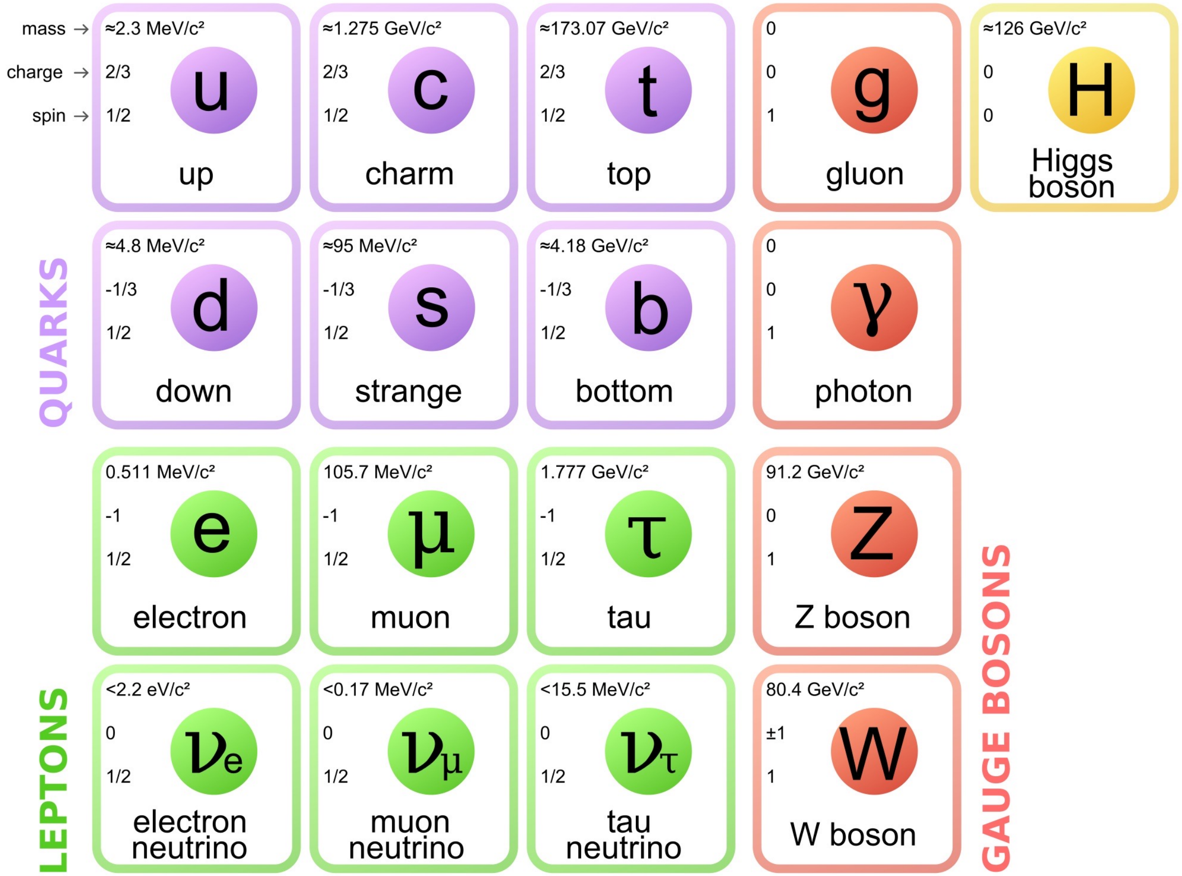
Heterogeneous, Multi-Task Models for Flavour Tagging in ATLAS

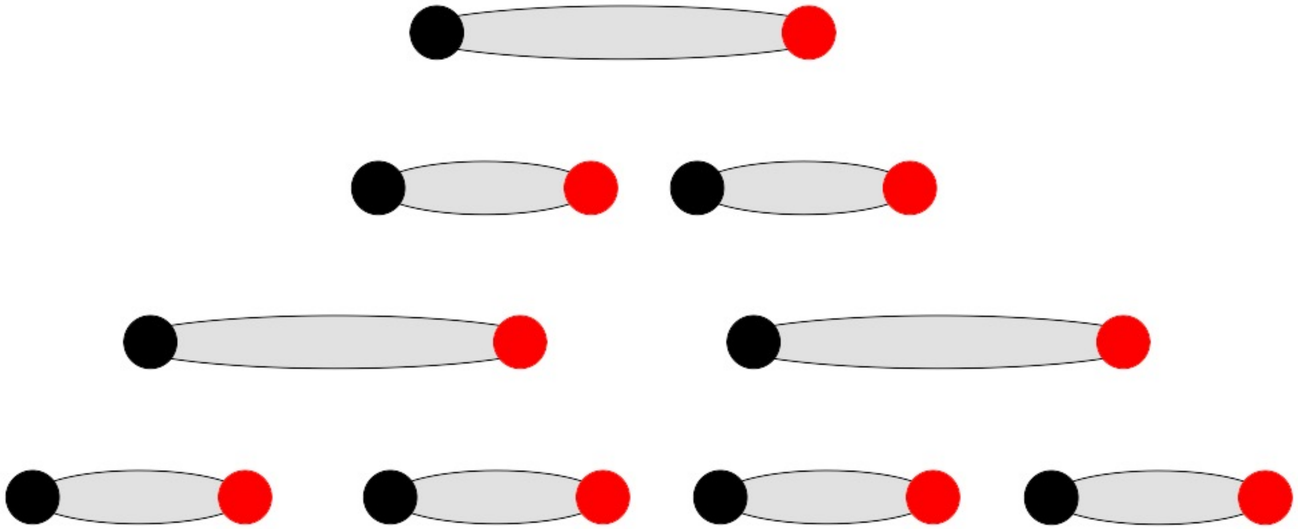
Jackson Barr
2024-10-02

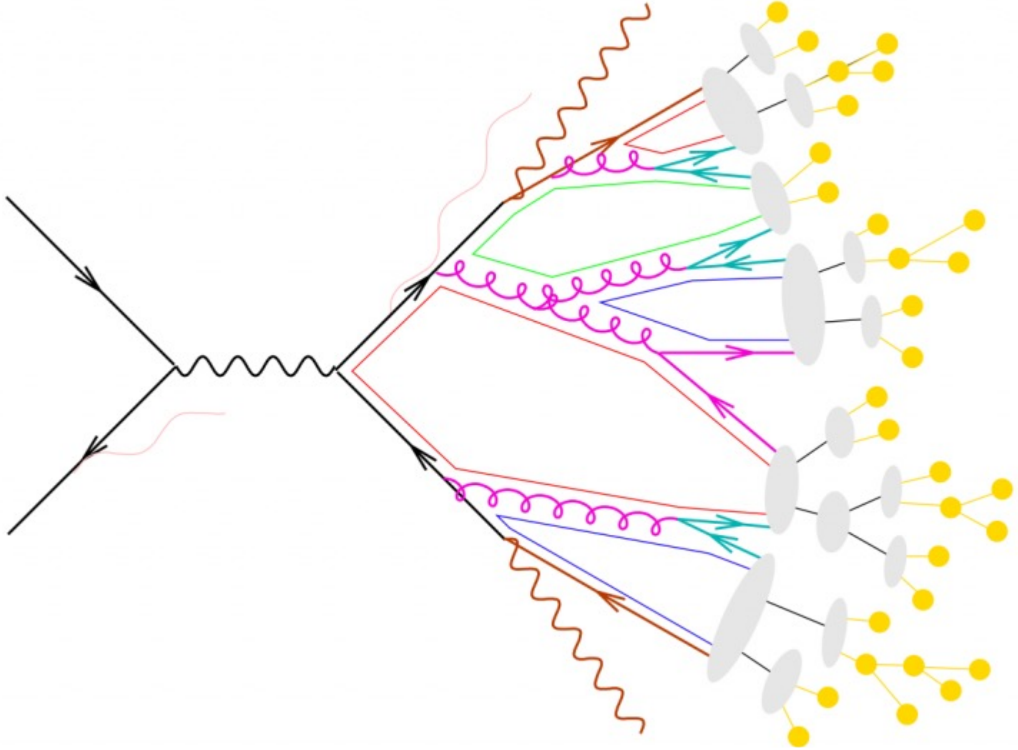
The Standard Model and the LHC



- The Standard Model (SM) of particle physics consists of 4 types of particles:
  - Quarks
  - Leptons
  - Gauge Bosons
  - The Higgs Boson
- At the LHC we collide protons which break apart upon interactions of the constituent quarks
- This results in collimated sprays of particles known as Jets

