NATIONAL RESEARCH UNIVERSITY

LAMBDA • HSE

Evolution of reconstruction for Upgrade II

ECAL Upgrade II Workshop @IJCLab, 12 December 2022

Alexey Boldyrev[1], Denis Derkach[1], Fedor Ratnikov[1,2], Andrey Shevelev[1]

1 — HSE University (Laboratory of Methods for Big Data Analysis); 2 — Yandex School of Data Analysis;

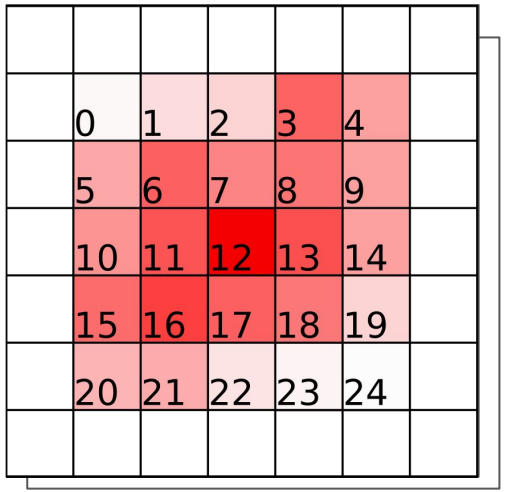
Motivation

- The conventional (parametric) Reco performs very well on Run1-2 data
- We assume that this might not work at high-pileup conditions
- Using additional information (5x5 clusters) and ML regressors may enhance the performance at high pile-up
- ML-based Reco based on 3 sets of regressors to estimate:
  - Position
  - Energy
  - Time
- Training using single photons sample (single readout and FTDR for ECAL Run 5)
- Repeati training on the reference physics sample (at given pileup conditions) $B_s^0 \to J/\psi(\to \mu^+\mu^-)\pi^0(\to \gamma\gamma)$

After clustering

- Search for local maxima in 3x3 ECAL cells
  - Define seed
- Consider 5x5 ECAL cells around the seed
- Recalculate hit and barycenter positions w.r.t the center of the seed
- Until converge:
  - Define a projection of the hit position in a depth (Z-direction)
  - Find the minimum of RMSE (3D) of the hit and barycenter
- Train the regressor to minimise spatial resolution

Regressor's raw features



5 cells * 5 cells * 2 layers = 50 raw features

50 raw features of energy deposits for:

- Spatial regressor
- Energy regressor

50 raw features of energy deposits &
50 raw features of timing in cells:

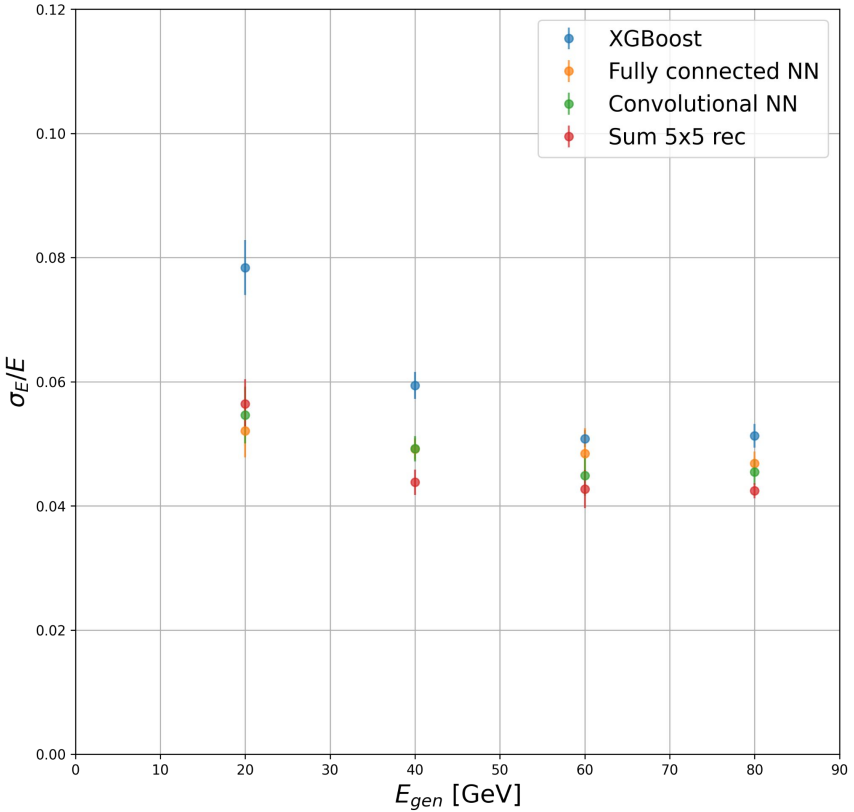
- Timing regressor

Regressors performance: energy resolution

Number of ML regressors were compared:

- XGBoost
- Fully-Connected NN
- Convolutional NN

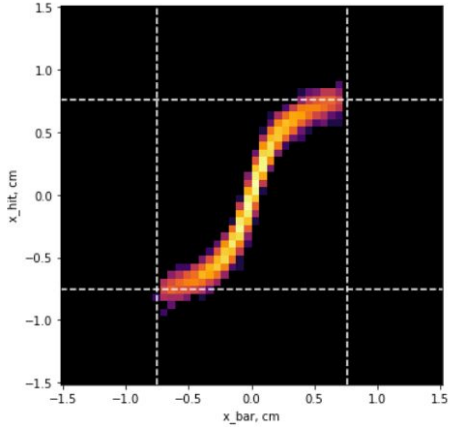
Total sum of energy deposited in 5x5(x2) ECAL cells was used as a reference.

