8	98.7	82.0
σ_E (GeV)	0.55	0.59	0.87	6.24
$\sigma_{i\phi}$ (crystal)	0.03	0.05	0.04	0.09
$\sigma_{i\eta}$ (crystal)	0.04	0.06	0.04	0.09
Splitting (%)	0.08	–	0.06	–
N_{bkg}/100k samples	73	312k	99	311k







*Other tested approaches include a 2-step strategy with separate CNN classifier and regression networks, either based on fixed distance-weighted message passing or attention, see backup

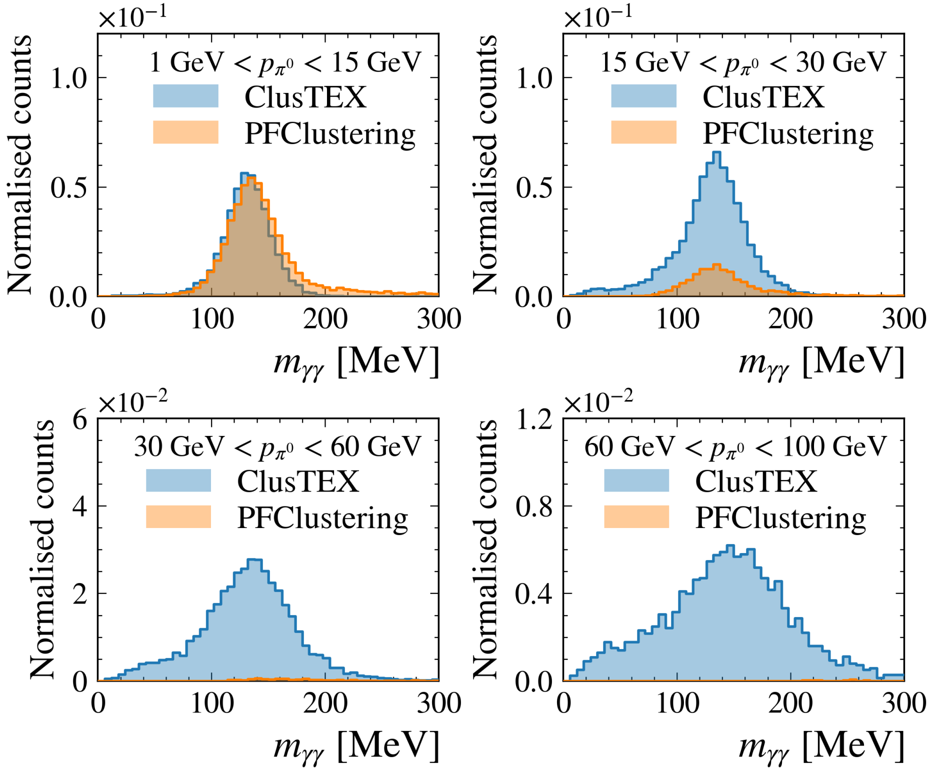
What do we gain from this improvement?

π^0 reconstruction example

$\pi^0 \to \gamma\gamma$ decay is simulated in the toy calorimeter and **used for evaluation only**

The energy distribution of π^0 is uniform in the 1-100 GeV range

Final-state γ become more collimated with increasing $|\mathbf{p}_{\pi^0}|$



$$m_{\gamma\gamma} = \sqrt{2E_1^{\text{reco}}E_2^{\text{reco}}(1-\cos\theta_{12})}$$

From the invariant mass reconstruction in four $|\mathbf{p}_{\pi^0}|$ bins we see that

ClusTEX significantly improves the reconstruction efficiency for $p_{\pi^0} > 15$ GeV

ClusTEX is now able to reconstruct pions with $p_{\pi^0} > 30$ GeV

This is inline with the improved performance (efficiency + energy + position) for overlapping showers

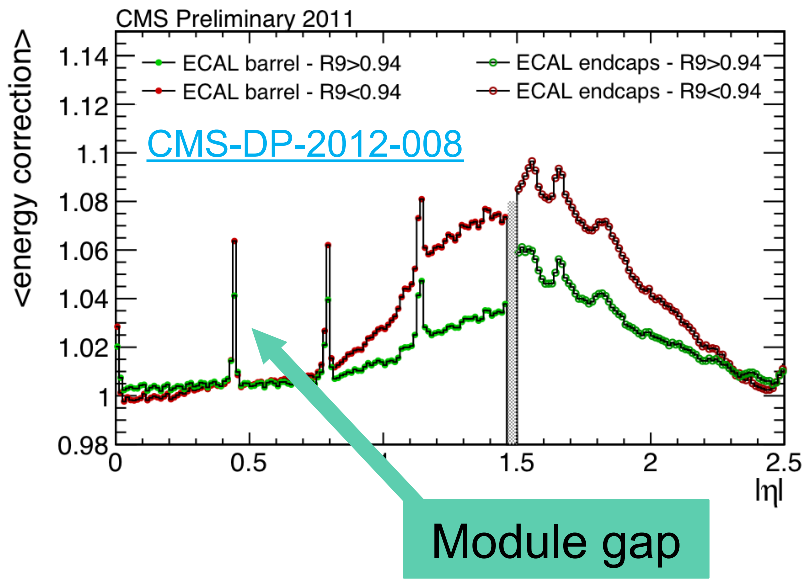
Moving to the full detector simulation

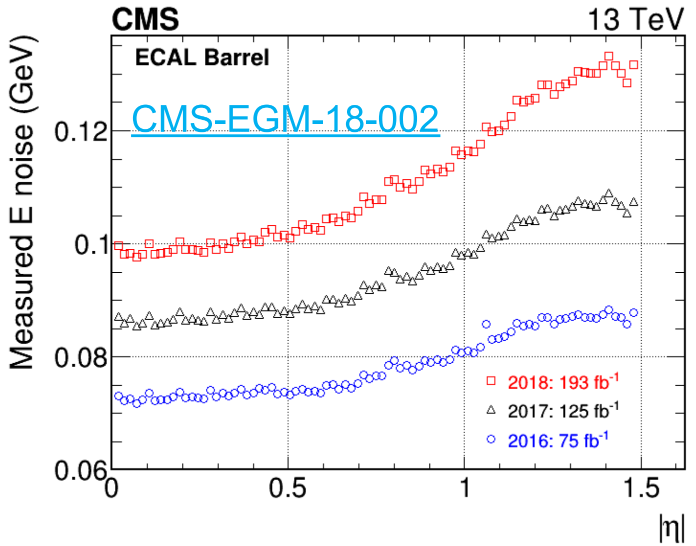
The toy calorimeter confirms the concept works. **Test the performance in the full detector simulation**

CMSSW (CMS Software) is the framework that contains the GEANT4 detector simulation and the full digisation+reconstruction chain from raw detector signals to the physics objects.

