Developments in b-tagging for the ATLAS upgrade and their impact on di-Higgs sensitivity

Leonardo Splendori

Supervisors: Arnaud Duperrin, Thomas Strebler

CPPM 3rd year seminar - 08/12/2025

ATLAS EXPERIMENT

CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE CPPM

A Toroidal LHC ApparatuS

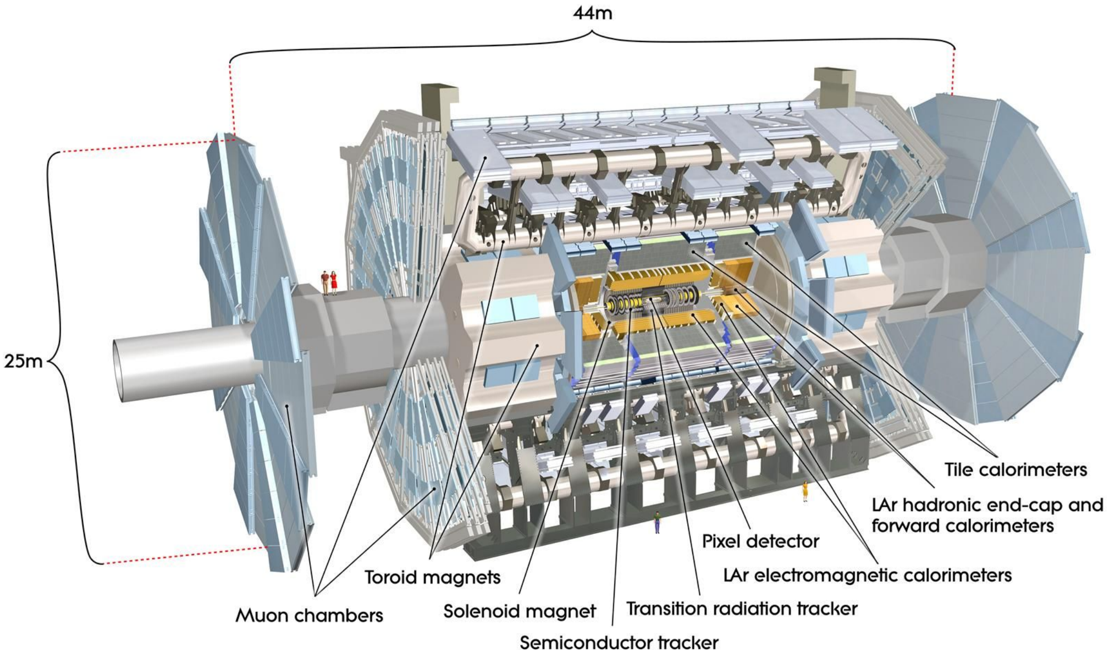
One of two **general purpose** experiments at LHC.

Beams cross in the center at **13.6 TeV** for proton-proton collisions

Composed of multiple systems:

- Inner Detector.
- Calorimeters.
- Muon Spectrometer.
- Magnets for magnetic fields.

Beams cross in the center → Collisions.



Flavour tagging at ATLAS

Collisions produce high energy and heavy hadrons (ex. b-hadrons).

Decays with multiple products, a single hadron can generate a multitude of tracks in the detector.

These tracks will have high impact parameters wrt the primary vertex → characteristic topology.

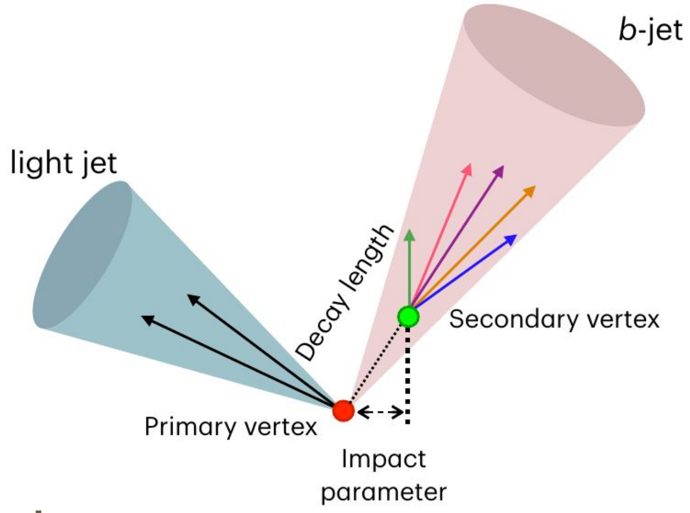
The original parton manifests as a collimated stream of particles.

