Universität Siegen

HEP Experimental High Energy Physics
CPPS Center for Particle Physics Siegen

ATLAS EXPERIMENT

$t\bar{t}H(c\bar{c})$: Charm-quark identification with machine learning methods

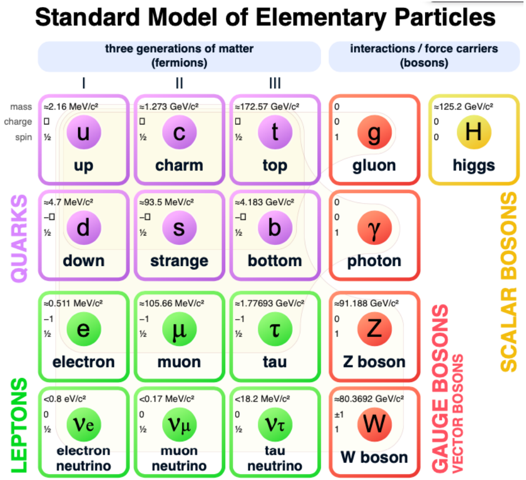
Emeline Olson

57. Herbstschule for High Energy Physics
September 2026

Motivation

Why $t\bar{t}H$?

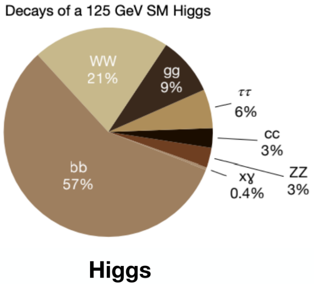
- Top quark lifetime: $\sim 10^{-25}$s
  - Decay before hadronization
- A direct measurement of top-Higgs Yukawa coupling
- High-precision measurement of Higgs' properties
  → see deviations from SM predictions



[1]

Why $H \to c\bar{c}$?

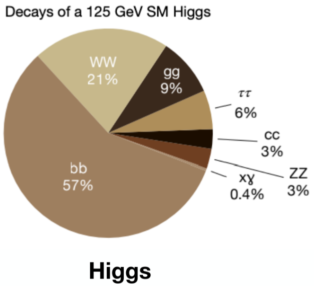
- H coupling to second-generation quarks is relatively unexplored:
  - 2019 $H \to \mu\bar{\mu}$ study [2]
  - 2025 CMS $t\bar{t}H \to c\bar{c}$ study [1]
- Welcome challenge in flavor-distinction
- Possible avenue for BSM signals to appear



[3]

Unexplored | **Rare** | **Flavor ID**

Why $H \to c\bar{c}$?



[3]

- H coupling to second-generation quarks is relatively unexplored:
  - 2019 $H \to \mu\bar{\mu}$ study [2]
  - 2025 CMS $t\bar{t}H \to c\bar{c}$ study [1]
- Welcome challenge in flavor-distinction
- Possible avenue for BSM signals to appear

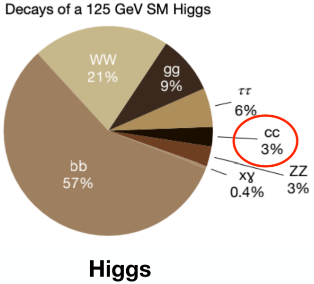
Unexplored
Furthers our knowledge of Higgs coupling to second-generation quarks; possibly deviates from SM with BSM signals

Rare

Flavor ID

Why $H \rightarrow c\bar{c}$?

- H coupling to second-generation quarks is relatively unexplored:
  - 2019 $H \rightarrow \mu\bar{\mu}$ study [2]
  - 2025 CMS $t\bar{t}H \rightarrow c\bar{c}$ study [1]
- Welcome challenge in flavor-distinction
- Possible avenue for BSM signals to appear



[3]

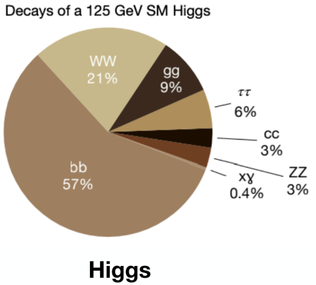
Unexplored
Furthers our knowledge of Higgs coupling to second-generation quarks; possibly deviates from SM with BSM signals

Rare
Tagging heavy flavor jets must be done efficiently to preserve statistics
→ improves software

Flavor ID

Why $H \to c\bar{c}$?

