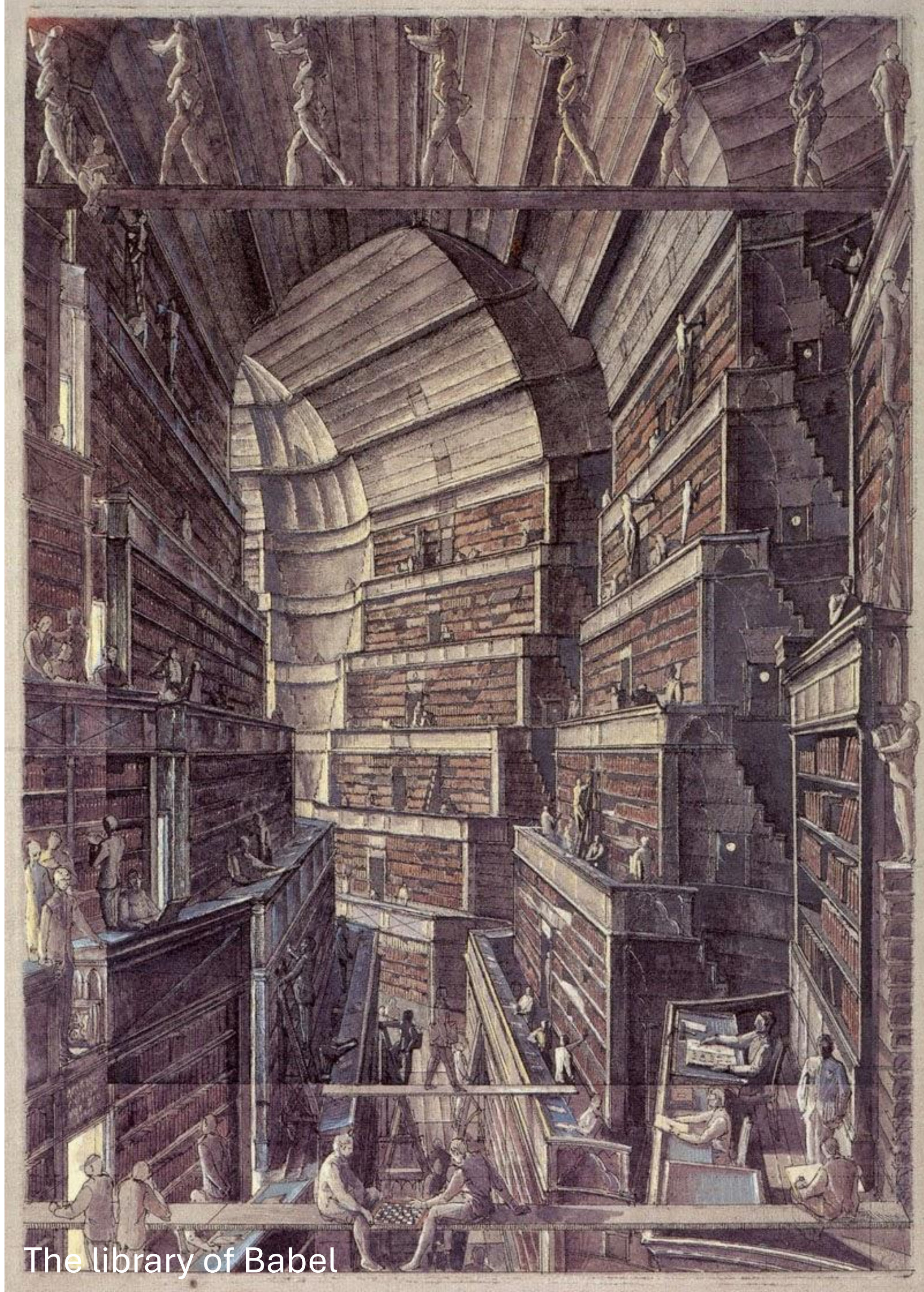
AI for LHC experiments

A personal perspective with some biases…

Francesco Armando Di Bello
Università di Pisa. 7/10/25

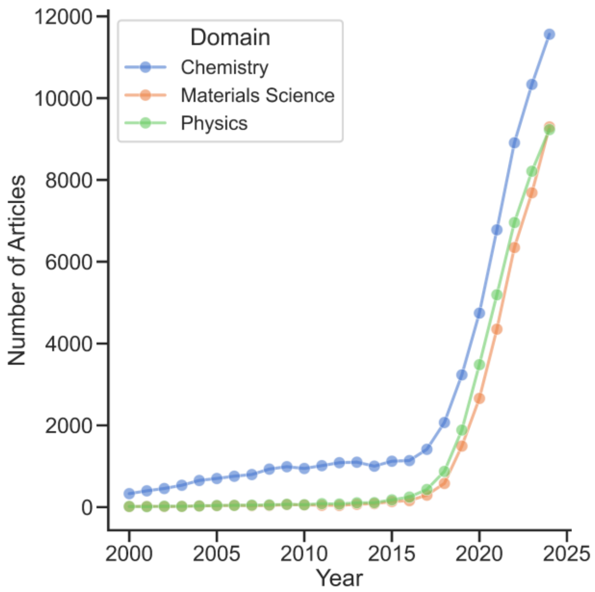


Outline of the talk

- We will discuss:

  1. State-of-the art machine learning algorithm used
  2. The problem of domain adaptation: calibrations
  3. Set-to-Set algorithms: particle-flow reconstruction and jet identifications
  4. Joining minimization all together? Differentiable programming for end-to-end pipelines

We will NOT discuss: fast generation of events, unfolding, Simulation based inference, uncertainty quantifications and fast inference



Number of papers are exploding with great majority originating from proof-of-principles

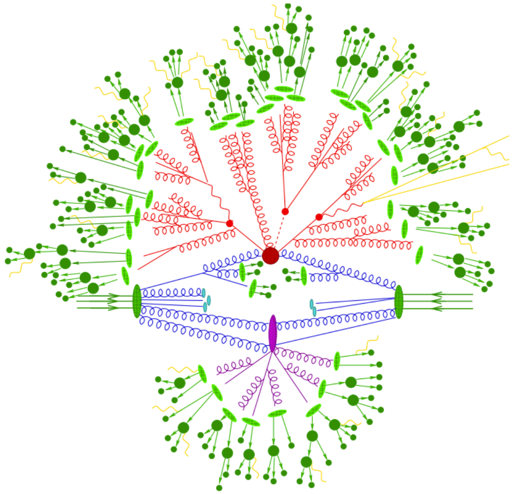
ml_publication_charts

Introduction

- Particle physics experiments are based on hierarchical, factorized pipelines whose goal is to allow unbiased parameters inference.

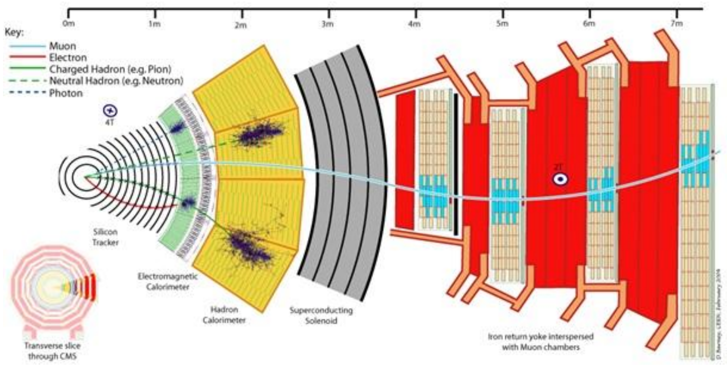
$$p(x \mid \theta) = \int dR'\, dR\, dH\, dz\; \delta\big(x - g(R')\big)\; \delta\big(R(H) - R\big)\, p_{\mathrm{sim}}(H \mid z)\, p_{\mathrm{gen}}(z \mid \theta)$$

$p_{\mathrm{gen}}(z \mid \theta)$



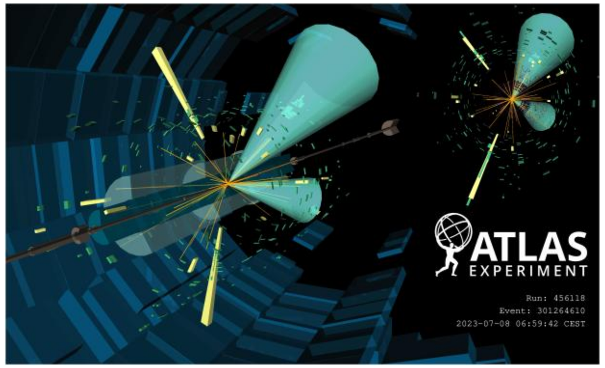
Event generation

$p_{\mathrm{sim}}(H \mid z)\, p_{\mathrm{gen}}(z \mid \theta)$



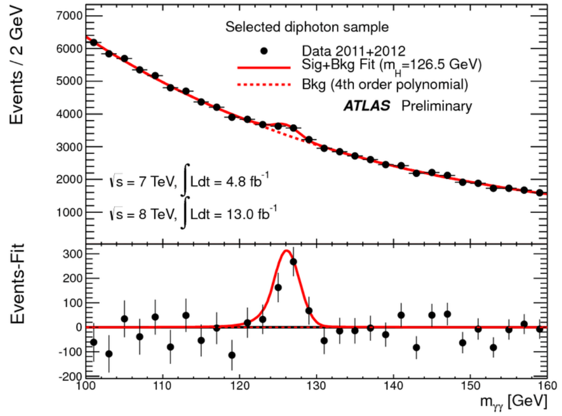
Detector interaction

$\delta\big(R(H) - R\big)\, p_{\mathrm{sim}}(H \mid z)\, p_{\mathrm{gen}}(z \mid \theta)$



Recontruction

$p(x \mid \theta)$

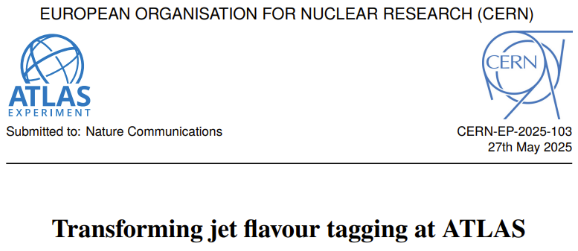


Inference

Jet identification

Peraphs the main example of the impact of ML for LHC experiments
This is used both for offline but also online data analysis

This is opening new doors in understanding for instance the Higgs boson

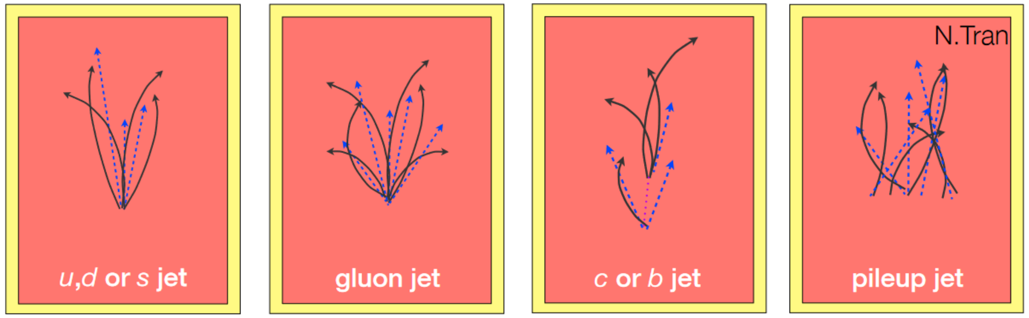

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: Nature Communications

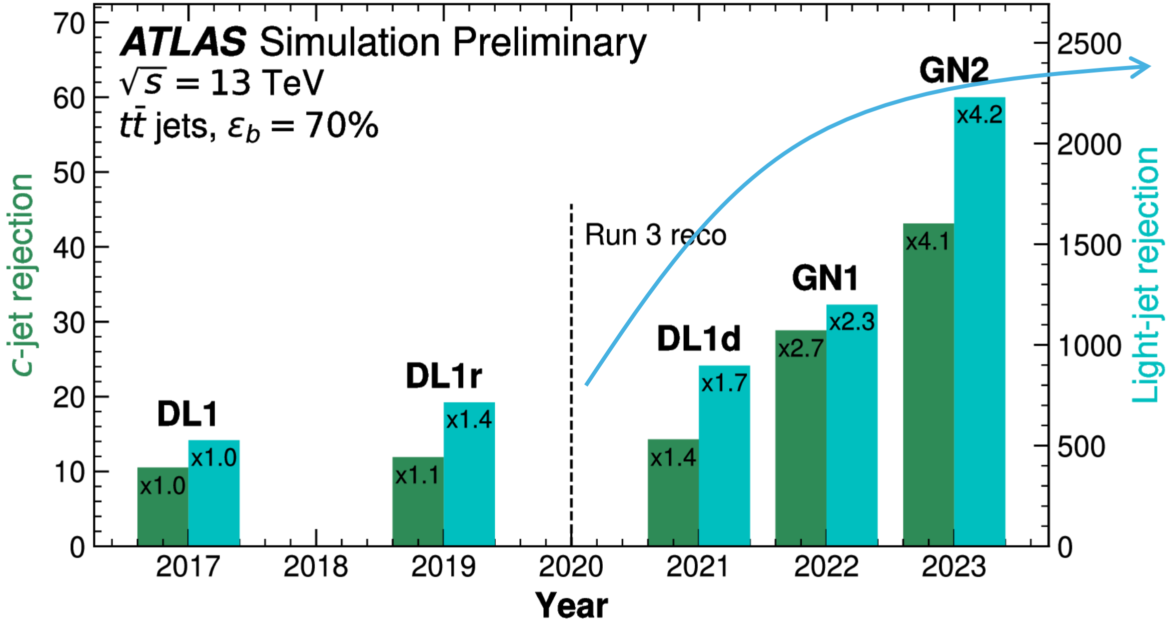
CERN-EP-2025-103
27th May 2025

Transforming jet flavour tagging at ATLAS

[CMSOpenData]