Hadronisation and Jets





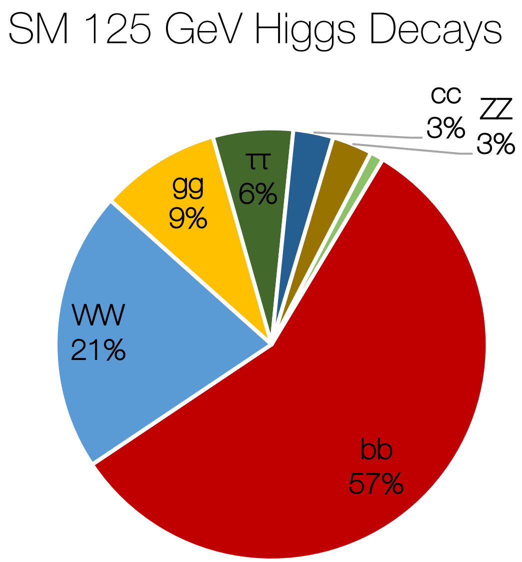
- Quarks experience Colour Confinement – pairs of quarks act springs, it takes energy to pull them apart. Eventually the energy is so great two new pairs form

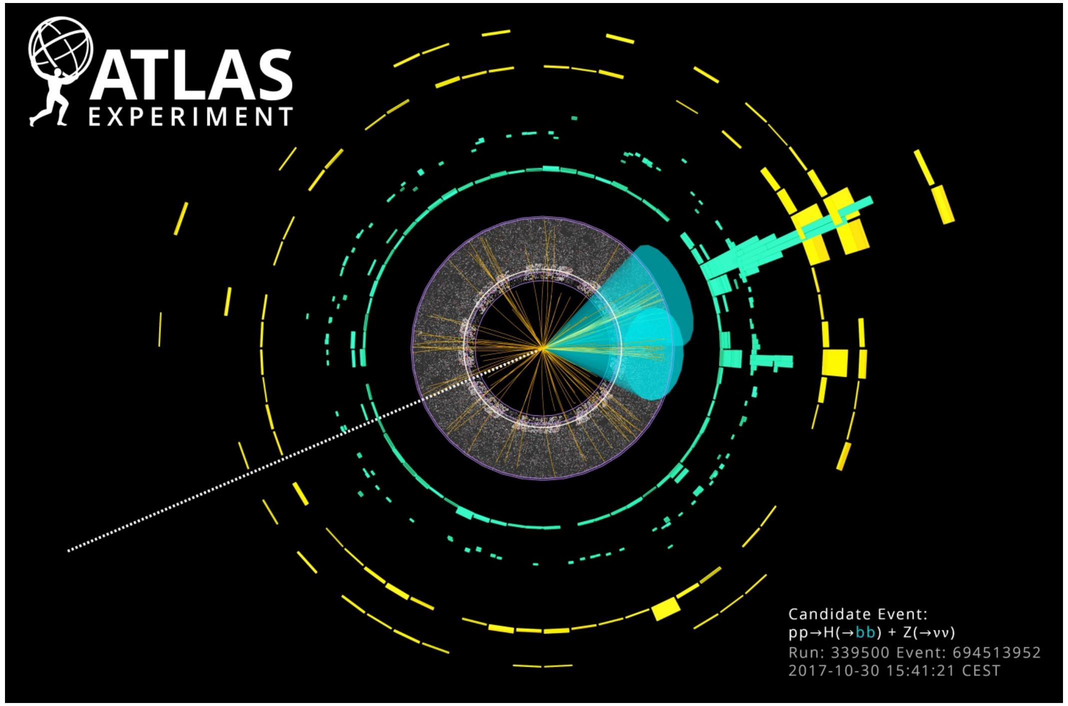
- The resultant hadrons interact slightly differently depending on which type of quark it is made of, the process of identifying this is called Flavour Tagging

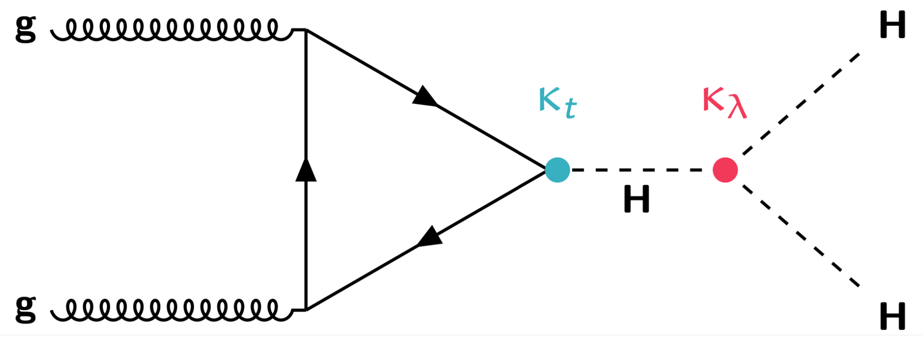
Flavour Tagging at Hadron Colliders

- Hadron colliders produce *a lot* of jets – the vast majority are produced from light flavour quarks

- Being able to identify heavy flavour jets is crucial for a wide range of interesting events e.g. H→bb/cc and t→Wb