A **hadronic jet**. Can be categorized by **flavour**:

- b-jets.
- c-jets.
- light-jets.
- tau-jets.

Can be identified from their topology and kinematics.

Modern flavour tagging techniques include **complex neural networks**.



[CERN-EP-2025-103]

Inside a neural network based tagger

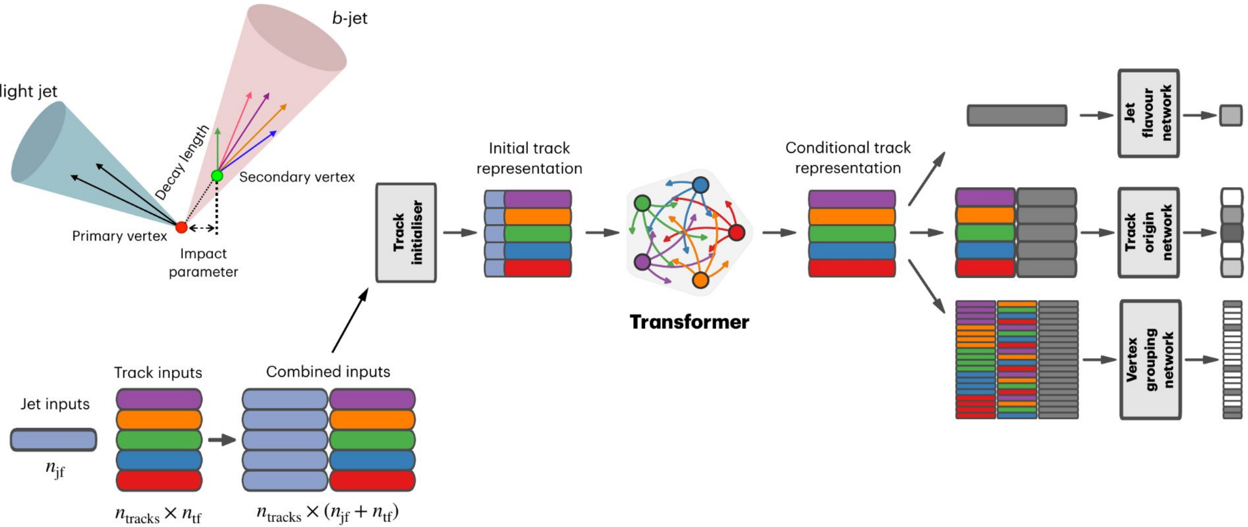
Currently deployed flavour tagging neural network is **GN2**.

Transformer based neural network.

Takes combined jet and track variables as inputs.

Embeds their representation in a transformer.

Loss is sum of multiple terms including flavour identification, track-origin association and vertex grouping.



[CERN-EP-2025-103]

Transformer based architectures are powerful but **expensive to train**.

Requires GPUs and large training datasets.

