

**FLAVOUR TAGGING AT
ATLAS**

Jackson Barr

15/10/2025



This talk in a nutshell

- ML is revolutionising HEP - we are replacing old and established methods with large black boxes
- This is happening ~everywhere you look
- I'm going to talk about how this has been seen in flavour tagging at ATLAS
- Flavour tagging within ATLAS has seen a rapid transformation in the past few years, far exceeding what we thought was possible



The Guardian

News Opinion Sport Culture Lifestyle

World UK Climate crisis Ukraine Environment Science

Large Hadron Collider
AI to revolutionise fundamental physics and 'could show how universe will end'

Exclusive: Cern's next director general Mark Thomson says AI is paving the way for huge advances in particle physics

Hannah Devlin in Geneva

Mon 3 Feb 2025 06.00 CET

Share

Advanced artificial intelligence is to revolutionise fundamental physics and could open a window on to the fate of the universe, according to Cern's next director general.

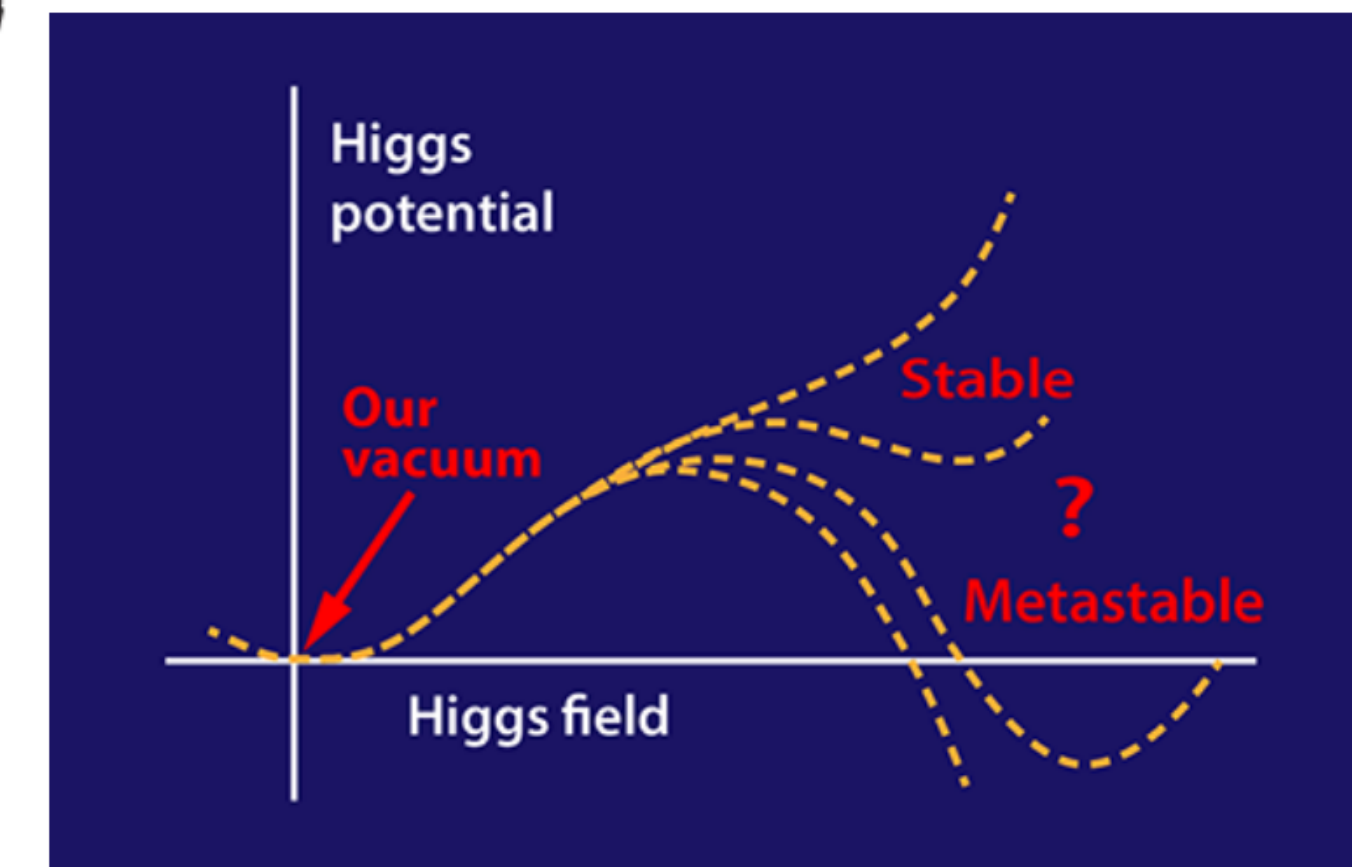
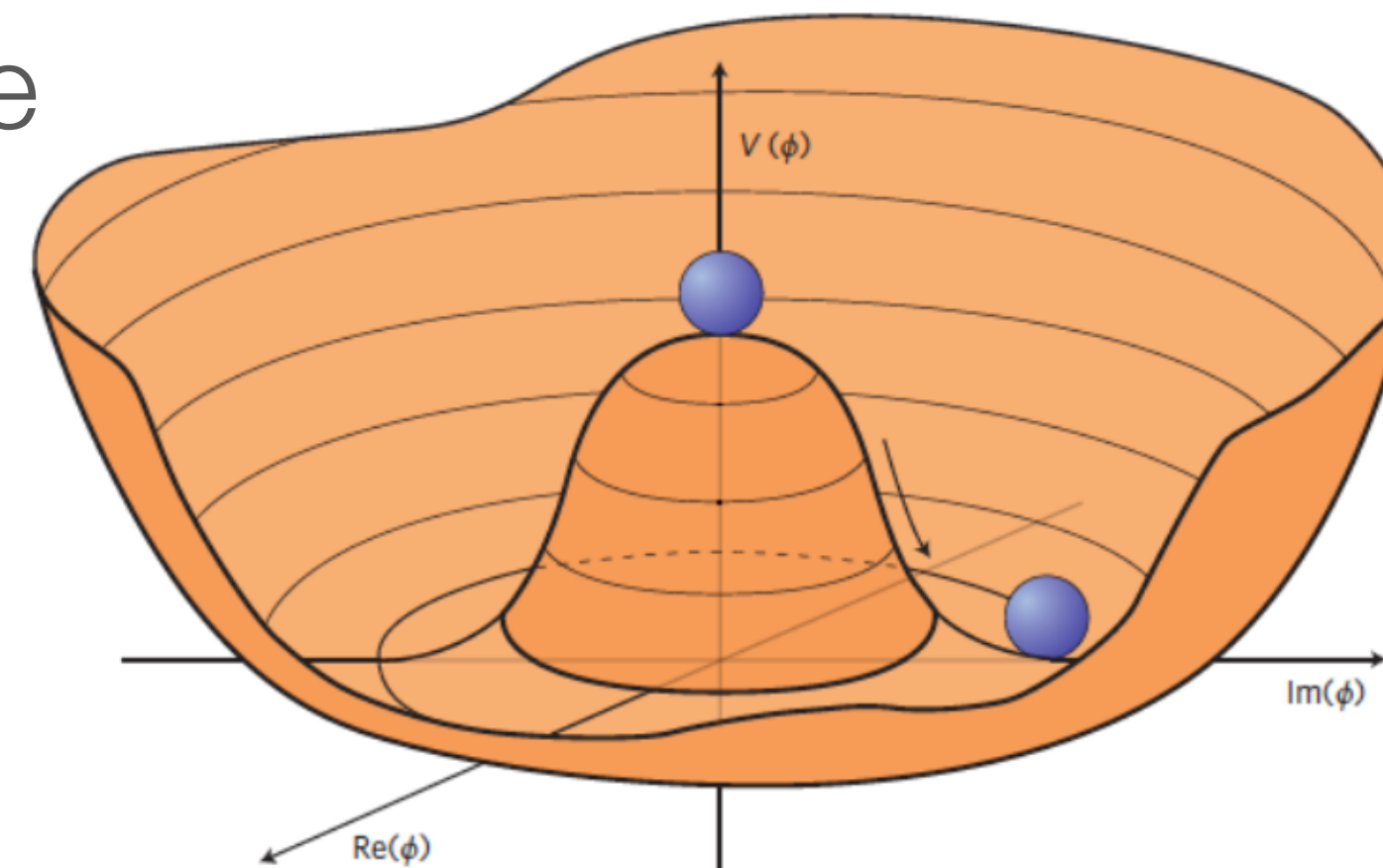
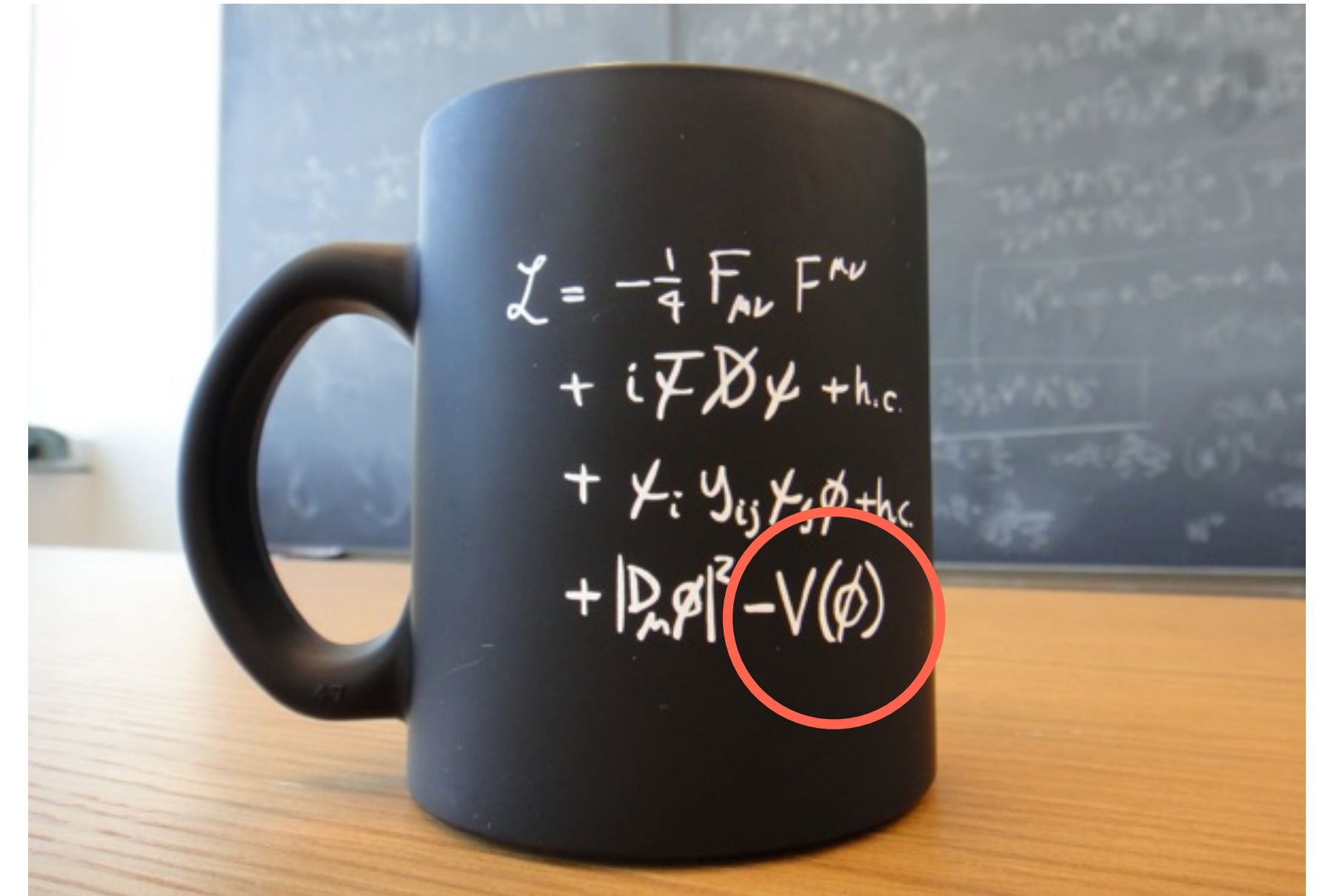
First, some physics

Intro

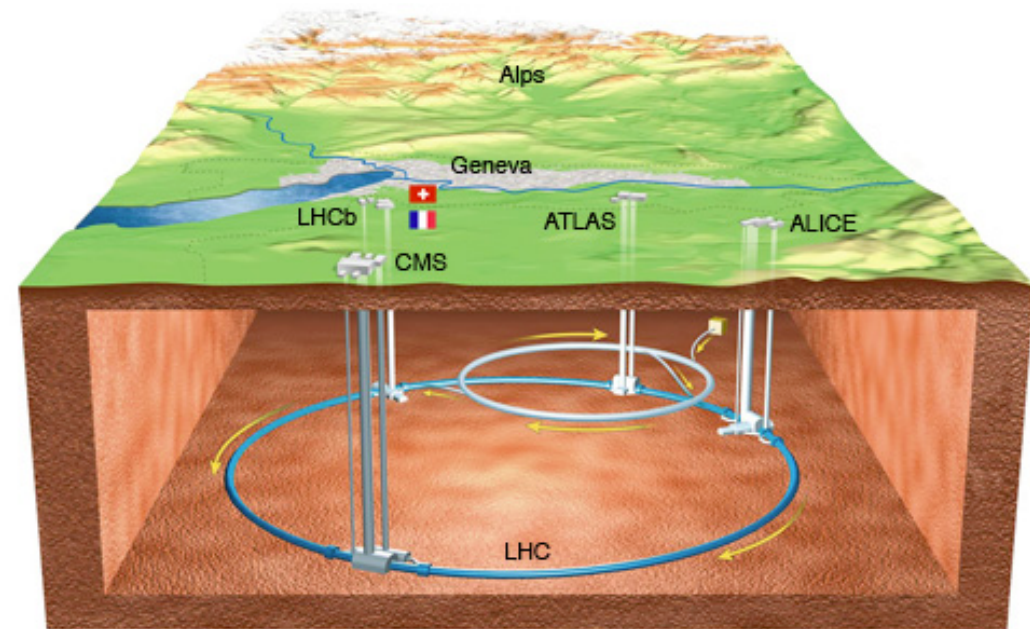
- Don't want to spend too much time talking about the physics
- I'm more interested in ML
- But we are paid to do physics after all
- So just to remind people of the basics

The Higgs

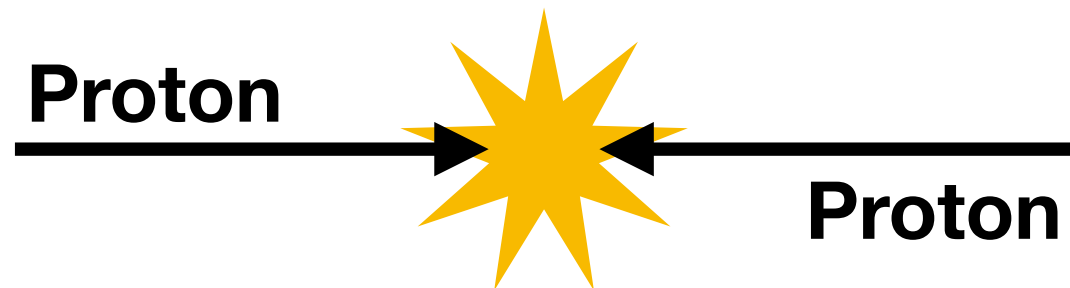
- The Higgs field is a scalar field with a value at every point in space-time
 - The **Higgs particle** is a local fluctuation in the **Higgs Field**
- In the Standard Model, particles gain mass through their interaction with the Higgs
- The properties of the Higgs sector have wide implications:
 - **Electroweak spontaneous symmetry breaking** at the beginning of the universe
 - Understand the **stability** of our universe
 - Origin of **Matter/antimatter asymmetry**



The LHC



- The LHC collides bunches of protons travelling at just 3ms slower than the speed of light 40M times per second



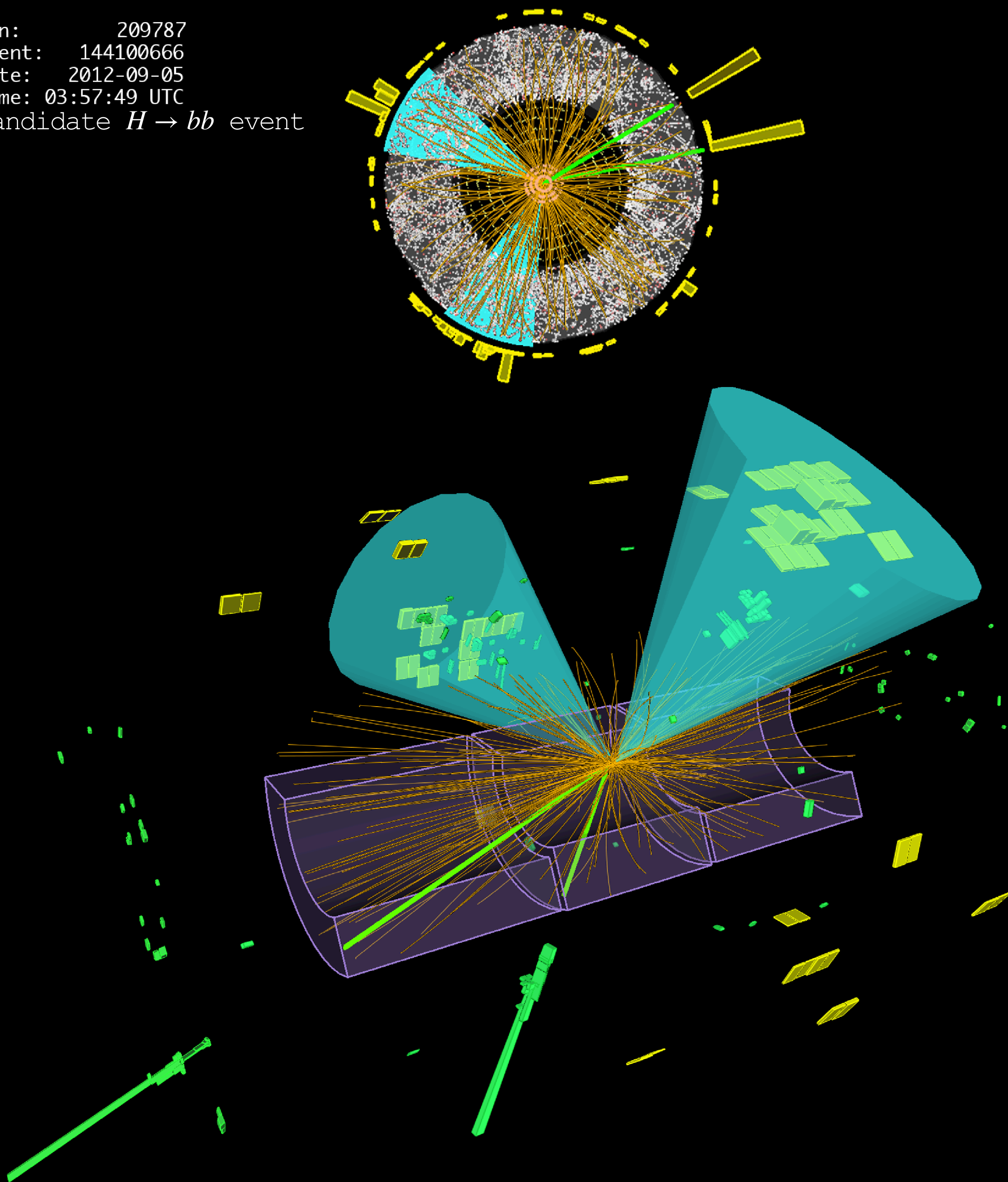
- New and interesting particles are produced in the collisions



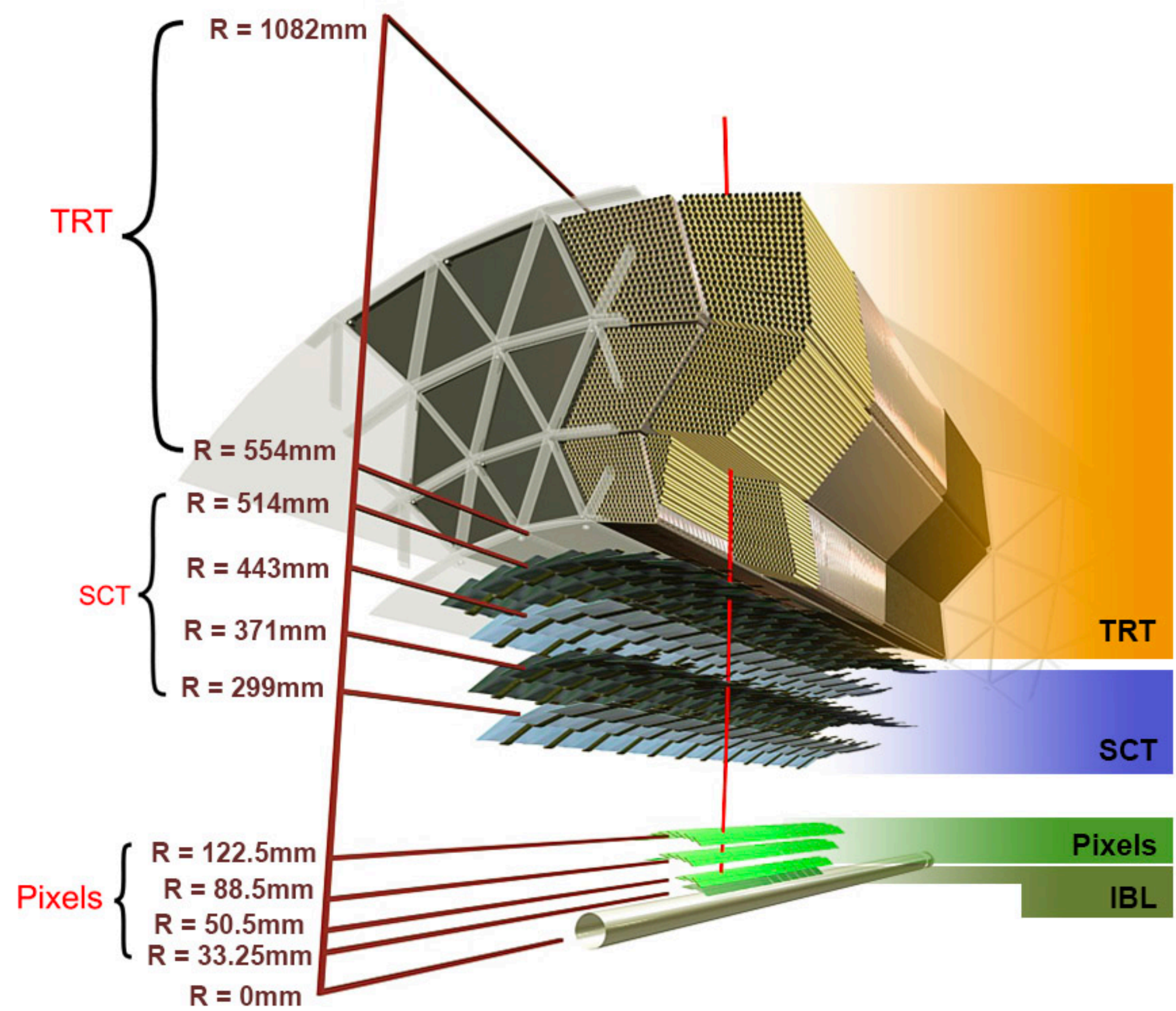
- These quickly decay to stable particles which are measured by specialised detector systems

- **Crucial task for collider physics:** Identifying and measuring (aka reconstructing) the produced particles

