Nikhef

Dutch Physical Society (NNV) – Subatomic division meeting in Lunteren – November 3rd, 2023

Advanced Machine Learning at the Service of Identifying Boosted Higgs bosons with ATLAS

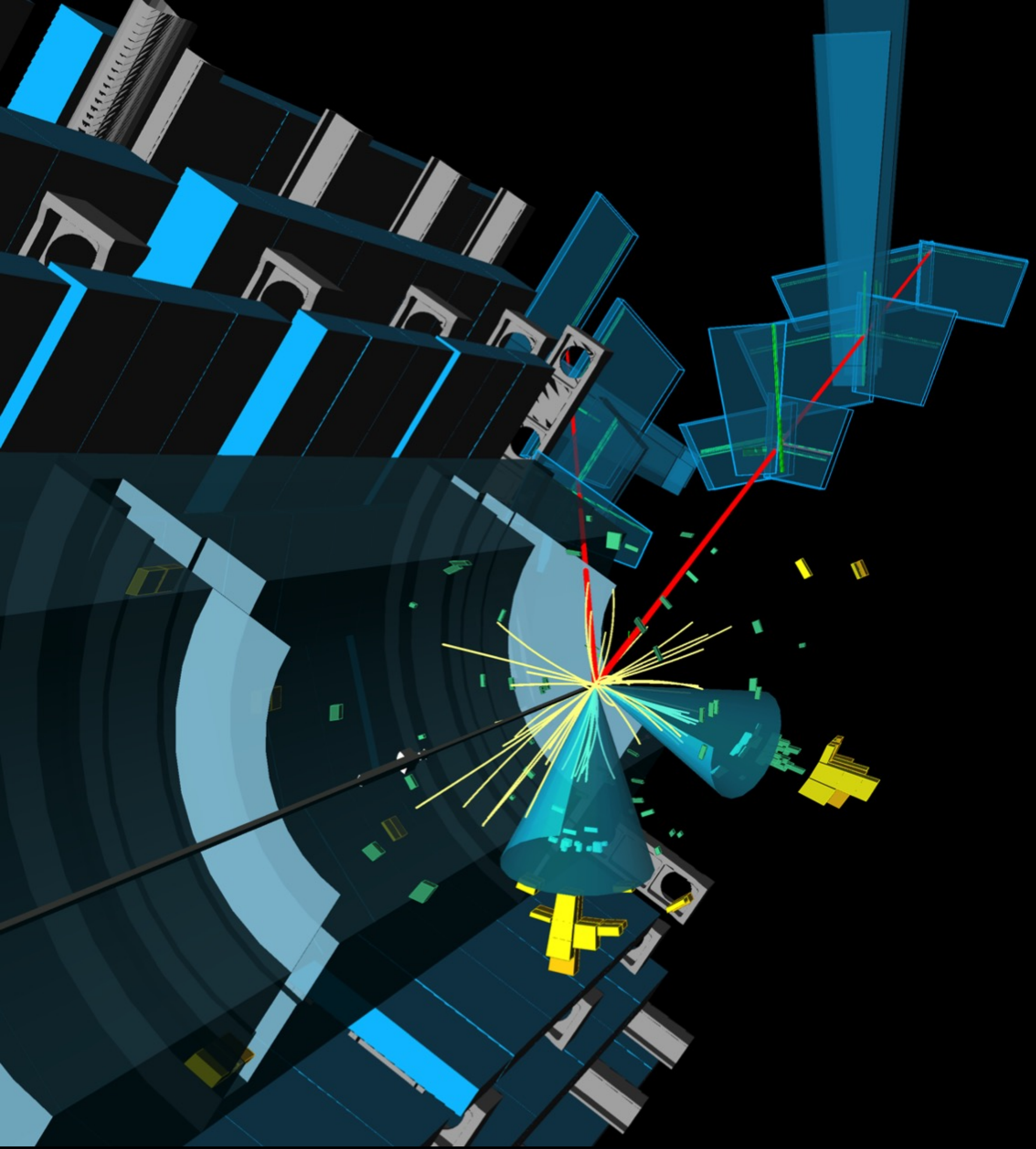
Geoffrey GILLES

Nikhef – Dutch National Institute for Subatomic Physics

ATLAS EXPERIMENT

NWO

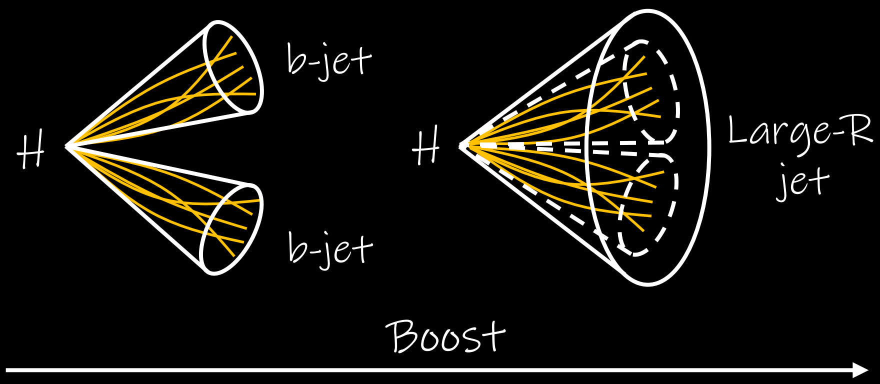
INTRODUCTION

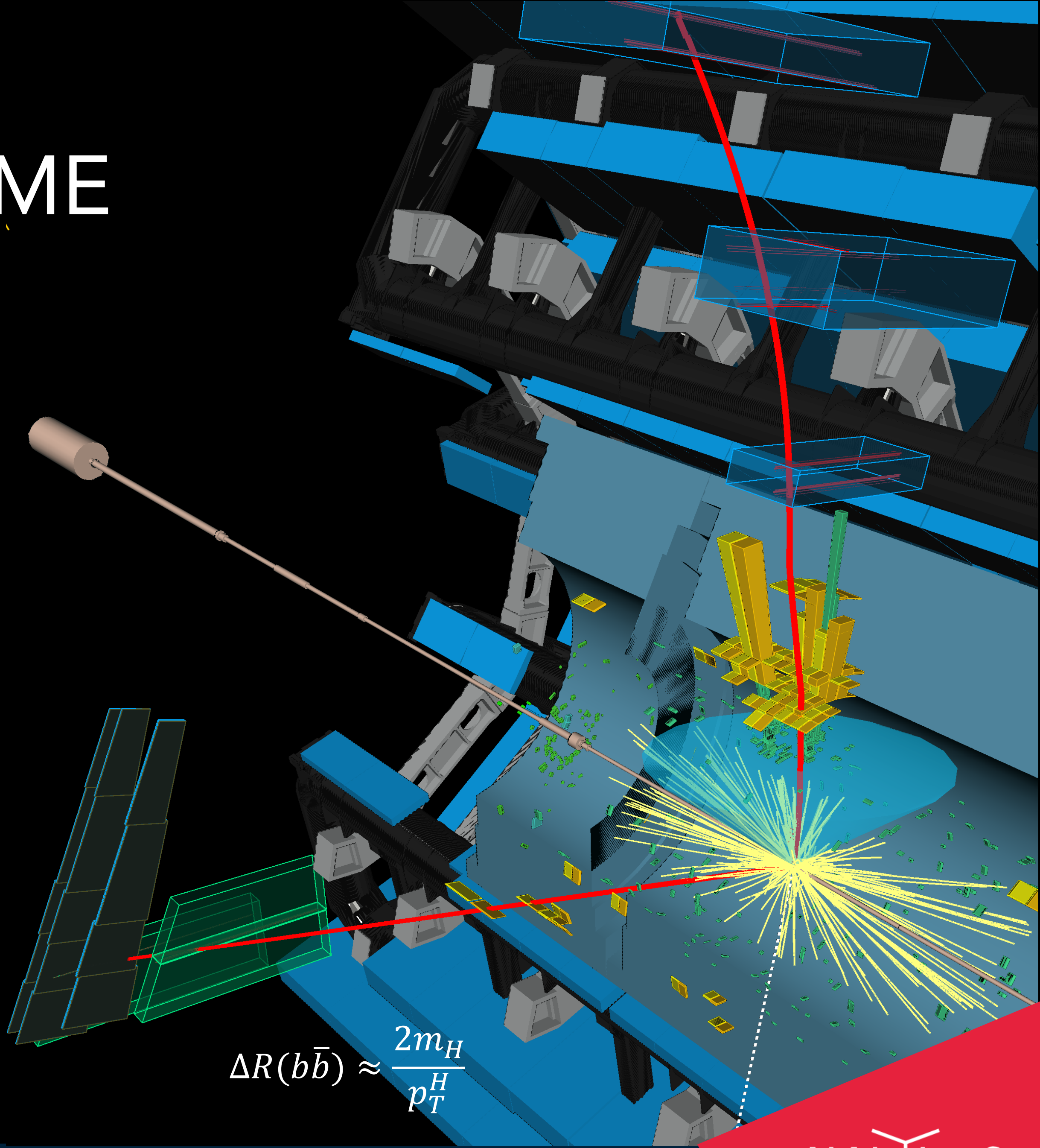


- Since its discovery, Higgs boson properties measured with precision
- Studying heavy-flavour decays remains central
- Observation of decay into b-quark pairs provided direct evidence of Yukawa coupling
- Challenging channel $H \to c\bar{c}$ still to be observed

THE HIGH-ENERGY REGIME

- Expected sensitivity to New Physics
