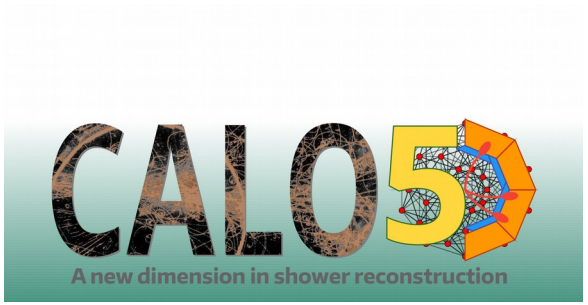
Work Update:
ML ParticleFlow with GravNet and Object Condensation

Uli Einhaus

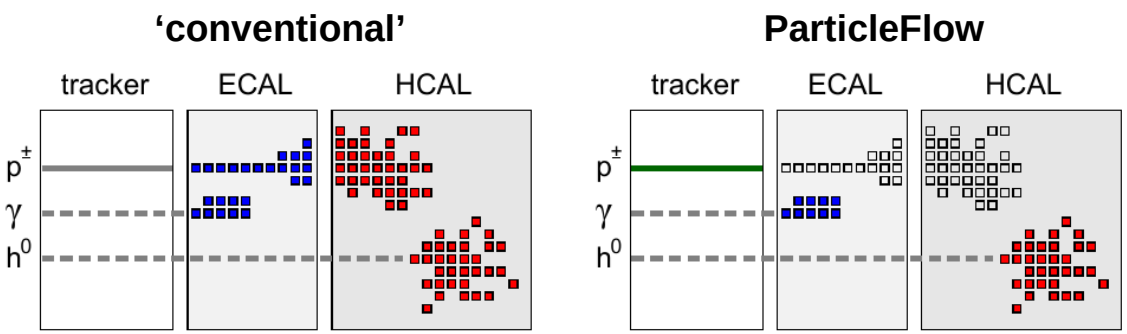
2nd CALO5D Face-To-Face Meeting

15.10.2025



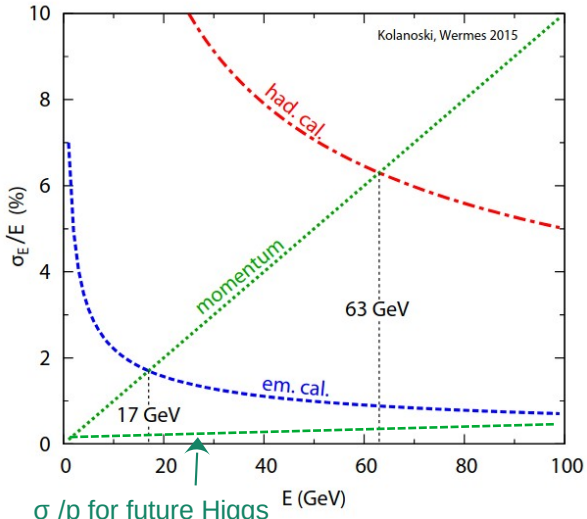
Calorimetry overview

- ParticleFlow: utilise the best subdetector for each particle / jet constituent
  - work horse: highly granular calorimeter
  - future collider detectors (e.g. ILD) developed with PFlow in mind, also applied to existing detectors (e.g. CMS HGCAL for HL-LHC)



$$\sigma_{jet} = f_{charged} \cdot \sigma_{Tracker} \oplus f_{\gamma} \cdot \sigma_{ECal} \oplus f_{neutral} \cdot \sigma_{HCal} \oplus \sigma_{confusion} \oplus \sigma_{leakage}$$