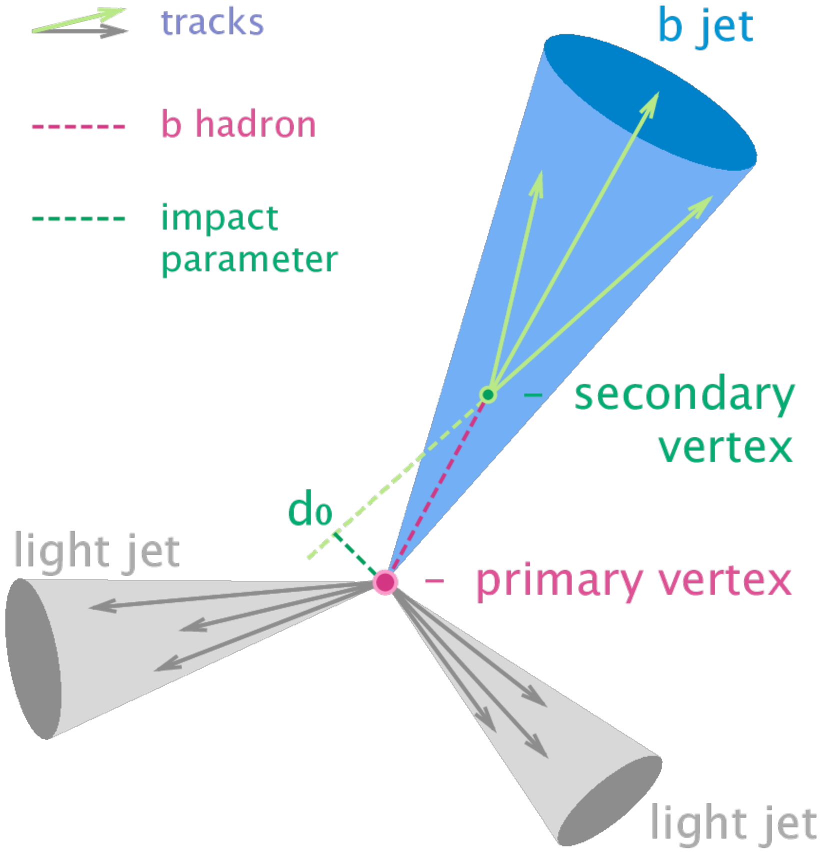
Physics of b-jets

- B-hadrons are relatively long lived and travel a measurable distance before decaying

- This creates multiple distinct vertices where charged particles originate from

- Secondary vertices will have a high mass – b-quark mass is significantly larger than light flavours

- High momentum muons are frequently found within the jet as well



History of ML in Flavour Tagging (Abridged)

MULTIVARIATE ANALYSIS METHODS TO TAG b QUARK EVENTS AT LEP/SLC

B. Brandl+, A. Falvard++, C. Guicheney++, P. Henrard++, J. Jousset++, J. Proriol++

First simple MLP for b-tagging at LEP

1992

b-Tagging and the Search for Neutral Supersymmetric Higgs Bosons at DØ

Tim Scanlon

Imperial College London

2005

First ML based b-tagging at a hadron-collider (Tevatron)

ATLAS NOTE

ATLAS-CONF-2014-046

July 3, 2014

Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data

ATLAS' first ML b-tagger MV1

2012

Jet Tagging via Particle Clouds

Huilin Qu*

Department of Physics, University of California, Santa Barbara, California 93106, USA

Loukas Gouskos†

CERN, CH-1211 Geneva 23, Switzerland

2019

CMS releases ParticleNet GNN tagger

ATLAS PUB Note

ATL-PHYS-PUB-2023-021

1st August 2023

Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS

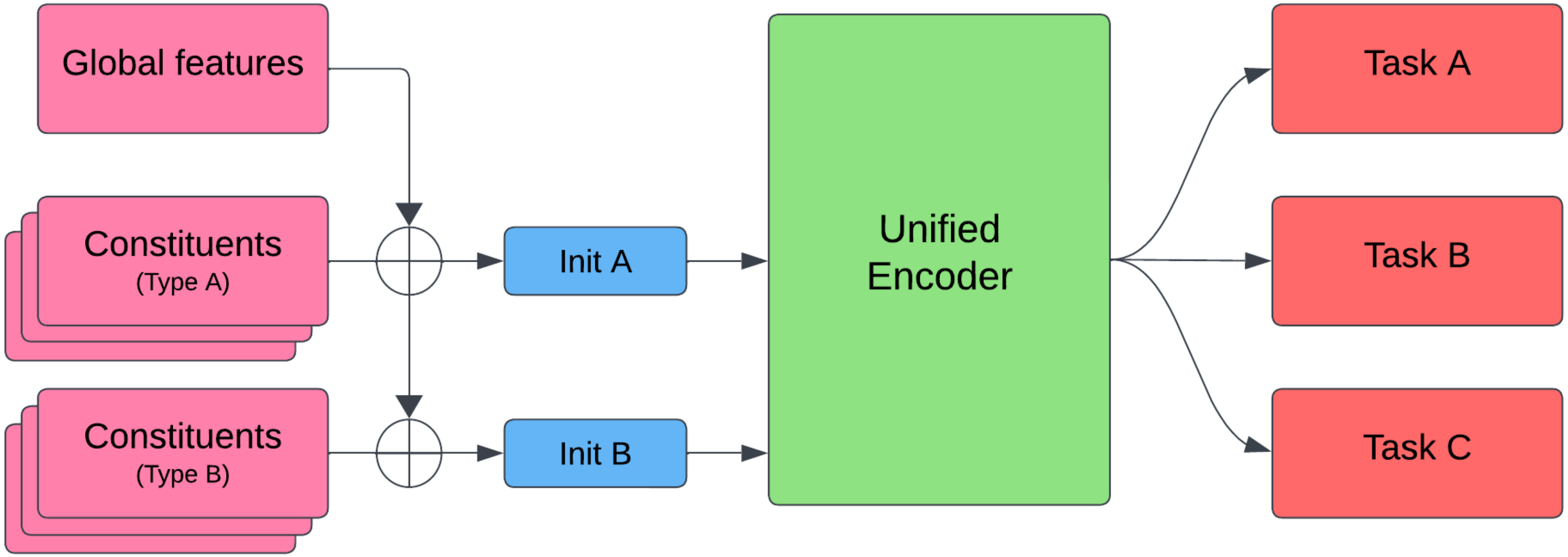
2022

Particle Transformer for Jet Tagging

Huilin Qu[1] Congqiao Li[2] Sitian Qian[2]

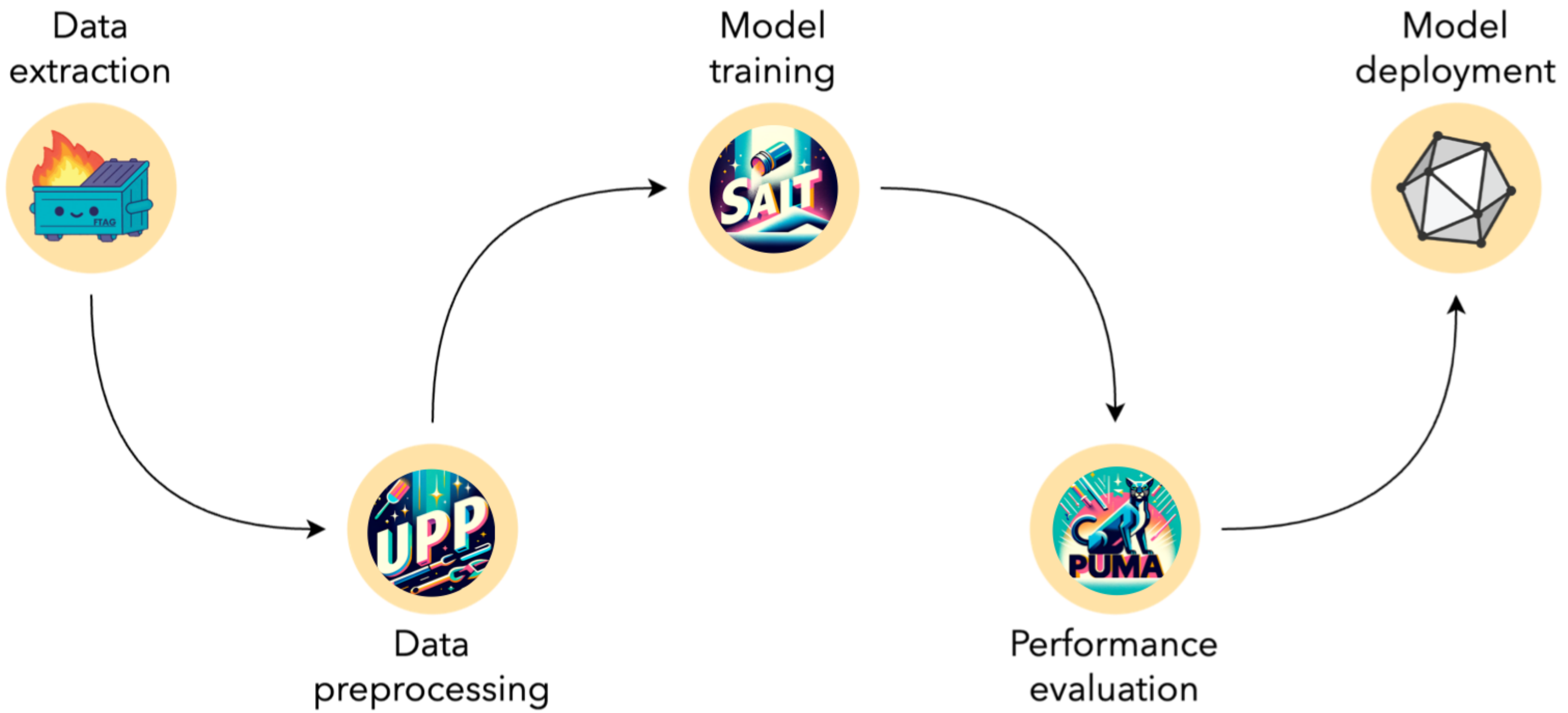
Transformers make their way to flavour tagging