High Luminosity LHC

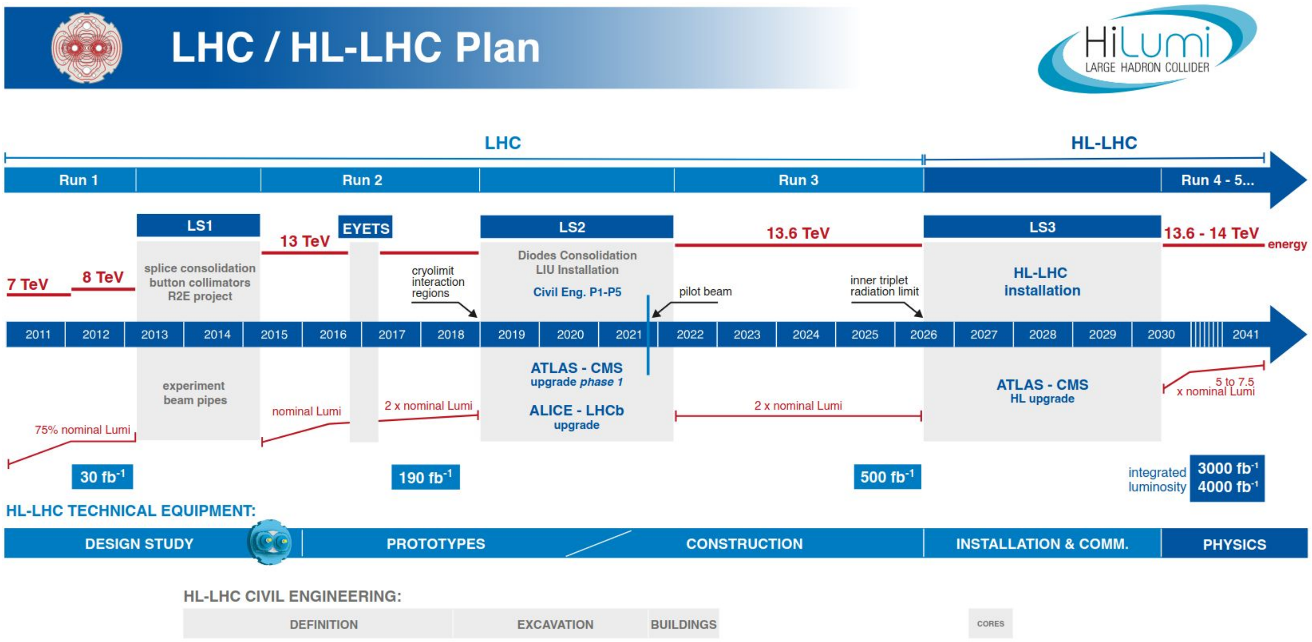
Starting in 2026, LHC and ATLAS to undergo **upgrading**.

Will push energy and luminosity frontiers.

Energy up to **√s = 14 TeV**.

Avg. interactions per crossing up to **200** (from 42 @ Run 3).

New hardware and software required to perform the physics analyses.



High Luminosity LHC

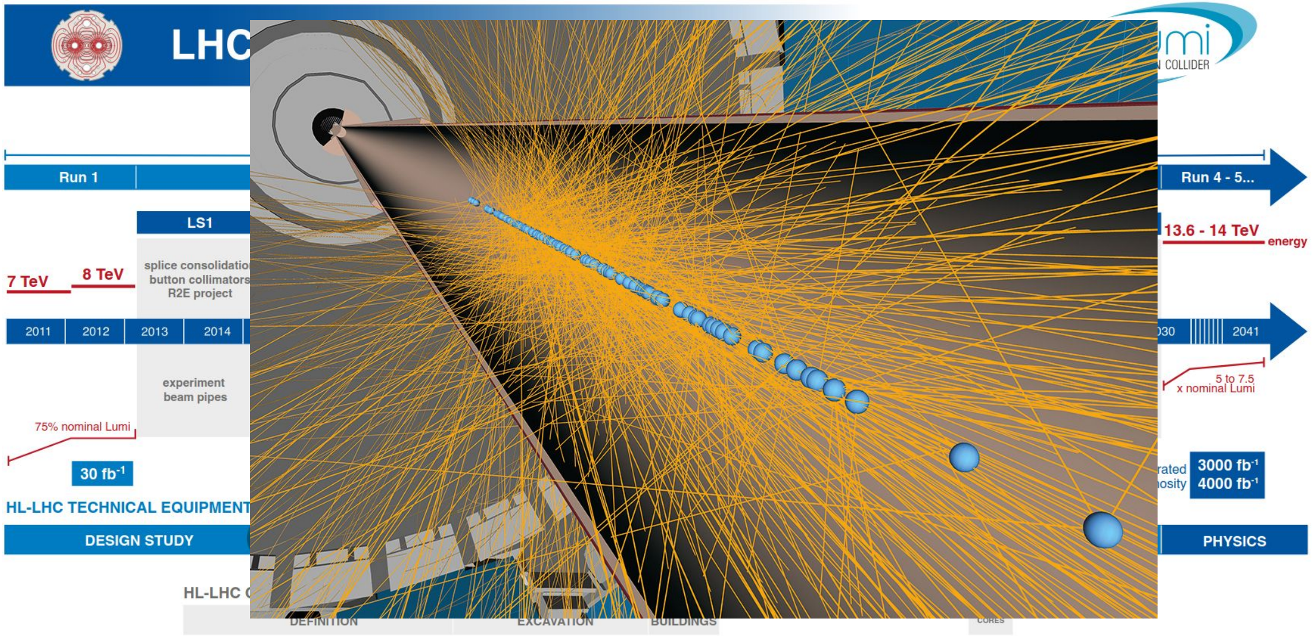
Starting in 2026, LHC and ATLAS to undergo **upgrading**.