- Helping to probe rare processes
- Highly boosted decay products become collimated, requiring dedicated tagging techniques





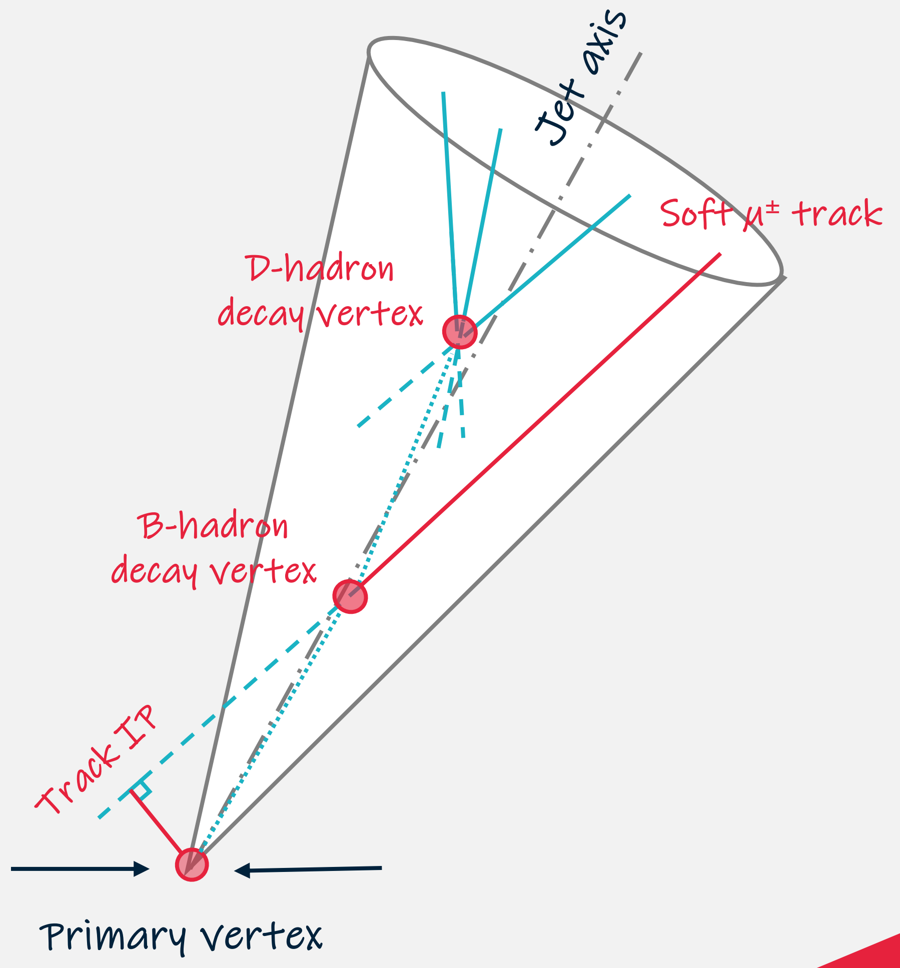
$$\Delta R(b\bar{b}) \approx \frac{2m_H}{p_T^H}$$

FLAVOUR-TAGGING

A KEY INGREDIENT

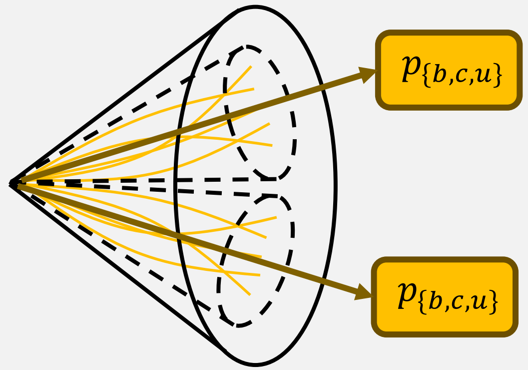
- Identifying heavy-flavour jets in inclusive approach
  - Discriminating b, c and light-flavour jets
- A key ingredient to a majority of physics analyses
  - From precision measurements to direct searches
  - To increase signal efficiency and background rejection
- Exploit specific topology of heavy flavour-jets
  - Lifetime, high-mass & decay multiplicity of B/D-hadrons
  - Using several complementary algorithms (c.f. backup)