Transformer Based Flavour Tagging



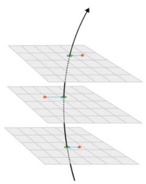
- The current generation of flavour tagging model used in ATLAS, called GN2, is a transformer based model
- Multiple input modalities are used and trained to do multiple different physics tasks

ML Pipeline



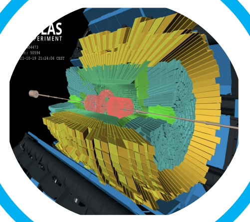
- Current models use in the regime of 100's of millions of training data points – a single training event contains O(100) constituents with O(20) features each. Models have up to O(10 m) parameters currently
- Multiple packages developed for the extraction, processing and training of models using ATLAS data (Salt training framework under review in JOSS

Heterogeneous Inputs



Tracks

- Left by charged particles going through inner detector
- Information on trajectories, hits in layers etc
- Neutral particles **don't** interact



Calorimeter

- Measures energy of particles
- Complements tracker information for charged particles
- Neutral particles **do** interact

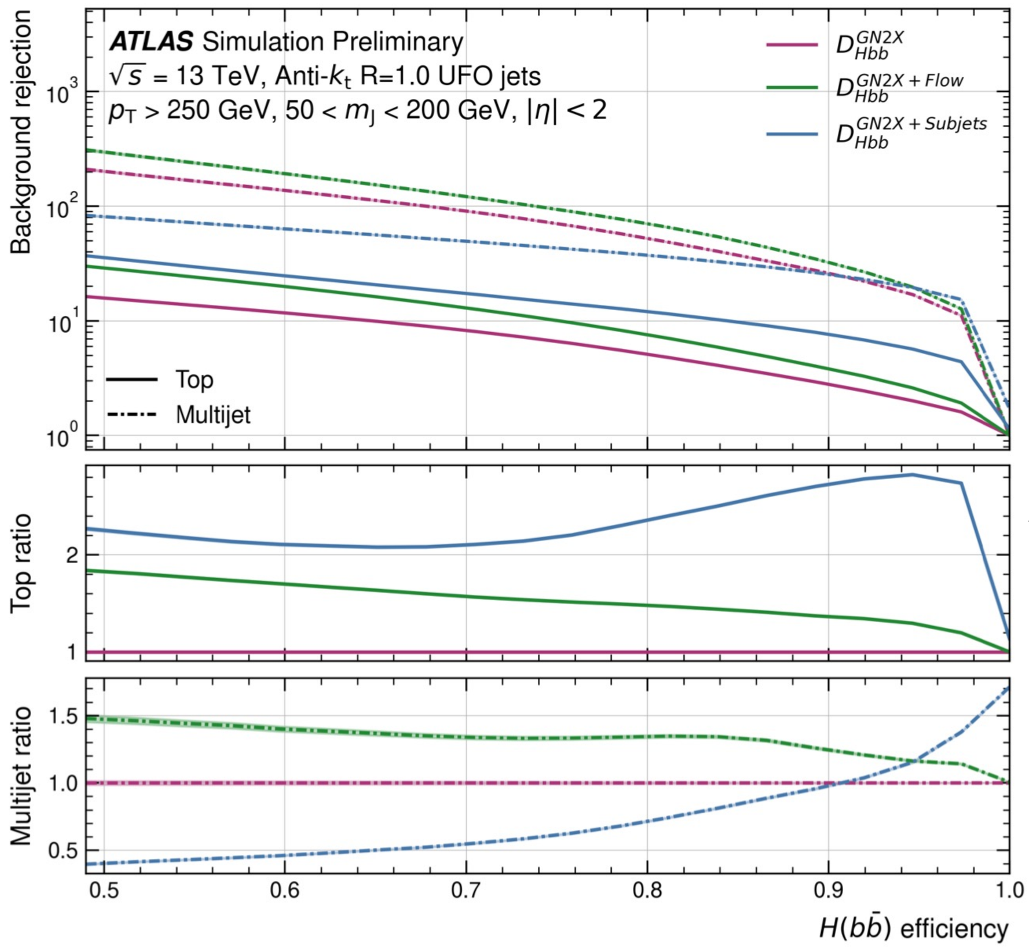
Leptons

- Electrons and muons are built from combining information from other parts of the detector
- Muons leave tracks in a dedicated muon spectrometer

- We have information from multiple different sub-detectors and reconstructed objects that can be used
- Different input types vary in both feature dimensions and multiplicity

Heterogeneous Inputs

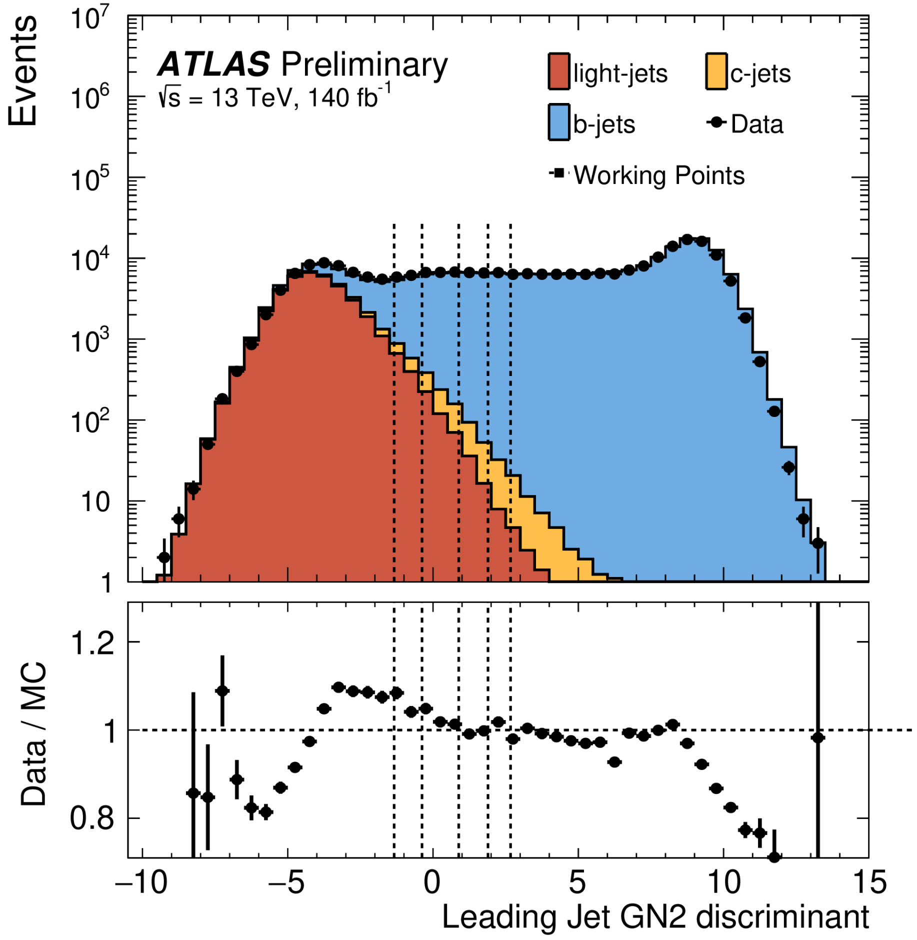
- In general, the more information we include the better the performance we see
- Still many open questions on how best to utilise different modalities – information can vary from having a lot of overlap to being very distinct
- We encode everything together at the same time rather than separate encoders for each input modality, perhaps more difficult to scale



