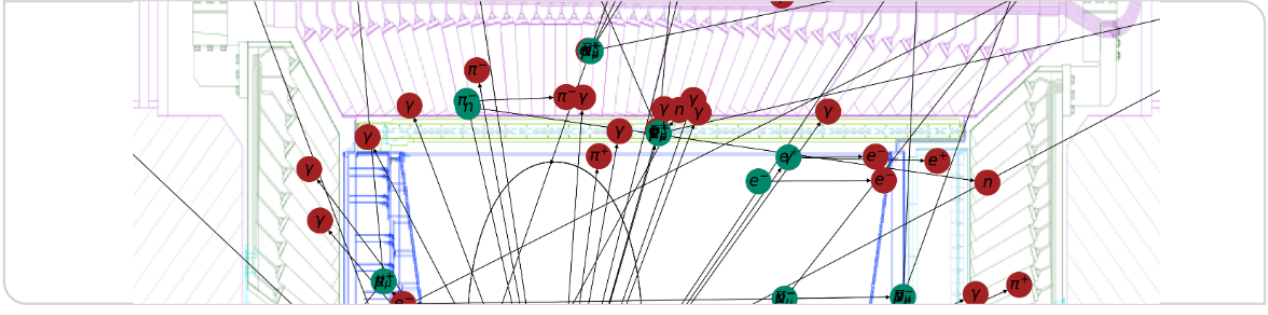
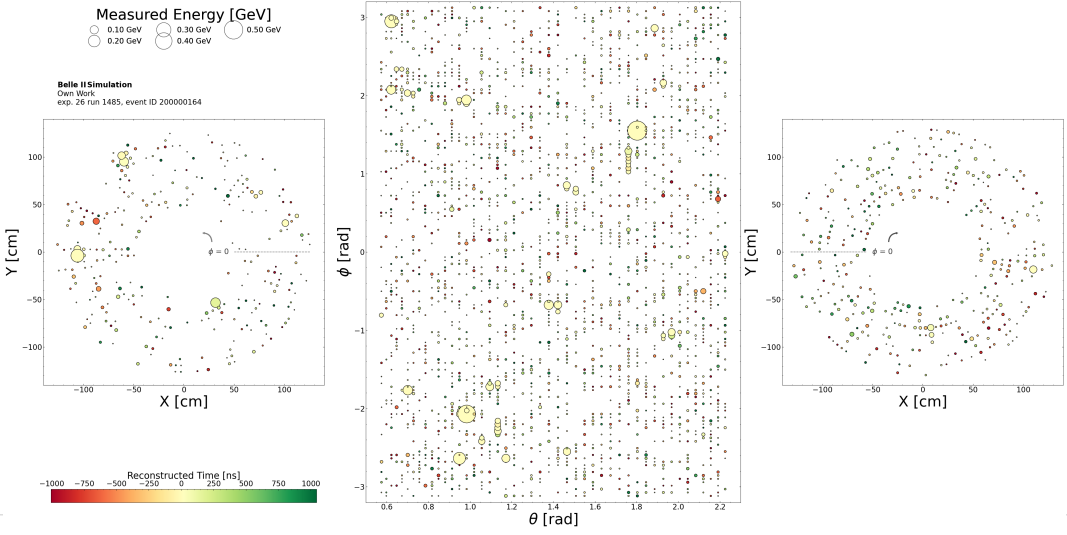
# Clustering in the Belle II ECL using Graph Neural Networks

