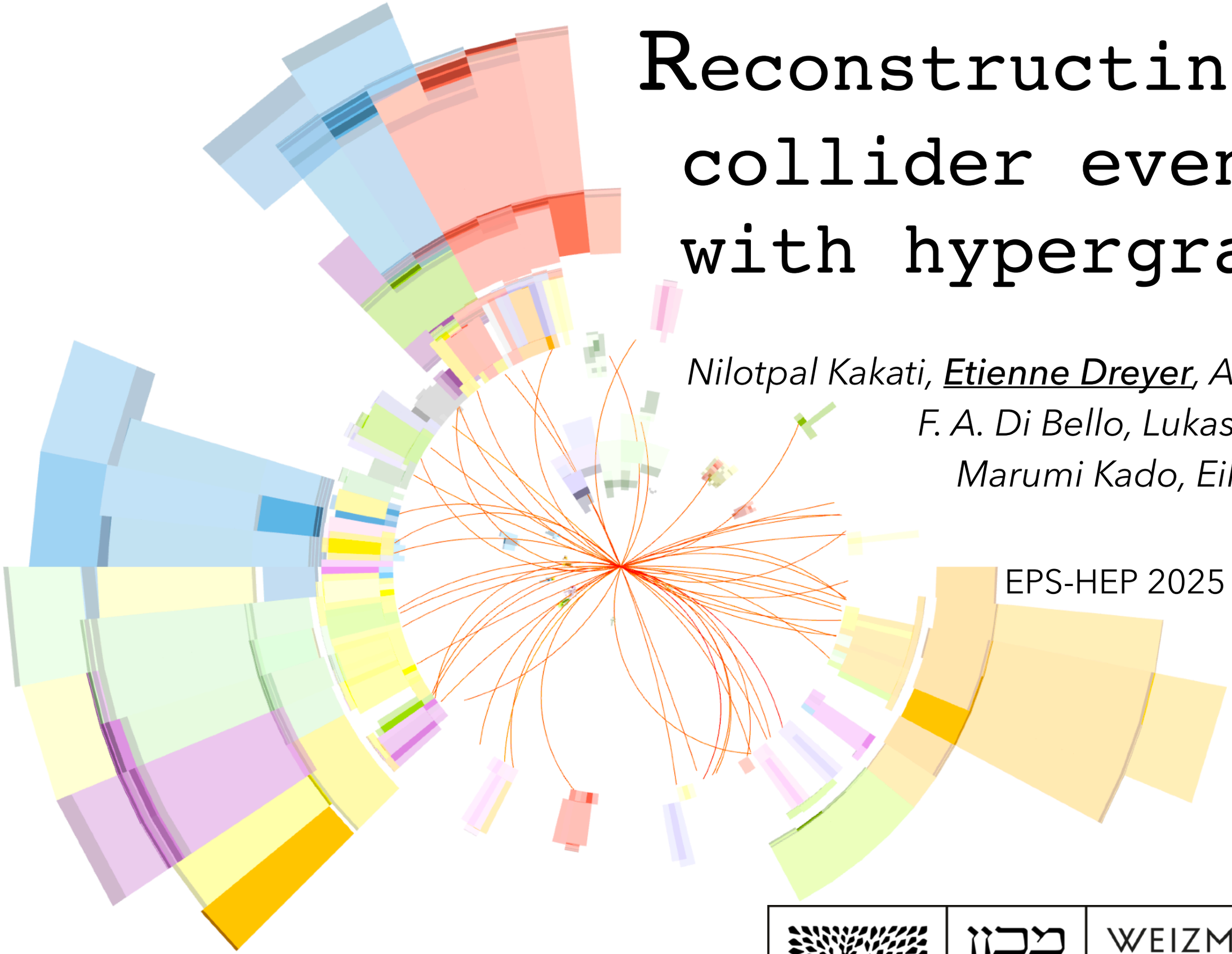
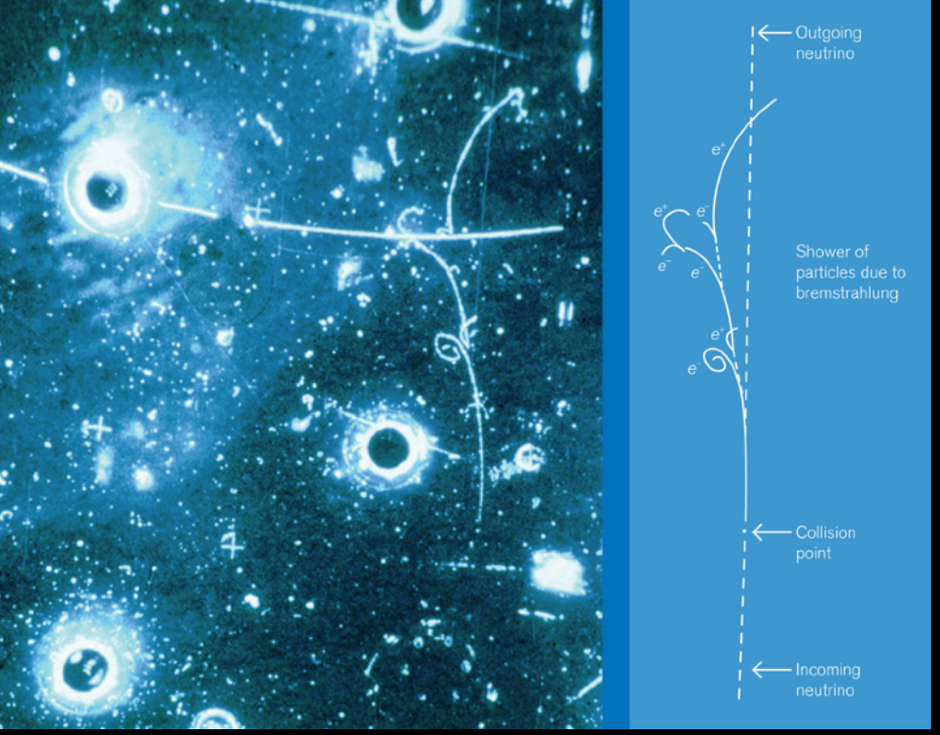


# Reconstructing collider events with hypergraphs

*Nilotpal Kakati, **Etienne Dreyer**, Anna Ivina, F. A. Di Bello, Lukas Heinrich, Marumi Kado, Eilam Gross*