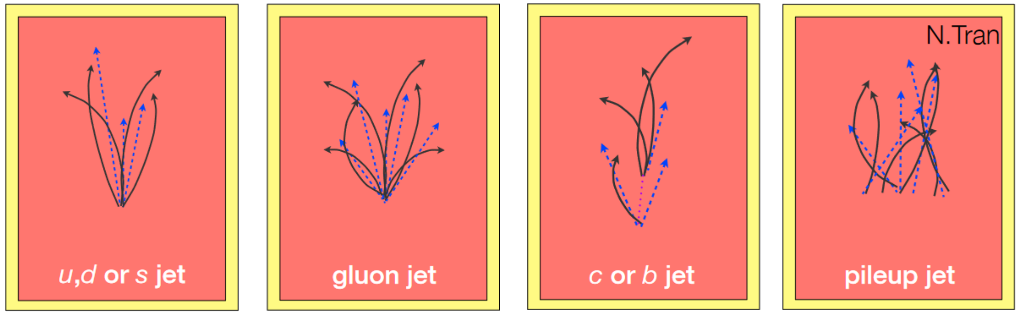
Are we saturating?

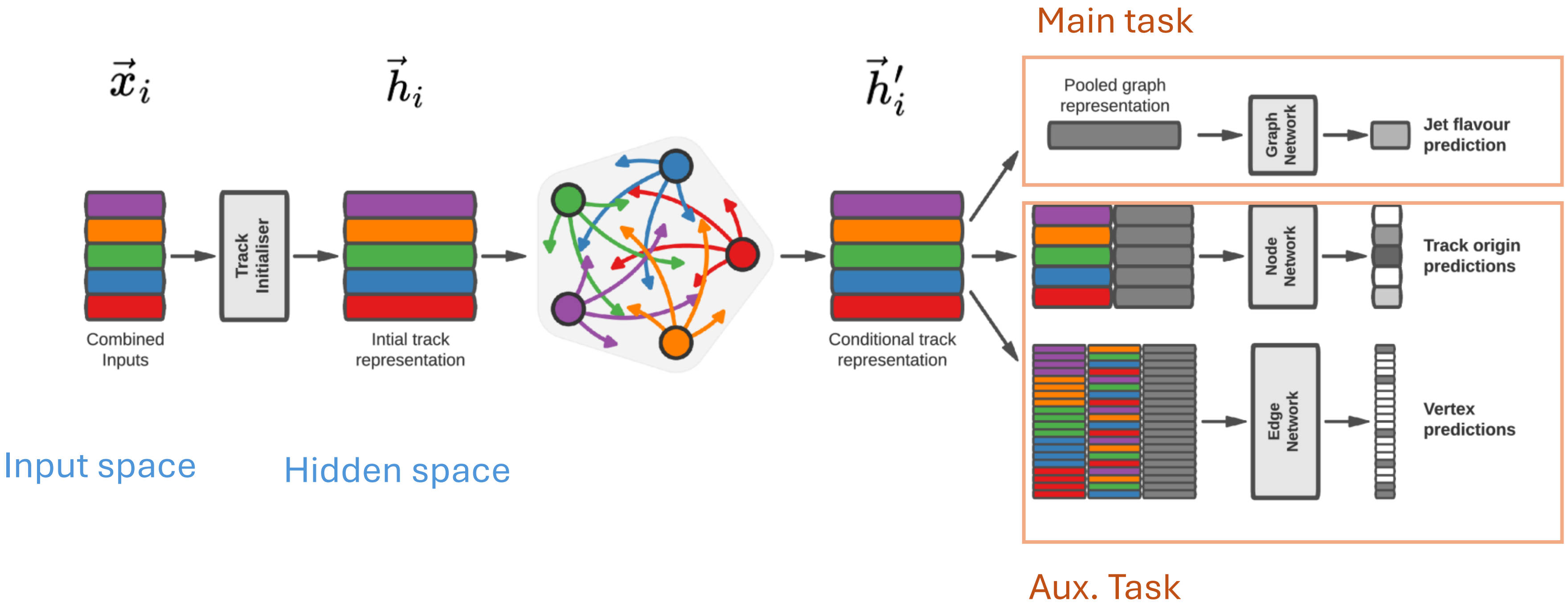


Accepted by Nature Comm.
[2505.19689] Transforming jet flavour tagging at ATLAS

Jet identification

- Inputs are recontructed low-level object: no vertex information
- Hidden dim. h is O(100)



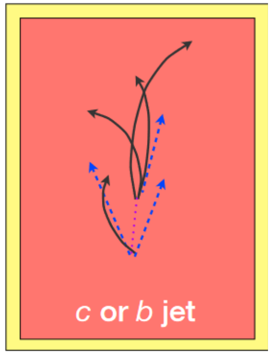


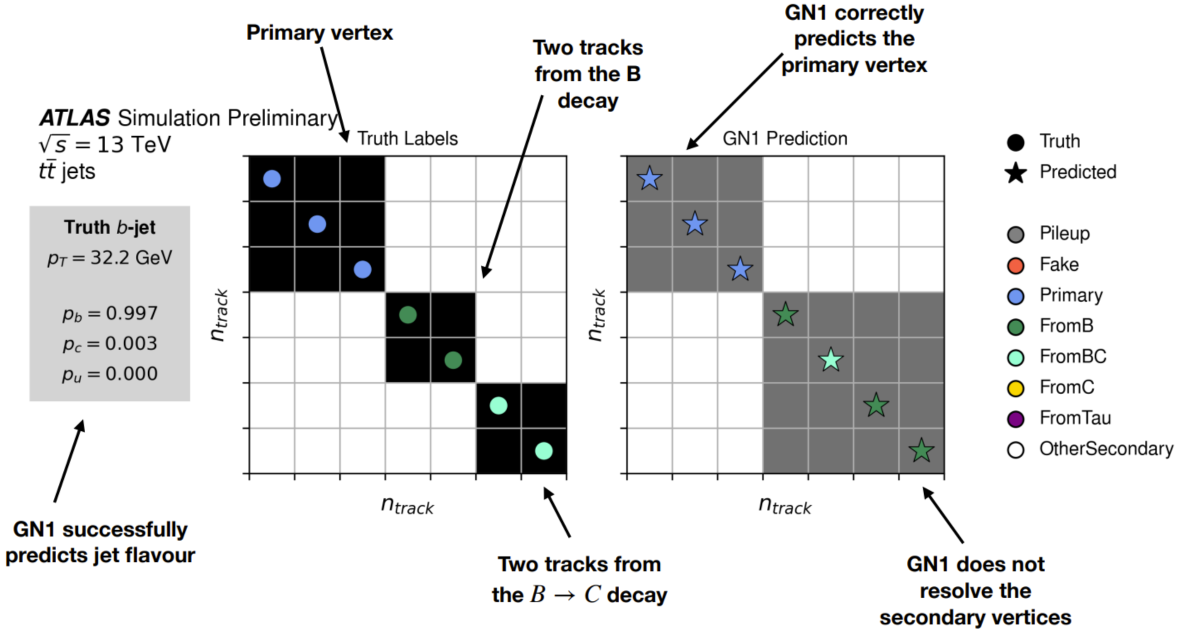
Total loss function

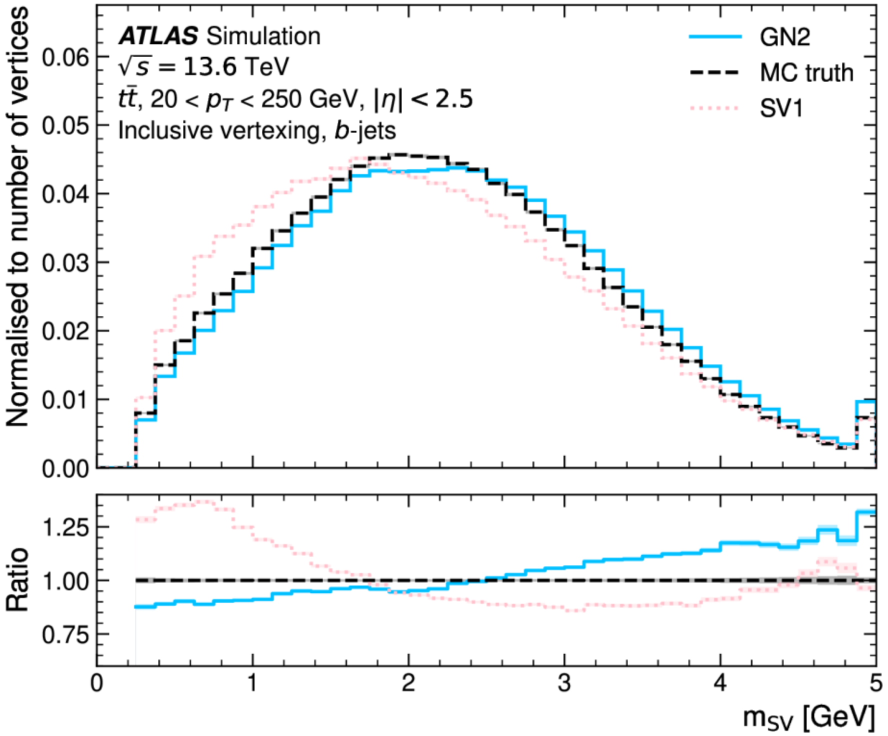
$$L_{\text{total}} = L_{\text{jet}} + \alpha L_{\text{vertex}} + \beta L_{\text{track}}$$

Jet identification

- Aux. tasks helps both performance and intepretability



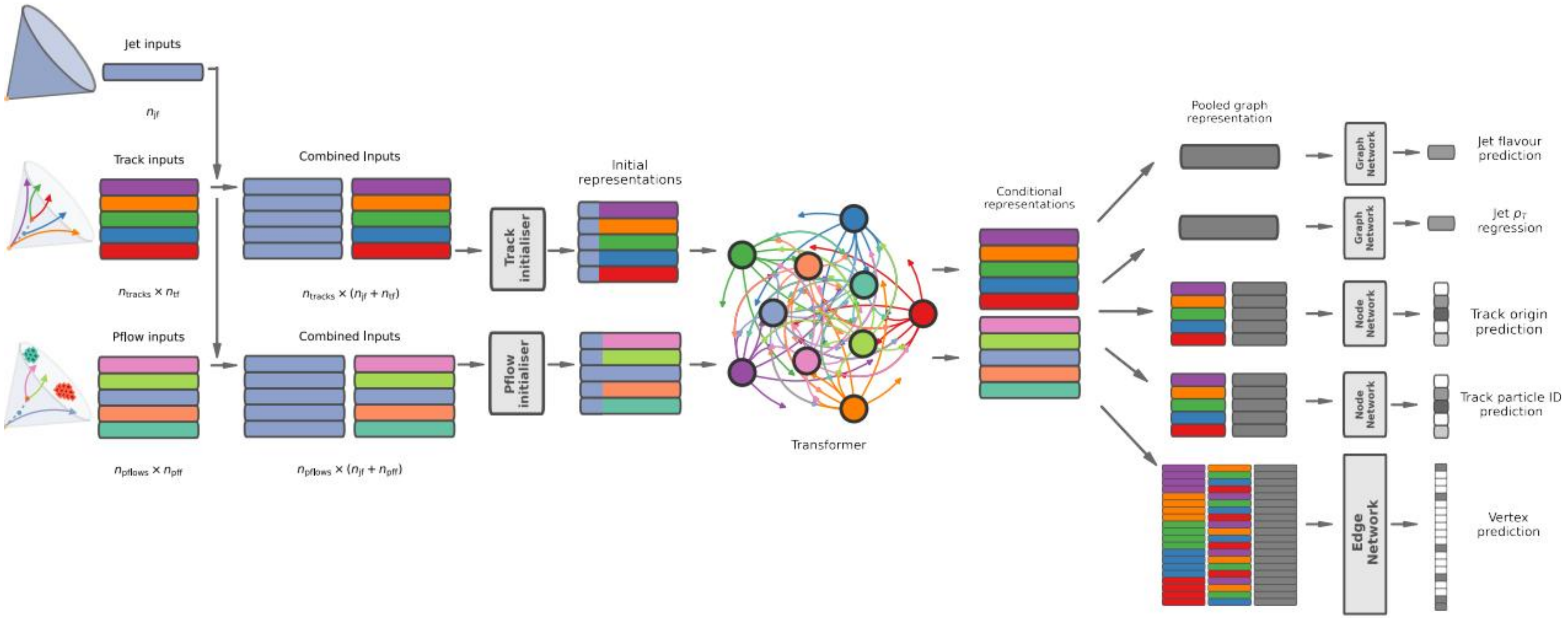




Examples from aux taks

The state-of-the art now

- State of the art flavour tagging algorithm. This start looking like a foundation model (yet a selected one)



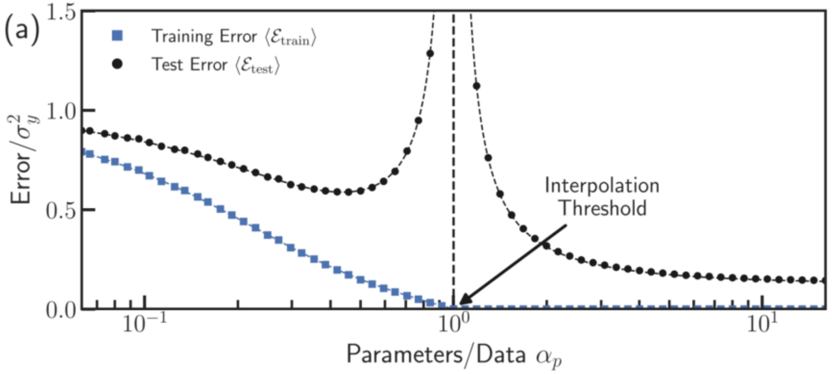
Energy regression to correct for jet momentum embedded into the model.