Jonas Eppelt, Torben Ferber | 28.September 2024



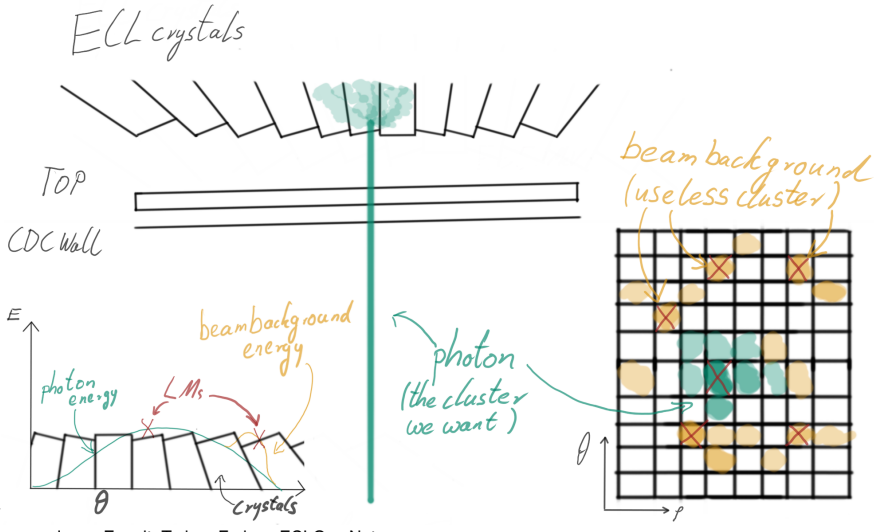
# Exemplary B-event



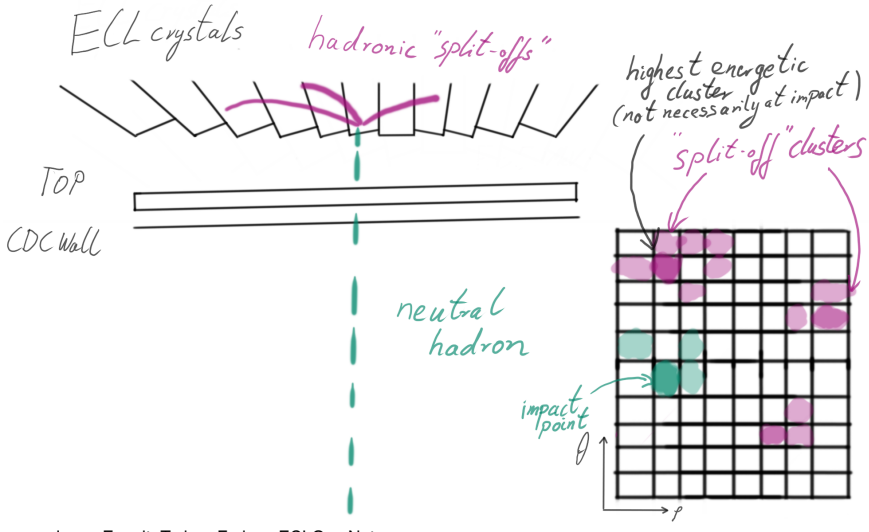
# Goals

- I want to identify localized depositions and group them together (clustering).
- I want to identify these depositions, that come from particles created in the collision.
- I want to remove as many clusters created by beam background as possible.