example for one particuar setup by Kolanoski & Wermes

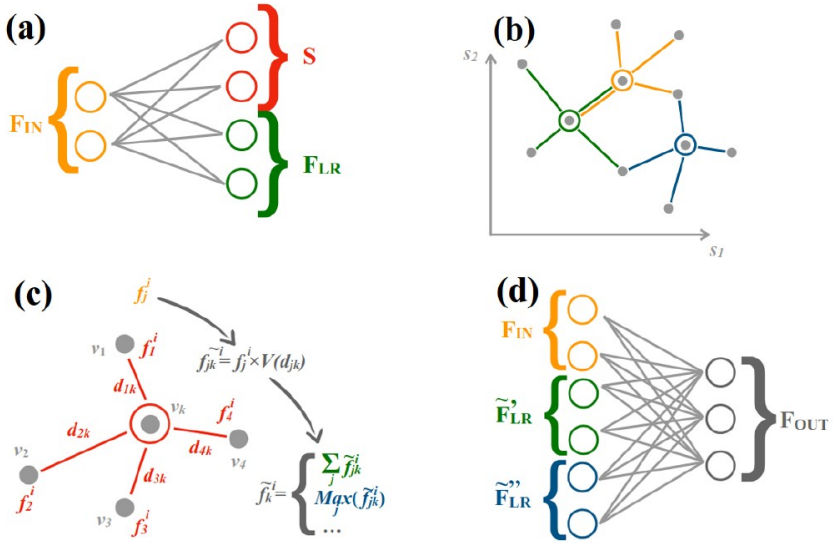


σ_p/p for future Higgs factory detectors

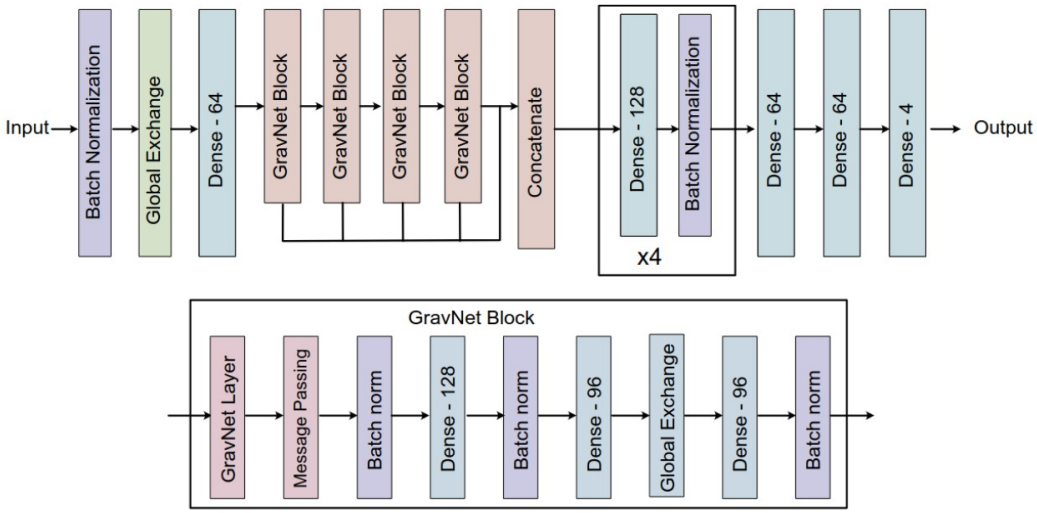
- Use Neural networks to reduce confusion term
- Issue with number of channels for fixed-geometry NN (e.g. CNN) → use GNN

Overview GravNet + Object Condensation

- ‘Holistic’ approach with end-2-end reconstruction - clustering and features reconstruction (position, energy, PID) in one integrated GravNet network
- Calorimeter hits are clustered in latent space S
- Feature properties F_{LR} of nearest neighbours in S are combined with different aggregator function to predict object features



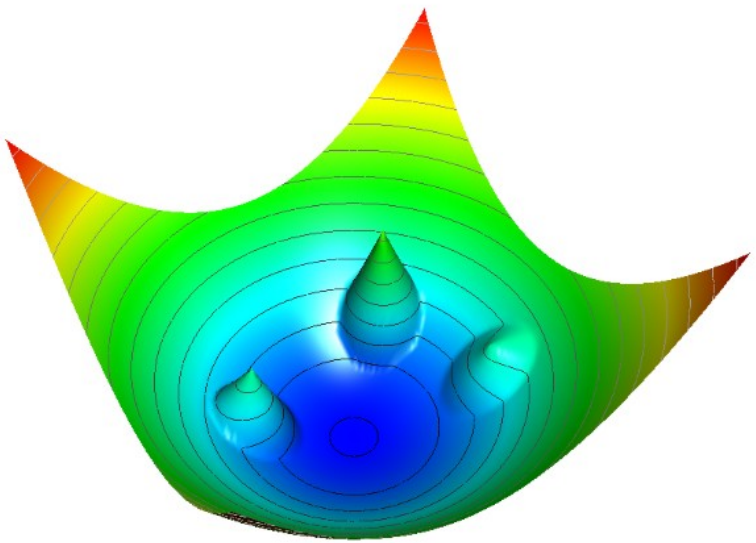
visualisation of the GravNet steps
arxiv:1902.07987