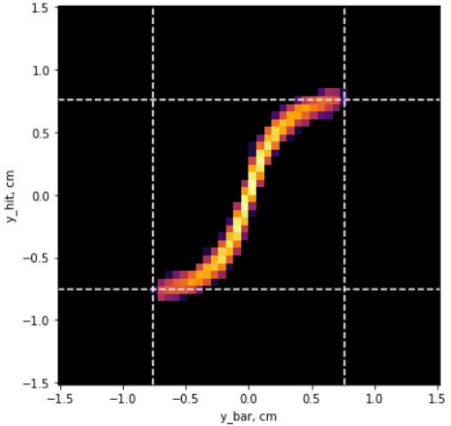


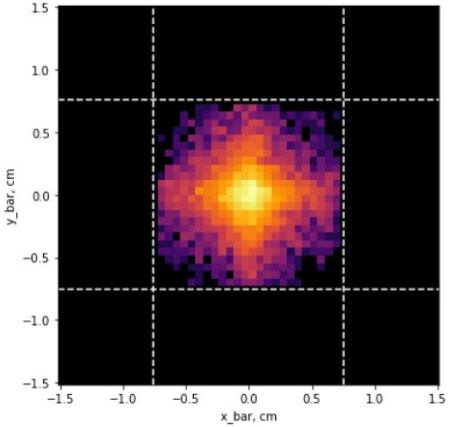
Regressors performance: spatial resolution

LAMBDA • HSE

Barycenter -based

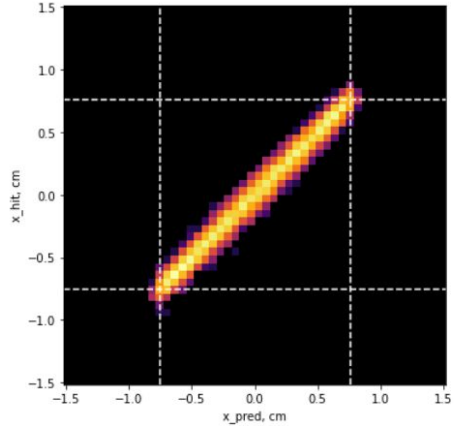


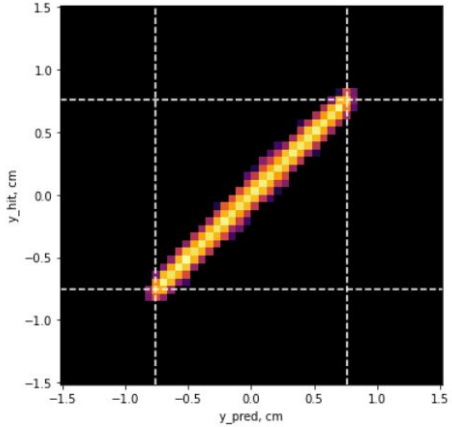


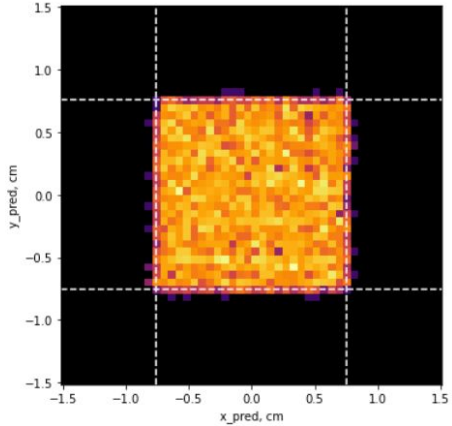


Spacal W/GAGG (1.515x1.515 cm^2 cells)