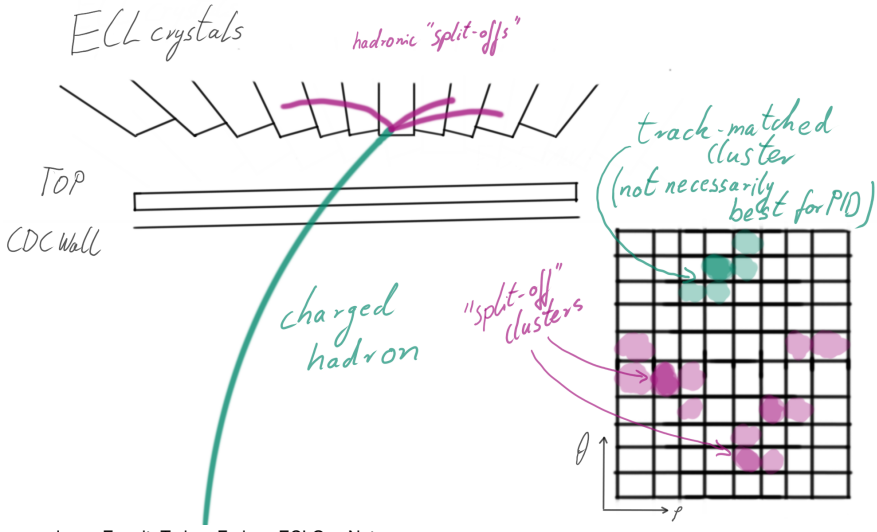
# Challenges from beam background



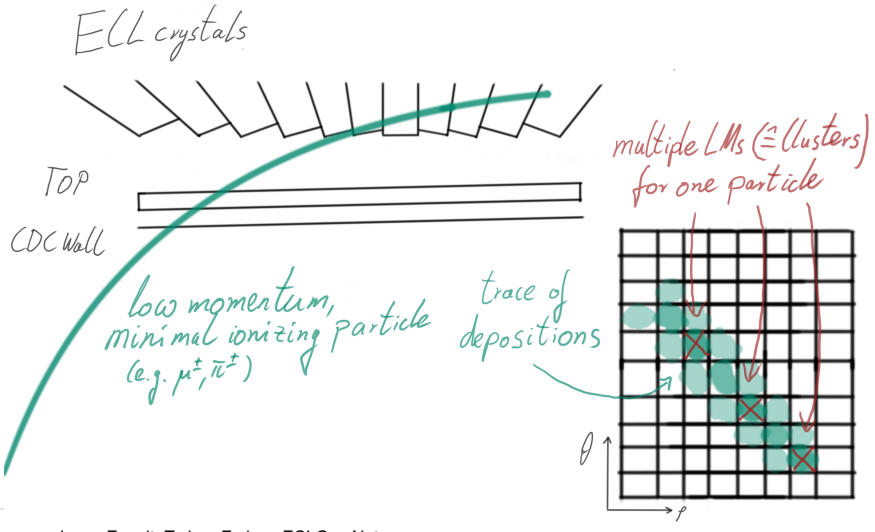
# Challenge of neutral hadrons



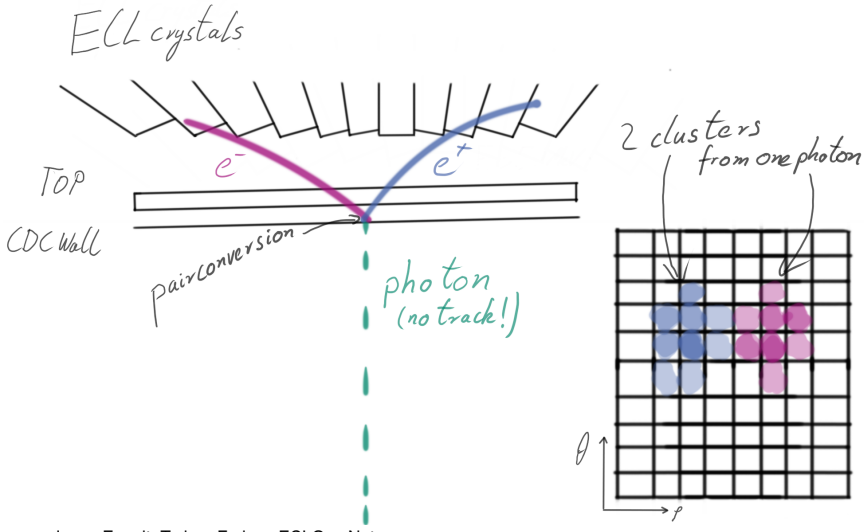
# Challenge of charged hadrons



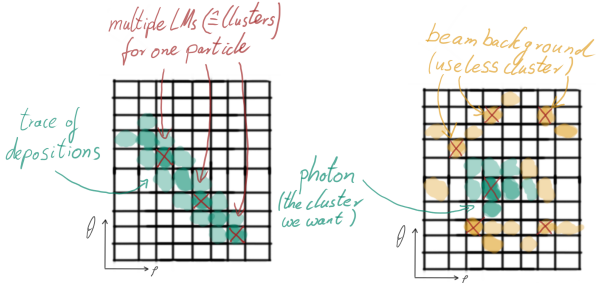
# Challenge of traces from minimal ionizing particles ("MIP-Worms")