GravNet model setup we used

Object Condensation

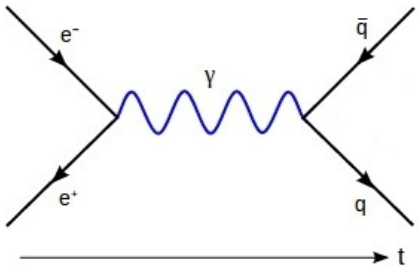
- For clustering introduce attractive potential ~ d^2 for all hits belonging to an object towards each other, and a repulsive potential ~ -d with short range towards all other hits
- To reduce computational effort, introduce β quantifying similarity to object properties, only use hit with highest β per object to generate combined potential
- Additional loss per object of 1-β of the condensation point encourages one hit per object to become representative of object and ‘gather’ related hits into a cluster



Object condensation attractive + repulsive potential visualisation
arxiv:2002.03605

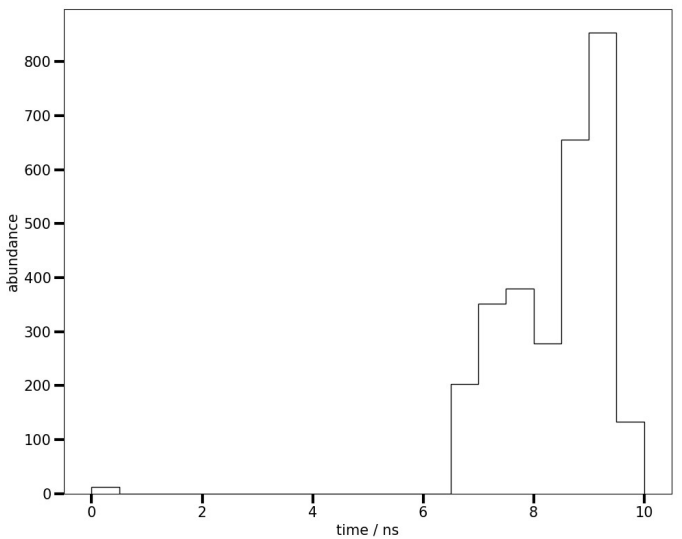
Data set

- Use data set of $e^+e^- \rightarrow (\gamma)q\bar{q}$ at 250 GeV (2fZhad), 45k events
- O(a few 1000) hits per event
- Truth: properties of MCParticle that entered calo and caused energy depositions
  - full MCParticle chain before calo from geant4 is available; in case of multiple contributions to the same hit, largest energy fraction is chosen
  - index, end point in calo (x,y,z), (energy, PID)
  - issue: particles that backscatter from the calorimeter sometimes are not registered, and hits are associated with the original particle that entered the calorimeter
- Simplification cut: all hits with $r_{\varphi,\theta} > .5$ m from true object are ommitted