Does it still make sense to do tau- and b-tagging separately? ML helps not only performance but also to streamline workflow

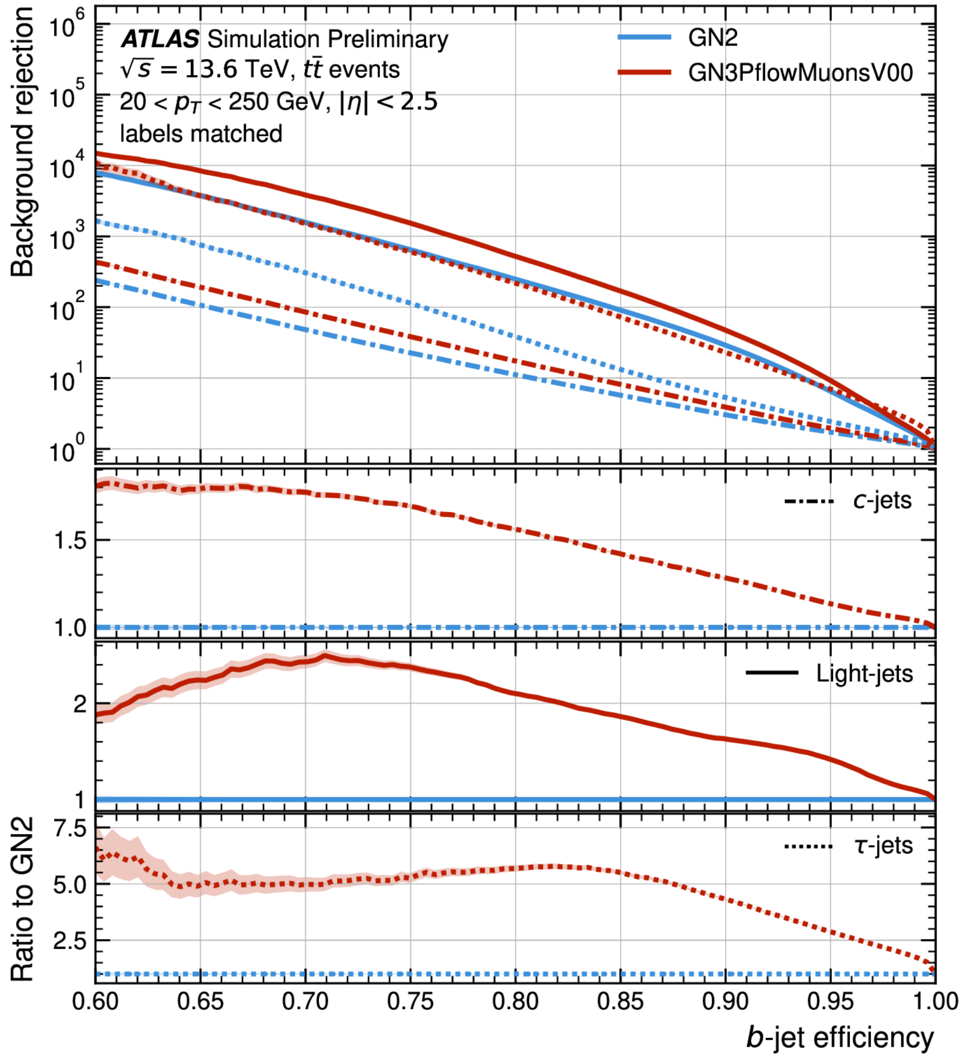
Jet identification

- We are not yet saturating yet…
- Main improvements from:
  - Larger architecture and data sample
  - Additional inputs (calo+soft leptons)

Where is the limit? Where is double-descent? Generalization?



[2509.01397] L. Henirich et al.

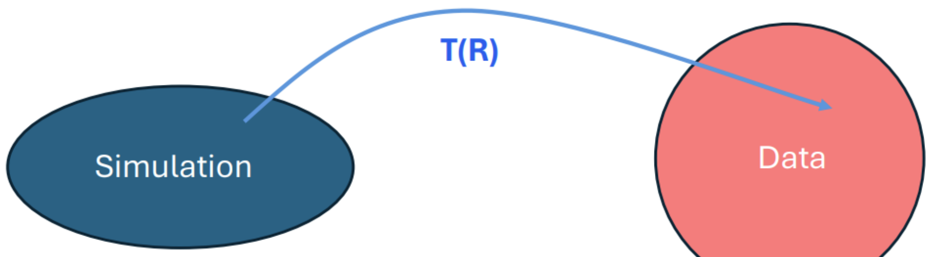


GN3, ATLAS plots

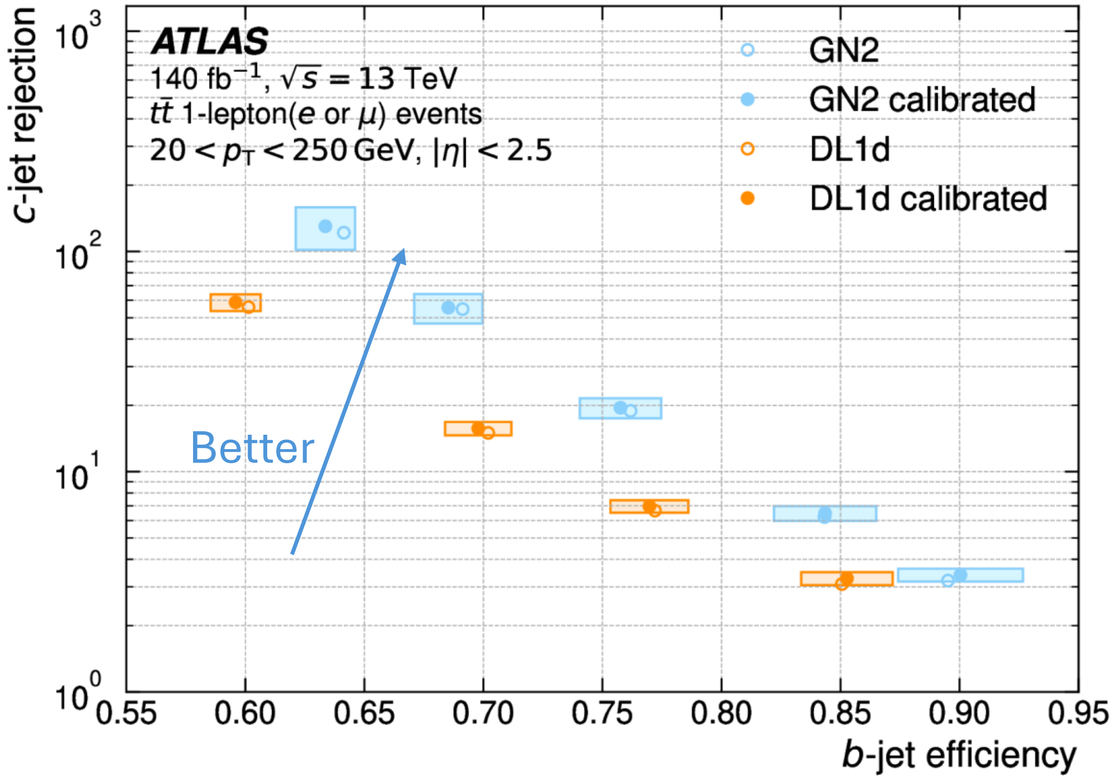
Domain shifts: calibrations

- How do we know it is well modeled in data?

$$p_{\rm calib}(x|\theta) = \int dR'\, dR\, dH\, dT\, dz\, \delta(x - g(R'))\boxed{\delta(R' - T(R))}\delta(R(H) - R)\, p_{\rm sim}(H|z) p_{\rm gen}(z|\theta)\,.$$

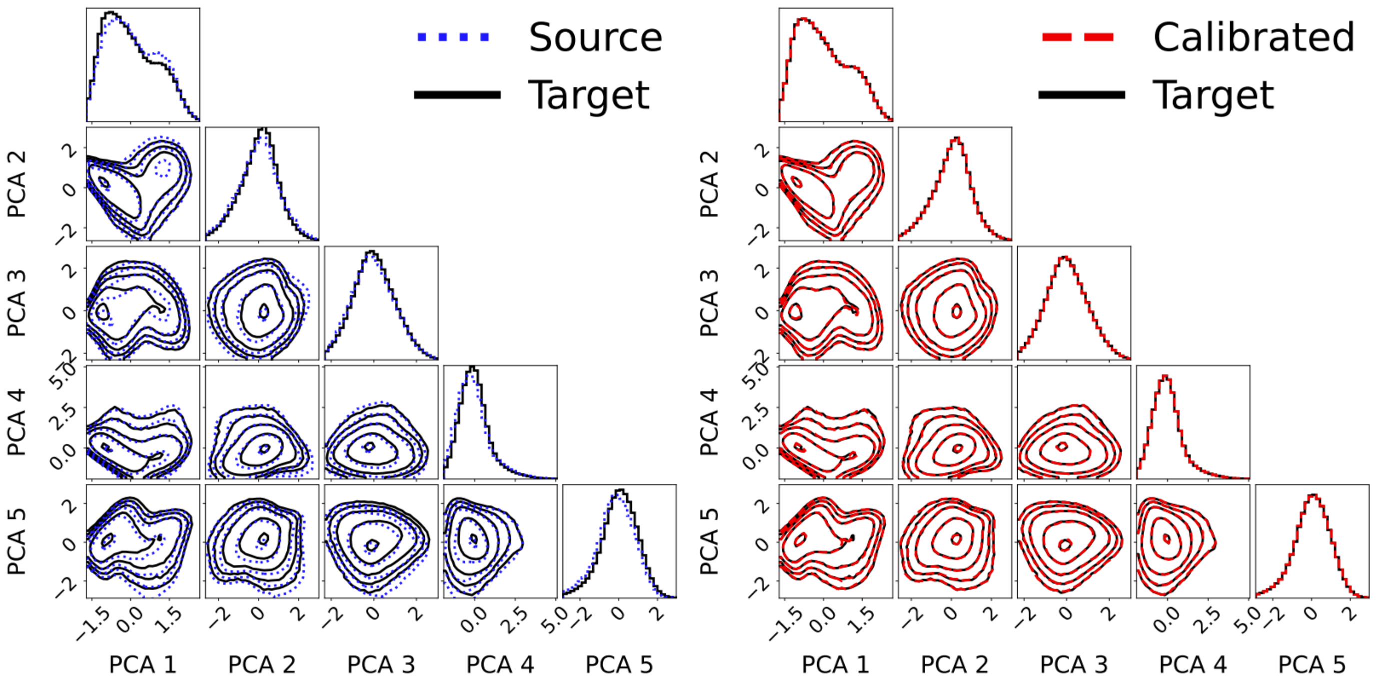


Calibrated performance



Improving calibrations with ML

- One way to use ML to improve calibrations is using optimal transport theory
- This is done and works on low-dimensional data [2505.13063], targeting quantities like: energy of a particle etc…
- Can be generalized to a broader class considering a physics object in its hidden representation space **h** (foundation model vision)
- Typical dimension of h is around 100 -- can we calibrate a 100 dimensional space?
- Looks like we can, and downstream tasks are also calibrated