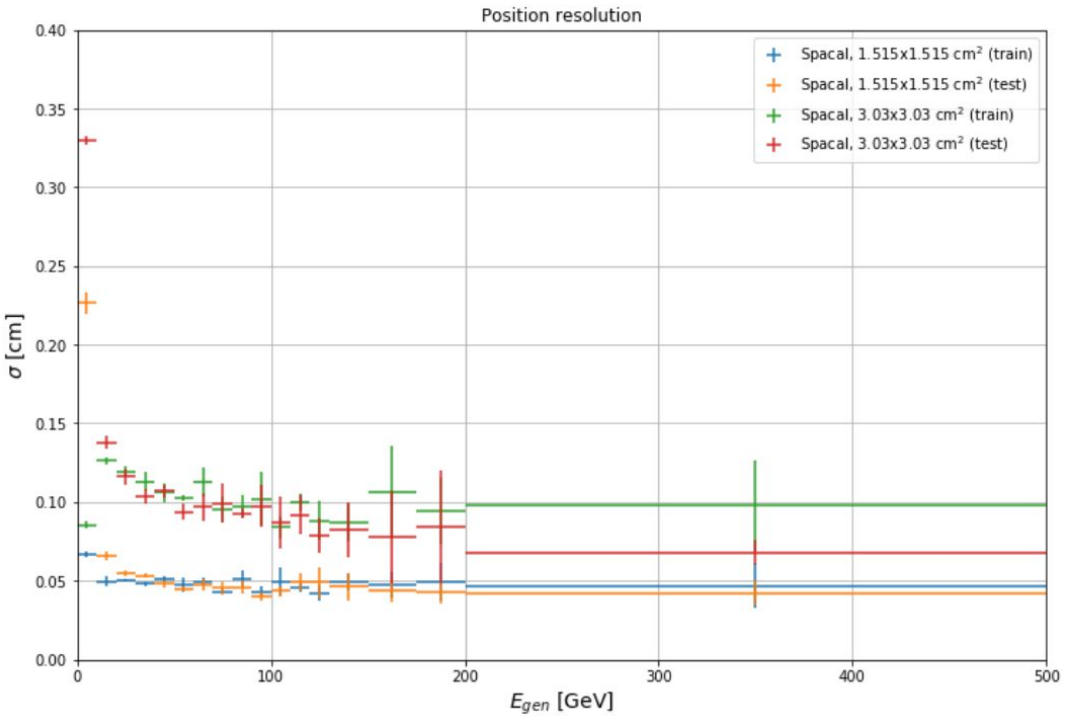
Predicted by the regressor







Regressors performance: Spatial resolution



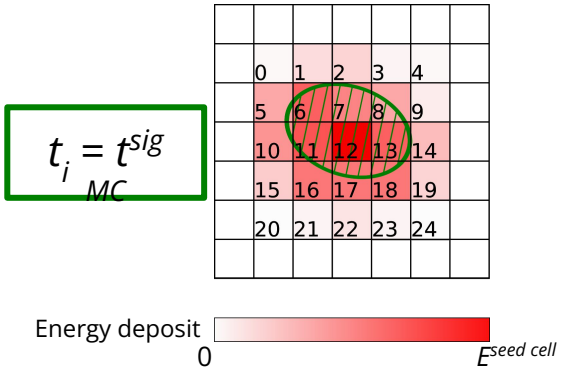
- Spacal modules W/GAGG
- Separate regressors on (Rec - Gen) RMSE for x & y positions of the hit
- Consider 5x5(x2) cells as features for the regressors

Spatial resolutions for the Spacal W/GAGG and Pb/Polystyrene modules are flat above 50 GeV
Stat. uncertainties are low.

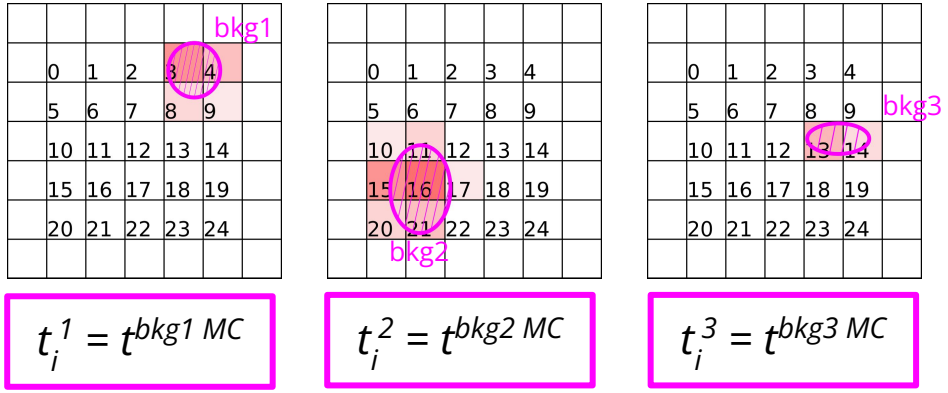
Approach for timing information

Suppose that there are 3 background contributions in the 5x5 cells vicinity of the seed cell and $t^{bkg1} < t^{sig} < t^{bkg2} < t^{bkg3}$.

Signal energy deposits and shower spot



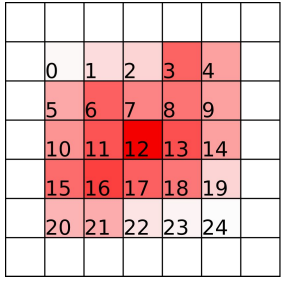
Background energy deposits and showers' spots



Resulting energy deposits:

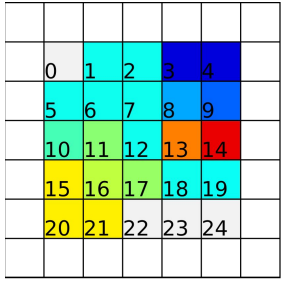
$E_i = E_i^{sig} + E_i^{bkg1} + E_i^{bkg2} + E_i^{bkg3}$

(Used as raw features for position and energy regressors)

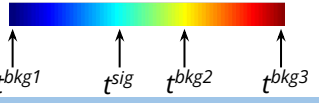


Resulting time of cells:

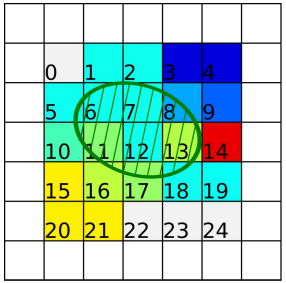
$$t_i = \frac{\sum_{i,j} t_i^j E_i^j}{\sum_{i,j} E_i^j}$$



Calculated cell time

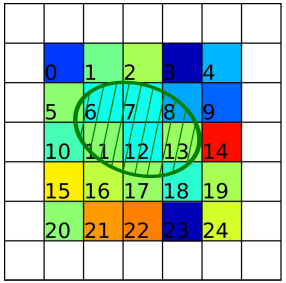


Regressors performance: timing resolution

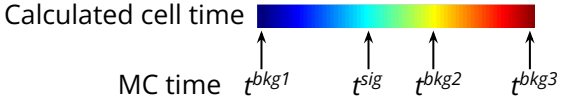


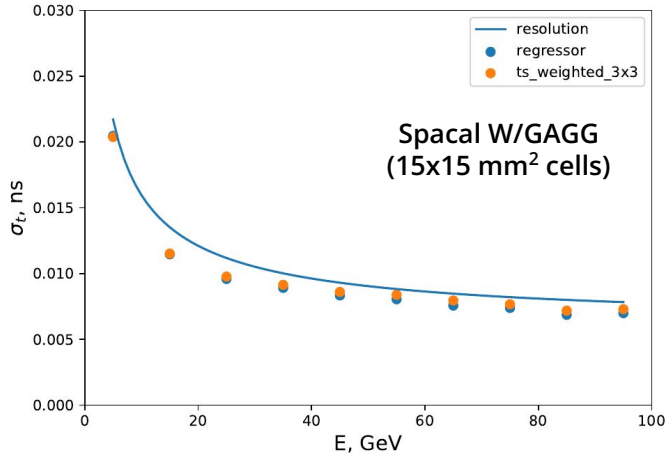
apply module time resolution

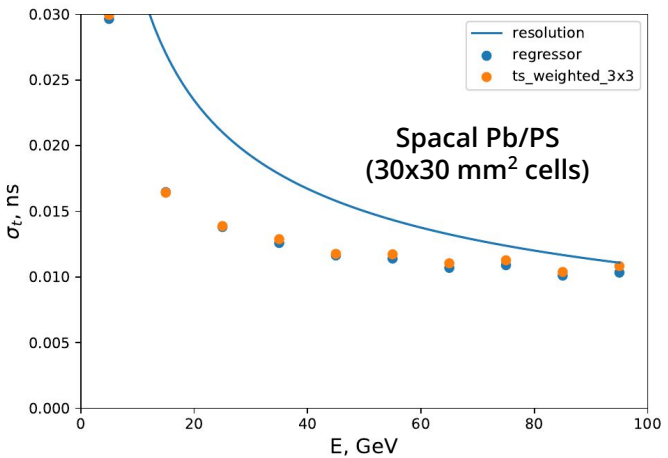
$$\sigma_t = A/\sqrt{E} \oplus B$$



(Used as raw features for time regressor)