Higher is better

Problems on the Horizon

- We use simulation to train these models but no simulation is perfect
- Expanding the number of modalities and granularity can expose more places where our simulation is lacking
- Potentially resolved with foundation models – unsupervised pre-training can be done on data, no simulation needed!

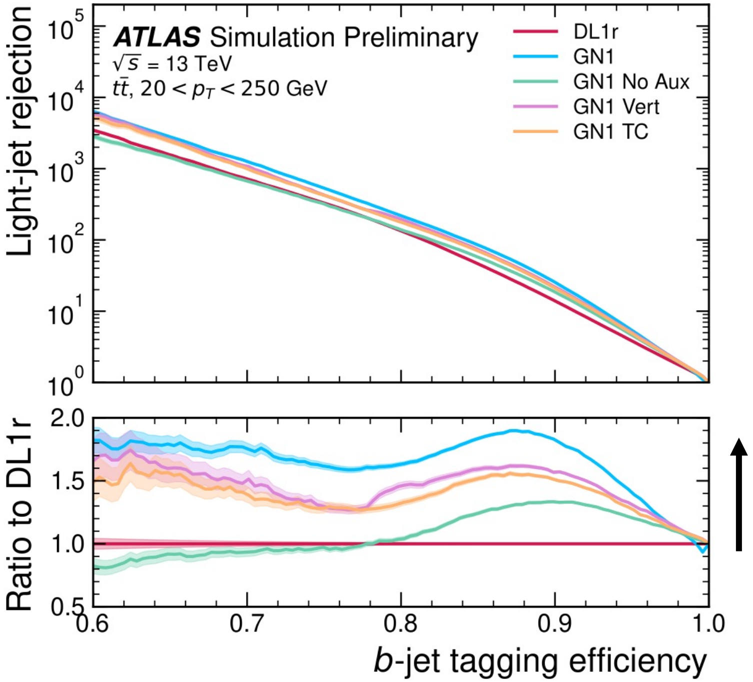


Log likelihood ratio of classifier in simulation and in real data

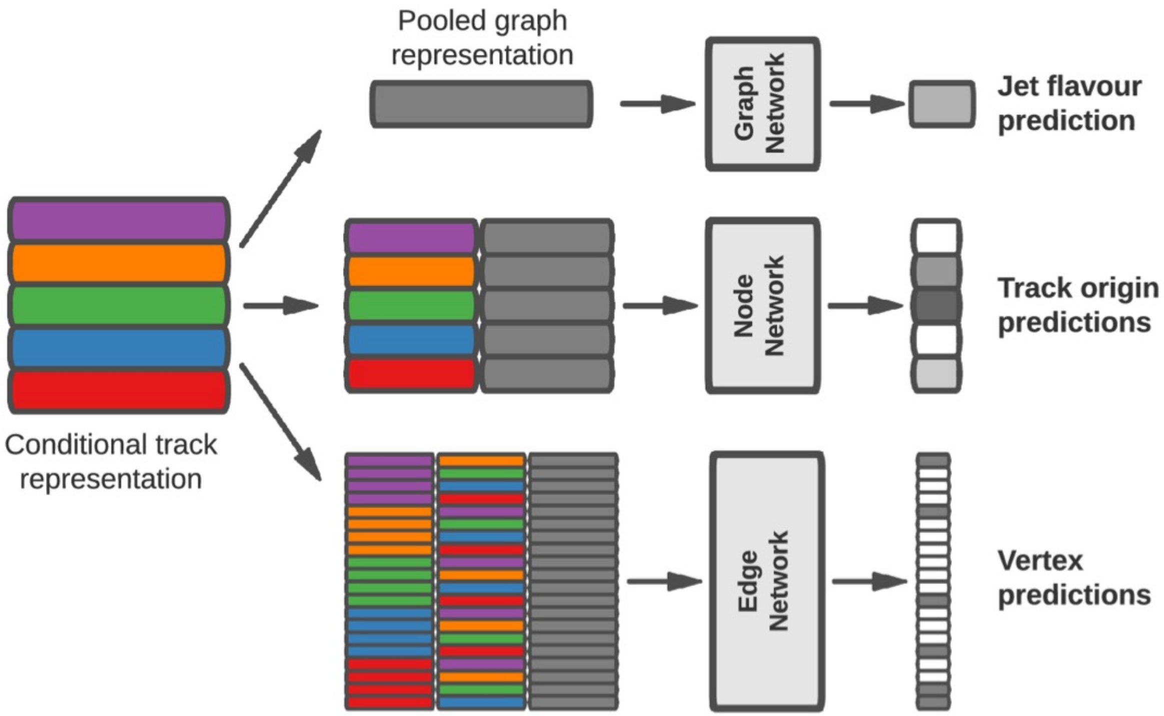
Multiple Tasks

- End goal is to identify the flavour of jets but the additional auxiliary tasks are shown to enhance performance of the main task