Anatomy of a b-jet

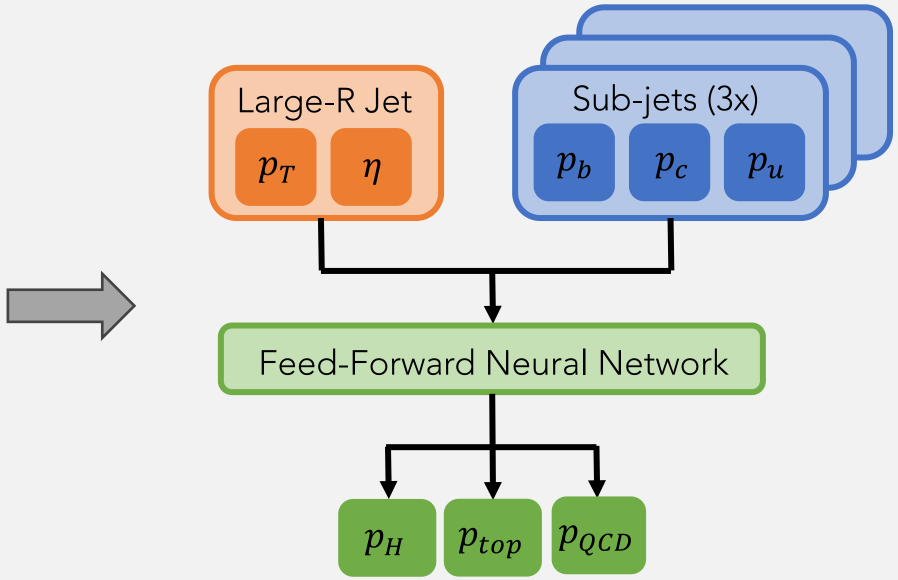


HISTORICAL ATLAS TAGGING STRATEGIES

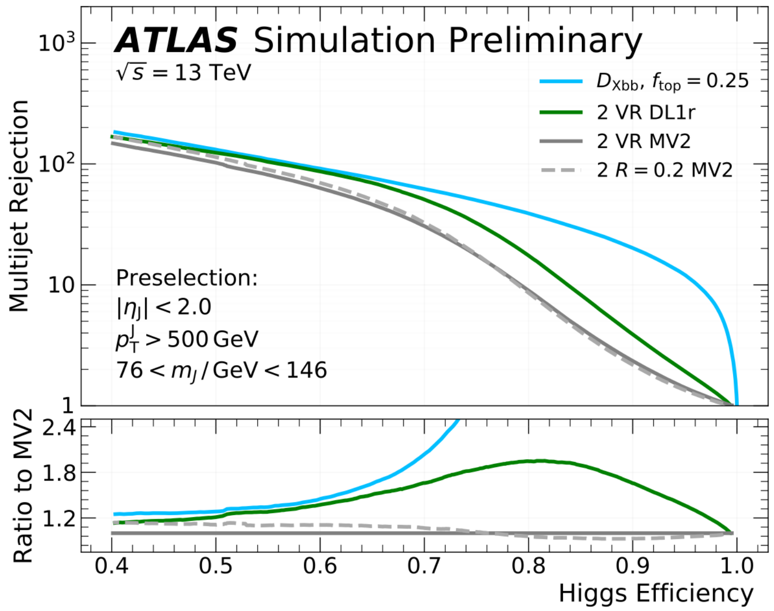
1st Generation
Flavour-tag sub-jet from large-R jet candidates



2nd Generation
Combined sub-jet tagging information and jet kinematics in NN



ATLAS-CONF-2021-051



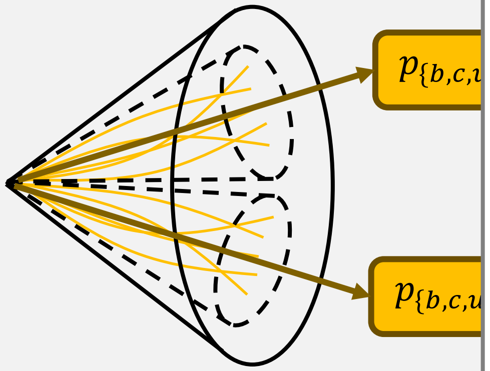
- Worked fine, but somehow blind to the entire picture of the jet activity

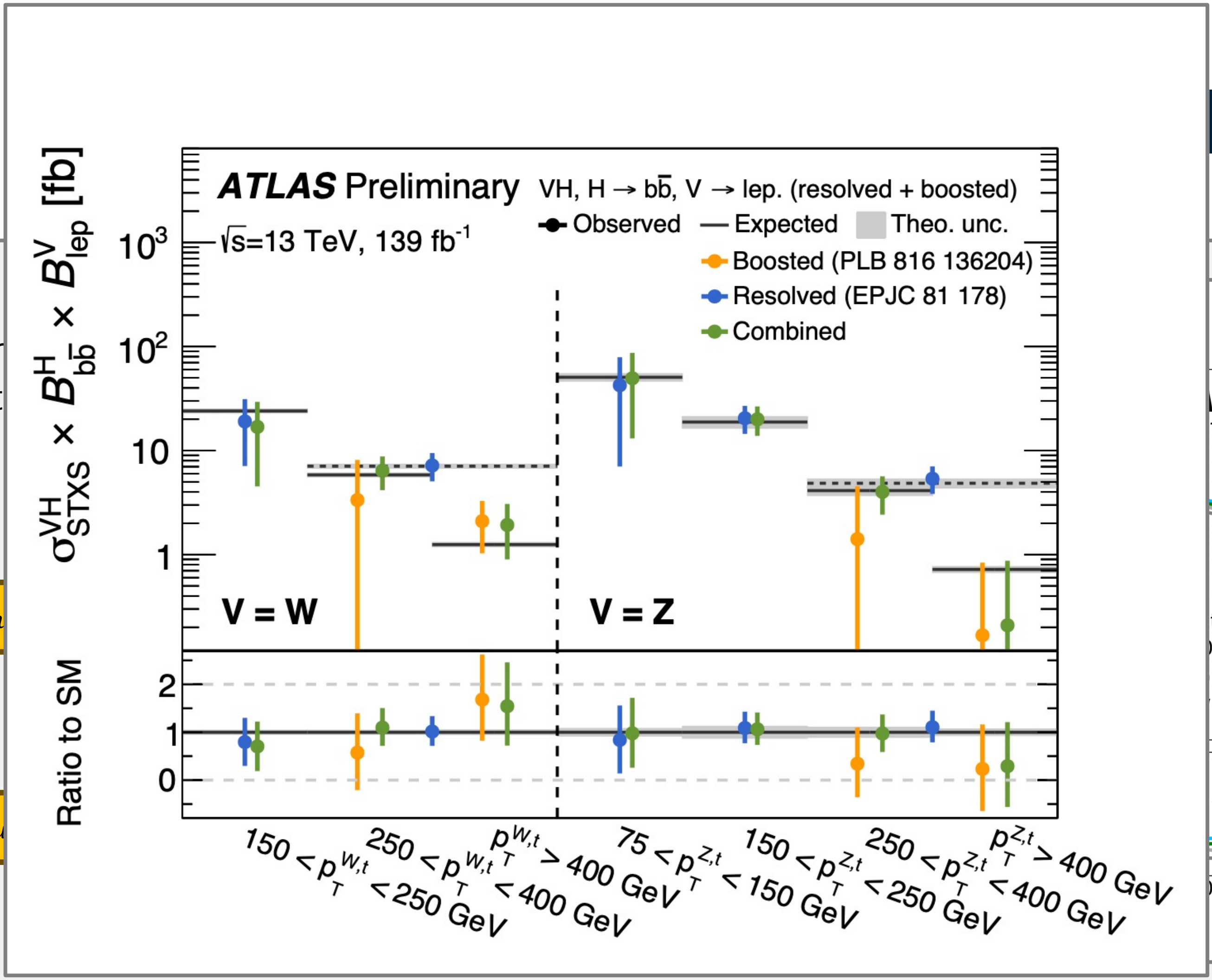
HISTORI EGIES

1st Generation

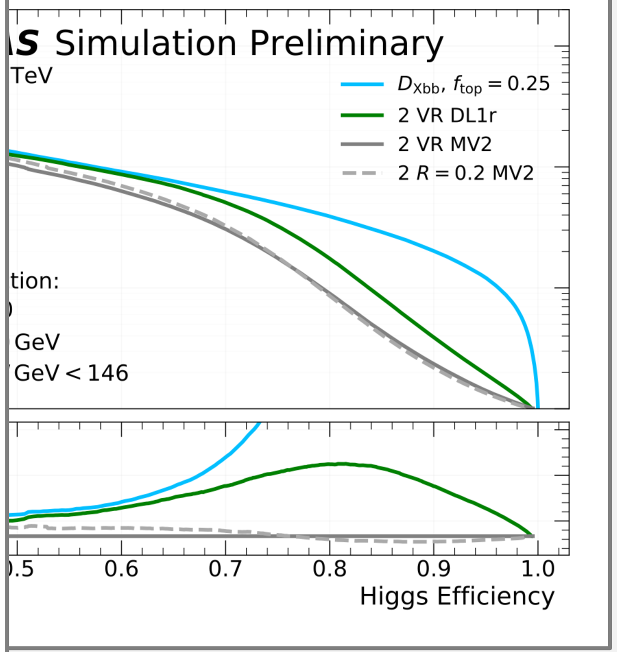
Flavour-tag sub-jet fr

large-R jet candidat





ATLAS-CONF-2021-051



- Worked fine, but somehow blind to the entire picture of the jet activity

THE TRANSFORMER EXPLOSION

Towards a 3rd Generation of Tagging Algorithm

- Recent progress in Machine Learning expanded realm of possibilities
  - Considering new generation of tagging algorithms
  - Learning from low-level information, avoiding intermediate reconstruction algorithms
  - ATLAS Combined performance groups at the forefront of adopting such technologies (e.g. flavour-tagging, tracking, etc.)
- Transformer introduced by a team at Google Brain in 2017 [link]
  - Emerged as a dominant paradigm in Machine Learning across various applications
  - Ability to model complex relationships and deliver outstanding performance