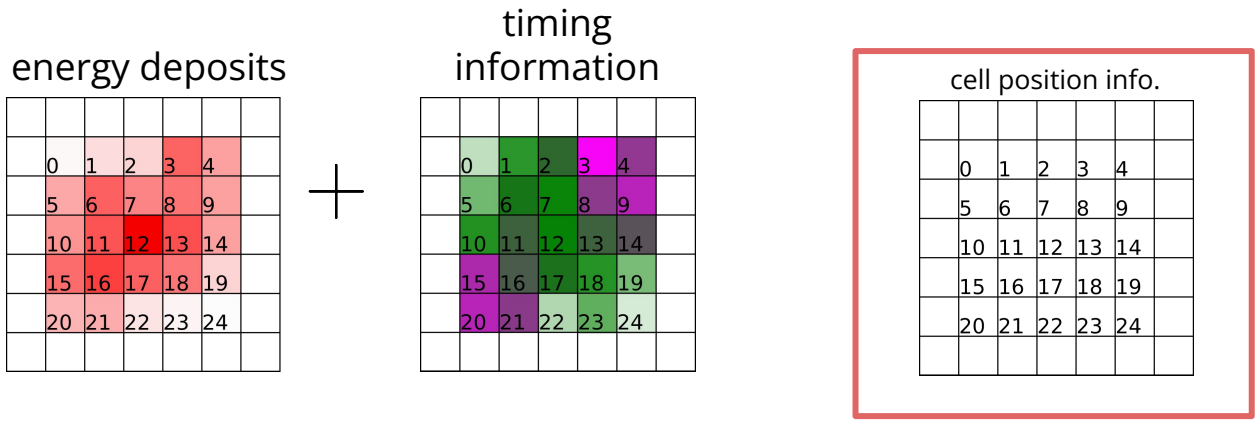


What we have so far

- Single ECAL module:
  - ML reco performance is compatible with conventional Reco performance using detailed simulation inputs
- Full ECAL
  - Requires geometrical irregularities
    - There are 4 borders between the regions of different granularity
    - Some modules have to be rotated due to technology limitations

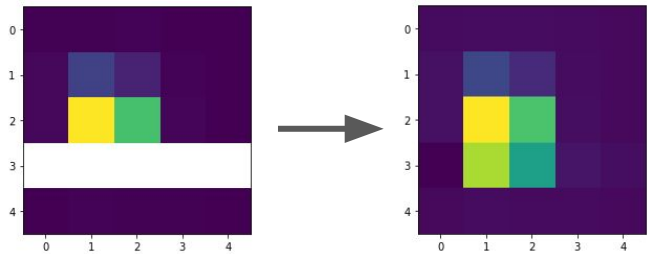
How does that fit in with the fact that reco algorithm needs to be geometry agnostic?

Strategies to have geometry agnostic inputs



Better strategy:

- Cell position matrix as addition input to ML regressors
- Interpolation of non-existing energy deposits in missing ‘virtual’ cells
- Interpolation of cells for equalization of granularity on both sides of the border



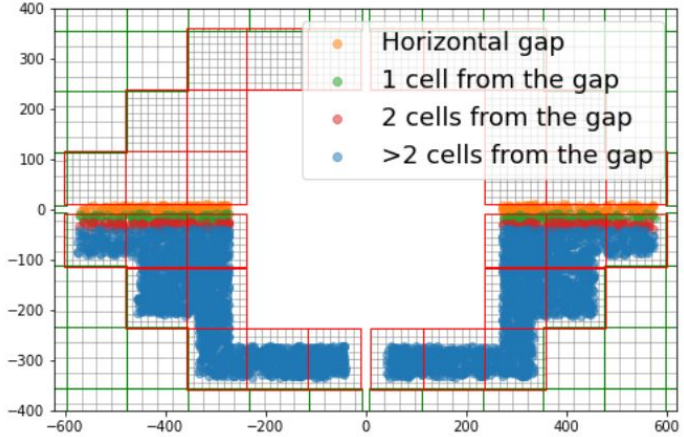
Energy reconstruction on geometry agnostic inputs

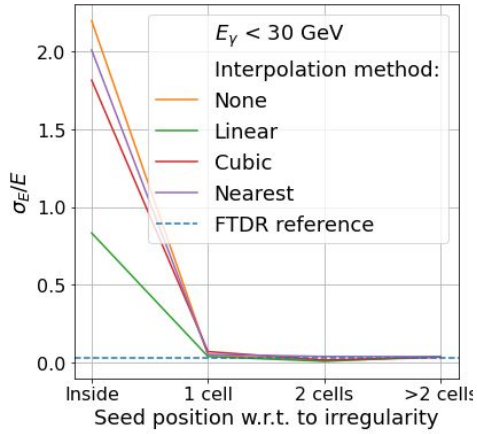
XGB-based ML regressor

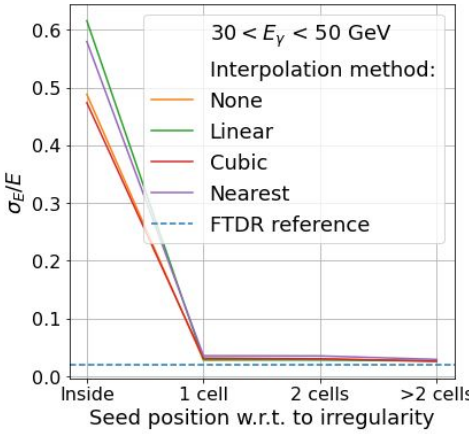
- 5x5 matrix of energy deposits
  - Missing cells recovered using