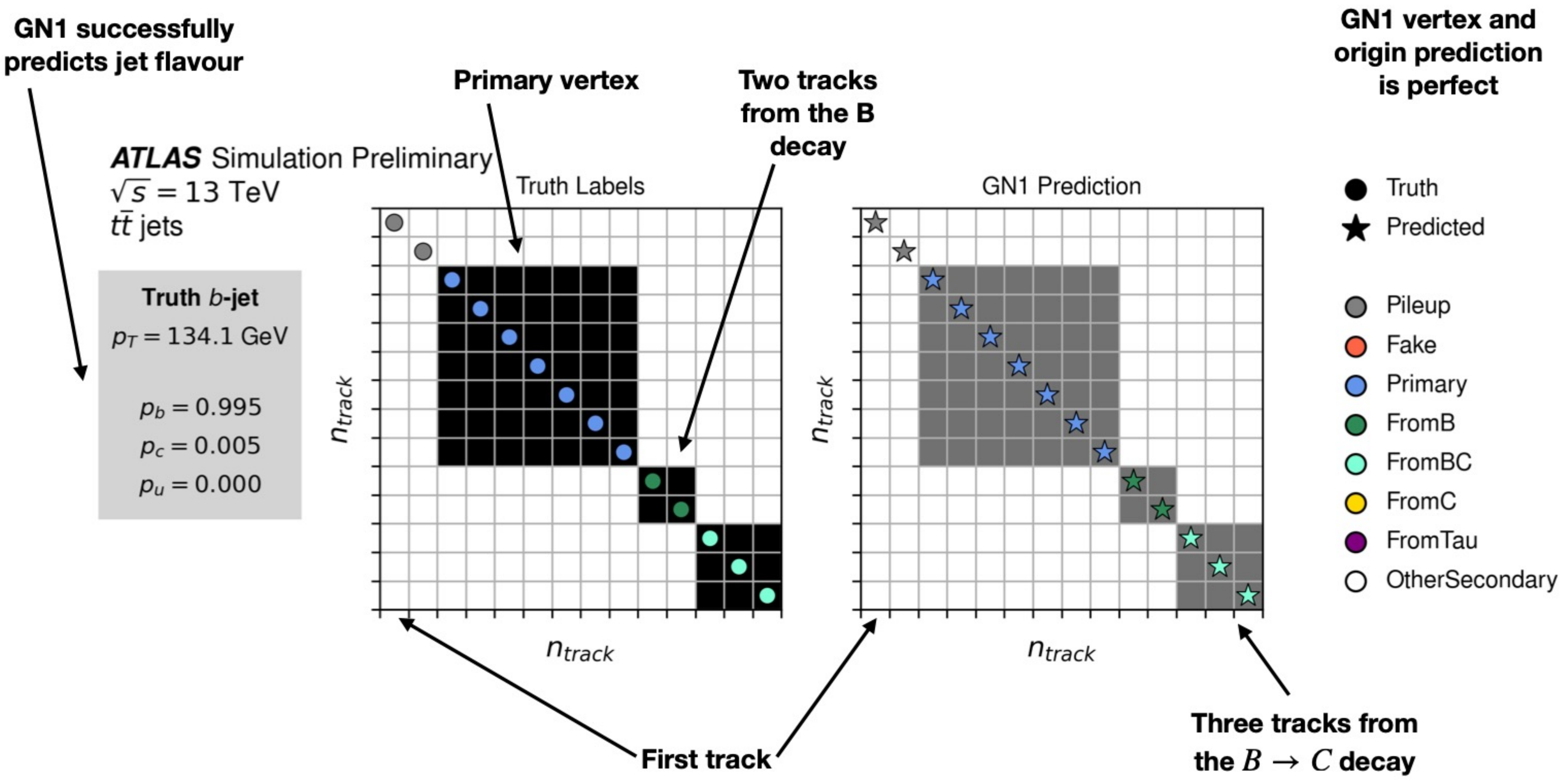
- Auxiliary tasks such as identifying which tracks originate from a common origin or identifying fake tracks act as stepping stones to the main task