Full CMSSW simulation adds **new challenges** to the ECAL reconstruction:

- η-dependent noise due to irradiation
- Detector module gaps
- Readout electronics (causes dead/noisy channels)





Full detector simulation: sample generation and training

Dataset

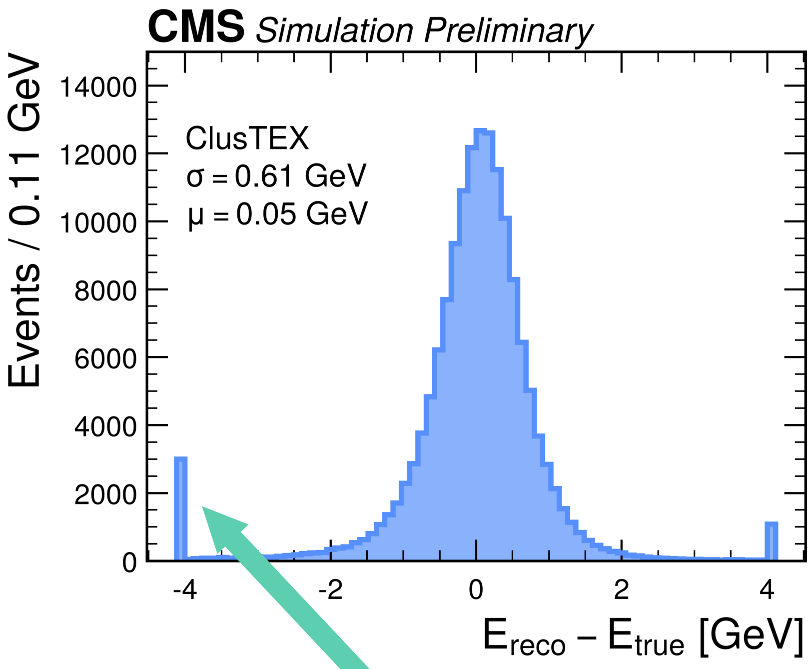
- Simulated using LHC Run 3 conditions. This includes noise but no pileup (subject of future work)
- Photon gun with uniform $(E_{\text{gen}}, \eta_{\text{gen}}, \phi_{\text{gen}})$ distributions with $1 < E_{\text{gen}} < 100$ GeV
- Select a subsample of **isolated photons** that did not convert before reaching the ECAL barrel surface (“unconverted”)
- **Simple to verify** and serves as a **direct comparison with PFClustering**

Training

- 1.6M events in the training sample, only single photons
- Same general architecture with 18M trainable parameters
- Additional input: status of each channel to indicate dead or noisy channels
- Lower noise threshold ($0.5 \rightarrow 0.23$ GeV) and more windows per graph ($8 \rightarrow 20$)

Full detector simulation: ClusTEX performance

- Inclusive result: $\sigma_E = 0.61$ GeV
- But $\sigma_E = 0.55$ GeV for events outside of the regions that are prone to energy loss (dead channels, module gaps) i.e. **ClusTEX recovers the nominal resolution from toy calorimeter simulation**



Mostly comes from detector module gaps

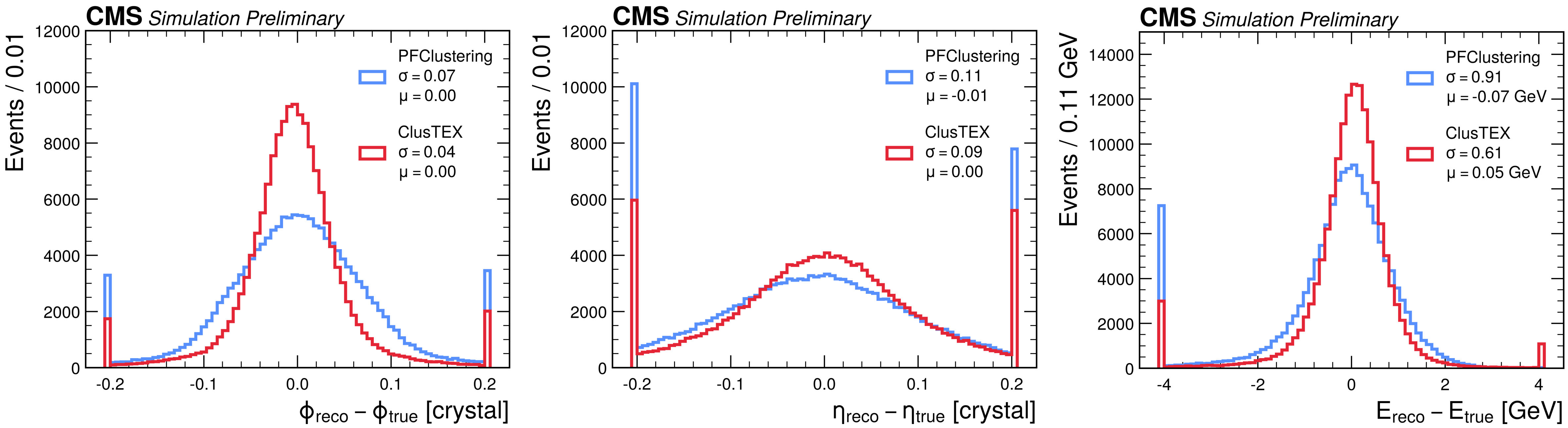
TOY CALORIMETER

Metric	1-photon sample	
	ClusTEX	PFClustering
Efficiency (%)	99.7	99.8
σ_E (GeV)	0.55	0.59

DPS note

Full detector simulation: comparison with PFClustering

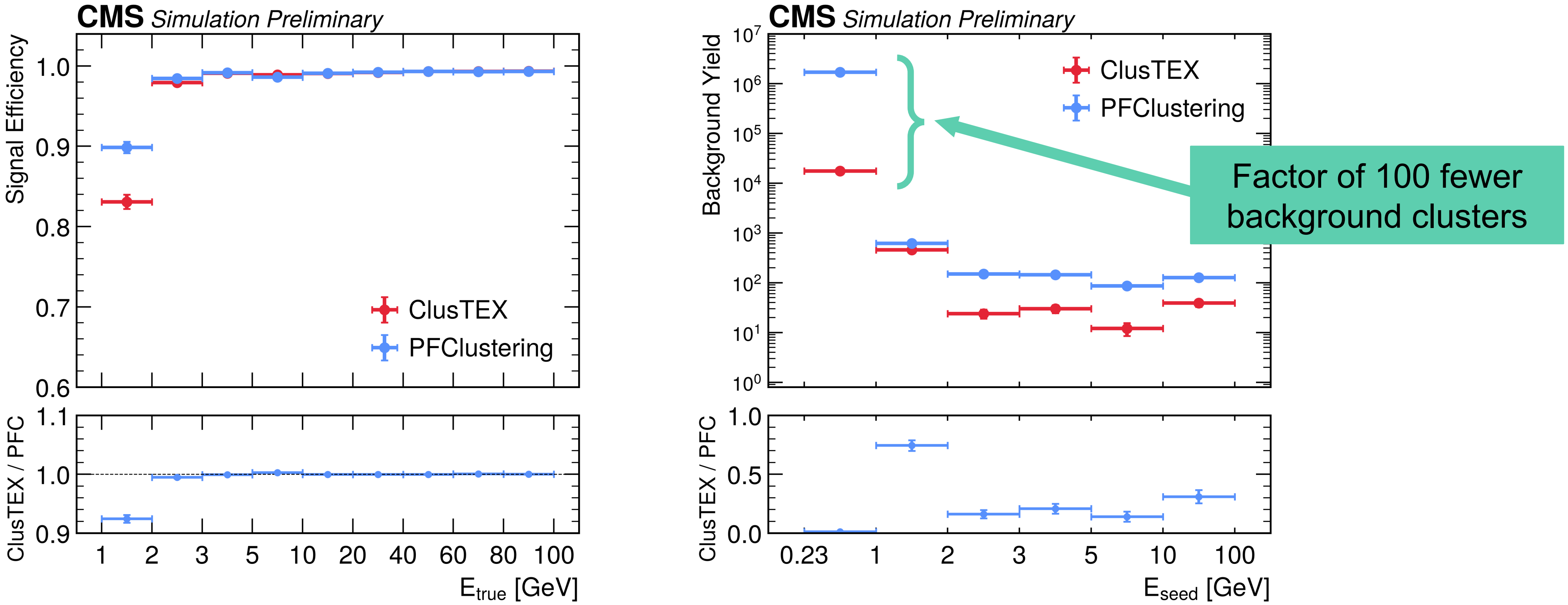
- $\sigma_{i\phi}$ and σ_E are much improved compared to the PFClustering baseline
- $\sigma_{i\eta}$ is better and $i\eta_{\mathrm{reco}} - i\eta_{\mathrm{true}}$ is more symmetric
- PFClustering σ_E degrades in the full sim: $\sigma_E = 0.79$ GeV in the absence of module gaps and dead crystals, compared to the nominal $\sigma_E = 0.59$ (toy calorimeter)