Understanding Transformer with a Social Network Analogy

Axiom

People with similar characteristics or interests tend to interact and influence each other more …

… one can imagine individuals paying attention to others based on their similarities

→ Attention mechanism

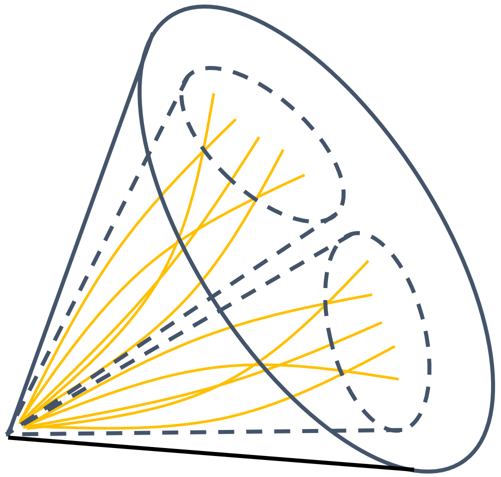
… individual have multiple perspectives or sources of influence

→ Multi-head attention concept

Now, let's map the analogy to large-R jets

1
Imagine each track as an individual in an "large-R jet" network

2
Their kinematics represent their unique characteristics or attributes



3
Tracks similarities measured using dot-products, indicating feature alignment or correlation
→ Attention mechanism

4
Runs parallel attention operations, attending to input parts differently
→ Multi-head attention concept

Nikhef

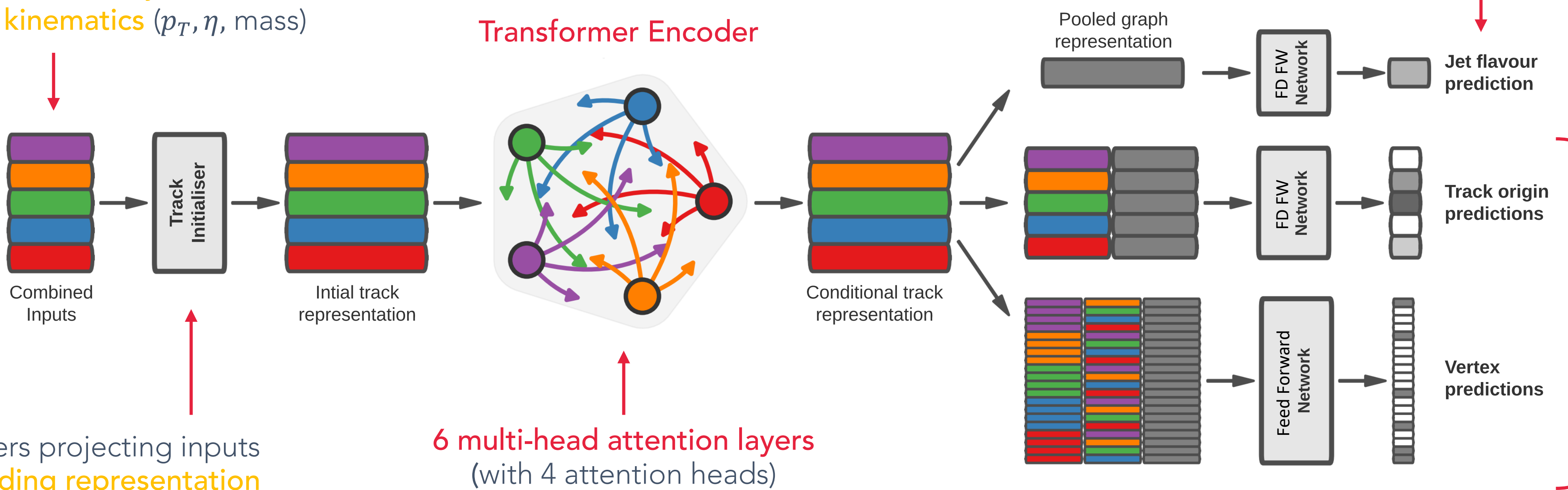
ATL-PHYS-PUB-2023-021

GN2X ARCHITECTURE

Sequence of 100 tracks / jet (~20 kinematics) and jet kinematics (p_T, η, mass)

Transformer Encoder

Jet classification as primary task



Auxiliary tasks

Dense layers projecting inputs into embedding representation

6 multi-head attention layers (with 4 attention heads)
Using scale dot-product attention

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

FOCUS ON JET CLASSIFICATION

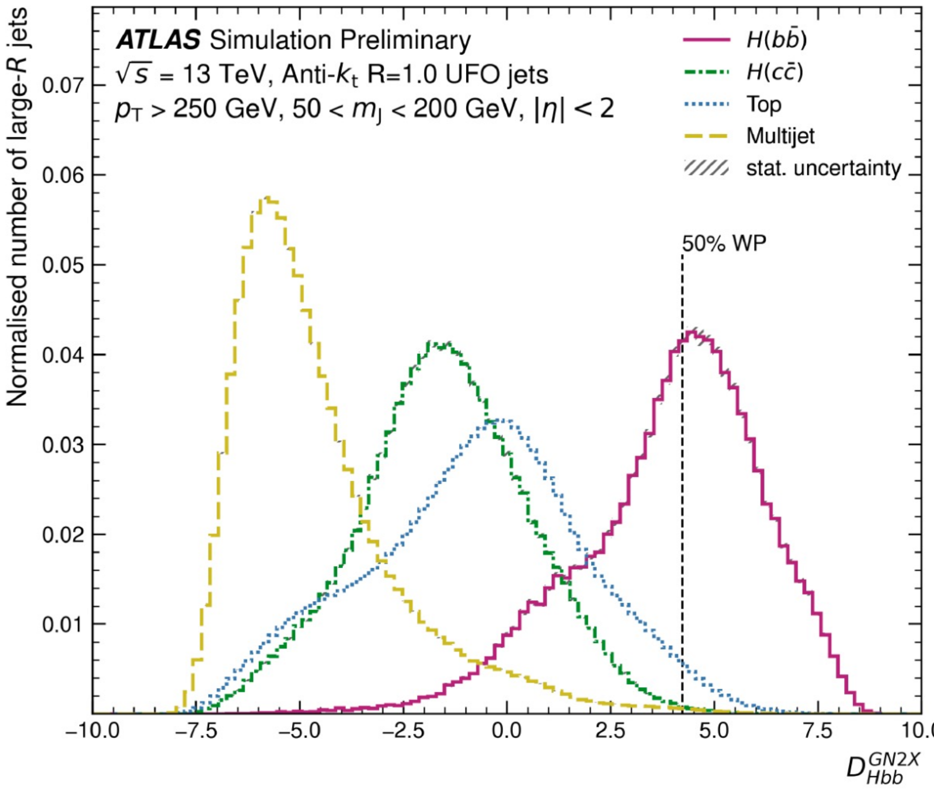
- Multi-class classification trained on 62 million jets
  - Generate probability scores for identifying $H \to b\bar{b}$, $H \to c\bar{c}$, Top and Multi-jet (QCD)