- H coupling to second-generation quarks is relatively unexplored:
  - 2019 $H \to \mu\bar{\mu}$ study [2]
  - 2025 CMS $t\bar{t}H \to c\bar{c}$ study [1]
- Welcome challenge in flavor-distinction
- Possible avenue for BSM signals to appear



[3]

Unexplored
Furthers our knowledge of Higgs coupling to second-generation quarks; possibly deviates from SM with BSM signals

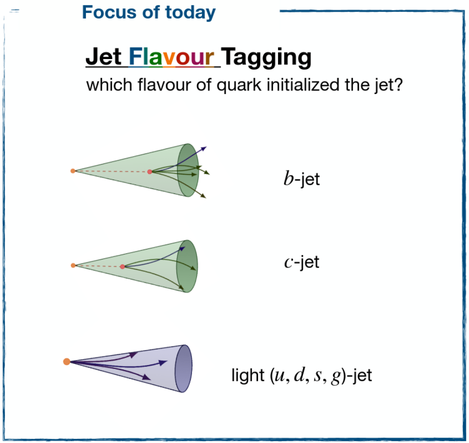
Rare
Tagging heavy flavor jets must be done efficiently to preserve statistics
→ improves software

Flavor ID
Distinguishing c-quarks from b-quarks
→ better flavor tagging algorithms for both

Machine Learning flavor tagging

What is flavor tagging?

- What kind of particle does this jet come from?
  - A classification task (ML is well-suited to do this)
- Distinguishing c vs b is necessary for a great many analyses
- Flavors can get muddied, misidentified
  - The rates of mis-IDs must be known to preserve healthy statistics, and
  - The rates of mis-IDs must be known to choose the right working point (tagging efficiency) for each analysis



[3]

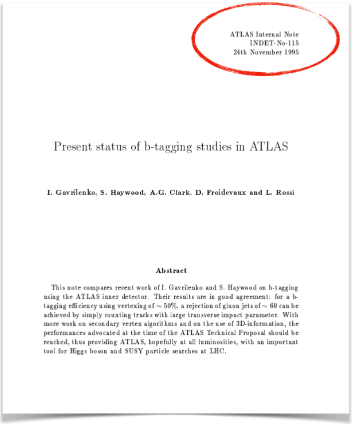
Previous tagging methods

Run 1

- Multivariate neural networks (MV1 2014) [4]

Run 2

- Boosted decision trees to begin (MV2 2018) [5]
- Deep Neural networks
  - DL1d; combines many underlying algorithms (2022) [6]
  - GN1; Graph NN (2022) [7]