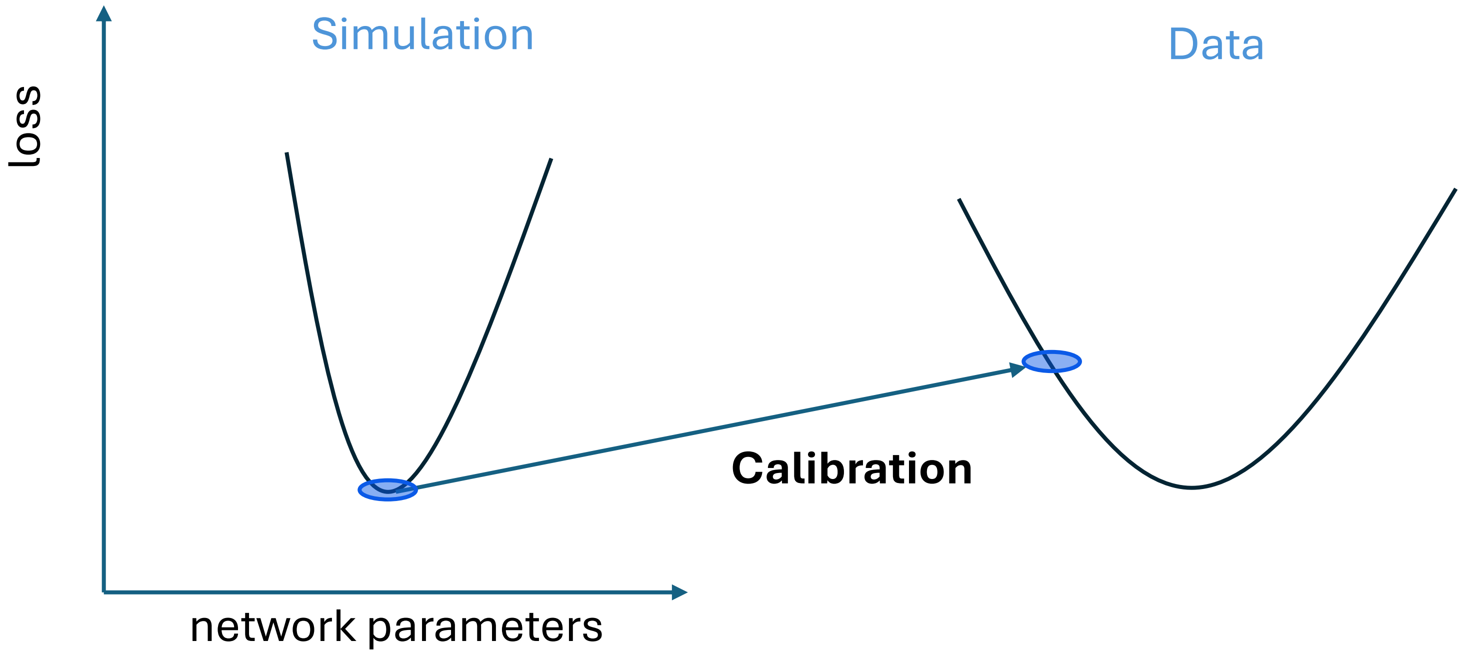


Alpegen, FDB et al 2507.08867

Domain shift: calibrating ML algorithms

A question rises naturally: if we train on simulation and correct them to look like the data, are we at the optimal performance in the data?

Clearly not in general…

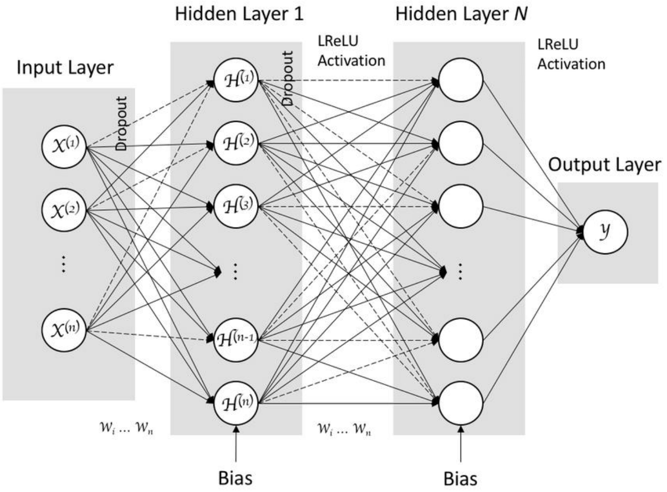


This is an active area of developements, still no clear strategy available.

Review of architectures

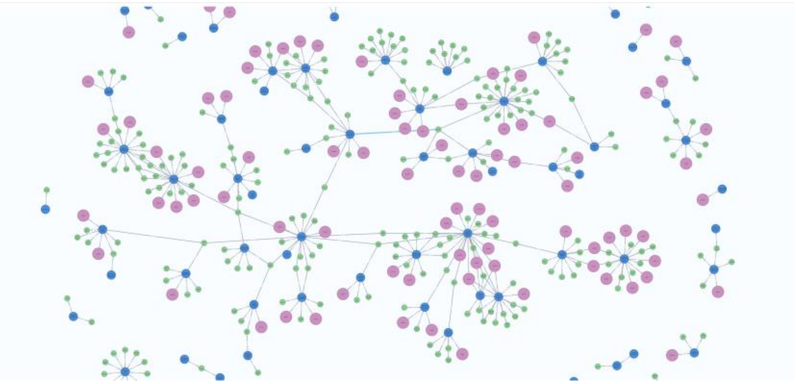
- Taking a step-back on the algorithms used in exp. HEP to improve data analysis
- A very large portion of problems are related to the more complex **Set-to-Set** category

Vector-to-vector



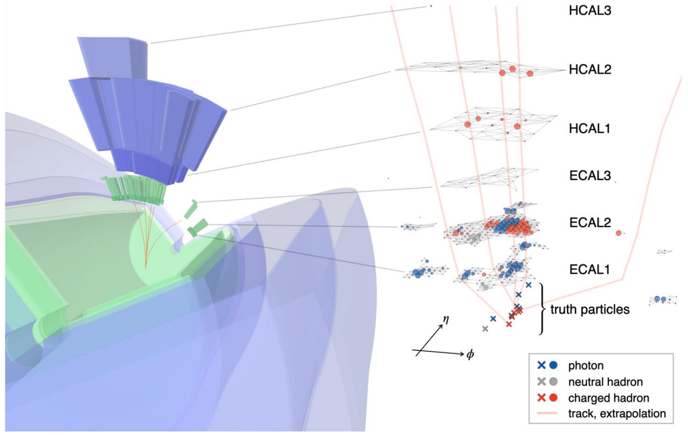
Known and used since many years

Set-to-Vector



Cardinality of the inputs is not fixed but output is

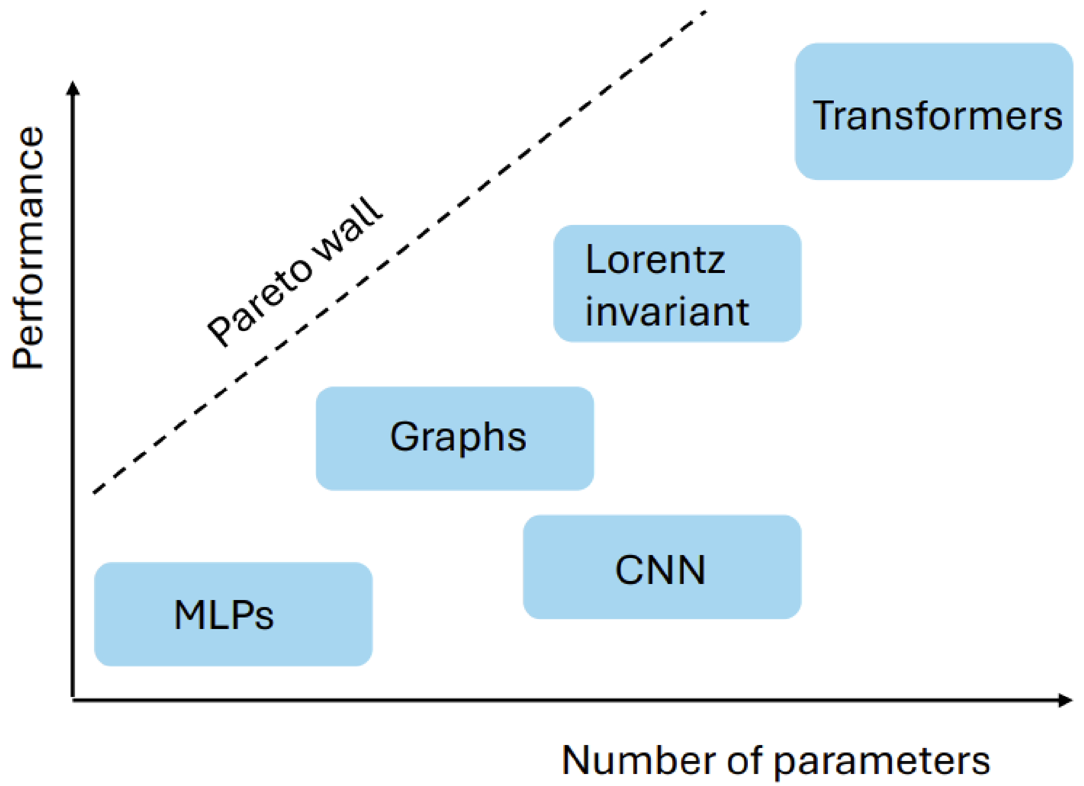
Set-to-Set



Cardinality of both the input and the output are not fixed

On Set-to-vector and symmetries

- From experience, nowadays set-to-vector is mostly just some variation of pure transformers.
- Embeding symmetries into the model (e.g. Lorentz) still needs some work



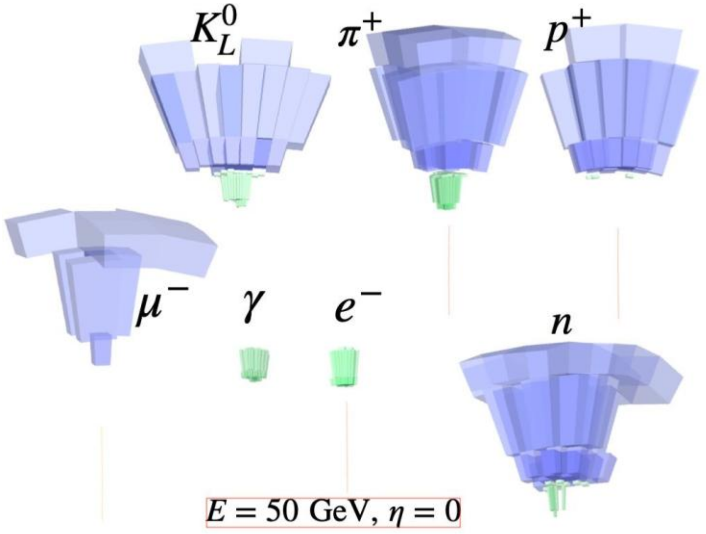
Set-to-Set algorithms: R&D using particle flow as an example

Particle-Flow reconstruction

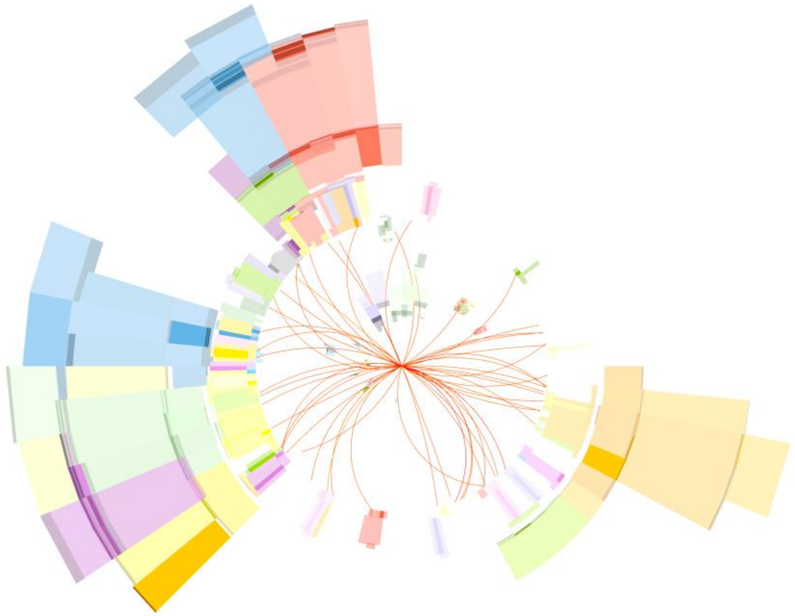
Particle flow can be phrased as follows:

Given the set of detector outputs, reconstuct the set of particles (and their properties) that generated them. This is in inverse problem.