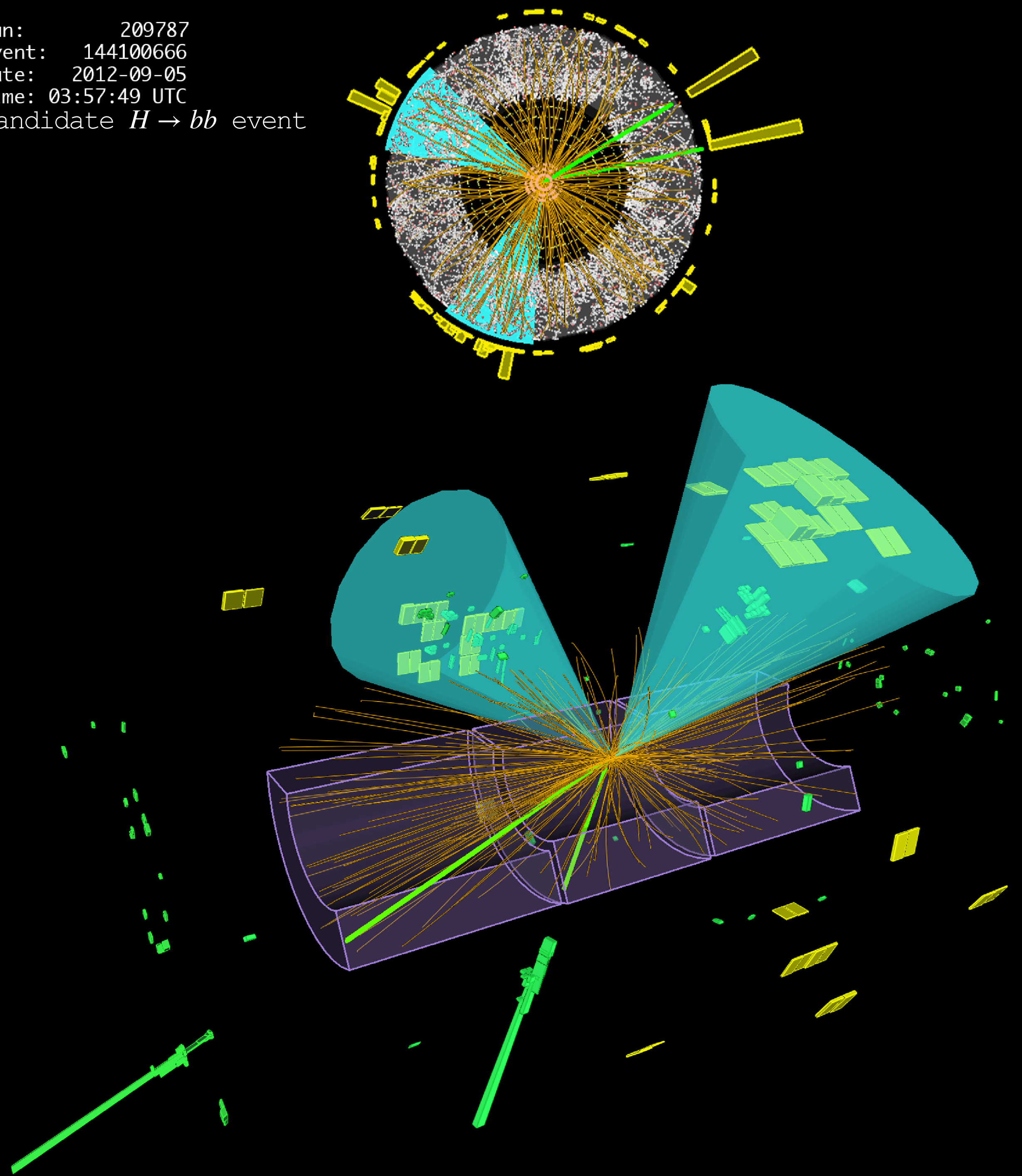
Run: 209787
Event: 144100666
Date: 2012-09-05
Time: 03:57:49 UTC
Candidate $H \rightarrow bb$ event



ATLAS



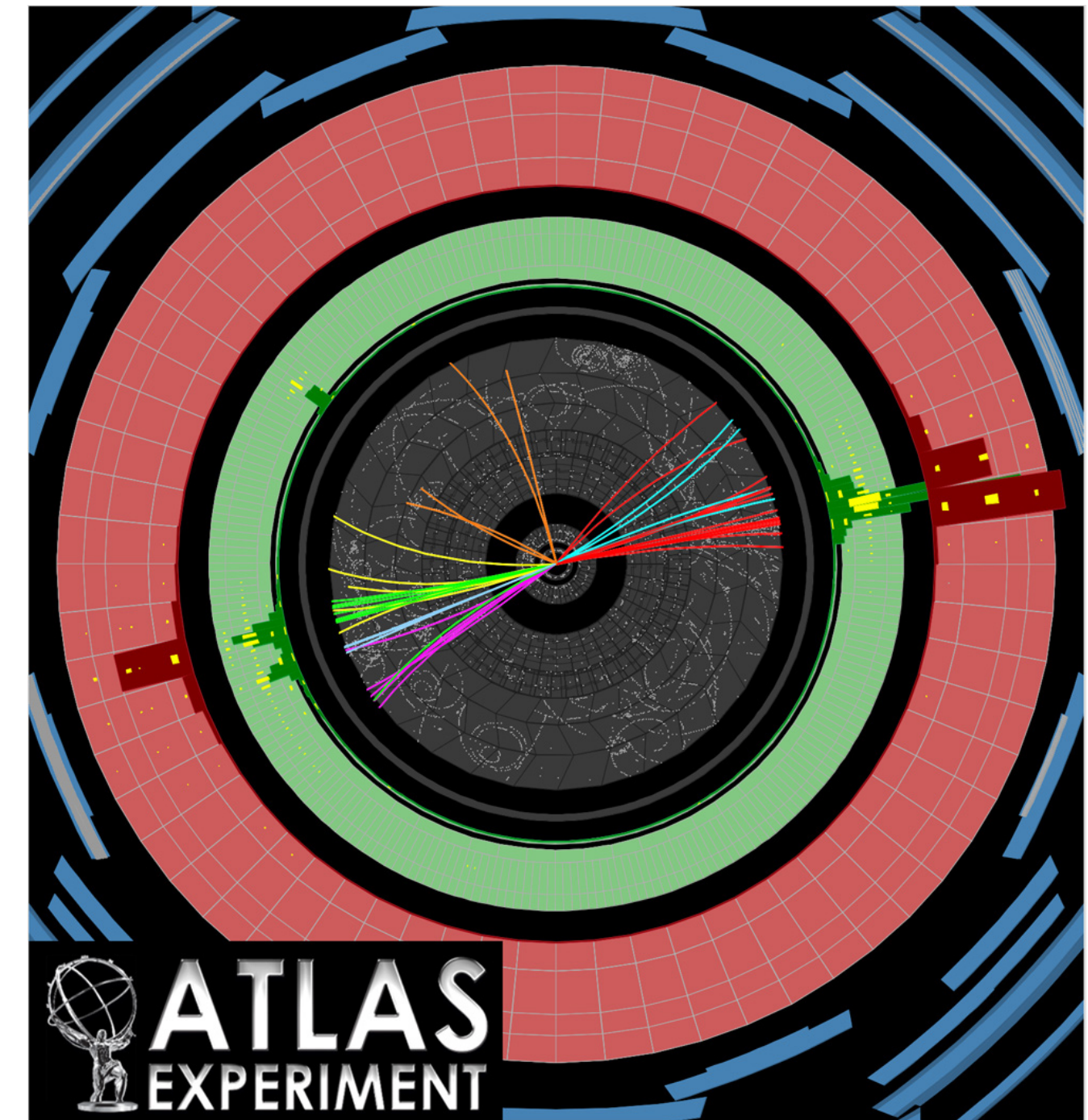
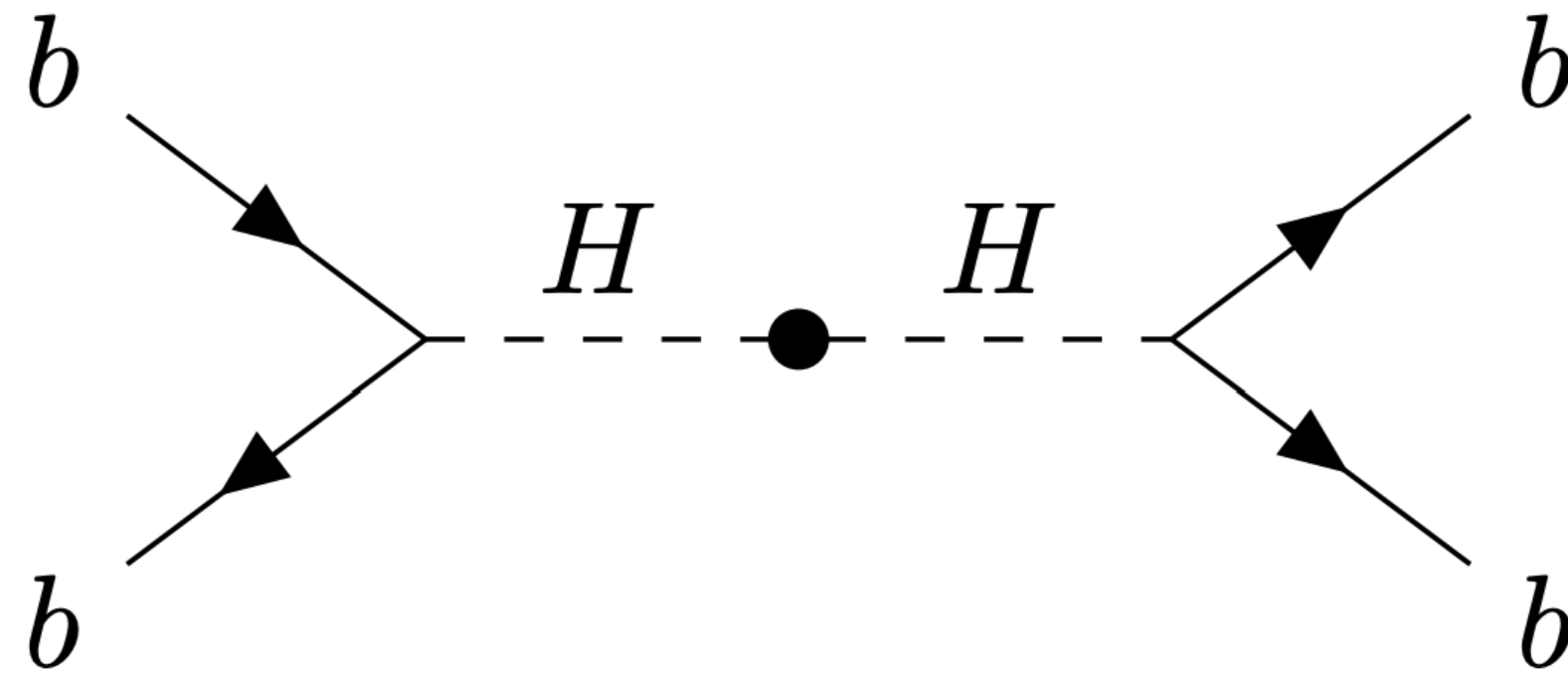
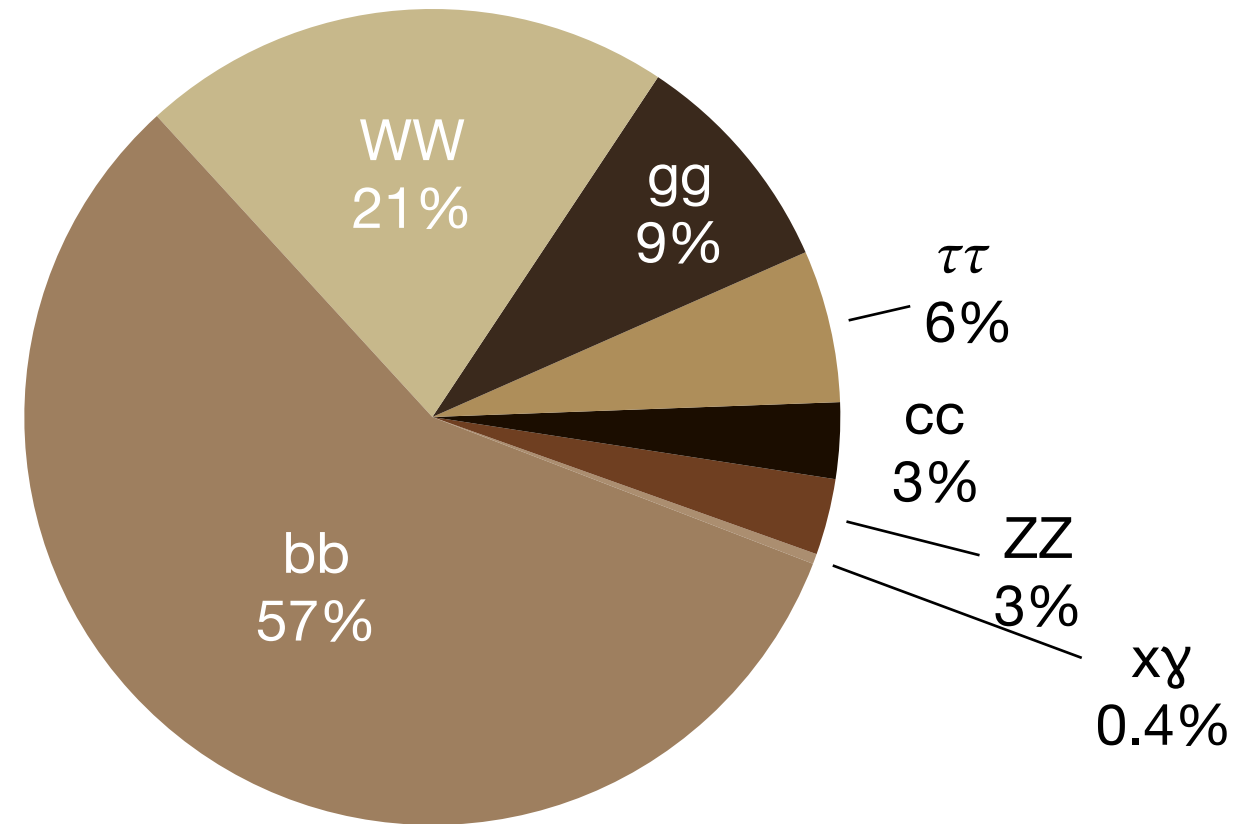
Run: 209787
Event: 144100666
Date: 2012-09-05
Time: 03:57:49 UTC
Candidate $H \rightarrow bb$ event



b-jets and the Higgs

The Higgs preferentially decays to a pair of a pair of b-hadrons. These are reconstructed in the ATLAS detector as **b-jets**

Decays of a 125 GeV SM Higgs



We can reconstruct the Higgs by identifying the b-jets it leaves behind
This is **crucial** for measuring the Higgs!

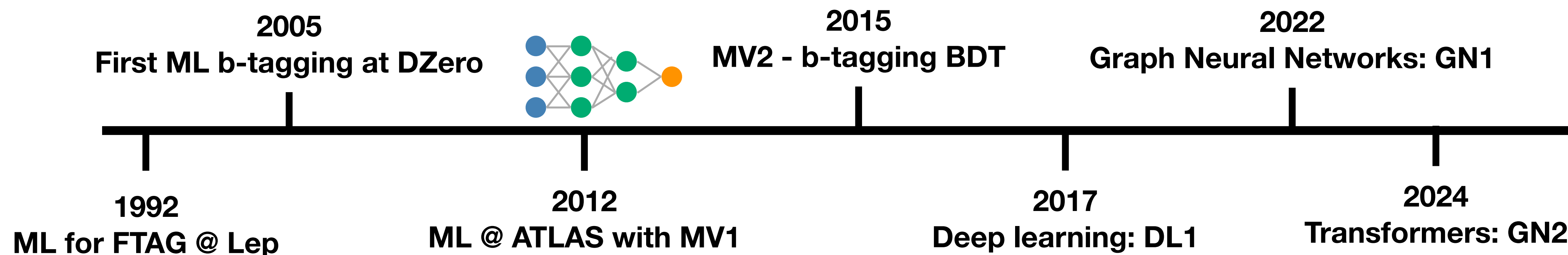
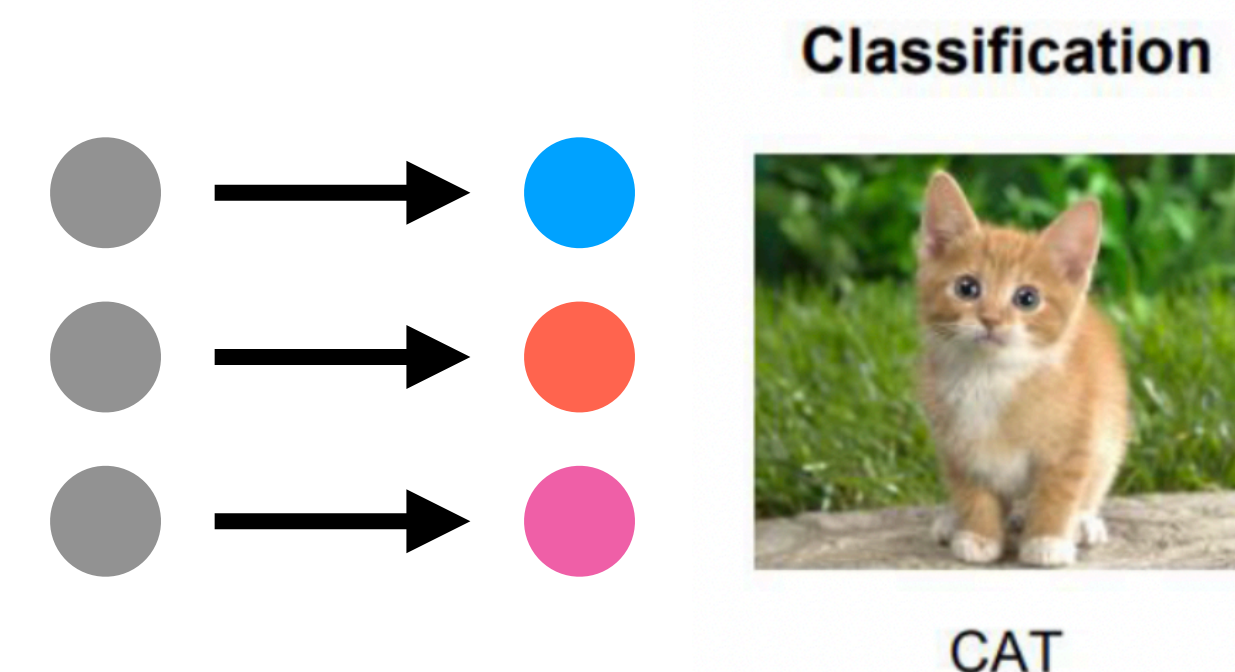
Flavour tagging

Multivariate Analysis Methods to Tag b Quark Events at LEP/SLC

B. BRANDL⁺, A. FALVARD⁺⁺, C. GUICHENEY⁺⁺,
P. HENRARD⁺⁺, J. JOUSSET⁺⁺, J. PRORIOI⁺⁺

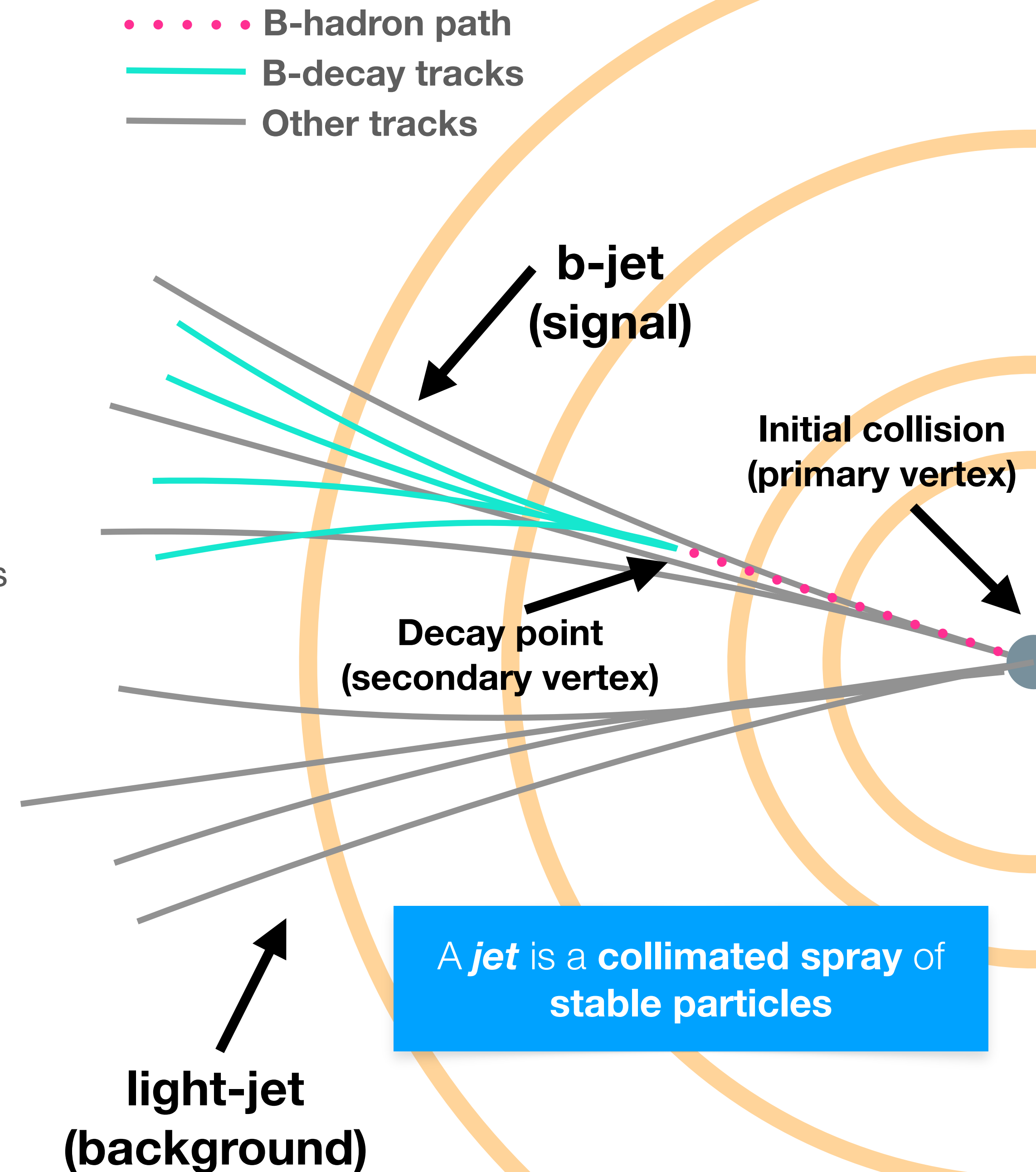
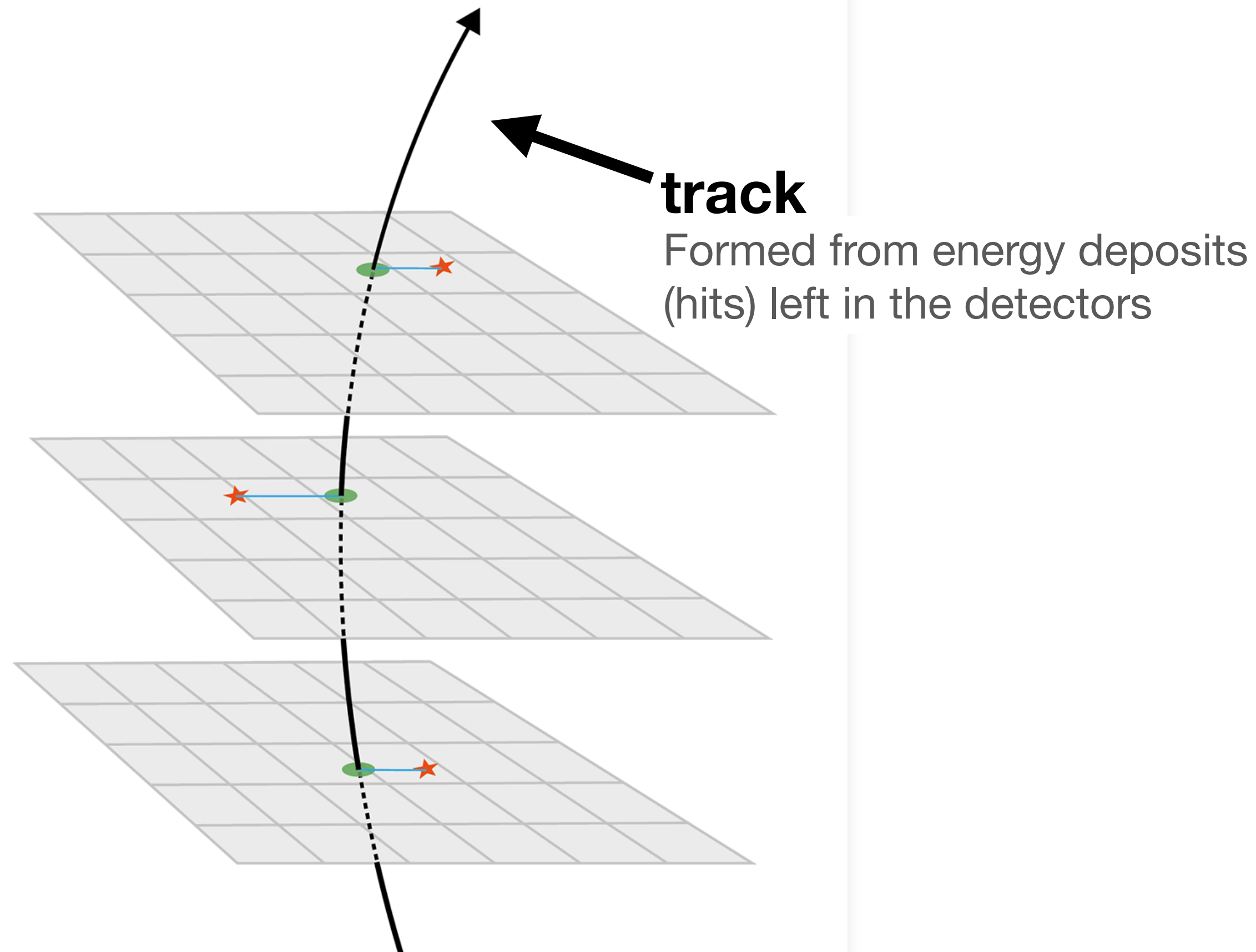
ML in Particle Physics