ATLAS Internal Note
INDET-No-115
24th November 1995

Present status of b-tagging studies in ATLAS

I. Gavrilenko, S. Haywood, A.G. Clark, D. Froidevaux and L. Rossi

Abstract

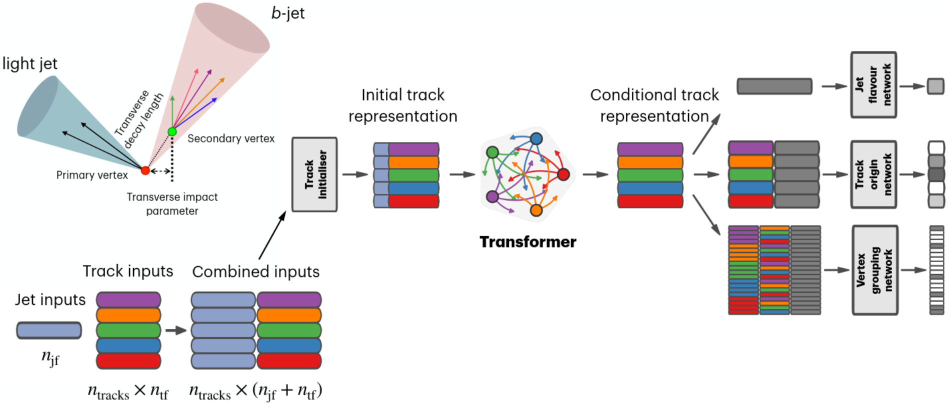
This note compares recent work of I. Gavrilenko and S. Haywood on b-tagging using the ATLAS inner detector. Their results are in good agreement: for a b-tagging efficiency using vertexing of ~ 50%, a rejection of gluon jets of ~ 60 can be achieved by simply counting tracks with large transverse impact parameter. With more work on secondary vertex algorithms and on the use of 3D-information, the performances advocated at the time of the ATLAS Technical Proposal should be reached, thus providing ATLAS, hopefully at all luminosities, with an important tool for Higgs boson and SUSY particle searches at LHC.

[3]

ATLAS' GN2 tagger

"Transforming jet flavour tagging at ATLAS" [8]

- Low-level track information as input (w/ physics-informed auxillary tasks)
- End-to-end architecture



[8]

ATLAS' GN2 tagger

"Transforming jet flavour tagging at ATLAS" [8]

- Low-level track information as input (w/ physics-informed auxillary tasks)
- End-to-end architecture

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

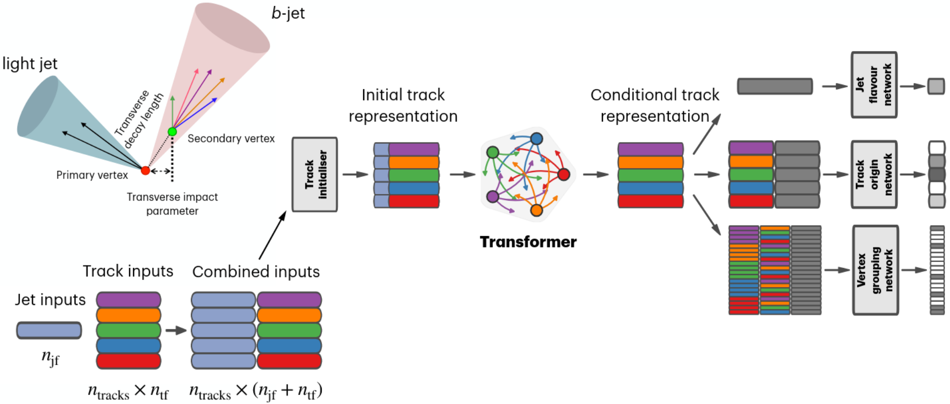
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

1 Introduction

ATLAS' GN2 tagger

"Transforming jet flavour tagging at ATLAS" [8]

- Low-level track information as input (w/ physics-informed auxillary tasks)
- End-to-end architecture

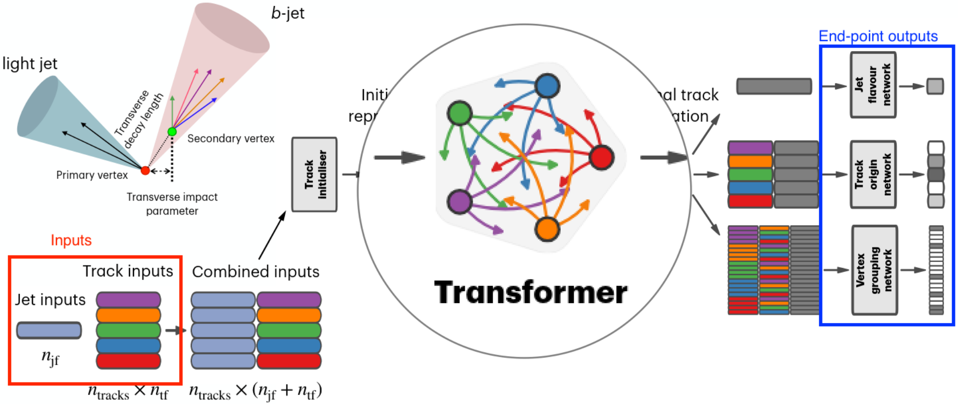


[8]