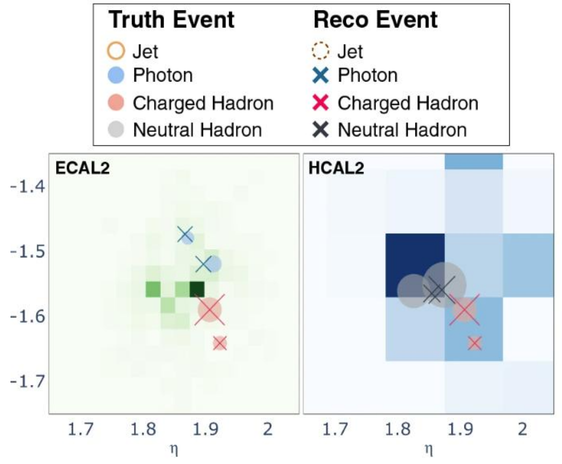
Particle signatures



Detector output



Solving the inverse problem



Particle flow generalities

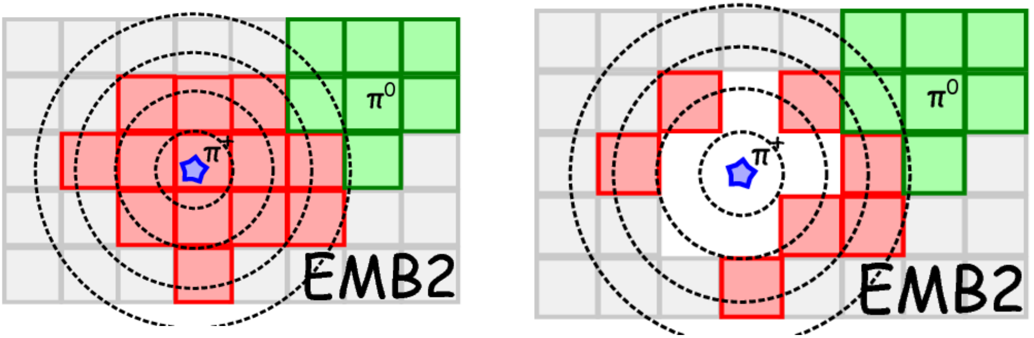
- Particle flow algorithms can optimally account for detector redudancies (e.g. calorimeter vs tracker)

- Typical implementation of Pflow algorithms are based on linearity: parametrize single particle shower

1. Can Machine Learning help Particle Flow reconstruction?

A non exhaustive list: PRD111.09(2025) Comm. Physics 124 (2024) EPJC596 (2023)

Eur. Phys. J. C 77 (2017) 466

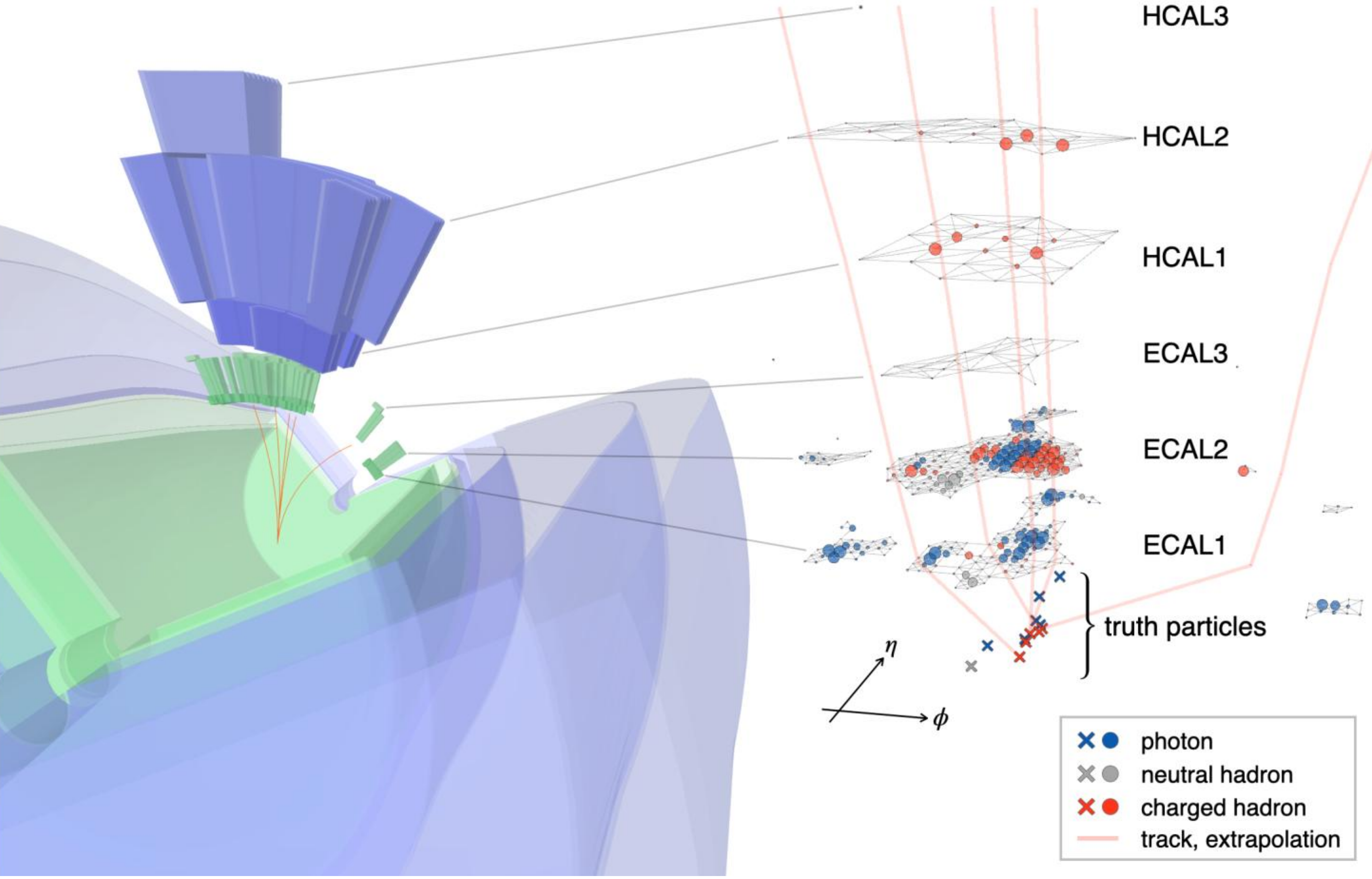


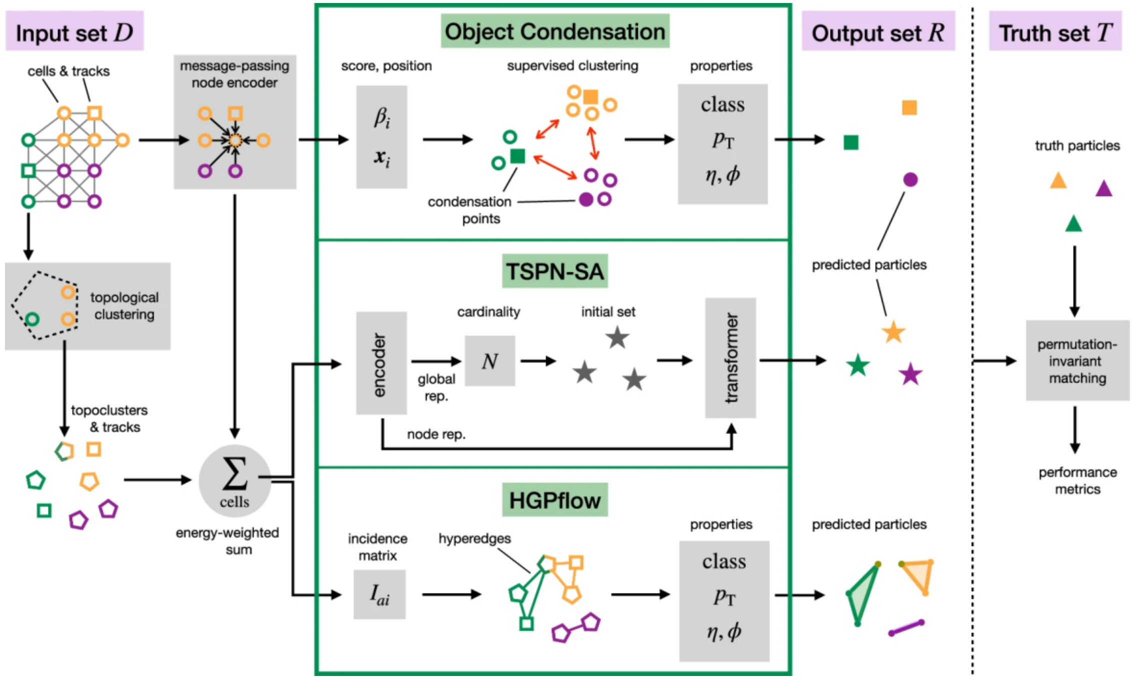
$$\frac{\sigma_p}{p} \approx a\,p$$

$$\frac{\sigma_E}{E} \approx \frac{a}{\sqrt{E}}$$

Constructing the inputs

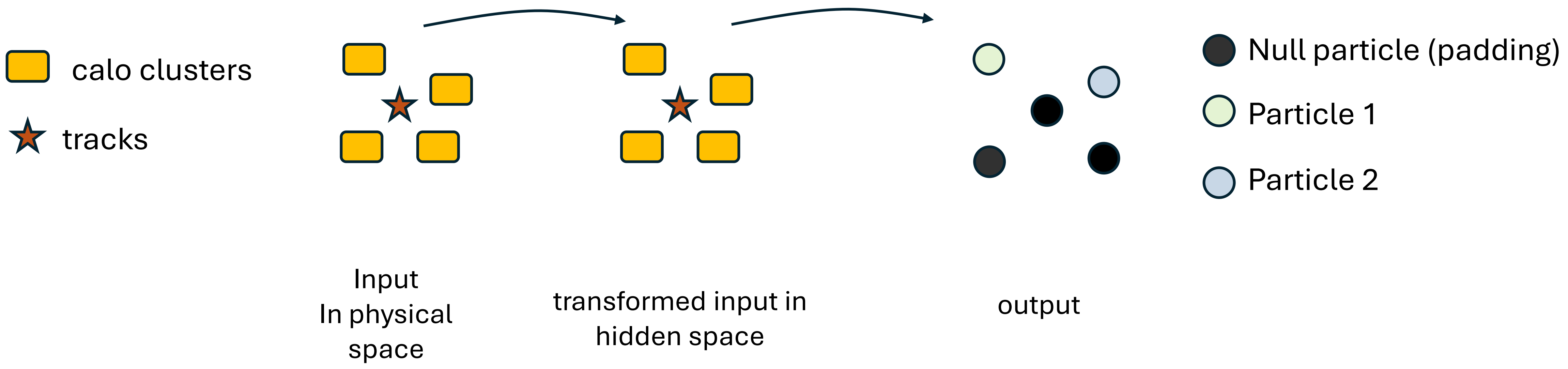
- A graph is build from the detector outputs





Machine learning for particle reconstruction

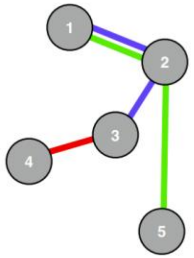
- Some ideas in the market: MLPF EPJC(2021)
- Take as an example a simplified scenarios, two particles, 4 calo-cluster, one track
- Need to define an ordering rule to associate which target particle belongs to a given input object



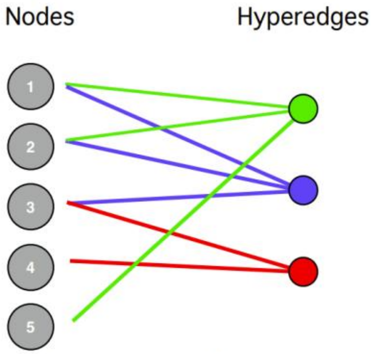
N.B: the definition of target particle is non trivial, requires some semi-fiducial definition

Hypergraps for particle reconstruction

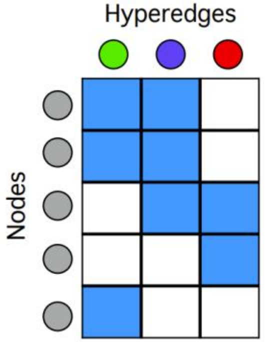
- Can we include inductive bias from the simulation in an explicit form?