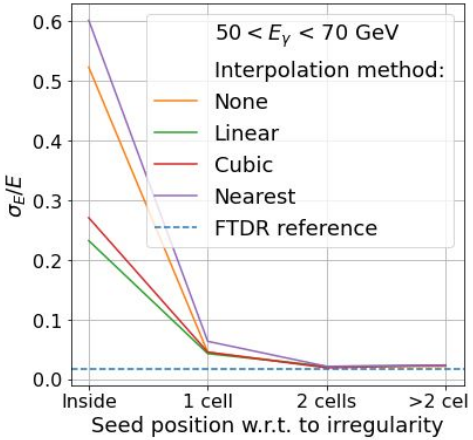
Linear, Cubic and Nearest-neighbor interpolation

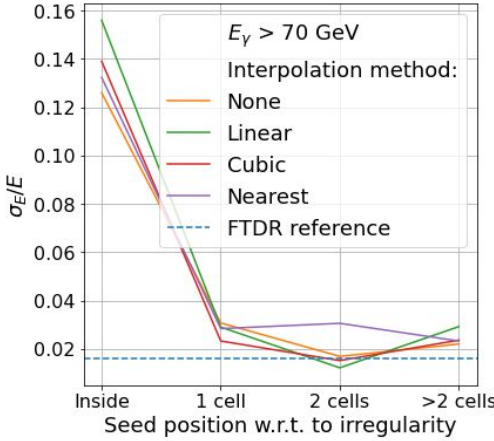
- Cell position info
- Additional features











Conclusions

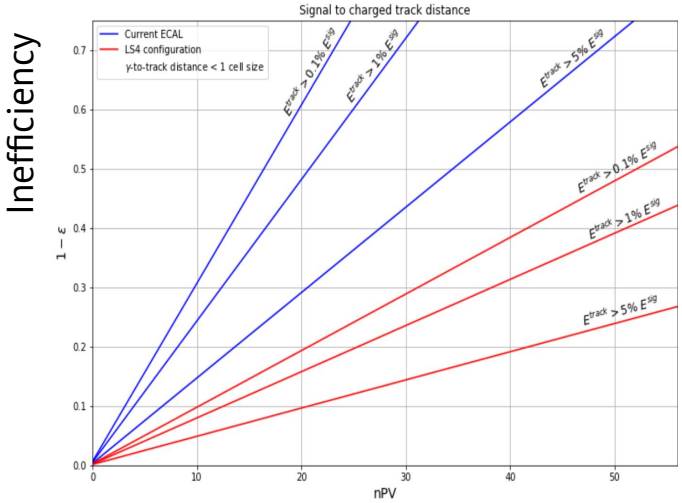
- The R&D process requires time consuming computation steps to evaluate physics performance for different detector techniques and configurations.
- ML reco is consistent with conventional reconstruction for single ECAL module and regular geometry
- ML reco is able to handle detailed simulation inputs using cell position matrix, interpolation of missing cells, and interpolation of low granularity cells close to high granularity cells
- Automatic training speeds up the turnover for the performance studies and ensures consistency and uniformity of obtained results

LAMBDA • HSE

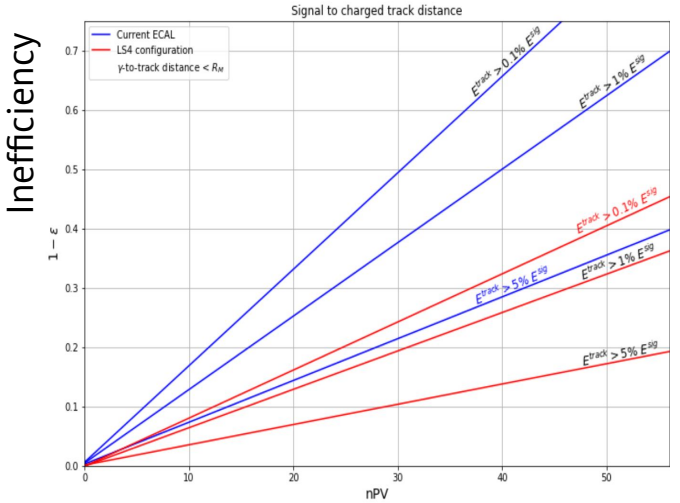
Backup slides

Charged tracks veto

Photon candidate requires no charged track nearby



distance < 1 cell size



distance < 1 Molière radius

1: Outer region, cell size = 12.12 x 12.12 cm^2
2: Middle region, cell size = 6.06 x 6.06 cm^2
3: Inner region, cell size = 4.04 x 4.04 cm^2

} R_M = 35.0 mm

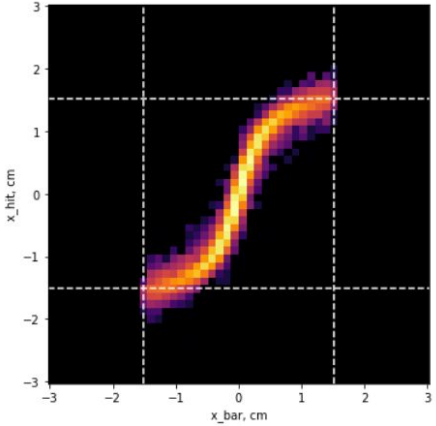
4: cell size = 3.03 x 3.03 cm^2 R_M = 14.5 mm