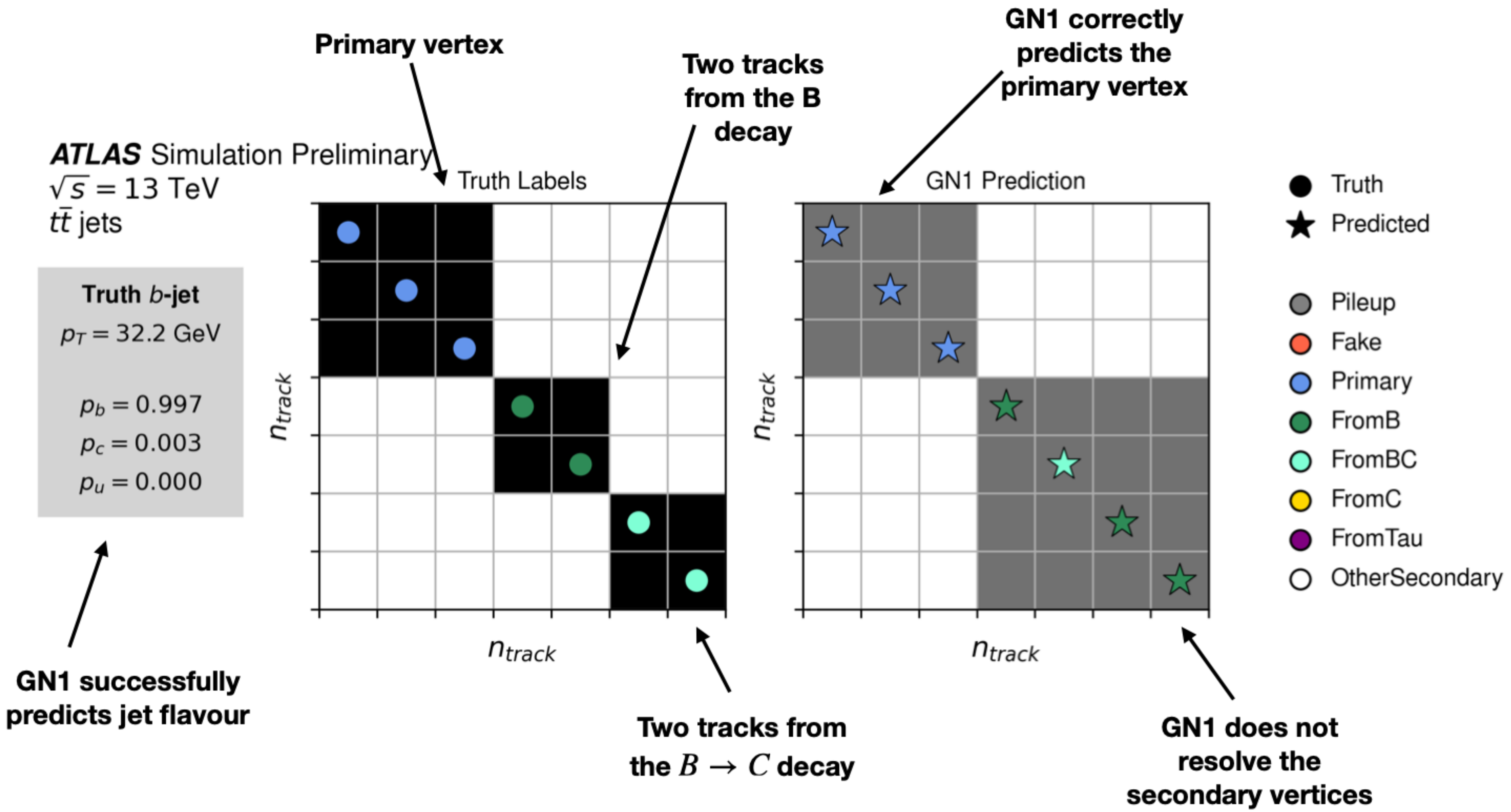
↑ Higher is better



Auxiliary Tasks for Interpretability



Auxiliary Tasks for Interpretability



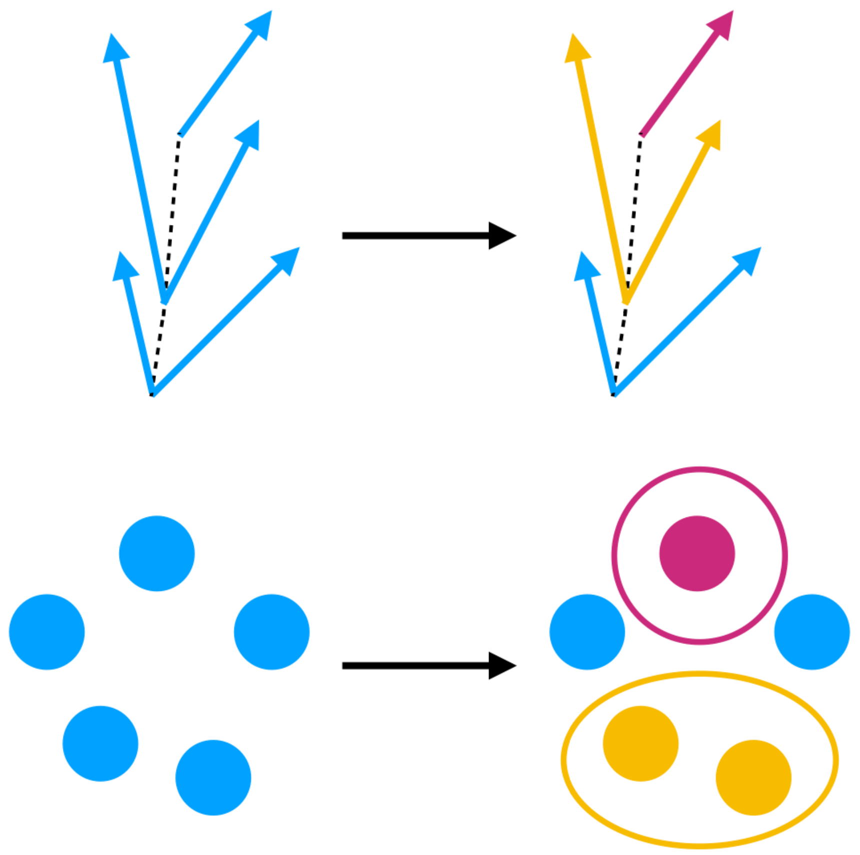
- More studies on interpretability in flavour tagging by Scott in the next talk using a different technique

Object Detection for Vertex Reconstruction

- The additional auxiliary tasks are beneficial to the main task performance – but can we extract even more information?

- e.g. trying to reconstruct the full decay chain of particles

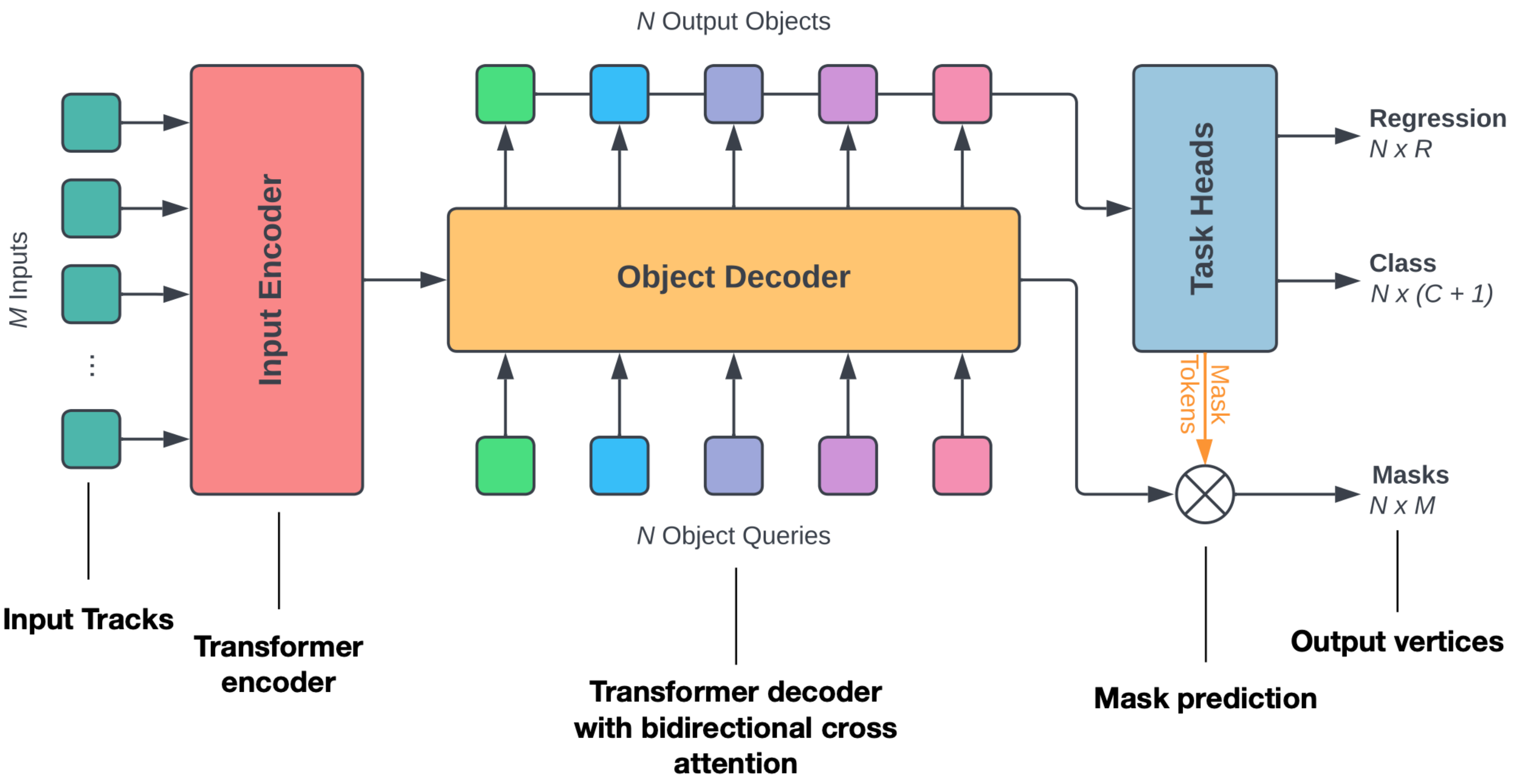
- Idea: use object detection techniques to identify vertices and the properties of the particles they belong to



Segment Anything

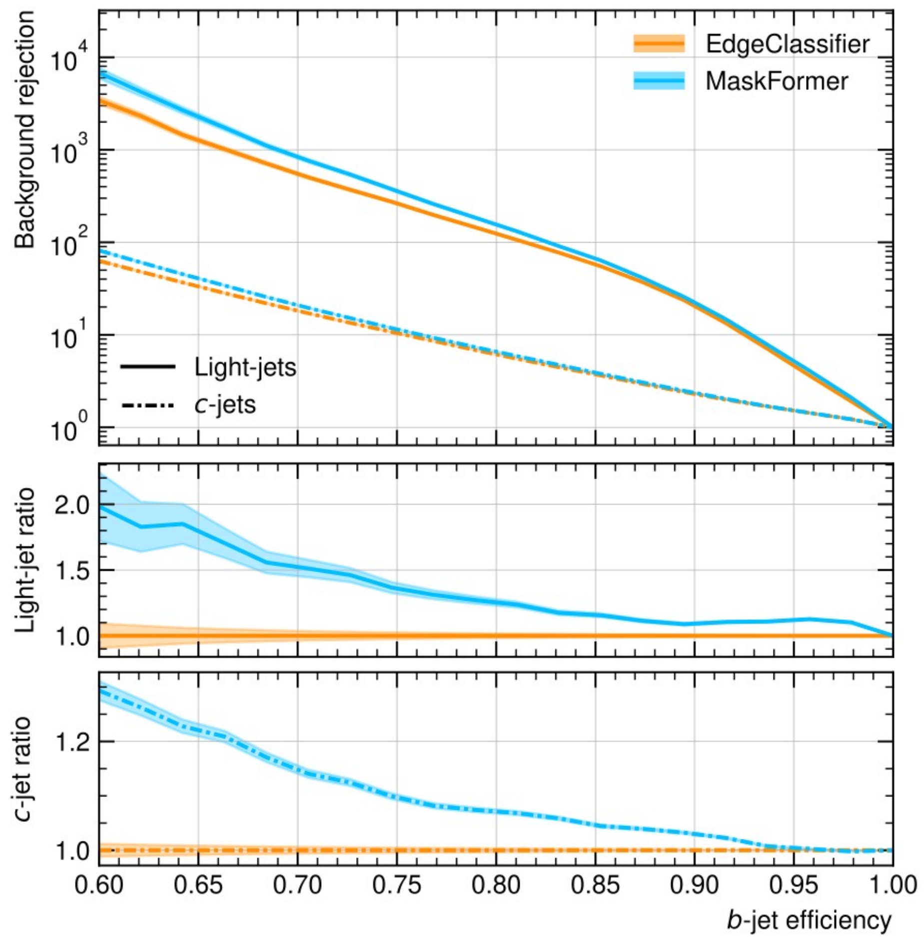
- Segment Anything is the current state of the art for identifying objects in images by learning binary masks over pixels
- Translated to flavour tagging we want to reconstruct vertices in jets by learning binary masks over tracks (and whatever else)

