Object Condensation Tracking

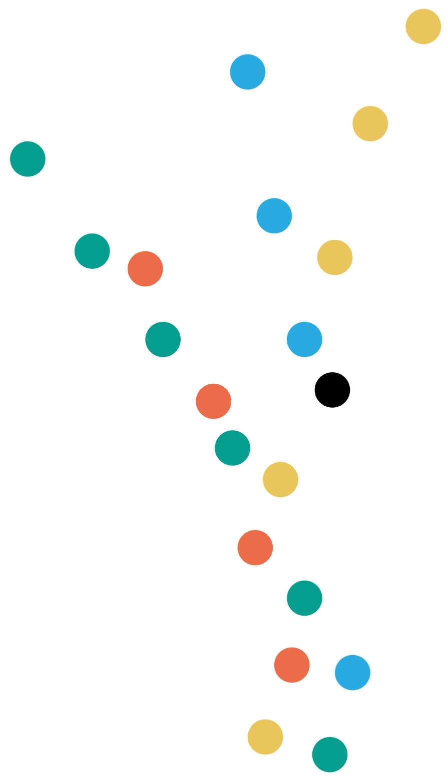

Kilian Lieret[1,2]
@klieret

Gage DeZoort[1]
@gageDeZoort

[1]Princeton University, [2]IRIS-HEP

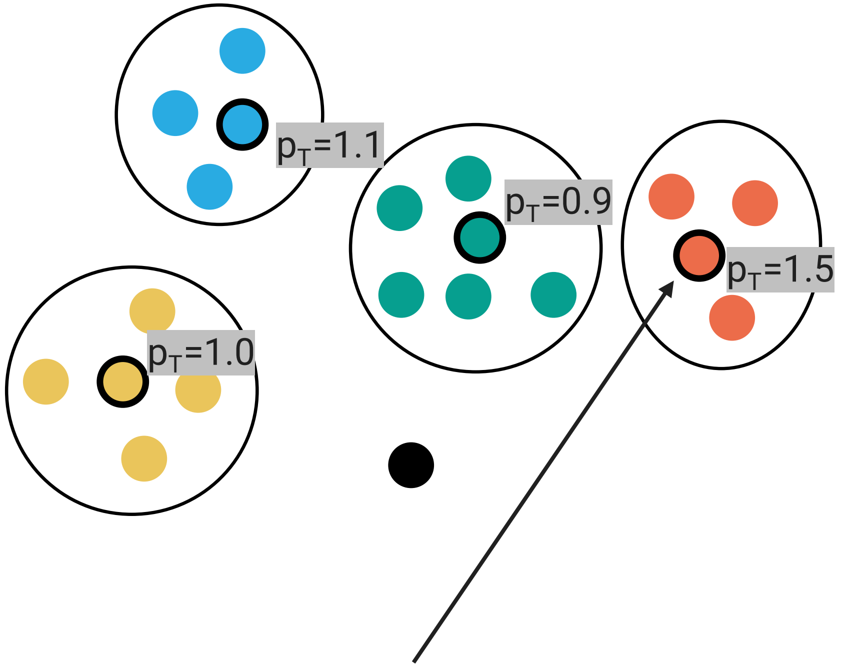
Vision: One-shot tracking

Point cloud



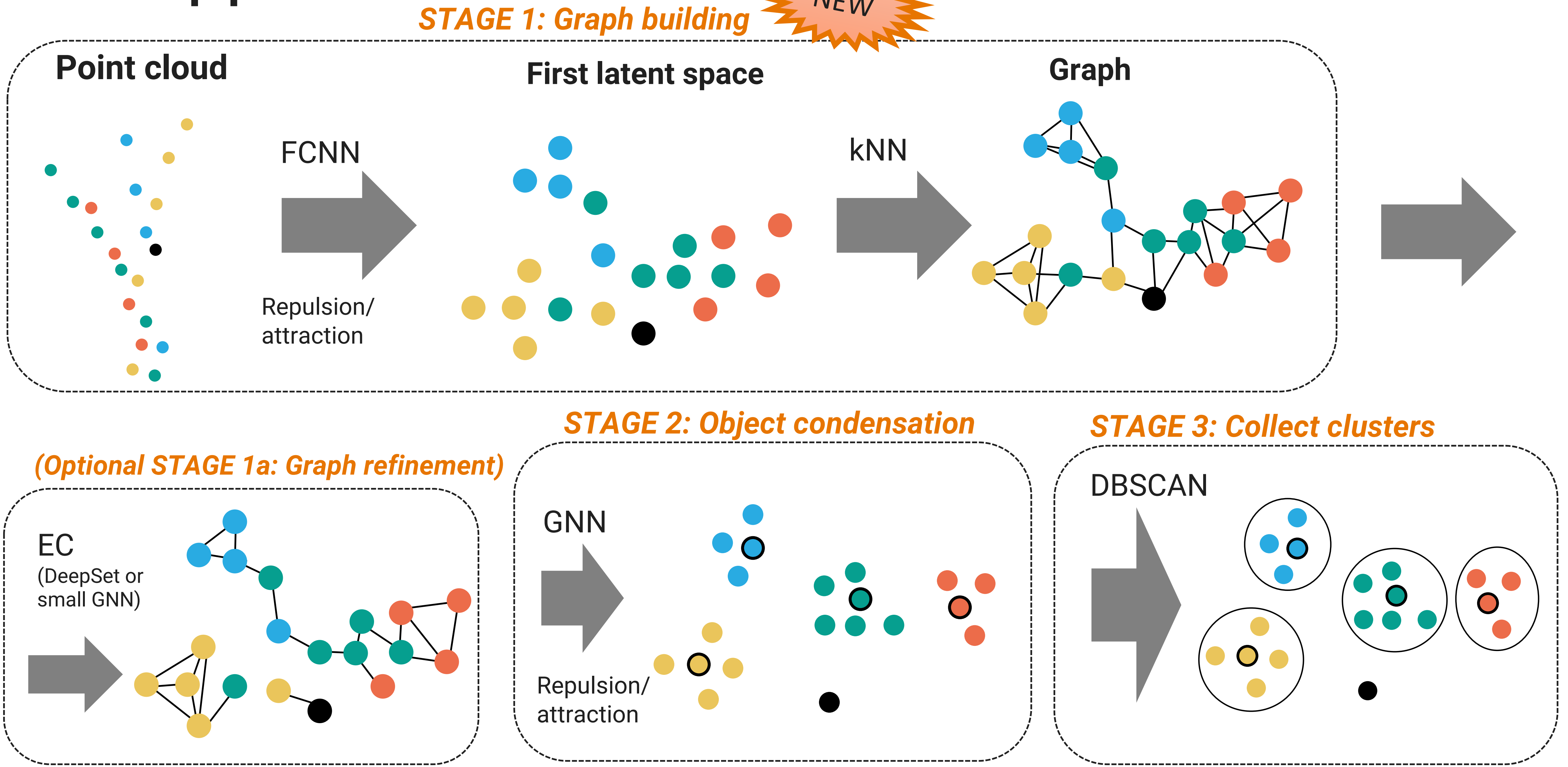
Repulsion & **attraction**
of points in latent space

Learnt latent space
Hits already clustered by particle;
Clusters can be collected trivially



Condensation point
Represents the track, can
learn track parameters like pT
(WIP)

Current pipeline



People

Actively looking for more collaborators, especially additional core team members (e.g., Ph.D. student)

Core team

Kilian Lieret
(Princeton)
Current development & training

Gage DeZoort
(Princeton)
Ideas, GNN know-how, original codebase

Transformer exploration

CS focused;
Tracking as application/benchmarking

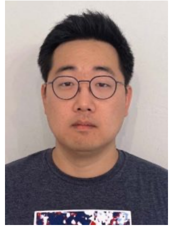
Siqi Miao
(Georgia Tech)

Pan Li
(Georgia Tech)

CMS LST Liaison

Application to CMS data, possibly combined with early LST stages

Jonathan Guiang
(UCSD)



Philip Chang
(Florida)

Summer fellows

Jian Park
(Chicago)

Devdoot Chatterjee
(Delhi Tech U)

Refilwe Bua
(Brown)