DPS note

Full detector simulation: efficiency and background yield

- ClusTEX has a **high signal efficiency**, comparable to PFClustering for $E_{\mathrm{true}} > 2$ GeV
- For $E_{\mathrm{true}} < 2$ GeV, the signal efficiency of ClusTEX is 7% lower but it produces **much fewer background clusters**, especially at the lowest energies

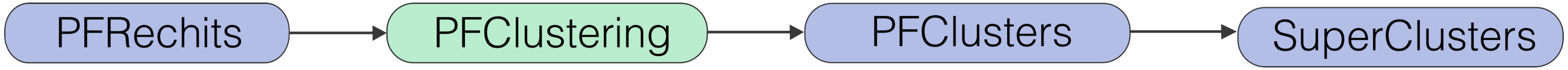


DPS note

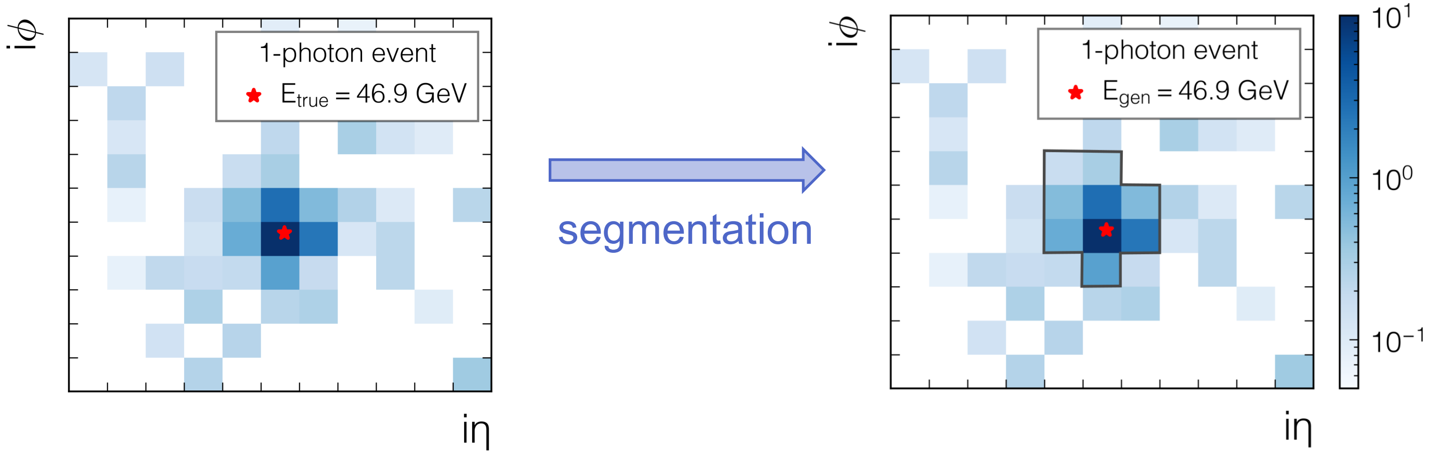
Segmentation

Current transformer output is $(E_{\text{reco}}, i\phi_{\text{reco}}, i\eta_{\text{reco}})$ of a cluster

Downstream reconstruction chain (superclustering) requires also a list of cluster hits



Add an output to the network that predicts the likelihood of each hit to belong to the corresponding cluster (**segmentation**)



Segmentation: definition

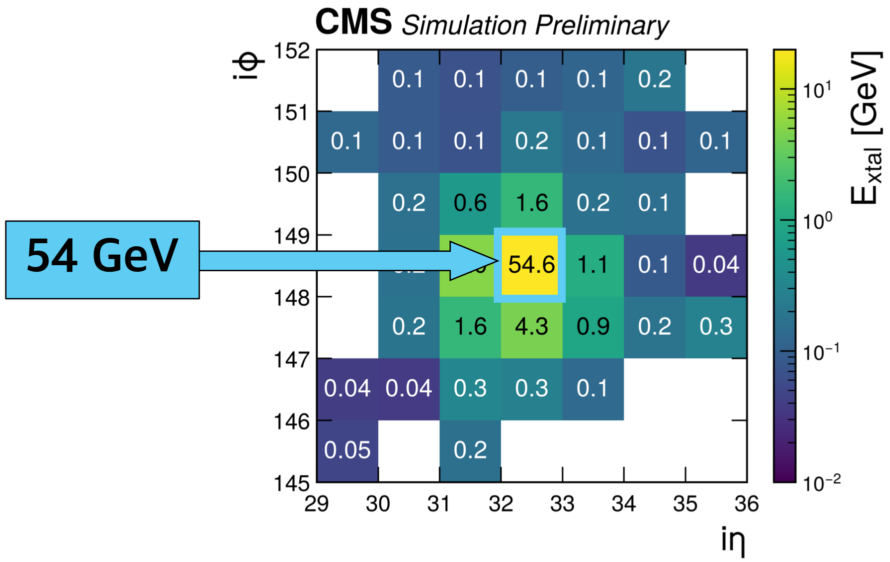
Each 7x7 seed window is either associated to a simulated particle or is considered as background (detector noise)