ATLAS' GN2 tagger

"Transforming jet flavour tagging at ATLAS" [8]

- Low-level track information as input (w/ physics-informed auxillary tasks)
- End-to-end architecture

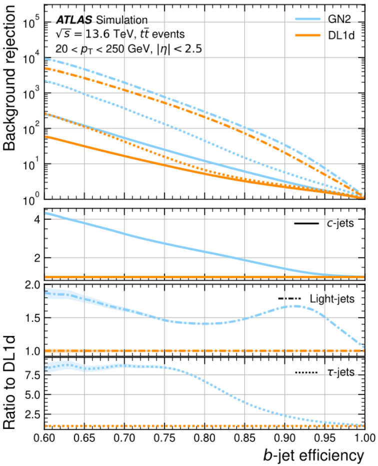


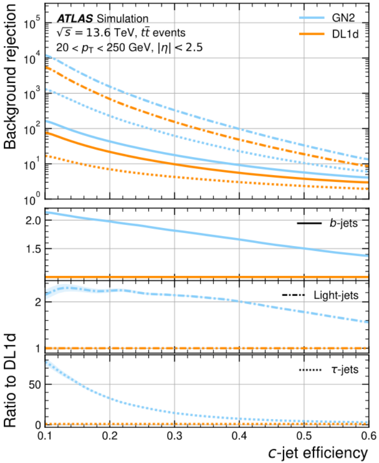
[8]

ATLAS' GN2 tagger

For $t\bar{t}$:

- 70% b-tagging efficiency
- 30% c-tagging efficiency



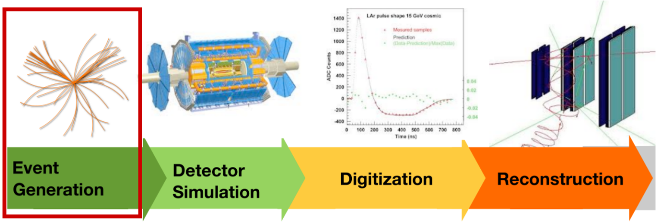


[8]

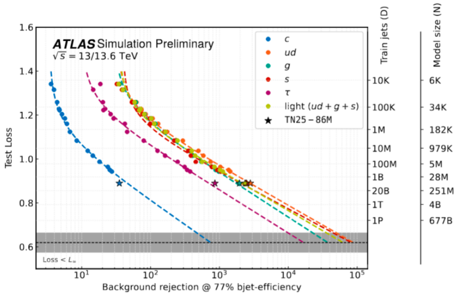
Generative ML for training GN3

ML scaling law + cheap generated data = cheaper GN3 tagger

- GN3 is a transformer model
  - ∴ it follows the ML scaling law
  - it now uses calorimetry info (expensive to simulate) in training inputs
- AtlFast3 → hybrid of Geant4 & generative ML (GAN) simulation