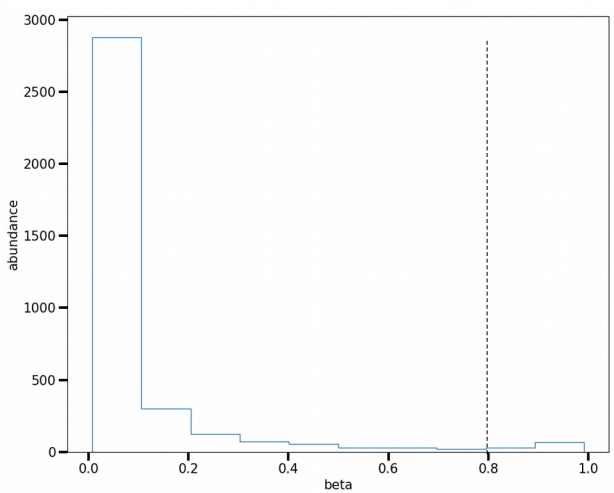
Properties used

- Input: tracks, ECal and HCal hits with 9 properties each
  - x, y, z, $r_{x,y}$, θ, η, E (p for a track), time, ID (isTrack)
  - three timing 'resolutions' are compared: perfect, smeared with 100 ps, none (t==0)
  - time tail cut: t [t>20ns] := -1
- Output: 12
  - target: β, cluster coordinates (x,y,z), position (x,y,z), energy, PID (4 channels)
- Loss:
  - object condensation → β, cluster coordinates
  - hit loss → position (energy and class not used atm)
- $L_{pos} = Huber\left(\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2 + 10^{-6}}, 0.1\right)$

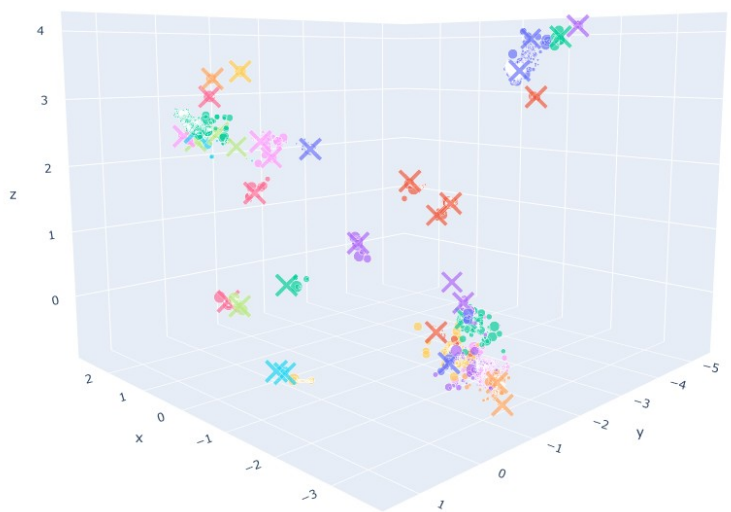


Post-Inference Thresholding

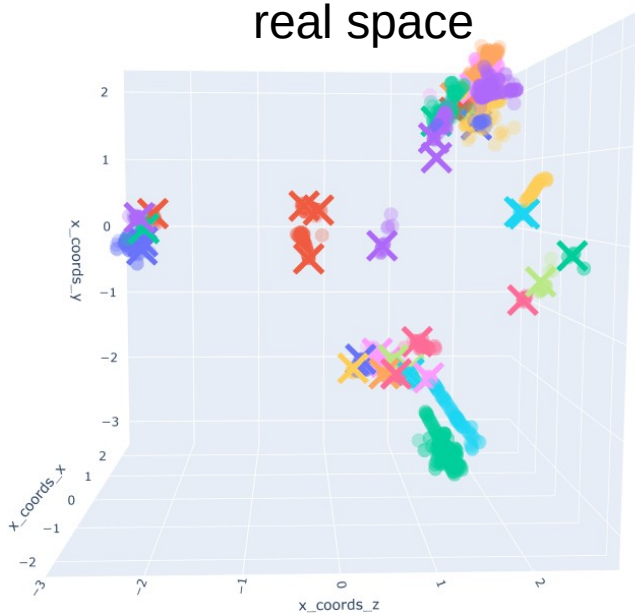
- For predicted objects, i.e. PFOs, thresholds of min beta and min dist to next PFO are chosen after running inference
- beta: few entries with values close to 1 represent the well reconstructed PFOs
- In space projections: matching colours belong to the same MCParticle, X mark predicted PFOs, here for β > 0.8