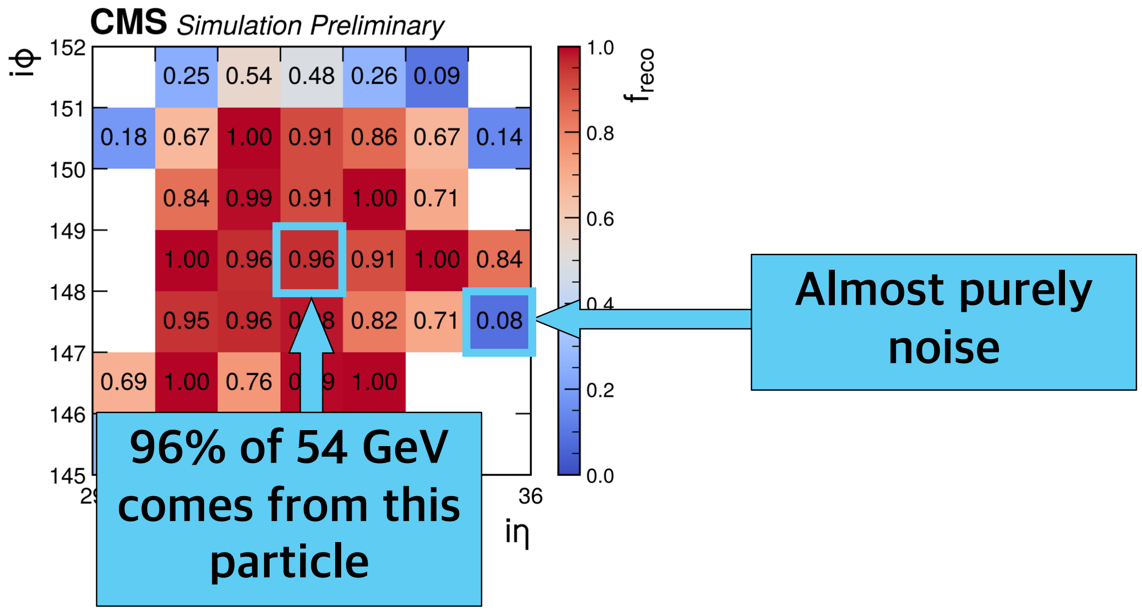
Define f_{true} for each cell of 7x7 window

f_{true} : the fraction of the cell's energy that belongs to the simulated particle

Since the simulated sample consists only of **single-photon events,** every cluster hit belongs to the same particle but each cell contains a non-zero **noise contribution**

e.g. $f_{\mathrm{true}} = 0.90$ means 90% signal and 10% noise



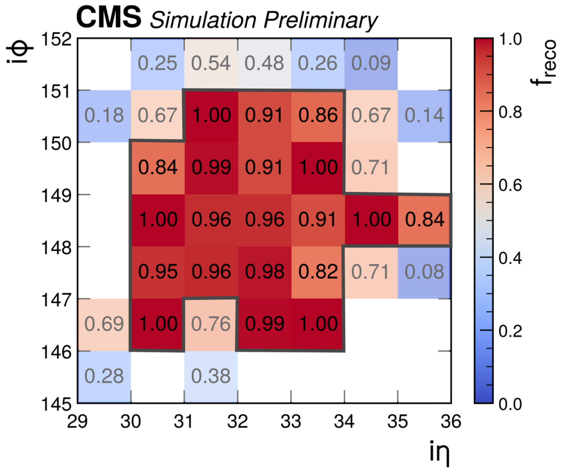


Segmentation: definition

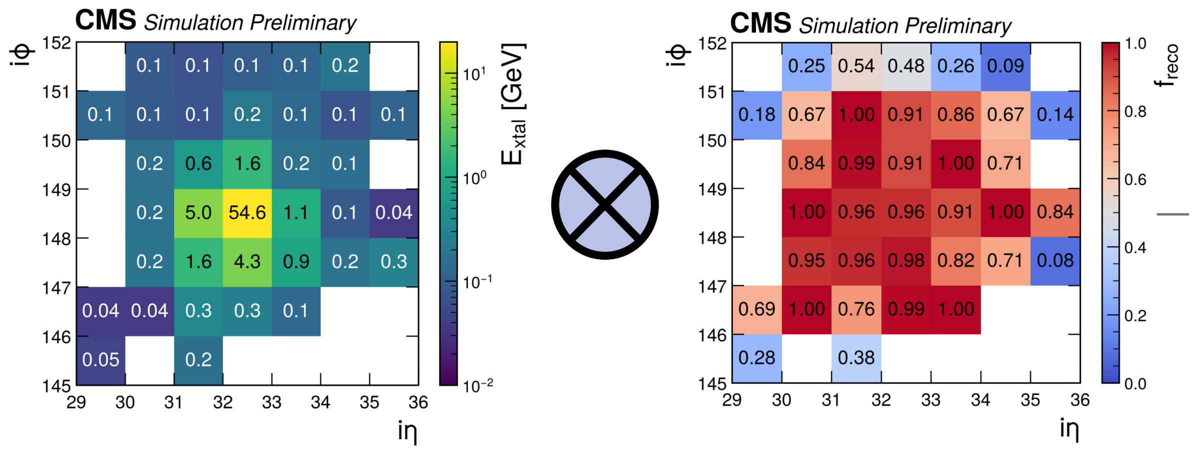
The goal of the segmentation task is to predict a fraction $f_{\mathbf{reco}}$ that is an estimate of $f_{\mathbf{true}}$

The value of f_{reco} supports **two interpretations**:

1. As a mask: thresholding f_{reco} gives **a list of hits for each cluster** required for downstream reconstruction

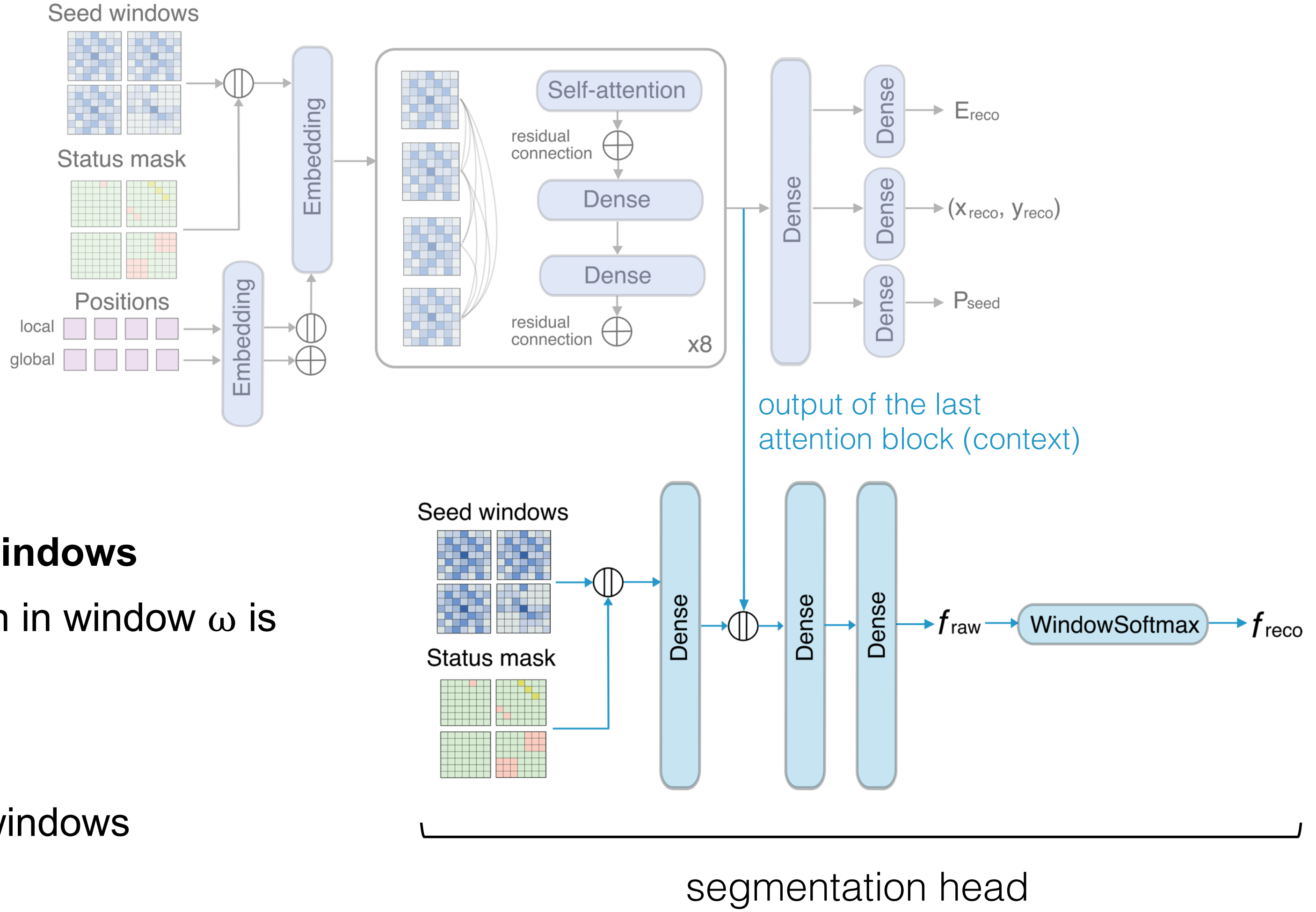


2. As a weight: multiplying f_{reco} by the map of energy deposits produces a **denoised cluster**



Segmentation: implementation

Freeze the weights of a trained model and train **a new segmentation head** (~200k weights)



WindowSoftmax: normalises f_{raw} across input windows

For a crystal i that appears in M windows, its fraction in window ω is

$$f_{i\omega} = \frac{e^{z_{i\omega}}}{\sum_{j=1}^{M} e^{z_{ij}} + e^{z_{\text{noise}}}}$$

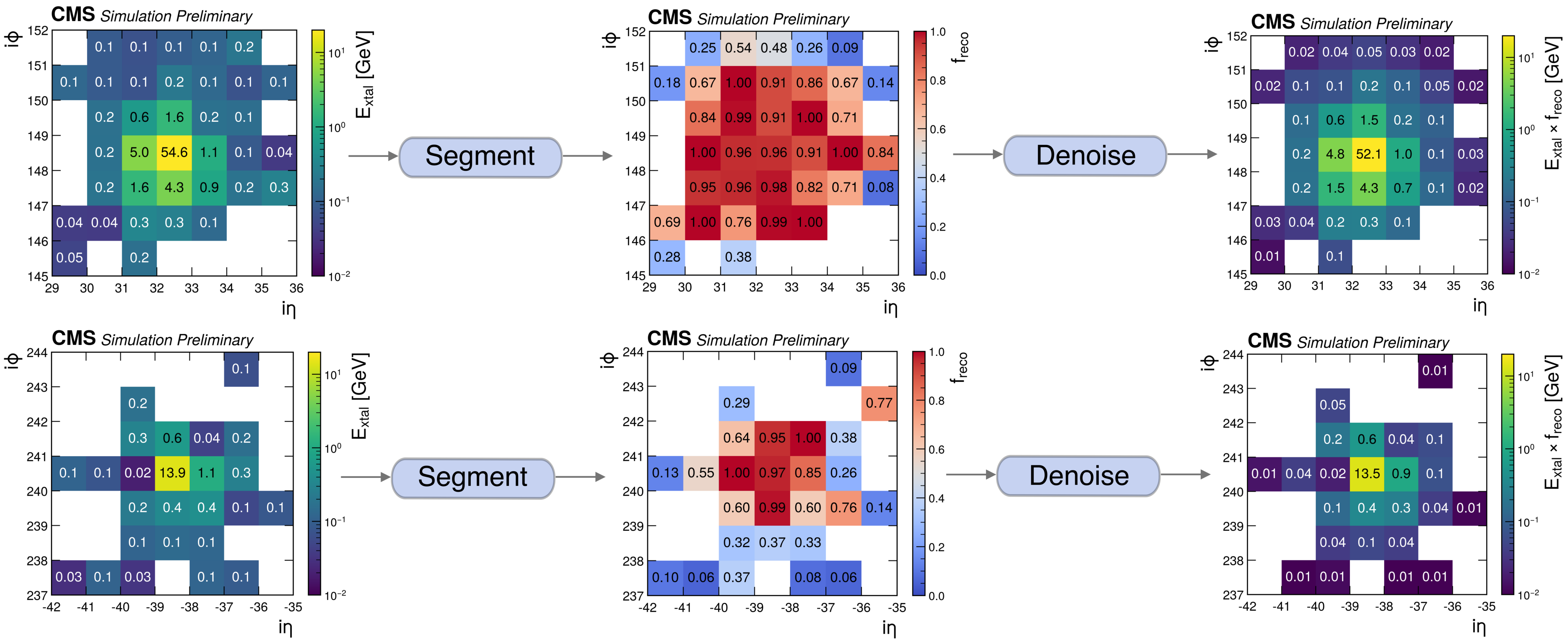
This ensures **proper energy sharing** between the windows

$$E_i = \sum_{j=1}^{M} E_{ij}$$

Segmentation: results

This small addition to the network is successful in learning the shower shape and size for clusters of all energies e.g. 72 GeV (top) and 17 GeV (bottom) clusters

The real test will be the energy sharing between overlapping clusters (to be continued)



Summary and next steps

Goal

Improve e/γ reconstruction for complex topologies (overlapping showers and detector effects)

Approach

- Toy calorimeter as a proof of concept → full detector simulation with LHC Run 3 conditions (without pileup)
- Graph-attention transformer for signal/background discrimination and $(E_{\text{reco}}, i\phi_{\text{reco}}, i\eta_{\text{reco}})$ regression, in one inference stage

Results

In the **toy calorimeter:**

- ClusTEX provides good performance on isolated photons
- ClusTEX significantly improves the reconstruction of overlapping showers compared to PFClustering

In the **full detector simulation:**

- ClusTEX shows promising results on isolated photons, improving resolution and background rejection