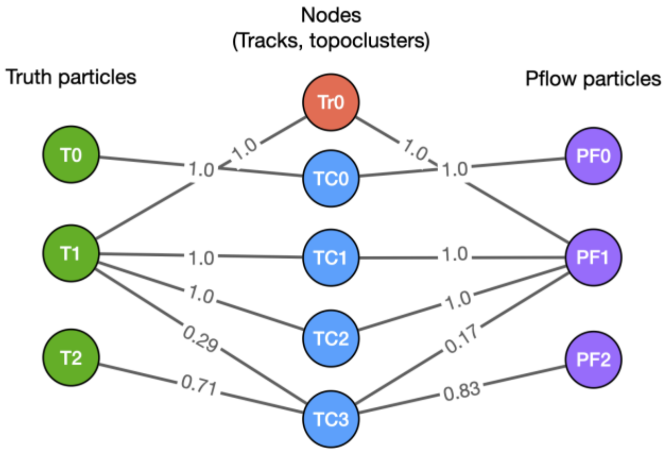
Hypergraph



Bipartite graph

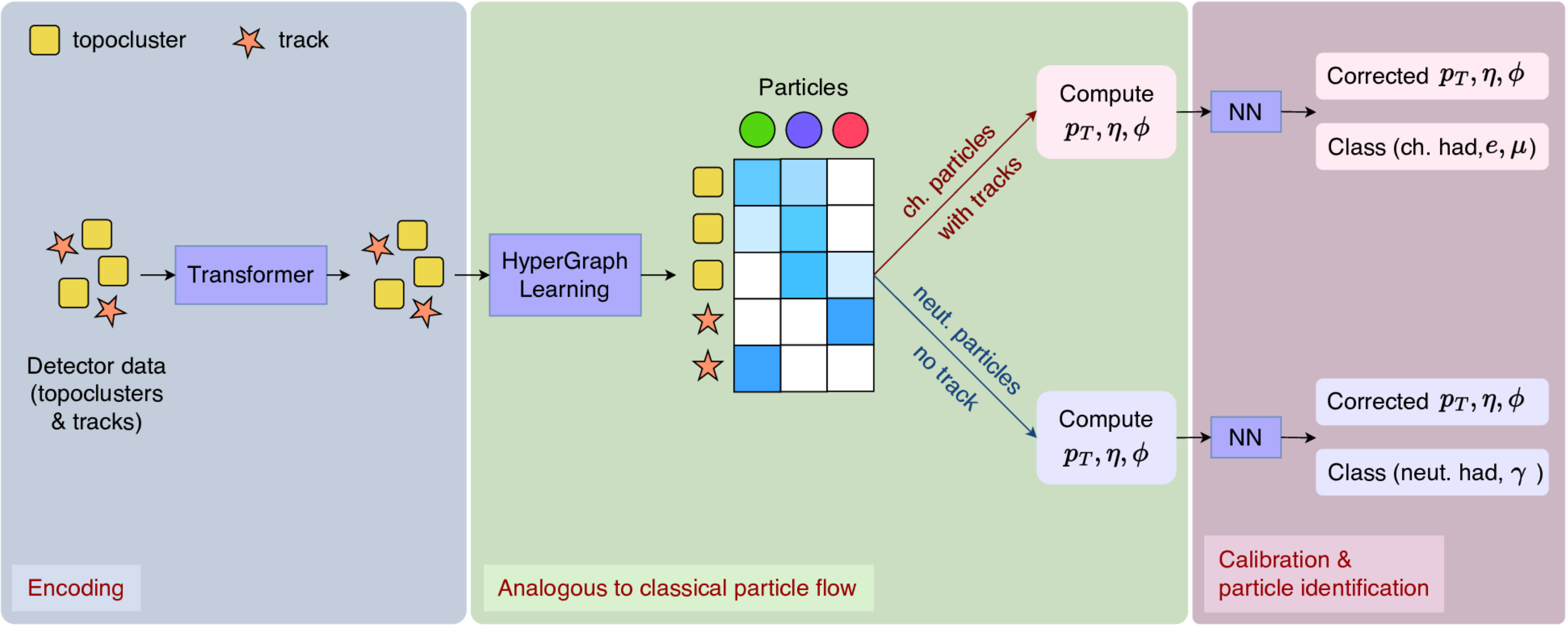


Incidence matrix



Learnable!

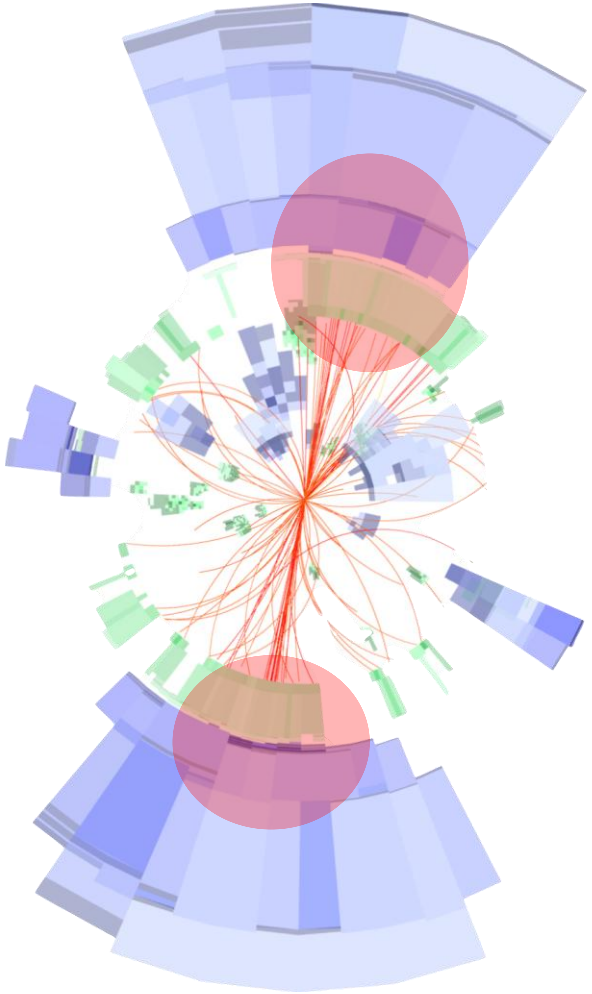
Hypergraphs

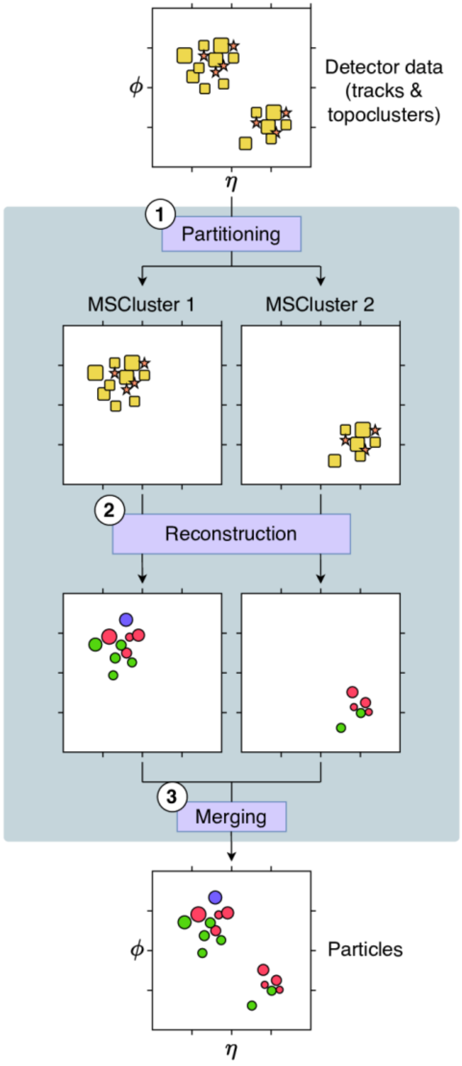


There are also other solutions to Set-to-Set problems, this is an active area of research.
Another open question is on the notion of locality, i.e. R(H) must only be local in order to generalize well.

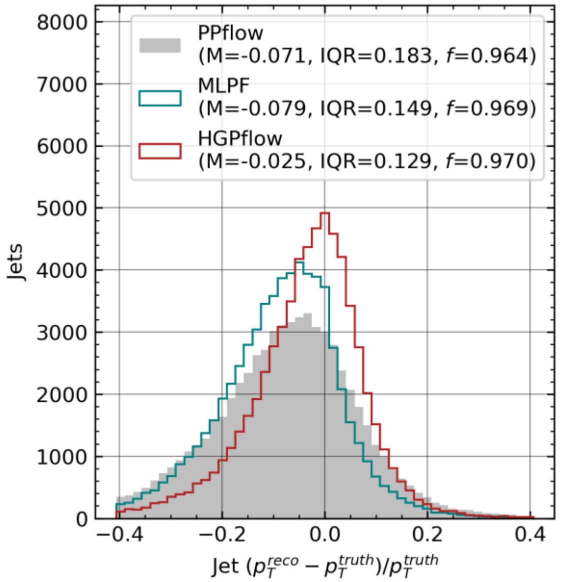
FDB et al EPJC

From single jet to full event reconstruction



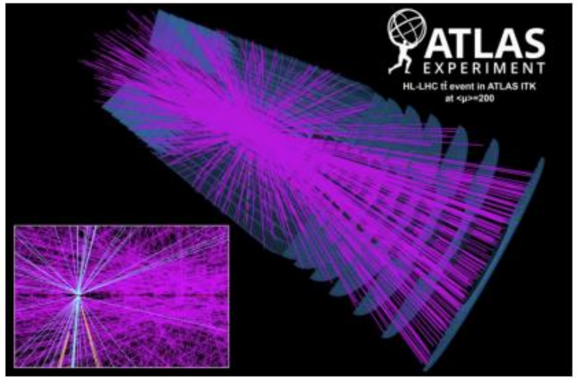


Momentum resolution

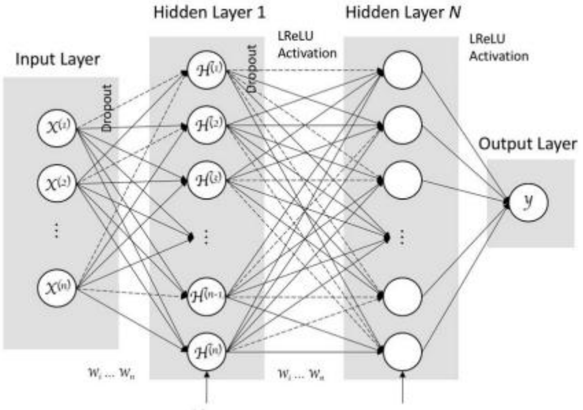


EPJC Dreyer, FDB, Heinrich, Kakati et al

End-to-end algorithms?



0101011101010101
0011010101010011
1111000101010101
1010100100001111
1000101010101010
1010101001010100
1100000011111100
1010101010010101



Is this a Higgs boson?

This is not a great idea if you ask me...

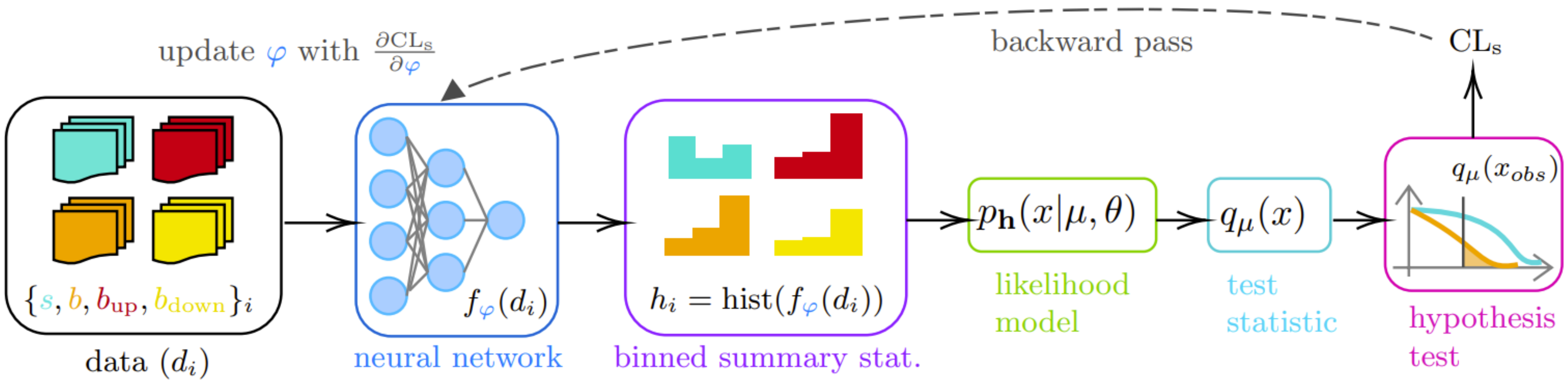
1. Dimensionality: need to probe from a space of around O(10^7) dimensions.
2. Data inherently have structures (locality) that are essential reduce the dimension of the problem as well as for generaliztion.
3. There is a problem of domain shift: we need a way to control mis-modelling of the simulations we use to train our algorithms.

But...