cluster space

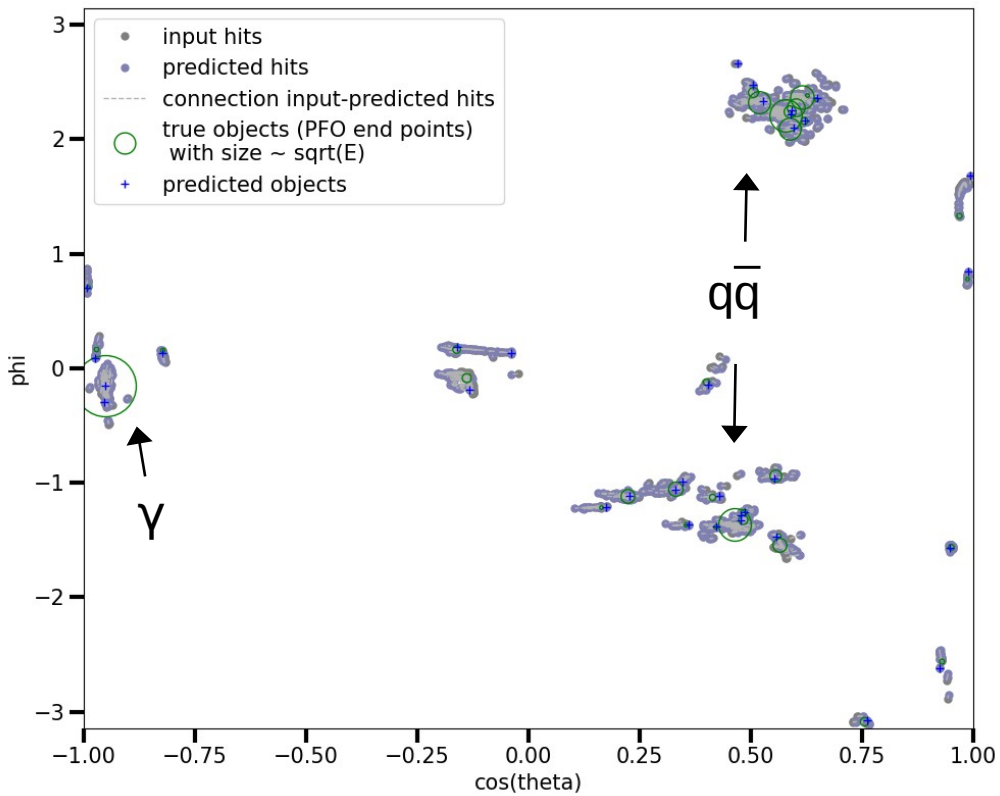


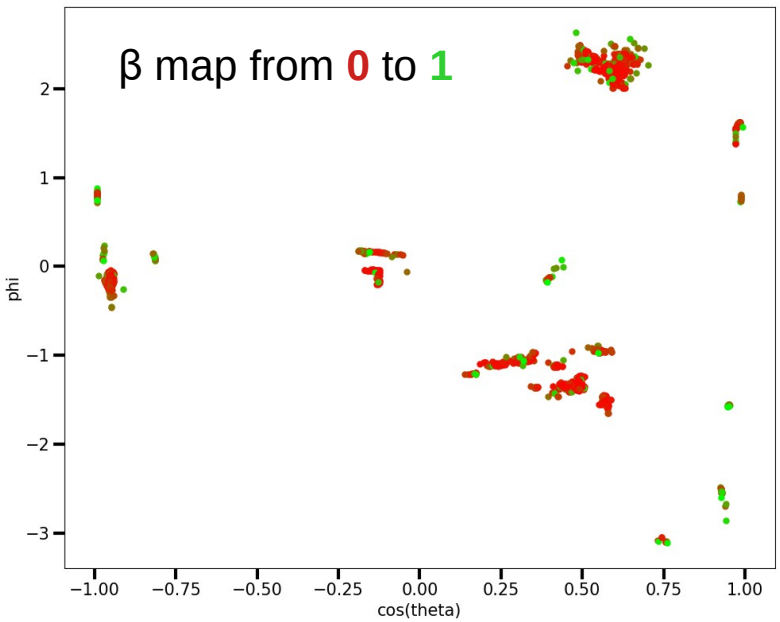
real space



Angular Projection of Example Event

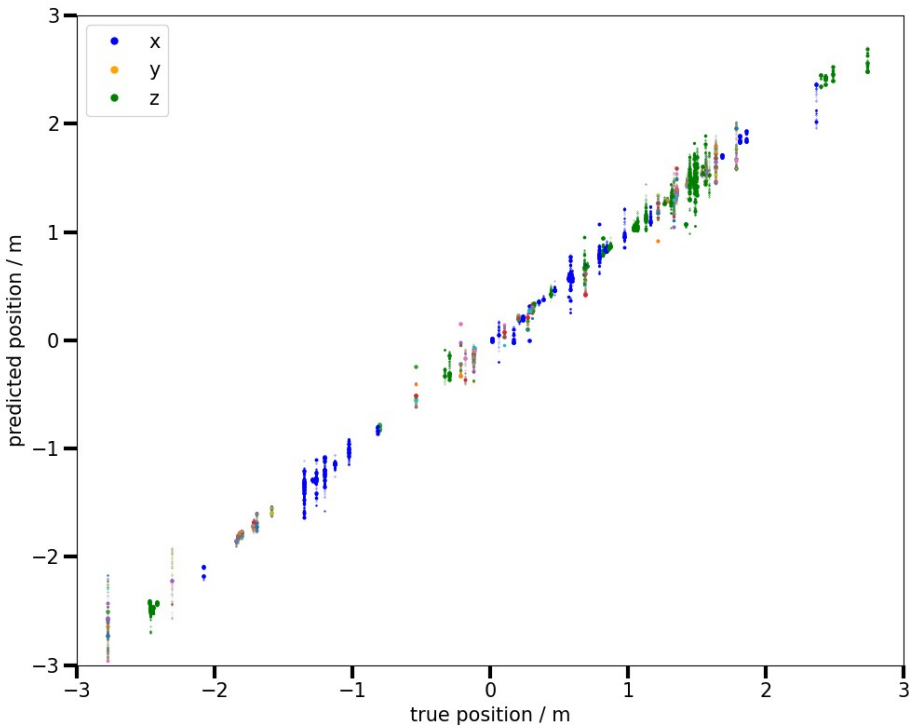
- Predicted positions excellently clustered around targets