EPS-HEP 2025 · Marseille

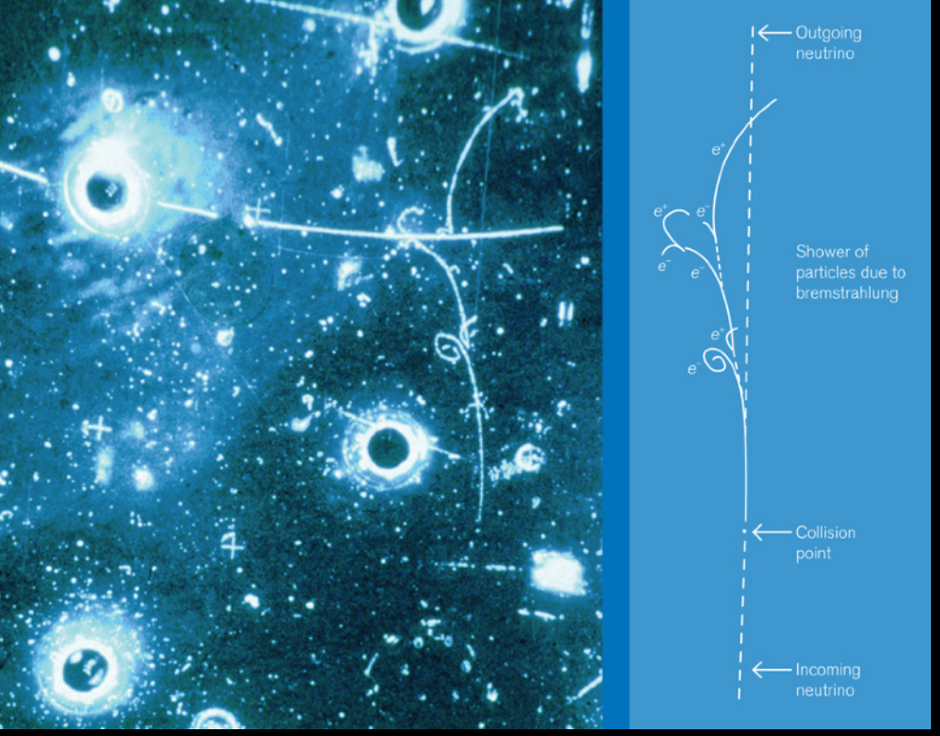
WEIZMANN INSTITUTE OF SCIENCE




## BEBC

1979

Weak neutral current




**BEBC**

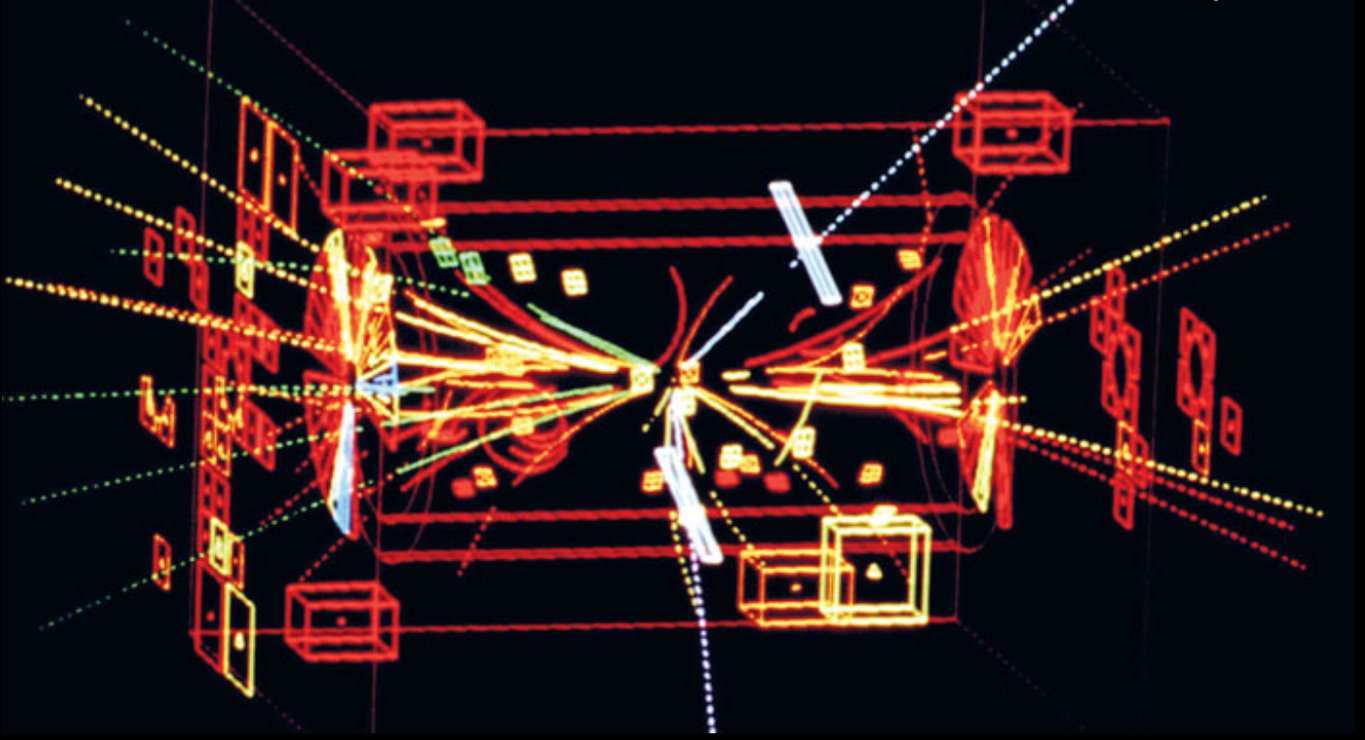
1979

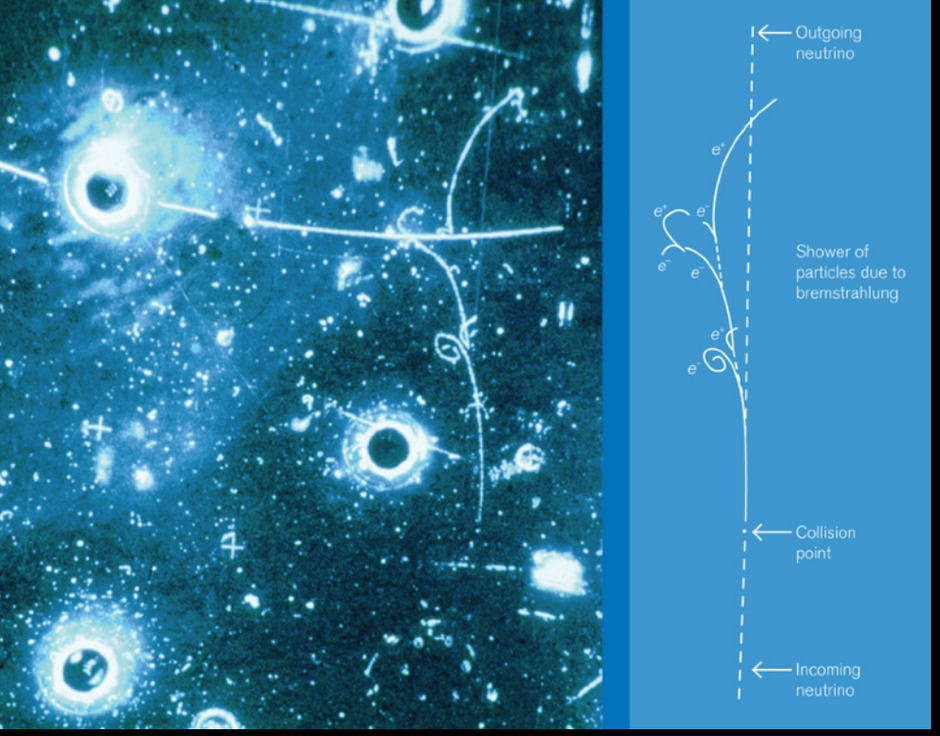
Weak neutral current

**UA1**

1984

W, Z bosons

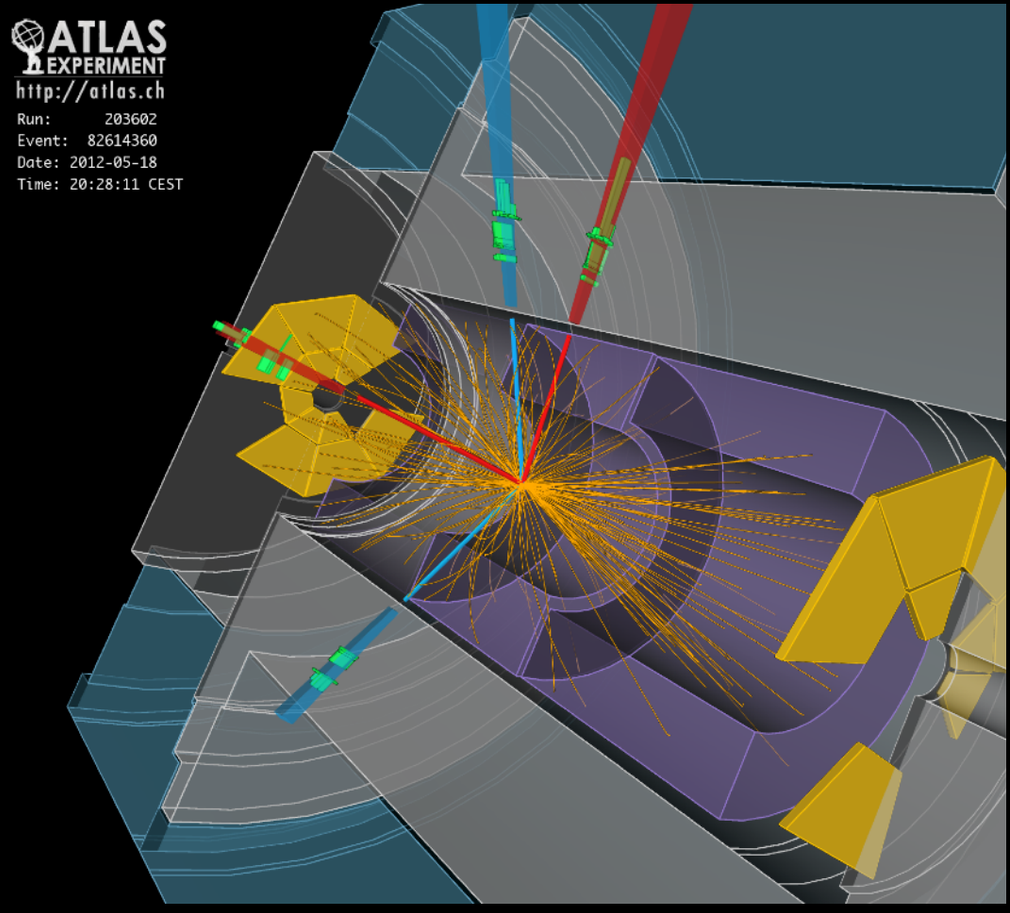





## BEBC

1979

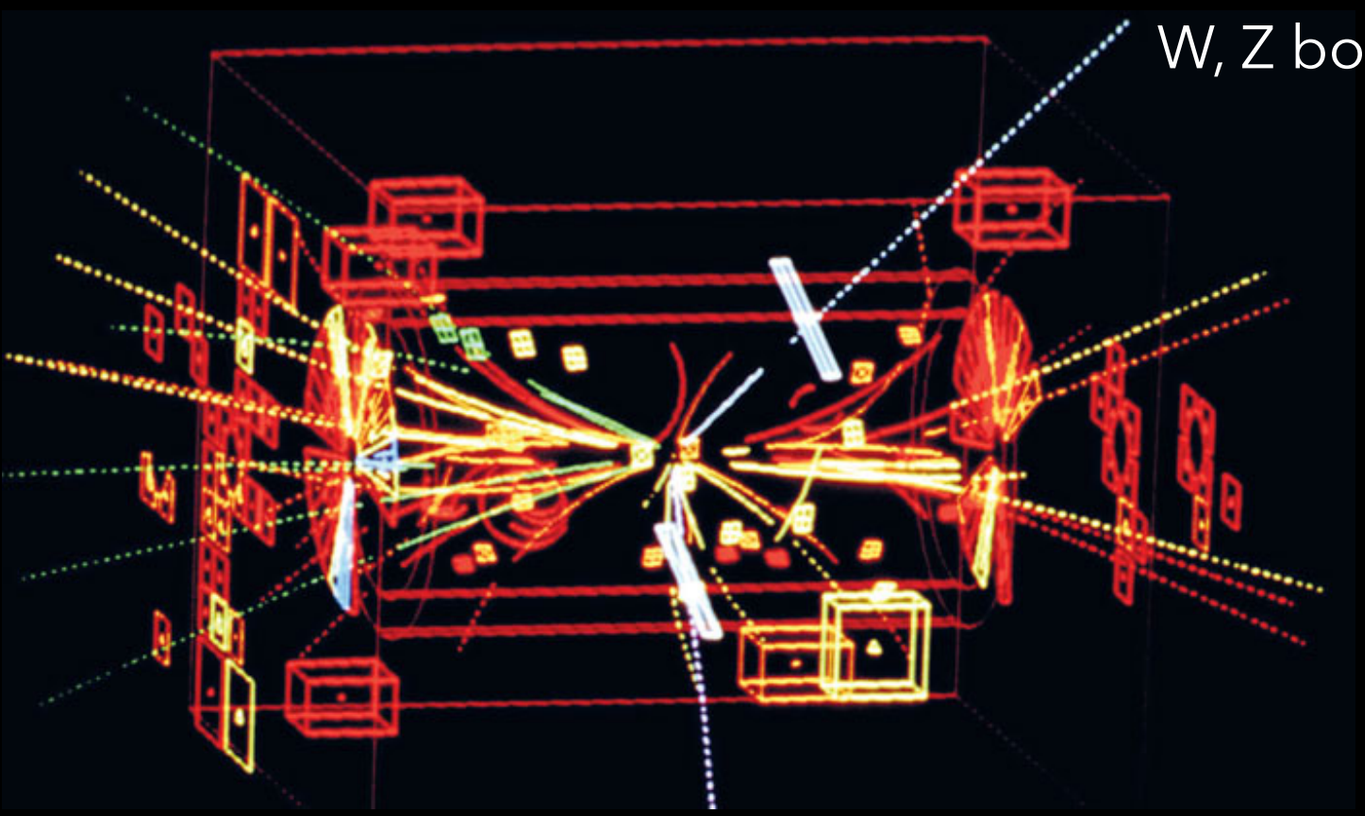
Weak neutral current




## UA1

1984

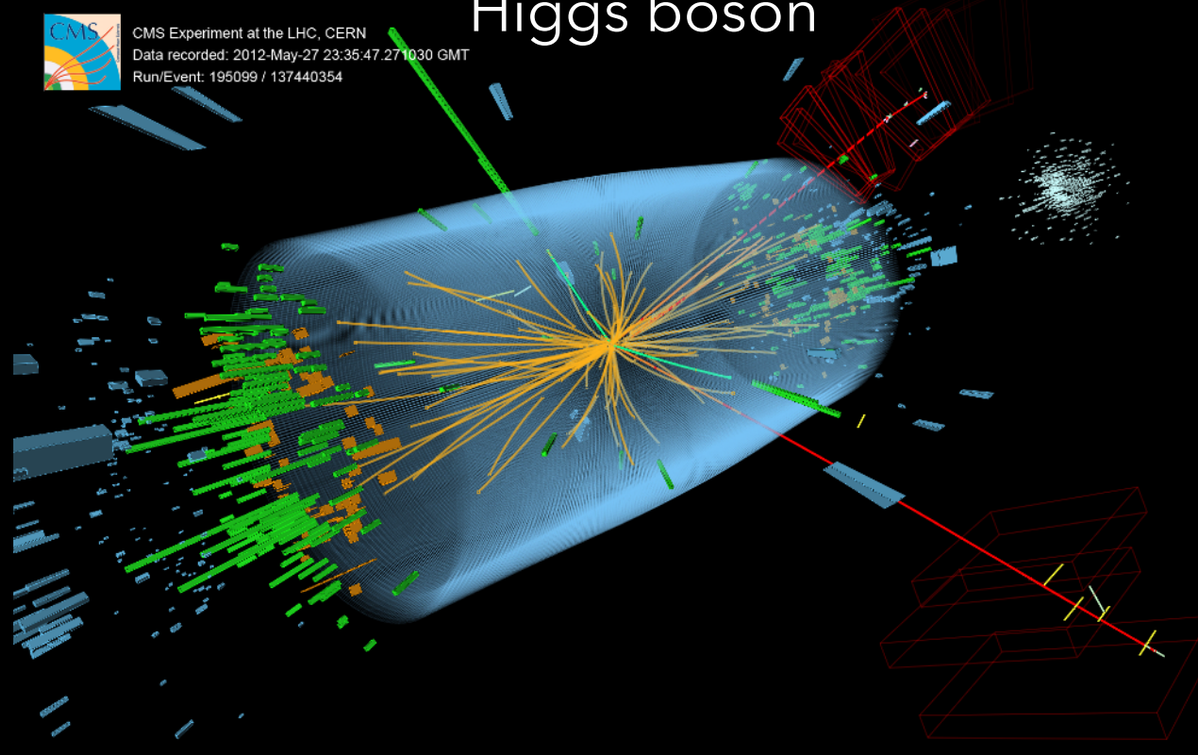
W, Z bosons



## ATLAS & CMS

2013

Higgs boson

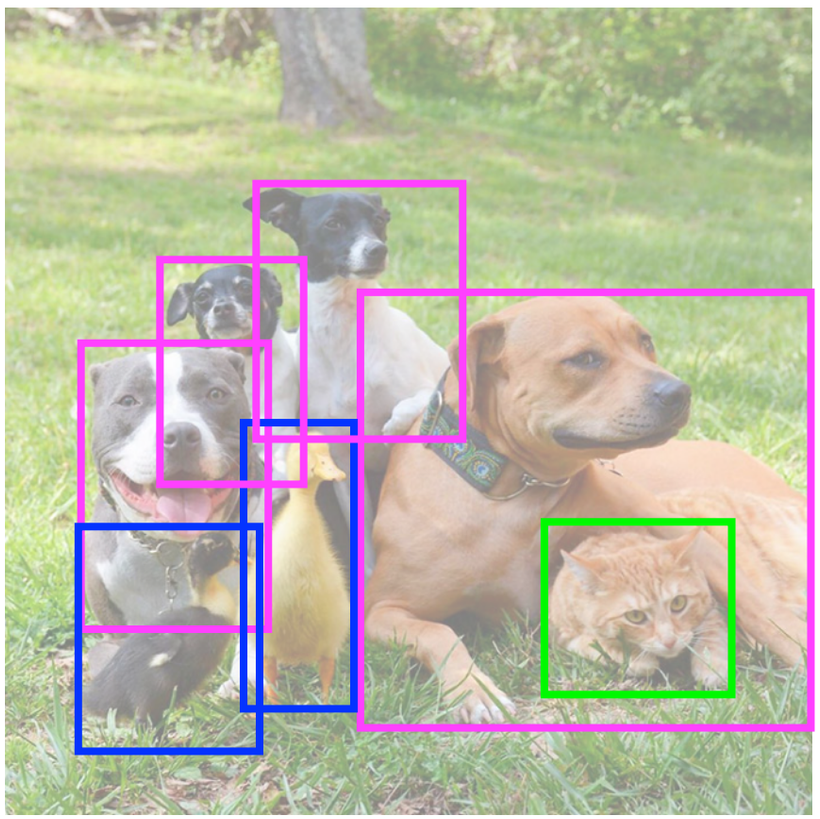


# Classic object detection

Input

Credit: BoredPanda

Output

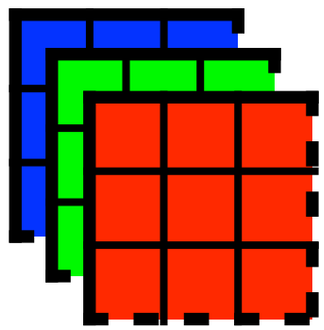


# Classic object detection

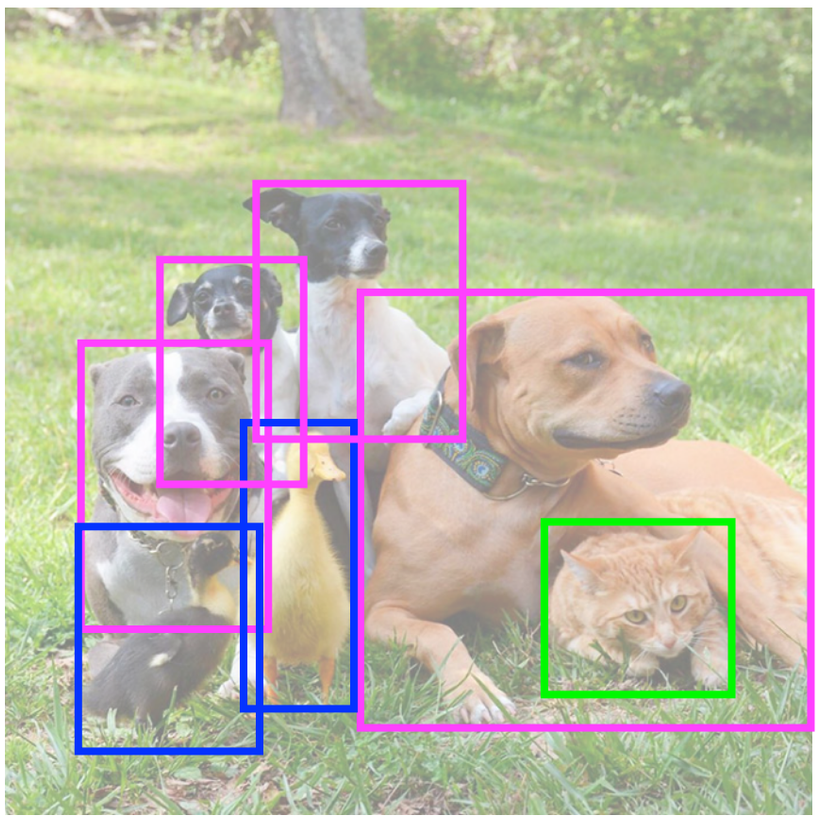
Input

Credit: BoredPanda

Features: `RGB value array`



Output

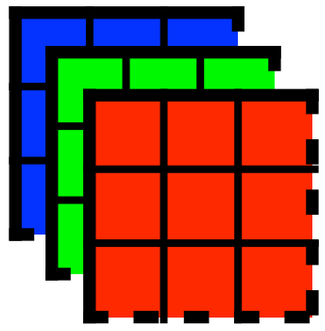


# Classic object detection

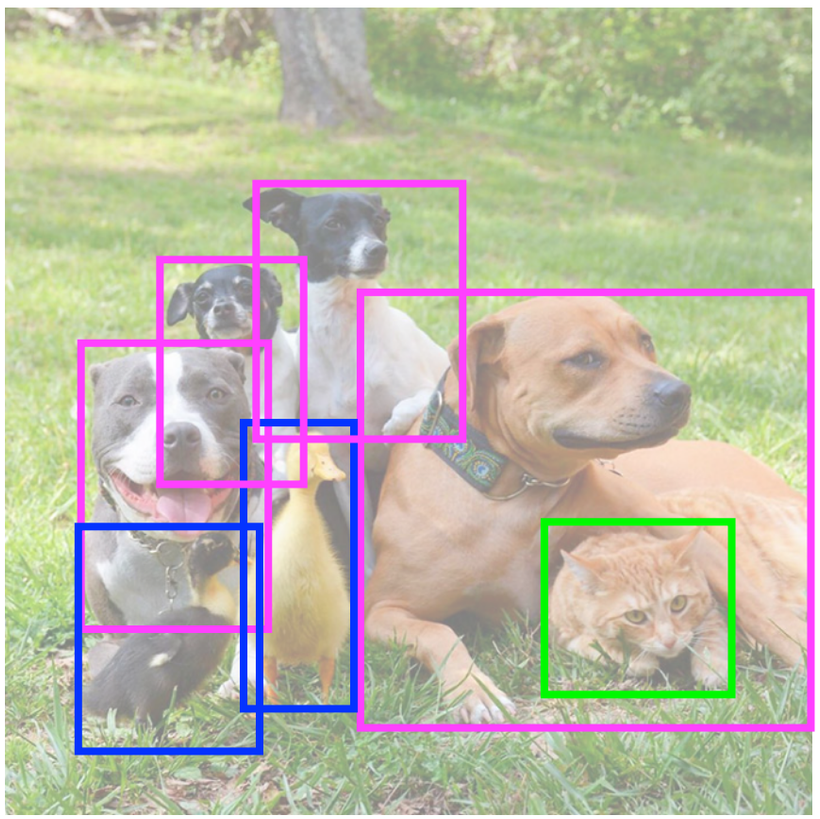
Input

Credit: BoredPanda

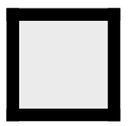
Features: `RGB value array`



Output

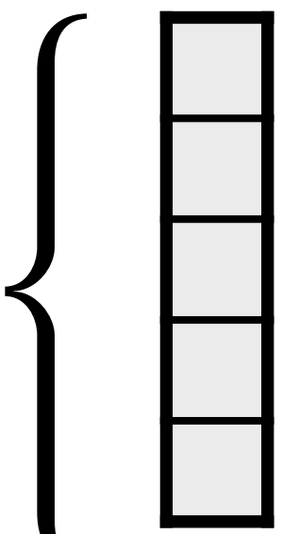


Cardinality



`7`

Class



`person`

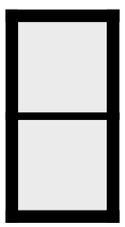
`dog`

`cat`

`duck`

`tree`

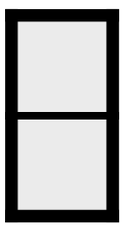
Position



`x`

`Y`

Size

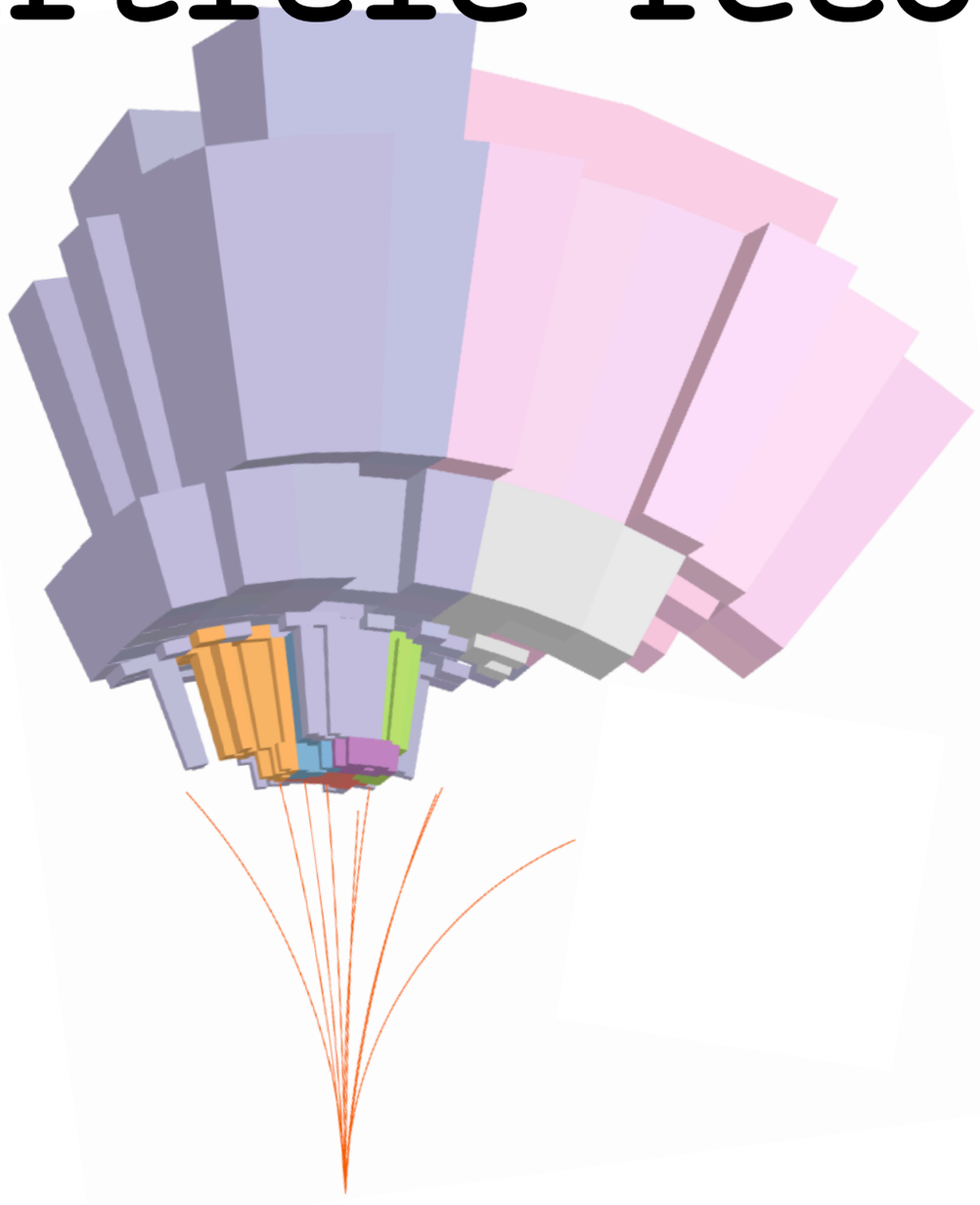


`h`

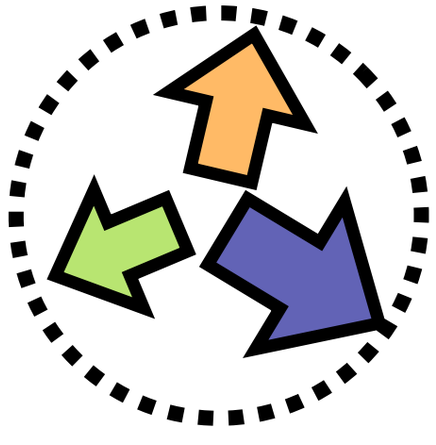
`w`

# Particle reconstruction

Input

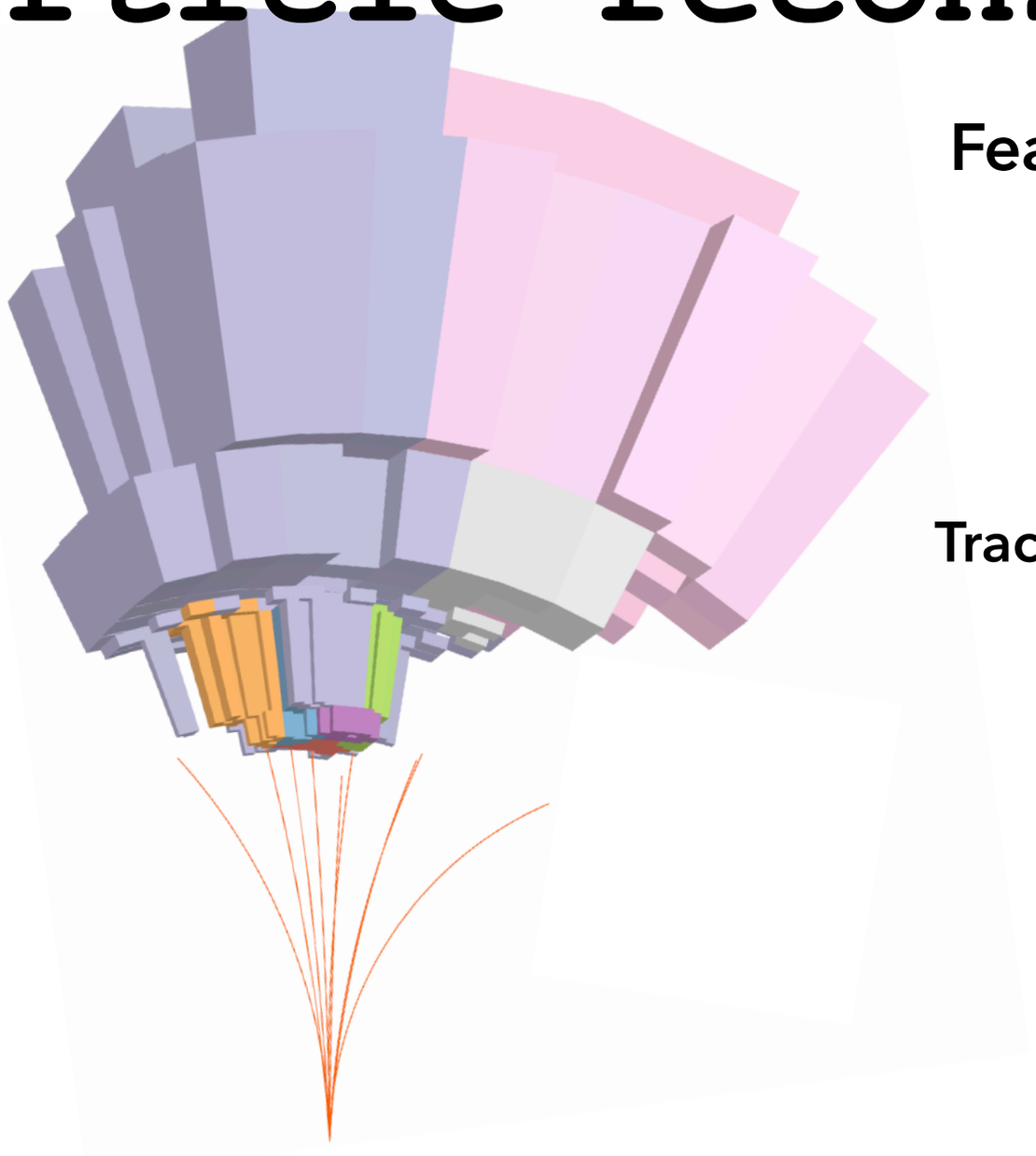


Output

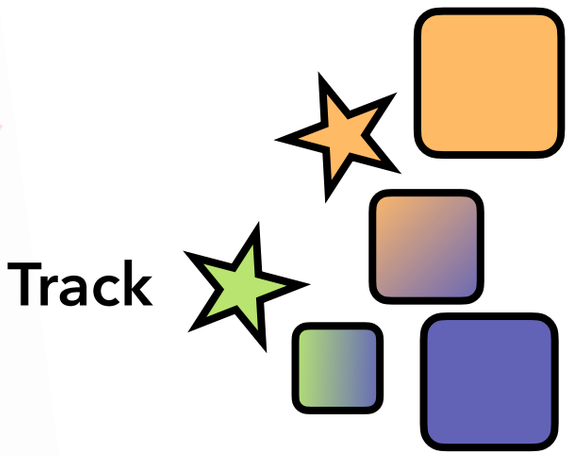


# Particle reconstruction

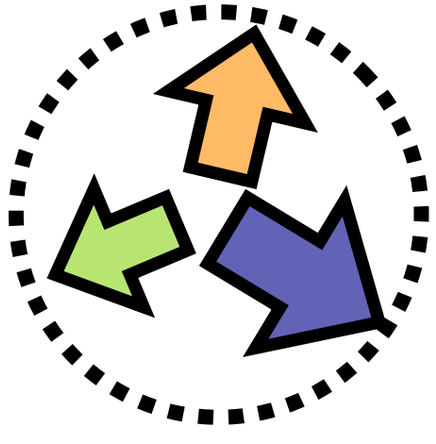
Input



**Features:** `[energy, location, …]`

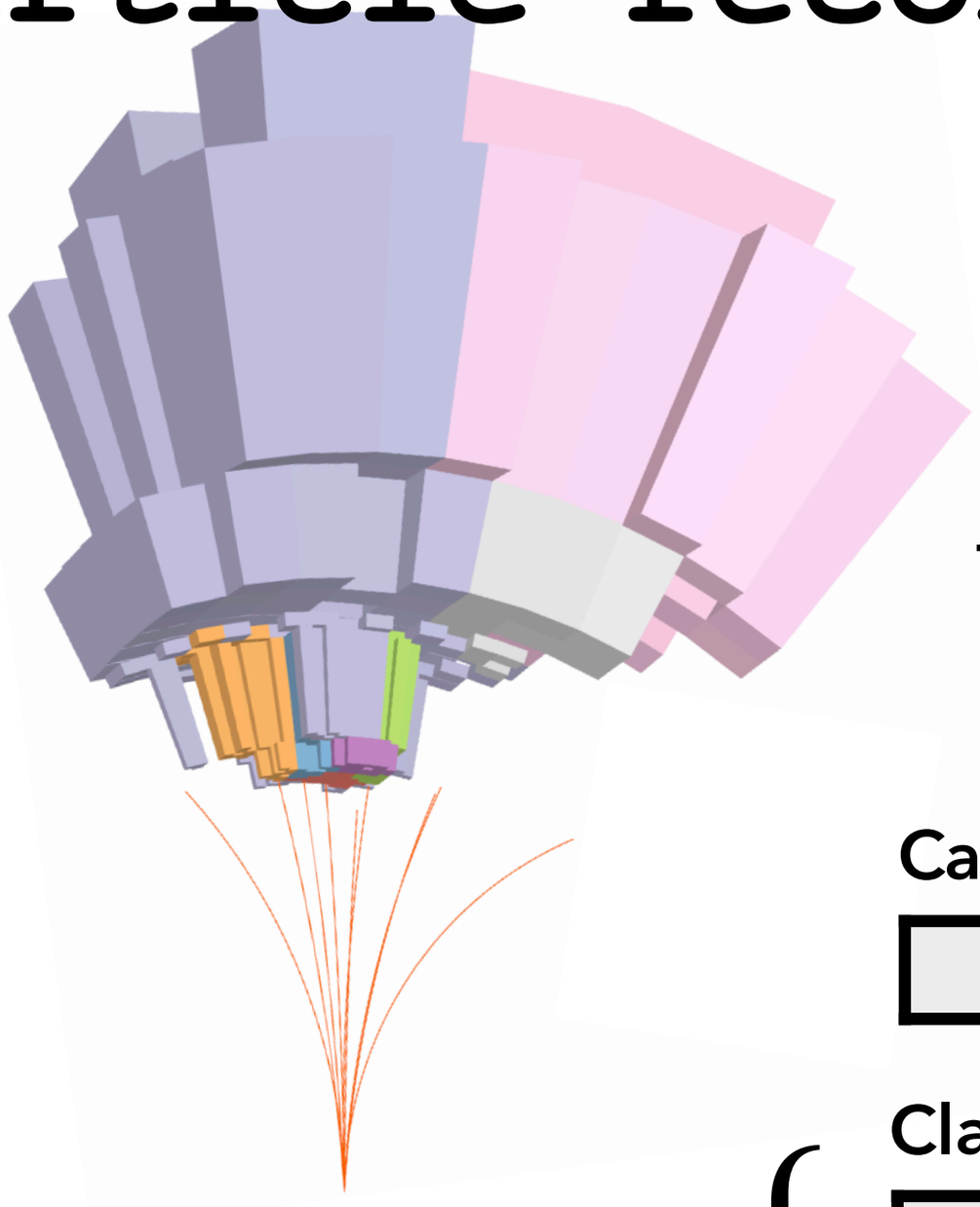



Output

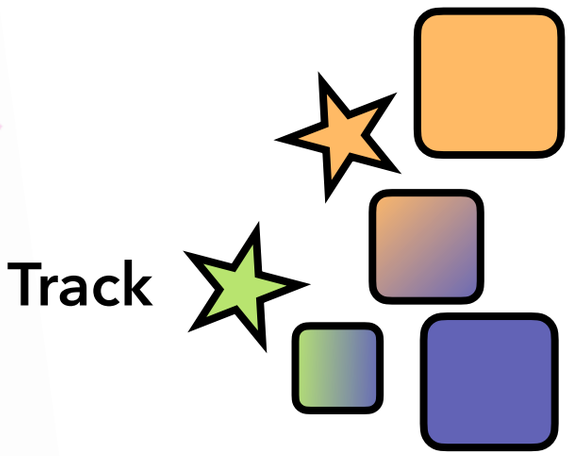


# Particle reconstruction

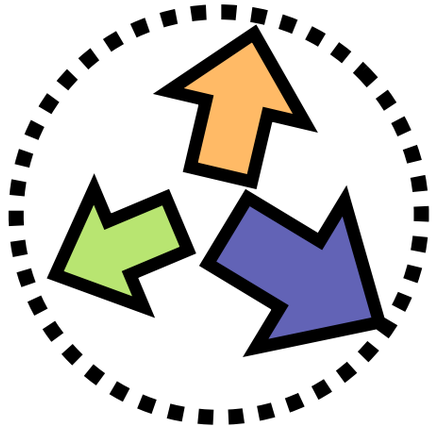
Input



Features: `[energy, location, …]`




Output

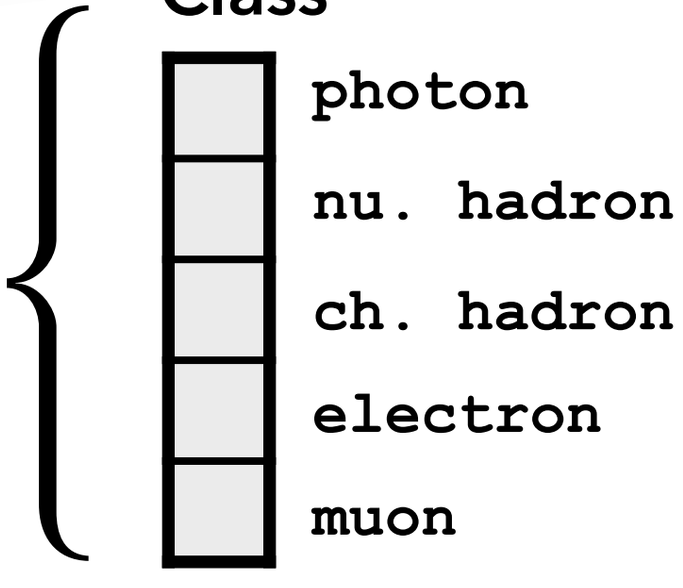


Cardinality

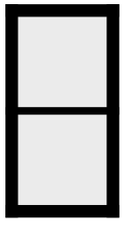


3

Class



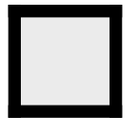

Direction



eta

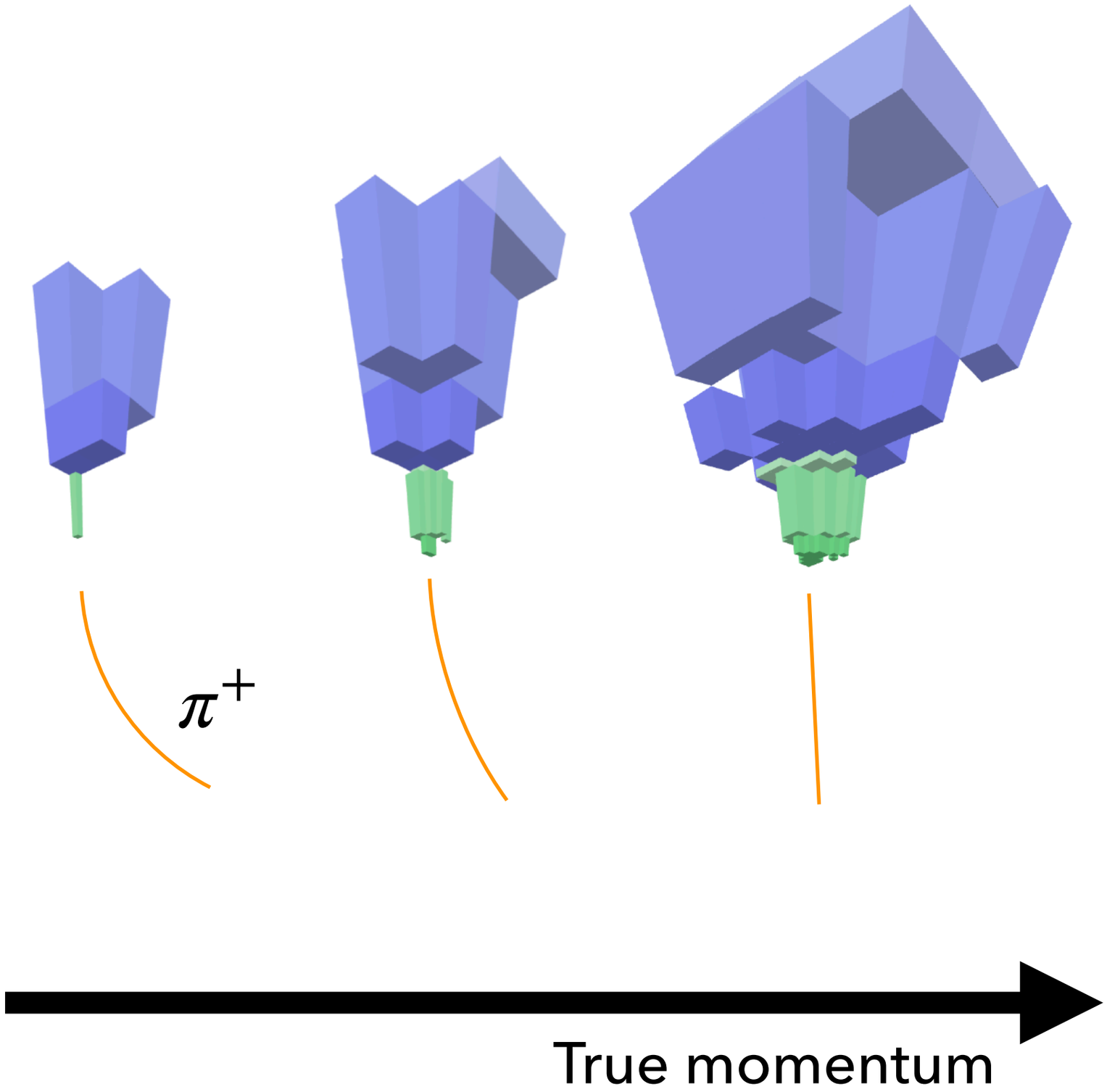
phi

Momentum



pT

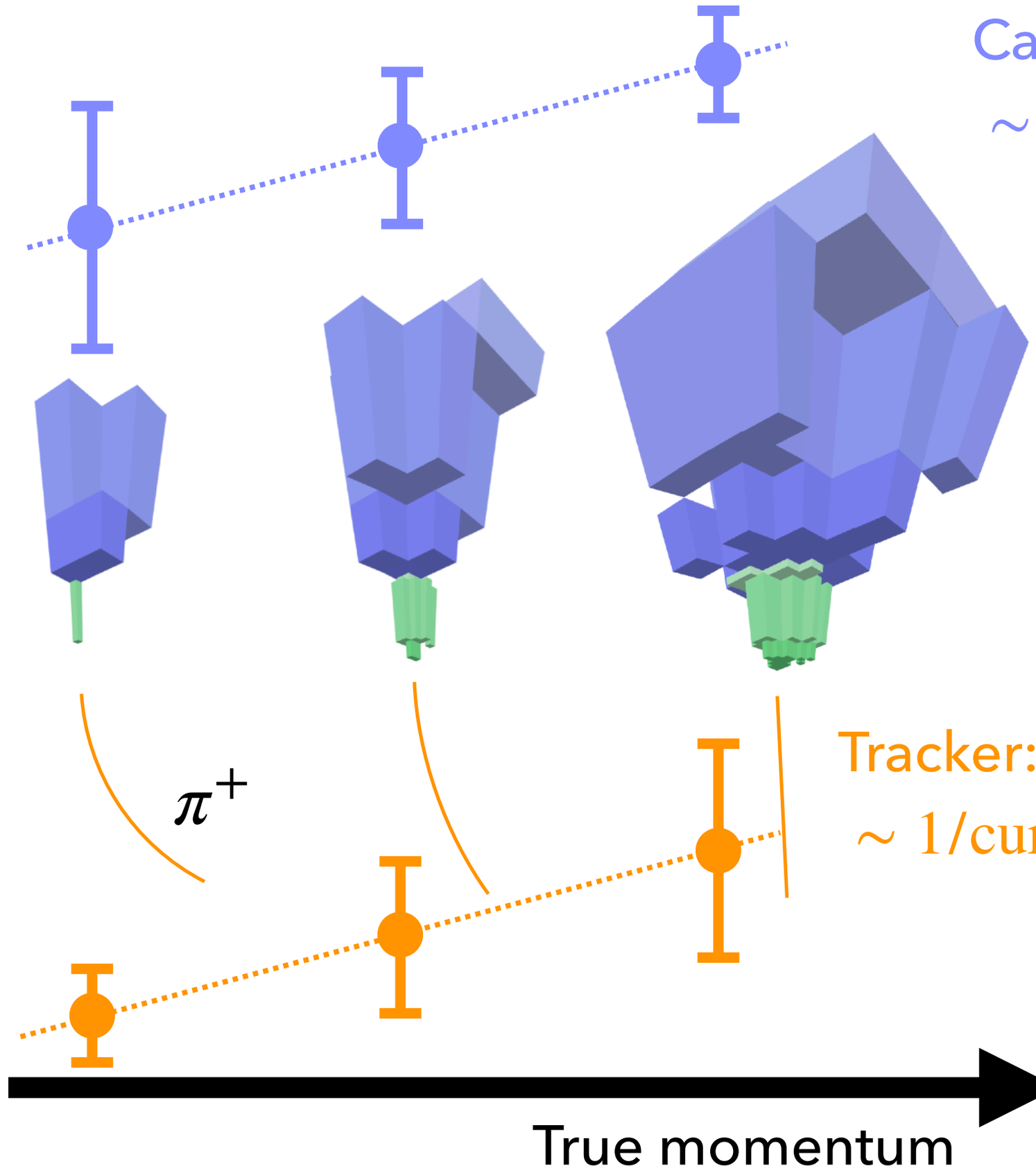
# Particle momentum measurement




## “Particle flow”

An algorithm that combines the information from both tracker and calorimeter to optimize the momentum prediction

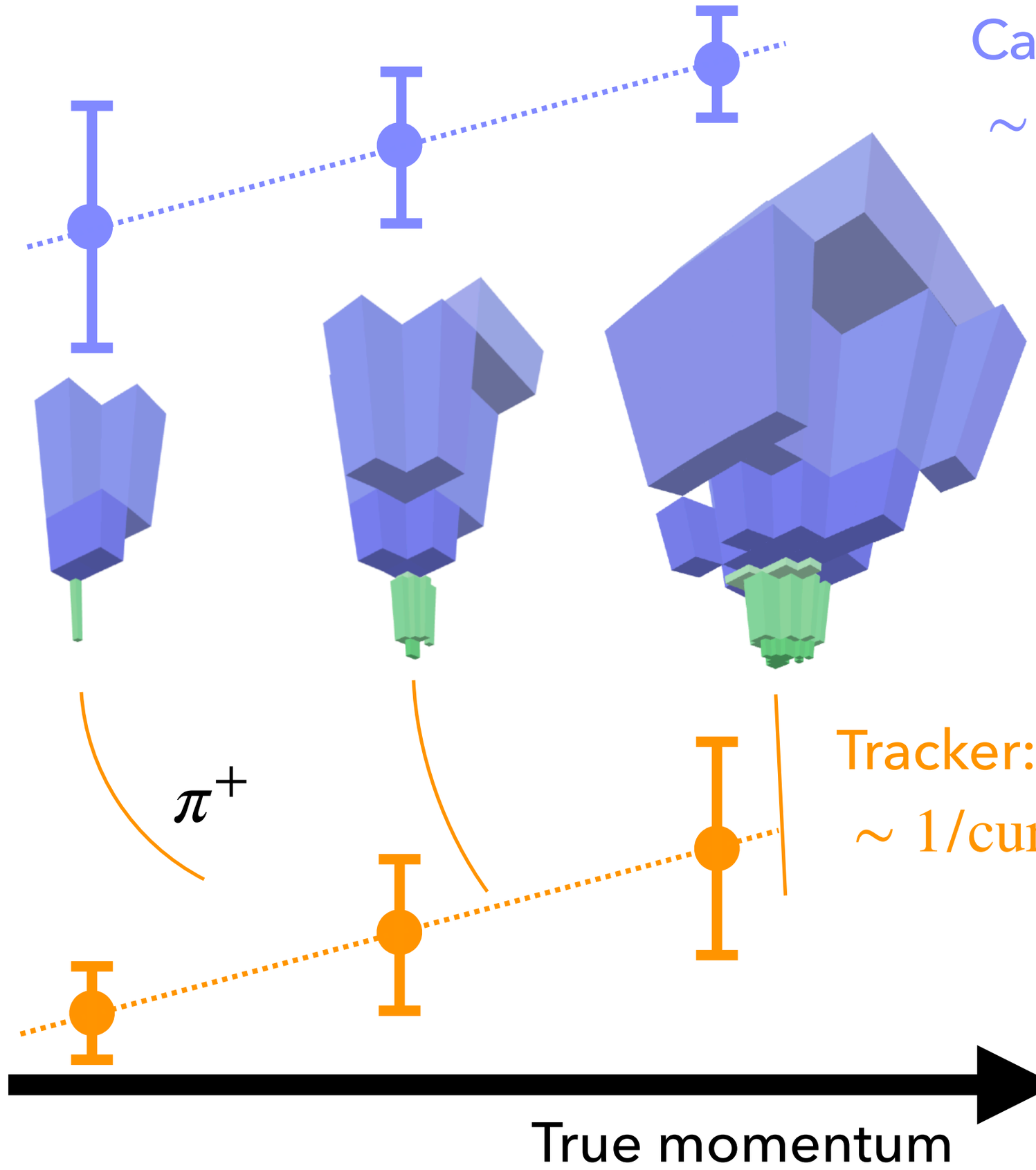
# Particle momentum measurement




## "Particle flow"

An algorithm that combines the information from both tracker and calorimeter to optimize the momentum prediction

# Particle momentum measurement




Calorimeter:

$$\sim \sum_{\text{cells}} E_i$$

## "Particle flow"

An algorithm that combines the information from both tracker and calorimeter to optimize the momentum prediction