Will push energy and luminosity frontiers.

Energy up to **$\sqrt{s}$ = 14 TeV**.

Avg. interactions per crossing up to **200** (from 42 @ Run 3).

New hardware and software required to perform the physics analyses.



Expectations for HL-LHC

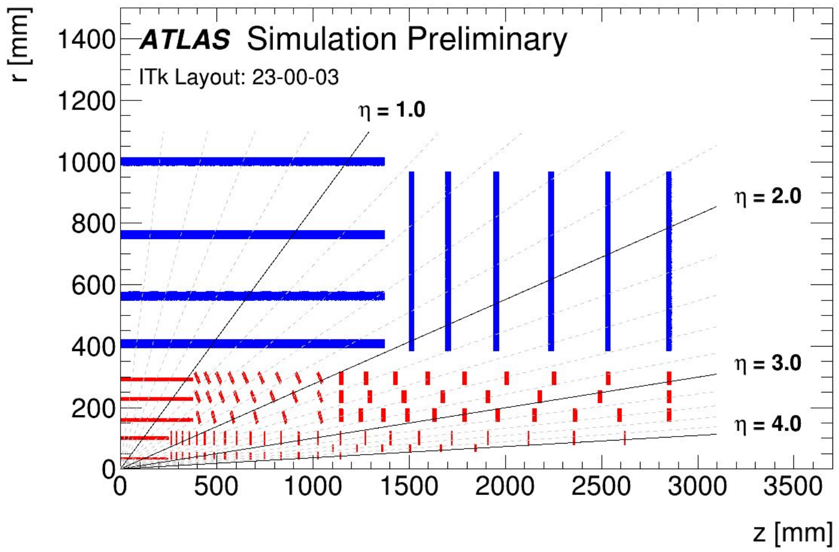
Different operating conditions:

- ITk → New detector geometry.
- Wider $|\eta|$ range ($|\eta| < 4$ vs $|\eta| < 2.5$).
- Higher luminosity → Increased pile-up.

With ITk, expected improvements in tracking and vertexing.

Finer detector granularity wrt Run 3.

Improved impact parameter resolution particularly useful in dense environments.



[ATL-PHYS-PUB-2021-024]

Training the tagger

We take the current architecture used in Run 3, GN2, and train it on Run 4 simulations.

High statistics training using ~115M jets. Run 3 training statistics ~192M jets.

Multiple steps to get results:

Training the tagger

We take the current architecture used in Run 3, GN2, and train it on Run 4 simulations.

High statistics training using ~115M jets. Run 3 training statistics ~192M jets.

Multiple steps to get results:

- **Data pre-processing**:
  - Prepares training/validation/test samples.
  - Performs resampling to reduce biases.