- When assessing tagging efficiencies, probability score combined into discriminant

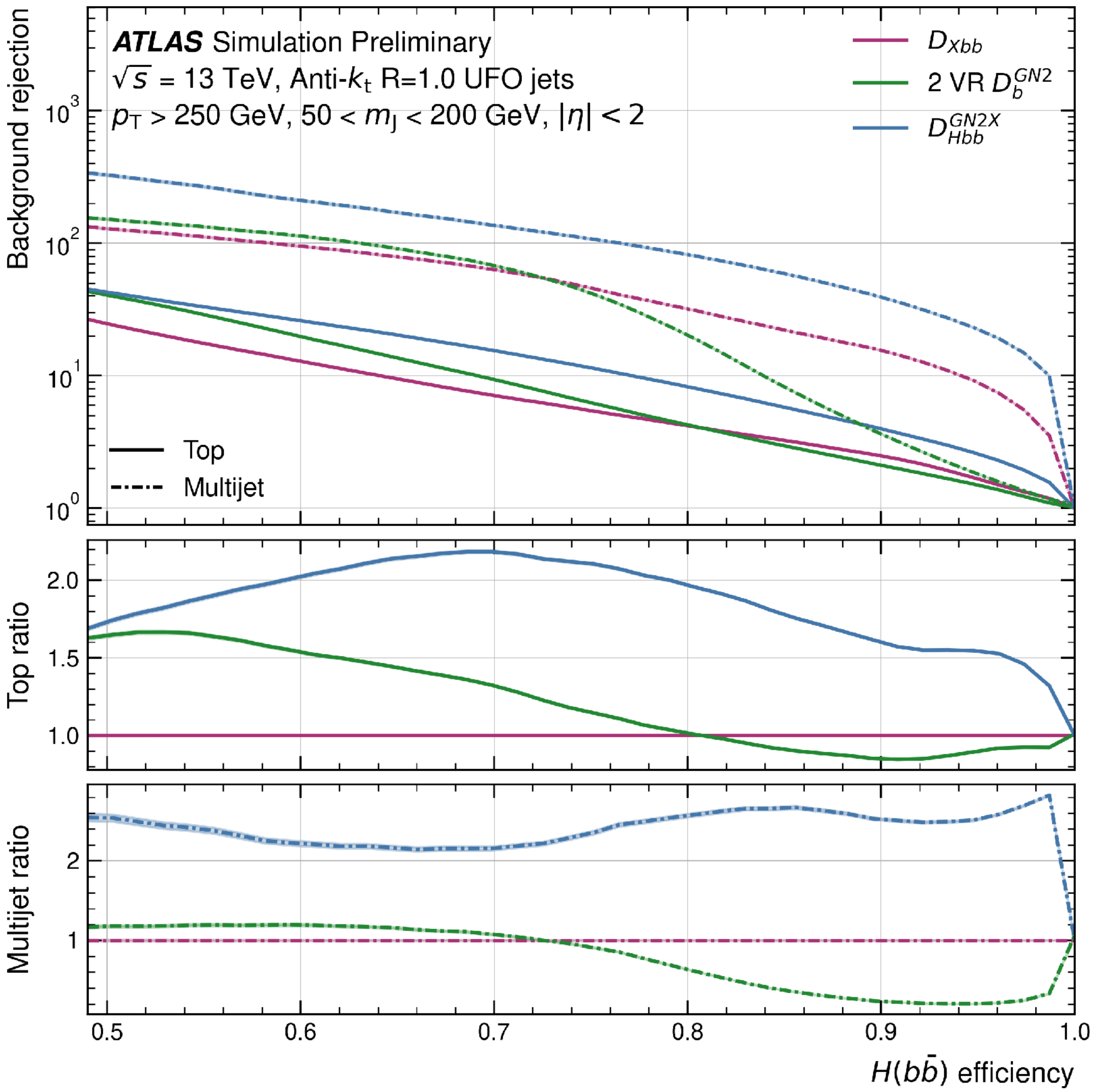
$$D_{\mathrm{Hbb}}^{\mathrm{GN2X}} = \ln\left(\frac{p_{\mathrm{Hbb}}}{\boxed{f_{\mathrm{Hcc}}} \cdot p_{\mathrm{Hcc}} + \boxed{f_{\mathrm{top}}} \cdot p_{\mathrm{top}} + (1 - f_{\mathrm{Hcc}} - f_{\mathrm{top}}) \cdot p_{\mathrm{QCD}}}\right)$$

Free parameters to control trade-off
among background rejections
(e.g. f_{Hcc} = 0.02, f_{top} = 0.25)

ATL-PHYS-PUB-2023-021



ATL-PHYS-PUB-2023-021



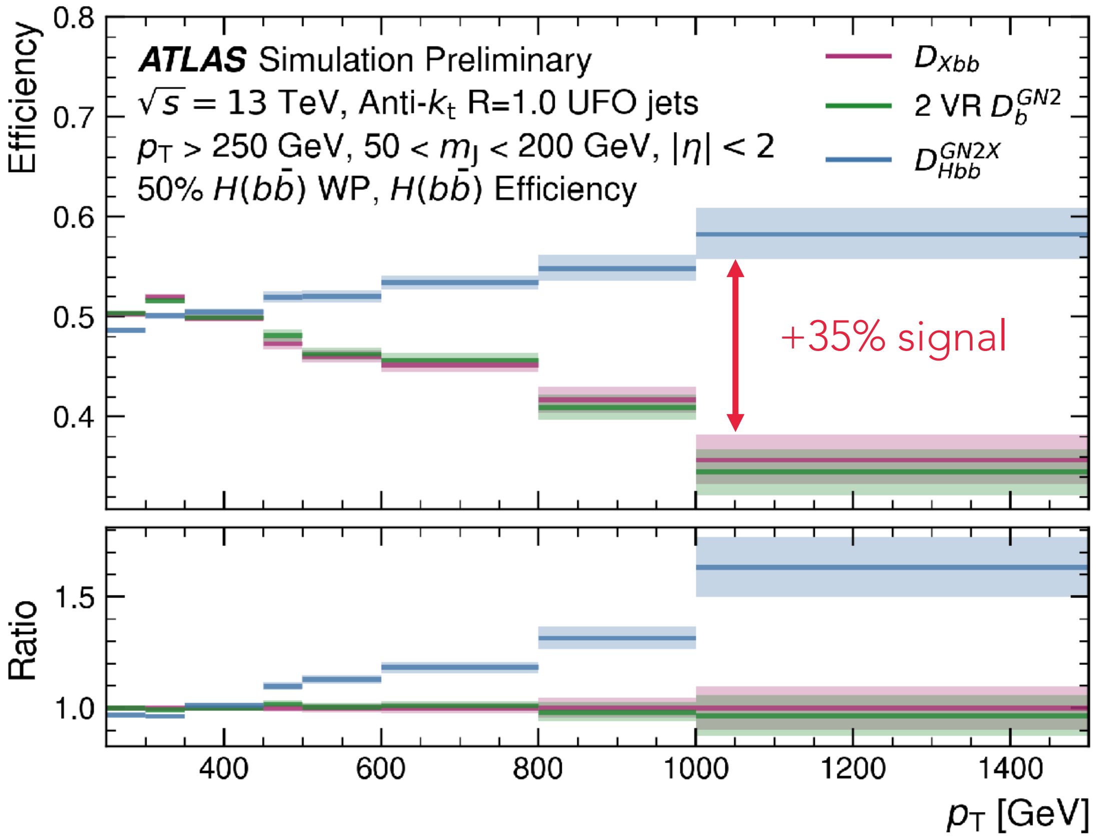
TAGGING PERFORMANCE

- Significantly outperformed previous models
  - 1.6x and 2.5x increase in top and multi-jet rejections respectively at 50% $H \to b\bar{b}$ efficiency