Feedback & intellectual support

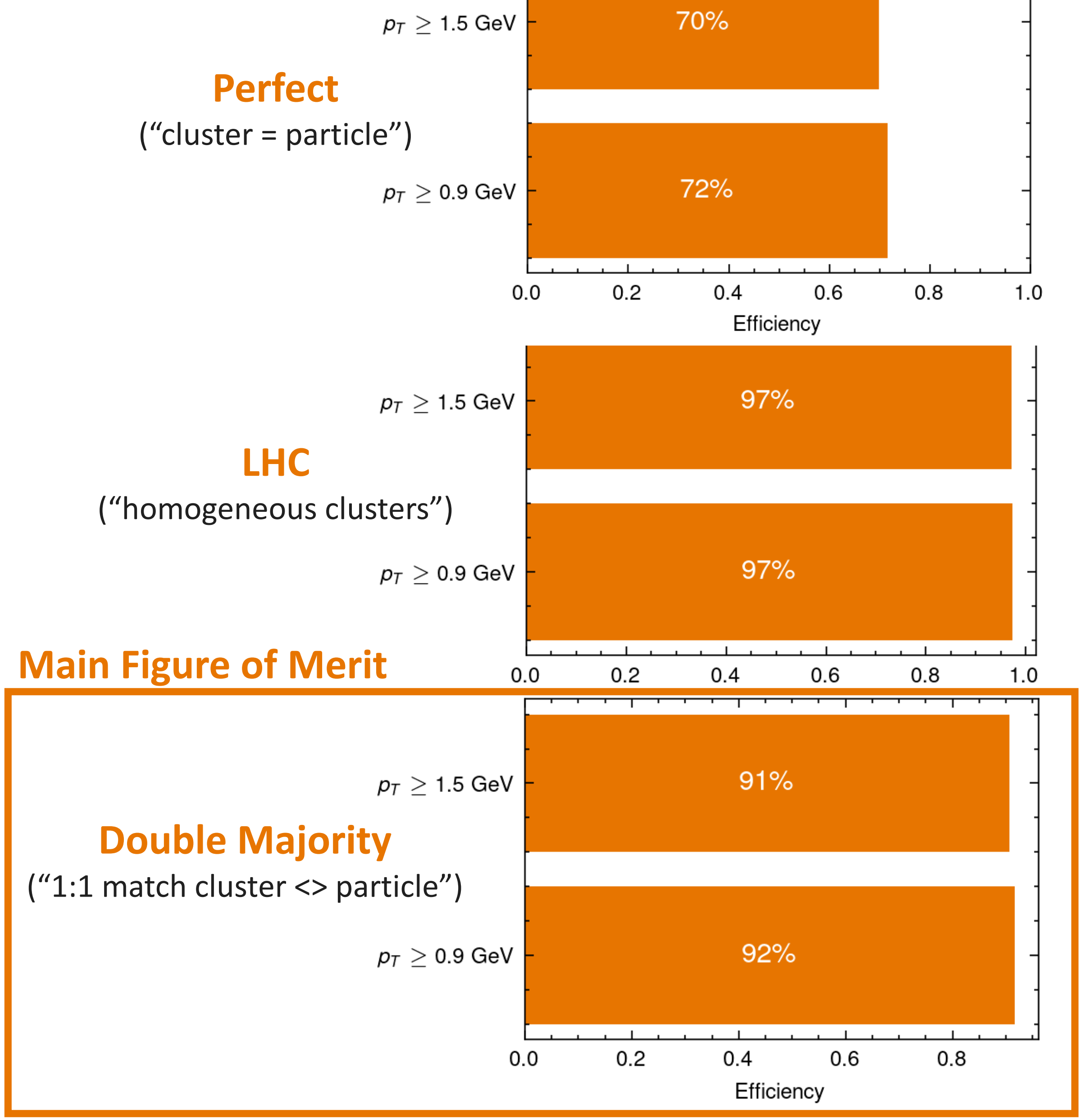
Javier Duarte
(UCSD)

Savannah Thais
(Columbia)

Status

- So far running on **trackML dataset pixel layers**
- **Results presented at CHEP:**
  - First results **without truth cuts**
  - OC results seem to **match/outperform comparable EC GNN pipeline** trained on same data when using connected components for evaluation
- **Significantly improved pipeline since CHEP:**
  - Using embedding + kNN for graph construction
  - Performance improved while reducing memory consumption
- **In progress:**
  - CHEP proceedings
  - Paper with results of new pipeline

CHEP results (superseded)