# Challenge of pair conversions before the ECL

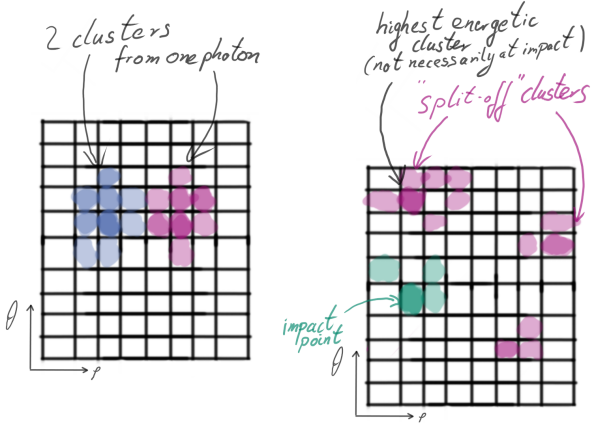


# Tackling beam background and MIP traces



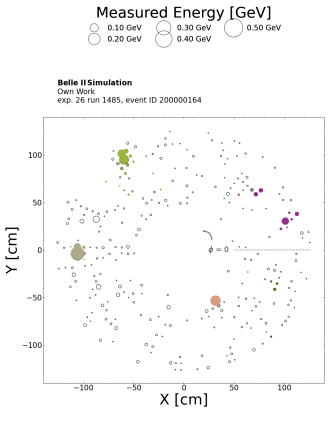

- What are energy depositions from signal particles and which come from beam background?
  → If there is no MCParticle matched to a Local Maximum (LM), it is marked as background.
- If a particle has multiple Local Maxima, which is the correct one?
  → Match every MCParticle, to the Local Maximum (LM) in which it deposited the most energy.

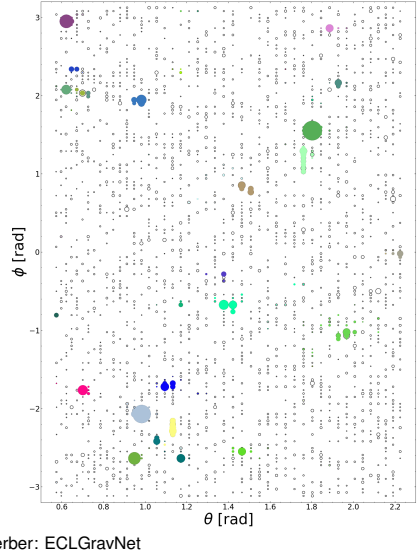
# Tackling hadronic split-offs and pair-conversion

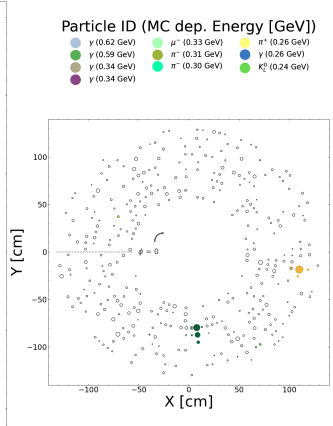



- Basf2 does not store information about the split-off particles from hadronic interactions.
  Instead, split-offs are attributed to their parent particle.
- I use a custom matching scheme, that:
  - keeps particles created before the ECL or,
  - particles that traveled at least 40 cm in the ECL.
- In both cases, the particles also need to have at least 20 MeV kinetic energy.