Nikhef

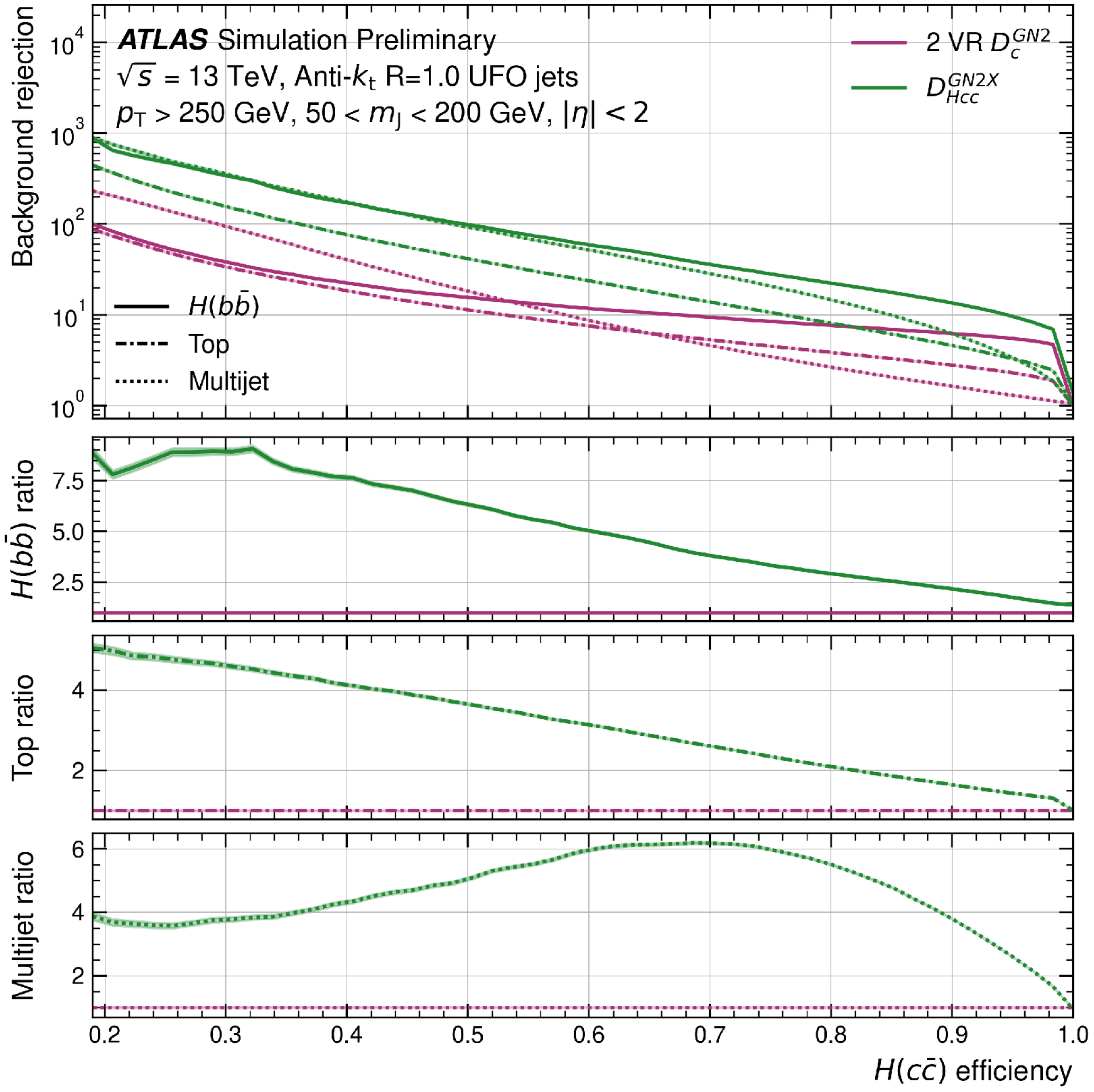
TAGGING PERFORMANCE

ATL-PHYS-PUB-2023-021



- Remarkable stable efficiency as a function of jet p_T

ATL-PHYS-PUB-2023-021



TAGGING PERFORMANCE

- $H \rightarrow c\bar{c}$ tagging by redefining discriminant

$$D_{\text{Hcc}}^{\text{GN2X}} = \ln\left(\frac{p_{\text{Hcc}}}{f_{Hbb} \cdot p_{\text{Hbb}} + f_{\text{top}} \cdot p_{\text{top}} + \left(1 - f_{\text{Hbb}} - f_{\text{top}}\right) \cdot p_{\text{QCD}}}\right)$$

e.g. with $f_{Hbb} = 0.3$, $f_{top} = 0.25$

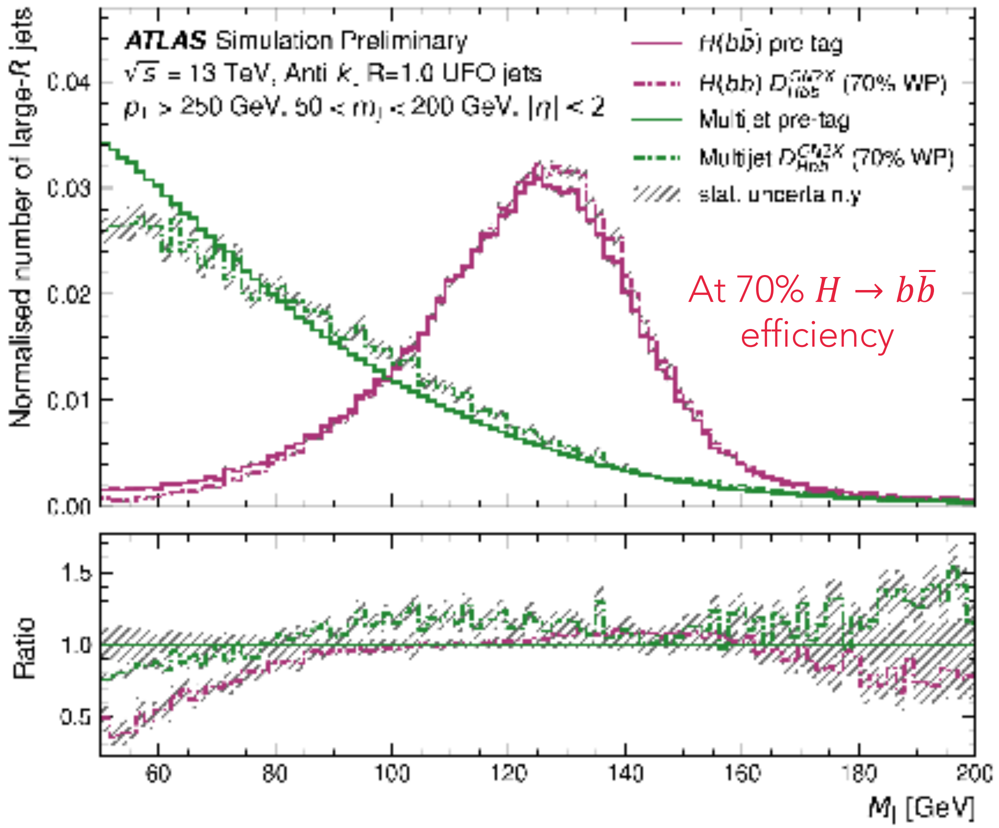
- Significantly outperformed previous models
  - 3x, 5x, 6x increase in top, multi-jet and $H \rightarrow b\bar{b}$ rejections respectively at 50% $H \rightarrow c\bar{c}$ efficiency

Nikhef

PRESERVING LARGE-R JET MASS

- Key variable to discriminate signal & background
  - Shapes should remain distinct after applying $H \to b\bar{b}/c\bar{c}$ tagging in analysis
- Tagging perf. to be decorrelated from jet kinematics
  - Training on "flat-mass" Higgs-boson samples
  - Resampling in p_T, η and mass
- Mass sculpting kept under control
  - Residual dependences under investigations

ATL-PHYS-PUB-2023-021



CONCLUSIONS

- Studying boosted Higgs-boson decays to heavy-flavours constitutes important aspect of ATLAS physics programme

- High-performance algorithm for tagging such Higgs-boson decays can play a crucial role in
  - Improving sensitivity of searches for New Physics
  - Precise measurements of the Higgs boson properties.