Training the tagger

We take the current architecture used in Run 3, GN2, and train it on Run 4 simulations.

High statistics training using ~115M jets. Run 3 training statistics ~192M jets.

Multiple steps to get results:

- **Data pre-processing**:
  - Prepares training/validation/test samples.
  - Performs resampling to reduce biases.
- **Model training**:
  - Actual training of the model using the training/validation samples.
  - Very computationally expensive and time consuming.

Training the tagger

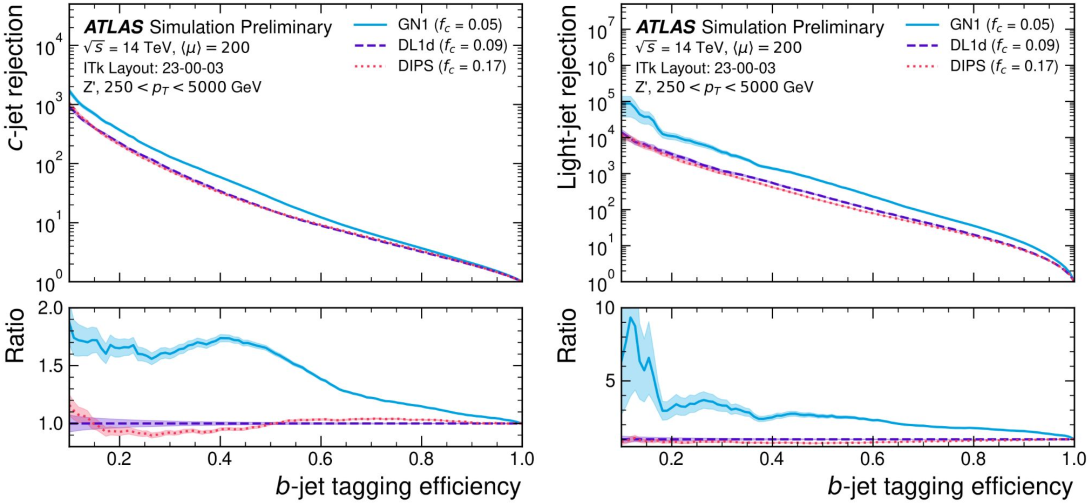
We take the current architecture used in Run 3, GN2, and train it on Run 4 simulations.

High statistics training using ~115M jets. Run 3 training statistics ~192M jets.

Multiple steps to get results:

- **Data pre-processing**:
    - Prepares training/validation/test samples.
    - Performs resampling to reduce biases.
- **Model training**:
    - Actual training of the model using the training/validation samples.
    - Very computationally expensive and time consuming.
- **Model evaluation**:
    - Model inference is computed on the test sample.
    - Production of performance plots and analyses.

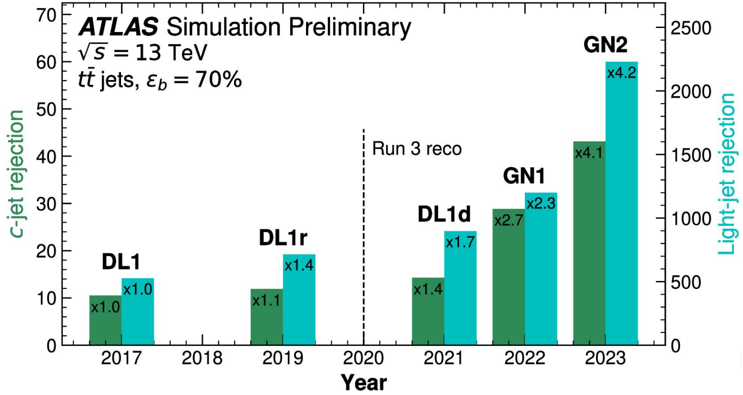
Previous Run 4 results



[ATL-PHYS-PUB-2022-047]