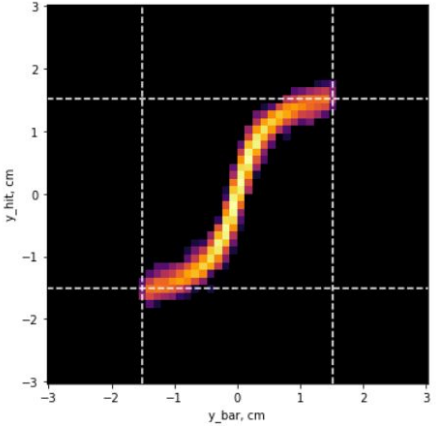
5: cell size = 1.515 x 1.515 cm^2 R_M = 29.5 mm

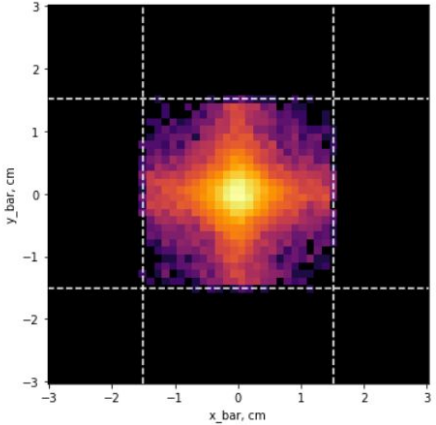
S-shape curves (downscope option)

Spacal Pb/PS (3.03x3.03 cm^2 cells)

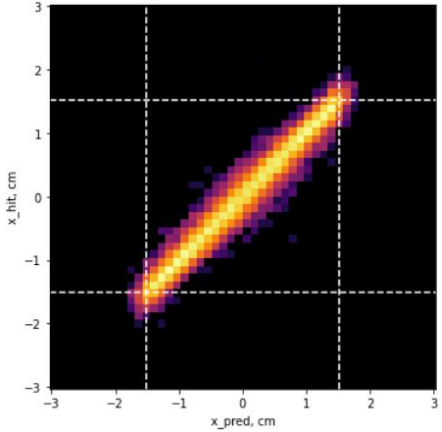
Barycenter -based

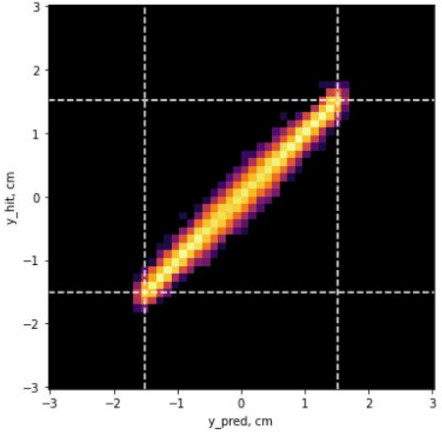


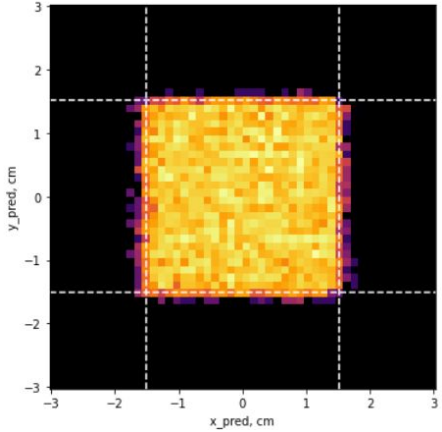




Predicted by the regressor



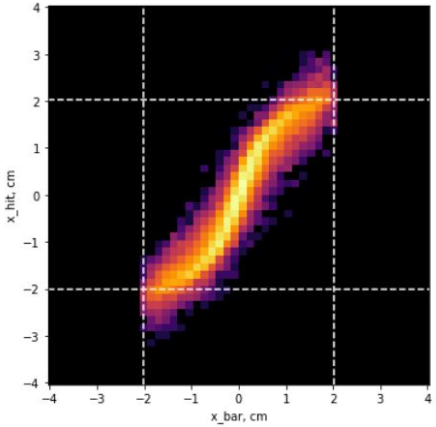


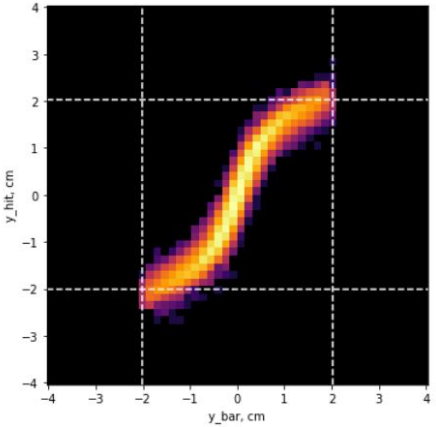


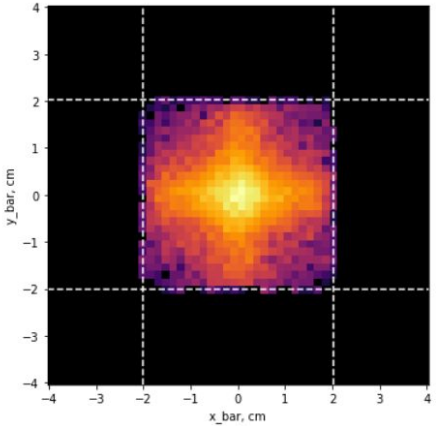
S-shape curves (downscope option)

Shashlik inner (4.04x4.04 cm^2 cells)