- ATLAS recently harnessed new cutting-edge tagging algorithm based on Transformer
  - Demonstrating remarkable performance improvements - ATL-PHYS-PUB-2023-021

- Algorithm being deployed in Physics analyses
  - Precise calibration now subject to all attention
  - New exciting results are yet to come, stay tuned

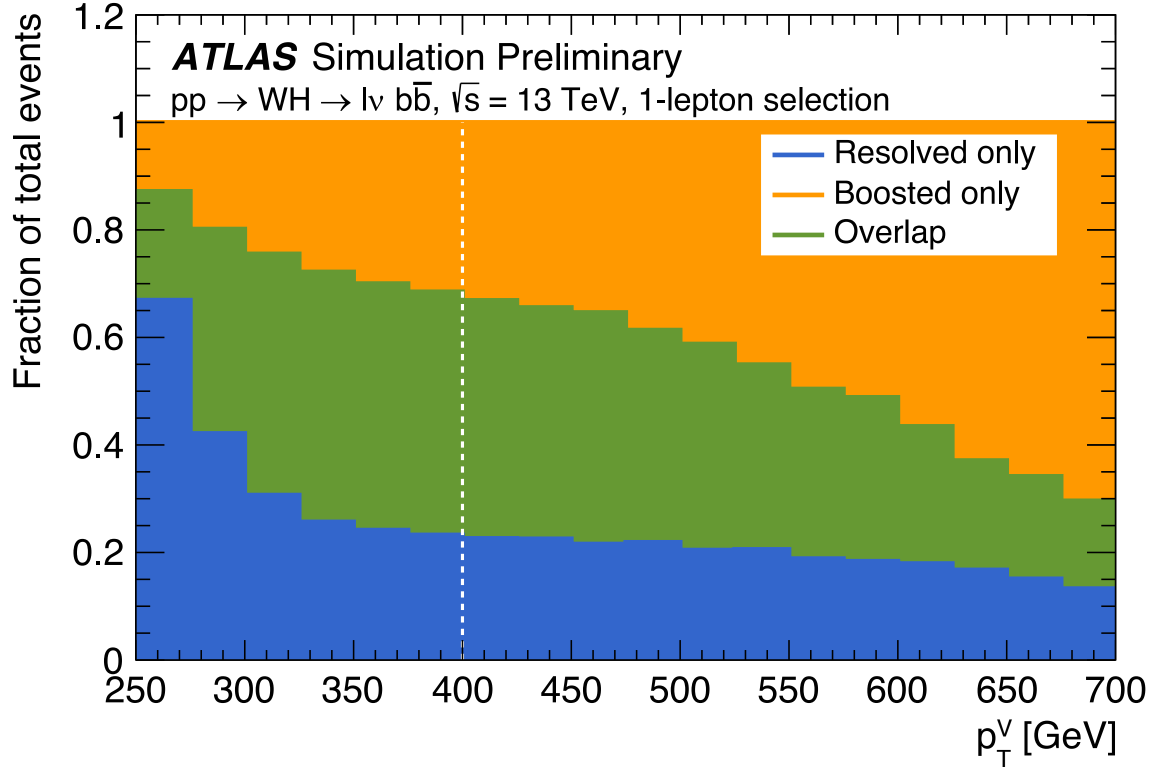
Nikhef

Thank you for your "Attention"

Back up

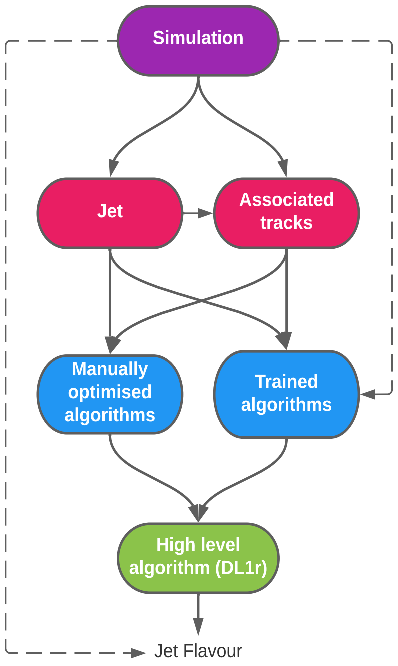
OVERLAP BETWEEN RESOLVED & BOOSTED TOPOLOGIES