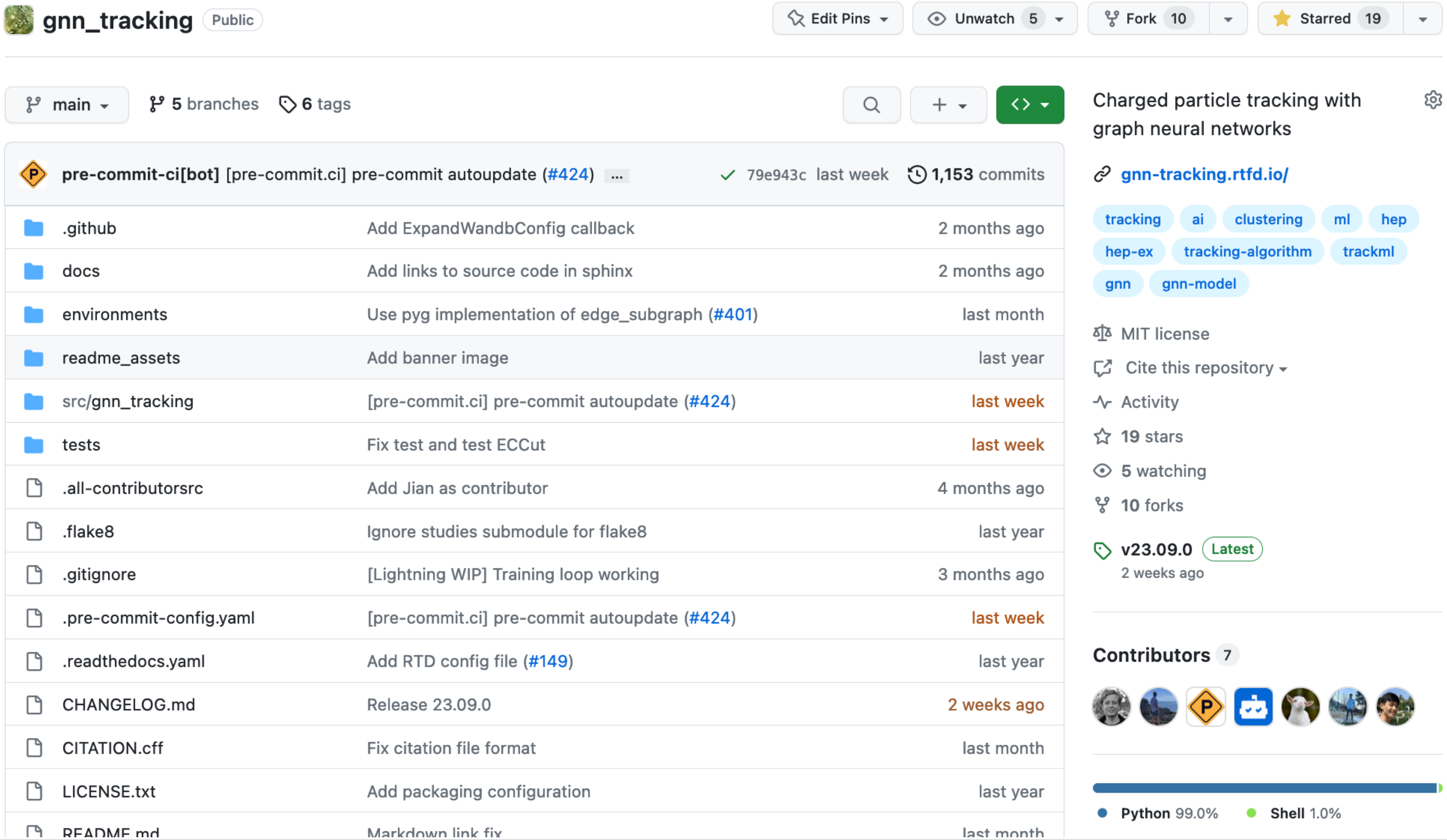


Plans

- **Core team is only ~1 FTE**
  - enough to sustain **exploration and R&D**
  - unlikely to bring this towards production fast
    (ExaTrkx has many times our resources and is only convering there)
- **2023 Milestones:**
  - **Publishing** current research
  - Running on **full-detector graphs** → This will result in more useful benchmarks & 🍏-🍎 **comparisons**
- **2023 Forks in the road**
  - Does our approach **match/outperform ExaTrkx**'s EC-based model on the full detector?
  - Can we apply OC tracking to MDs from **LST project**? Do we see significantly better features in **CMS data**?
  - Will **local transformers** be competitive in our approach?
  - Does it make sense to merge frameworks with **ExaTrkx**?
- **2024 Milestones:**
  - Adding **track parameter prediction**
  - Training on **CMS data** and getting first benchmarks/results
  - First **performance tests for speed**
- **2025 Milestones:**
  - **Performance optimization** with specific accelerators
  - **CMSSW** integration
  - **Physics** studies with CMS data
- **End of project (3-5y):**
  - OC Tracking is used/tested in **production**