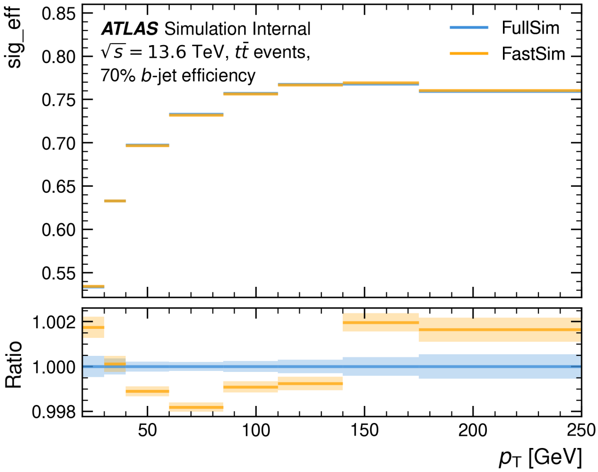


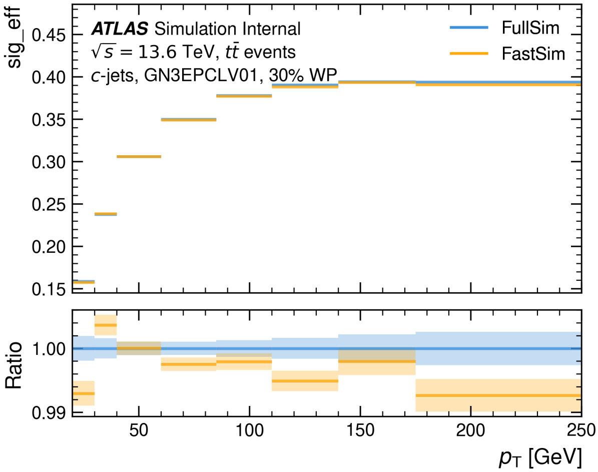
[9]



[10]

Generative ML for training GN3





The analysis

Decay channels

- Each $t \to Wb$ w/ 99.7% chance
- $t\bar{t}H \to b\bar{b}$ and $t\bar{t}$ are the leading-order backgrounds
  - The two of them combined result in a signal-to-background ratio of $\sim 4 \cdot 10^{-5}$ for the full Run 2 integrated luminosity

0L

2L

1L $\lightning$

Decay channels

- Each $t \to Wb$ w/ 99.7% chance
- $t\bar{t}H \to b\bar{b}$ and $t\bar{t}$ are the leading-order backgrounds
  - The two of them combined result in a signal-to-background ratio of $\sim 4 \cdot 10^{-5}$ for the full Run 2 integrated luminosity

0L

- Both W → qq (no leptons present)
- High stat, difficult selection

W^+ DECAY MODES		Fraction (Γ_i/Γ)
$\ell^+\nu$	[c]	(10.86 ± 0.09) %
$e^+\nu$		(10.71 ± 0.16) %
$\mu^+\nu$		(10.63 ± 0.15) %
$\tau^+\nu$		(11.38 ± 0.21) %
hadrons		(67.41 ± 0.27) %

2L

1L ⚡

Decay channels

- Each $t \to Wb$ w/ 99.7% chance
- $t\bar{t}H \to b\bar{b}$ and $t\bar{t}$ are the leading-order backgrounds
  - The two of them combined result in a signal-to-background ratio of $\sim 4 \cdot 10^{-5}$ for the full Run 2 integrated luminosity

0L

2L

- Both W $\to e,\mu$
- Supreme selection, low stat

W^+ DECAY MODES		Fraction (Γ_i/Γ)
$\ell^+\nu$	[c]	(10.86 ± 0.09) %
$e^+\nu$		(10.71 ± 0.16) %
$\mu^+\nu$		(10.63 ± 0.15) %
$\tau^+\nu$		(11.38 ± 0.21) %
hadrons		(67.41 ± 0.27) %

1L ↯

Decay channels

- Each $t \to Wb$ w/ 99.7% chance
- $t\bar{t}H \to b\bar{b}$ and $t\bar{t}$ are the leading-order backgrounds
  - The two of them combined result in a signal-to-background ratio of $\sim 4 \cdot 10^{-5}$ for the full Run 2 integrated luminosity

0L

2L

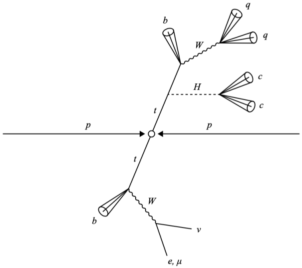
1L ↯

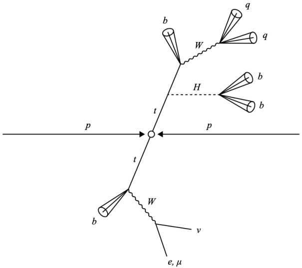
- One W $\to qq$, the other to e,μ
- ! Balance of selection w/ statistics

W^+ DECAY MODES		Fraction (Γ_i/Γ)
$\ell^+\nu$	[c]	(10.86 ± 0.09) %
$e^+\nu$		(10.71 ± 0.16) %
$\mu^+\nu$		(10.63 ± 0.15) %
$\tau^+\nu$		(11.38 ± 0.21) %
hadrons		(67.41 ± 0.27) %

Pre-selection (1L)

- $t\bar{t}H \rightarrow c\bar{c}$ events must follow the following pre-selection reqs:
  - Each event must contain at least five reconstructed jets.