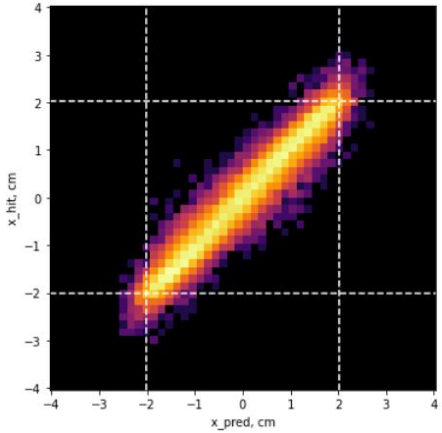
Barycenter -based

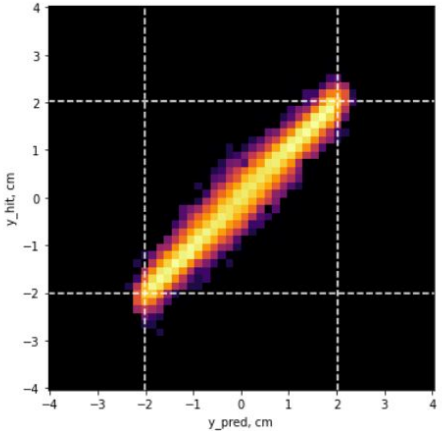


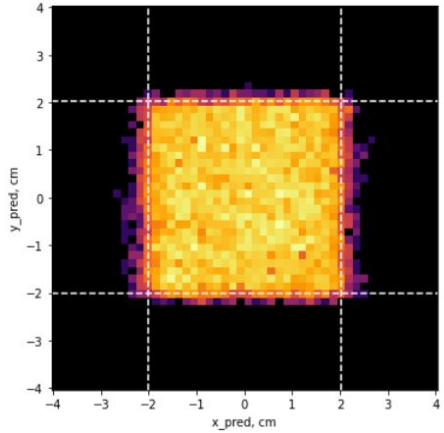




Predicted by the regressor



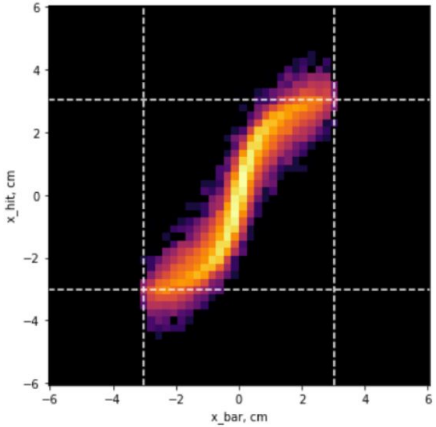


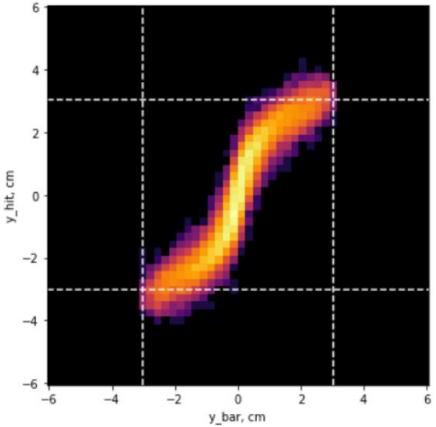


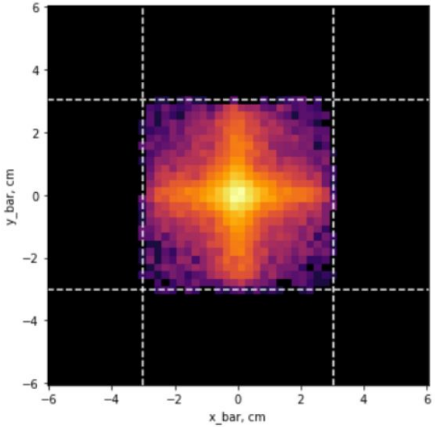
S-shape curves (downscope option)

Shashlik middle (6.06x6.06 cm^2 cells)