***N.B. cannot naively "add" tracks!***

# The core challenge: overlap

How to correctly split clusters among the tracks to disentangle particles

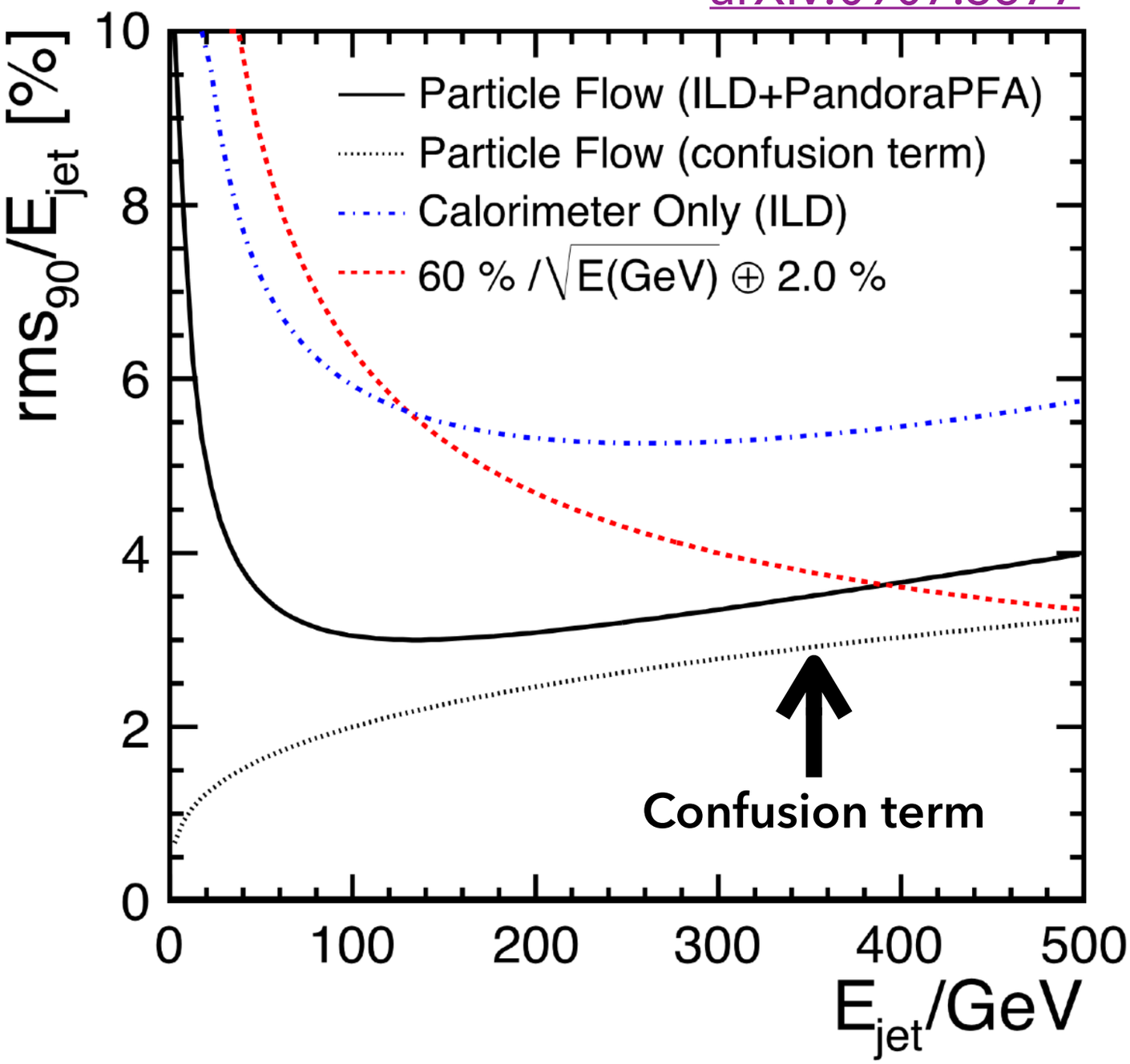
# The core challenge: overlap

How to correctly split clusters among the tracks to disentangle particles

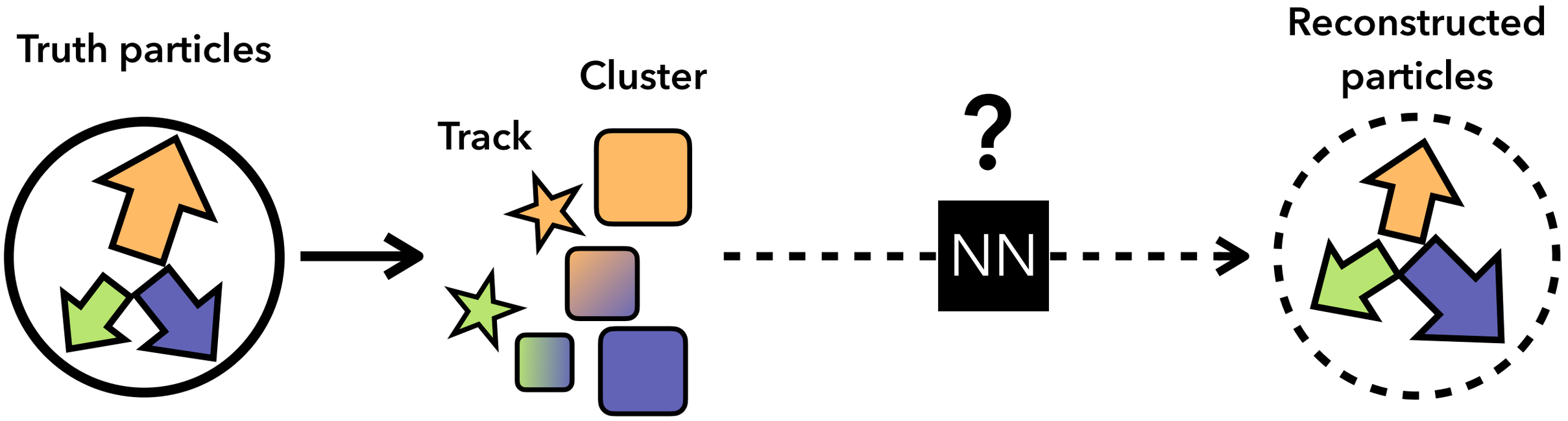
# The core challenge: overlap

How to correctly split clusters among the tracks to disentangle particles

arXiv:0907.3577



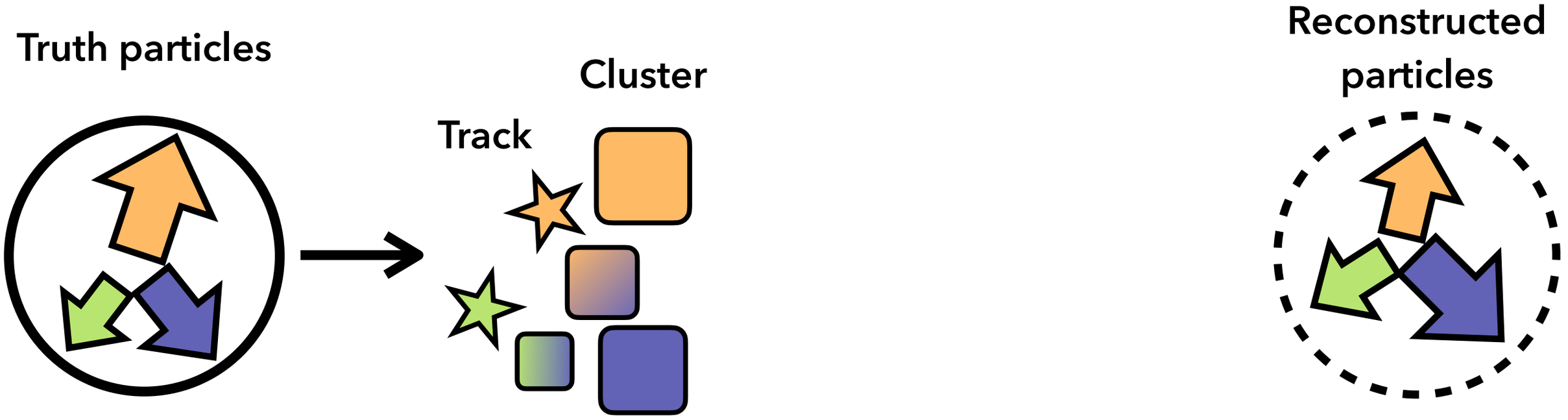
# Set-to-set ML architecture



# Benchmark: MLPF

arXiv:2101.08578 , arXiv:2309.06782

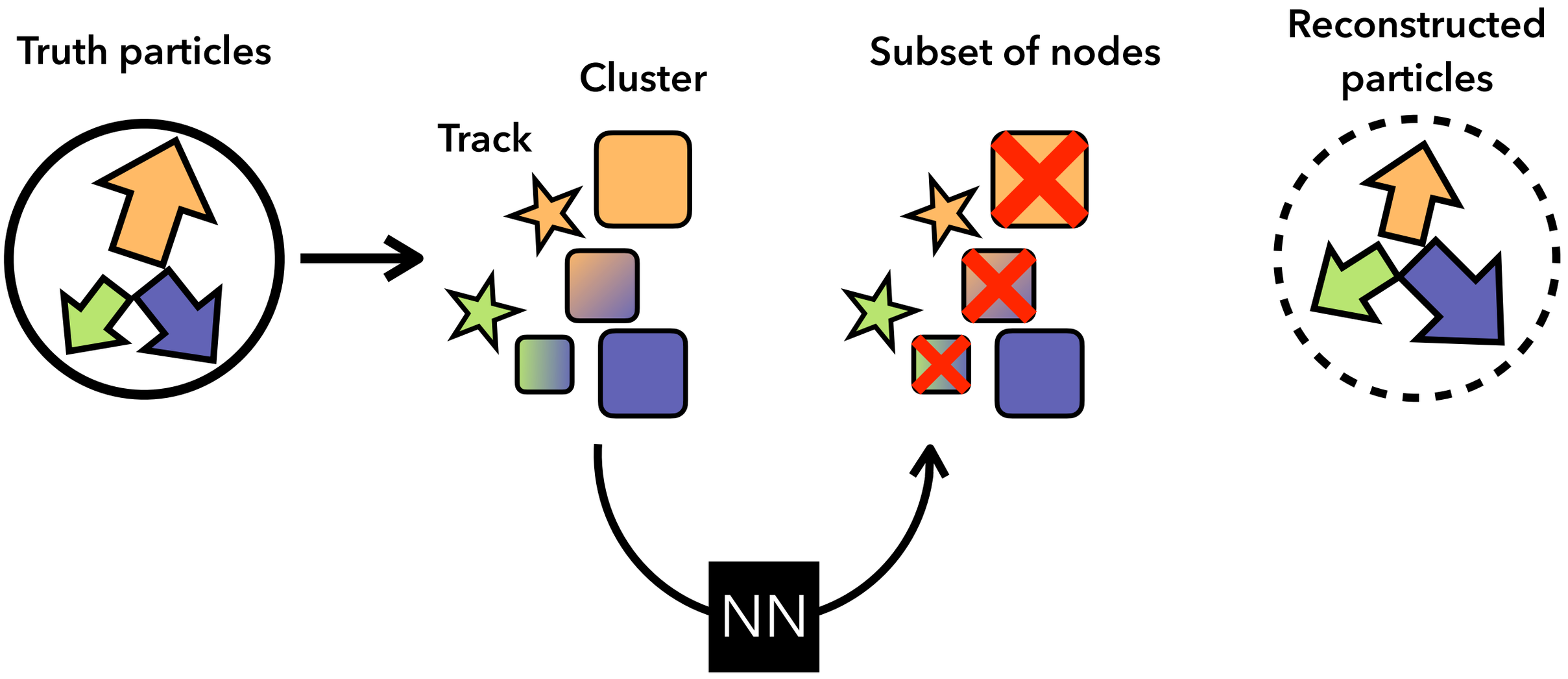
See Farouk's talk yesterday



# Benchmark: MLPF

arXiv:2101.08578 , arXiv:2309.06782

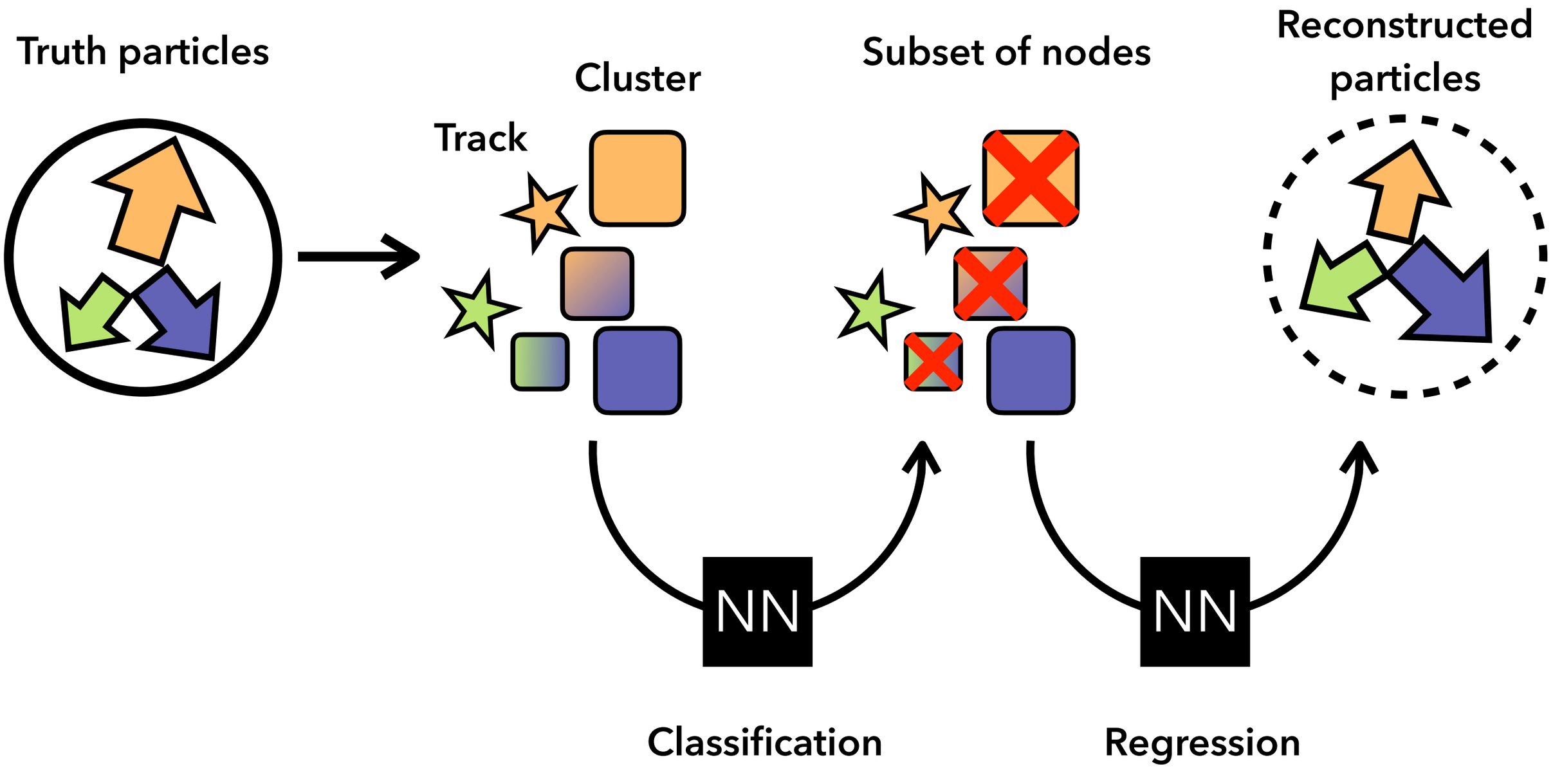
See Farouk's talk yesterday



# Benchmark: MLPF

arXiv:2101.08578 , arXiv:2309.06782

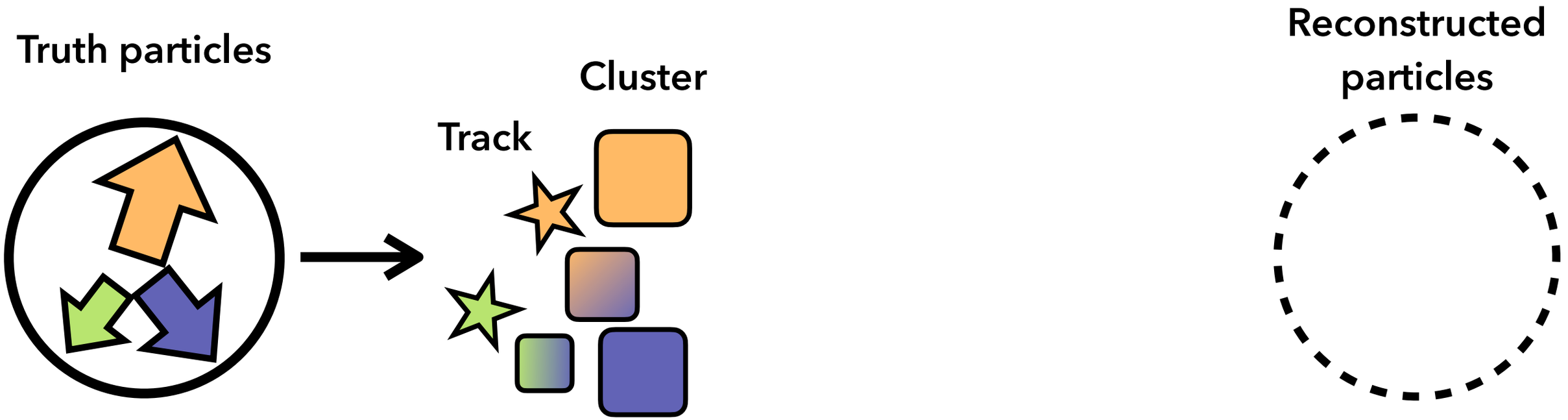
See Farouk's talk yesterday




**[ N.B. *in practice the tasks are simultaneous* ]**

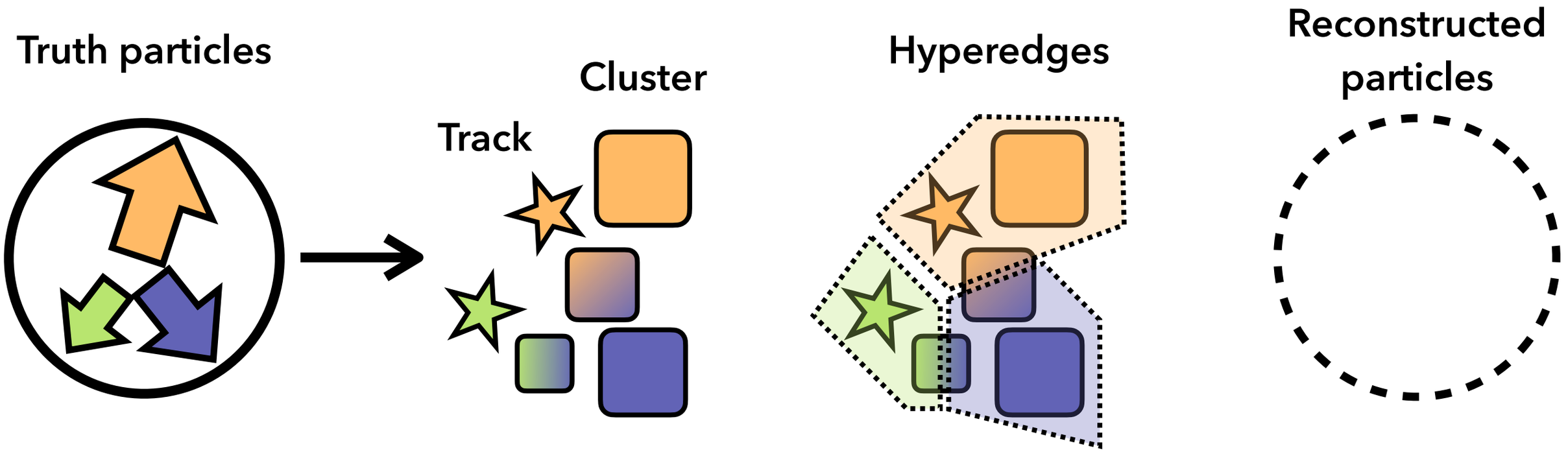
# Ours: HGPflow

arXiv:2212.01328 , arXiv:2410.23236



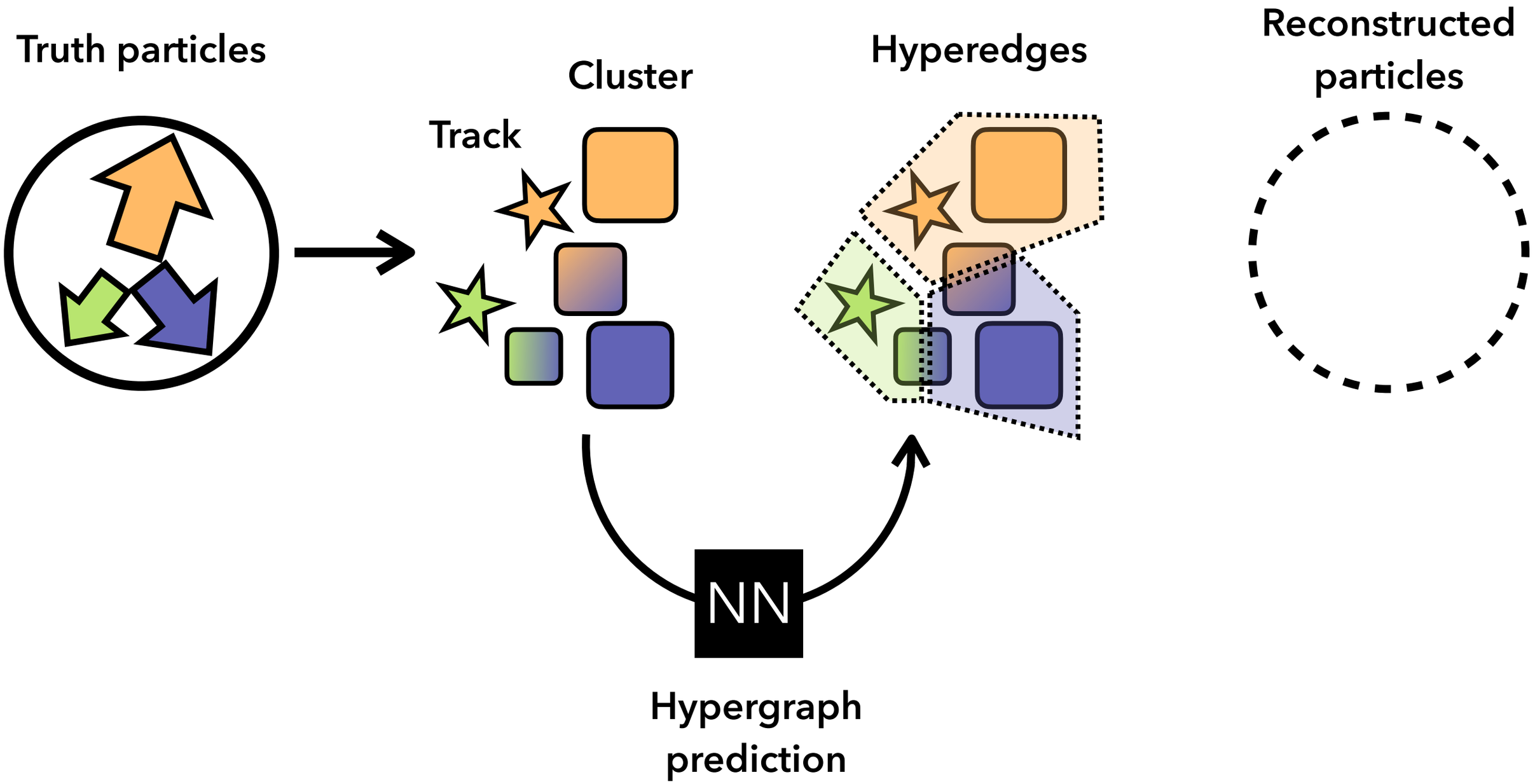
# Ours: HGPflow

arXiv:2212.01328 , arXiv:2410.23236



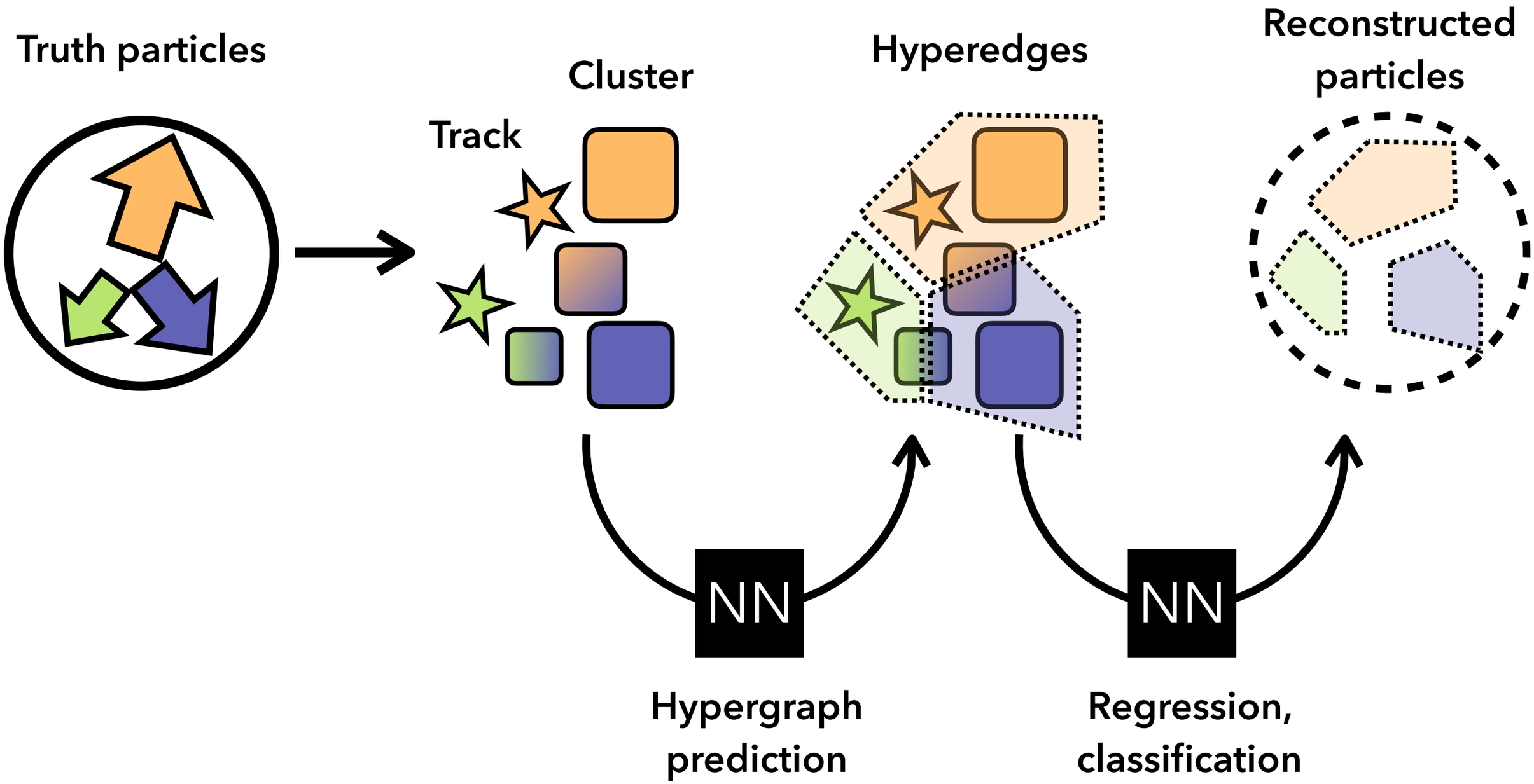
# Ours: HGPflow

arXiv:2212.01328 , arXiv:2410.23236

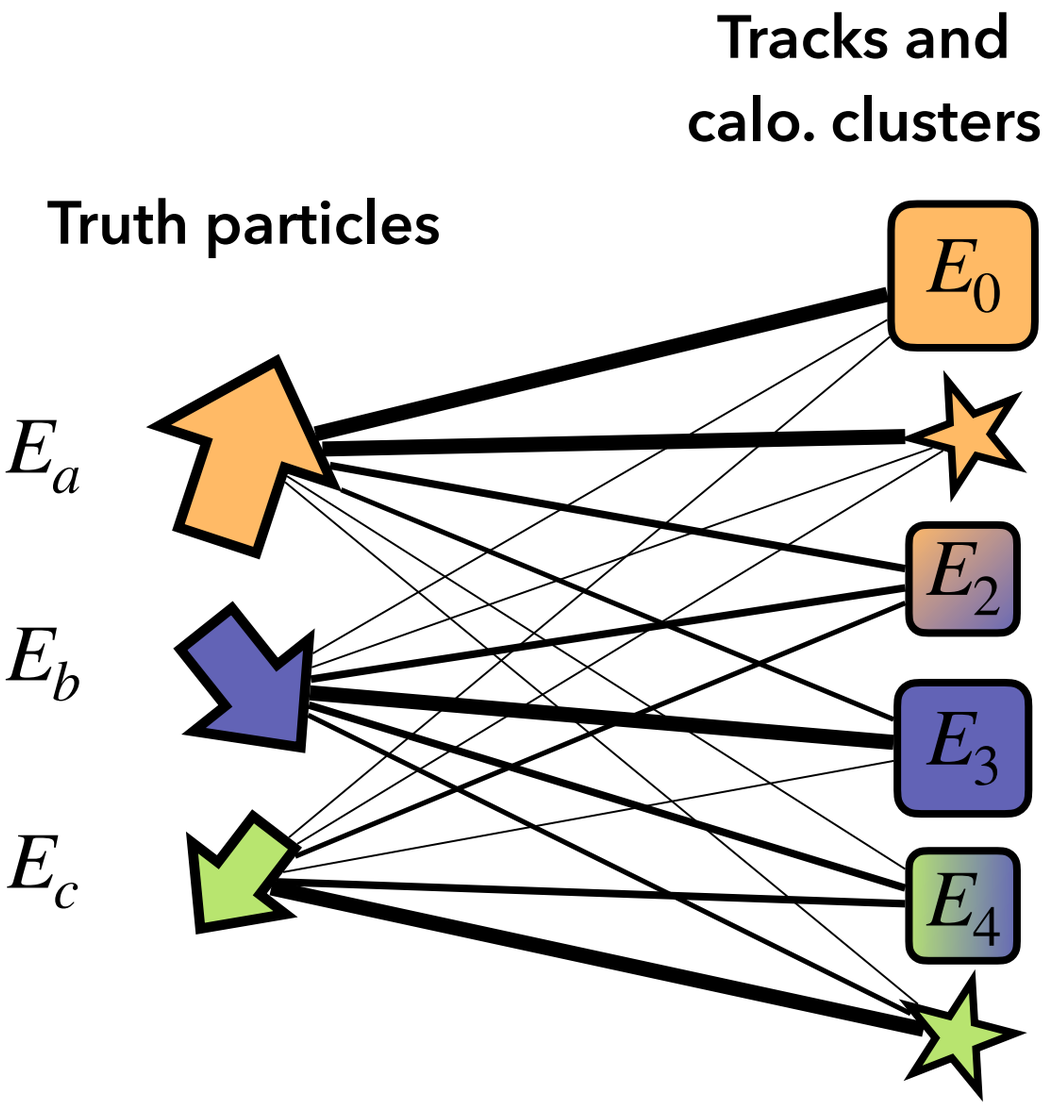


# Ours: HGPflow

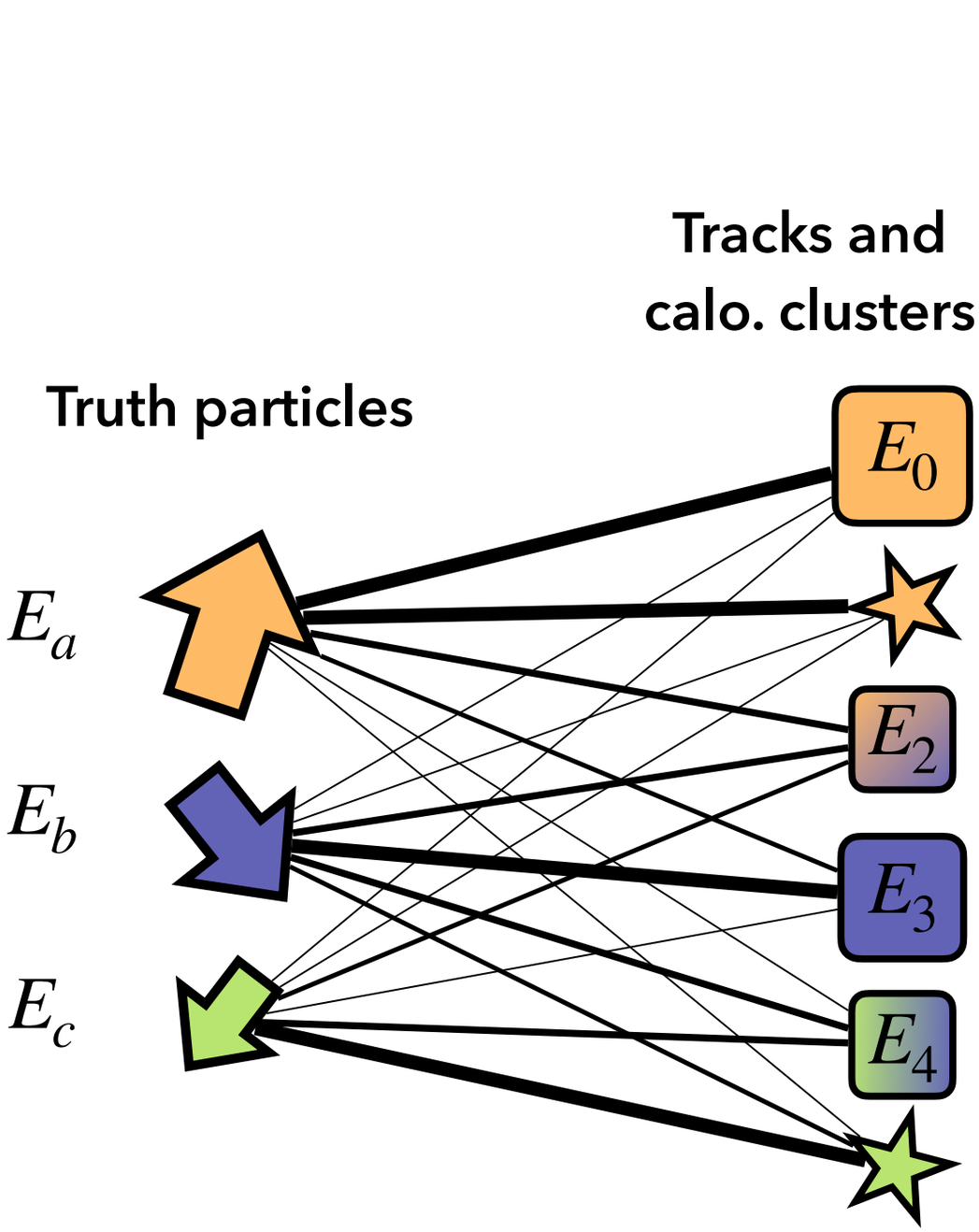
arXiv:2212.01328 , arXiv:2410.23236



# How to predict a hypergraph?



# How to predict a hypergraph?




Incidence matrix

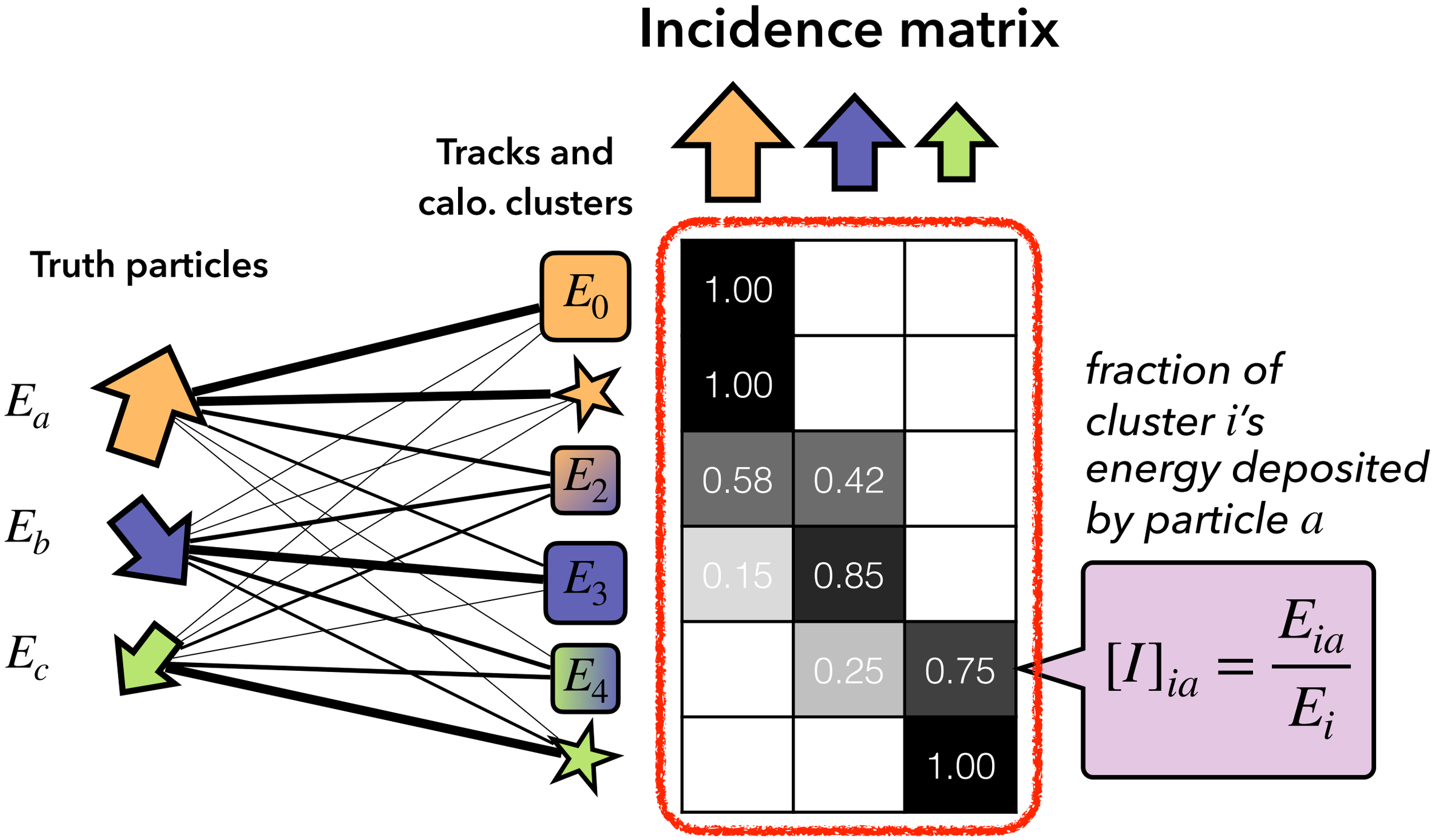
Tracks and calo. clusters

| | | |
|---|---|---|
| 1.00 | | |
| 1.00 | | |
| 0.58 | 0.42 | |
| 0.15 | 0.85 | |
| | 0.25 | 0.75 |
| | | 1.00 |