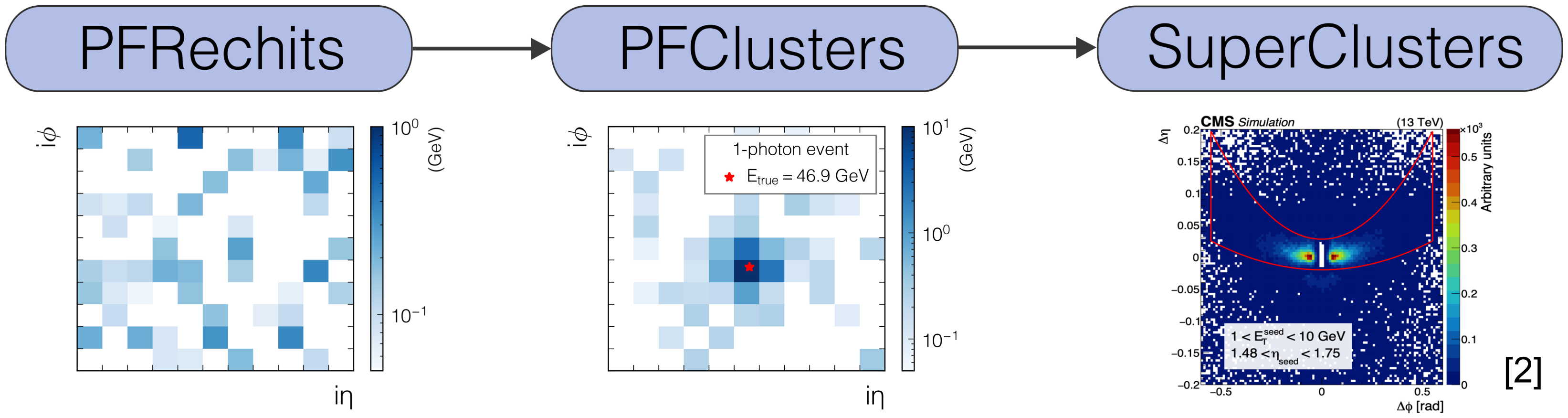
Outlook

- Ongoing developments in CMSSW following the promising preliminary results

Backup

e/γ reconstruction in CMS ECAL

[1] *JINST* 12 (2017) P10003. 10.1088/1748-0221/12/10/P10003
[2] *JINST* 16 (2021) P05014. 10.1088/1748-0221/16/05/P05014
[3] *J. Phys.: Conf. Ser.* (2023) 2438 012077. 10.1088/1742-6596/2438/1/012077



- **PFRechits:** Calibrated signal amplitudes recorded in individual ECAL crystals.

- **PFClusters**: e/γ passing through ECAL deposit energy in cluster-like formations around their impact position. The goal of a **clustering algorithm**, is to gather the signal from nearby crystals and interpret them as a single particle. The standard clustering algorithm is **PFClustering** [1], which builds topological clusters around local energy maxima.

- **SuperClustering**: Interactions that occur in the upstream tracker material create secondary ECAL clusters around the primary particle of interest. Each secondary cluster in itself is a PFCluster. A superclustering algorithm gathers PFClusters coming from the same primary particle and provides an improved estimate of the energy resolution. The standard superclustering algorithm is **Moustache** [2] but there exists an ML-based alternative [3] considered for Phase-2

Introduction: standard clustering algorithm

Reconstruction of e/γ in ECAL relies on a **clustering algorithm** that interprets a cluster of calorimeter hits as a single particle

The standard algorithm used by CMS in the ECAL barrel is **PFClustering** [1]

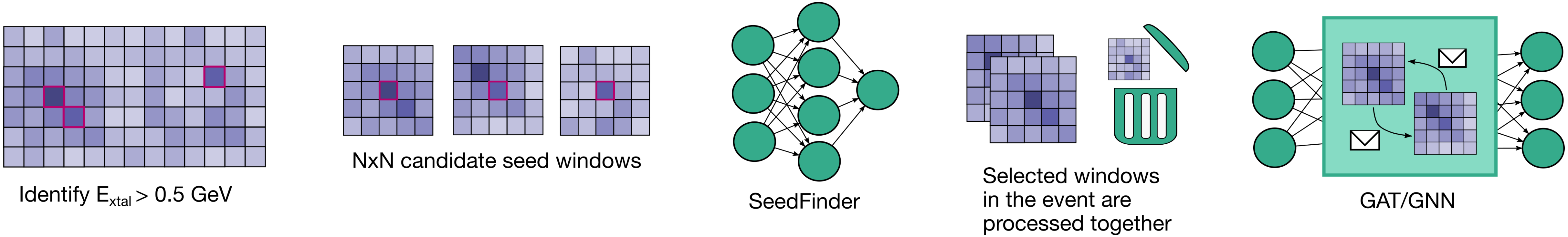
P. Simkina (2023). Event reconstruction and analysis in CMS using artificial intelligence. HAL

1. Identify crystals with measured energy > $\mathrm{E}_{\mathrm{threshold}}^{\mathrm{seed}}$. Every local maximum becomes a **cluster seed**
2. Starting from the highest-energy seed, grow a **topological cluster** by gathering neighbouring crystals > $\mathrm{E}_{\mathrm{threshold}}^{\mathrm{gather}}$
3. Use an **iterative algorithm** based on a Gaussian-mixture model to reconstruct clusters within a topological cluster

1 seed = 1 cluster

[1] *JINST* 12 (2017) P10003. 10.1088/1748-0221/12/10/P10003

Toy calorimeter: 2-step approach



Goal: simulated calorimeter windows → $(\mathbf{E_{reco}}, \mathbf{x_{reco}}, \mathbf{y_{reco}})$ of each cluster

1. Classifier

CNN classifier between signal (photon) and background (noise)

2. Regressor

Predicts $(E_{\mathrm{reco}}, x_{\mathrm{reco}}, y_{\mathrm{reco}})$ from selected windows

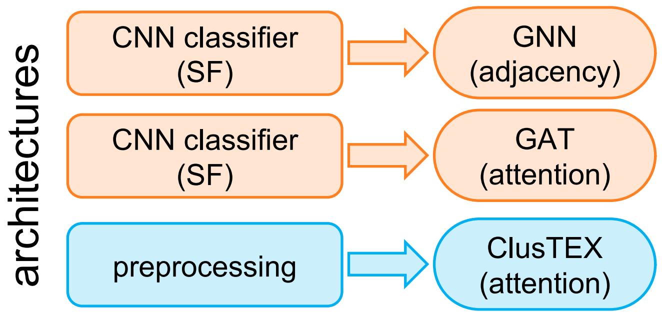
Neighbouring windows show the same cluster: implement **message passing** to ensure information exchange between them

Consider two schemes for message passing:

- adjacency matrix of inverse Euclidean distances between windows centers [1]
- scaled dot product attention [2]

[1] *EPJC* 84, 639 (2024). 10.1140/epjc/s10052-024-12978-1
[2] *EPJC* 86, 873 (2026). 10.1140/epjc/s10052-026-16097-x

Toy calorimeter: results



	1-photon sample			
Metric	ClusTEX	SF+GAT	SF+GNN	PFClustering
Efficiency (%)	99.7	99.8	99.8	99.8
σ_E (GeV)	0.55	0.58	0.57	0.59
$\sigma_{i\phi}$ (crystal)	0.03	0.03	0.03	0.05
$\sigma_{i\eta}$ (crystal)	0.04	0.04	0.04	0.06
Splitting (%)	0.08	0.17	0.30	–
N_{bkg}/100k samples	73	26	102	312k