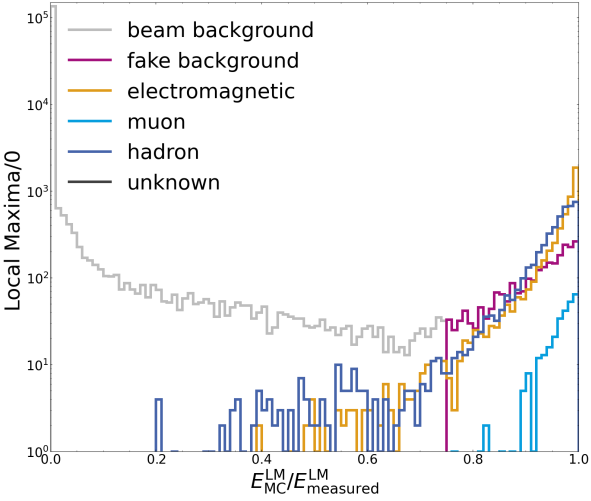
# Examplary B event with MC information



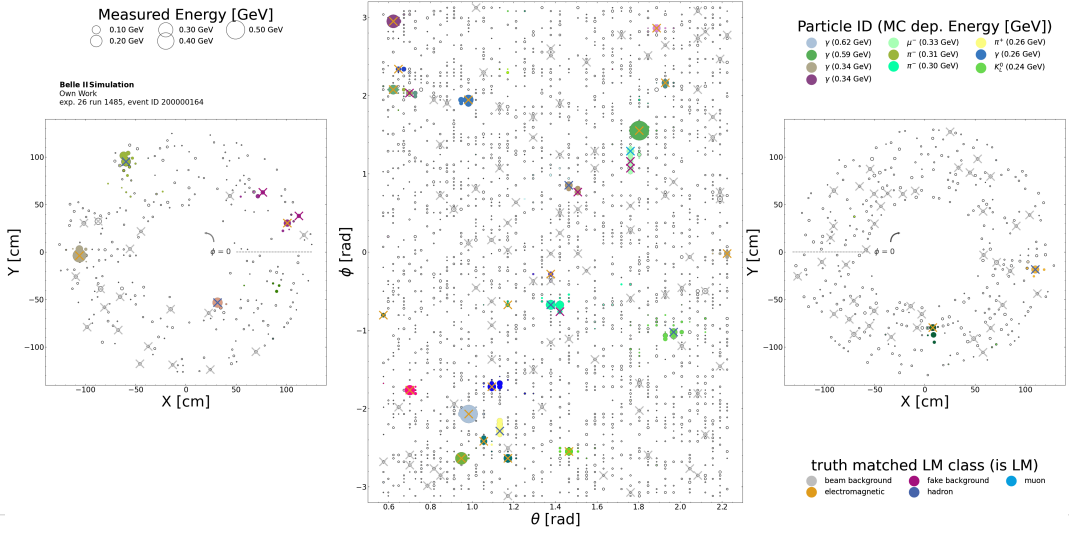



# Definition of target classes

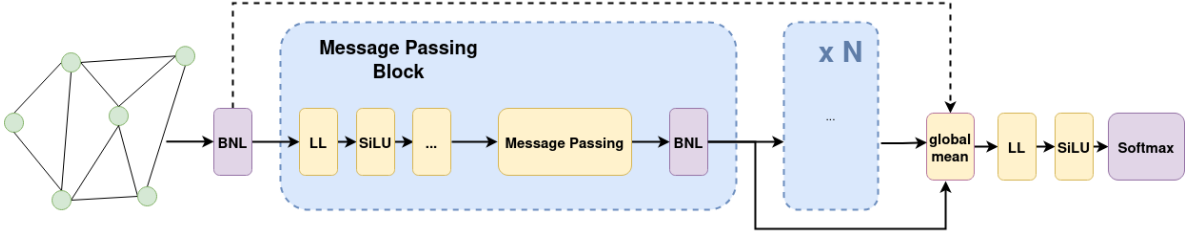



- beam background LMs:
  The LM is not matched to any MCParticle and $\frac{E^{\mathrm{dep}}_{\mathrm{MC}}}{E_{\mathrm{measured}}} < 75\%$.
- fake background LMs
  The LM is not matched to any MCParticle and $\frac{E^{\mathrm{dep}}_{\mathrm{MC}}}{E_{\mathrm{measured}}} > 75\%$.
- electromagnetic LMs
  The LM is matched to a photon, electron or positron.
- hadronic LMs
  The LM is matched to any hadronic MCParticle.
- muonic LMs
  The LM is matched to a muon.