- In many cases one predicted PFO (+) per MCParticle (o), here: 35 and 34

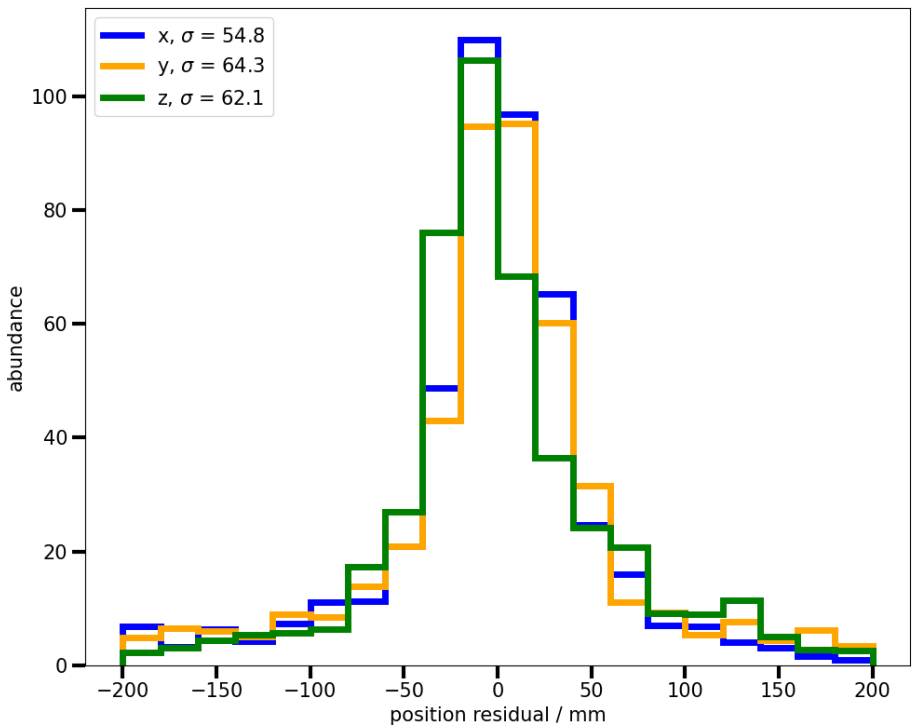




Position Reconstruction

- Position per hit, weighted with beta
- Good correlation, PFO position uncertainty about 6 cm with current training setup