ATLAS-CONF-2021-051

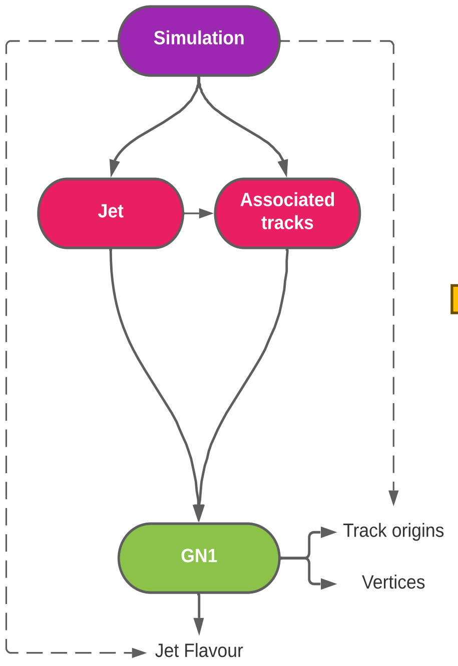


CHANGE OF FLAVOUR-TAGGING PARADIGM

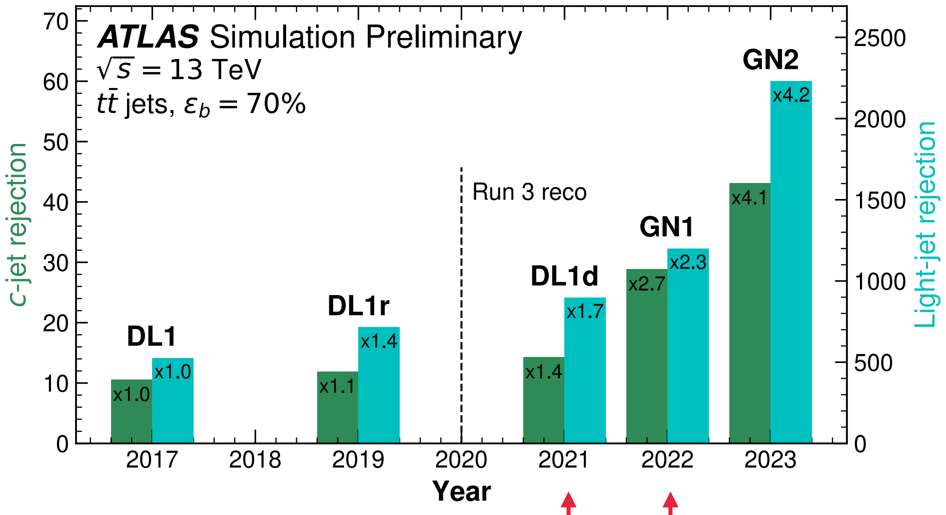
ATLAS historical approach



New paradigm



Transformer-based tagger

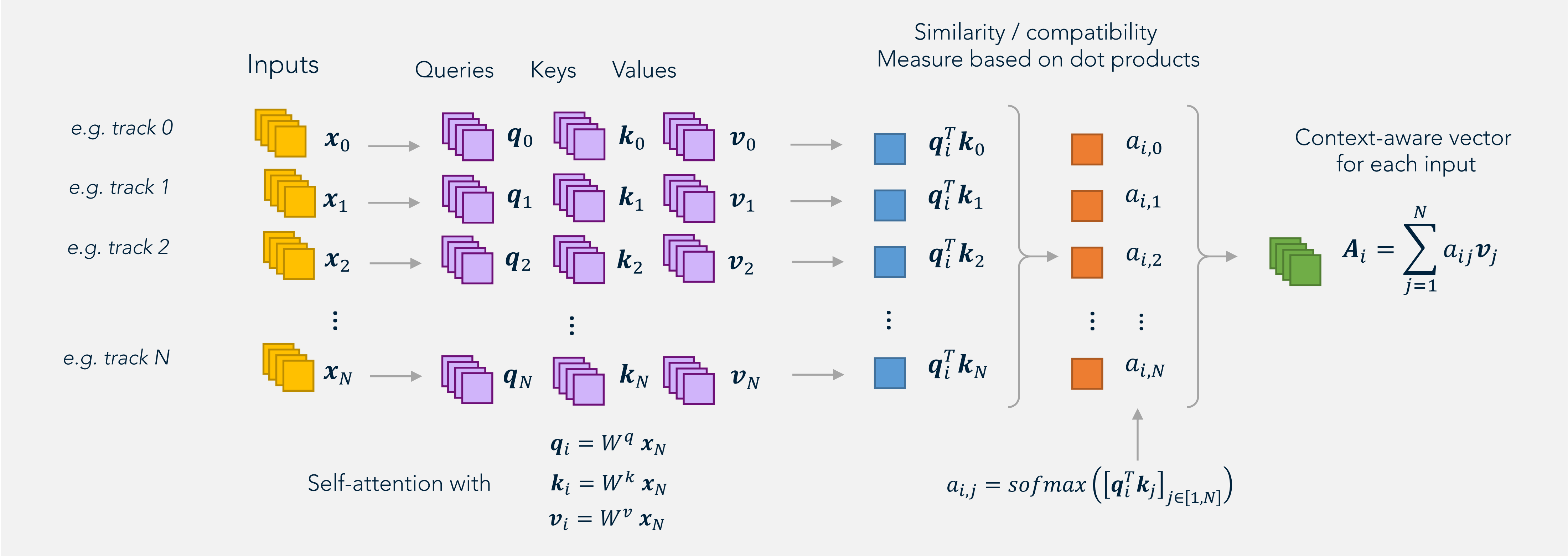


Our latest contributions to single b-tagging algorithms

First All-in-one GNN-based tagger

THE SELF-ATTENTION MECHANISM

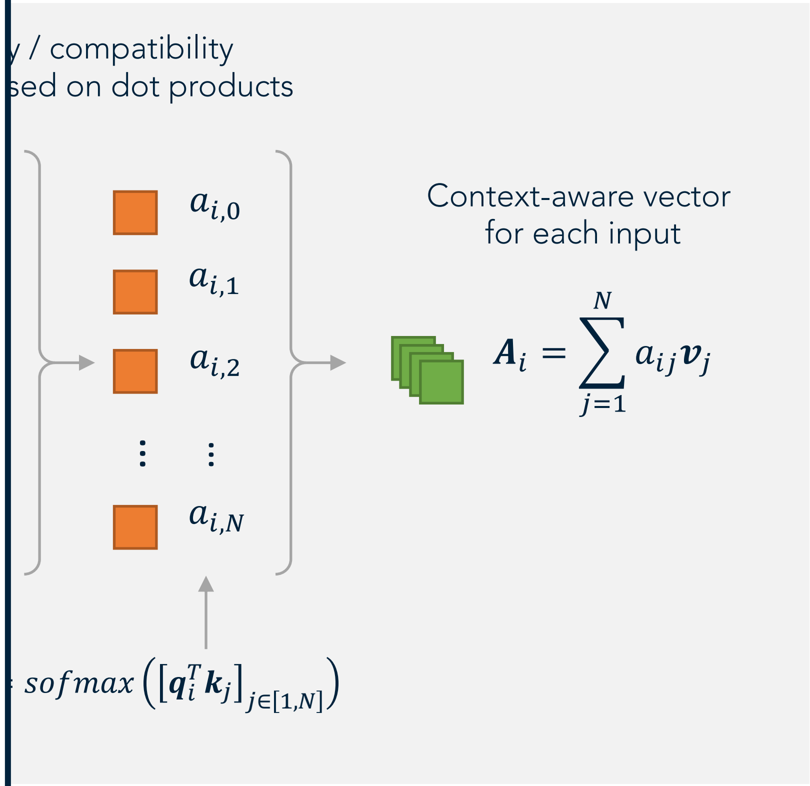
IN A NUTSHELL



NB: Self attention can be seen as an example of message passing on a fully connected graph

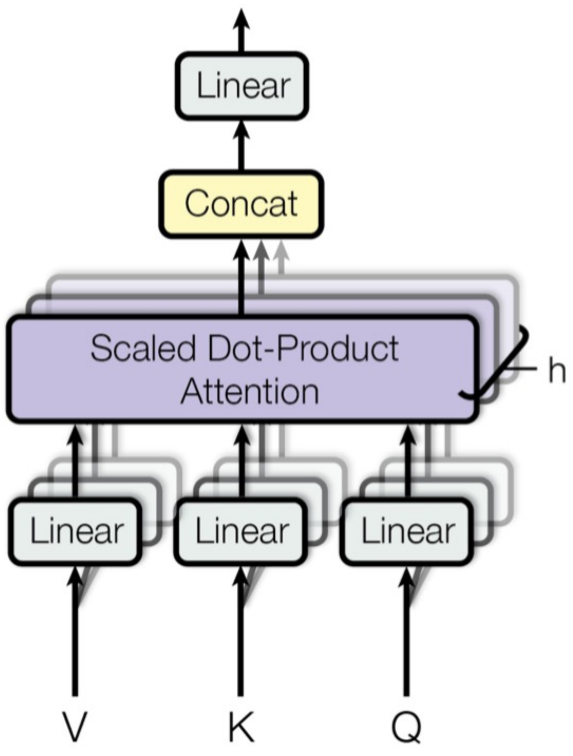
THE SELF-AT… CHANISM

IN A NUTSHELL



MULTI-HEAD ATTENTION

PRINCIPLE



Using scale dot-product attention

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

NB: Self attention can be seen as an example of message passing on a fully connected graph

GN2X INPUT KINTEMATICS

Jet Input	Description
p_T	Large-R jet transverse momentum
η	Signed large-R jet pseudorapidity
mass	Large-R jet mass

Track Input	Description
q/p	Track charge divided by momentum (measure of curvature)
$\mathrm{d}\eta$	Pseudorapidity of track relative to the large-R jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the large-R jet ϕ
d_0	Closest distance from track to primary vertex (PV) in the transverse plane
$z_0 \sin\theta$	Closest distance from track to PV in the longitudinal plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0 \sin\theta)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits

ATL-PHYS-PUB-2023-021