- ML has been widely adopted in particle physics: massive improvements over hand-crafted data analysis algorithms!
- Typical ML applications are **per-object** (classification or regression)
- Jet flavour identification (aka “flavour tagging”) is a critical challenge and well suited for ML and has seen constant incremental advancements over two decades



Inputs: Tracks & Jets

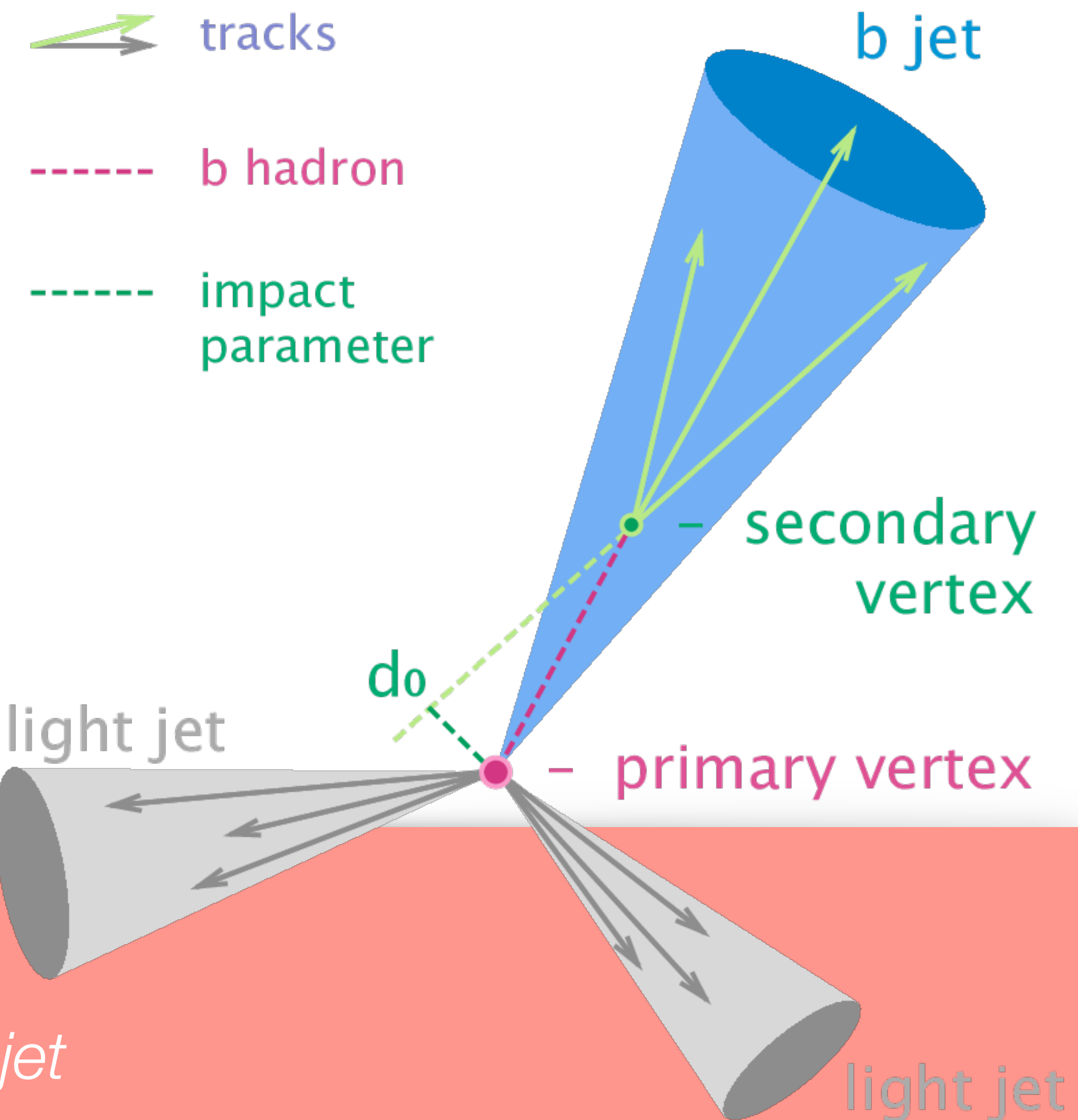
A **track** is the **reconstructed trajectory** of a **charged particle**



b-jets

The (relatively) long lifetime of b-hadrons ($\sim 1.5\text{ps}$) can allow for significant displacement before decay, leading to the following typical signatures (compared to ‘light-flavour’ jets (which make up $\sim 90\%$ of all jets))

- The b-hadron decay products contain most of the jet energy
- Displaced secondary vertices (with high mass)
- the subsequent displaced decay of the c-hadron $B \rightarrow C$
- Large track impact parameters (d_0)
- Missing hits on inner detector layers
- A muon or electron in the decay



Flavour tagging (FTAG) algorithms (“taggers”):

Hypothesise the jet flavour using the tracks associated to the jet

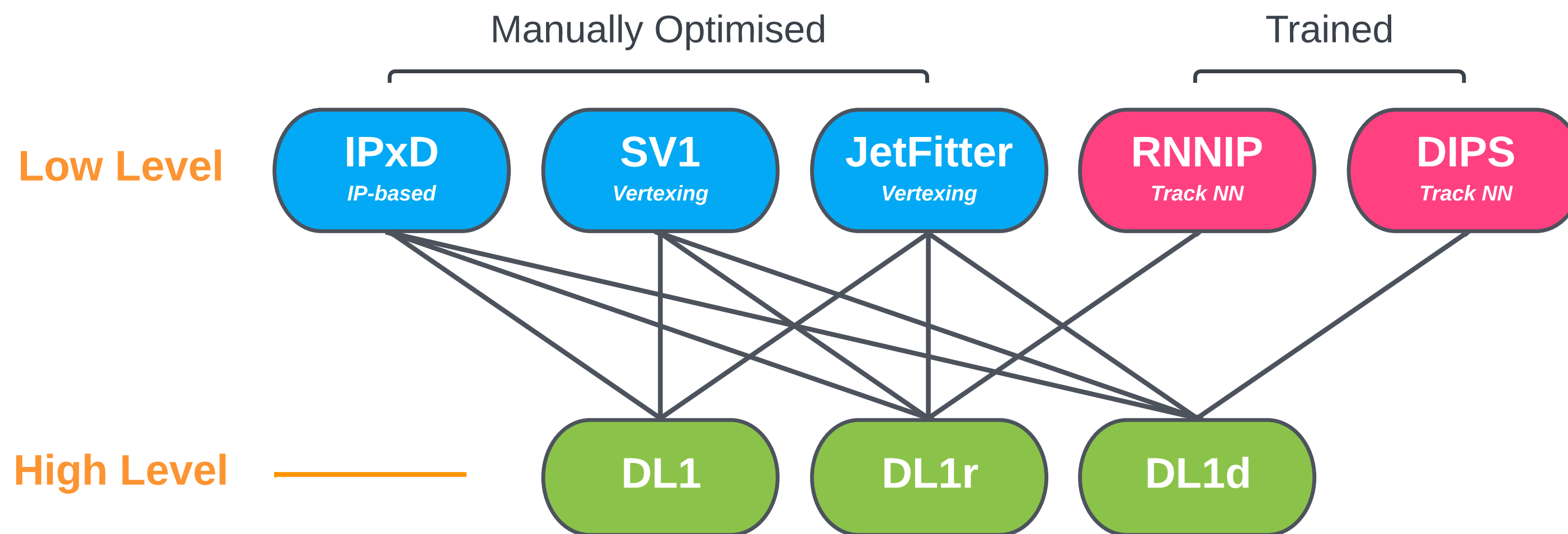
Flavour Tagging @ ATLAS

Inputs are fed to **low level taggers**:

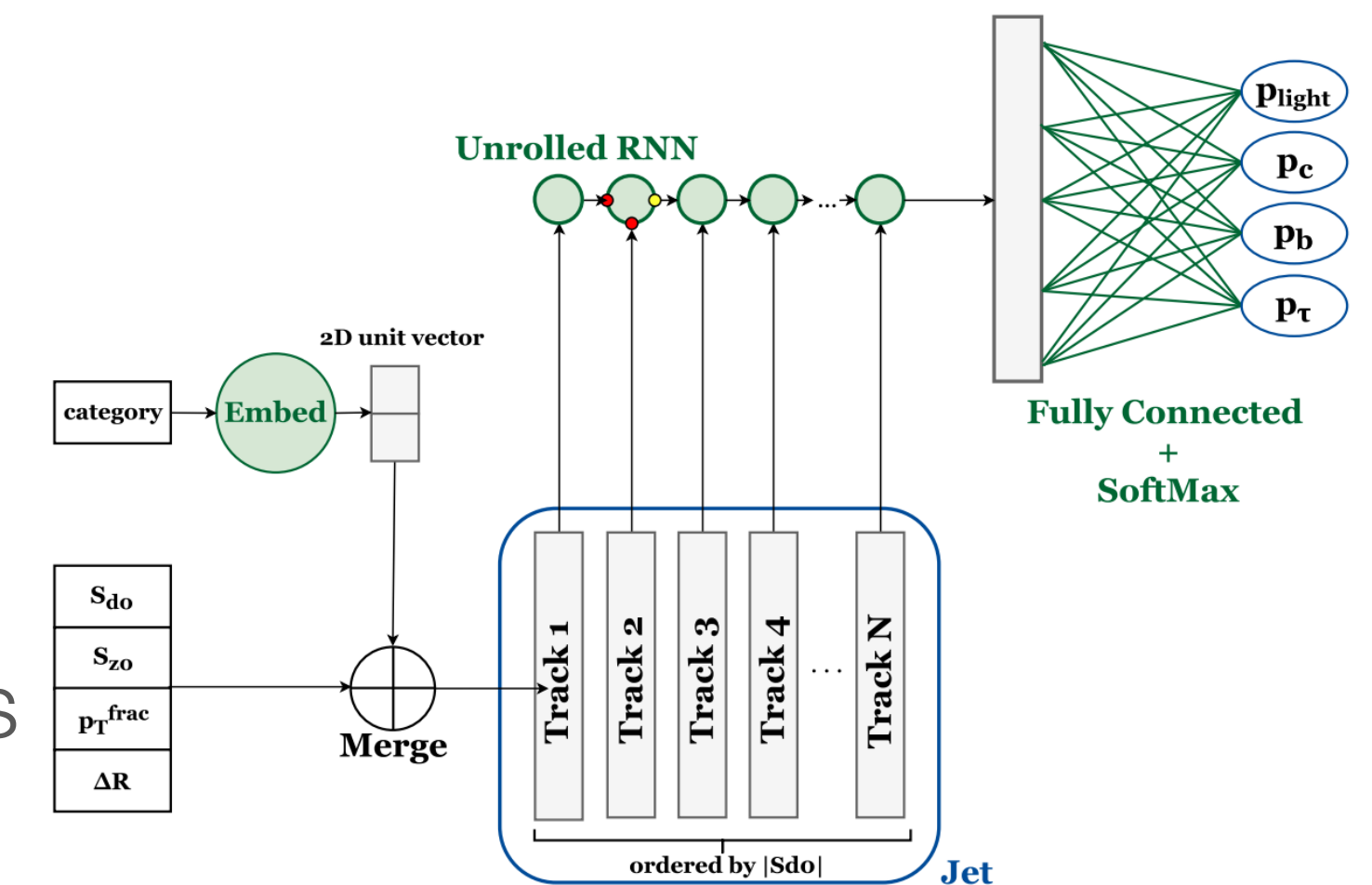
- **Algorithms using physics knowledge** - developed over decades
- **ML models** - developed more recently

Low level taggers are ensembled into **high level taggers** — BDTs or NNs

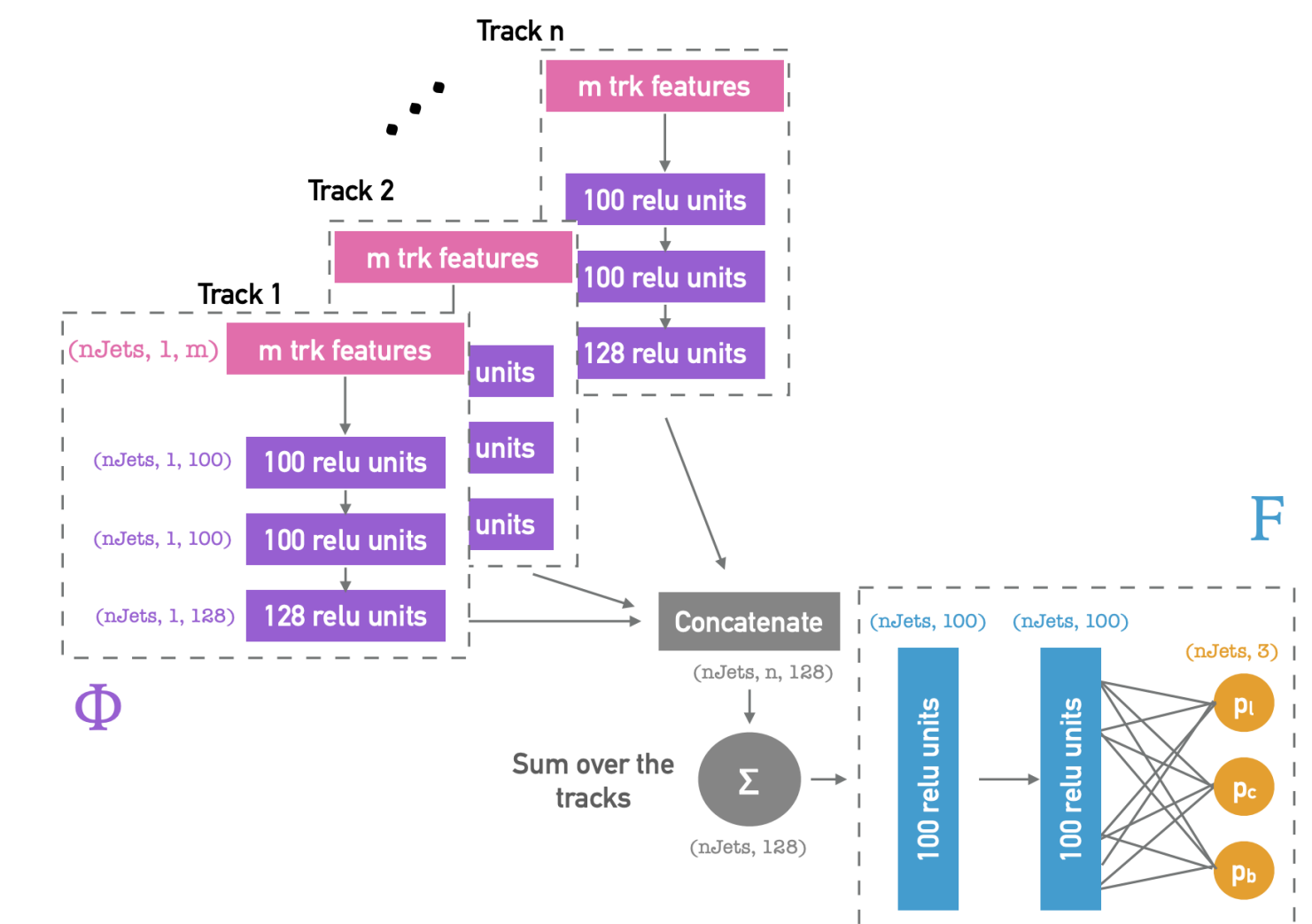
- ~40 inputs from the low level outputs
- Outputs: probabilities for each jet flavour class: p_b , p_c , p_l