# Examplary B-event with target LMs

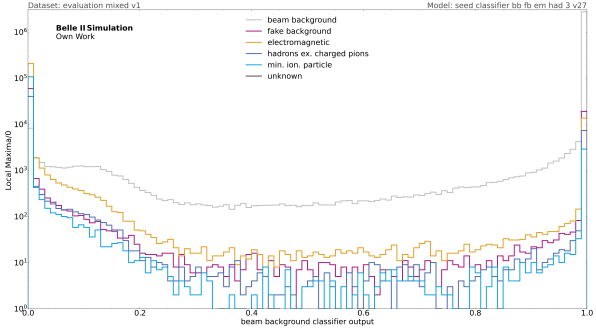


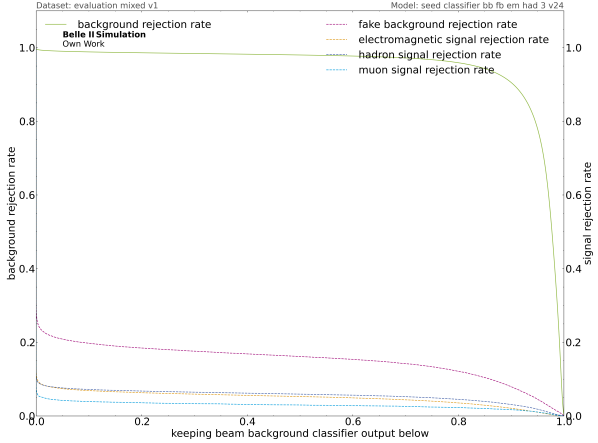
# The Seedclassifier model



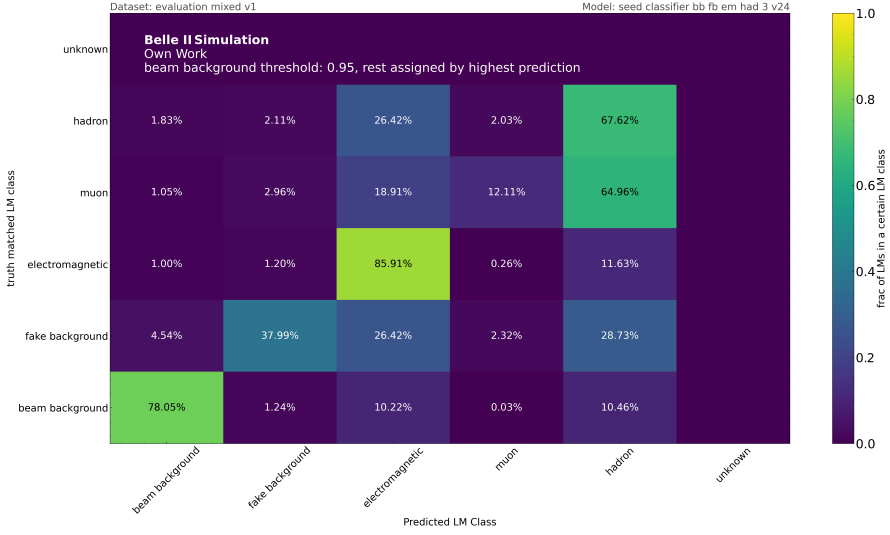

- Message passing blocks collect information about the neighborhood.
- A Focal Loss (weighted cross-entropy) accounts for a large imbalance towards background samples.
- The training is done on Particle Gun events enriched with low energetic particles.