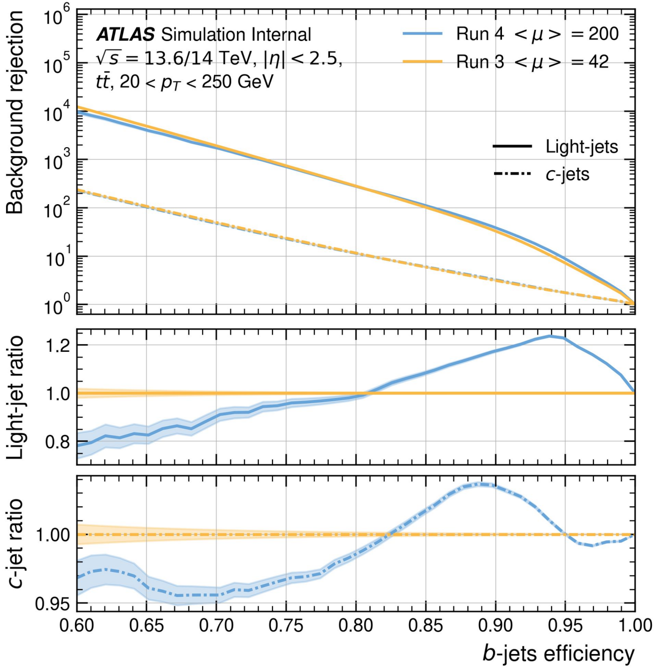
Left: Plots showing background rejections as functions of signal efficiency (ROC curves) for previous taggers.

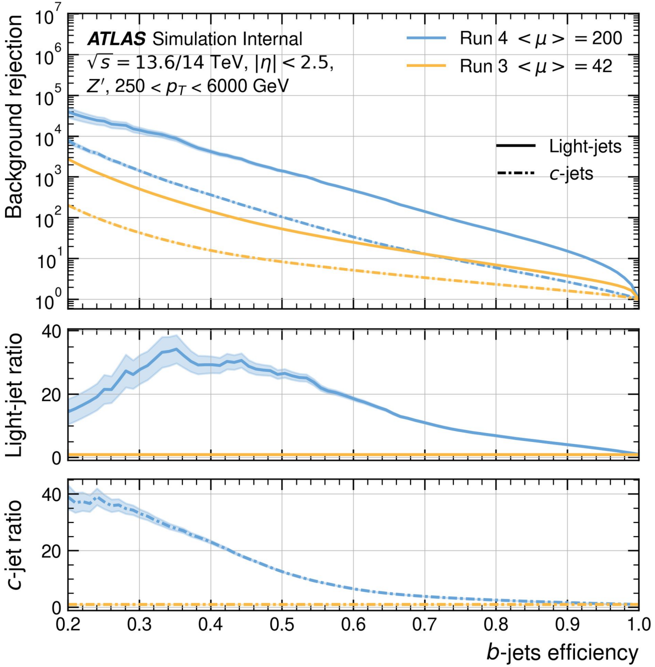
GN2 overall outperforms all previous taggers.



[FTAG-2023-01]

Performance: Run 3 vs Run 4





Large improvement at high p_T (right plot).

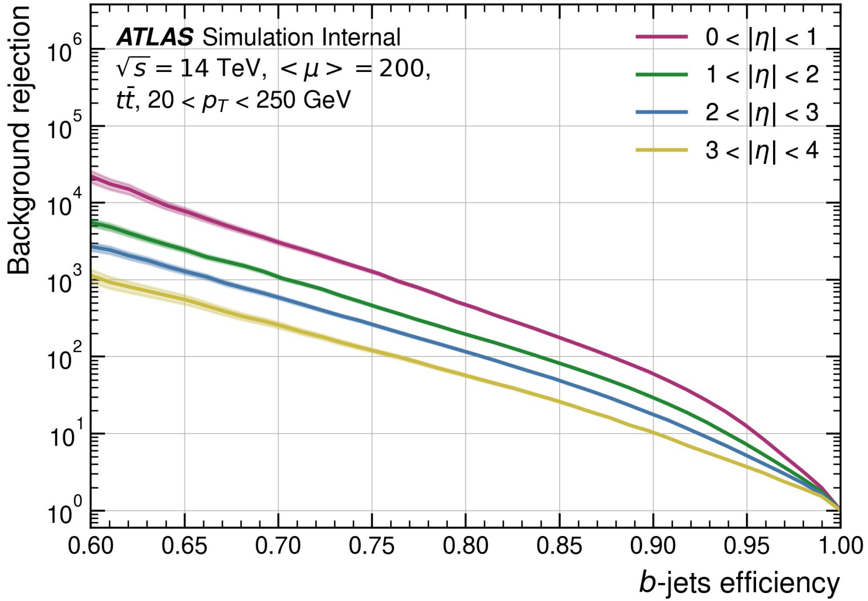
Up to ~40x in rejection.