*fraction of cluster $i$'s energy deposited by particle $a$*

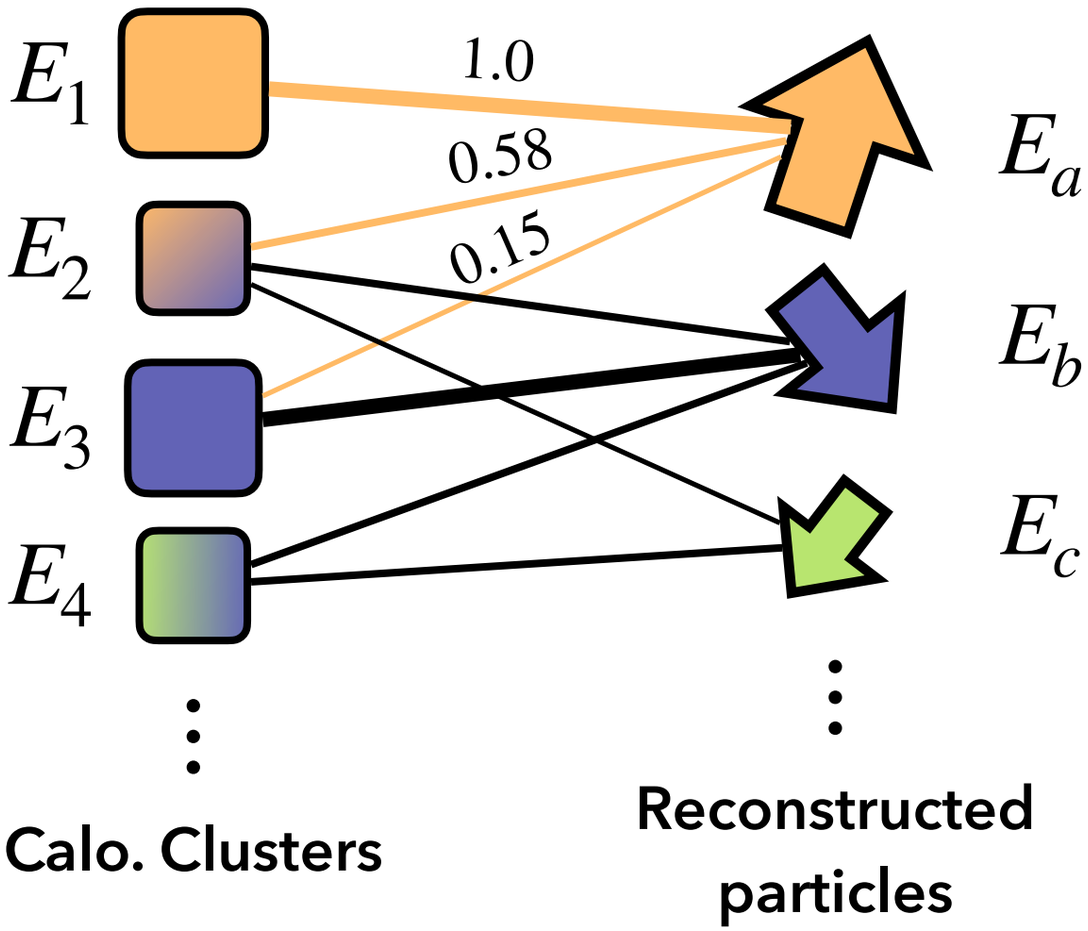
$$[I]_{ia} = \frac{E_{ia}}{E_i}$$

# How to predict a hypergraph?



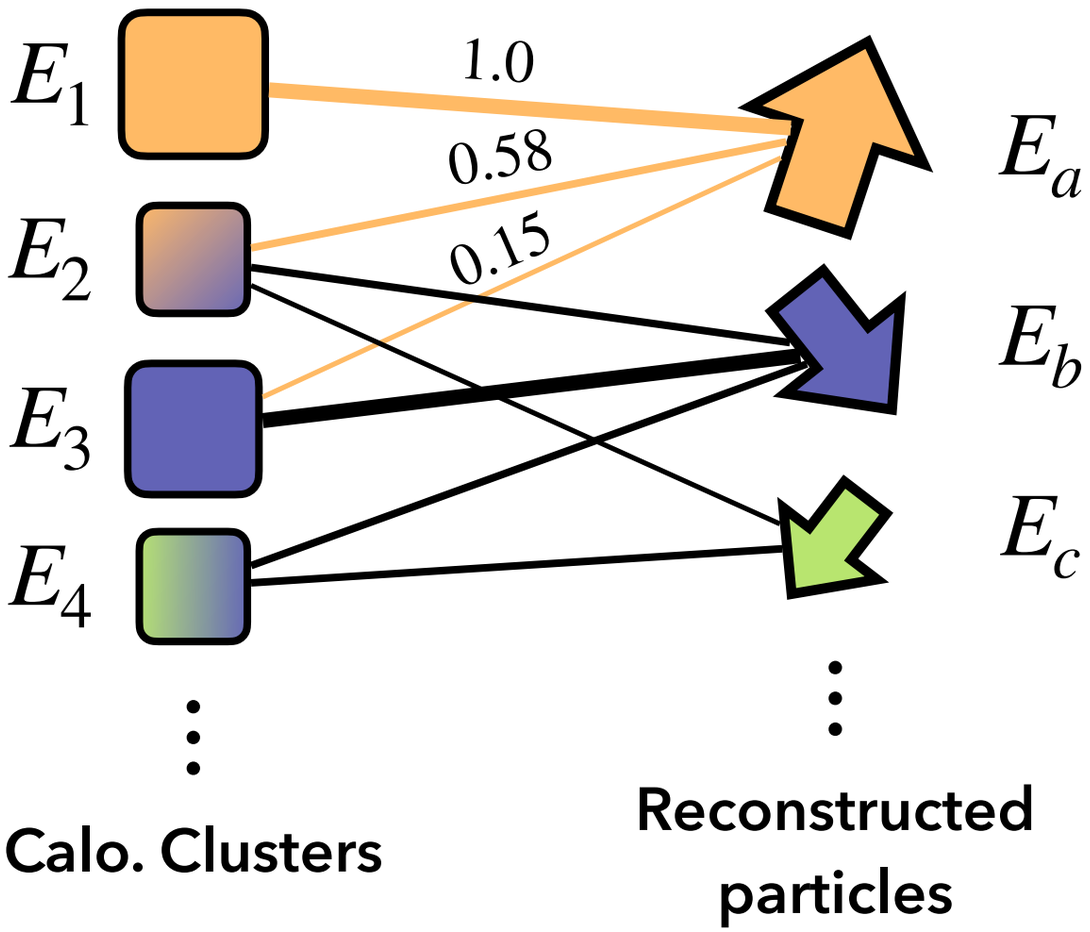
# Perks of learning incidence matrix

*Assuming we predicted the incidence matrix correctly…*



# Perks of learning incidence matrix

*Assuming we predicted the incidence matrix correctly…*

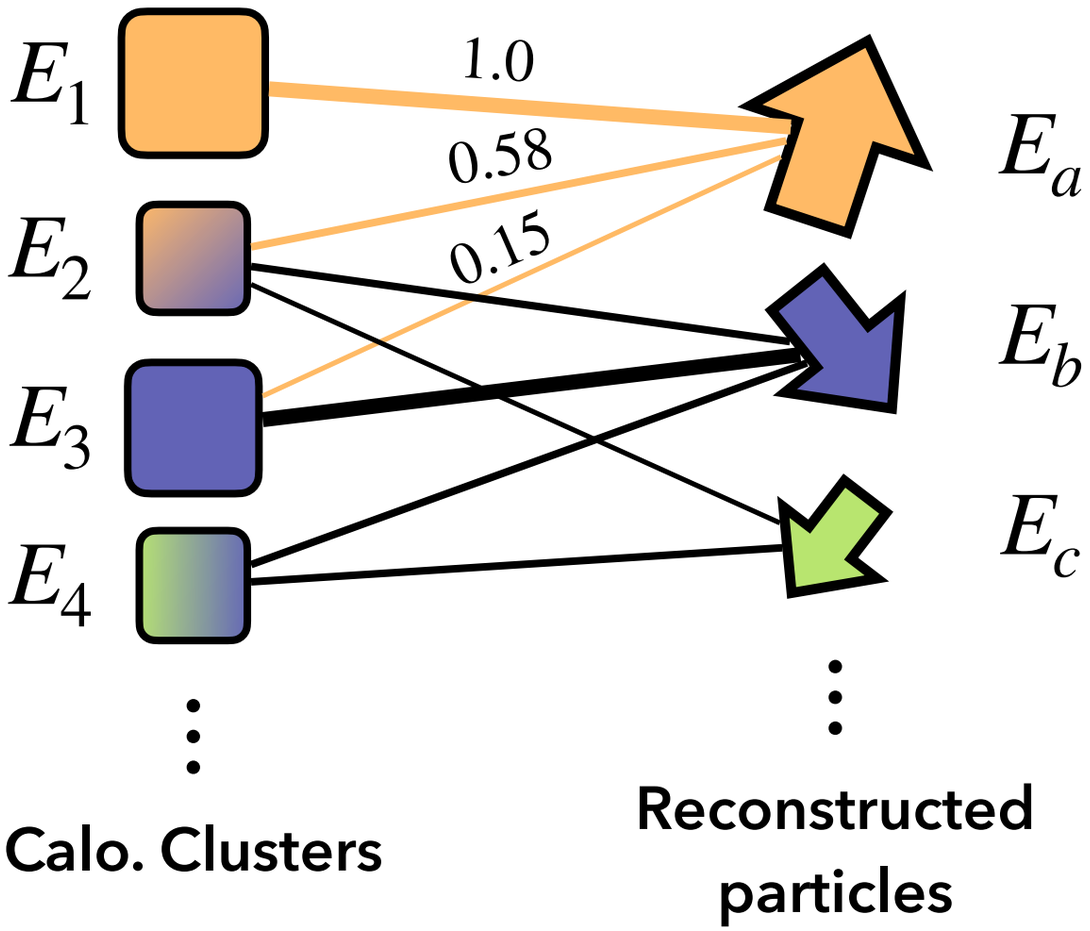



*… then we can already estimate the properties of the particles:*

$$E_a \simeq E_1 + (0.58 \cdot E_2) + (0.15 \cdot E_3)$$

# Perks of learning incidence matrix

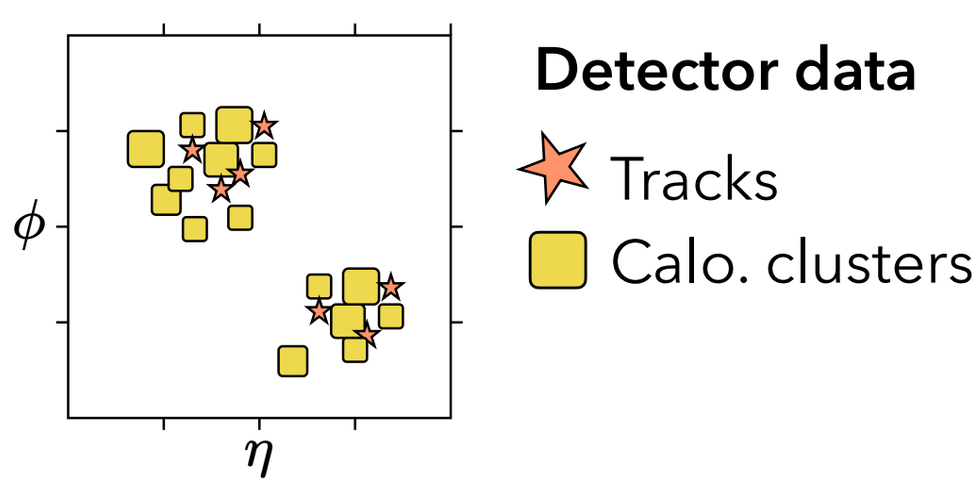
*Assuming we predicted the incidence matrix correctly…*



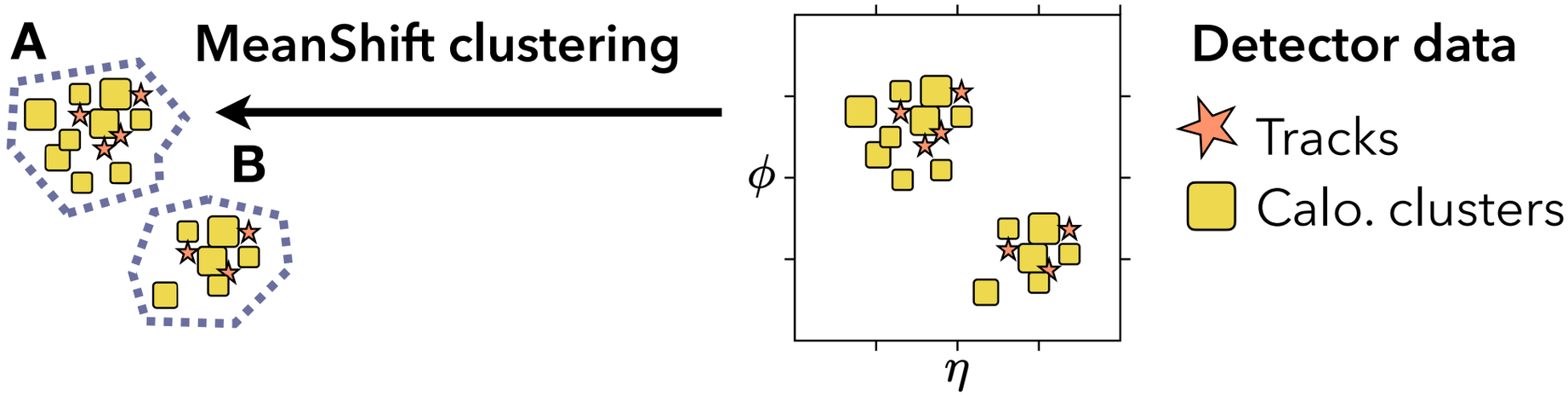

*… then we can already estimate the properties of the particles:*

$$E_a \simeq E_1 + (0.58 \cdot E_2) + (0.15 \cdot E_3)$$

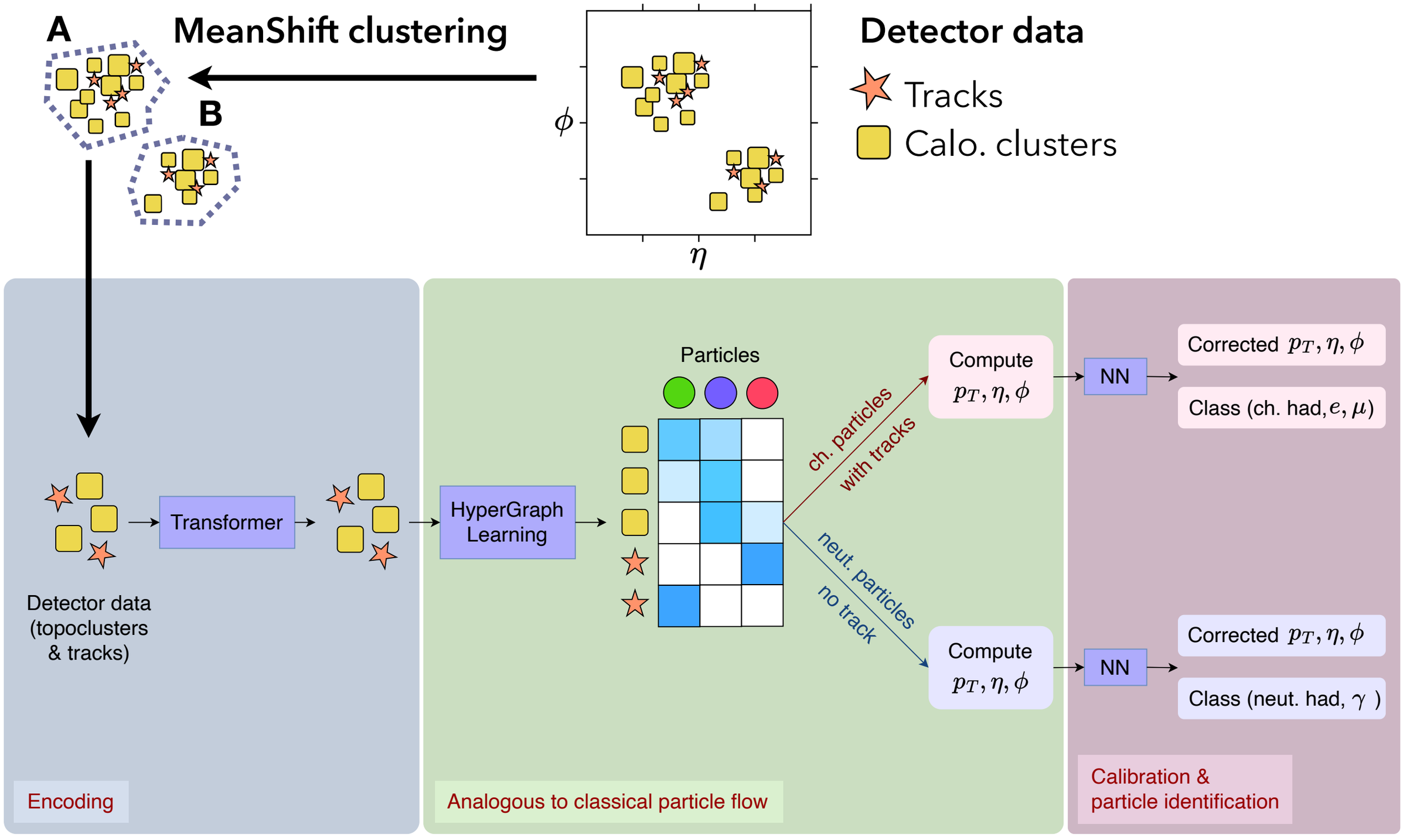
Learning the energy-based incidence matrix is an **inductive bias** that makes predictions more accurate and **fully interpretable**



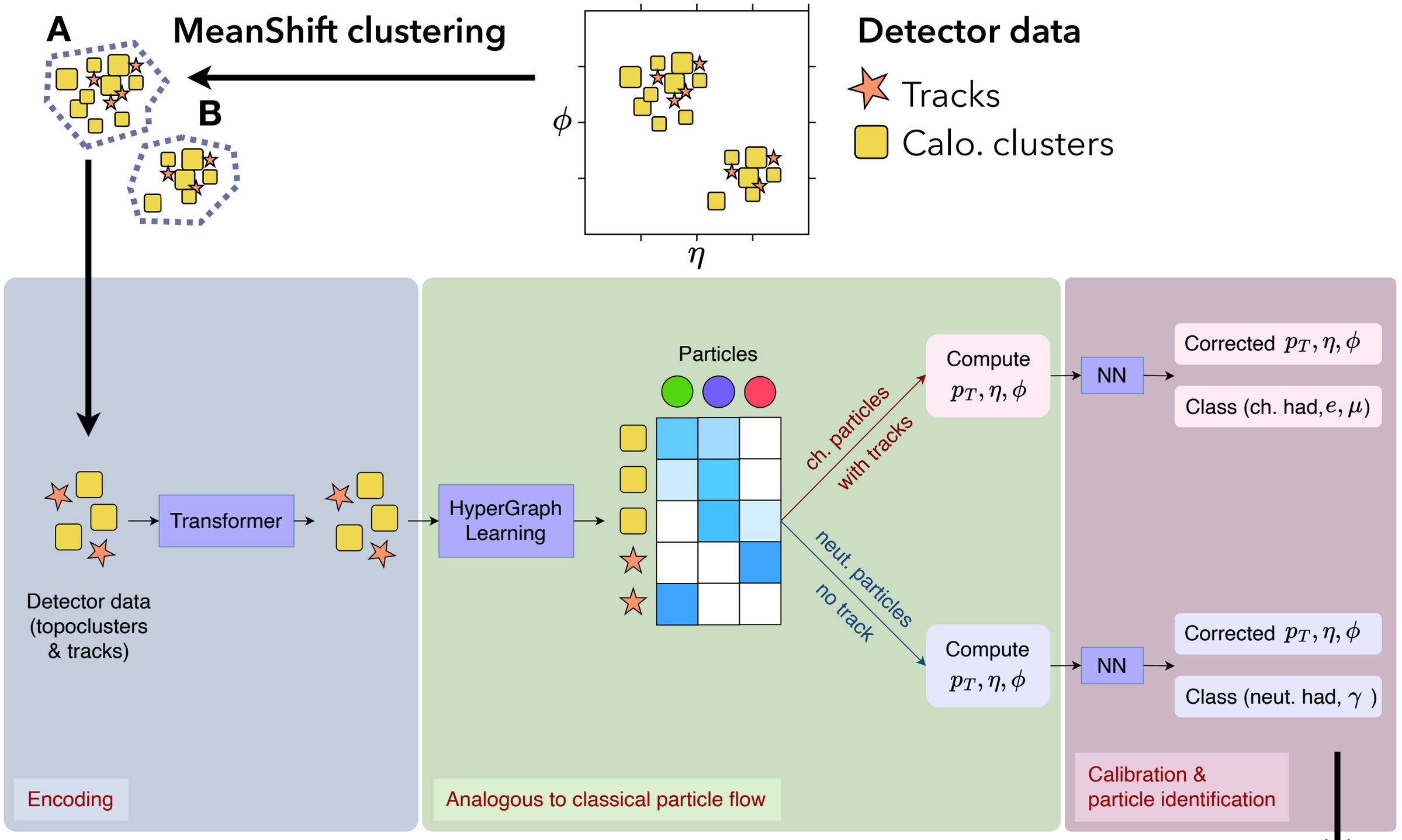

# HGPflow algorithm




# HGPflow algorithm

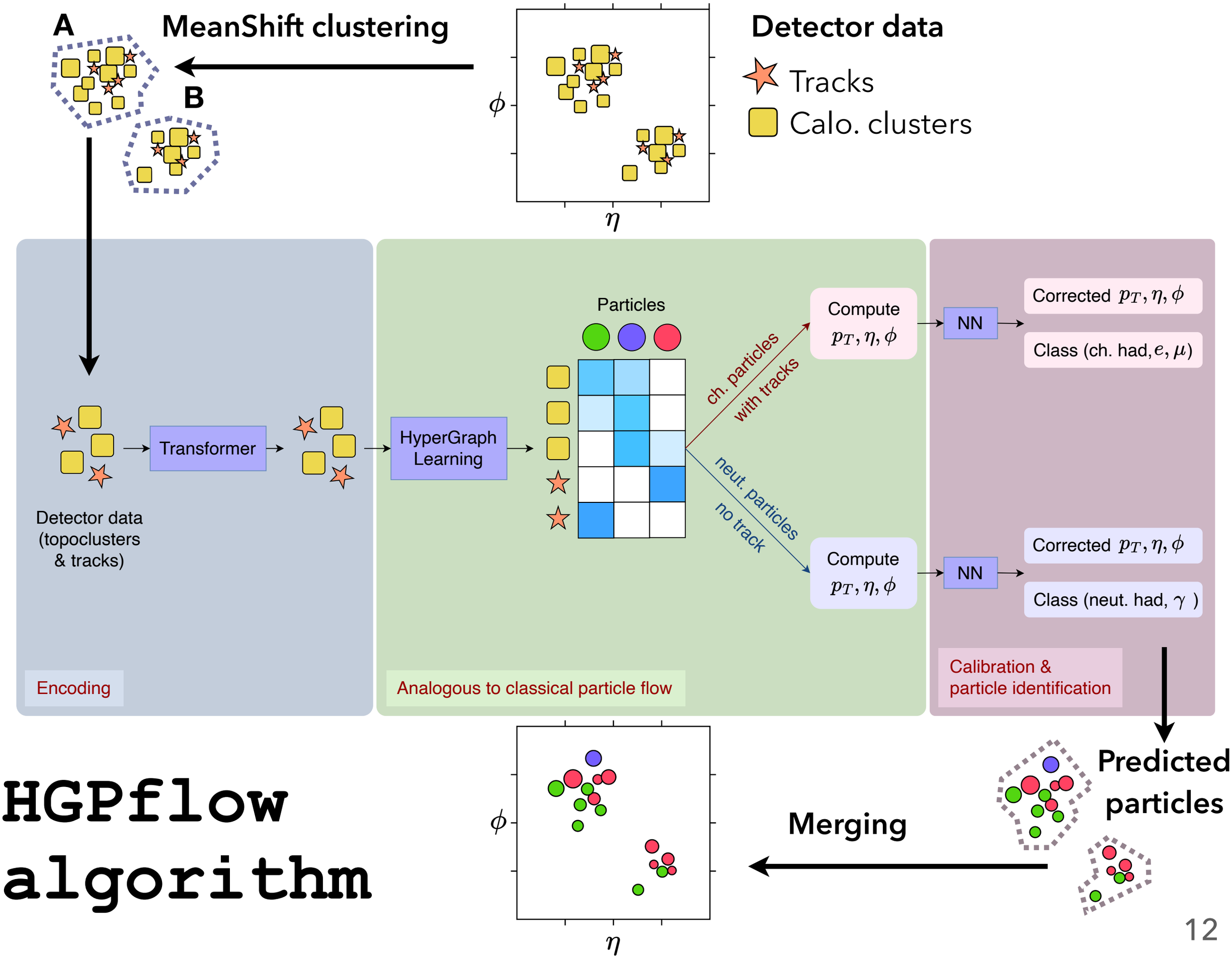



# HGPflow algorithm




# HGPflow algorithm

# HGPflow algorithm



# arXiv:2410.23236 (*soon to appear in EPJC*)

arXiv > hep-ex > arXiv:2410.23236

Search... Help | Adv

High Energy Physics – Experiment

*[Submitted on 30 Oct 2024]*

## HGPflow: Extending Hypergraph Particle Flow to Collider Event Reconstruction

Nilotpal Kakati, Etienne Dreyer, Anna Ivina, Francesco Armando Di Bello, Lukas Heinrich, Marumi Kado, Eilam Gross

In high energy physics, the ability to reconstruct particles based on their detector signatures is essential for downstream data analyses. A particle reconstruction algorithm based on learning hypergraphs (HGPflow) has previously been explored in the context of single jets. In this paper, we expand the scope to full proton–proton and electron–positron collision events and study reconstruction quality using metrics at the particle, jet, and event levels. Rather than operating on the entire event in a single pass, we train HGPflow on smaller partitions to avoid potentially learning long-range correlations related to the physics process. We demonstrate that this approach is feasible and that on most metrics, HGPflow outperforms both traditional particle flow algorithms and a machine learning–based benchmark model.

WEIZMANN INSTITUTE OF SCIENCE

Technical University of Munich

Università di Genova

MAX-PLANCK-INSTITUT FÜR PHYSIK

# Datasets

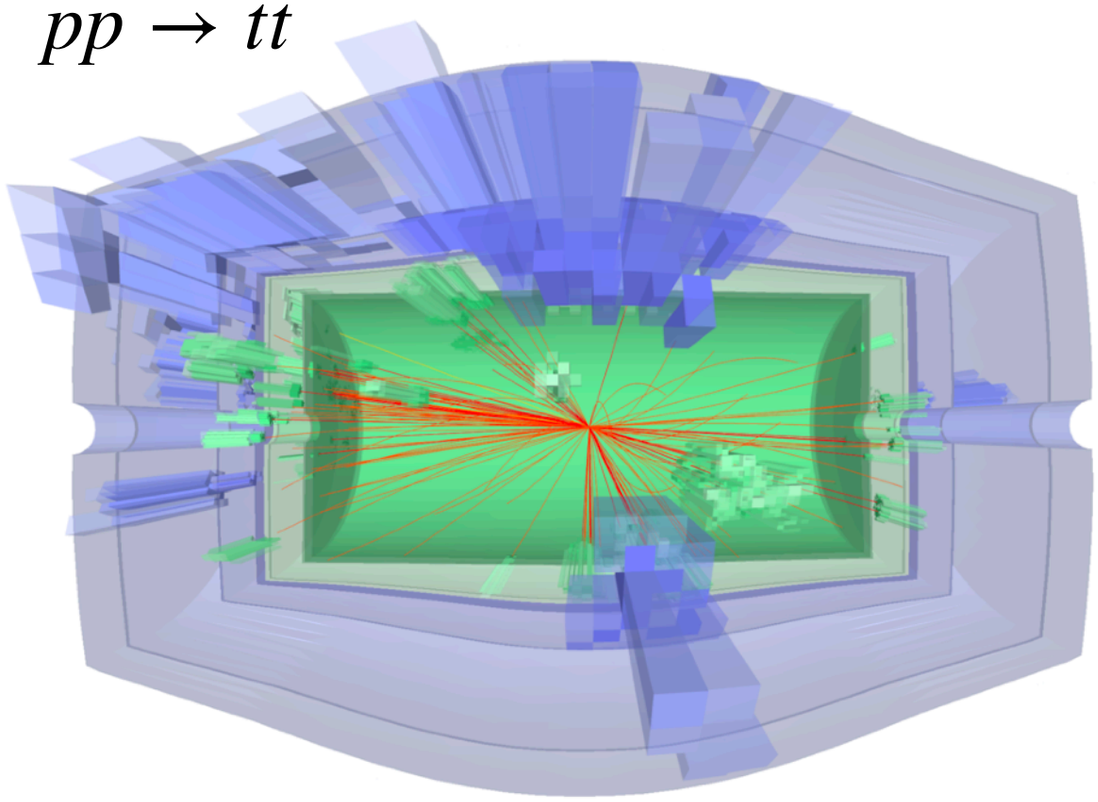
COCOA (2023) MLST 4 035042

- Similar to ATLAS
- Relatively low granularity
- Comes with basic particle flow algorithm

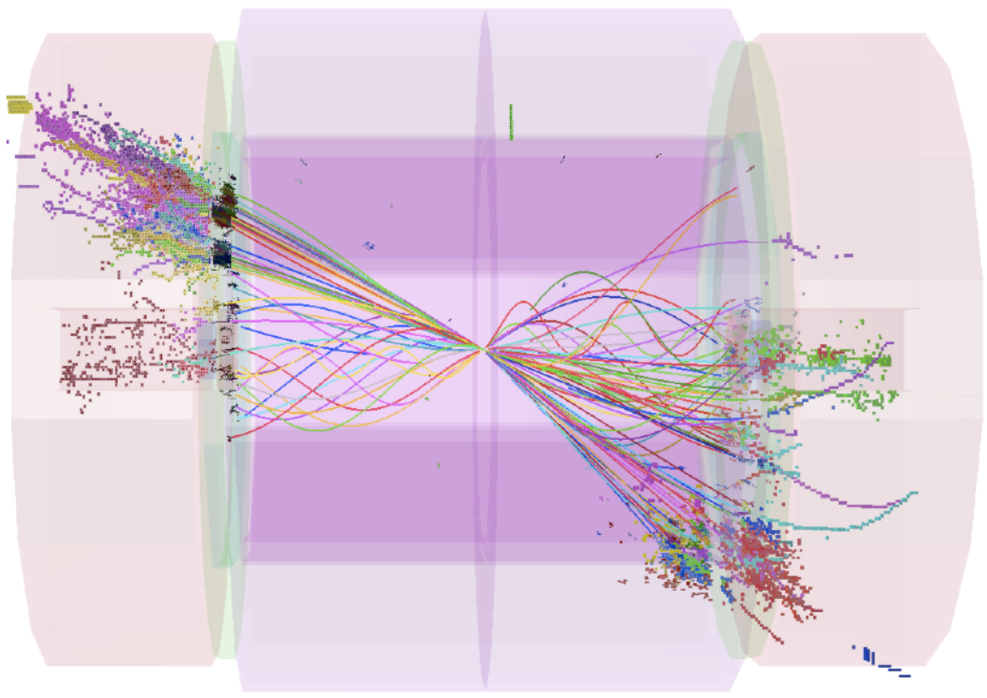
CLICdet arXiv:812.07337

- Publicly-available dataset: zenodo/8260741
- High granularity
- Sophisticated Pandora particle flow algo.

$pp \to t\bar{t}$



$e^+e^- \to t\bar{t}$

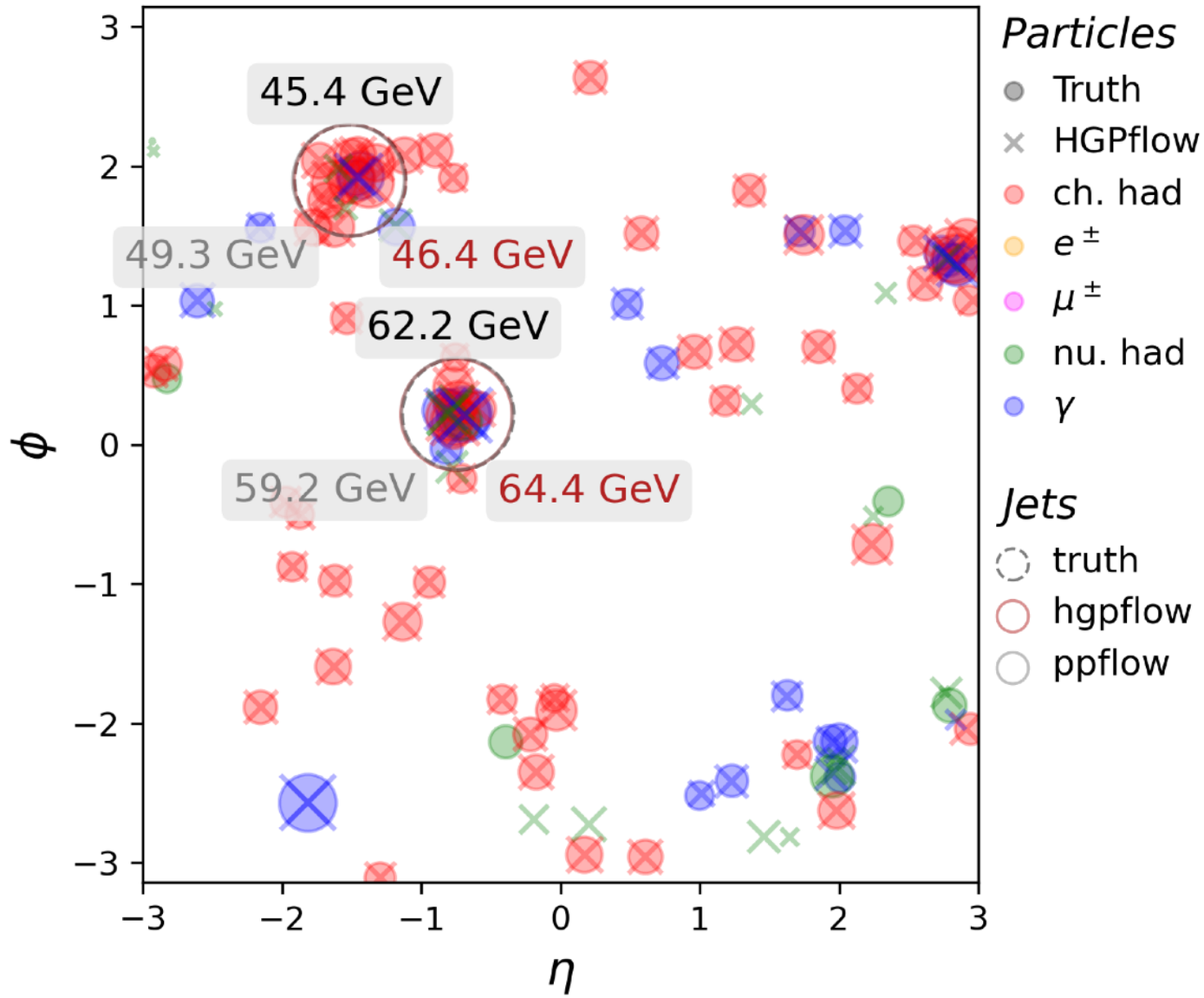


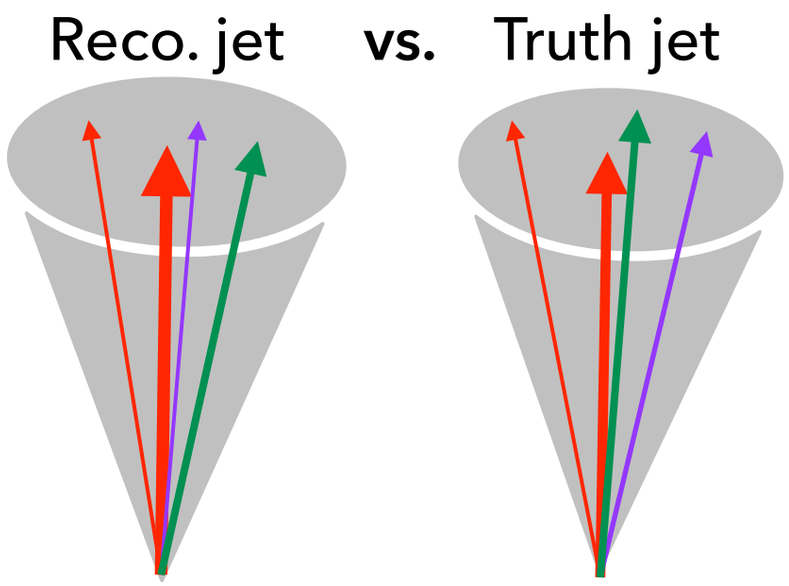
Source: arXiv:1208.1402

| Detector | Process | Statistics train | val. | test |
|---|---|---|---|---|
| COCOA | $p^+p^+ \to q\bar{q}$ | 250k | 10k | 35k |
| | single $\pi^+$ | – | – | 30k / $p_T$ bin |
| | $p^+p^+ \to t\bar{t}$ | – | – | 20k |
| | $p^+p^+ \to Z(\nu\bar{\nu})H(b\bar{b})$ | – | – | 10k |
| CLIC | $e^+e^- \to q\bar{q}$ | 1M | 5k | 20k |

# Performance: dijet (COCOA)

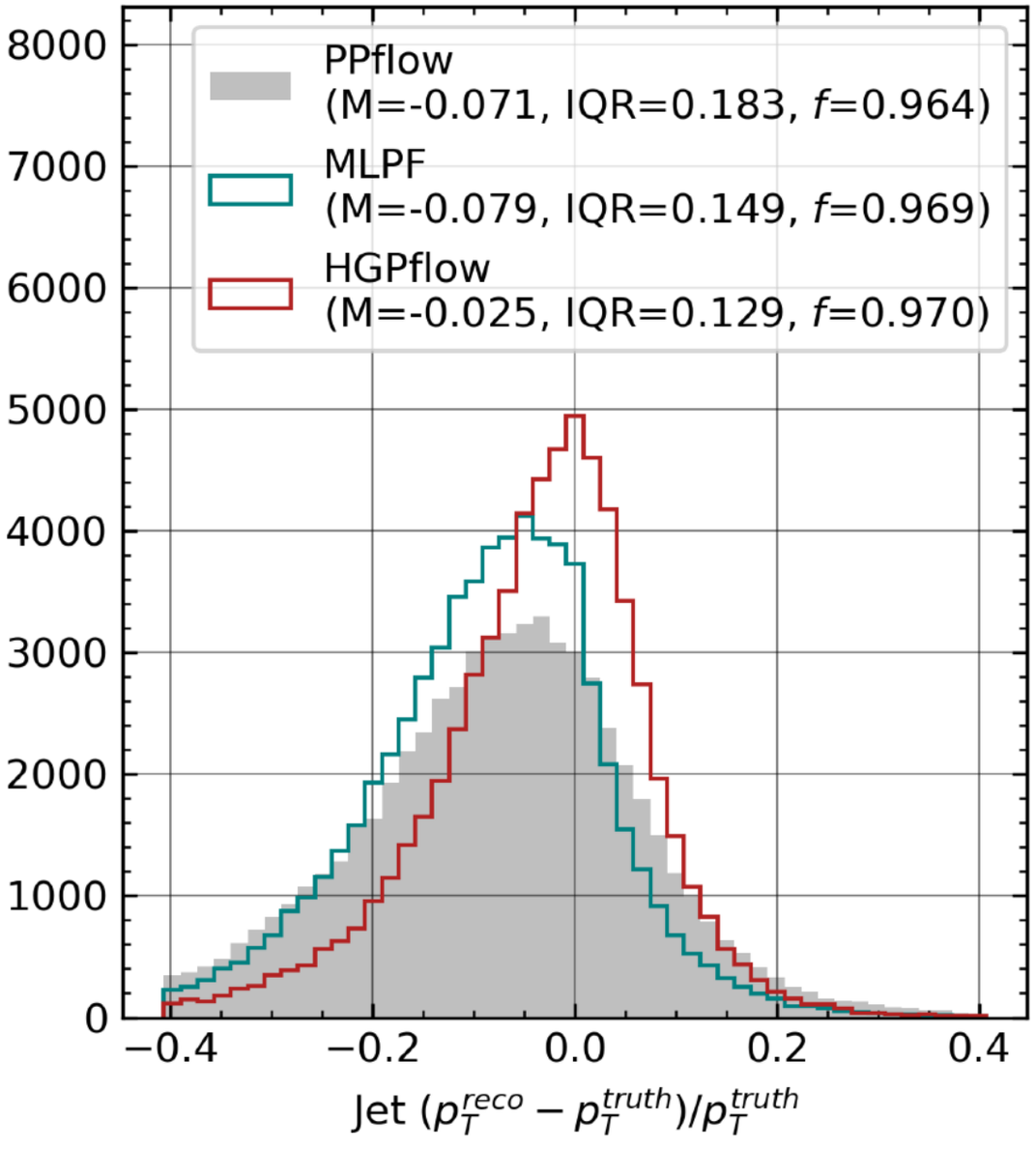
Trained on 250k and tested on 35k events