MaskFormer


N Output Objects
M Inputs
Input Encoder
Object Decoder
Task Heads
Regression
N x R
Class
N x (C + 1)
Mask Tokens
Masks
N x M
N Object Queries
Input Tracks
Transformer encoder
Transformer decoder with bidirectional cross attention
Mask prediction
Output vertices

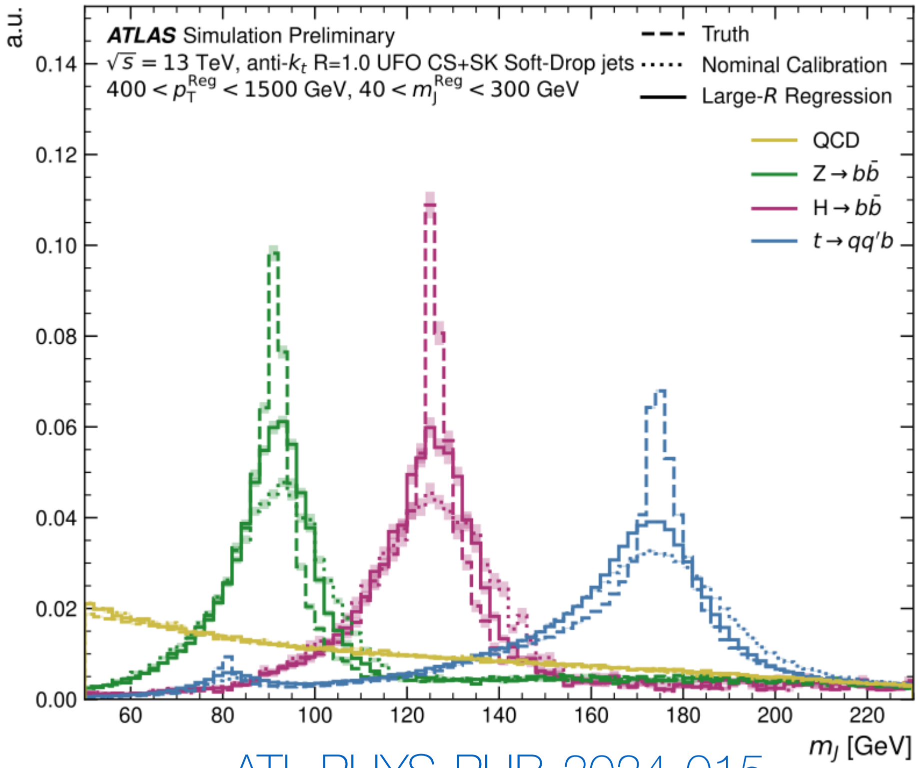
MaskFormer Results

- The MaskFormer style vertexing further improves in the main flavour tagging task performance over the simpler edge classifier task
- Multi-task learning is a good way of including our physics knowledge of what things are related as inductive biases within the model
- But will this still hold as we scale larger or does the bitter lesson continue to be true?

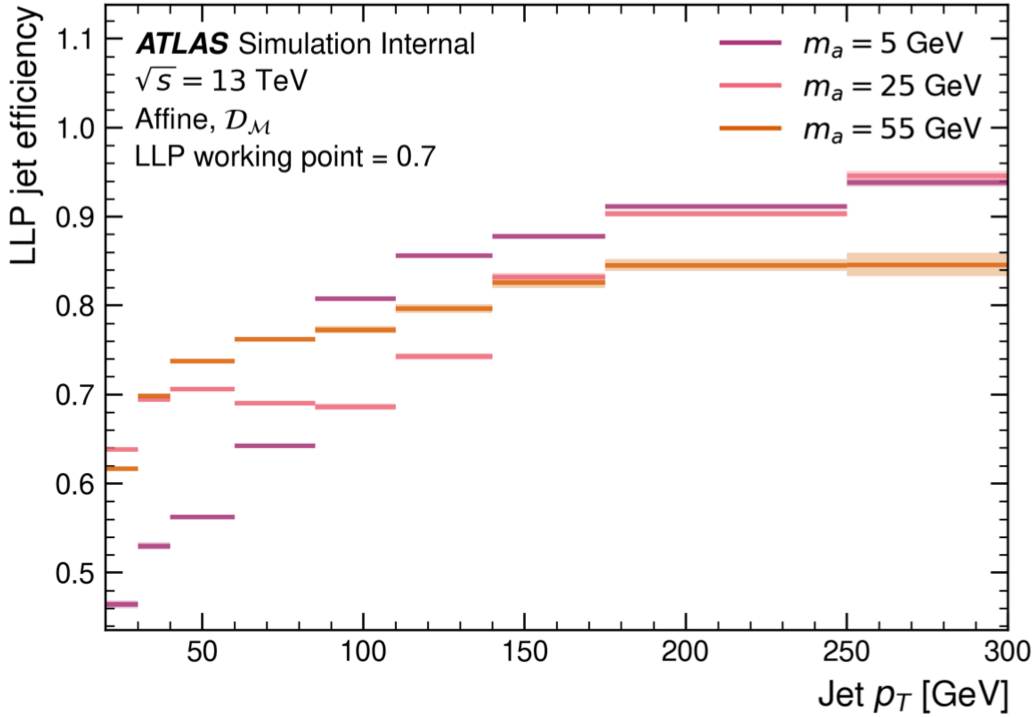


arXiv:2312.12272

Beyond Flavour Tagging



ATL-PHYS-PUB-2024-015



E. Haines Thesis

- The tools and pipelines used are fairly agnostic to the task at hand – it has been adapted to a range of other ATLAS physics results including Jet regression and Long Lived Particle searches
- It's easier to consider combining lots of different things in a foundation model if we're all speaking the same language to begin with

Conclusion

- Particle physics detectors give us a wide range of types of data ranging from low-level information straight from sub-detectors to high-level information reconstructed by different algorithms

- Flavour tagging is an example use case that uses this information and has seen significant improvements in recent years as we try to exploit more and more of this information

- Multiple tasks help generalisation and can improve performance on any individual task – another reason why foundation models are an interesting goal