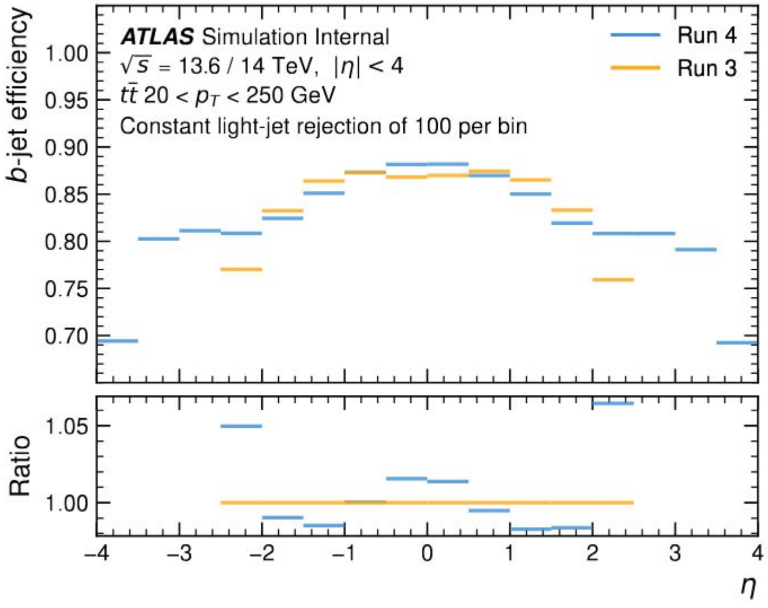
Possible thanks to ITk:

- Improved impact parameter resolution
- Improved tracking efficiency at high p_T.

Promising results for analyses targeting boosted topologies!