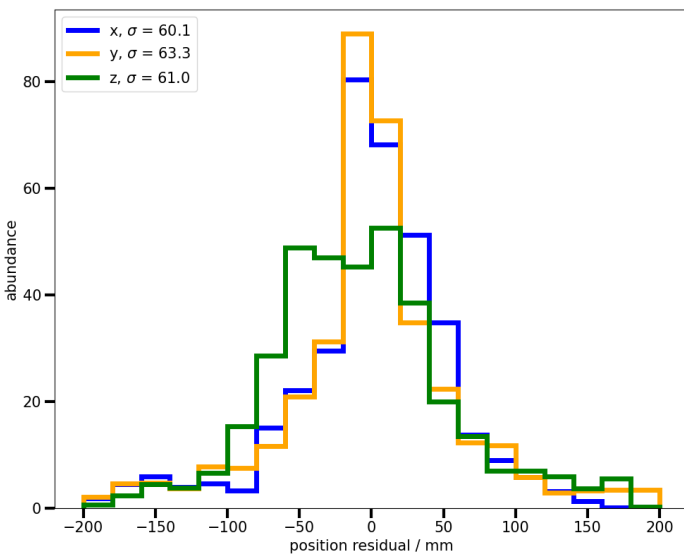




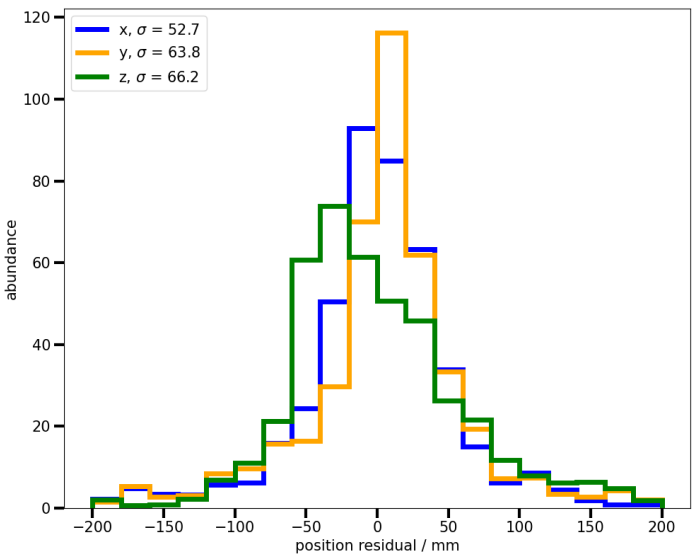
Impact of Timing

- [here: 1 event, event-to-event fluctuation larger than timing impact]