# Seed classifier results - beam background identification



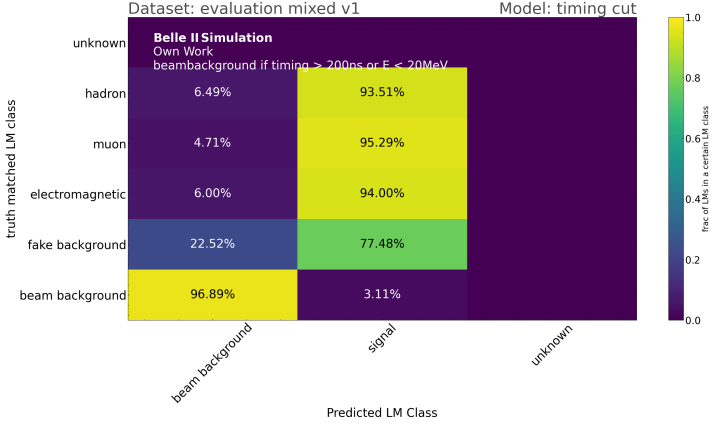



# Seed classifier results - confusion matrix



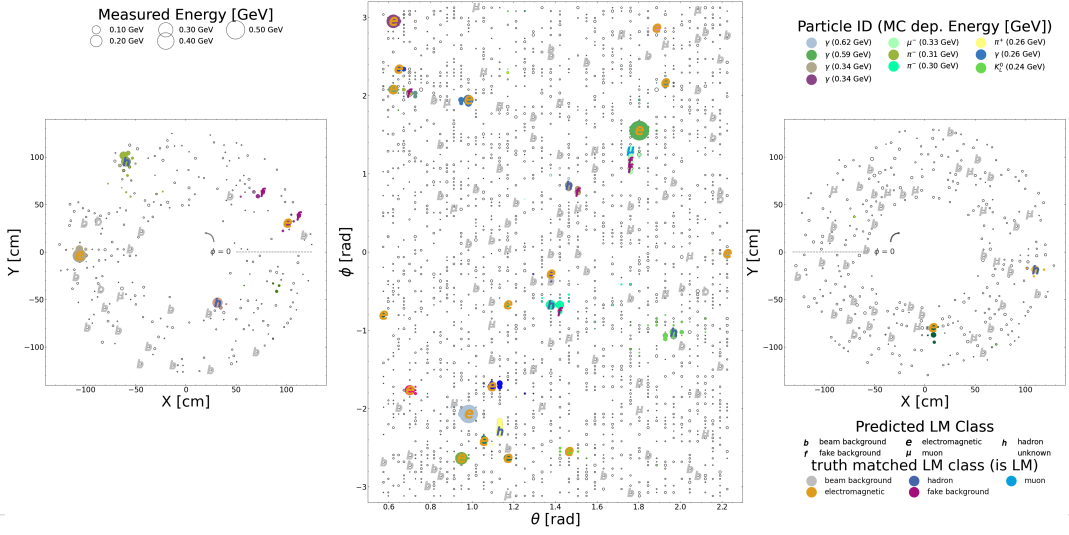

- Every LM, that has a beam background prediction value $> 0.95$, is assigned as such.
- All other LMs are assigned the most likely class (except beam background).

# Comparison to a Timing Cut

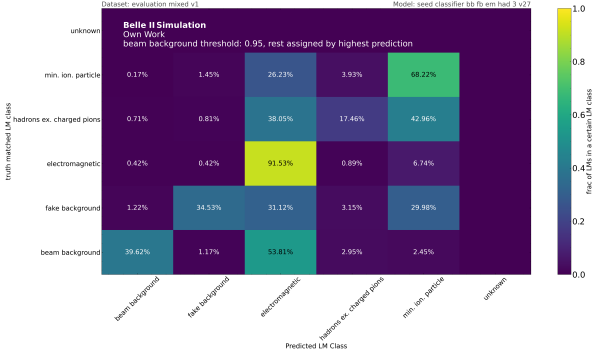



- Every LM, that has less then 20 MeV or whose timing is below 200 n is labeled beam background.
- This selection can be implemented in the next run (see Torben's talk).

# Seed classifier results - exemplary B-event



# Summary and outlook




## Summary

- Using GNNs, we can reduce the number of beam background LMs.
- Further, we can identify electromagnetic LMs.

## Outlook

- Seperating hadrons and muons is more challenging. I plan to use different training targets: *minimal ionizing particle* and *other hadrons*.
- Using these labels, I plan to build clusters from them.