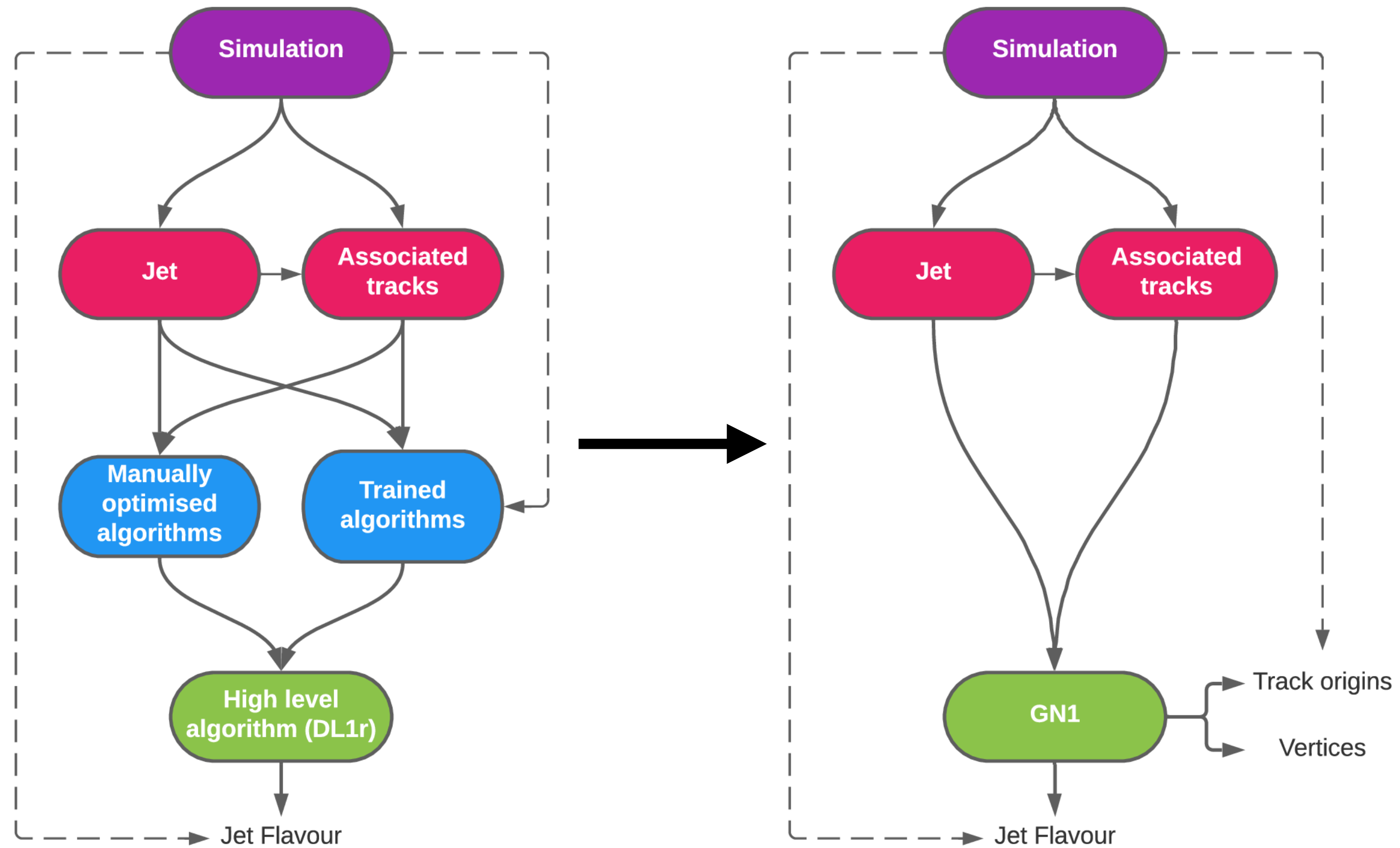
ATL-PHYS-PUB-2017-003



ATL-PHYS-PUB-2020-014



A New Approach



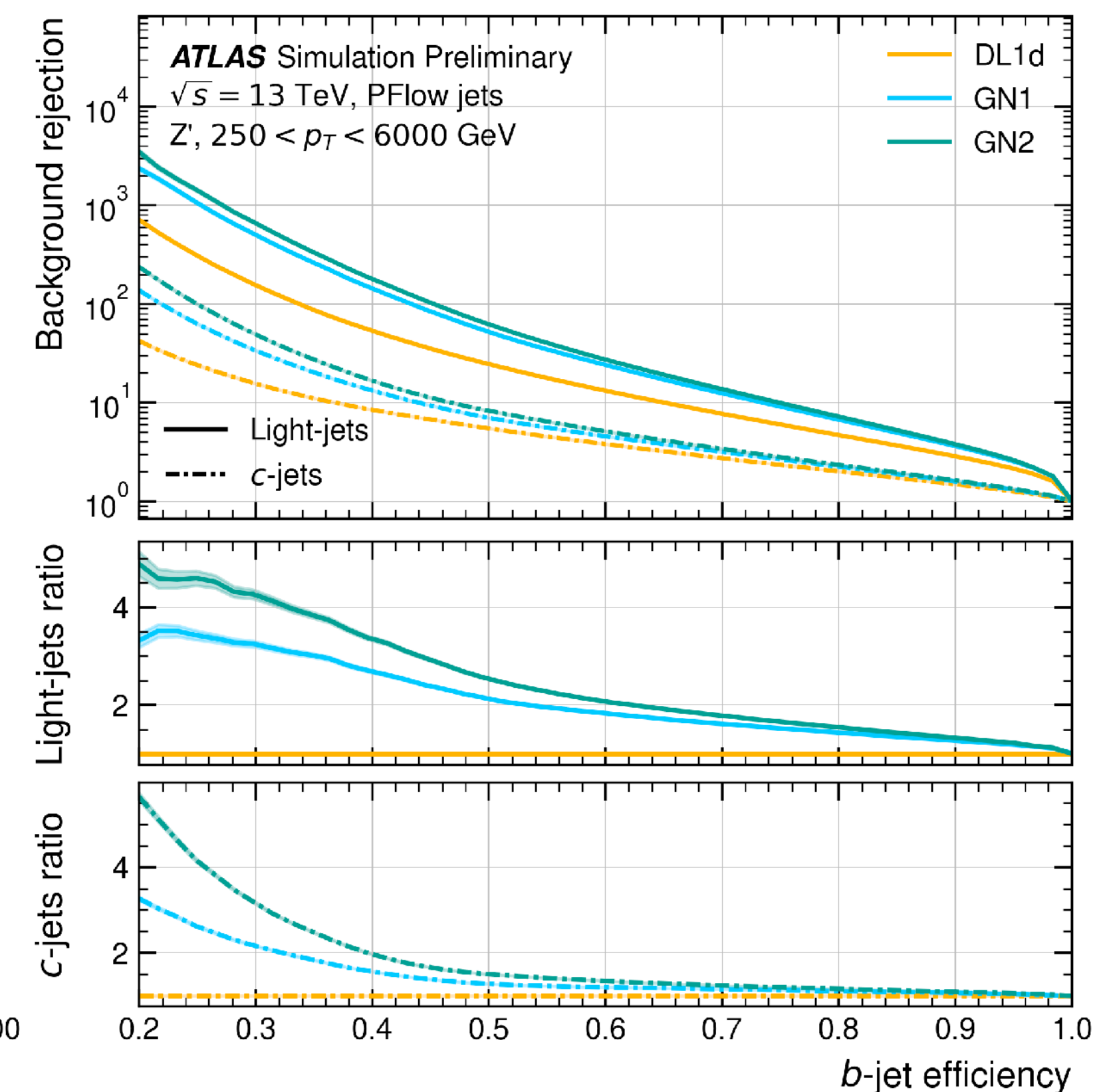
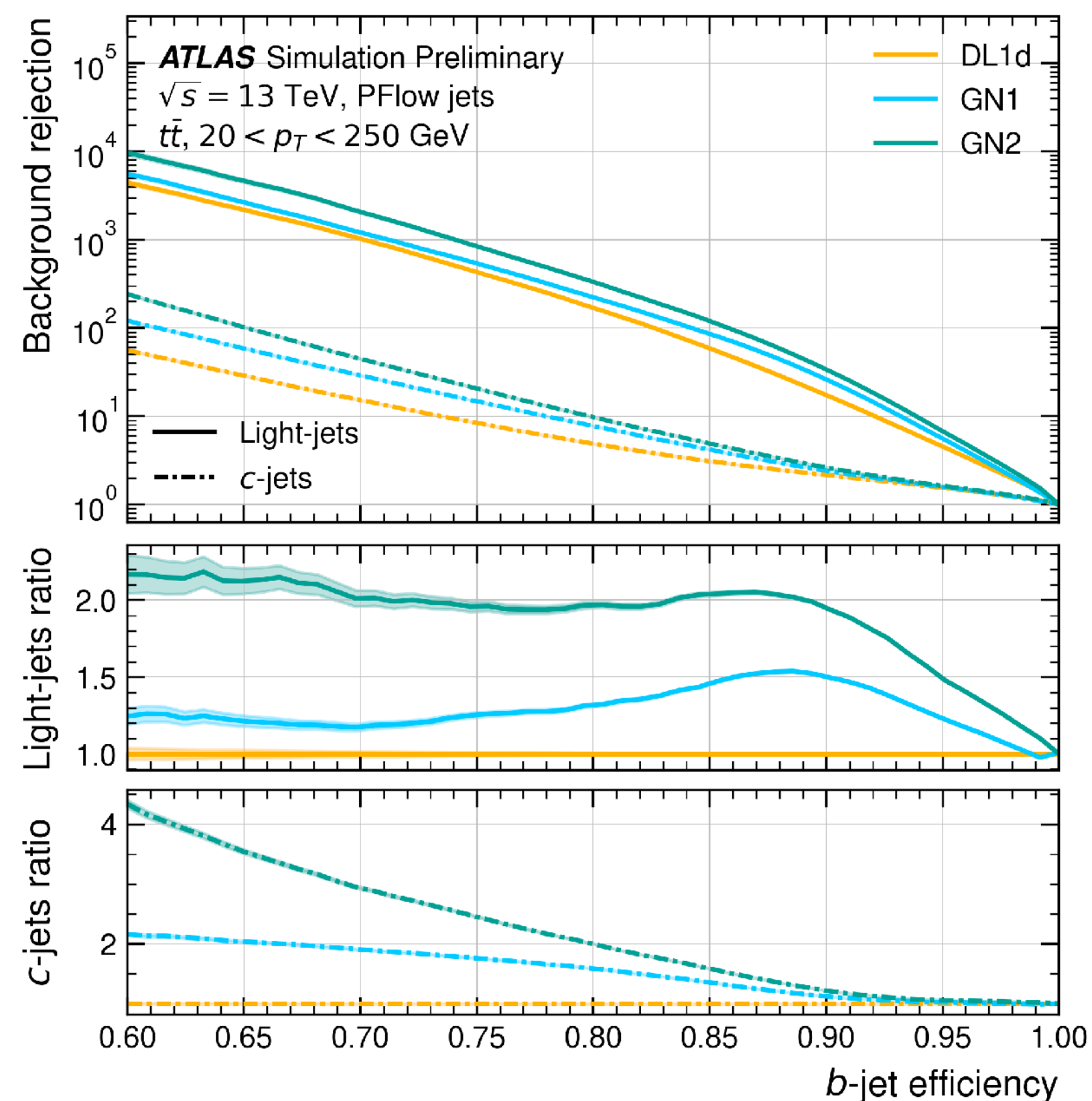
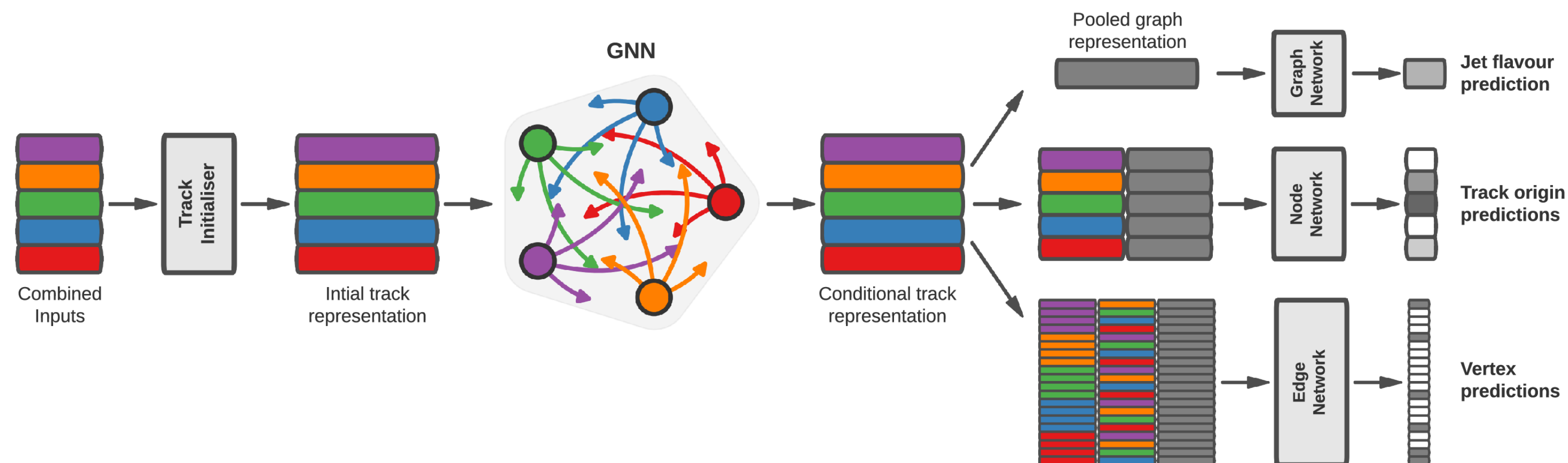
Based on previous work by [J. Shlomi et al, 2007.13681](#)

GN2 Tagger

CERN-EP-2025-103

Advanced ML achieves a step change in tagging performance

- Replace **5** models with **1**
- Extract physics information from jet - **simultaneous training** of jet flavour, track origin, and vertexing objectives
- **Improved performance** & **interpretability**
- **Easy optimisation** for different use cases
- **Future-proof:** integration of inputs from other sub-detectors, e.g. **leptons**, **neutrals**, **hits**



Aside: transformers

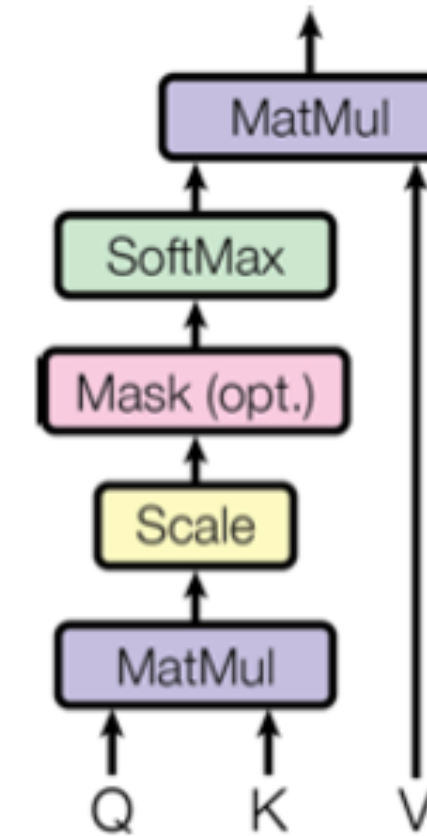
Attention Is All You Need

GN2 uses a transformer model

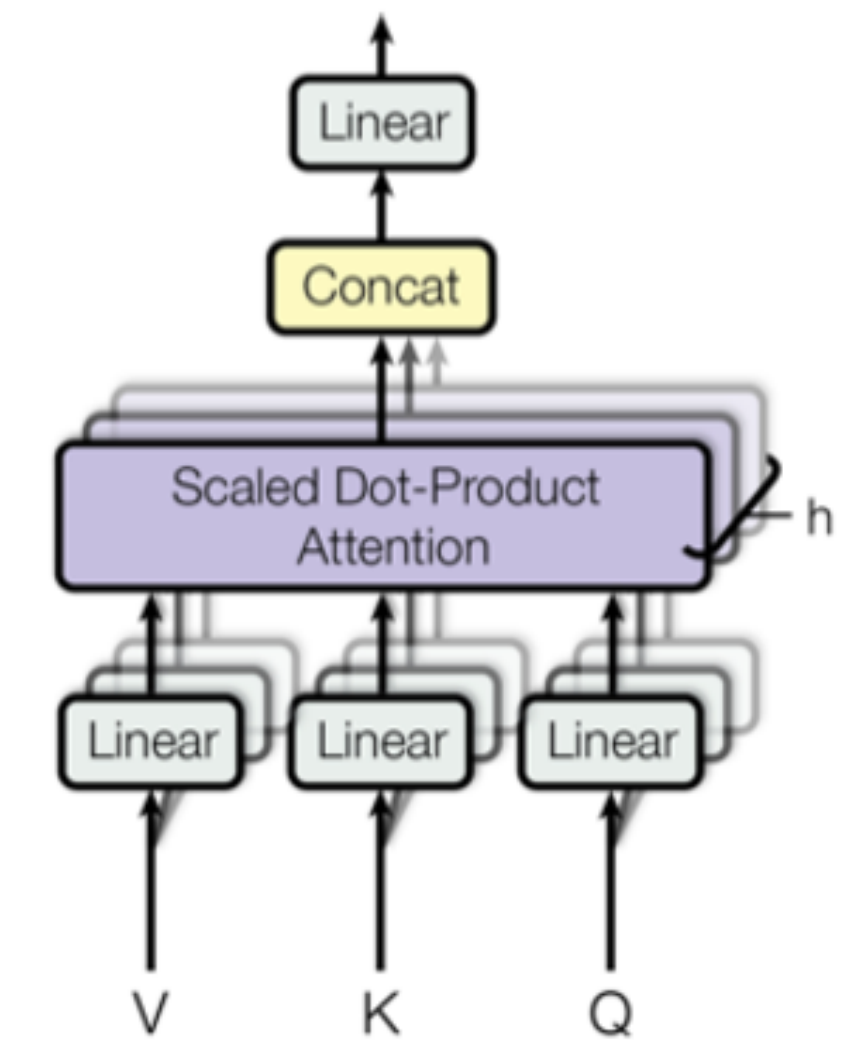
- Originally arose from the field of natural language processing
- Transformers are what power current large language models like ChatGPT
- The key to the transformer architecture is the **Attention Mechanism** – calculates the pairwise relevance between outputs
- Tracks are updated by an attention weighted average of all other tracks in the jet

$$Output_i = \sum_j \alpha_{ij} v_j$$

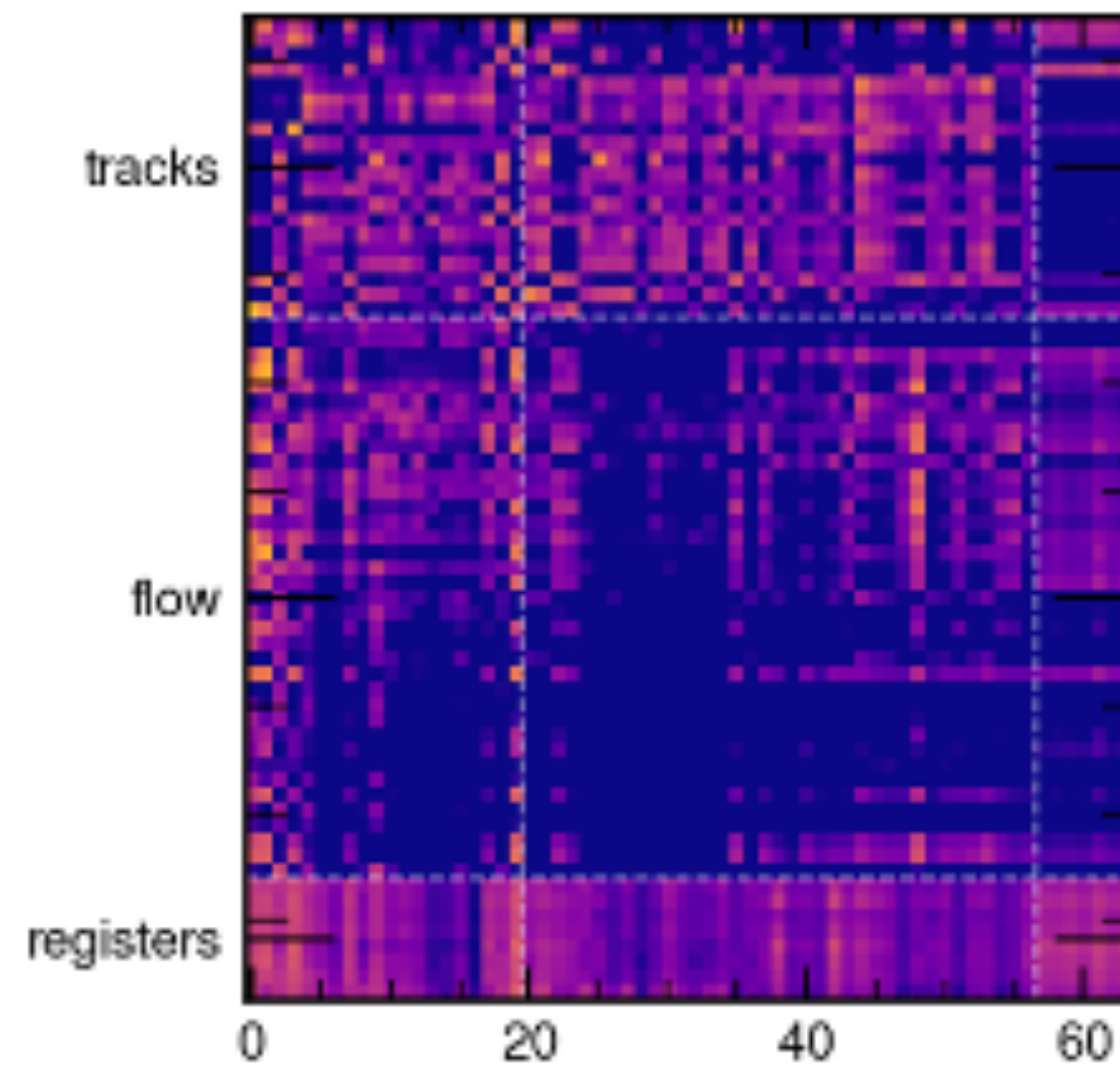
Scaled Dot-Product Attention



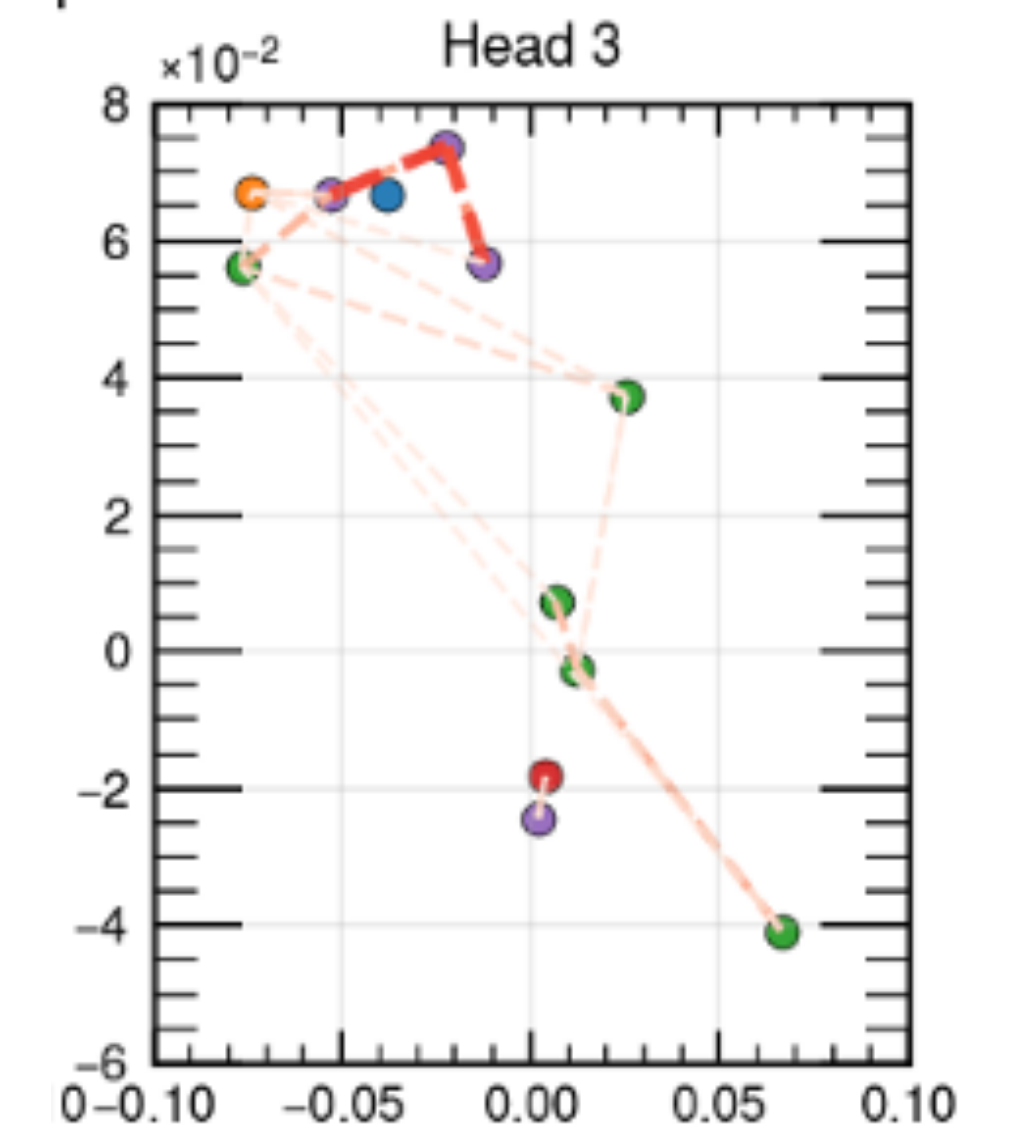
Multi-Head Attention



Head 0

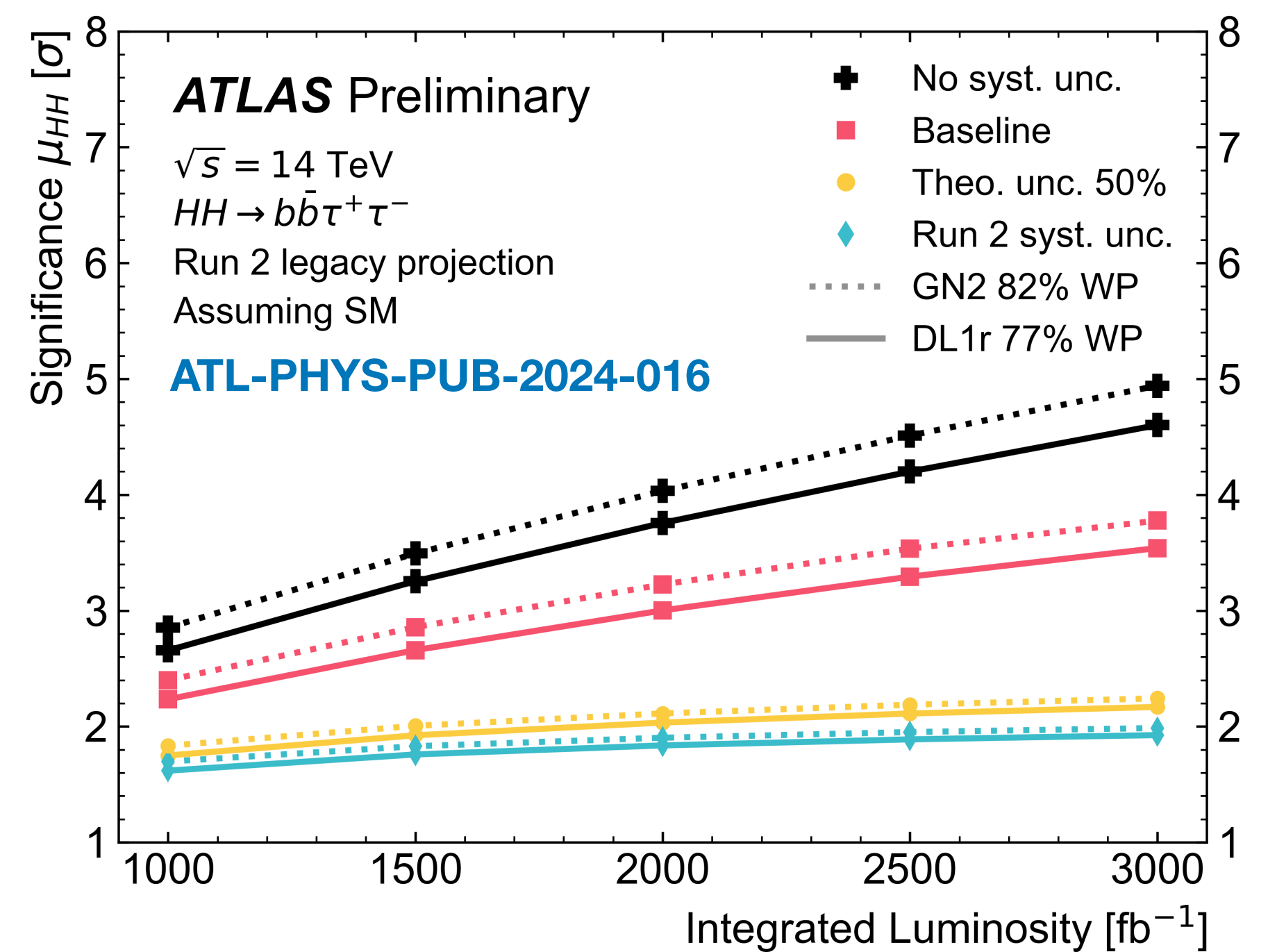
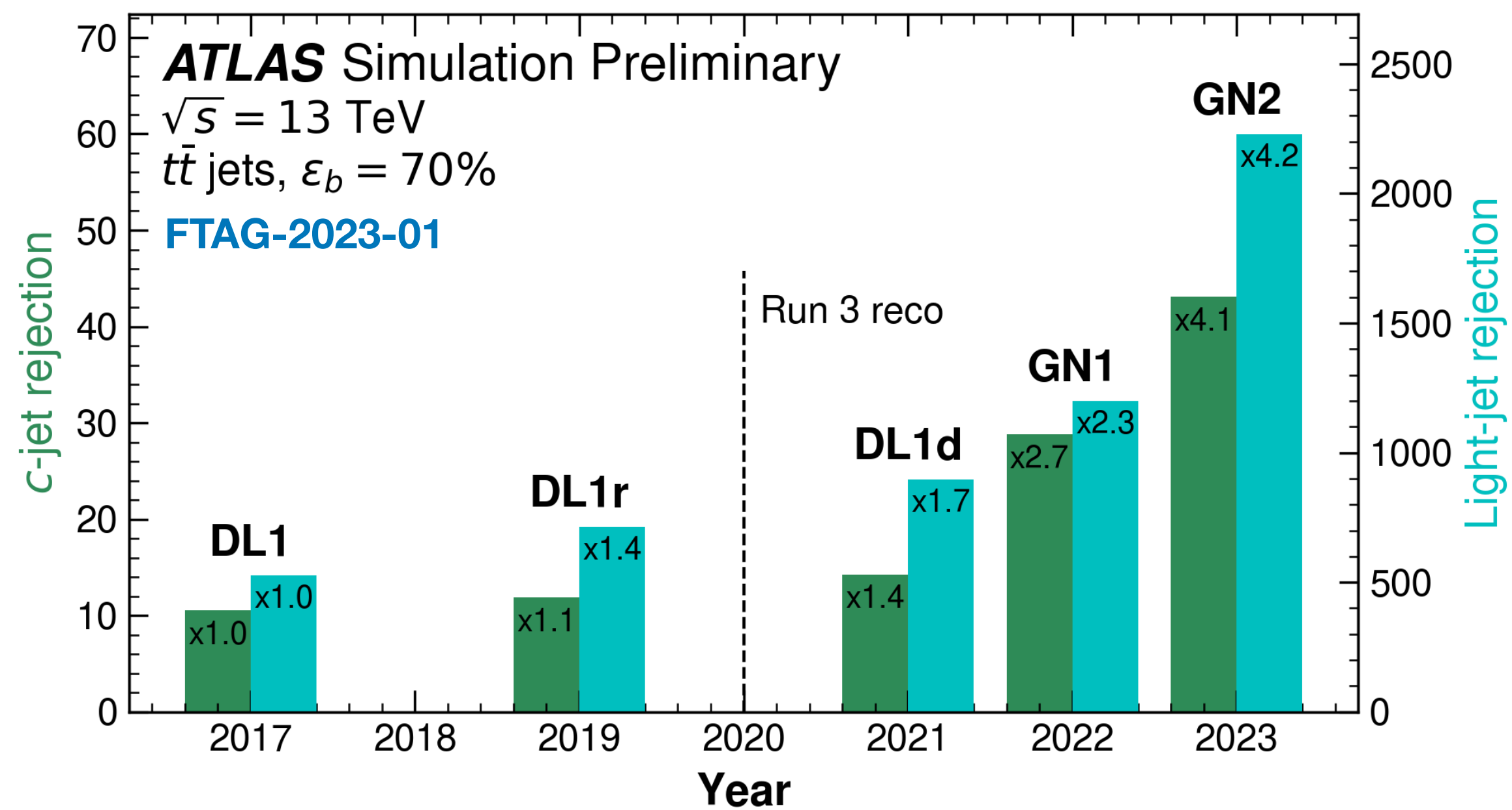


| hbb



Steady Progress...