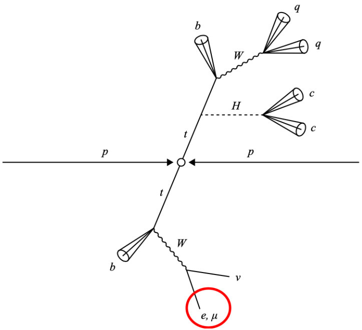


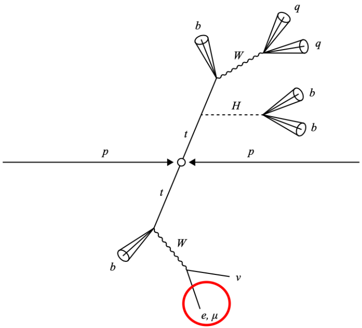


- $H \rightarrow c\bar{c}$ w/ 3% chance
- $H \rightarrow b\bar{b}$ w/ 58% chance

Pre-selection (1L)

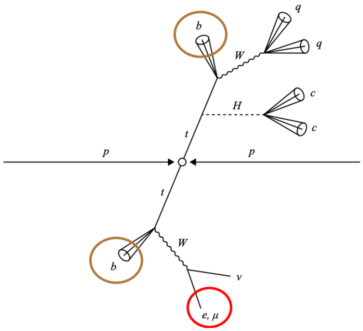
- $t\bar{t}H \rightarrow c\bar{c}$ events must follow the following pre-selection reqs:
  - Each event must contain at least five reconstructed jets.
  - Exactly one electron or muon must be present in the event.

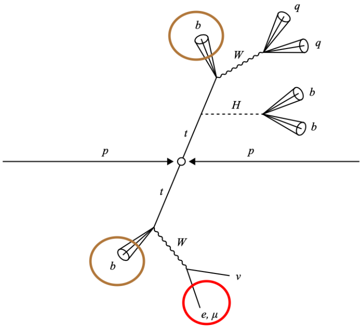




Pre-selection (1L)

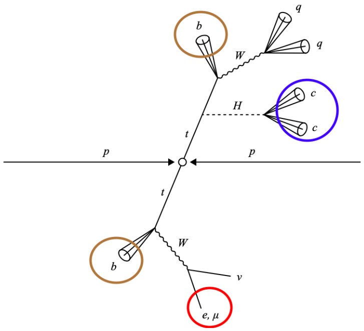
- $t\bar{t}H \rightarrow c\bar{c}$ events must follow the following pre-selection reqs:
  - Each event must contain at least five reconstructed jets.
  - Exactly one electron or muon must be present in the event.
  - At least two jets must fulfill a b–tagging requirement (at least loose b–tagging)

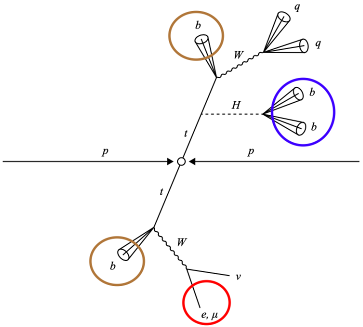




Pre-selection (1L)

- $t\bar{t}H \to c\bar{c}$ events must follow the following pre-selection reqs:
  - Each event must contain at least five reconstructed jets.
  - Exactly one electron or muon must be present in the event.
  - At least two jets must fulfill a b–tagging requirement (at least loose b–tagging)