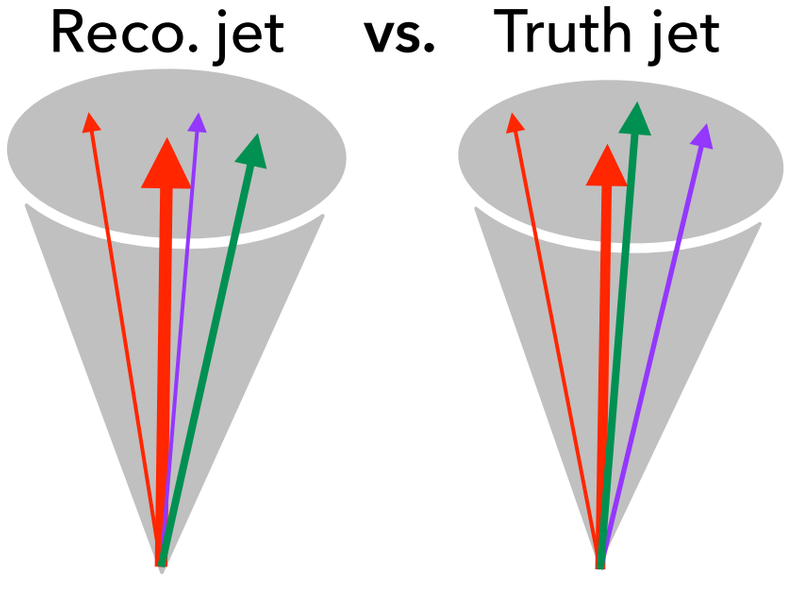


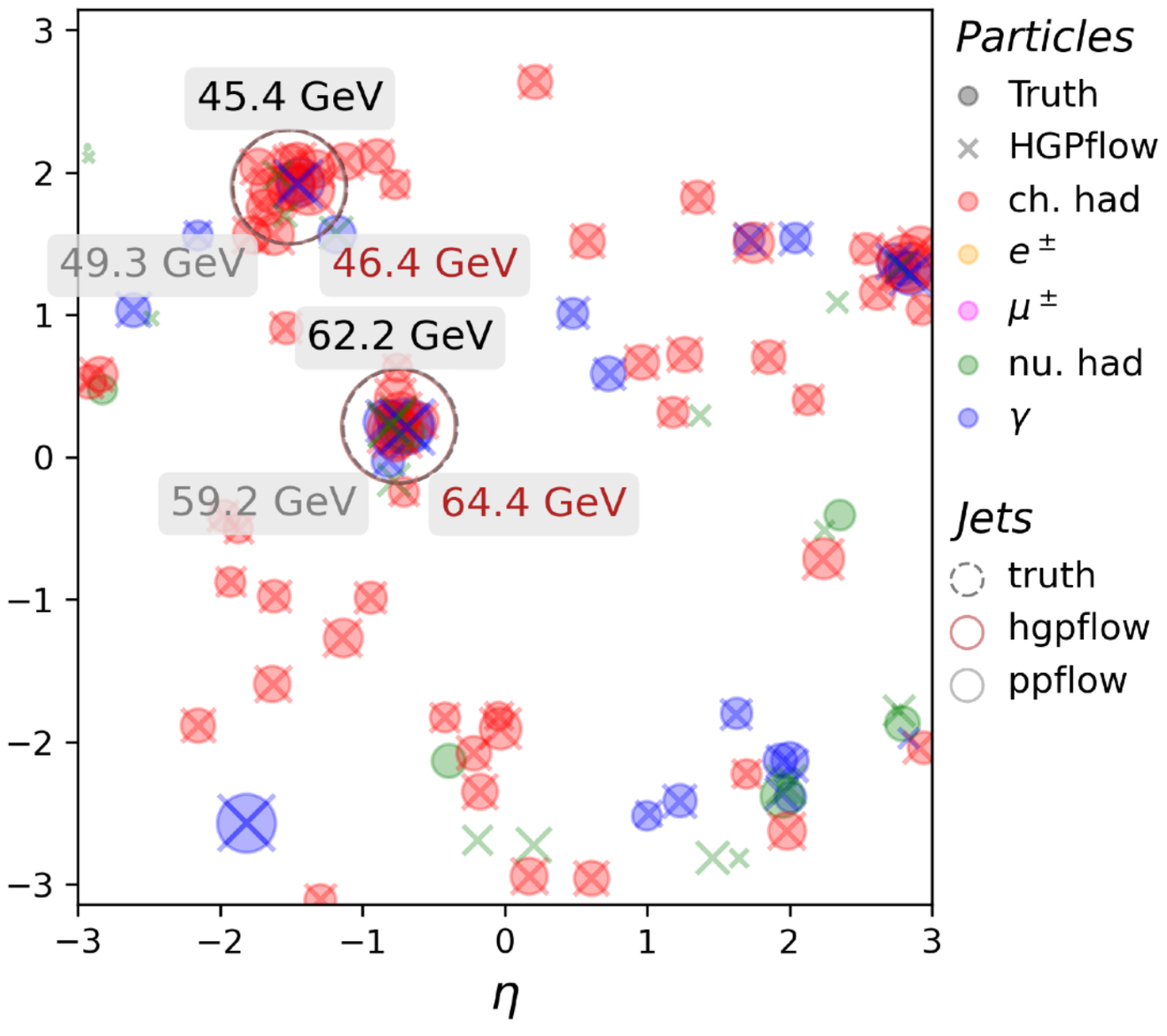



# Performance: dijet (COCOA)

Trained on 250k and tested on 35k events

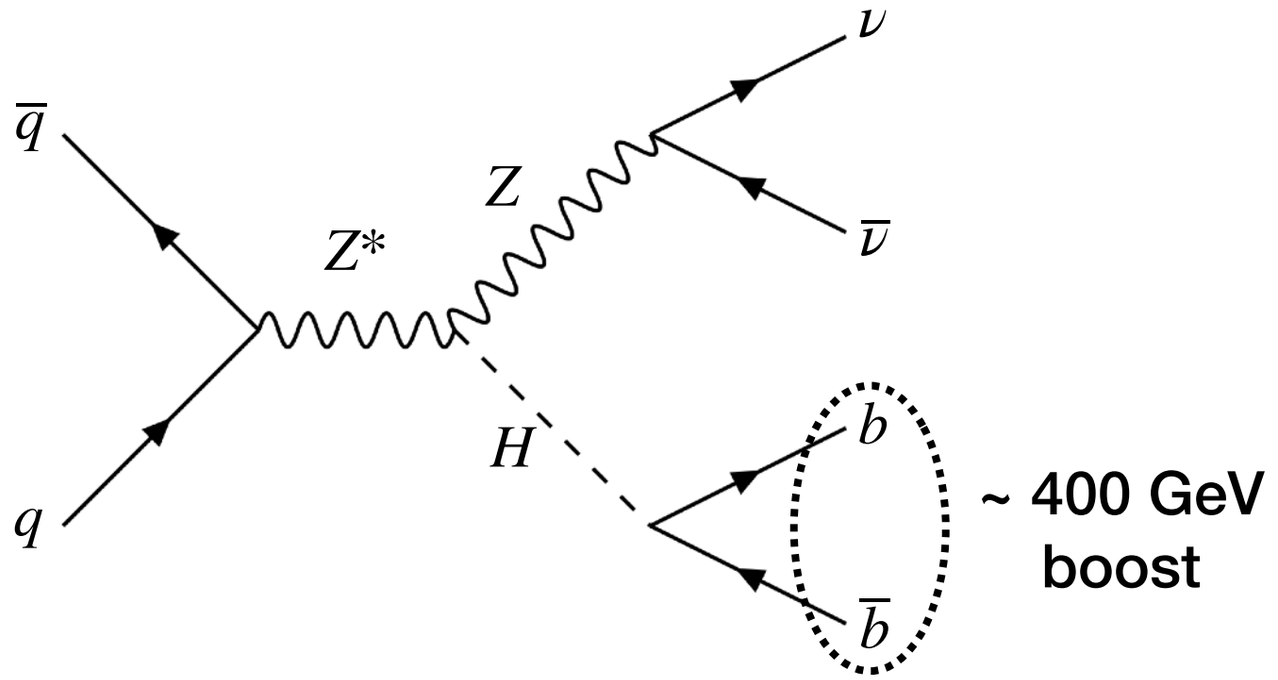


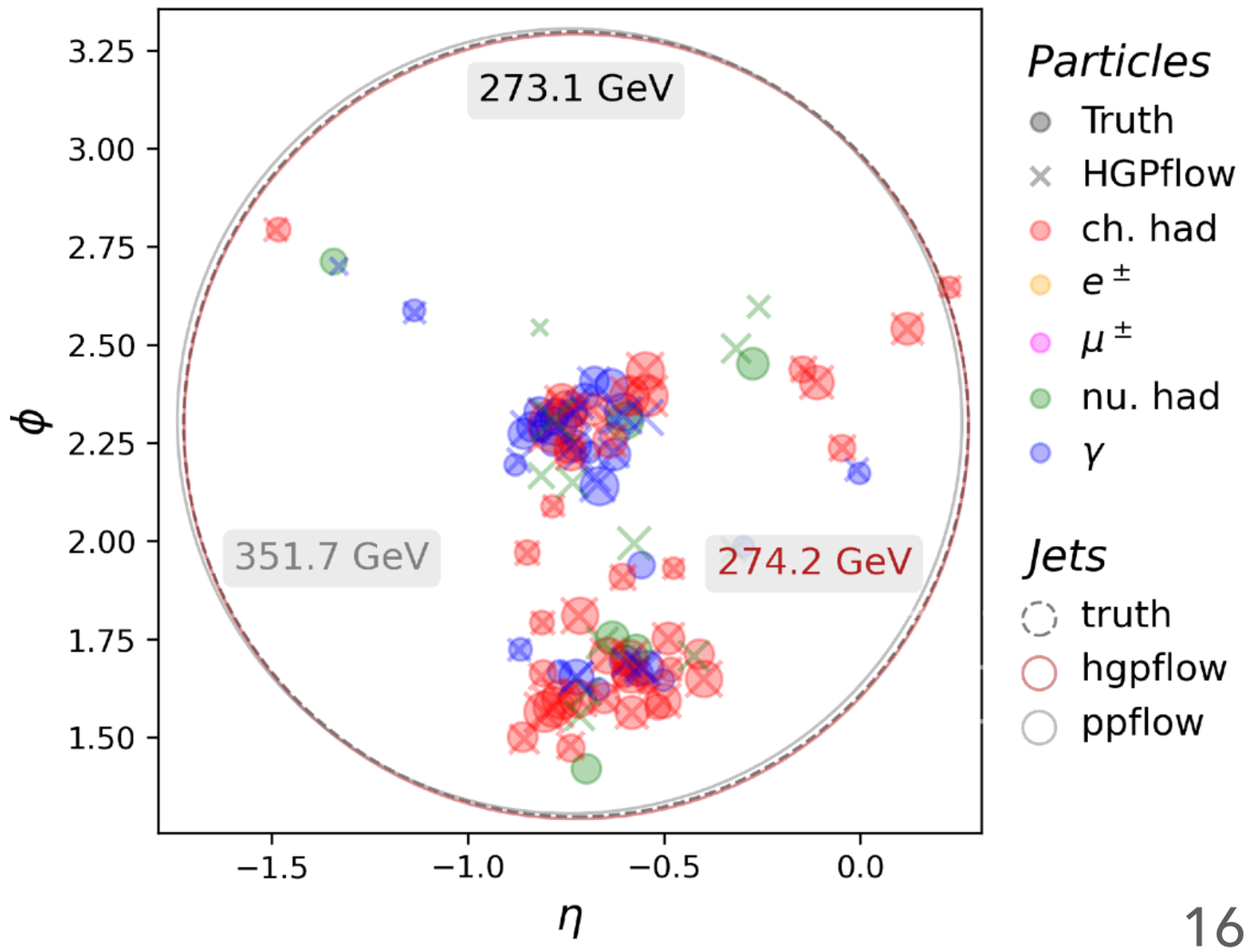



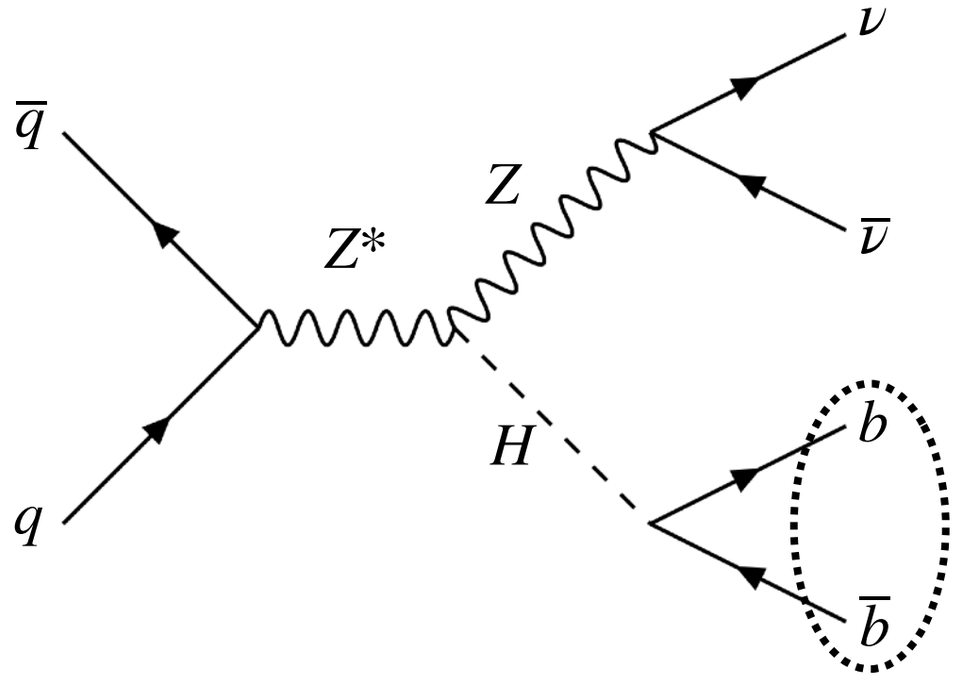
# Performance: boosted $ZH(bb)$ (COCOA)

Not encountered during training!



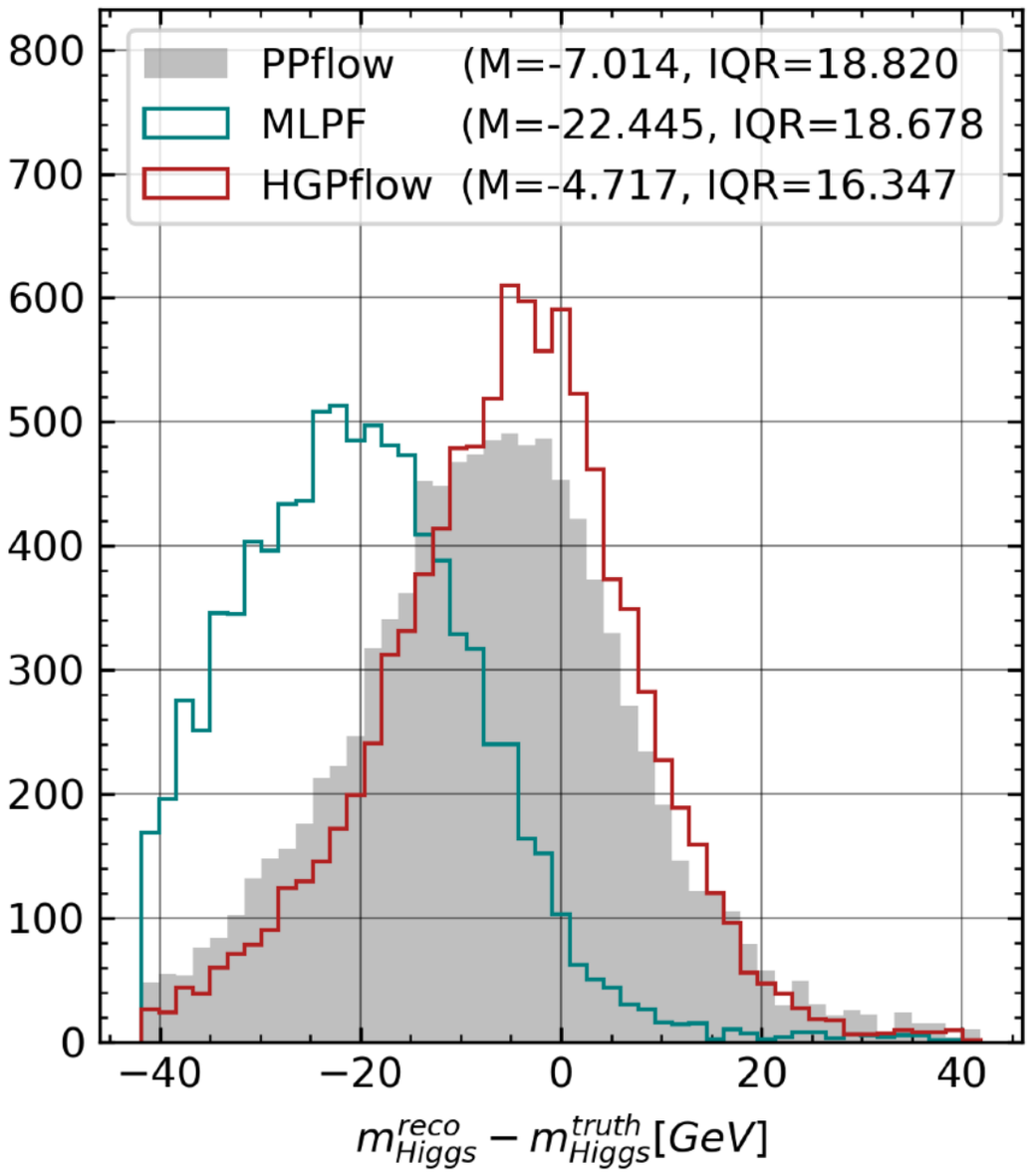



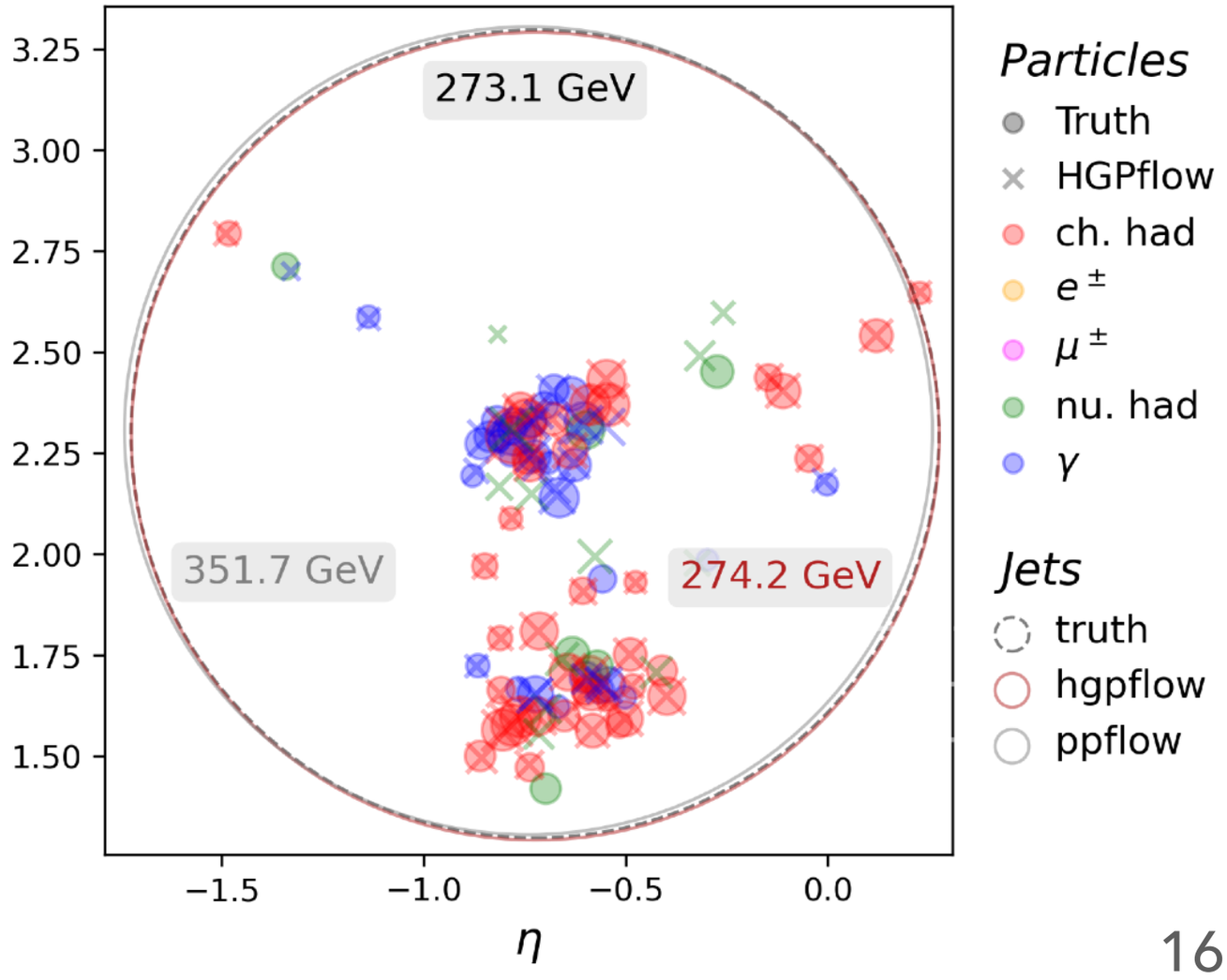
# Performance: boosted $ZH(bb)$ (COCOA)



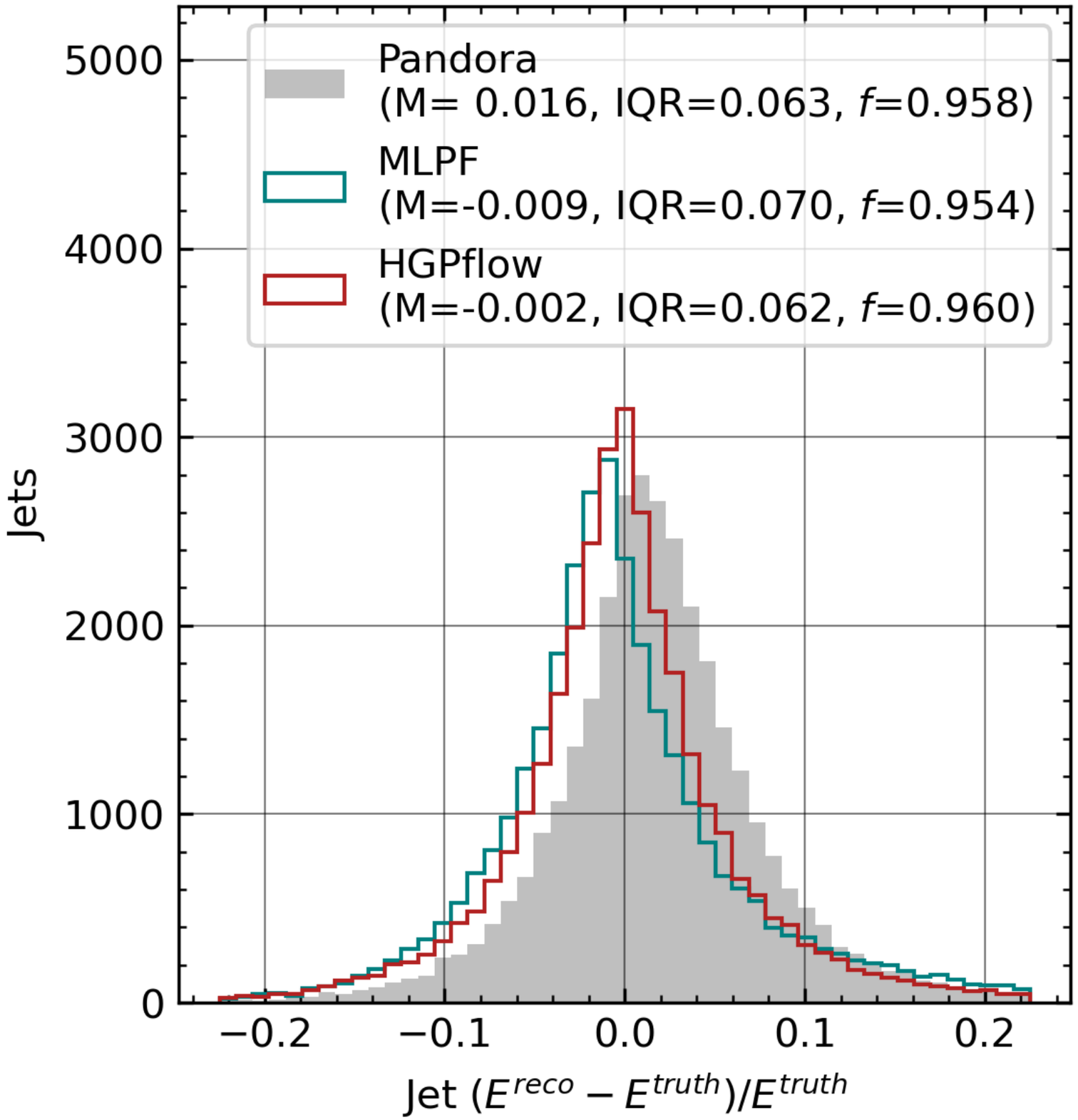

~ 400 GeV boost

Not encountered during training!



# Performance: $e^+e^- \to q\bar{q}$ (CLIC)

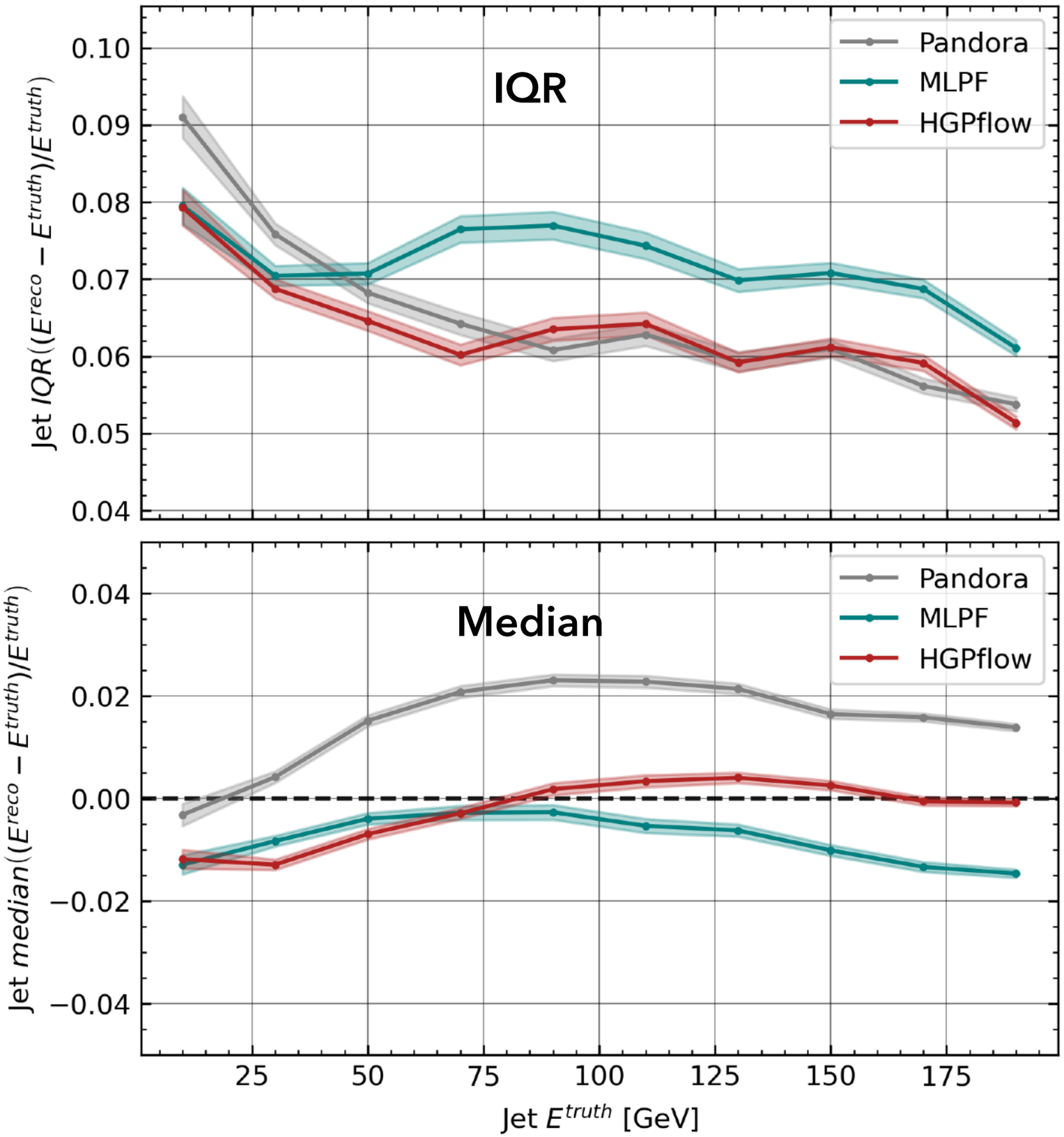



Trained on 1M, tested on 20k

HGPflow works for high-gran. calorimeters too

- Slightly outperforms Pandora
- Promising for existing and future facilities

# Performance: $e^+e^- \to q\bar{q}$ (CLIC)




Trained on 1M, tested on 20k

HGPflow works for high-gran. calorimeters too

- Slightly outperforms Pandora
- Promising for existing and future facilities

# Summary

**Particle reconstruction** is a set-to-set problem that is foundational for collider physics

**Hypergraph learning**

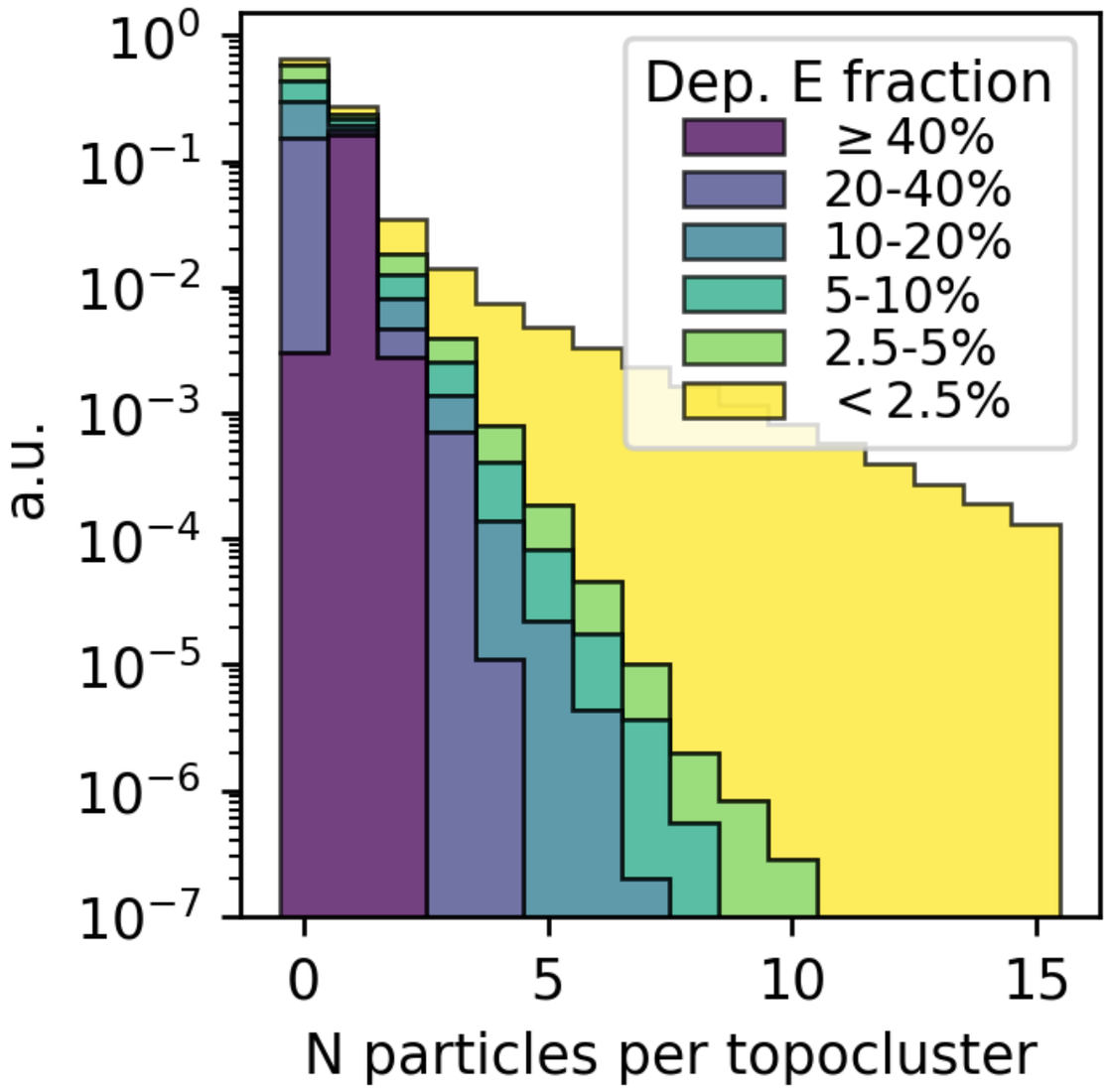
- addresses the core problem of overlap
- is fully interpretable in terms of energy flow
- can be readily scaled to full $p^+p^+$ and $e^+e^-$events