4x improvement in background rejection since DL1,
directly contributes to improved discovery significance at the HL-LHC

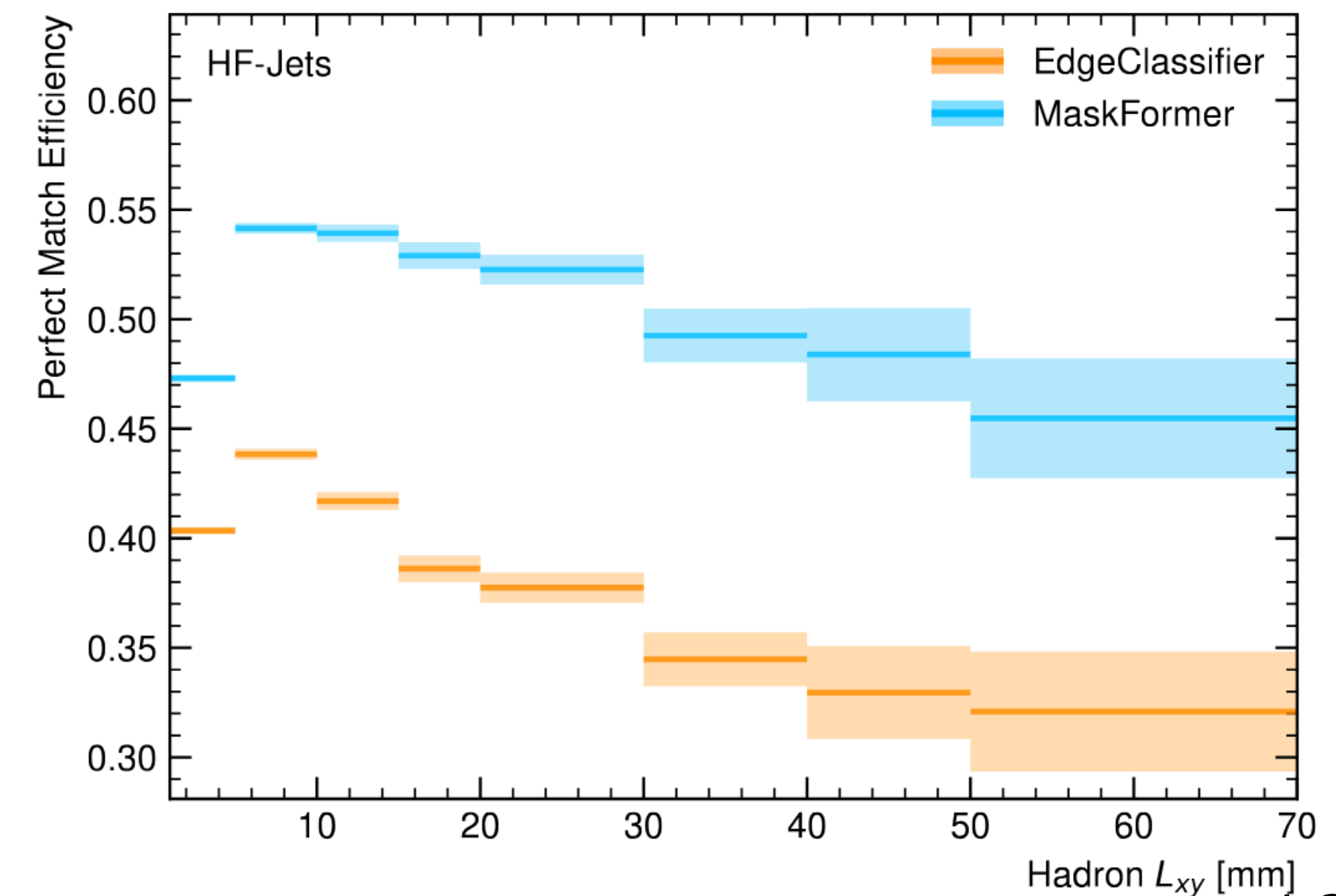
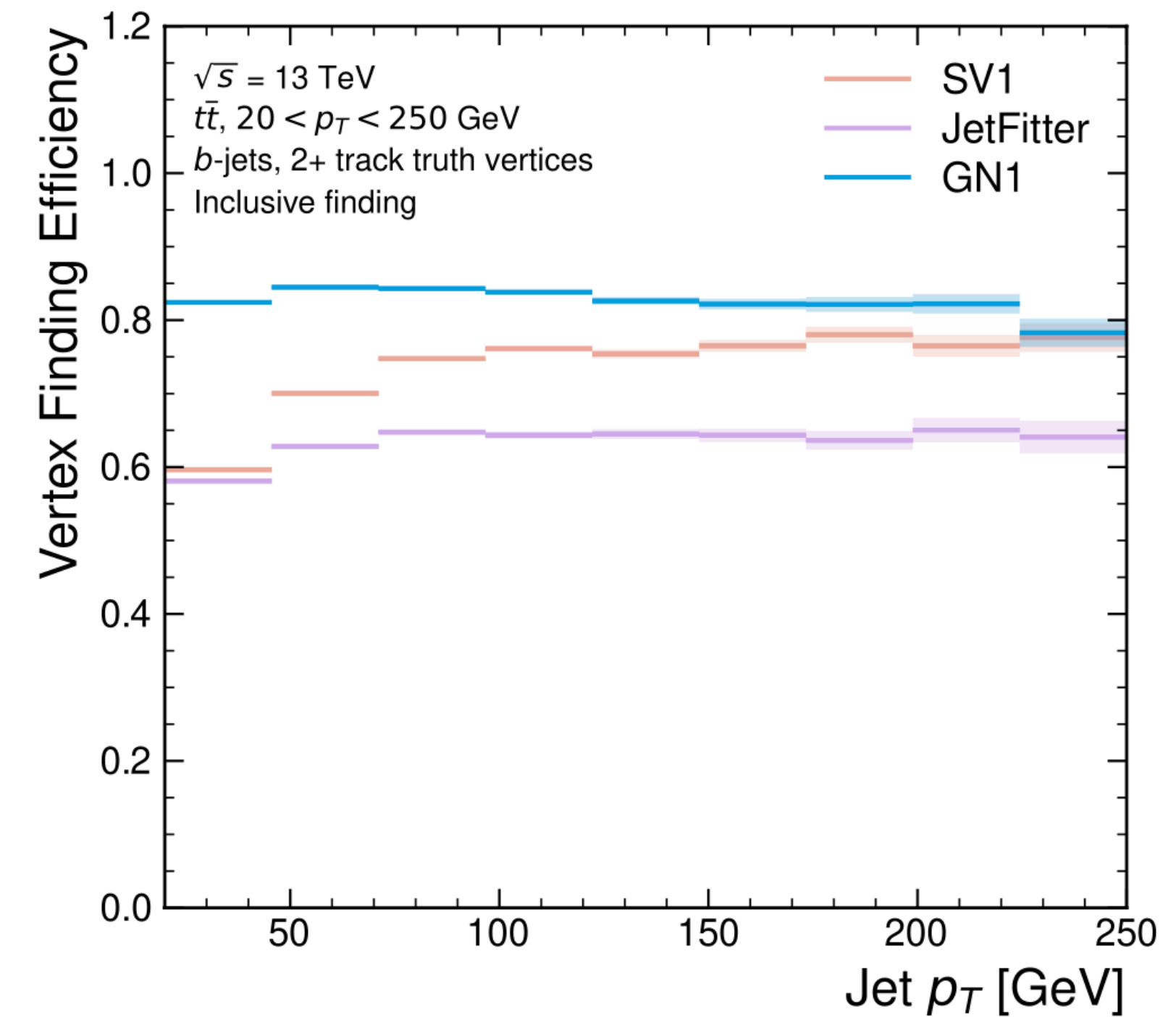


Understanding

Vertexing

GN2 is trained with two auxiliary tasks: track classification and vertexing

- These improve the performance of the main task, and provide additional understanding of what the model is doing
- They replace inputs from low level vertexing algorithms: SV1 and JetFitter
- More recently: using MaskFormer architecture to improve vertexing **2312.12272**
 - Powerful new type of model!

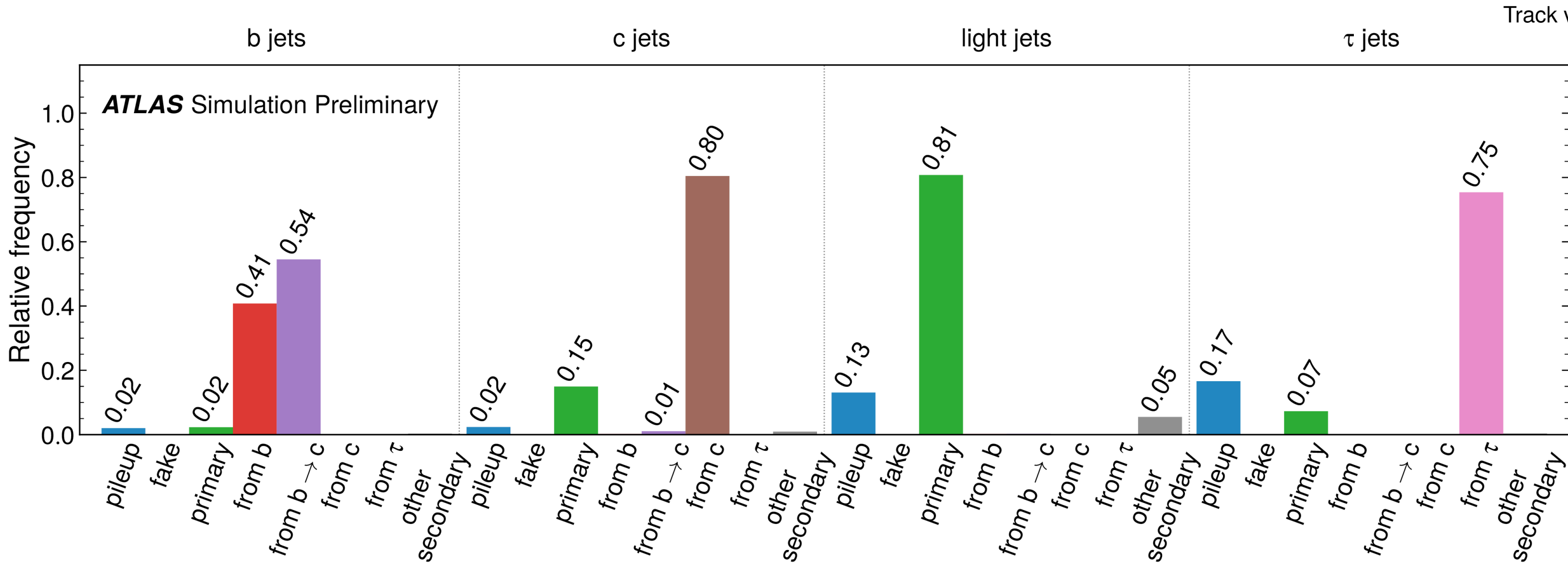
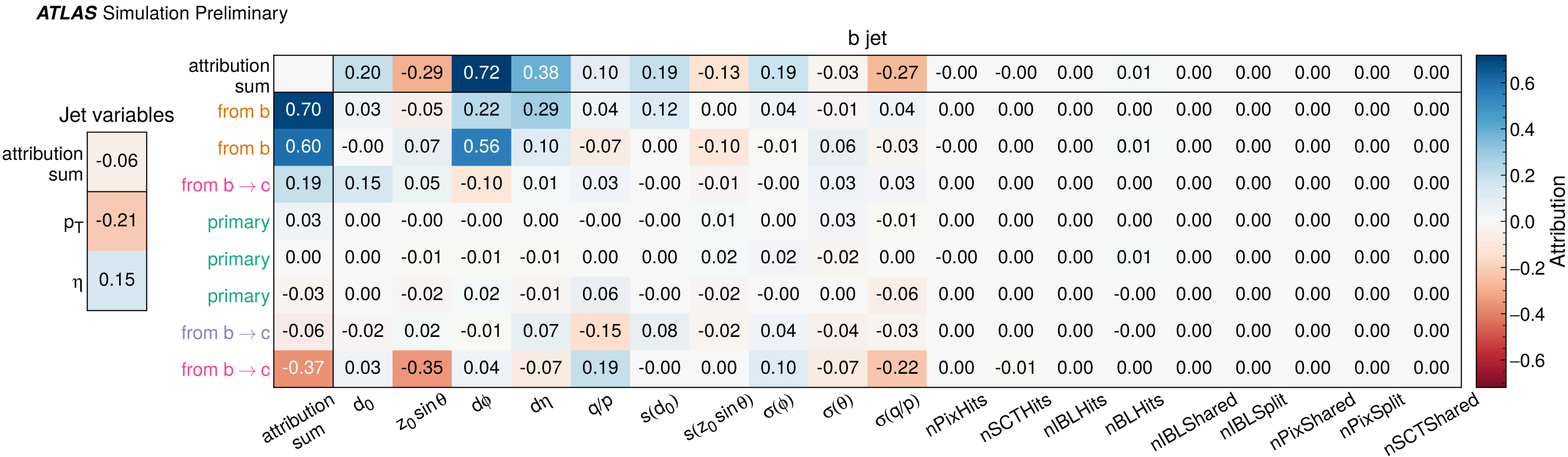


Breaking open the black box

FTAG-2024-01

Approaches like **integrated gradients** can help us to understand these complicated models

Track and feature attributions in a single jet



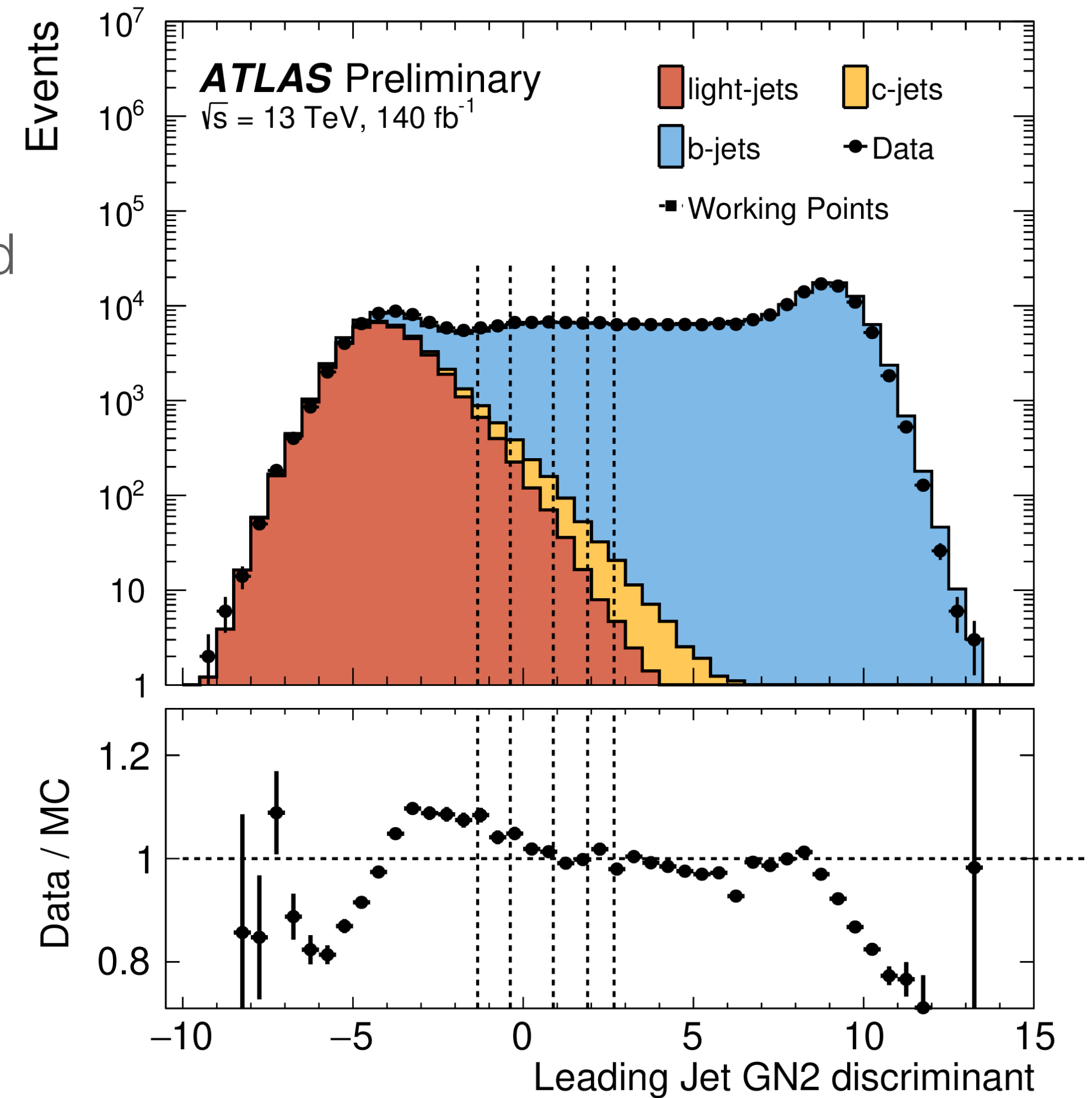
Average track attributions for different flavours and track origins

Calibration

How do we know performance gain observed in simulation is real?

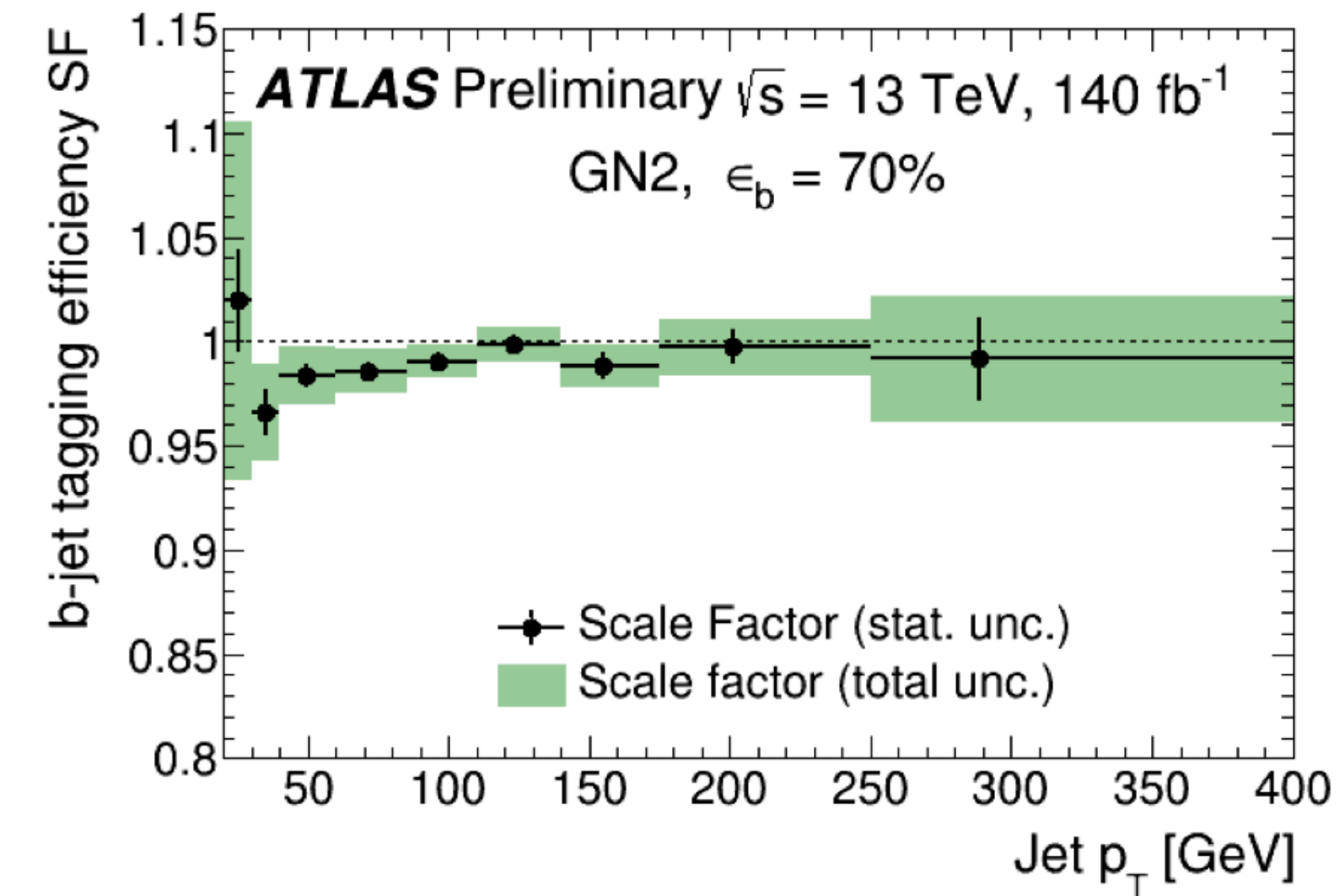
- Need to **calibrate** model by measuring performance in data and comparing to simulation
- We derive correction factors (**scale factors**) to match the simulation to data
- Three main calibrations are used to measure:
 - The b-jet efficiency in di-leptonic ttbar events
 - The c-jet mis-tag rate in semi-leptonic ttbar events
 - The light-jet mis-tag rate in Z+jets events