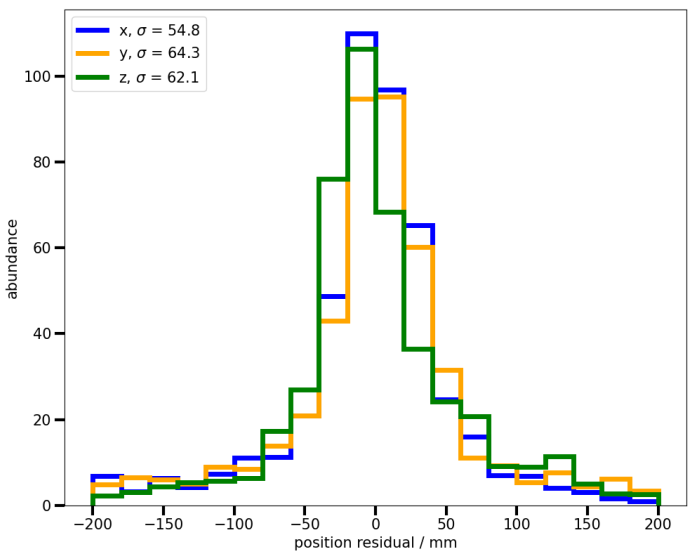
perfect timing



smearing with 100 ps



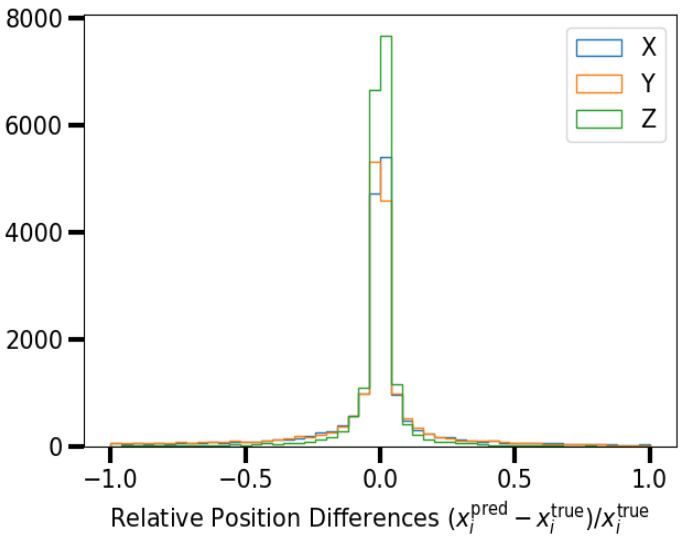
no timing



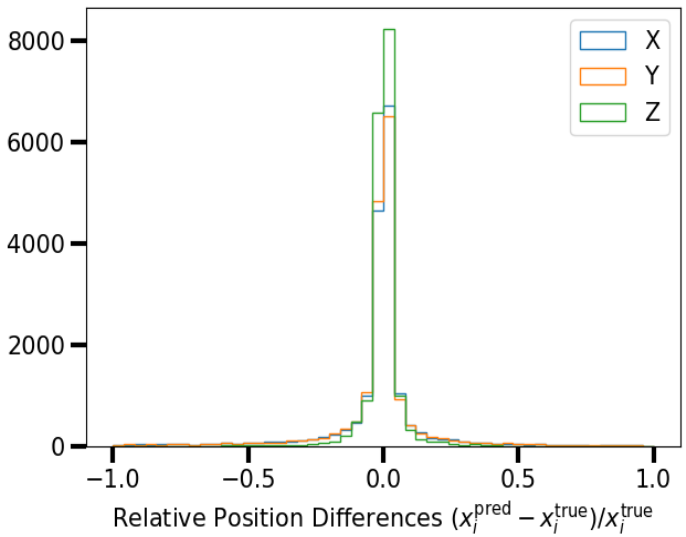
Impact of Timing

- [statistics for 400 test events, width missing]

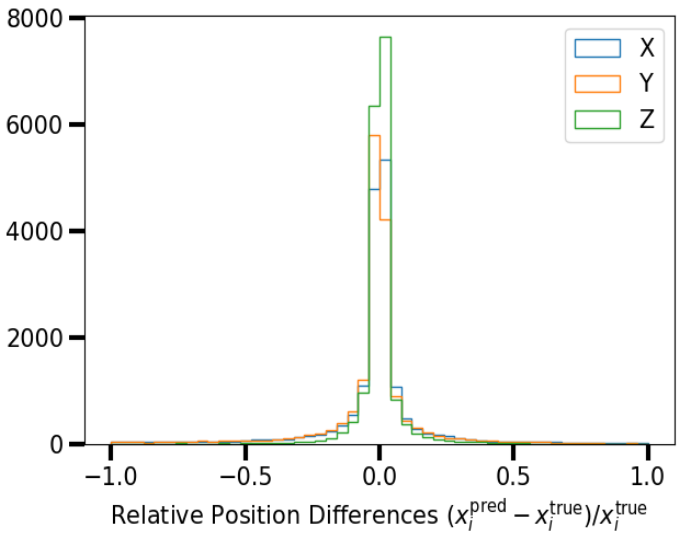
perfect timing



smearing with 100 ps



no timing



Conclusion & Outlook

- **Promising results** with GravNet + Object Condensation for highly granular calorimeters: clustering and position prediction work well, optimisation ongoing
- **More statistics** and variation of the hyperparameters (and possibly the model), proper assessment of clustering efficiency/efficacy
- **Additional features** energy and PID to be added next, then go to jet energy resolution
- **Application** to physics channel