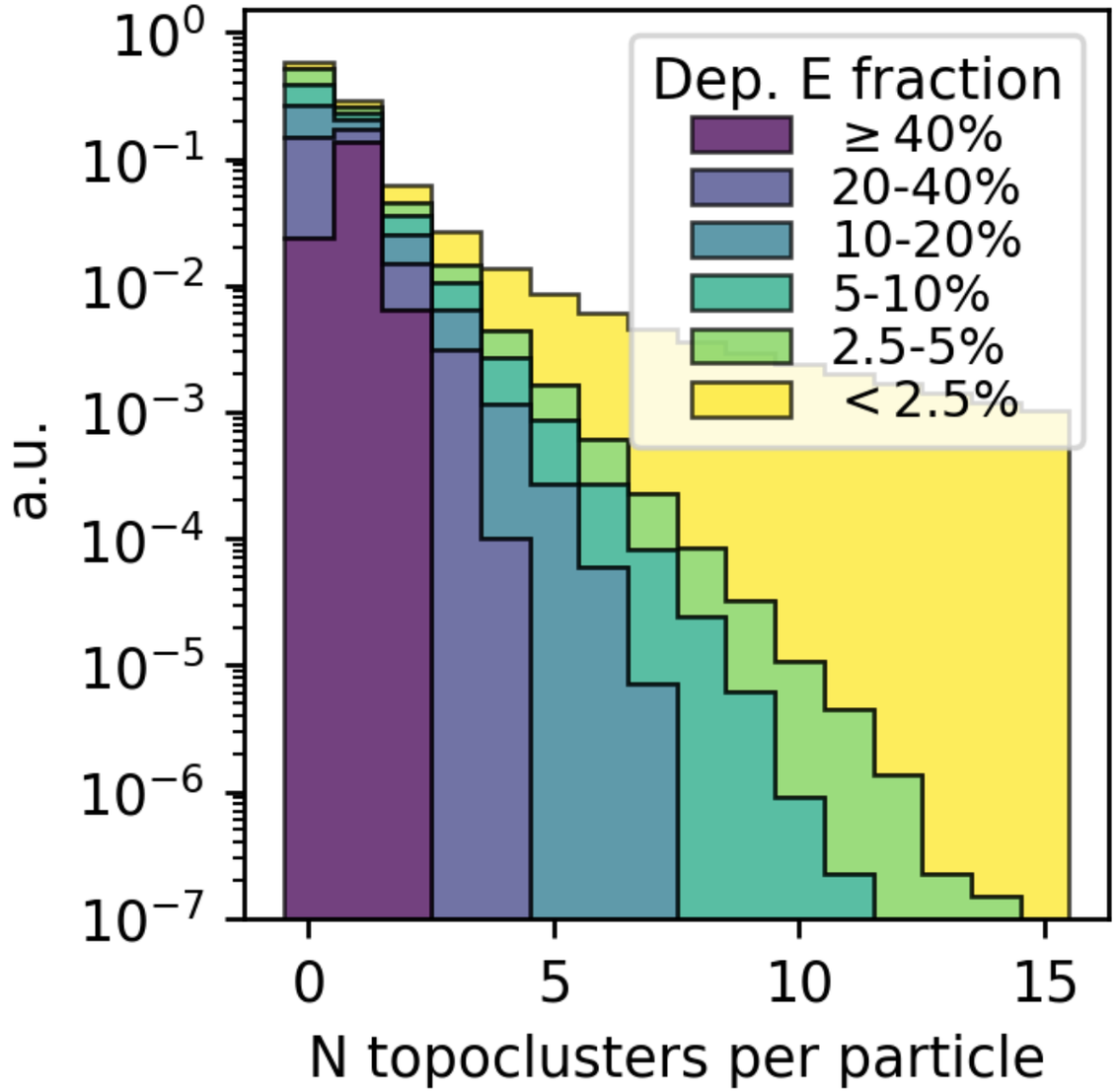
**Next steps**:

- implement at LHC – work on pileup, isolated lepton perf.
- further studies for future colliders, e.g. FCC

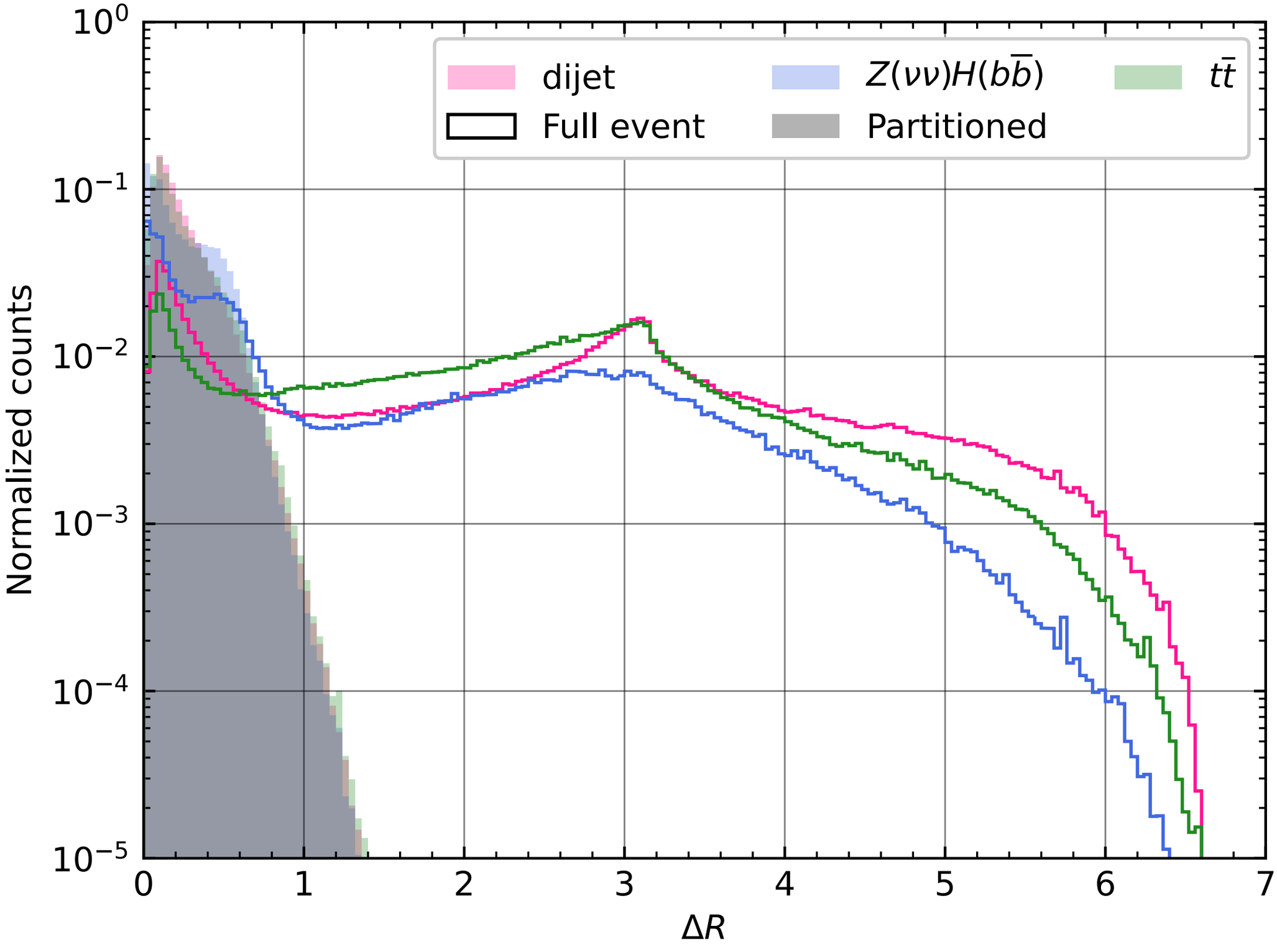
# Bonus

# Particle-cluster multiplicity



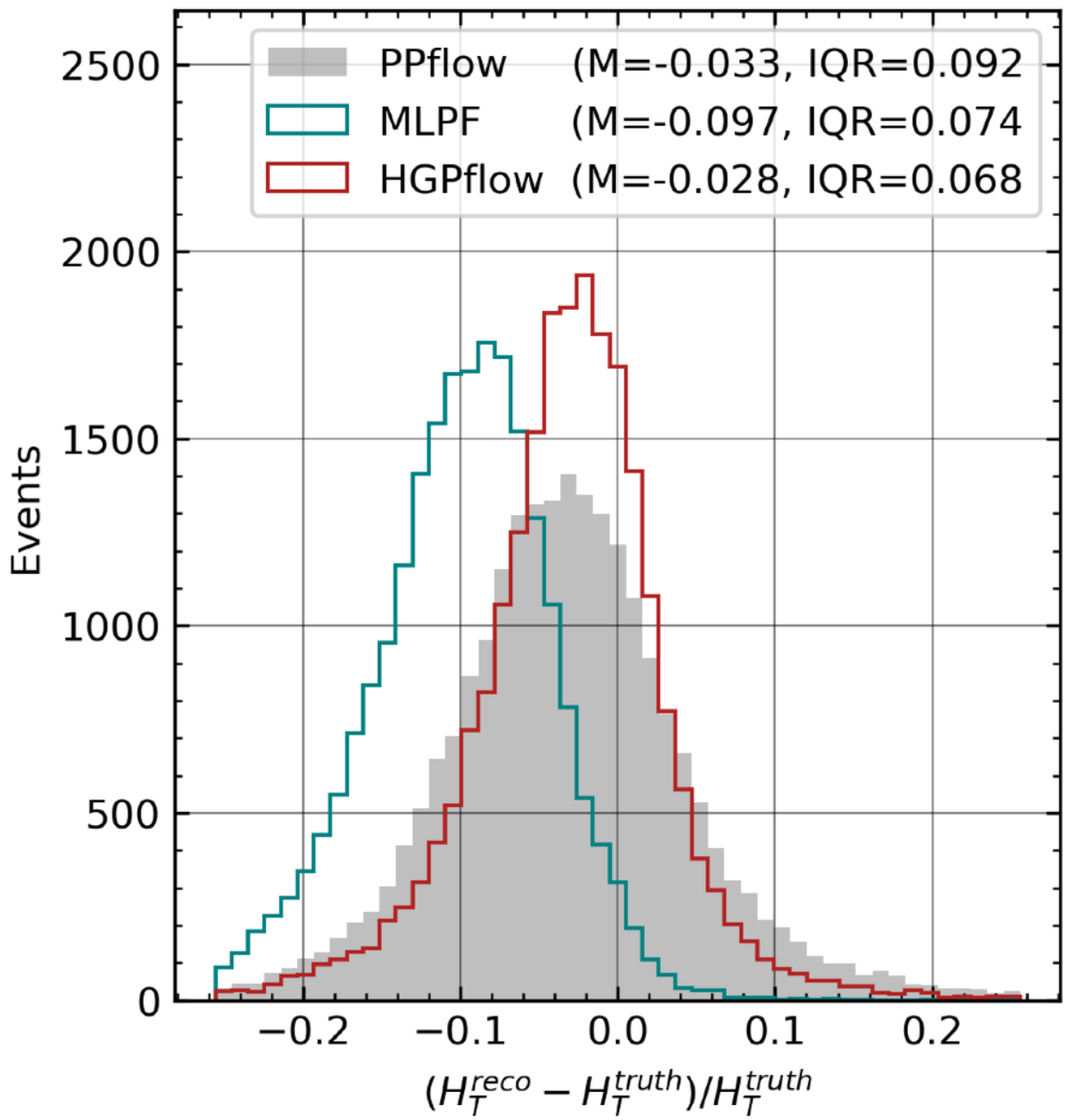



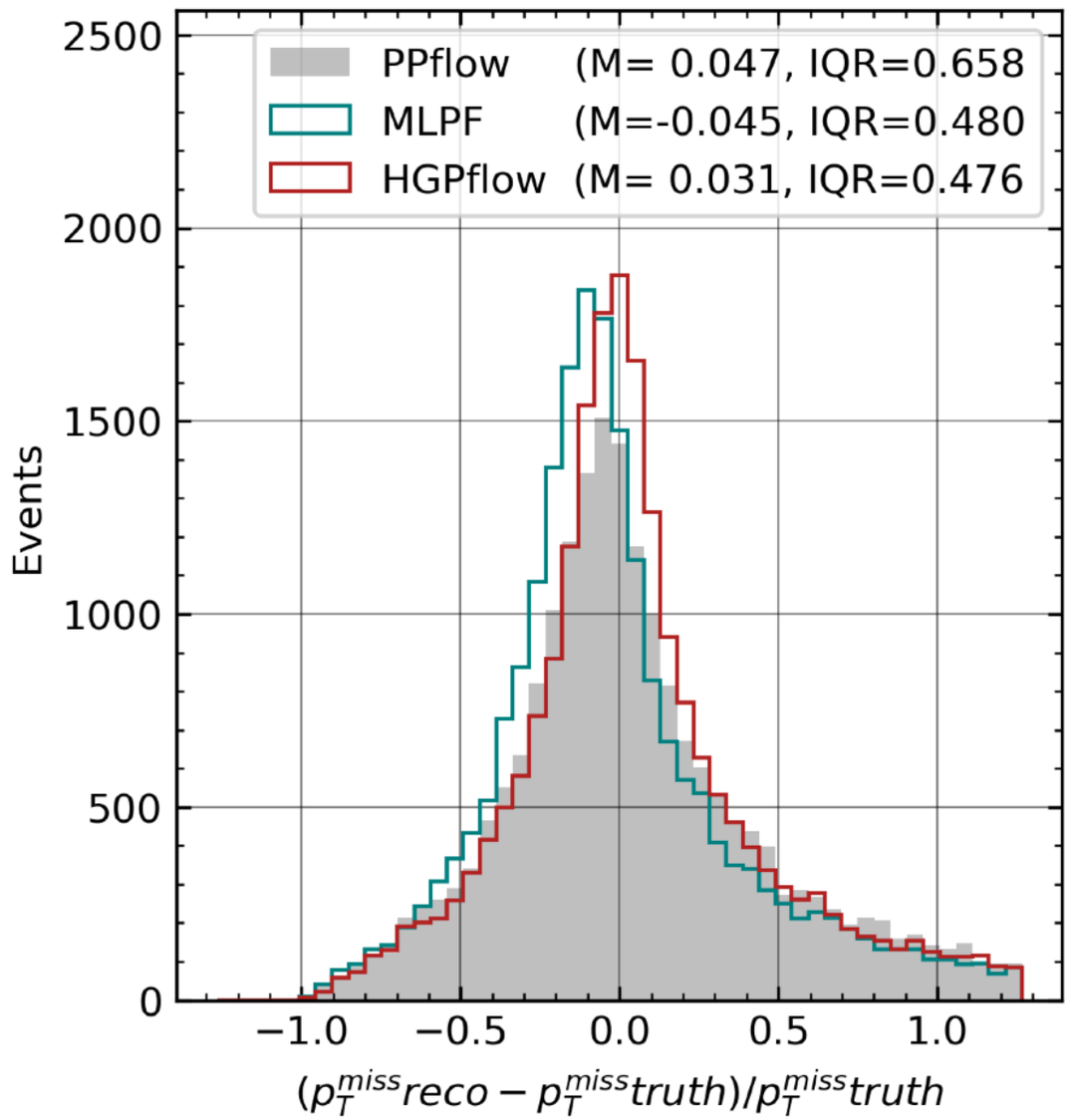
# Long-range correlations



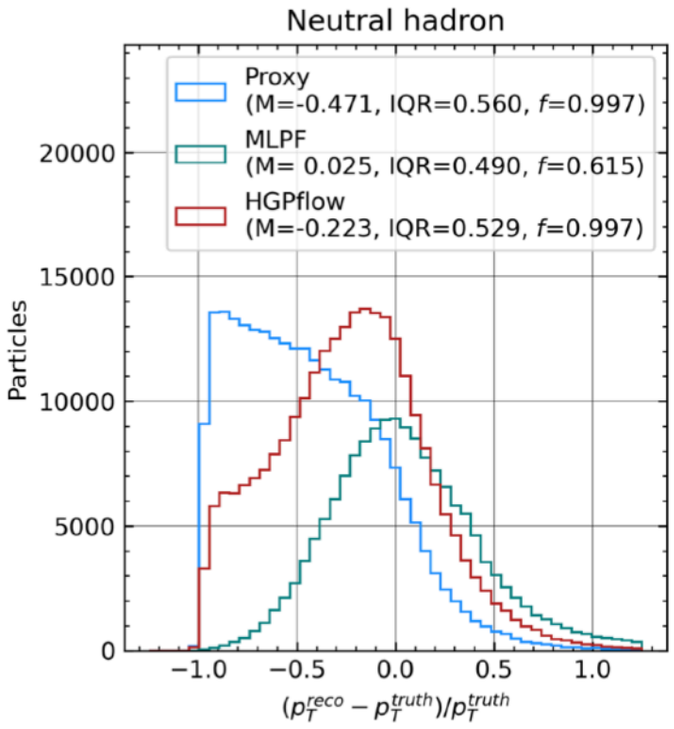
# Performance: $t\bar{t}$ (COCOA)

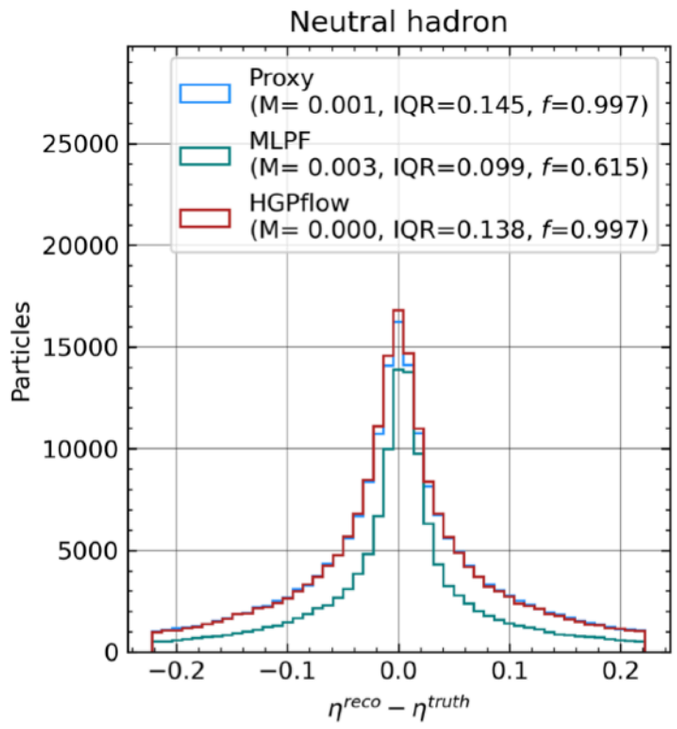
Not encountered during training!

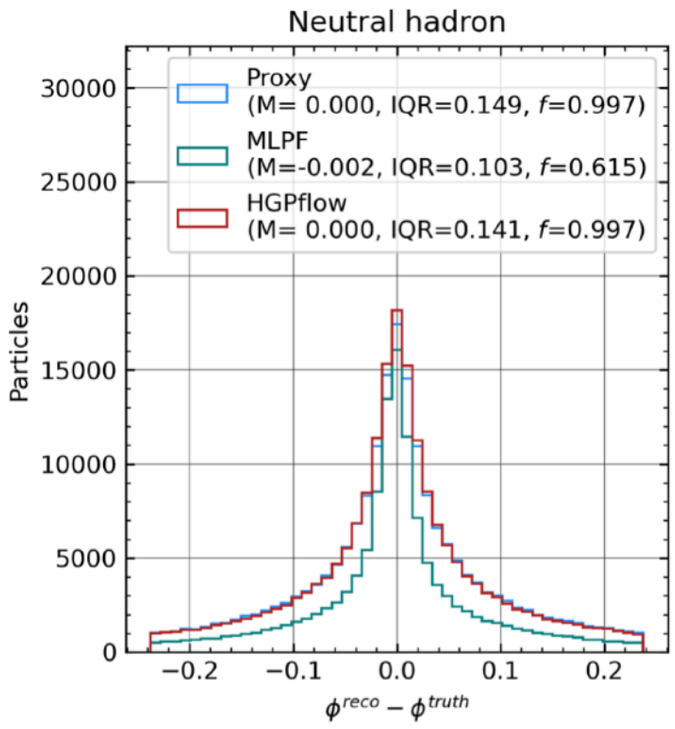


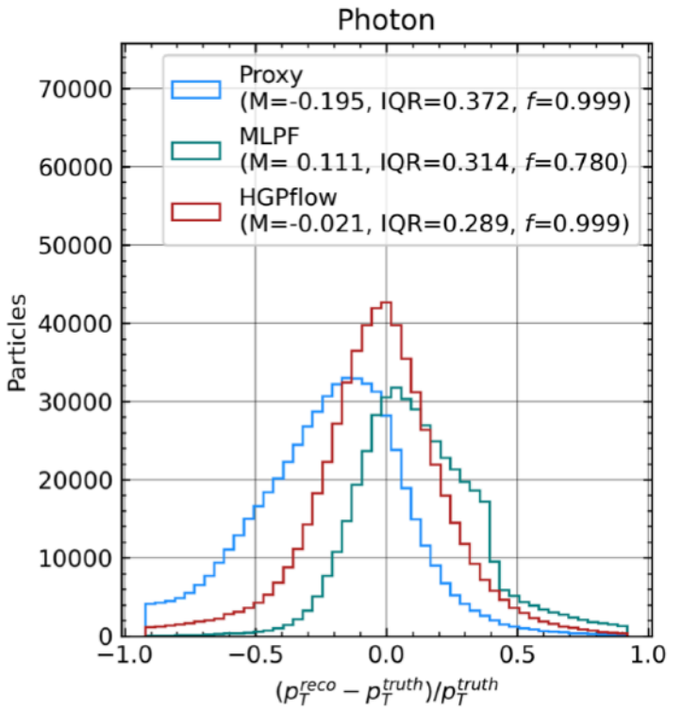



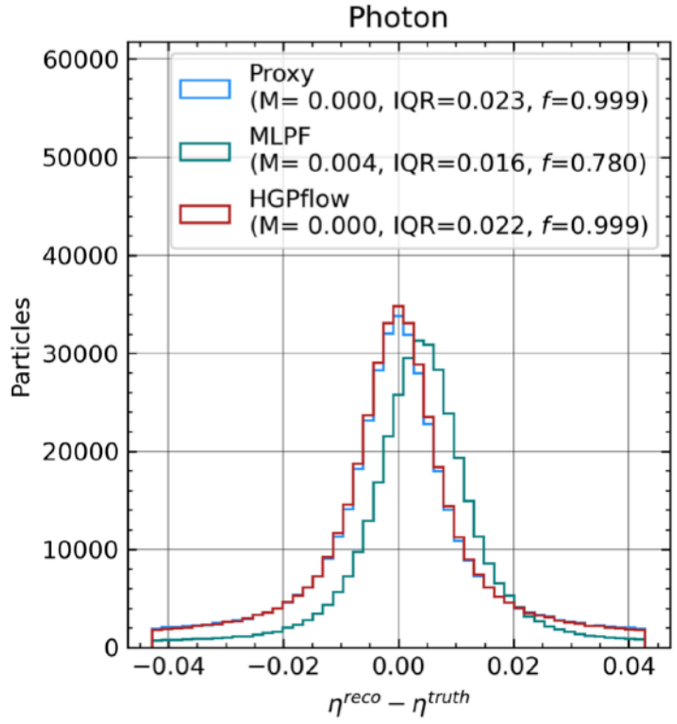
# Performance: dijet, particle-level (COCOA)