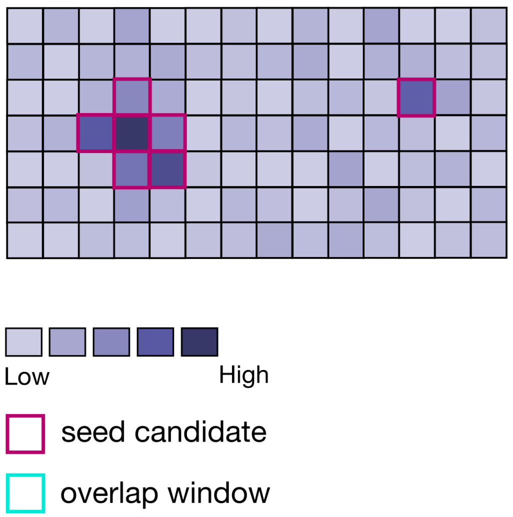
	2-photon sample			
Metric	ClusTEX	SF+GAT	SF+GNN	PFClustering
Efficiency (%)	98.7	98.3	98.3	82.0
σ_E (GeV)	0.87	1.10	1.17	6.24
$\sigma_{i\phi}$ (crystal)	0.04	0.04	0.04	0.09
$\sigma_{i\eta}$ (crystal)	0.04	0.04	0.04	0.09
Splitting (%)	0.06	0.31	0.59	–
N_{bkg}/100k samples	99	84	178	311k

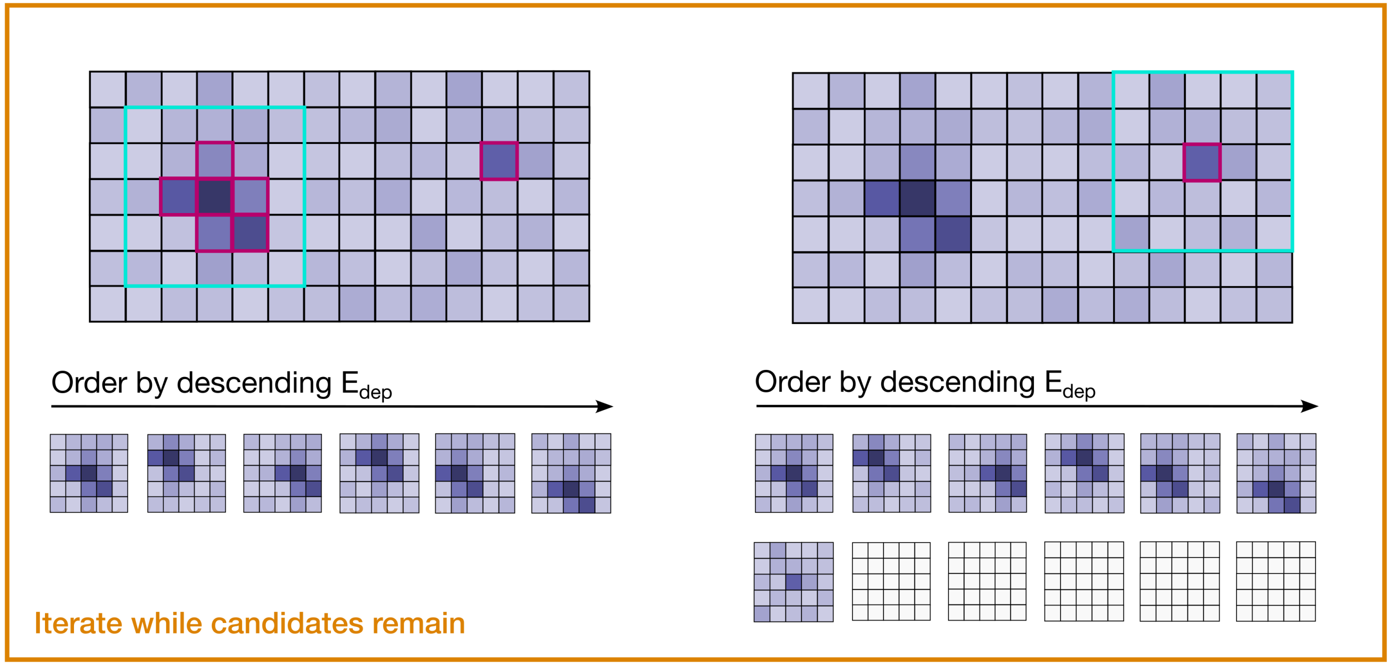
1-step

2-step attention/GNN

Baseline

1-step approach: graph construction

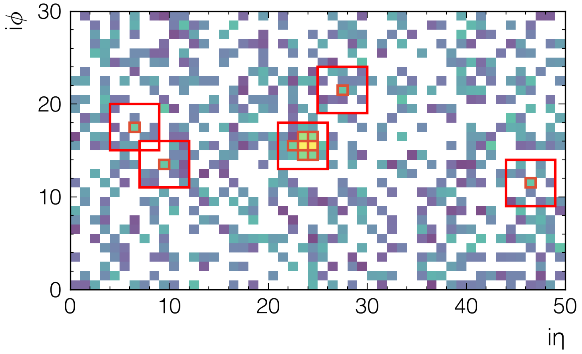




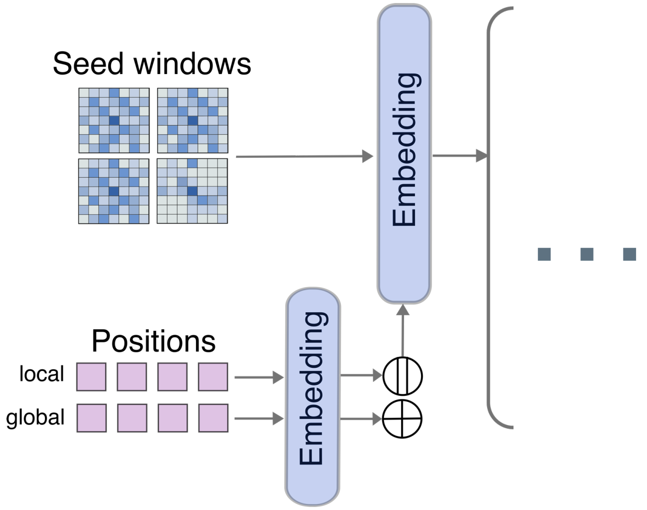
Each graph is an independent network input

1. Identify cells with E>0.5 GeV as **seed candidates**. Order them by descending energy
2. Select the highest-energy seed as an **anchor** and build an **overlap window** around it
3. For each seed in the overlap window, extract a $d_{crop} \times d_{crop}$ **seed window**
4. Eliminate processed seeds from candidate list. Iterate until none remain

Example: real calorimeter window



ClusTEX: positional encoding scheme



Real ECAL: non-trivial geometry, η-dependent irradiation, detector module gaps, dead cells, ...