End-to-end pipelines

- Can exploit the hierarchical structure in some detail to combine algorithms into an end-to-end pipeline.
- Examples include analysis dominated by systematic uncertanties; specialized recontruction (high energy)
- NNPDF is a great example of differentiable pipelines

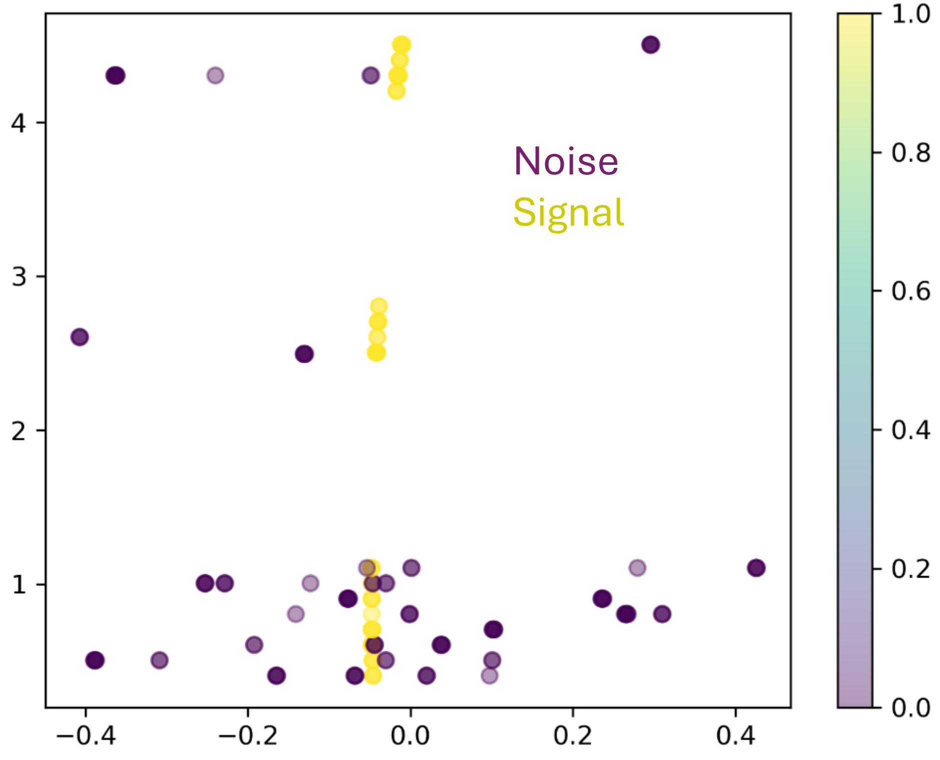
Example of a unified differentiable pipeline



2203.05570 Henirich et al.

Differentiable programming to build end-to-end pipelines

- Differentiable programming is a way to stack together different reconstruction pipeline in a single minimization
- To make it more clear we use a simple example: tracking of a single particle in a magnetic field. Goal: estimate the particle momentum



L. Rambelli

Traditional pipeline

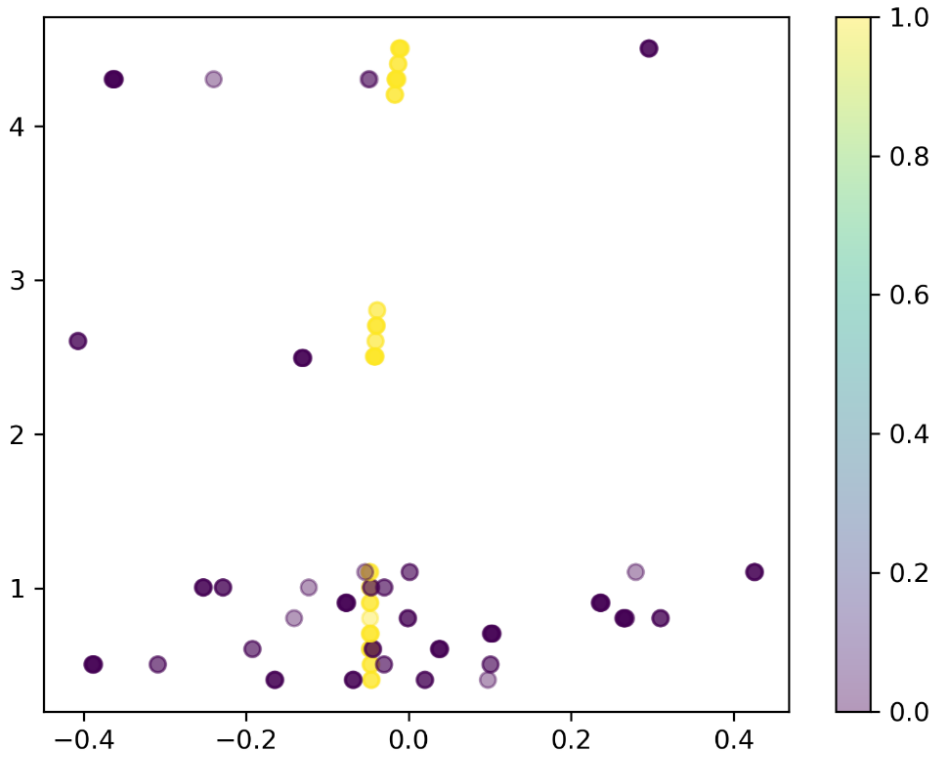
$$\mathcal{L}_\omega = \frac{1}{N}\sum_{h=1}^{N} \mathrm{BCE}(\hat{l}_h(\omega),\, l_h)$$

pattern reco → tracking → Particle properties $\hat{p}$

$$\hat{p} = \arg\min_p \chi^2_{\hat{\omega}}(p)$$

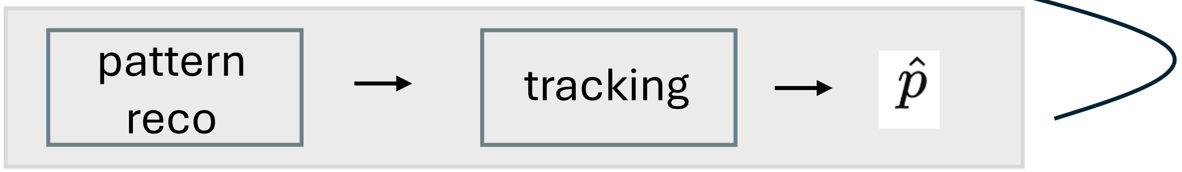
Differentiable programming to build end-to-end pipelines

- Can build an end-to-end architecture provided that pattern reco and tracking are both differentiable
- No need to learn that a particle in a magnetic field moves as a helix. Can embed this into the model.



Differentiable pipeline

$$\hat{\omega} = \arg\min_{\omega} \mathcal{L}(\omega) = \arg\min_{\omega} \left(\hat{p}(\omega) - p_{\text{true}}\right)^2$$



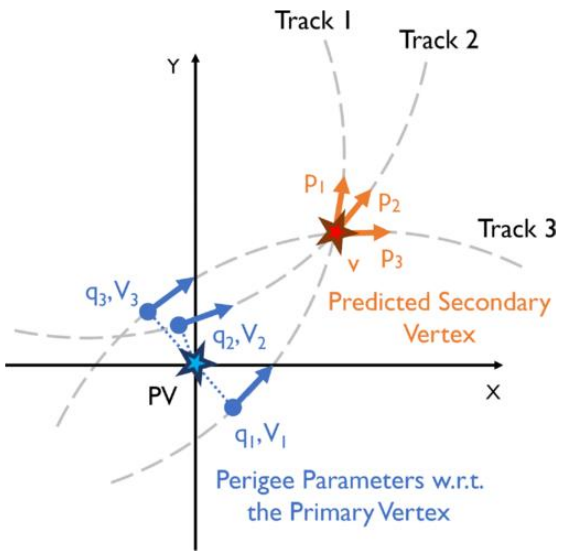
Single end-to-end model

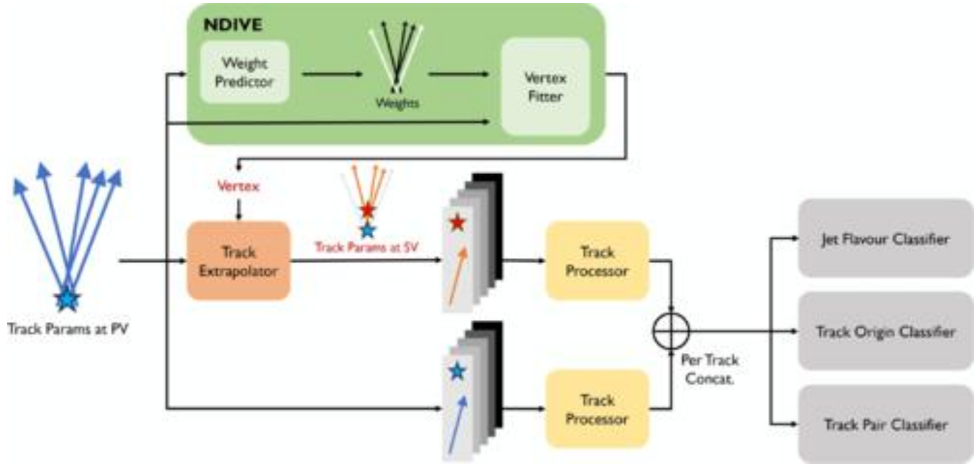
$$\chi^2(p;\omega) \xrightarrow{\arg\min_p} \hat{p}(\omega)$$

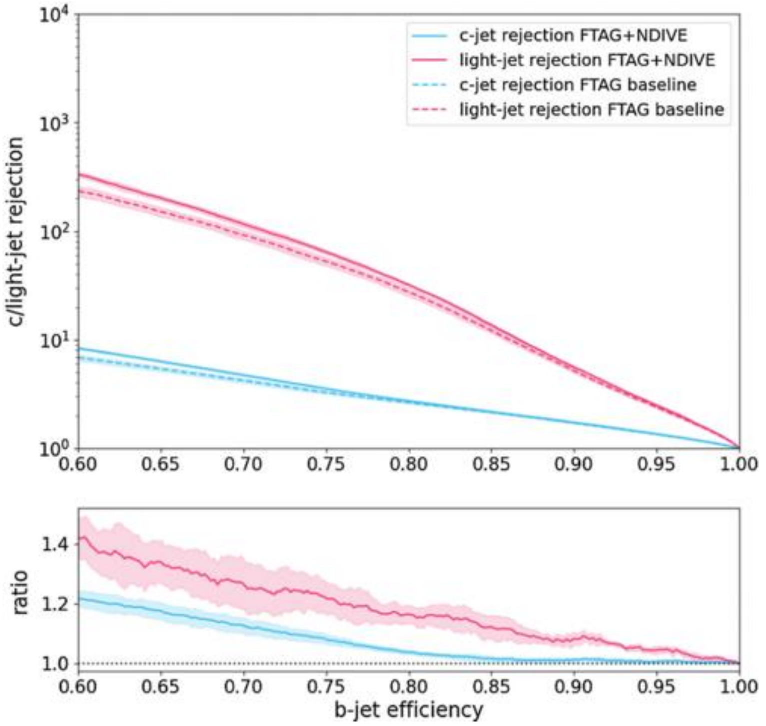
Need to solve a ‘complex’ nested minimization problem
To compute backprop.

Differentiable programming to build end-to-end pipelines

- This is a more complete example of including differentiable vertexing into the training of the network







This kind of algorithm can be re-iterated for any hierarchical pipeline and shows improvement and major interpretability

kagan et al 2310.12804

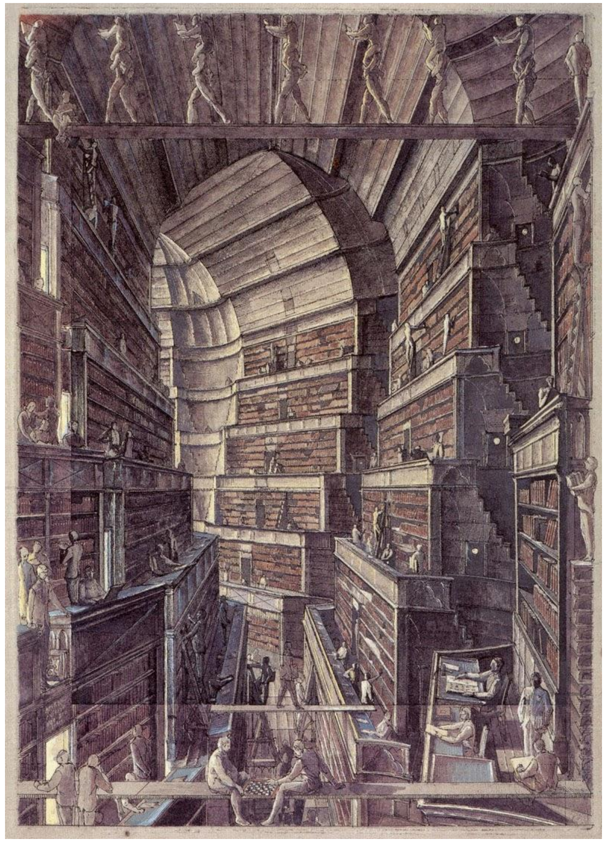
Conclusion

- Machine learning is already today helping us uncovering physics.
  Measurements of the Higgs self-coupling and Charm Yukawa among major examples

- Set-to-set problems will be tackled in the next years, relevant for a variety of tasks: tracking, particle-flow, unfolding and more