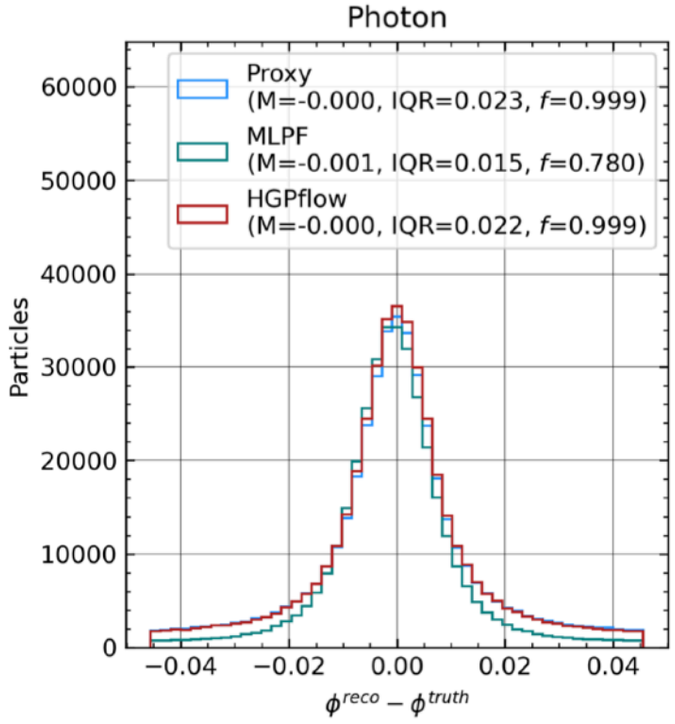


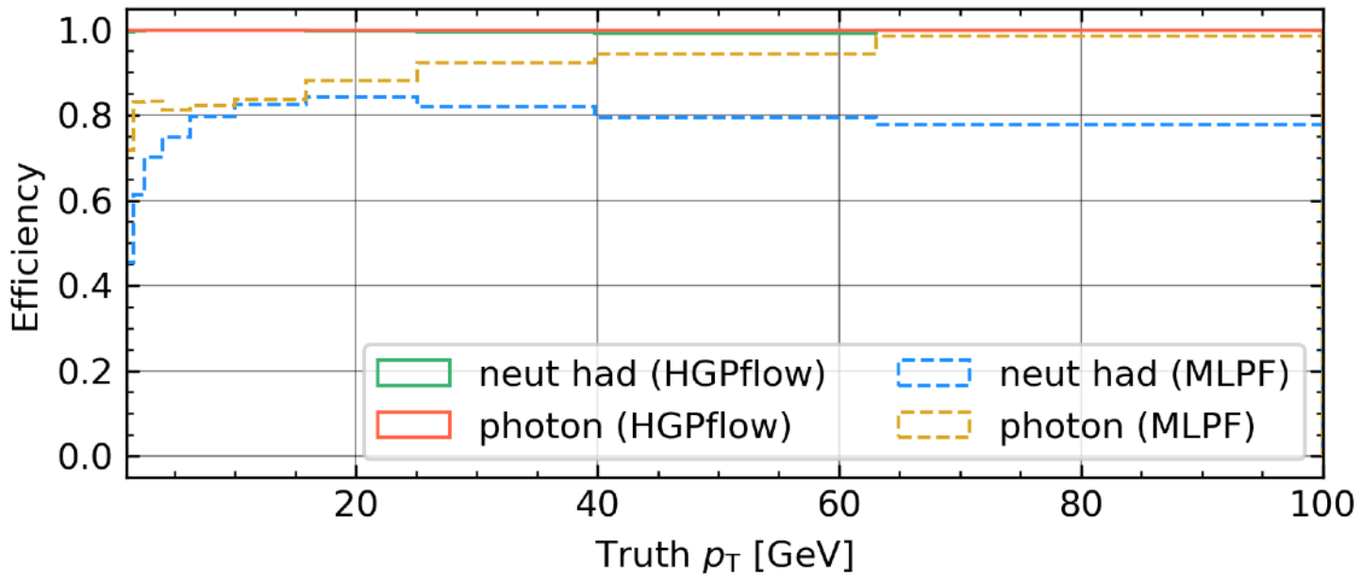



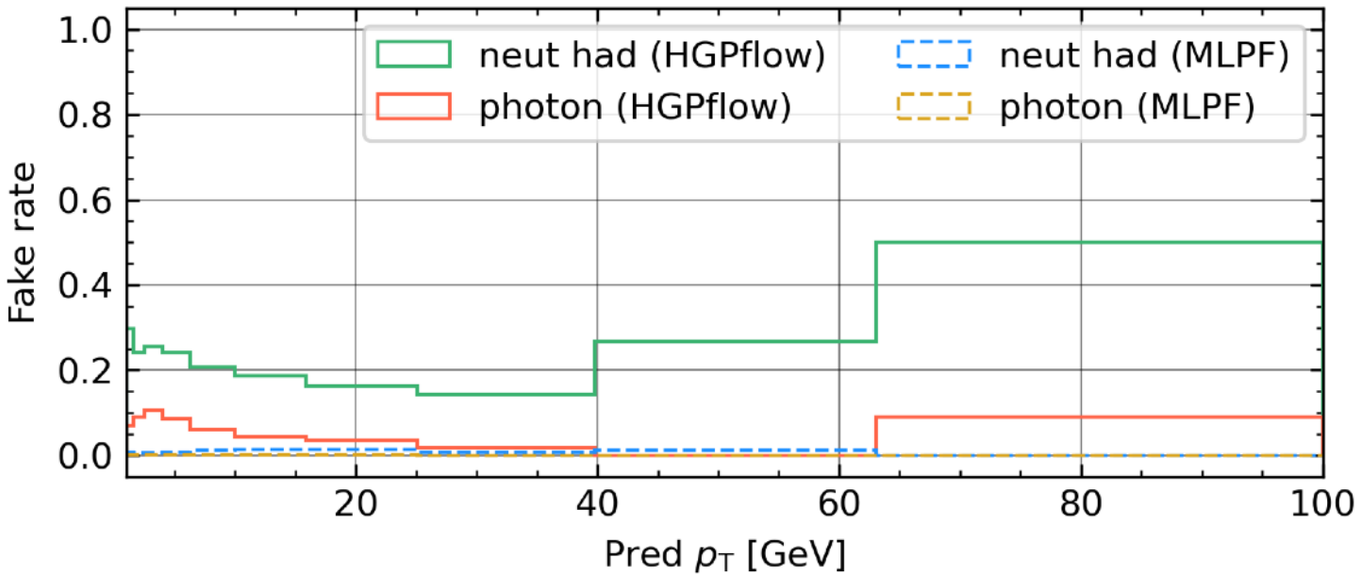



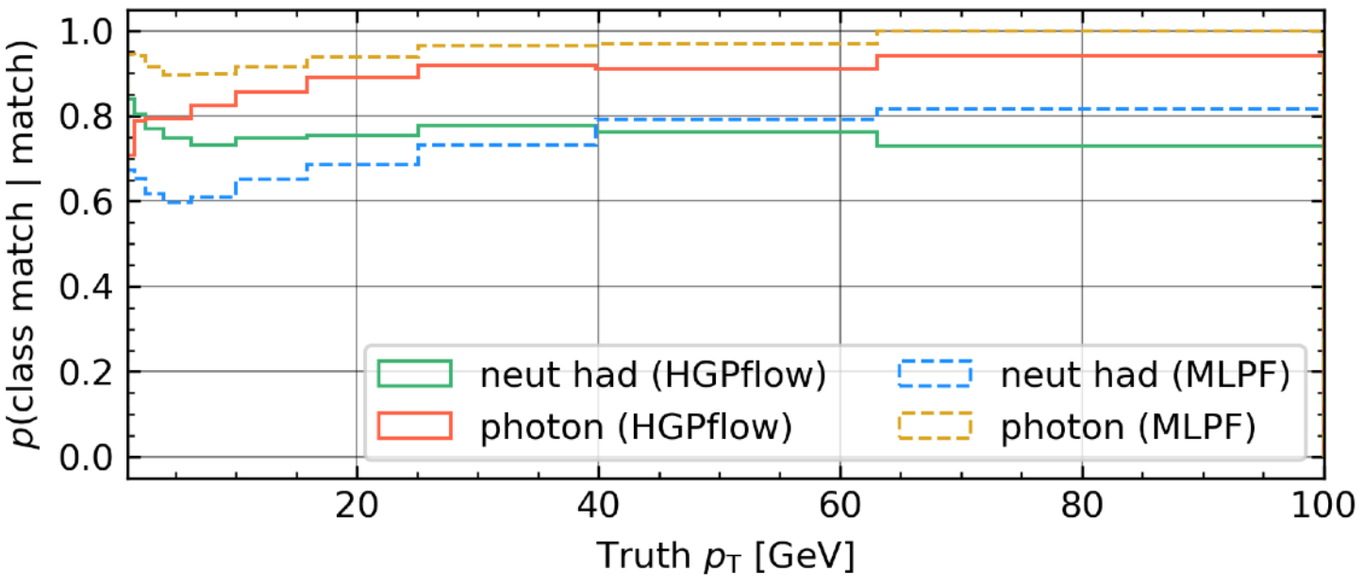



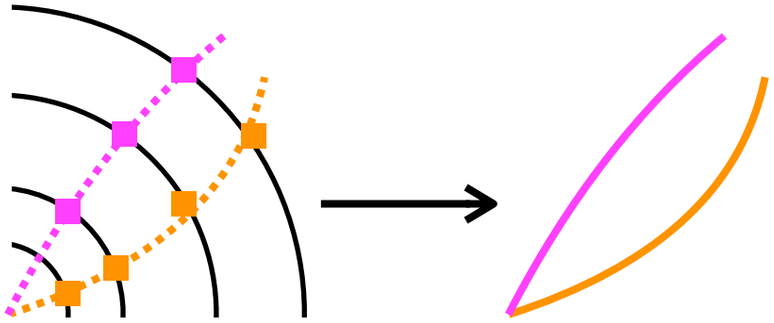


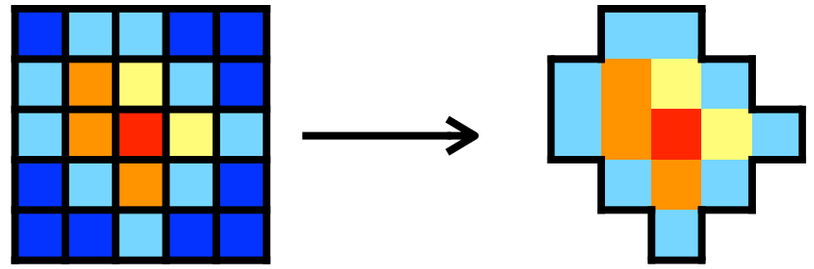

# Performance: dijet, particle-level (COCOA)

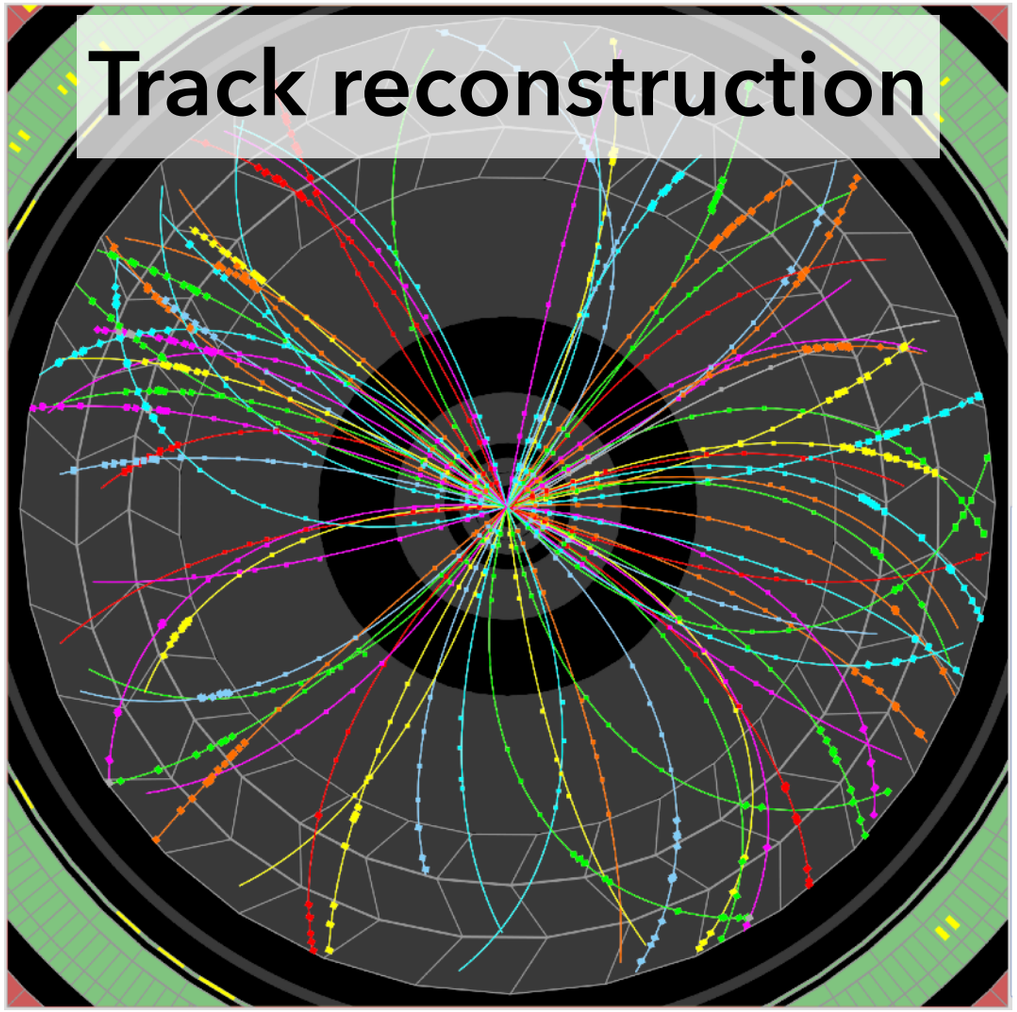


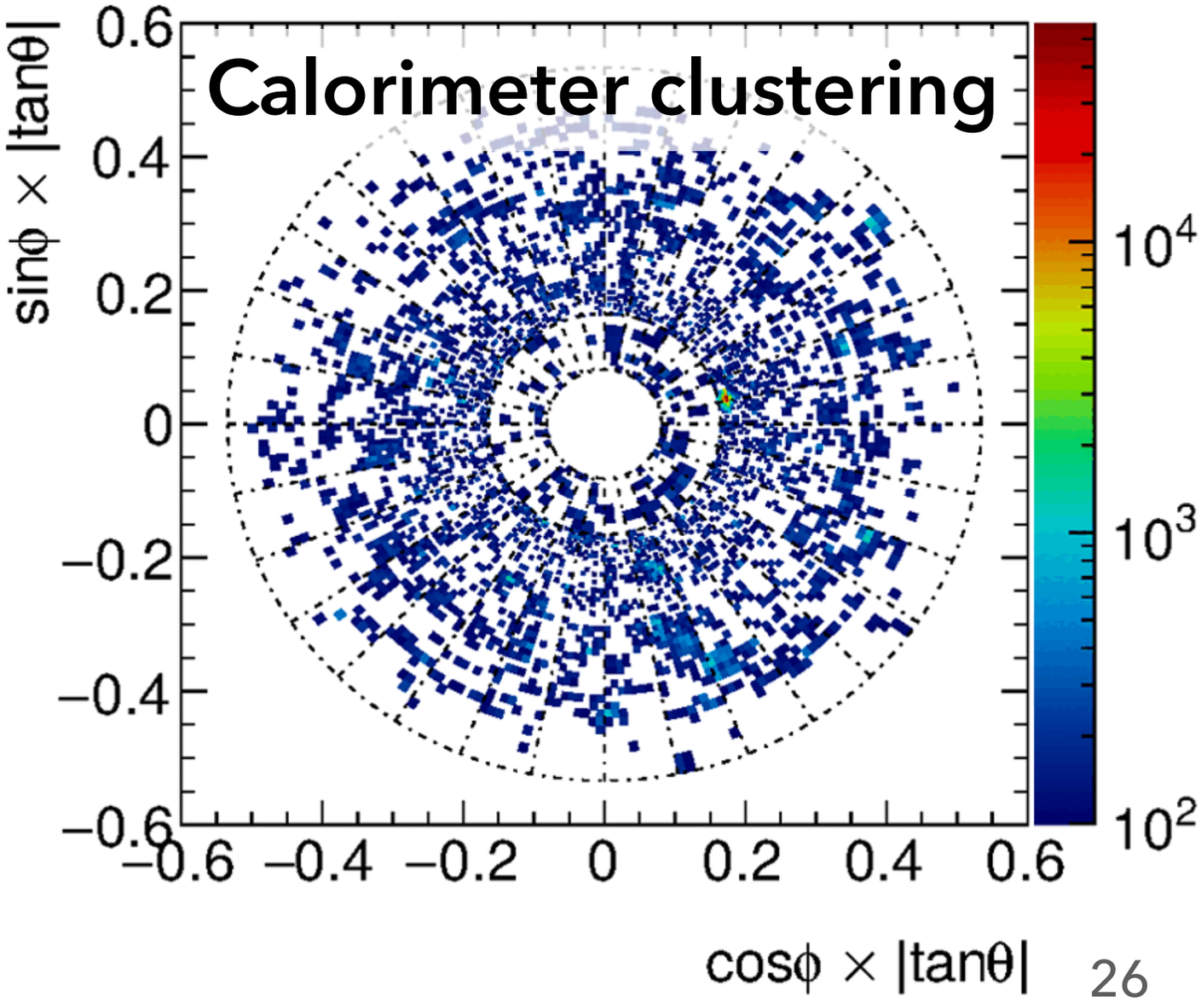



# Dimensionality reduction

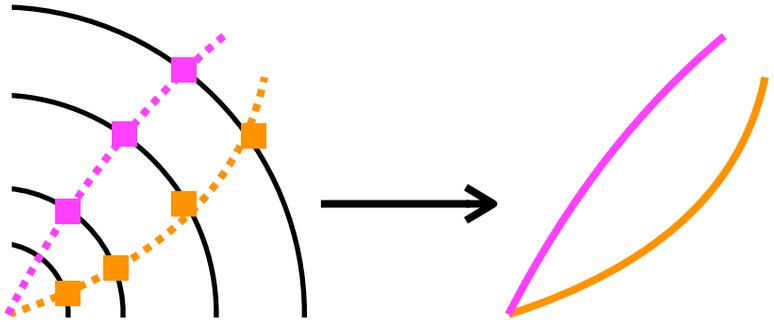


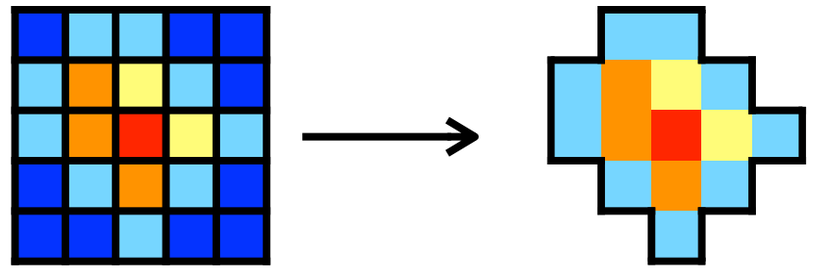


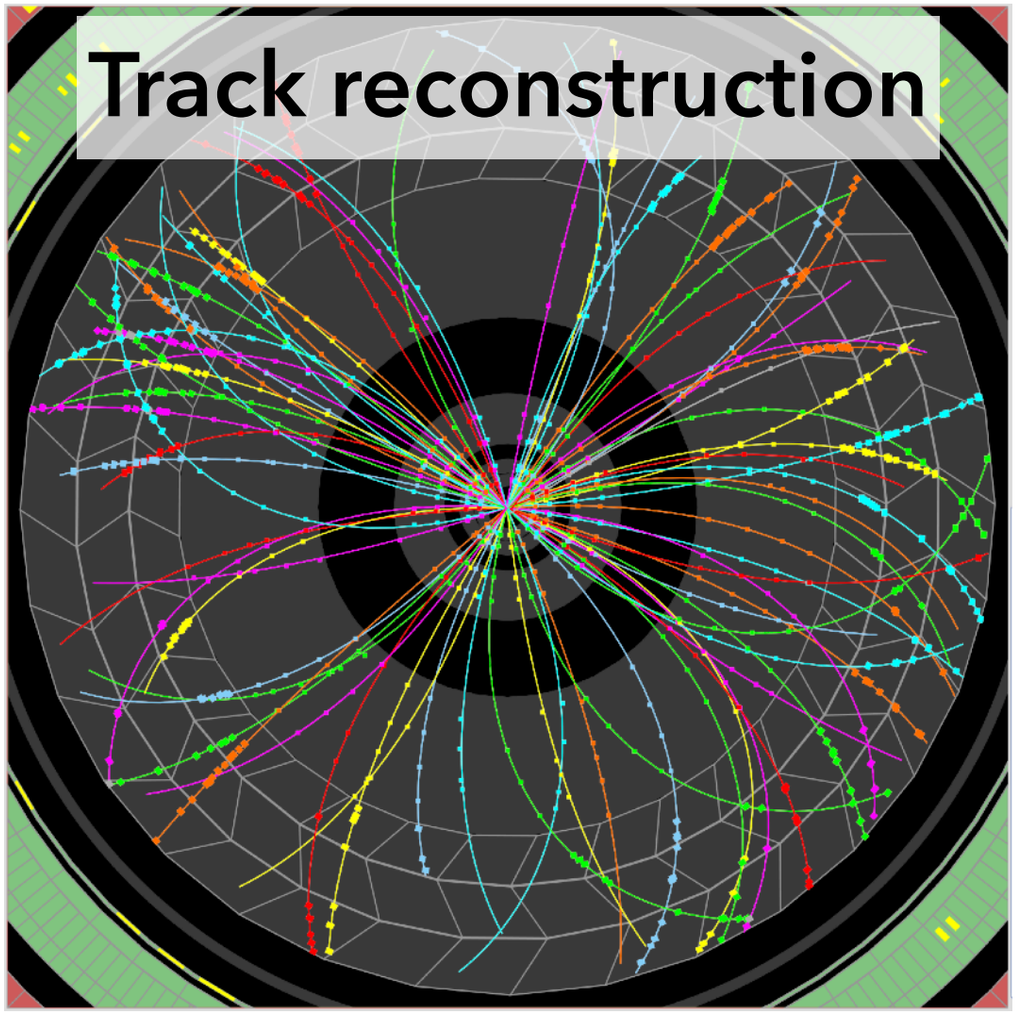


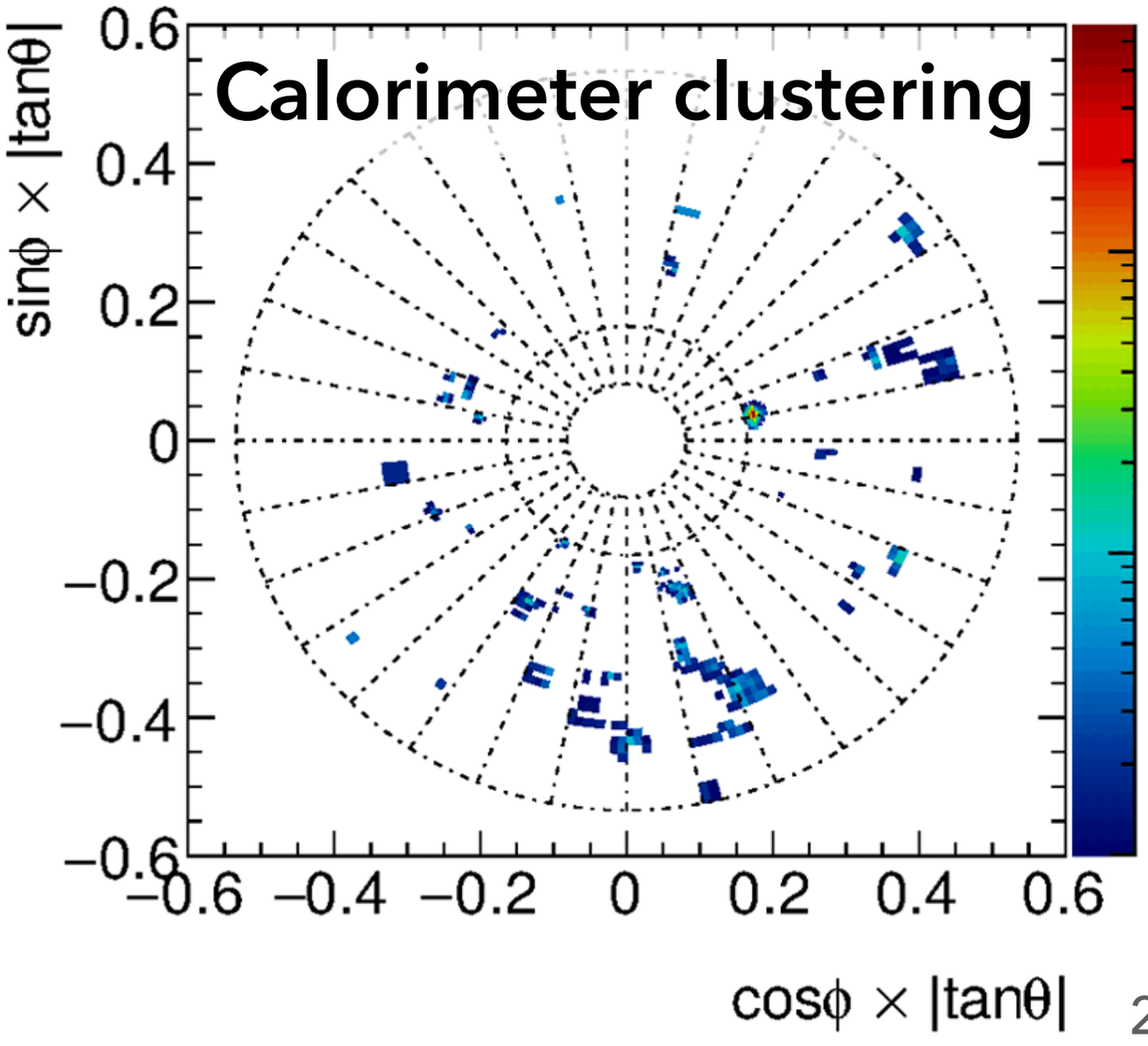



# Dimensionality reduction







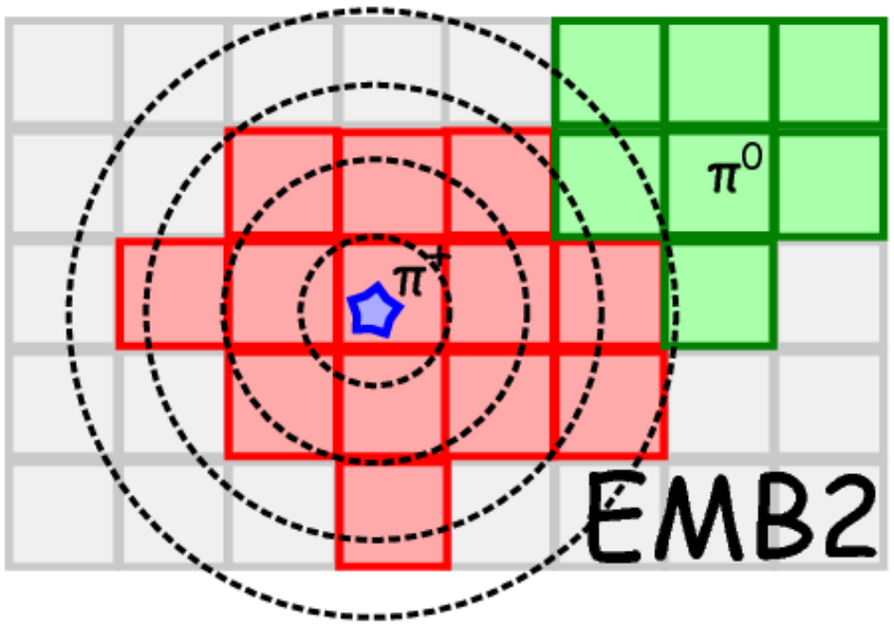



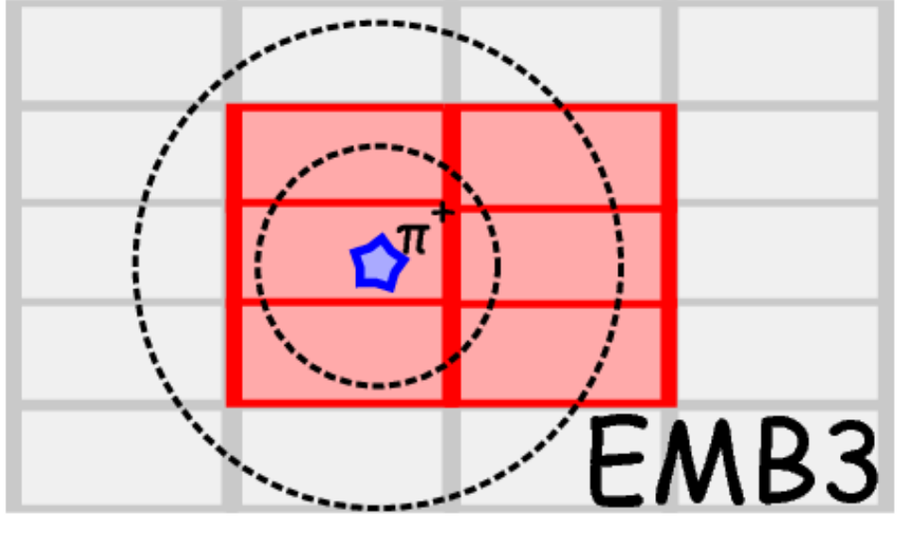
# ATLAS particle flow

*We want to use tracks at low momentum (better resolution)…*

*… but we first need to remove their <u>expected</u> contribution*

Subtract energy from cells
in rings around the track



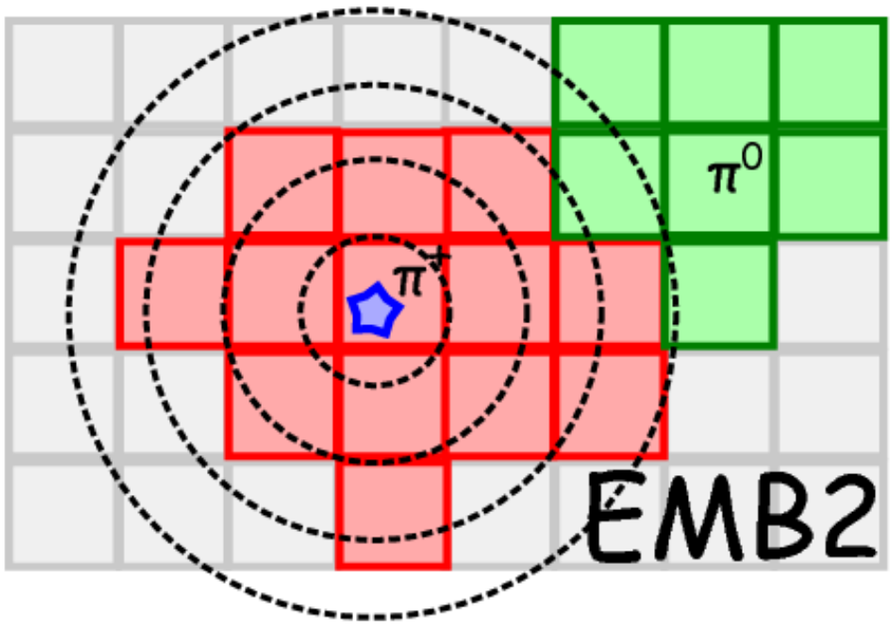



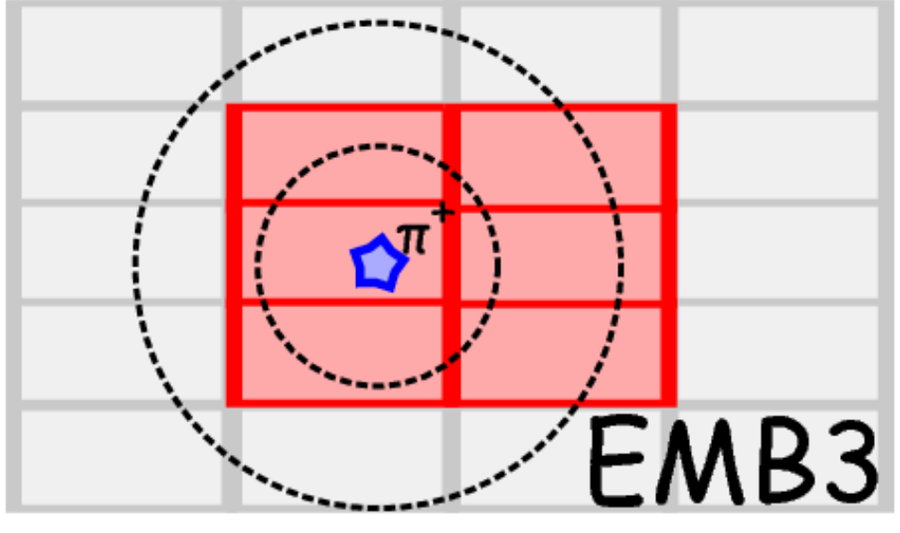
# ATLAS particle flow

*We want to use tracks at low momentum (better resolution)…*

*… but we first need to remove their <u>expected</u> contribution*

Subtract energy from cells
in rings around the track



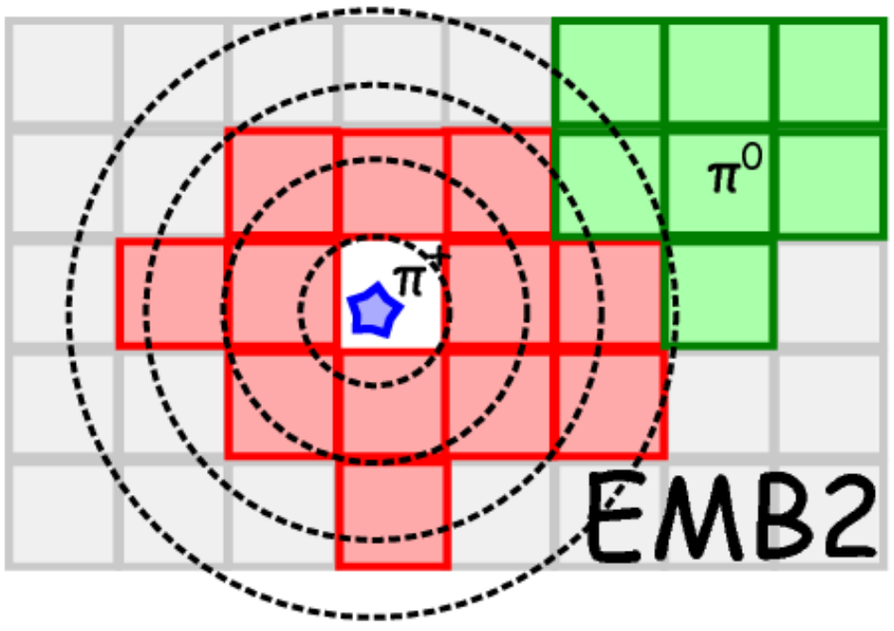



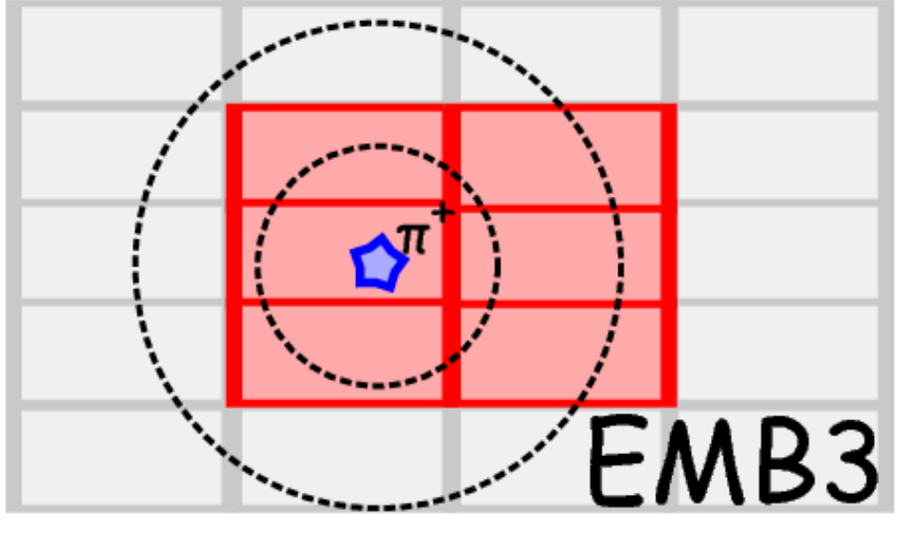
# ATLAS particle flow

*We want to use tracks at low momentum (better resolution)…*

*… but we first need to remove their <u>expected</u> contribution*

Subtract energy from cells
in rings around the track



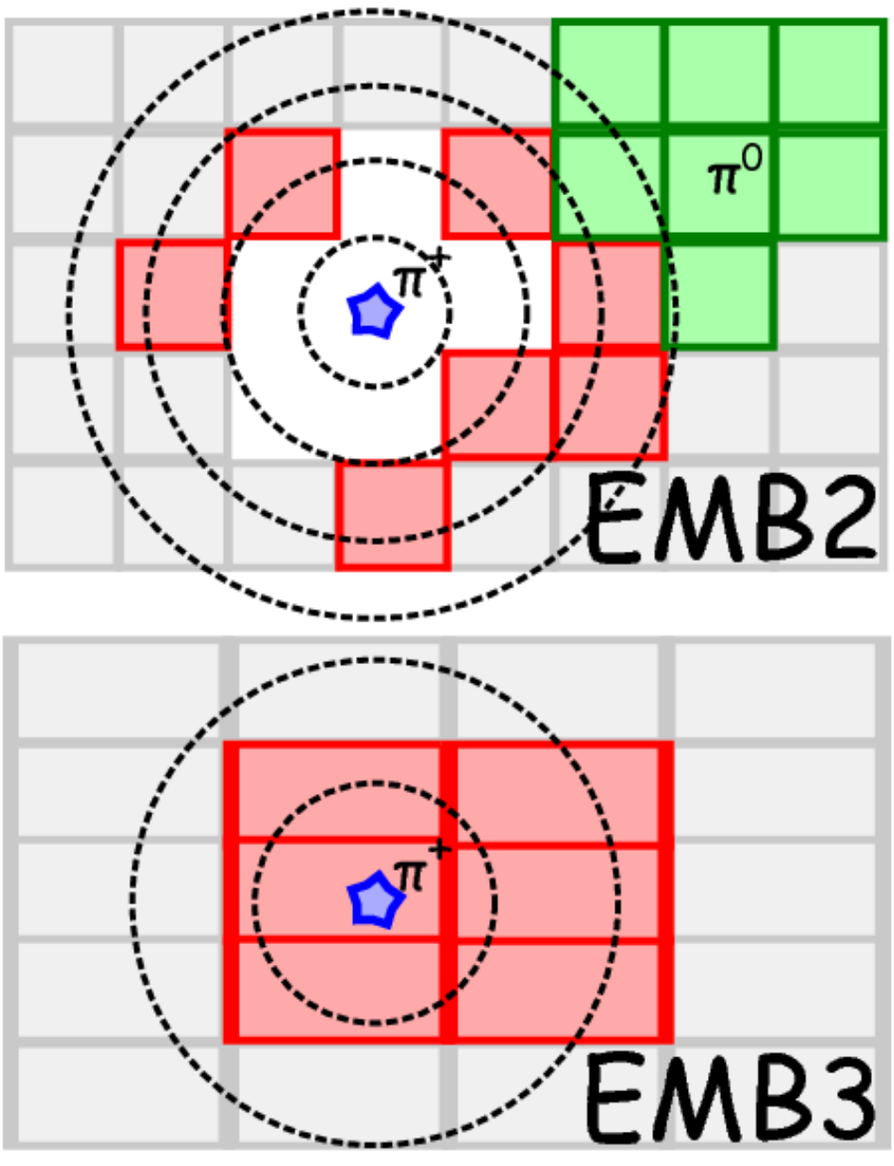



# ATLAS particle flow

*We want to use tracks at low momentum (better resolution)…*

*… but we first need to remove their <u>expected</u> contribution*

Subtract energy from cells
in rings around the track

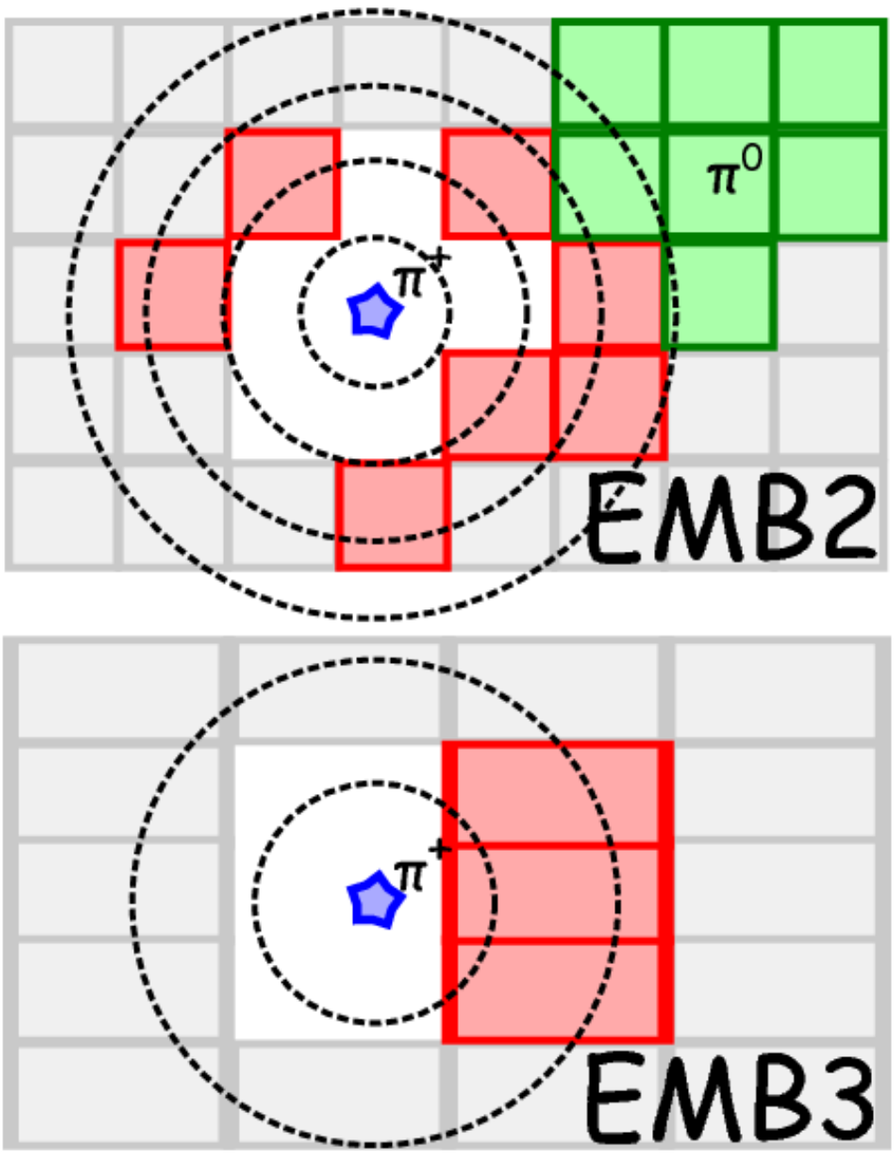


# ATLAS particle flow

*We want to use tracks at low momentum (better resolution)…*