$$SF_f = \frac{\epsilon_f^{Data}}{\epsilon_f^{MC}}$$

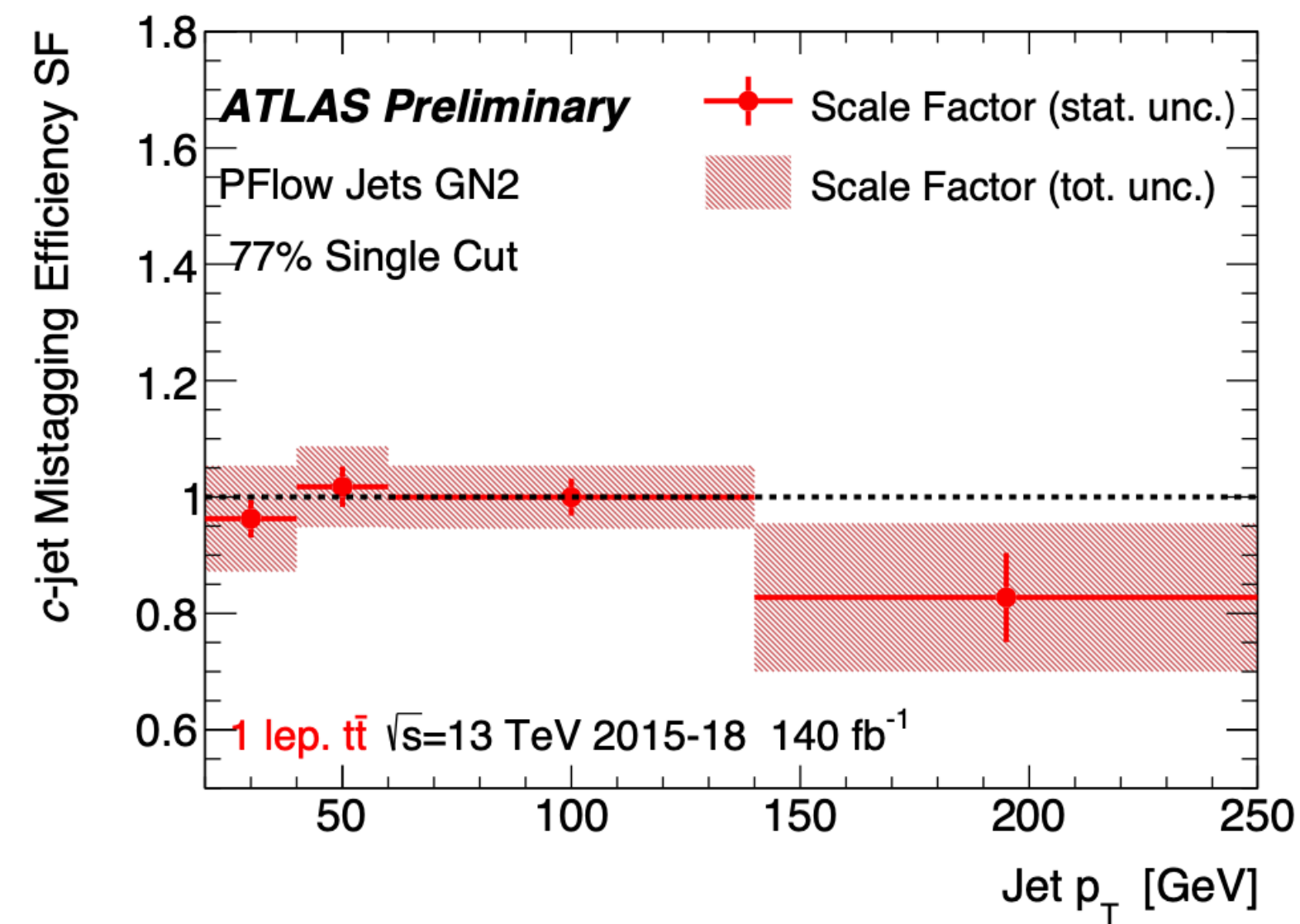


Performance in data

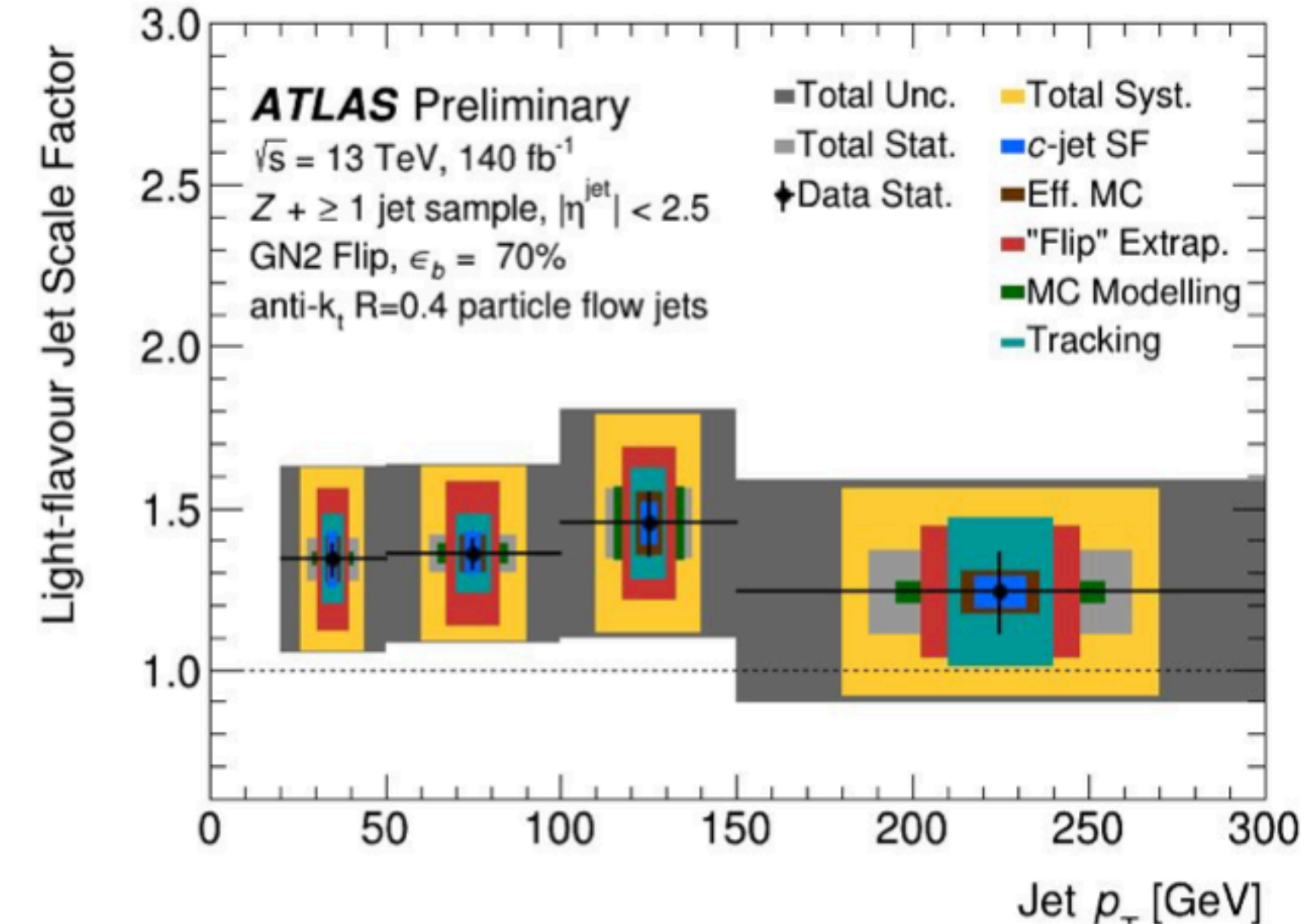
b-jets



c-jets



light-jets

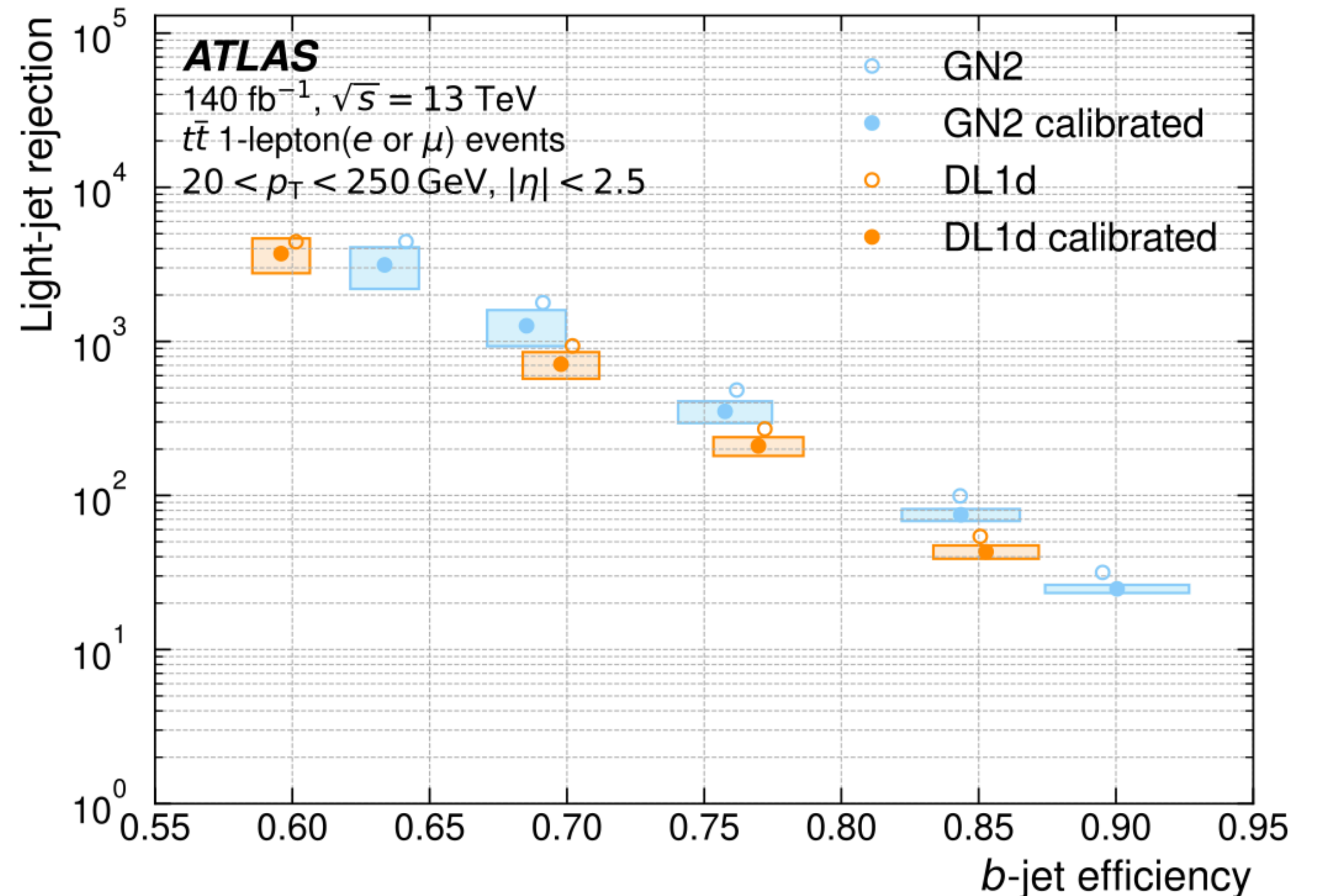


Good modelling observed for
both b-jets and c-jets —
biggest difference in light-jets

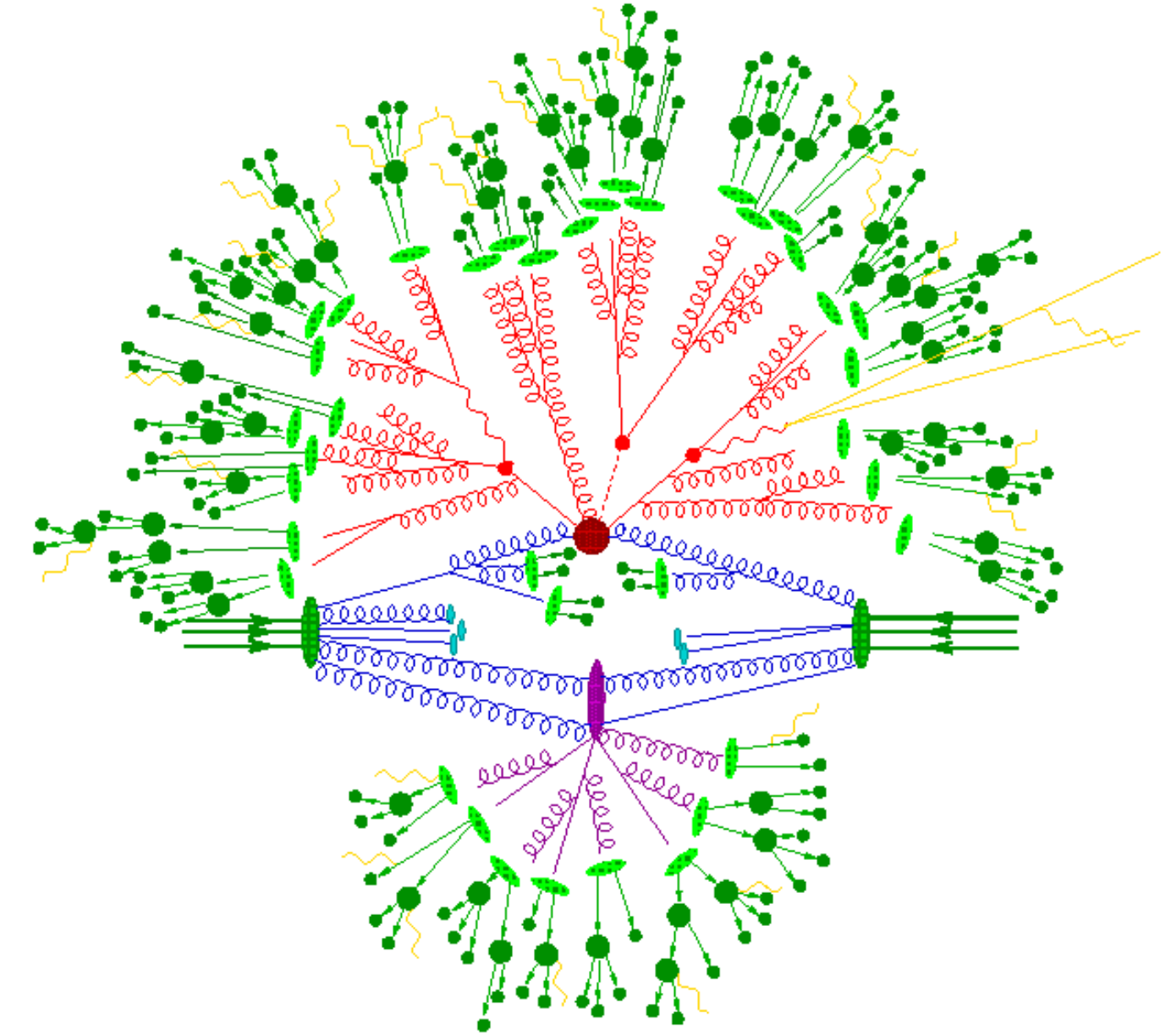
Calibration adjusted performance

- Model overestimates model performance in data compared to simulation
- Real performance in data still shows impressive improvements in performance over previous models
- Level of data-to-simulation agreement also comparable to older models

We're still extracting real physics from these more powerful models!



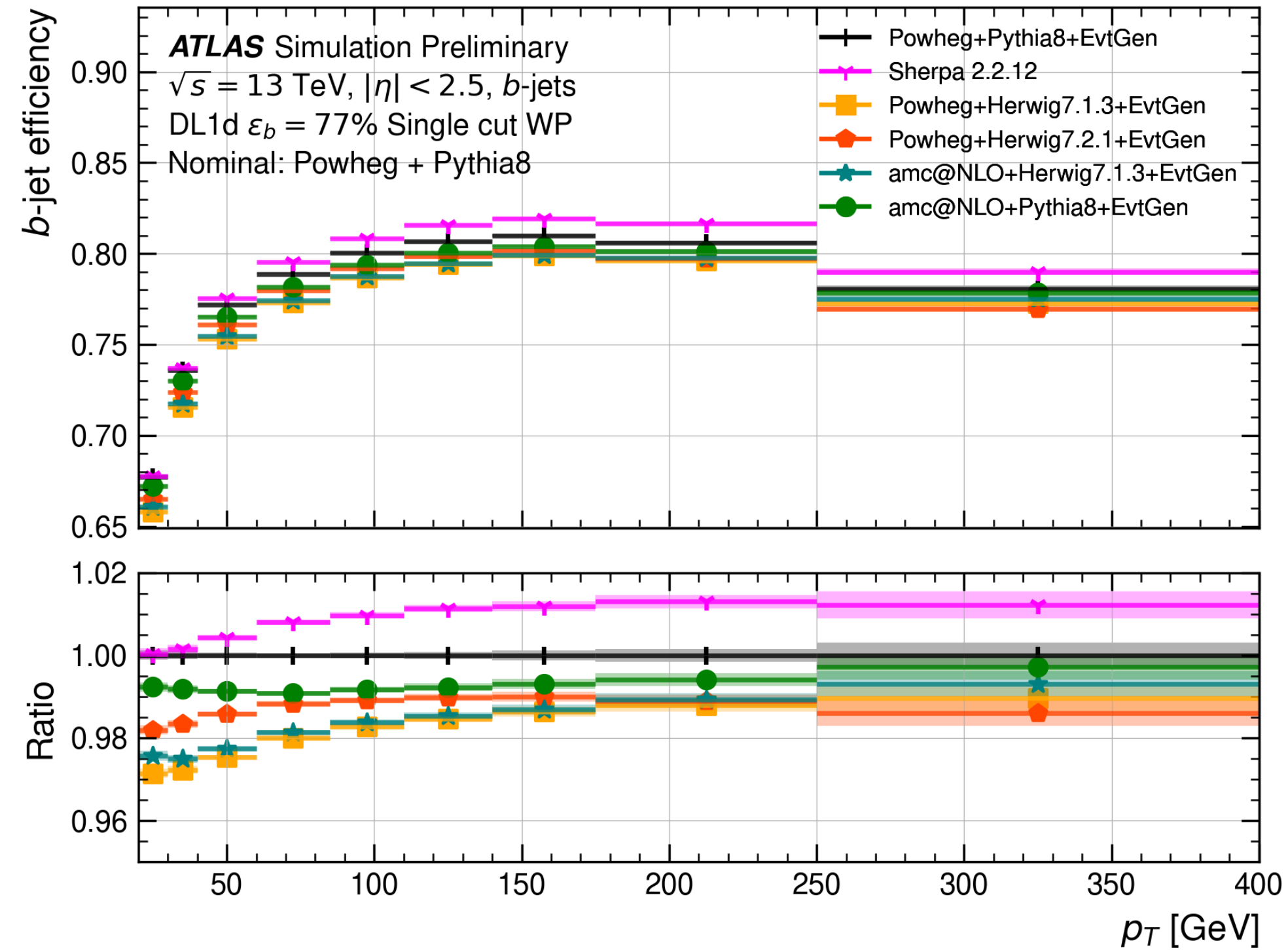
Generator Dependence



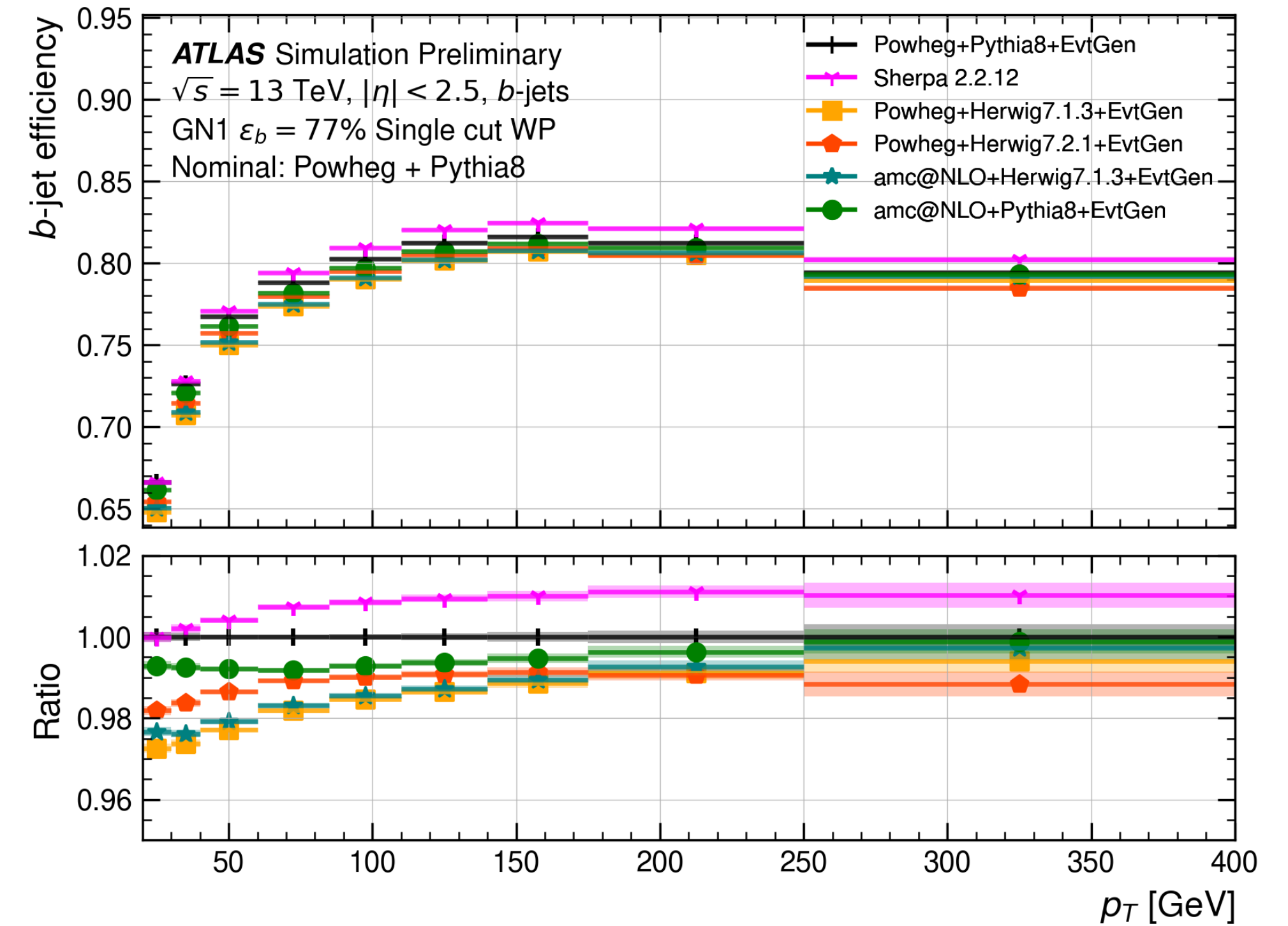
- Models are trained on “*best-knowledge*” samples:
 - Low- p_T $t\bar{t}$: **POWHEGBOX + PYTHIA8 + EVTGEN**
 - High- p_T Z' : **PYTHIA8 + EVTGEN**
- But no generator is perfect!
- Testing the model on other MC samples allows for an understanding of the generator dependence
- **Essential** to verify that the a more sophisticated model such as GN2 is not learning generator dependent information

Generator Dependence - b-jets

DL1d



GN1

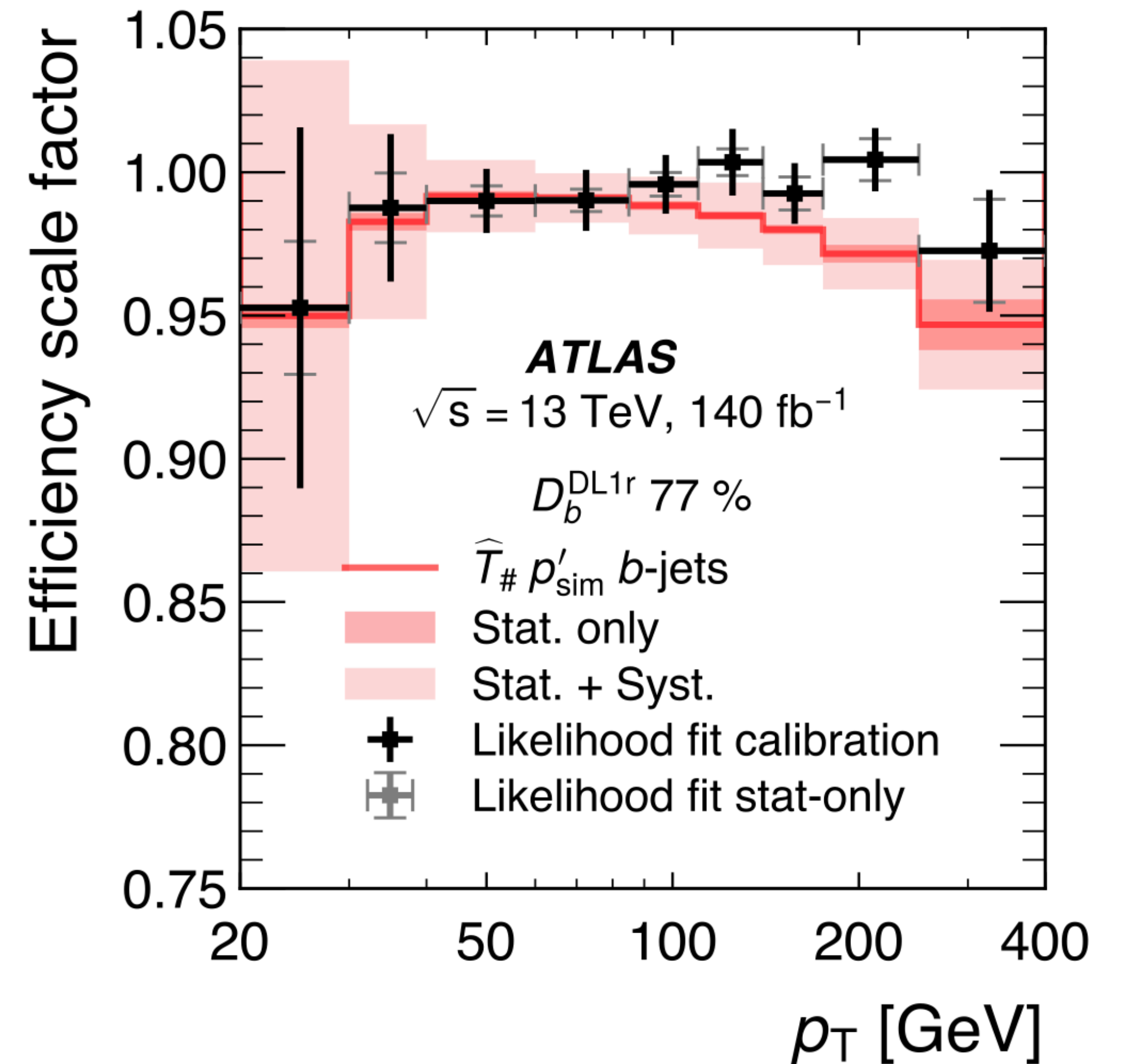


- Overall generator dependence: $O(3\%)$
- Variations between DL1d and GN1 are very similar
- Indicates that the more sophisticated model is not exploiting generator specific information

Continuous calibration

[CERN-EP-2025-026](#)

- Current calibration methods work using fixed **working points**, we only calibrate those specific efficiencies — for GN2 these are 65%, 70%, 77%, 85% and 90%
- Analysis specific working points or directly using the model predicted probabilities in downstream tasks are therefore not allowed
- New **optimal transport** approach learns to map the entire probability distribution between simulation and data
- **This is a fully continuous calibration**



Newest Developments*

*that are public

What comes next?