Discriminants, working points

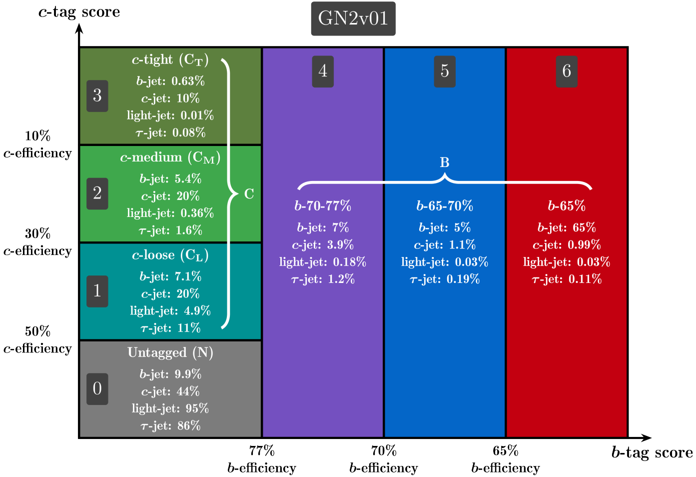
Discriminants for the analysis are defined as:

$$D_b = log(\frac{p_b}{f_c \cdot p_c + f_\tau \cdot p_\tau + (1 - f_c - f_\tau) \cdot p_u})$$

$$D_c = log(\frac{p_c}{f_b \cdot p_b + f_\tau \cdot p_\tau + (1 - f_b - f_\tau) \cdot p_u})$$

with f_c, f_τ, and f_b being constants defined as needed for efficient tagging, and the probabilities given by GN2.

Discriminants, working points



- Each jet is assigned to one of these seven categories
- As a default for comparison, we use 30% c-jet efficiency and 70% b-jet efficiency

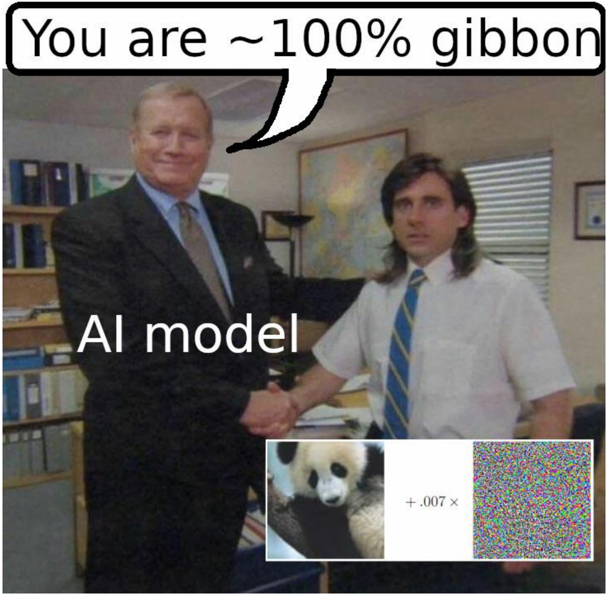
Conclusion

Conclusion

The analysis will yield:

- Higgs-top coupling understanding
- Further constrained limits on BSM signals
- ML flavor tagging improvements
- (part of) A PhD thesis for yours truly

Thank you kindly!



References

[1] The CMS Collaboration. "Search for Higgs boson decay to a charm quark-antiquark pair via ttH production". In: (2025). URL: https://cds.cern.ch/record/2929444.

[2] The CMS Collaboration. "Search for the Higgs boson decaying to two muons in proton-proton collisions at s = 13 TeV". In: (2019). URL: https://arxiv.org/abs/1807.06325.

[3] Binbin Dong. *Jet flavor Tagging in ATLAS*. July 2024. URL: https://indico.cern.ch/event/1418291/attachments/2888239/5090663/FTAG-lecture.pdf.

[4] ATLAS Collaboration. "Performance of *b*-Jet Identification in the ATLAS Experiment". In: *Journal of Instrumentation* 11.04 (Apr. 2016), P04008. DOI: 10.1088/1748-0221/11/04/P04008. arXiv: 1512.01094 [hep-ex]. URL: https://doi.org/10.1088/1748-0221/11/04/P04008.