*… but we first need to remove their <u>expected</u> contribution*

Subtract energy from cells
in rings around the track

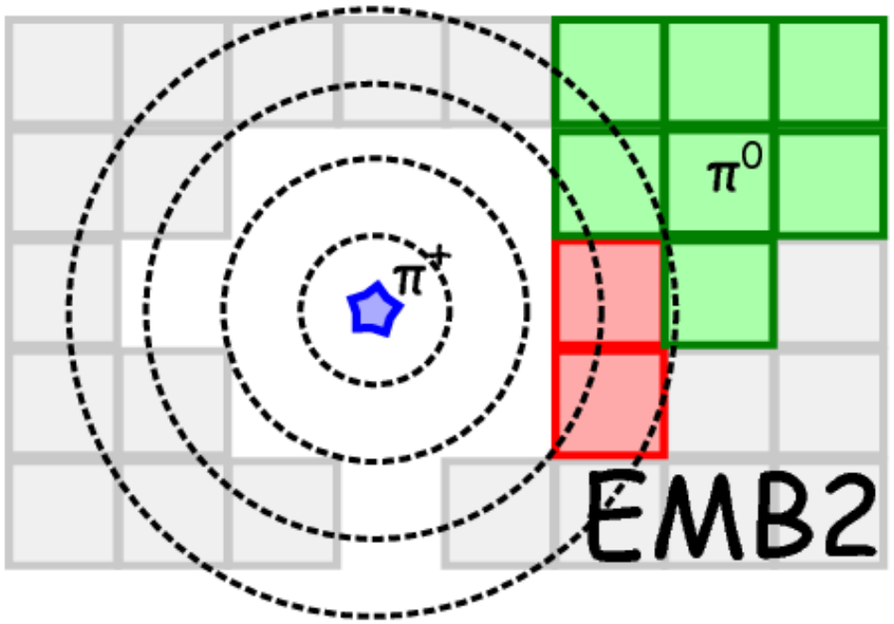


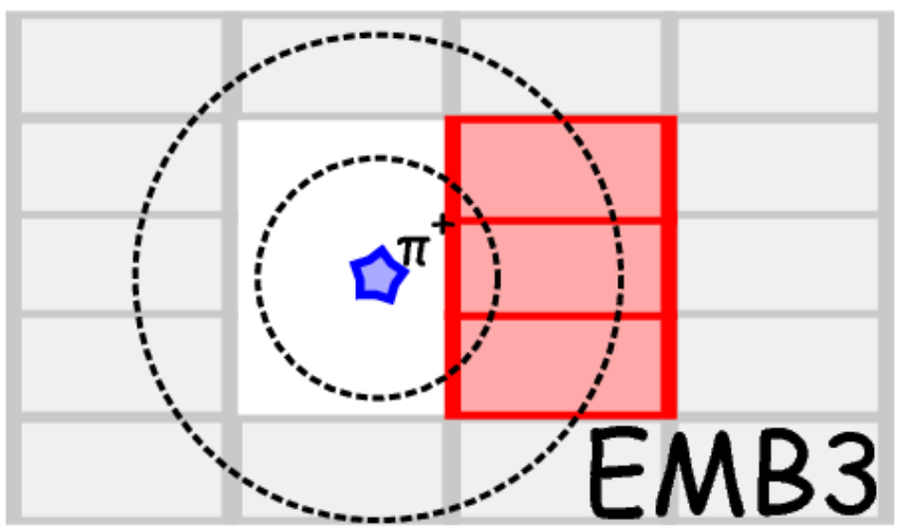
# ATLAS particle flow

*We want to use tracks at low momentum (better resolution)…*

*… but we first need to remove their <u>expected</u> contribution*

Subtract energy from cells
in rings around the track



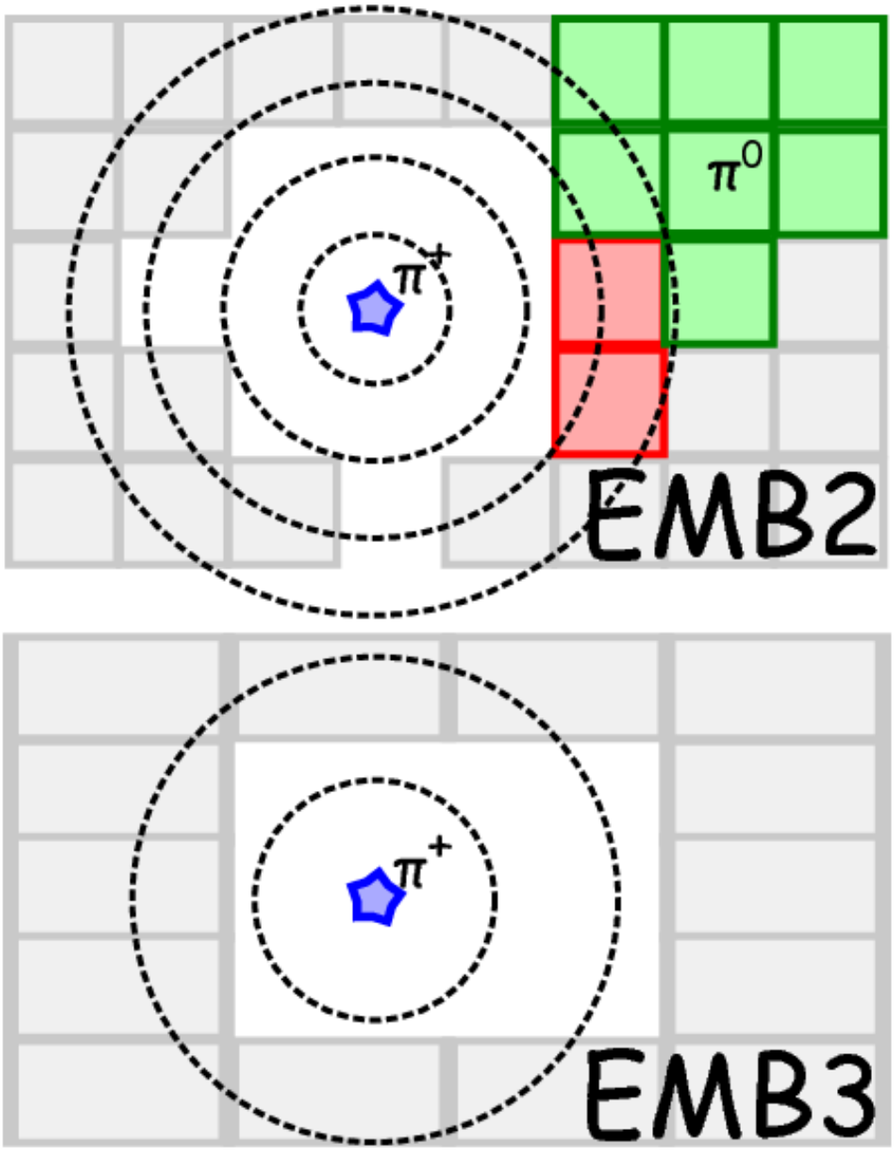



# ATLAS particle flow

*We want to use tracks at low momentum (better resolution)…*

*… but we first need to remove their <u>expected</u> contribution*

Subtract energy from cells
in rings around the track

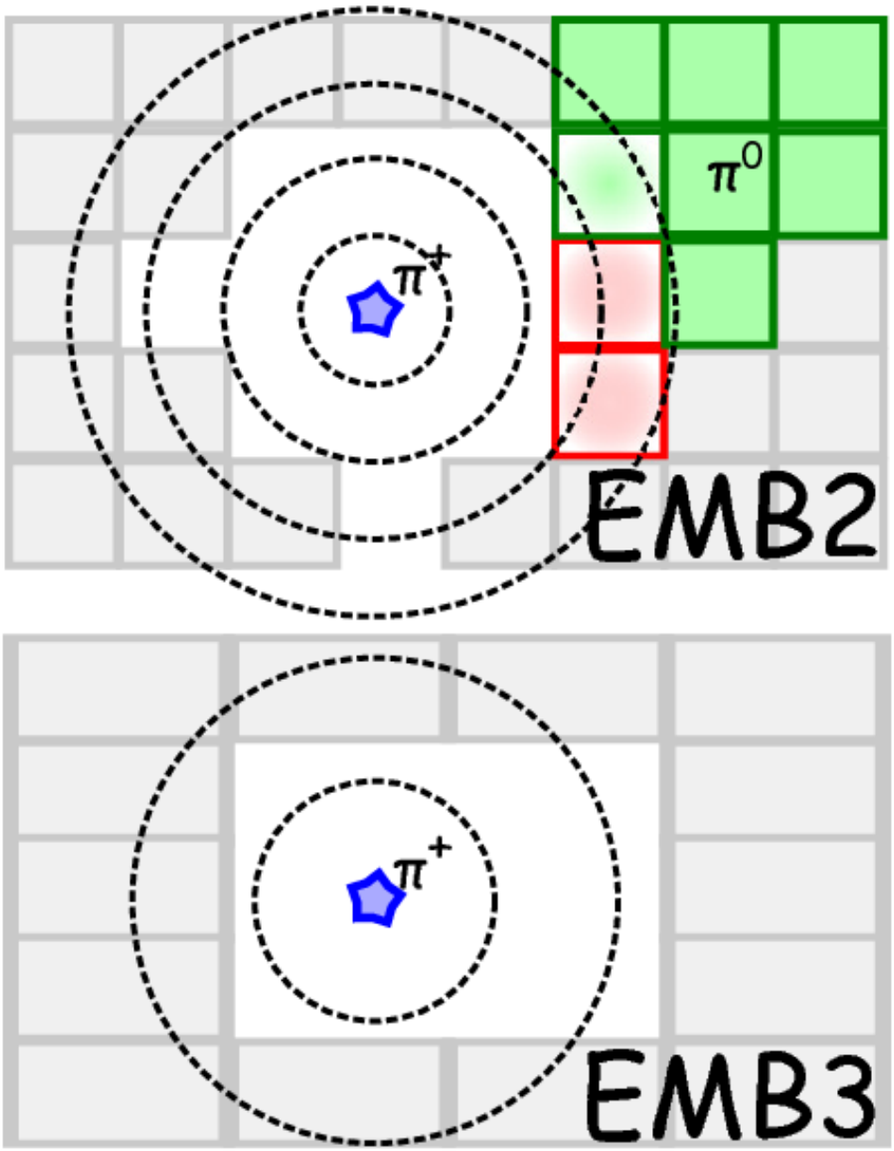


# ATLAS particle flow

*We want to use tracks at low momentum (better resolution)…*

*… but we first need to remove their <u>expected</u> contribution*

Subtract energy from cells
in rings around the track

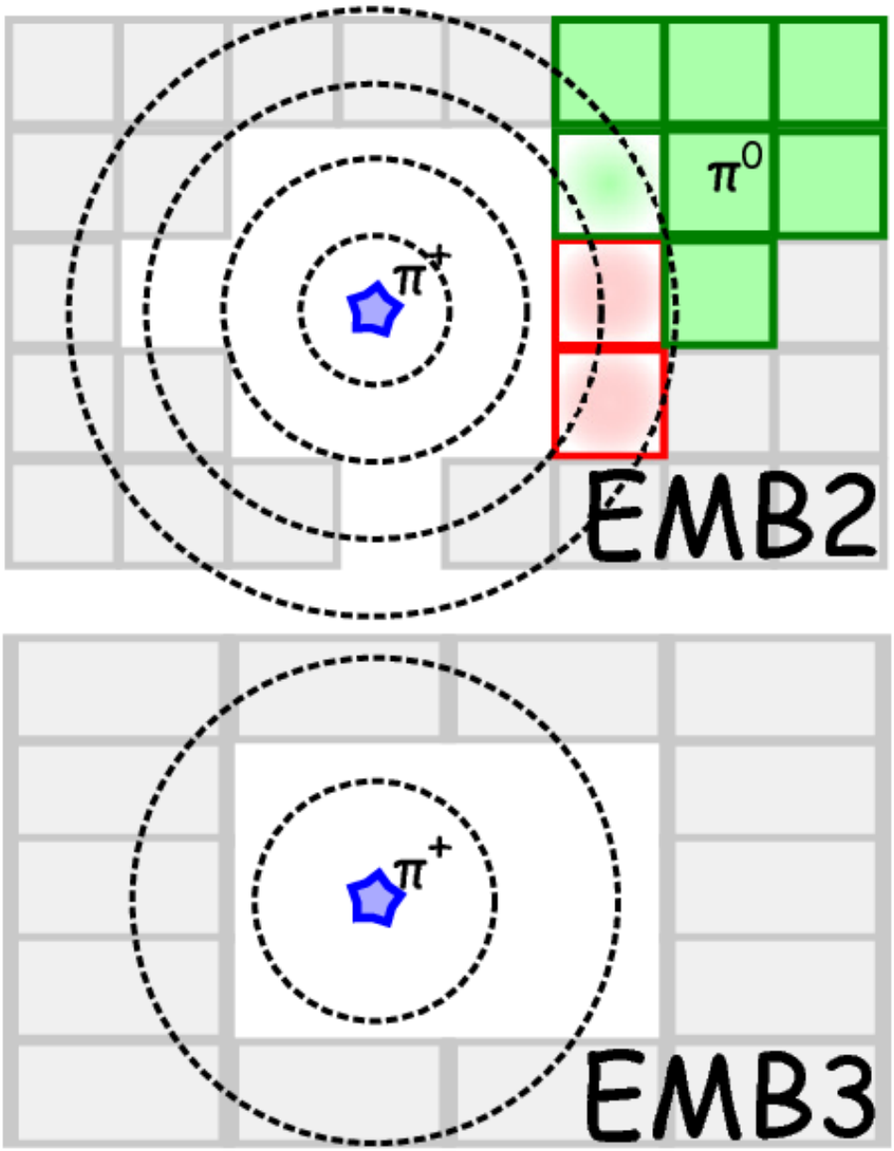


# ATLAS particle flow

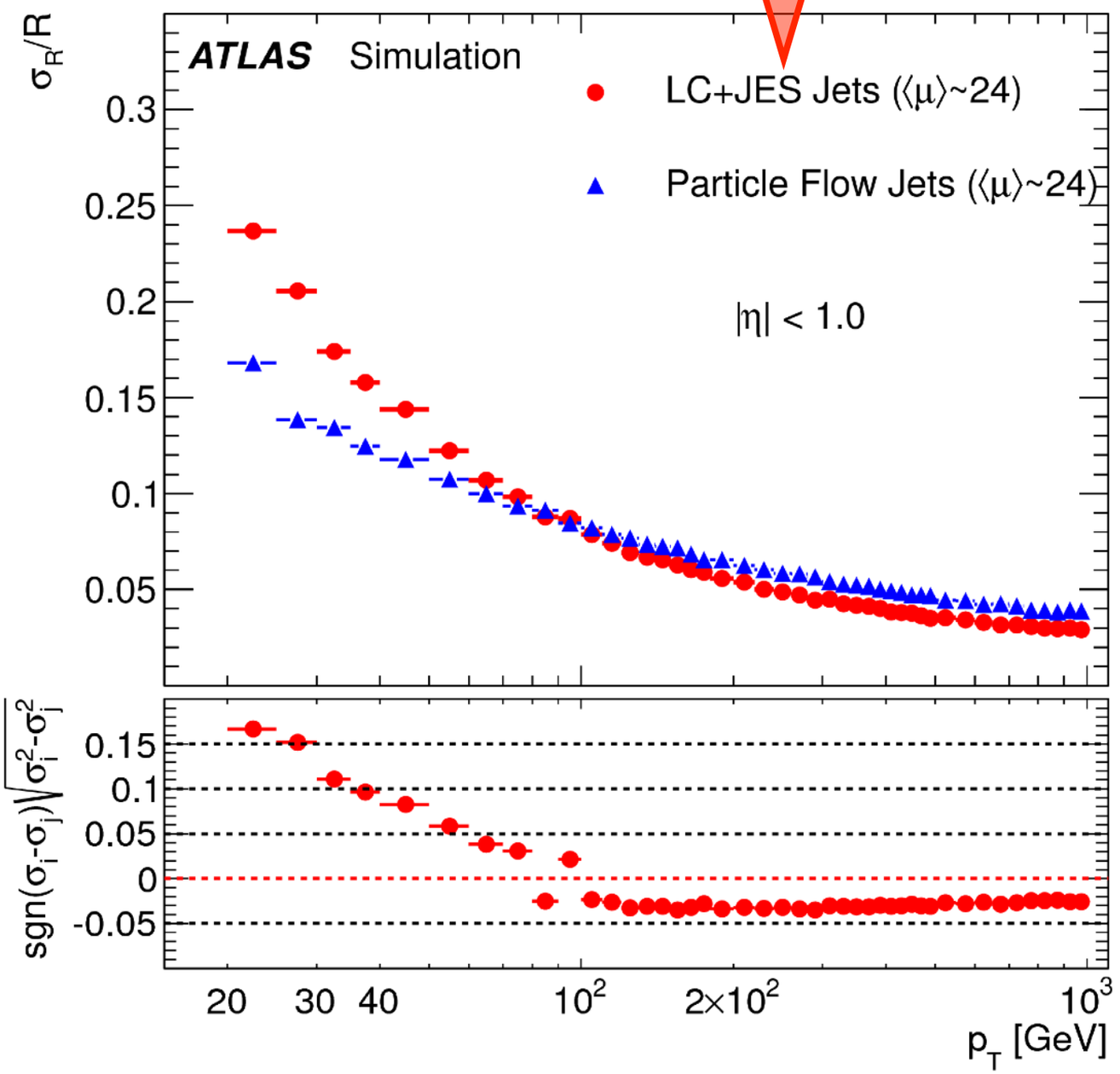
***We want to use tracks at low momentum (better resolution)…***

***… but we first need to remove their <u>expected</u> contribution***

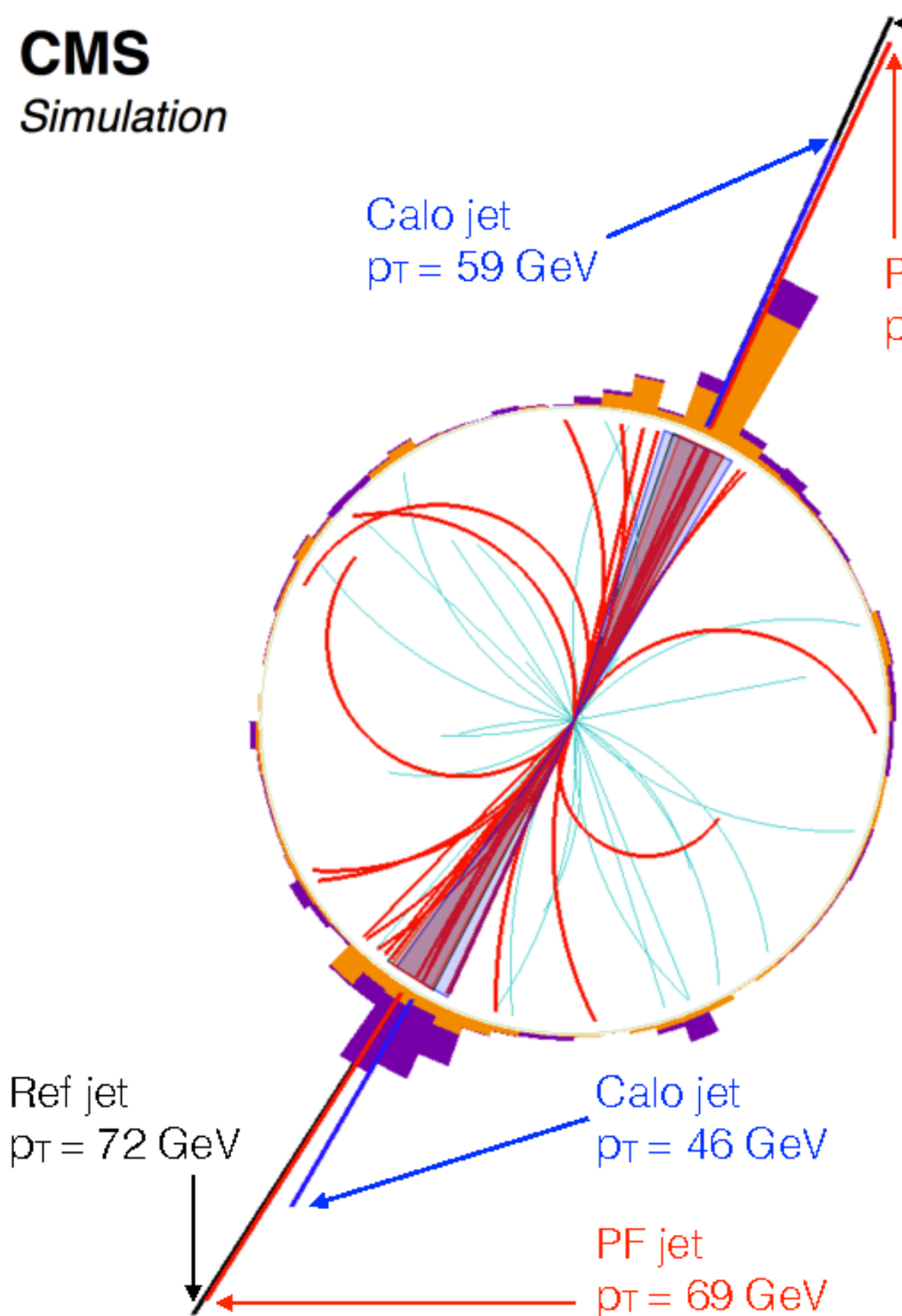
Subtract energy from cells
in rings around the track




N.B. Comparing calibrated and uncalibrated jets



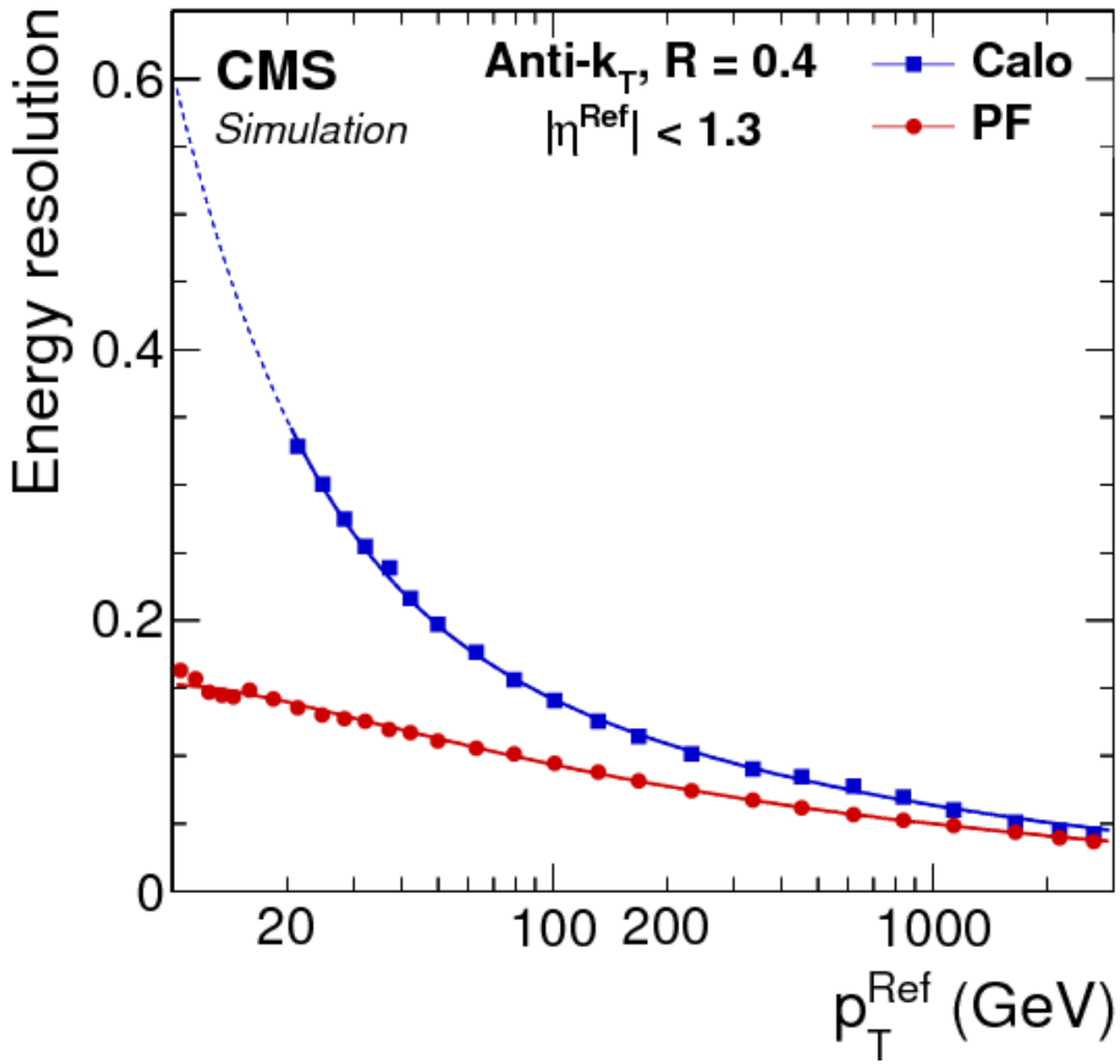
# CMS particle flow



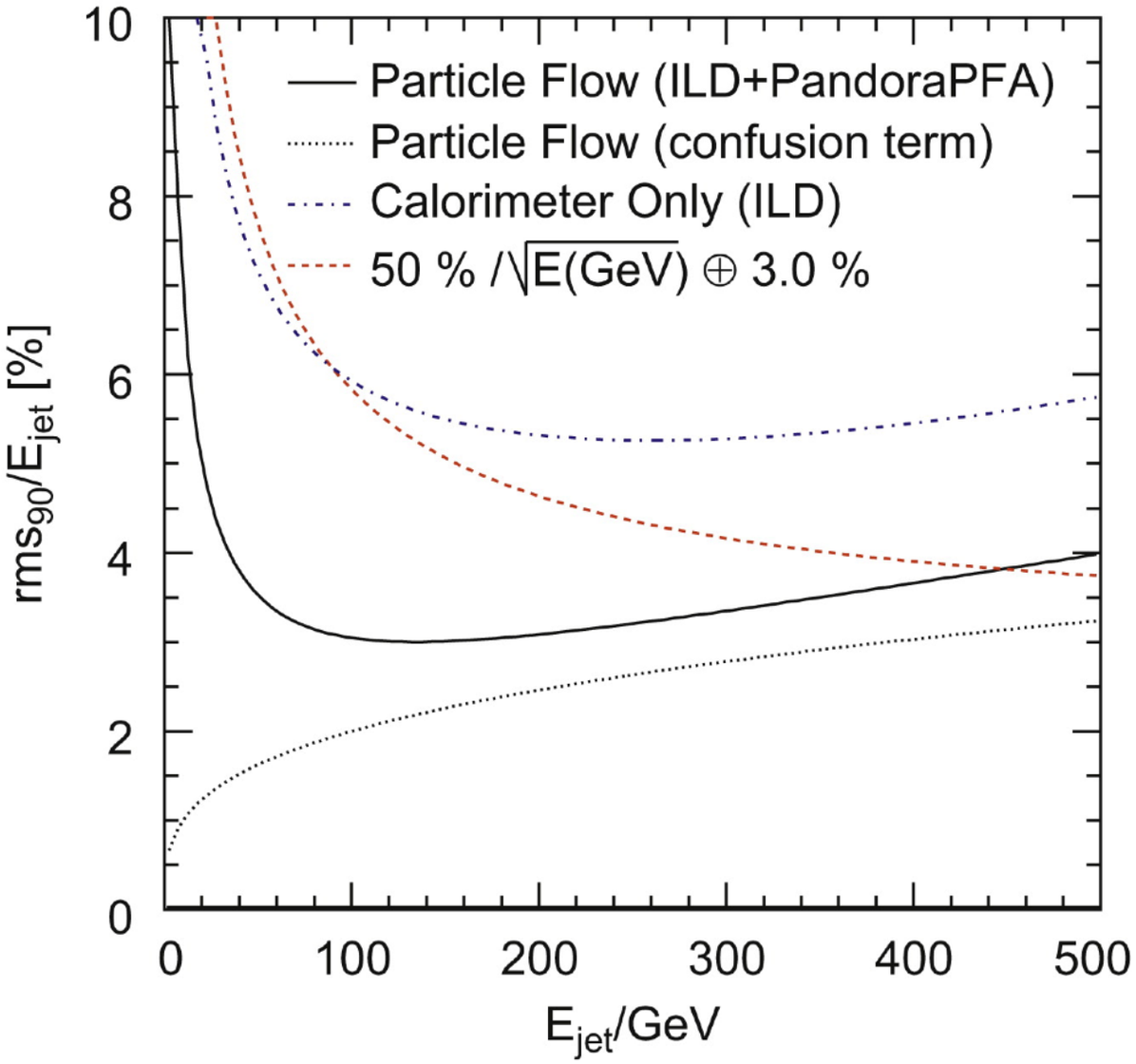

**Calorimeter-only**

**vs.**

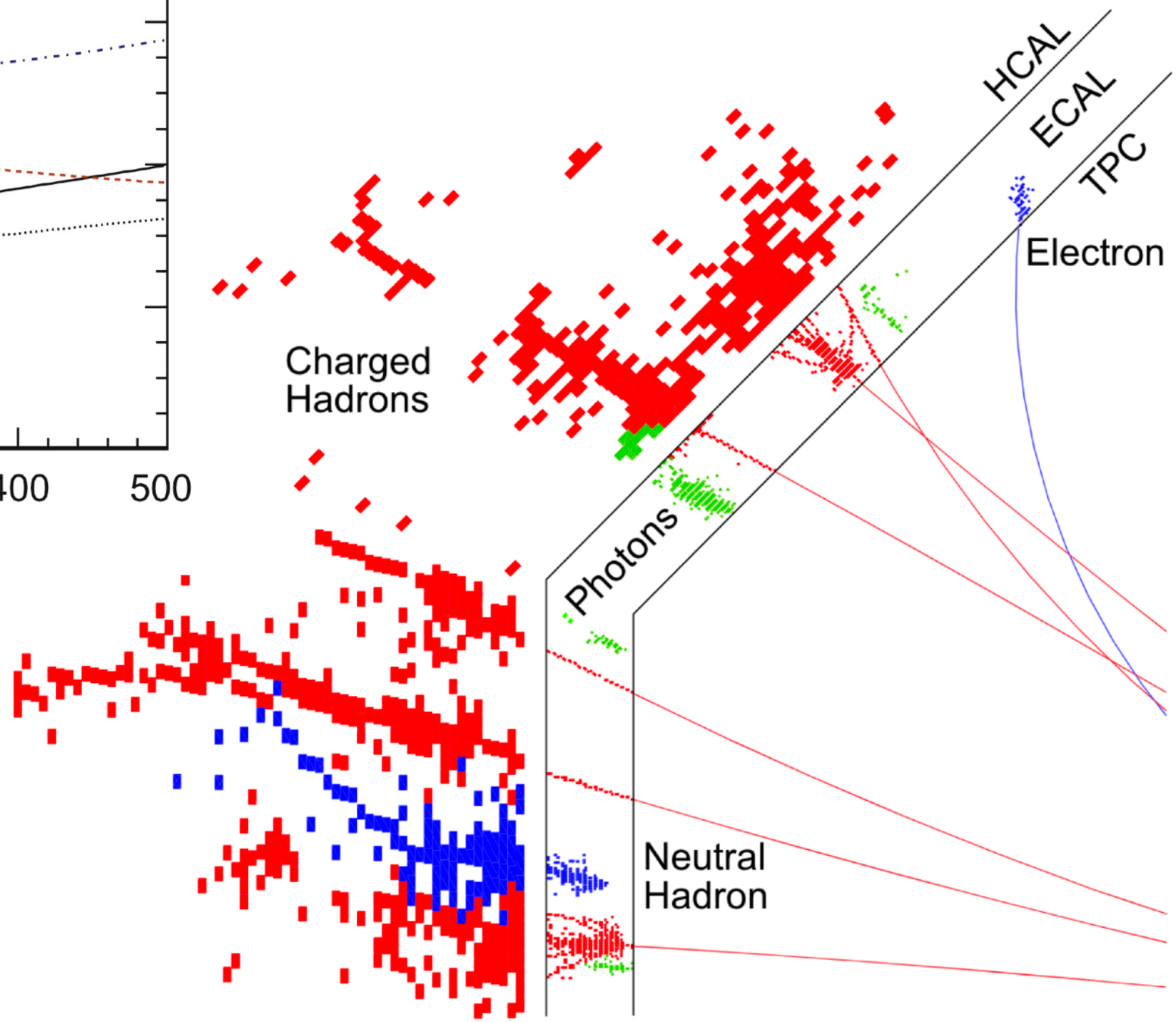
**Particle flow**



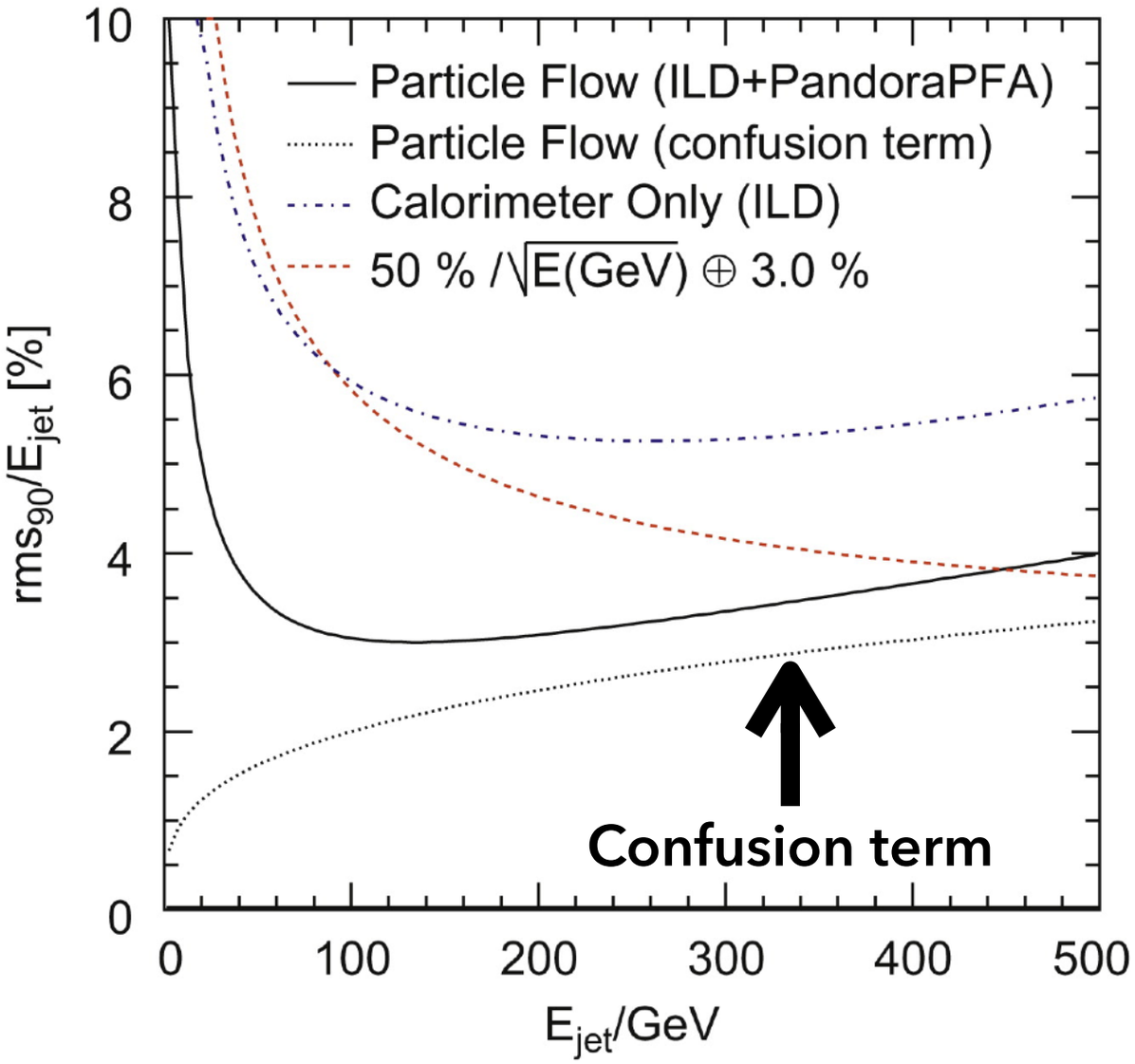
# Pandora: particle flow for CLIC




- Multiple pattern recognition steps
- Highly-granular calorimeter
- Cleaner $e^+e^-$ collision environment



# Pandora: particle flow for CLIC




- Multiple pattern recognition steps
- Highly-granular calorimeter
- Cleaner $e^+e^-$ collision environment