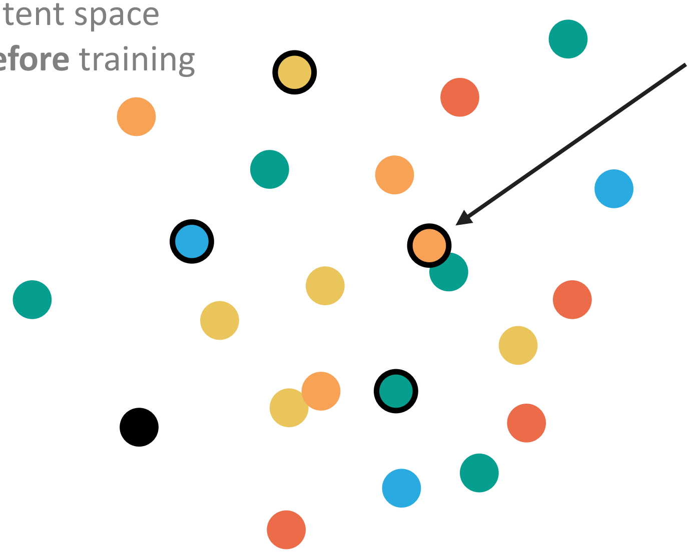
Thanks!

Find us on github!
https://github.com/gnn-tracking



Object condensation: Training losses

Latent space
before training



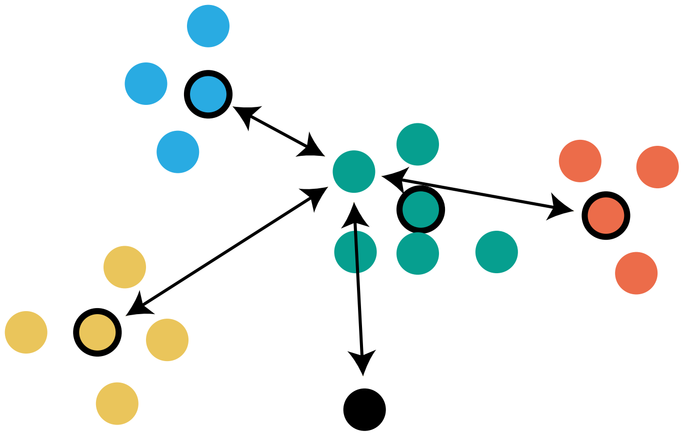
GNN predicts **condensation likelihoods (CL)** for every hit.
Hit with max CL for particle* is **condensation point (CP)**

*during inference: for cluster

Repulsive loss function

penalizes hits close to <u>other</u> CP

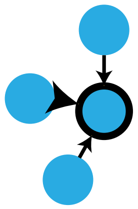
hinge loss: no more repulsion after certain distance
repulsion stronger for strong CP CLs