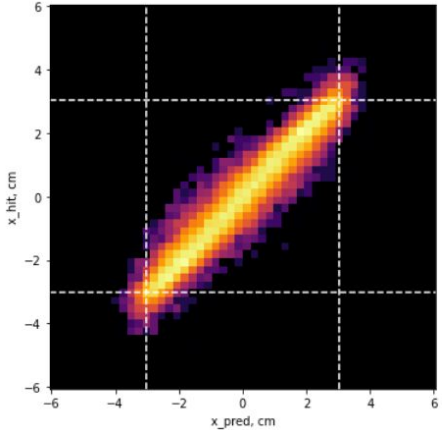
Barycenter -based

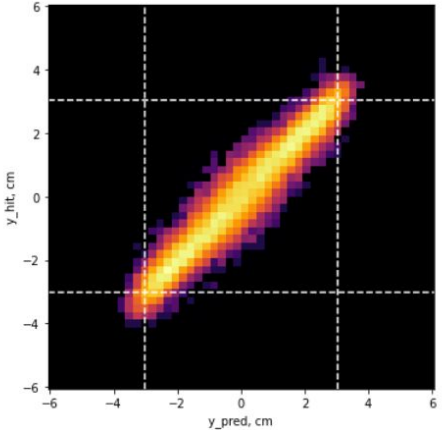


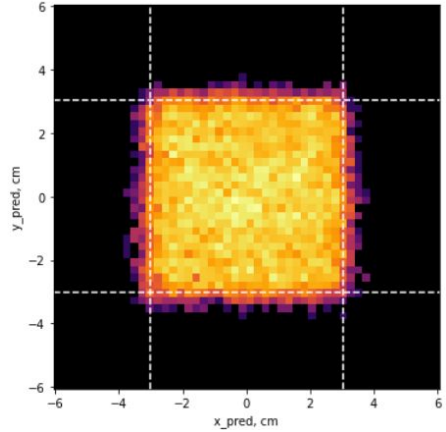




Predicted by the regressor



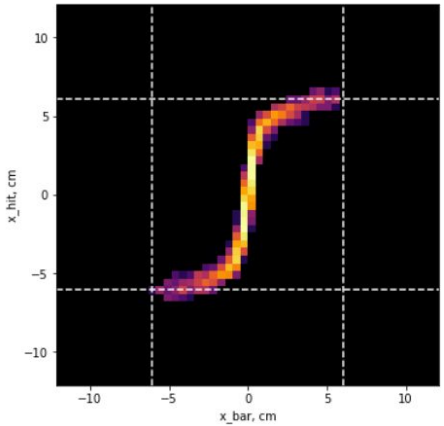


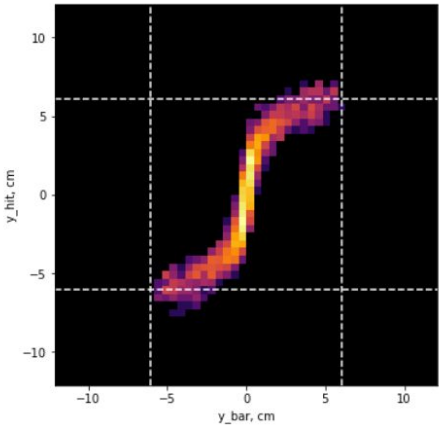


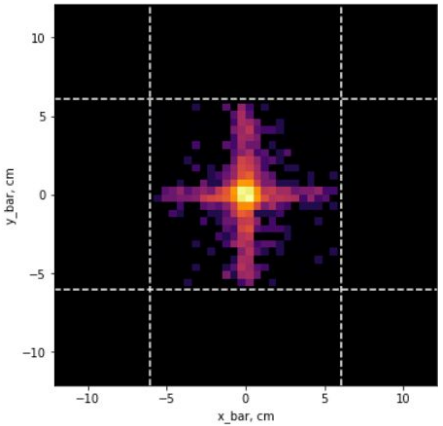
S-shape curves (downscope option)

Shashlik outer (12.12x12.12 cm^2 cells)

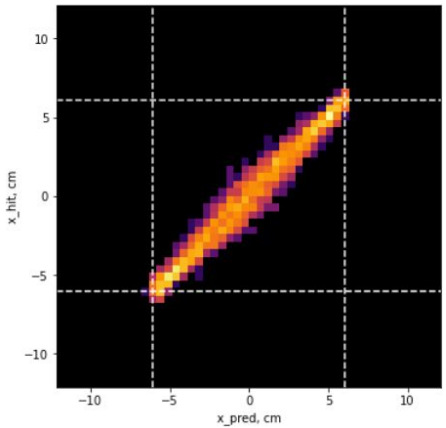
Barycenter -based

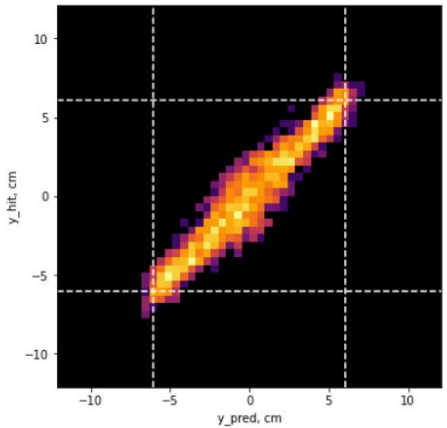


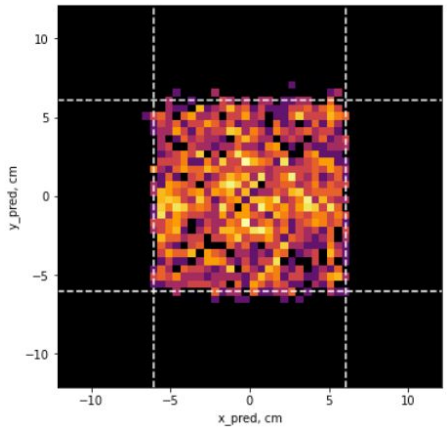




Predicted by the regressor





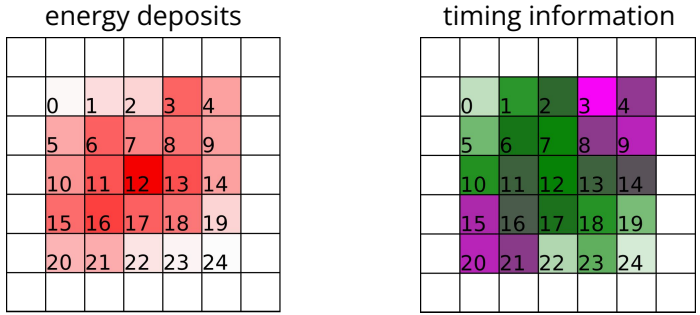


Regressors performance: Timing resolution

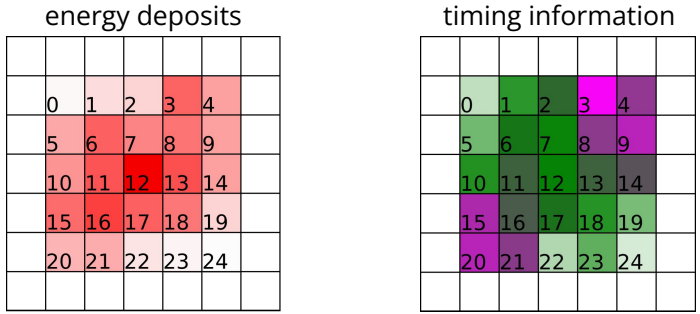
Features used:

- Energy and position reco:
  - Energy deposits are calibrated using Log fitting
  - Energy deposits of 25 cells around the seed cell (2*25 in case of long.-segmented modules)
  - Barycenter position
  - Sums, squared sums, rings, etc. of energy deposits
- Timing regressor:
  - All of the above + 25(50) timing cells

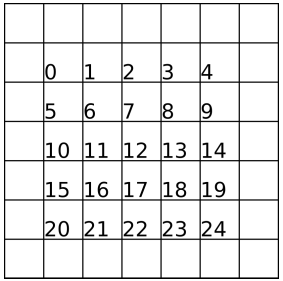
before:



now:



+



Features used:

- All of the above (barycenter position is updated)
- 25 position cells