References

[5] ATLAS Collaboration. "Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2018.8 (Aug. 2018), p. 089. DOI: 10.1007/JHEP08(2018)089. arXiv: 1805.01845 [hep-ex]. URL: https://doi.org/10.1007/JHEP08(2018)089.

[6] ATLAS Collaboration. "ATLAS flavor-tagging algorithms for the LHC Run 2 *pp* collision dataset". In: *The European Physical Journal C* 83.7 (July 2023), p. 681. DOI: 10.1140/epjc/s10052-023-11699-1. arXiv: 2211.16345 [physics.data-an]. URL: https://doi.org/10.1140/epjc/s10052-023-11699-1.

[7] ATLAS Collaboration. *Graph Neural Network Jet flavor Tagging with the ATLAS Detector*. Tech. rep. ATL-PHYS-PUB-2022-027. Geneva: CERN, July 2022. URL: https://cds.cern.ch/record/2811135.

References

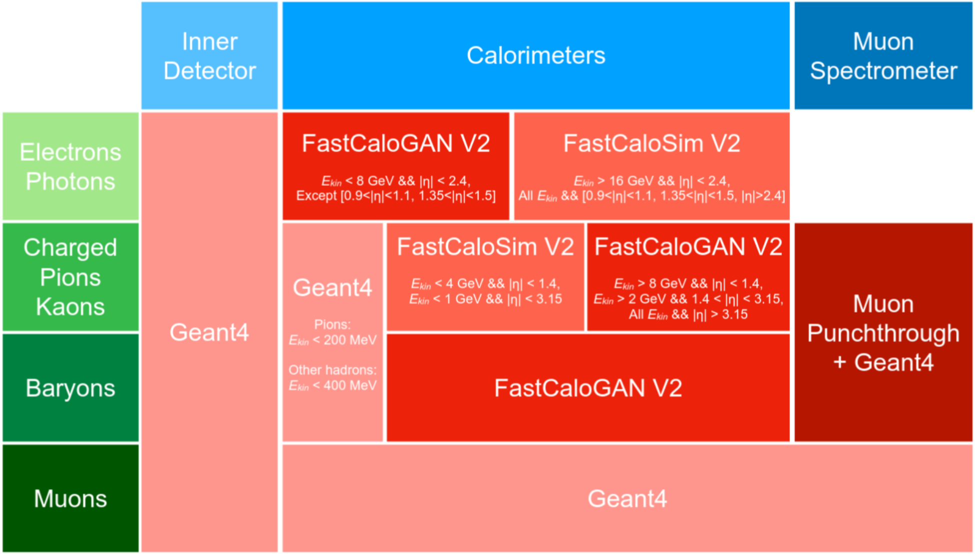
[8] ATLAS Collaboration. "Transforming jet flavor tagging at ATLAS". In: *Nature Communications* 17.1 (2026), p. 541. DOI: 10.1038/s41467-025-65059-6. arXiv: 2505.19689 [hep-ex]. URL: https://doi.org/10.1038/s41467-025-65059-6.

[9] E Ritsch and the ATLAS Collaboration. "Concepts and Plans towards fast large scale Monte Carlo production for the ATLAS Experiment". In: *Journal of Physics: Conference Series* 523.1 (June 2014), p. 012035. DOI: 10.1088/1742-6596/523/1/012035. URL: https://doi.org/10.1088/1742-6596/523/1/012035.

[10] ATLAS Collaboration. "Carpe Datum: Scaling behavior of transformers for heavy hadron flavor identification". In: *CDS* ATL-SOFT-PUB-2026-002 (2026).

Backup

A map of AtlFast3



Analysis feature definition

- The analysis dataset is defined by up to 10 jets per event, with each jet having the following features:
  - Kinematics (E, p_T, η, ϕ)
  - Flavor-tagging probabilities (for jet to originate from whichever flavor)–GN2 provided
  - True flavor
- And also, some event-level features:
  - number of reconstructed jets
  - number of b-tagged and light-tagged jets
  - Missing transverse energy and ϕ of MET vector
  - true process label ($t\bar{t}H \rightarrow c\bar{c}$, etc)