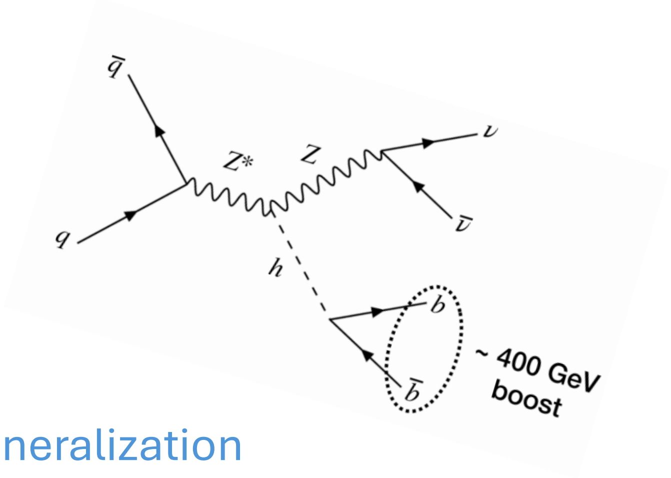
- I think many topics are still in their infancy but have the potential to bring large gains. Some (personal) selection of these algorithms shown today

For any discussions: fdibello@cern.ch

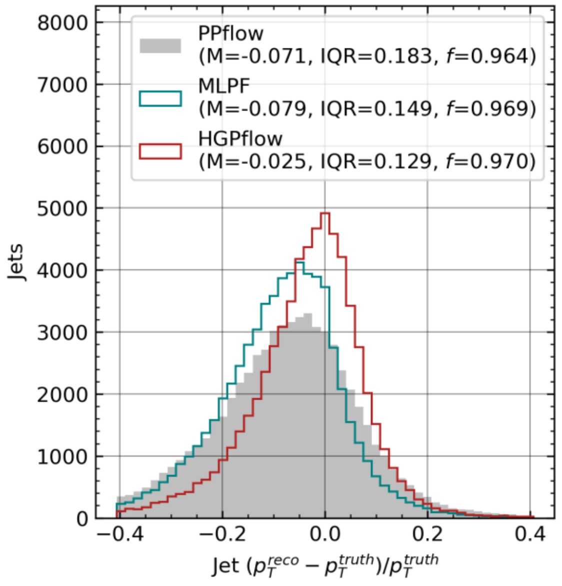


Results

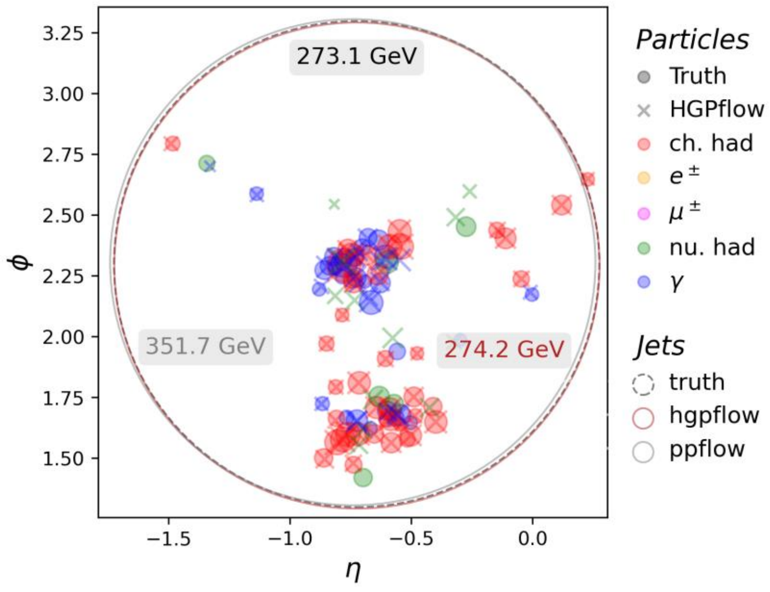
- Several topologies have been tested



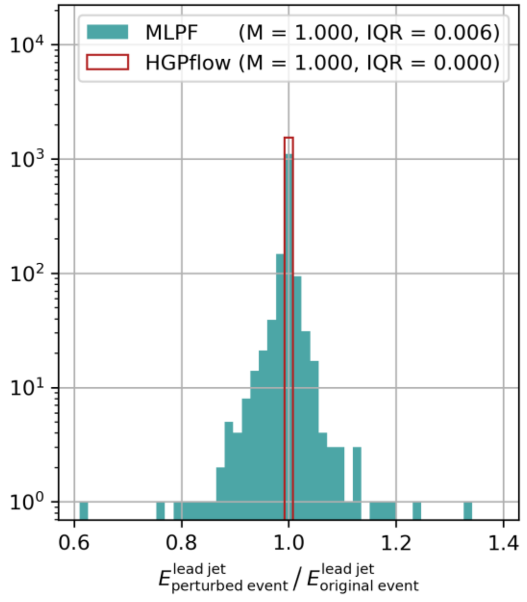
Momentum resolution



Generalization



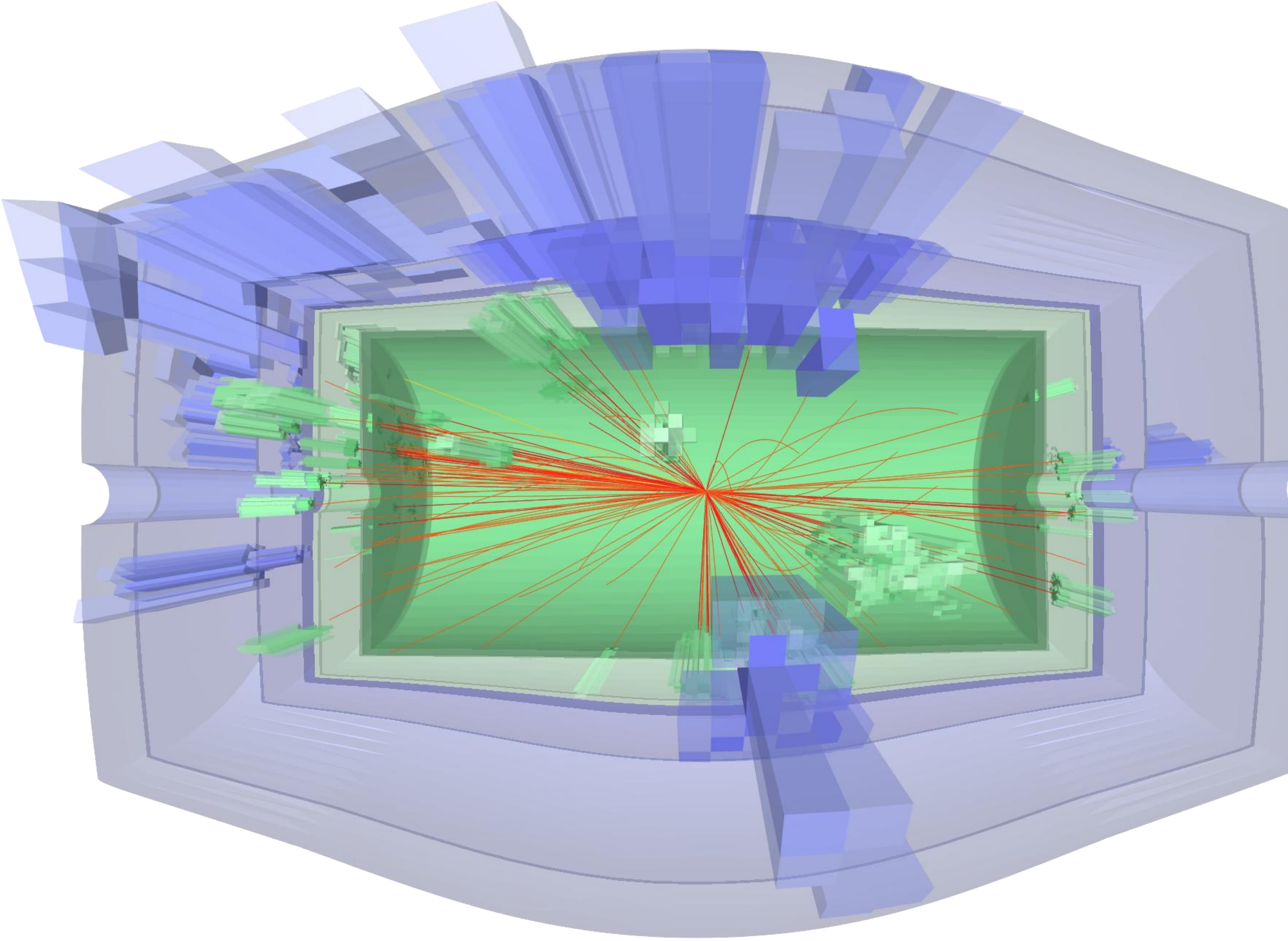
Locality



The simulation: COCOA

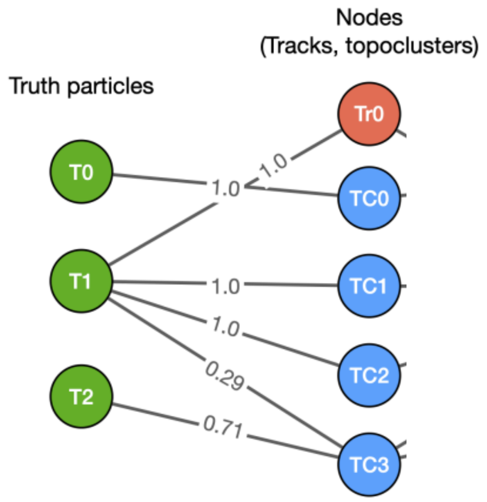
Table 1 Samples used for training and performance testing

Detector	Process	Statistics		
		Train	Val	Test
COCOA	$p^+p^+ \to q\bar{q}$	250k	10k	35k
	Single π^+	–	–	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

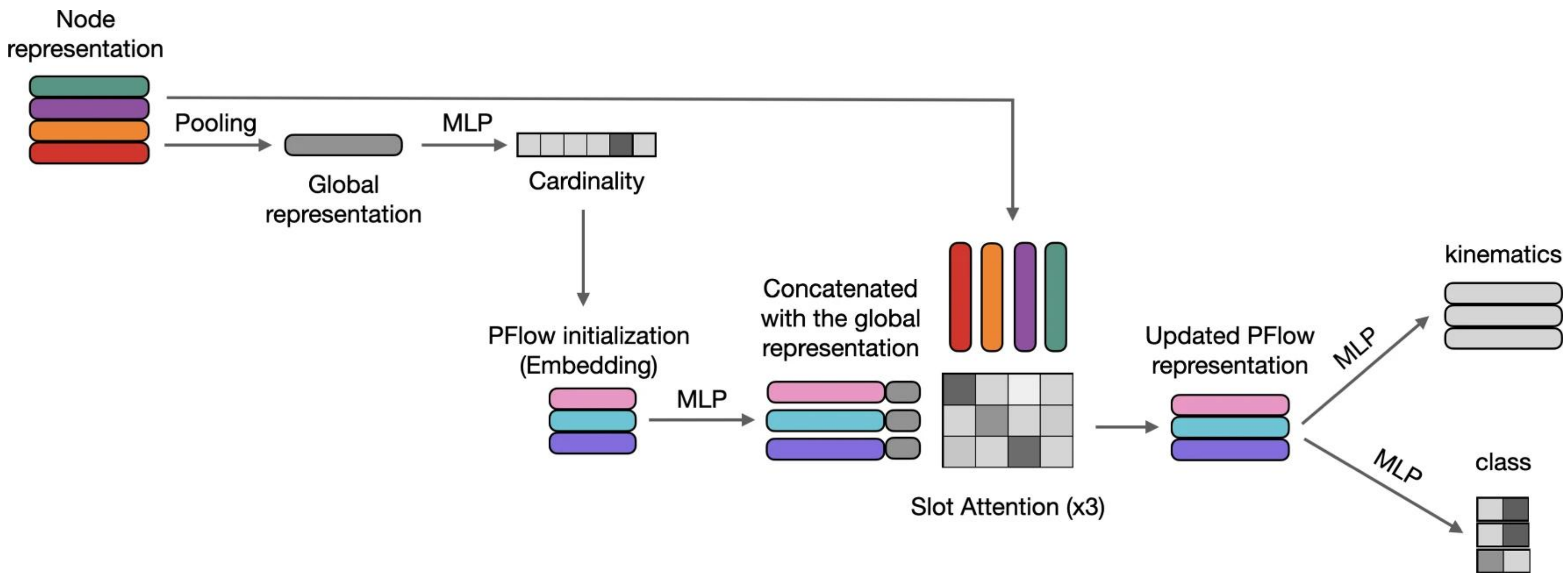


Machine learning for particle reconstruction

- Can we include inductive bias from the simulation in an explicit form?
- Starting point was to build a cross-attention between particles and input objects
- Provided reasonble results but still left some unanswered questions: interpretability and auxialiry information from the simulation



Can we use this knowledge?



Prospects at HL-LHC

- Can we aim to have a 3sigma evidence on kc at the HL-LHC?
- What precision can we expect on kb at the HL-LHC?
- Projections at the HL-LHC from VH analyses with improvements expected by GN2
- We contributed to the extrapolation for the EU strategy update of the c- and b-yukawa: EU Strategy update