- Flavour tagging models traditionally considered three different types of jets
 - **B-jets**
 - **C-jets**
 - **Light-jets**
- GN2 also added the ability to identify **hadronic taus**
- The models also have only been using information from inner detector tracks

Can we do better?

What comes next?

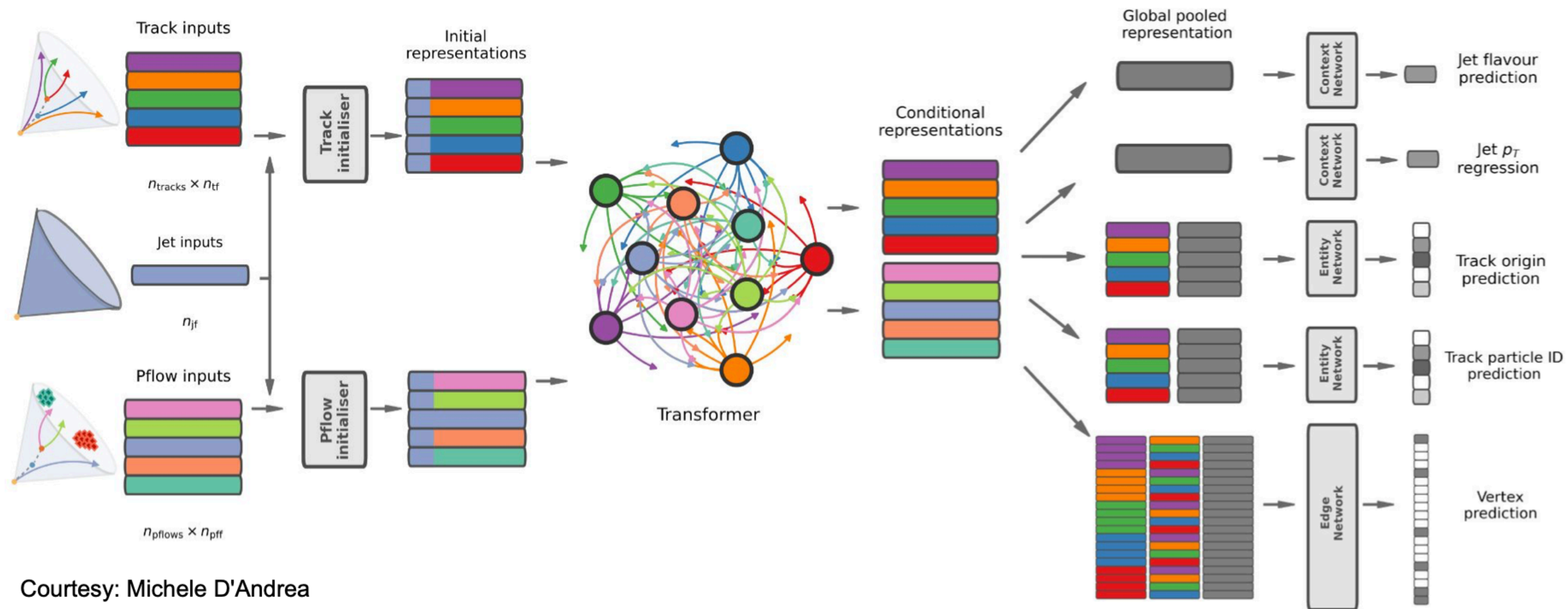
[FTAG-2025-01](#)

- Flavour tagging models traditionally considered three different types of jets
 - **B-jets**
 - **C-jets**
 - **Light-jets**
- GN2 also added the ability to identify **hadronic taus**
- The models also have only been using information from inner detector tracks
- **GN3** is the next generation flavour tagging model (creative naming scheme)
- GN3 extends our tagging capabilities to include **strange** and **gluon** jets
- We now include calorimeter information in the form of particle flow constituents
- Soft lepton information has been added to account for semi-leptonic decays
- Extending beyond just tagging capabilities — now including regression tasks

Can we do better?

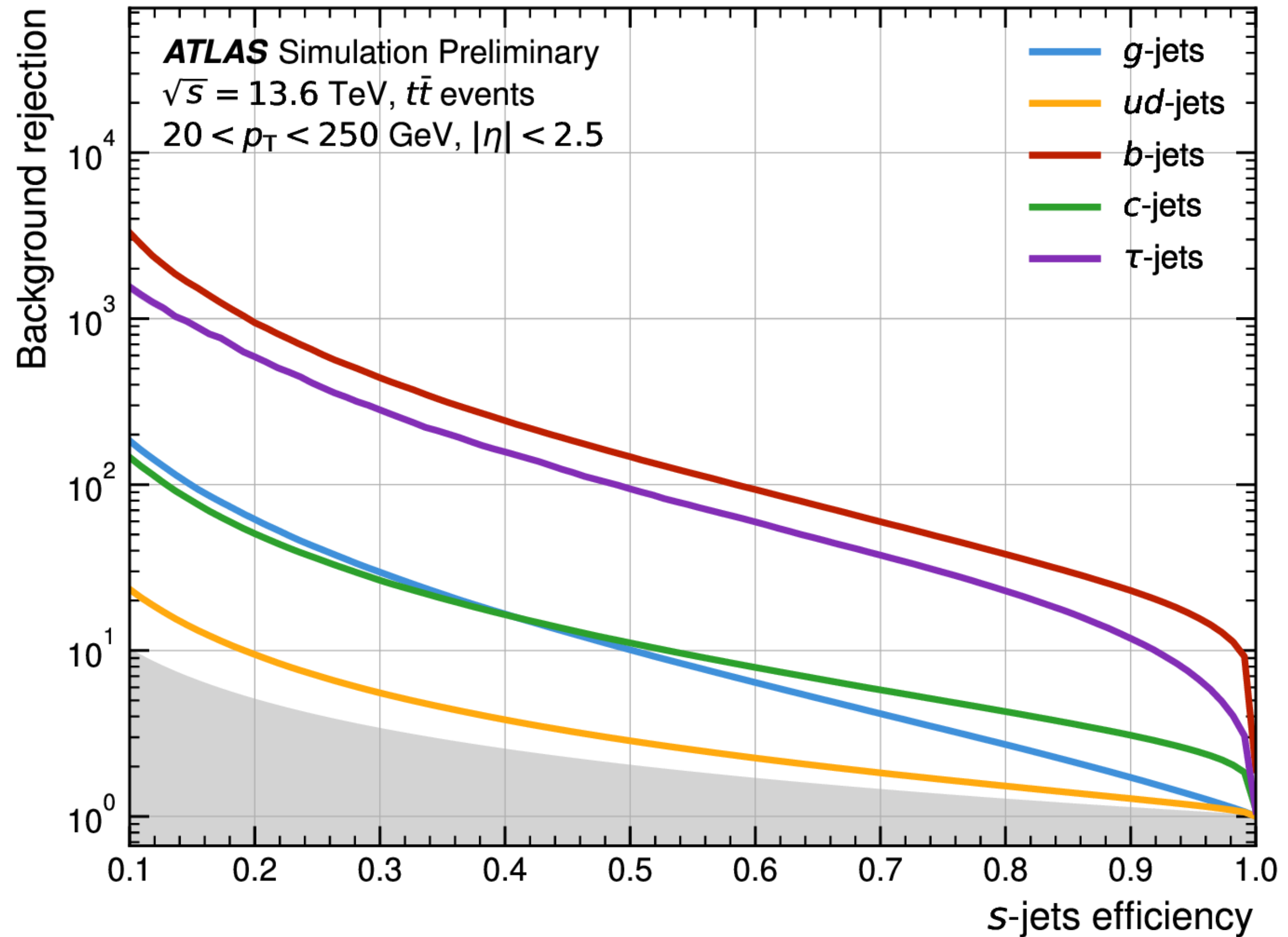
Yes

GN3 model



Strange + Gluon jets

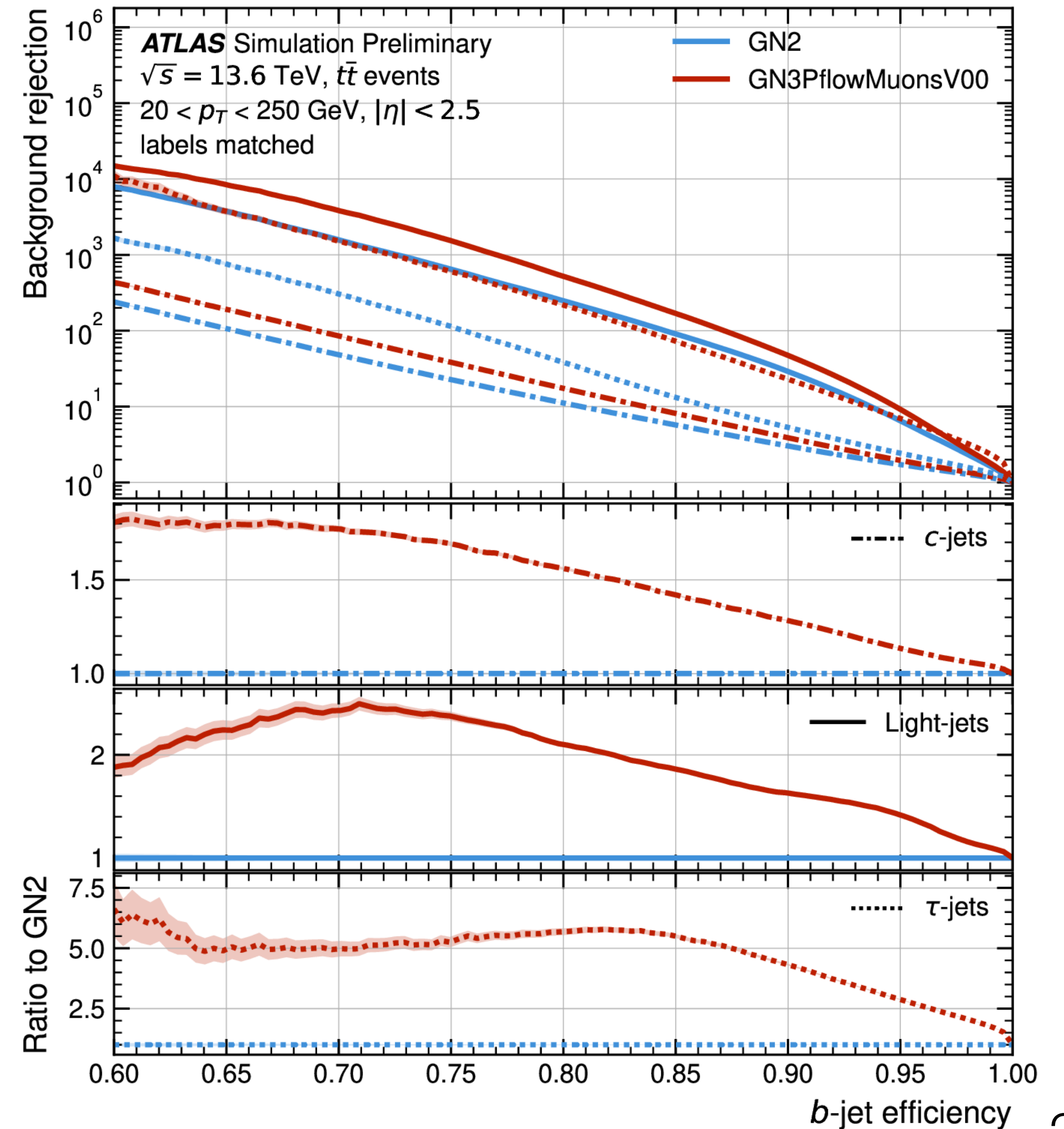
- There are existing quark vs gluon taggers in use within ATLAS but GN3 is the first to do both flavour tagging as well
- For the first time we can see separation power between strange and ud-jets



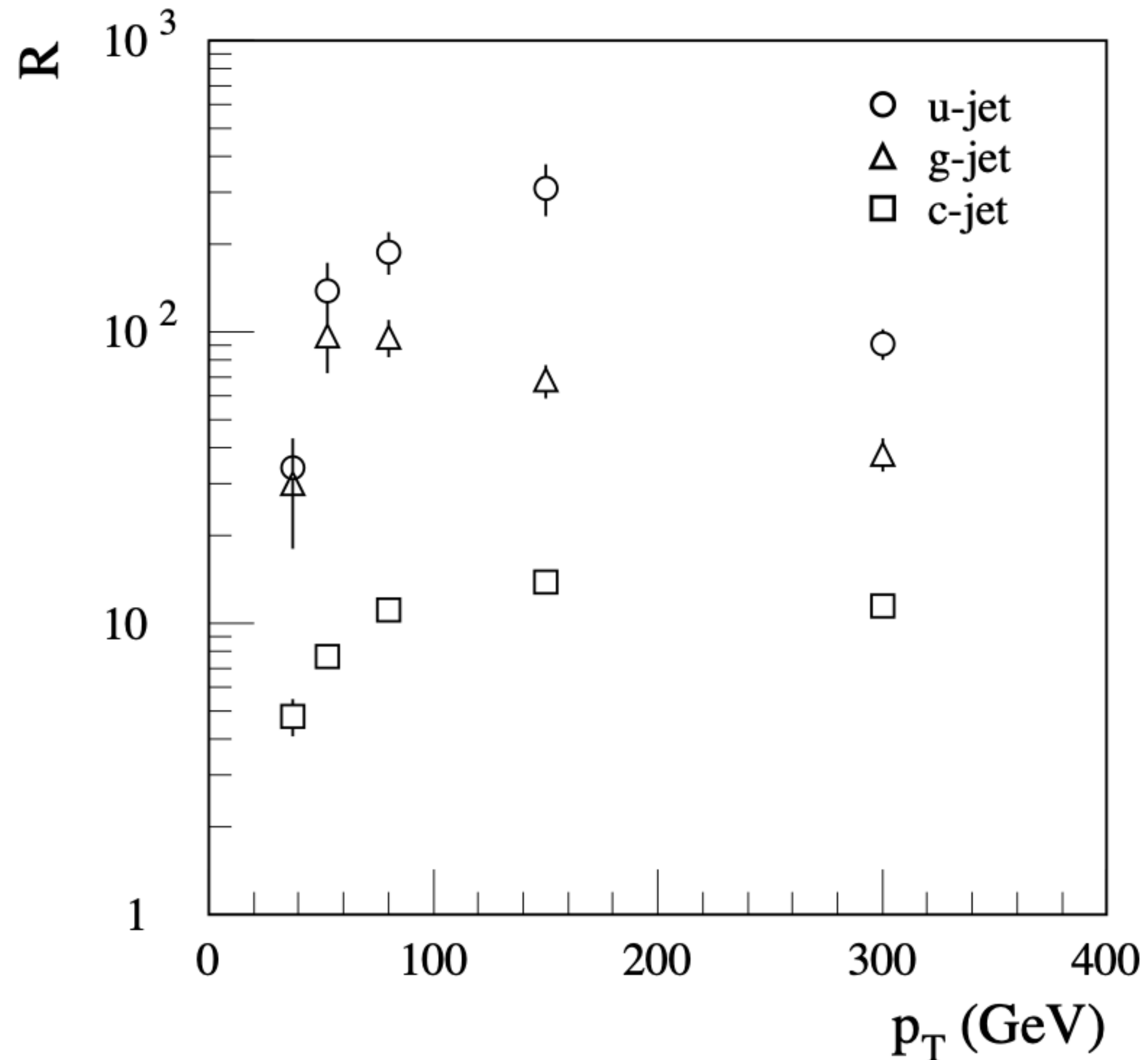
GN3 performance

GN3 is set to significantly outperform GN2, further doubling our light jet rejection

- But will this survive the calibration?
- Lots of new inputs, how well modelled are they?