Desired feature: geometrical awareness injected with a positional encoding block

Each input seed window i is characterised by its

- **Local position** with respect to the anchor of the graph $\mathbf{x}^i_{\mathrm{local}} = \left(\frac{x_i - x_a}{d_{\mathrm{effective}}}, \frac{y_i - y_a}{d_{\mathrm{effective}}}\right)$
- **Global position** within ECAL (flattened) $x^i_{\mathrm{global}} = x_i + y_i \cdot S_{\mathrm{detector}}$

(x_a, y_a) : graph anchor

$d_{\mathrm{effective}} = 13$: effective size of the graph

$S_{\mathrm{detector}} = 85$: number of crystals in $i\eta$

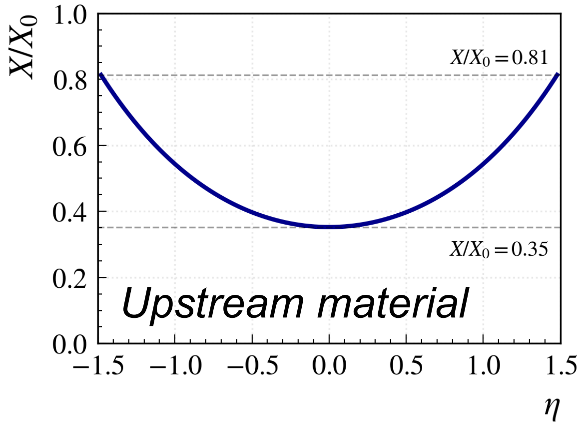
- Embeddings of $\mathbf{x}^i_{\mathrm{local}}$ are **concatenated** with input seed window embeddings $\boldsymbol{h}_i$ $\tilde{\boldsymbol{h}}_i = \boldsymbol{h}_i \mid \boldsymbol{f}(\mathbf{x}^i_{\mathrm{local}})$
- Embeddings of x^i_{global} are injected with a **summation** $\boldsymbol{z}_i = \tilde{\boldsymbol{h}}_i + \boldsymbol{g}(x^i_{\mathrm{global}})$

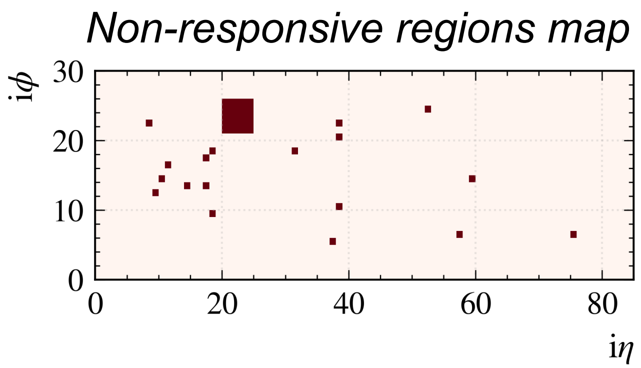
This separation allows the network to remain sensitive to the relative structure of candidates in the graph while simultaneously learning detector-position dependencies

ClusTEX: positional encoding scheme

Toy calorimeter already has a shower shape variation with η from the geometry and a 33-mm slab of silicon upstream

Test the positional encoding scheme by **introducing also non-responsive channels** to the simulation



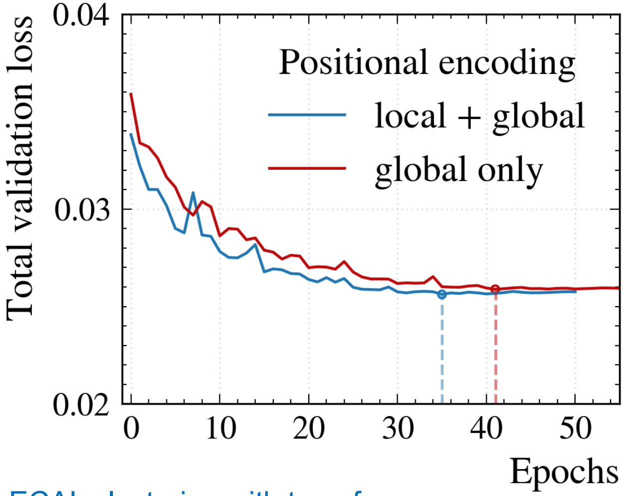


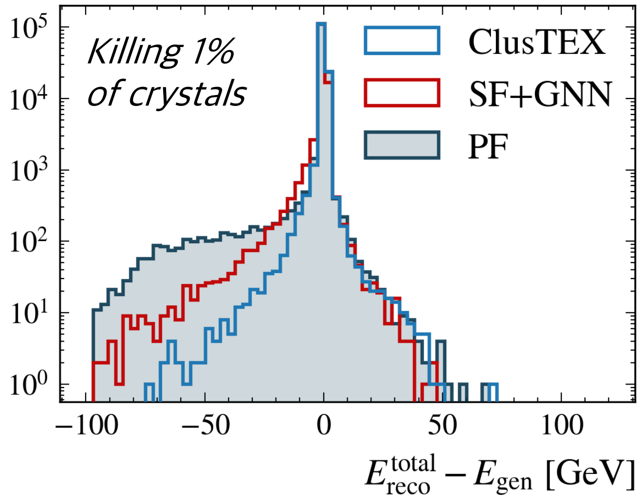
~1% of individual channels + one 5x5 matrix

Training/validation/testing samples have the same mapping
(alternative is to add it as an input feature)

In the presence of non-responsive channels:

- **Local+global positional encoding is more effective** than global-only encoding
- **ClusTEX proves to be a robust approach** recovering the missing energy better than PFClustering and SF+GNN





Formal definition of evaluation metrics

X resolution (σ_X)	Half-width of the central 68% interval of the corresponding $\mathrm{X_{reco}} - \mathrm{X_{true}}$ distribution from the percentile differences
True-reco matching metric	$r_{\mathrm{match}} = \sqrt{\left(\frac{i\phi_{\mathrm{reco}} - i\phi_{\mathrm{true}}}{\sigma_{i\phi}}\right)^2 + \left(\frac{i\eta_{\mathrm{reco}} - i\eta_{\mathrm{true}}}{\sigma_{i\eta}}\right)^2 + \left(\frac{E_{\mathrm{reco}} - E_{\mathrm{true}}}{\sigma_E}\right)^2}$
Signal efficiency	The fraction of gen particles matched to a reconstructed object that satisfies $\mathrm{P_{seed}} > \mathrm{P_{seed}^{threshold}}$ and $\Delta\mathrm{R} = \sqrt{(\mathrm{i}\phi_{\mathrm{reco}} - \mathrm{i}\phi_{\mathrm{true}})^2 + (\mathrm{i}\eta_{\mathrm{reco}} - \mathrm{i}\eta_{\mathrm{true}})^2} < \mathrm{R_{threshold}} = 1.5$ $\mathrm{P_{seed}^{threshold}}$ is model-dependent, ~0.3-0.35
Background yield	The number of reconstructed objects that satisfy $\mathrm{P_{seed}} > \mathrm{P_{seed}^{threshold}}$ and $\Delta\mathrm{R} > 1.5$
Splitting	The fraction of gen particles that are matched to >1 reconstructed objects