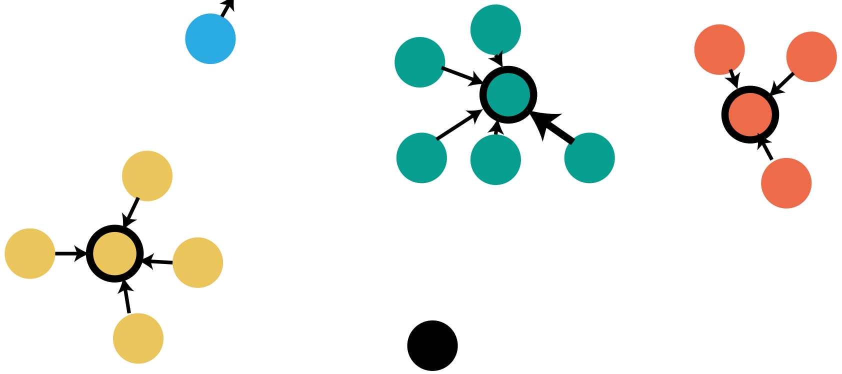


Attractive loss function

rewards hits close to <u>their</u> CP

quadratic potential
Attraction stronger if CP's CL is high





Background loss function

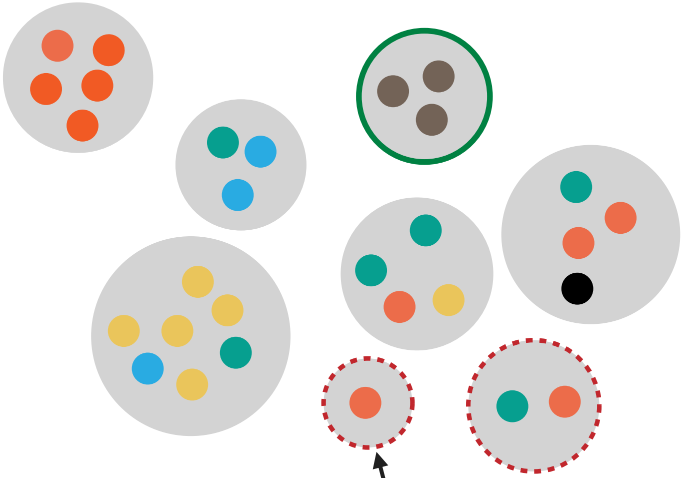
noise hits should have low CL

Loss functions implemented from Kieseler 2020 ([2002.03605](2002.03605))

Metrics

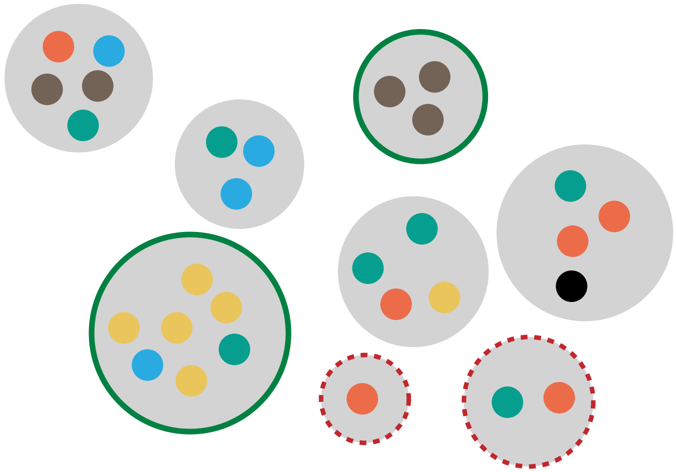
Perfect

Cluster contains only hits from one particle
and
no hits outside of cluster



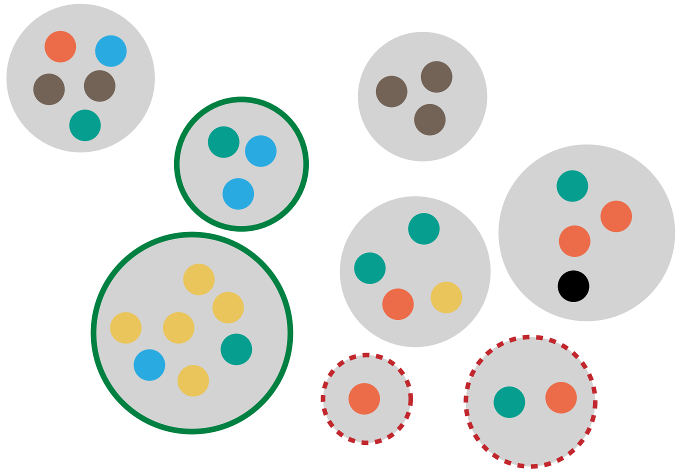
LHC

Cluster contains >= 75% hits from one particle



Double Majority

Cluster contains >= 50% hits from one particle
and
This particle has < 50% of its hits outside



Clusters with < 3 hits or non-reconstructable majority particle are discarded

#reconstructable particles

Perfect efficiency = **1/5**
Perfect fakes = **5/5**

#clusters with >= 3 hits & majority particle reconstructable

LHC efficiency = **2/5**
LHC fakes = **4/6**

#reconstructable particles

DM efficiency = **2/5**
DM fakes = **4/5**

We also evaluate these **metrics at pT thresholds**: pT cut is applied to majority particle of cluster or particle (this is not a truth cut on the data, but simply a efficiency vs pT study)

Reconstructable: >= 3 hits