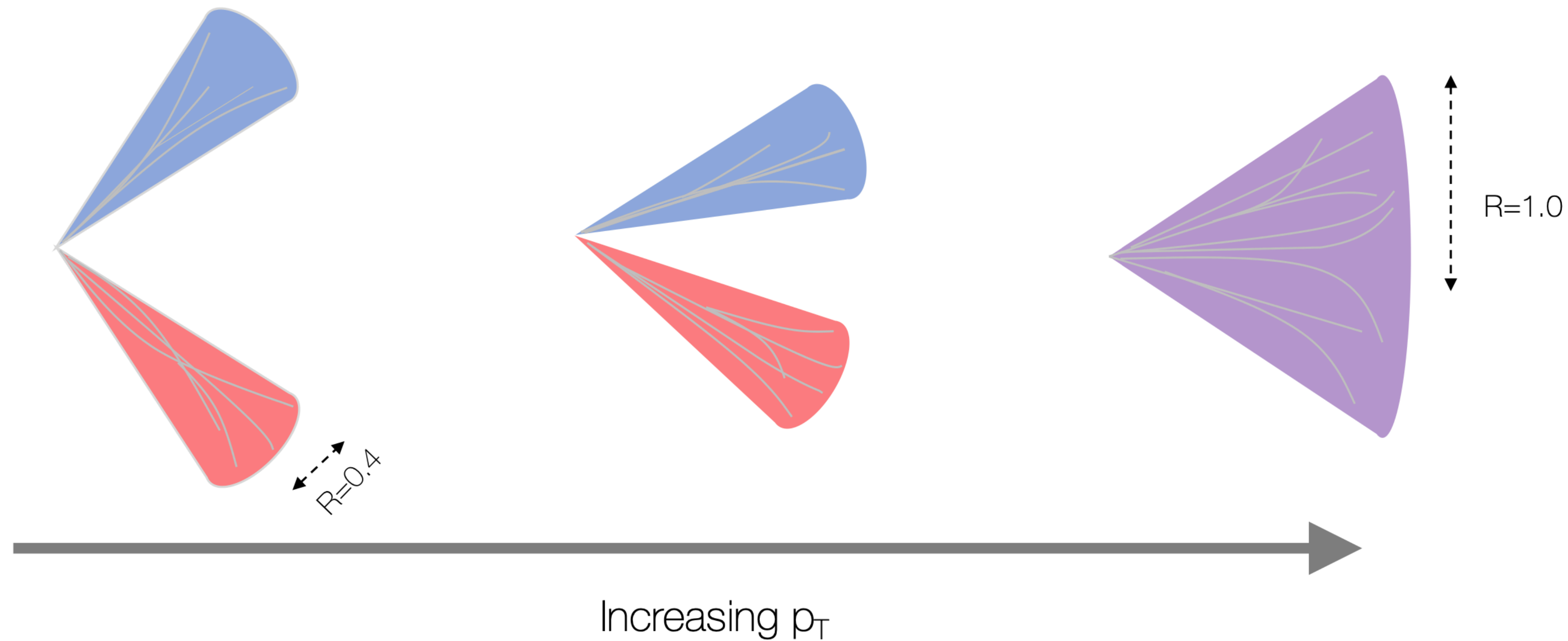
How far we've come



Expected flavour tagging
performance from ATLAS TDR at a
50% signal efficiency

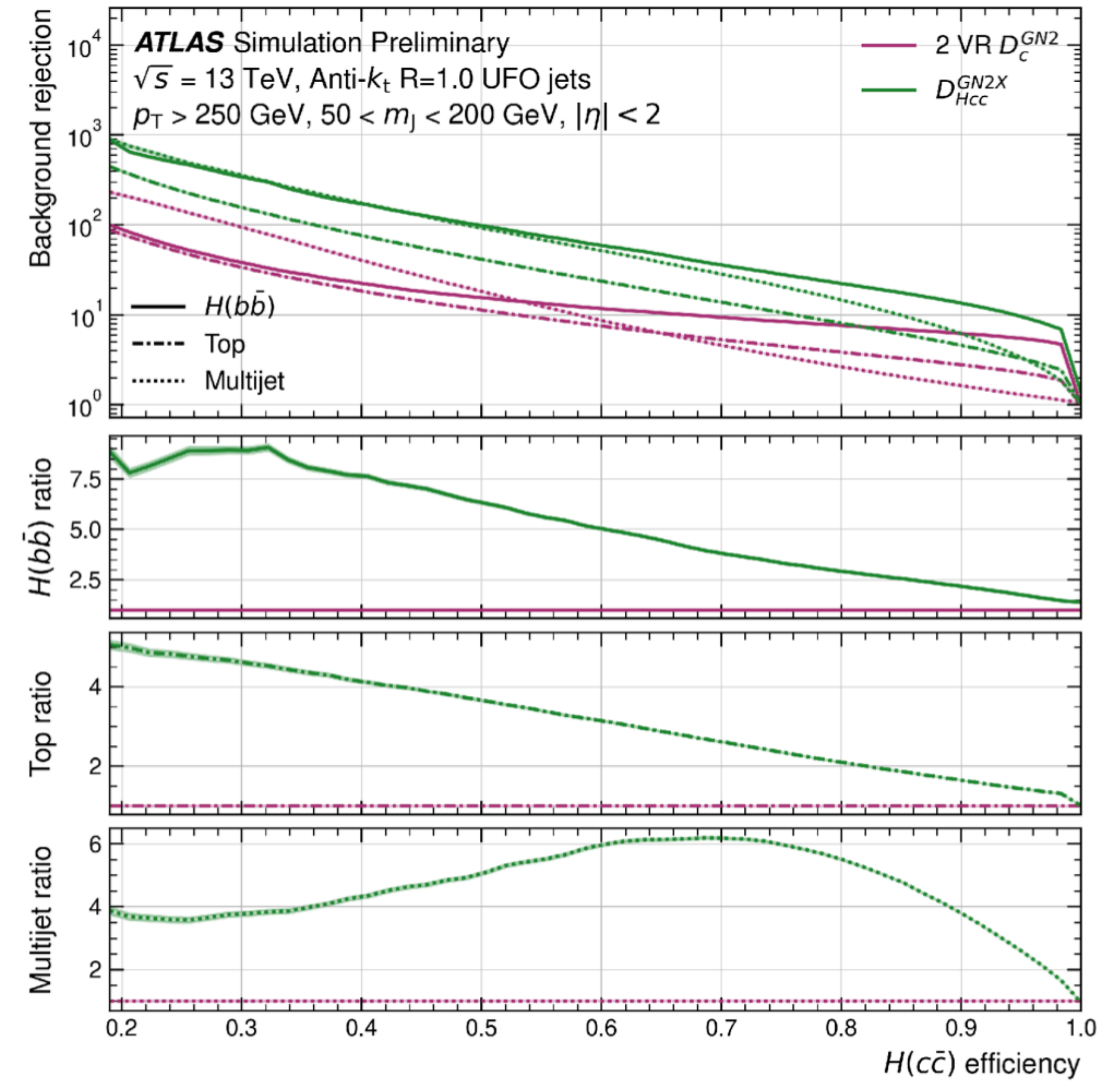
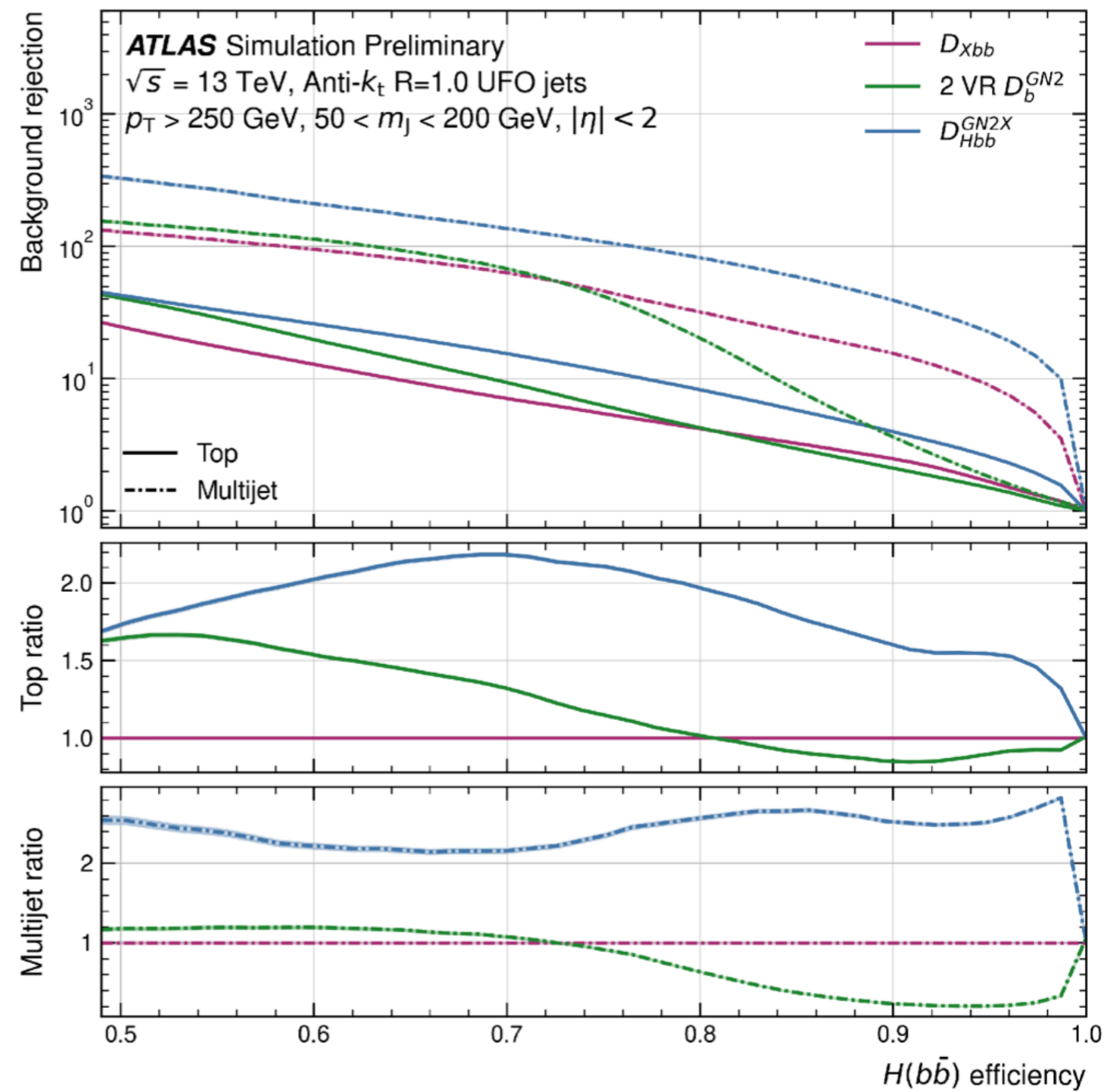
Beyond Flavour Tagging

Boosted Higgs



- In the high transverse momentum regime, the Higgs is boosted and so the decay products are collimated
- At some the two b-jets merge into a single large-radius jet

GN2X



- Tagging the large radius jet directly improves performance!
- Can identify both $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ (which still hasn't been observed)

HL-LHC

Compared with **GN1**: another step change in performance

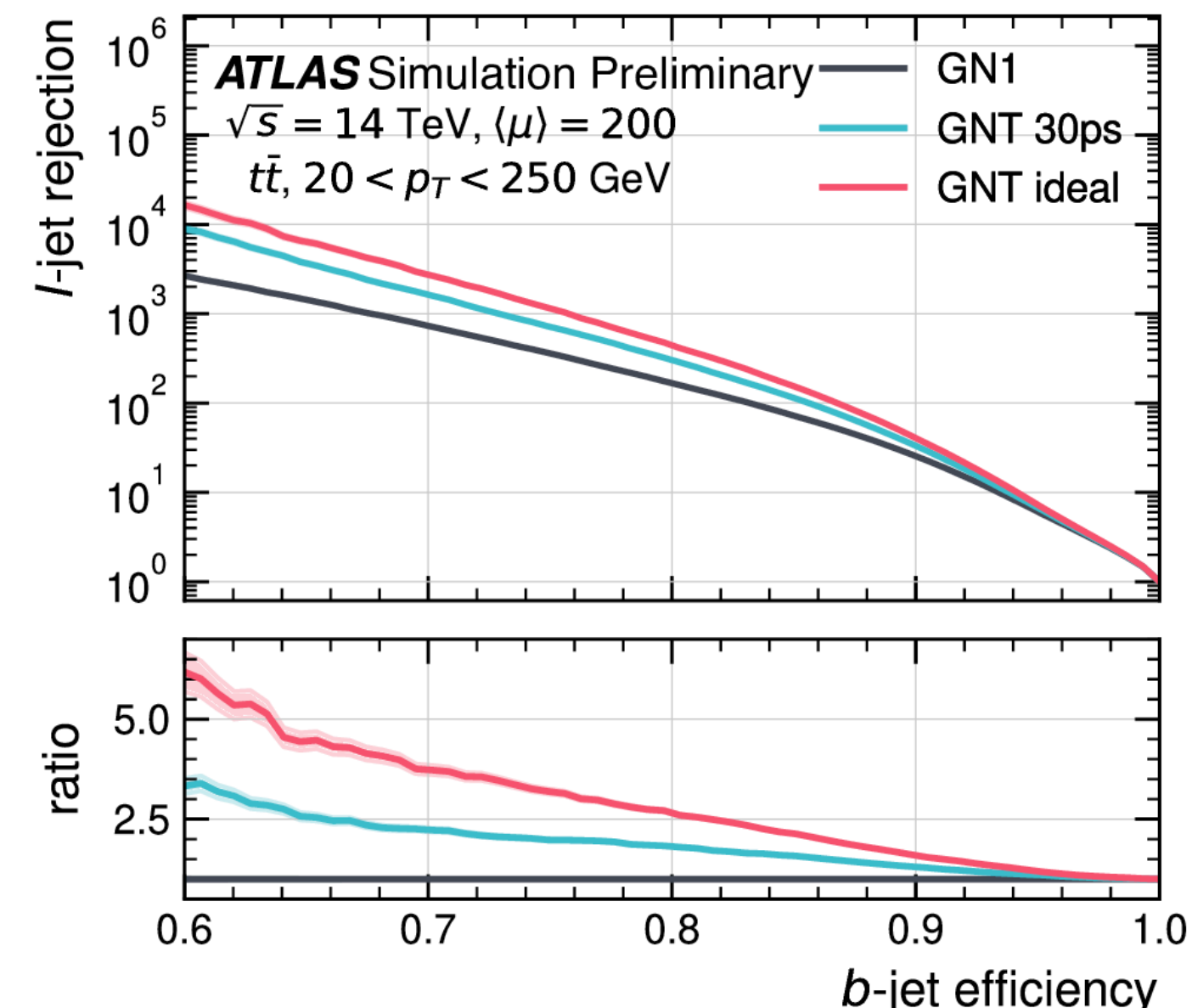
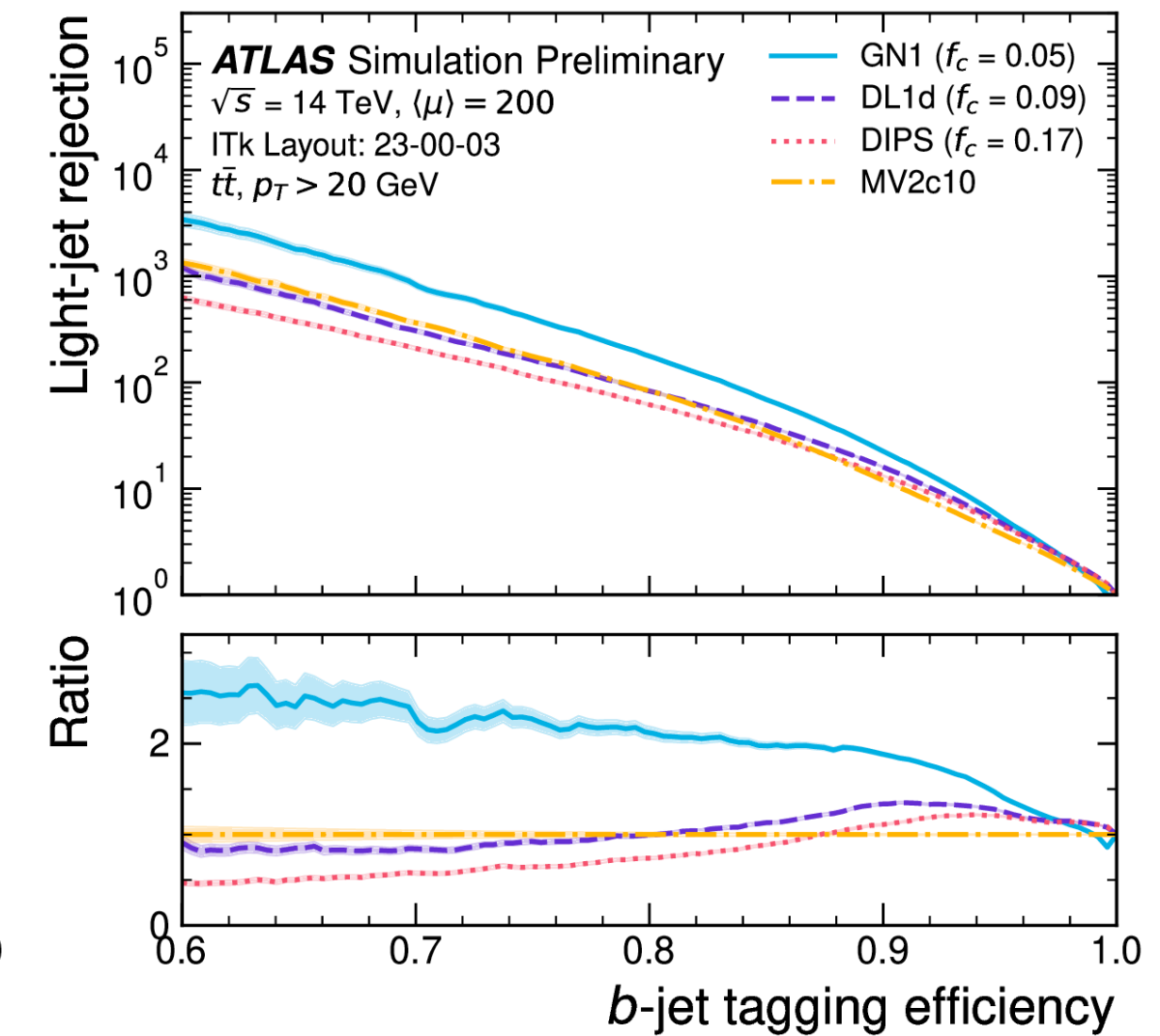
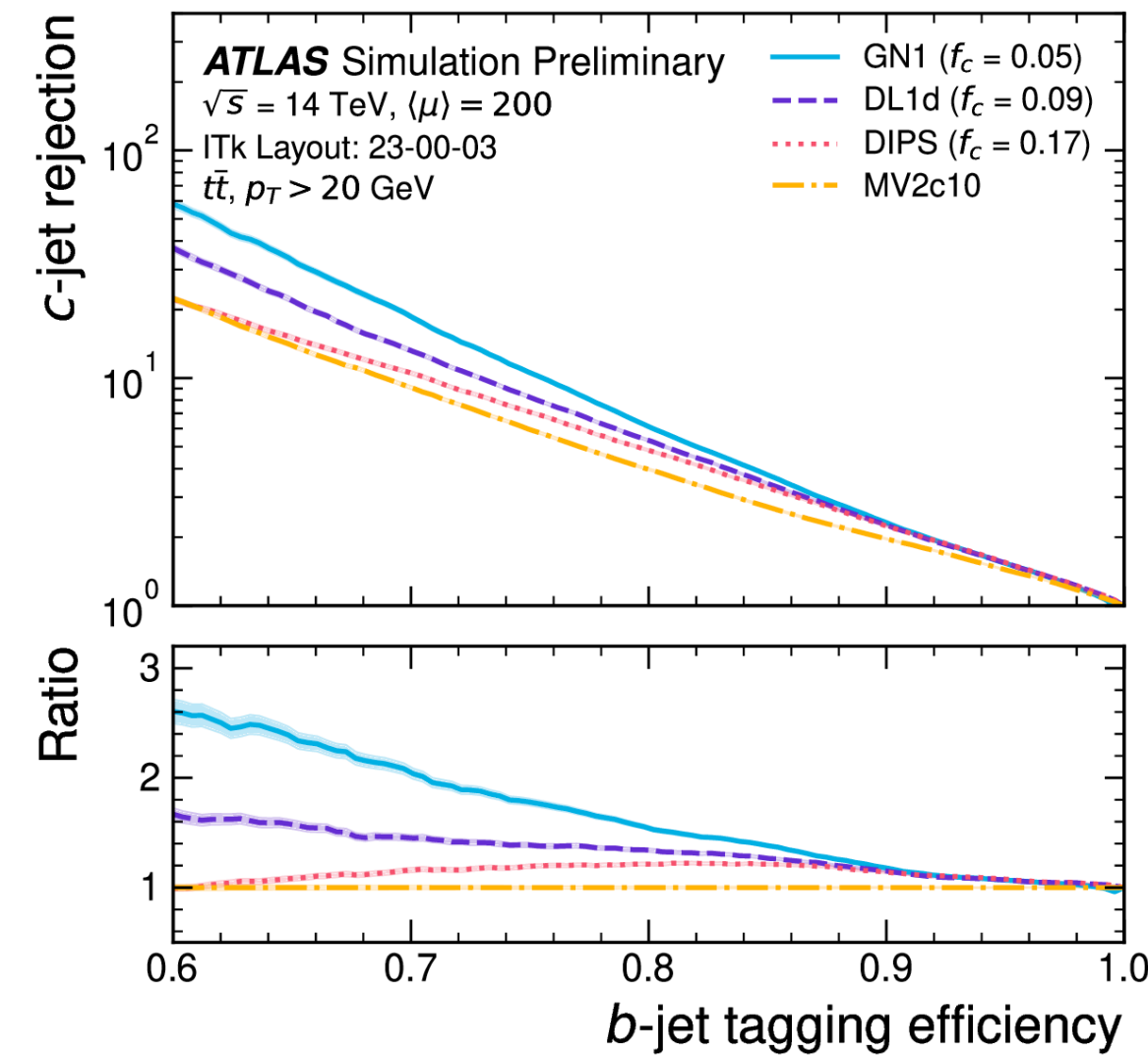
$t\bar{t}$ 70% WP

- **1.5x** c-rejection
- **2x** light-rejection

Z' (equivalent 20% WP)

- **1.75x** c-rejection
- **1.2x** light-rejection

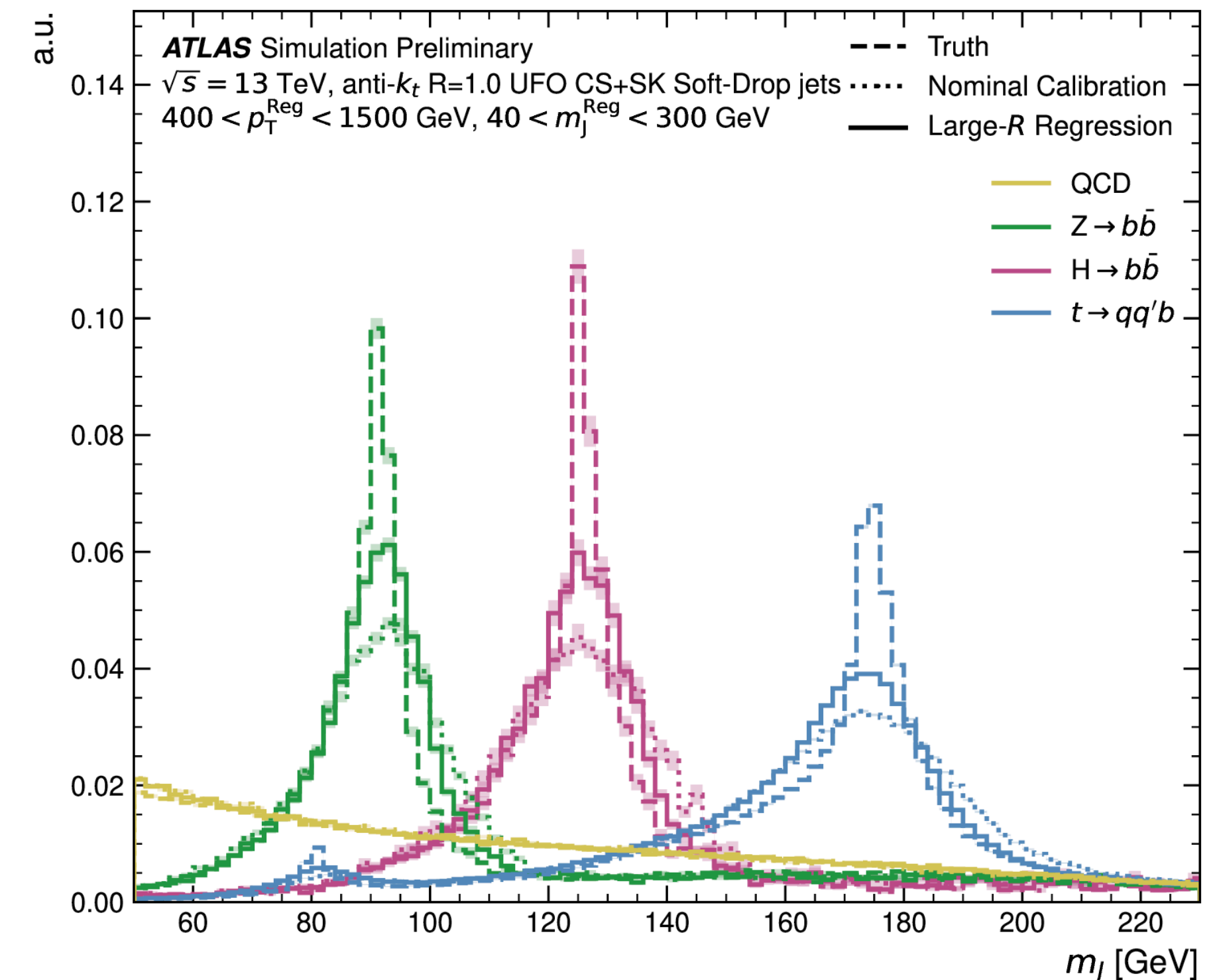
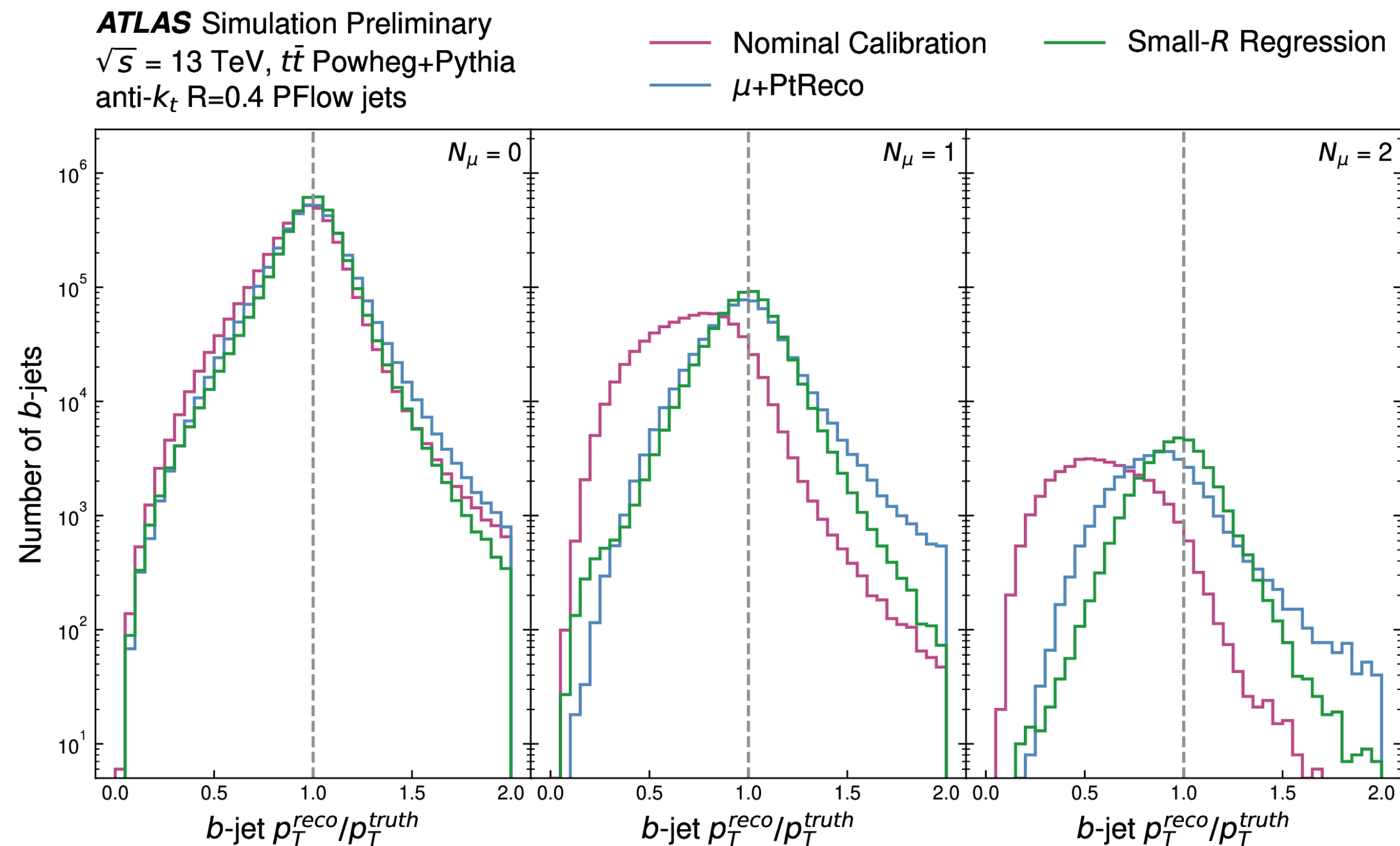
Note: ratio is to DL1d



Jet calibration

GN2-style models are also being used to calibrate jet kinematics $\mathcal{L} = |y - \hat{y}|$

- Exactly the same as the tagging model with a change of labels and loss function
- Paradigm shift! Historically state of the art energy measurement would have required a very different approach to tagging.



Building an *ecosystem*...

All these results are powered by the same technology.
A sufficiently powerful ML model can have wide applications!

Benefits of taking a ecosystem wide view:

- Increase collaboration & inclusivity
- Share tools and workflows
- Reduce barrier to entry
- Iterate quickly on best practices and advanced techniques
- Accelerate physics performance!



Summary

The 2024 Nobel Prize in Physics was awarded for foundational ML

- Physicists helped create modern ML/AI
- These tools are becoming ubiquitous at collider experiments
- In the near-ish future the reconstruction stack could be end-to-end ML

Should we be worried?

- Do we need to understand what's inside the box?
- Will the value of physics understanding decrease over time?
- Will we lose our intuition?
- How will this affect progress?

In the meantime

- These methods have brought significant gains to our physics programme and are pushing the limits of what we thought could be achieved at the LHC
- So strap in and enjoy the ride?