Performance vs kinematics: η



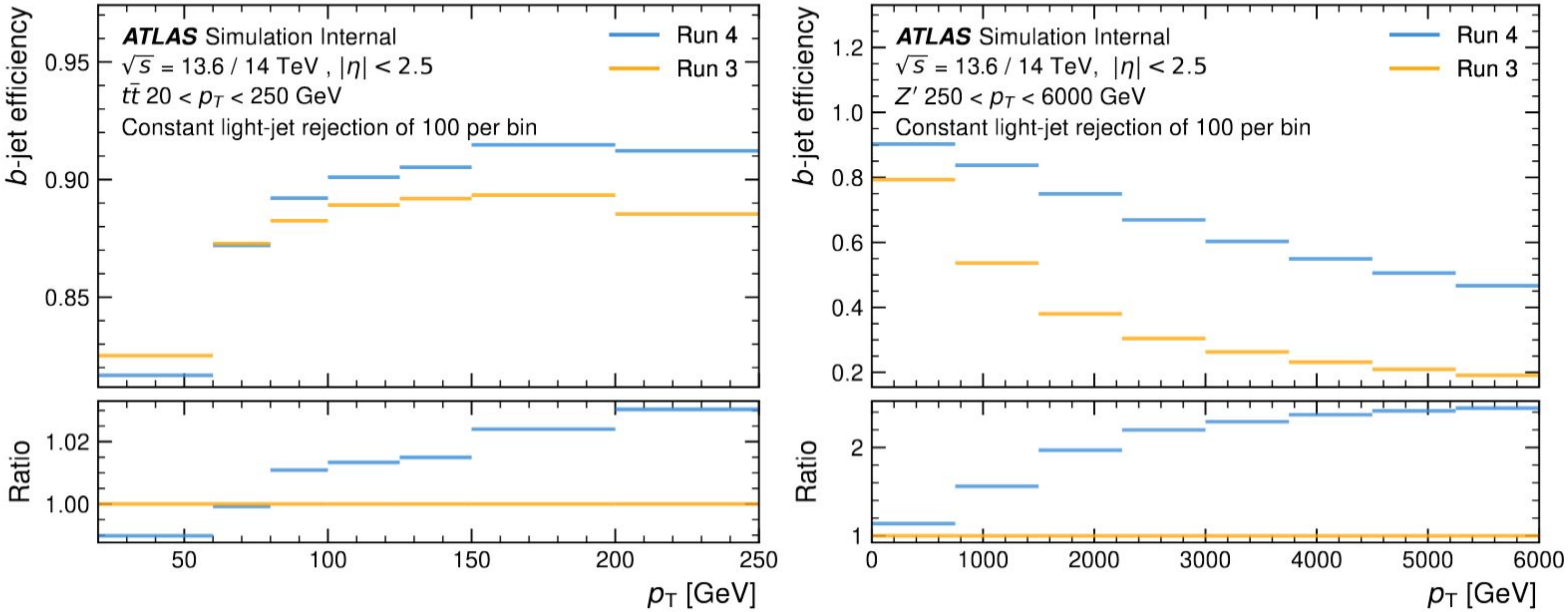


Left: ROC curves for Run 4 tagger in multiple |η| slices. Right: b-efficiency for Run 4 and Run 3 taggers as functions of η.

Left plot shows performance degradation in forward regions → expected from higher material budget.

Right plot illustrates a modest 2% improvement in the central region, up to 6% for 2 < |η| < 2.5

Performance vs kinematics: p_T

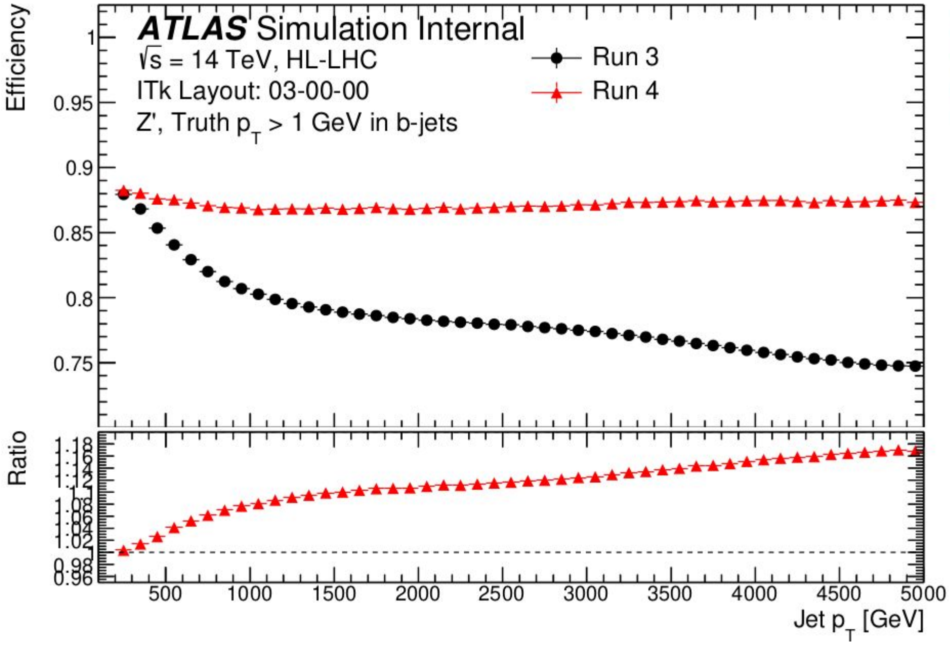


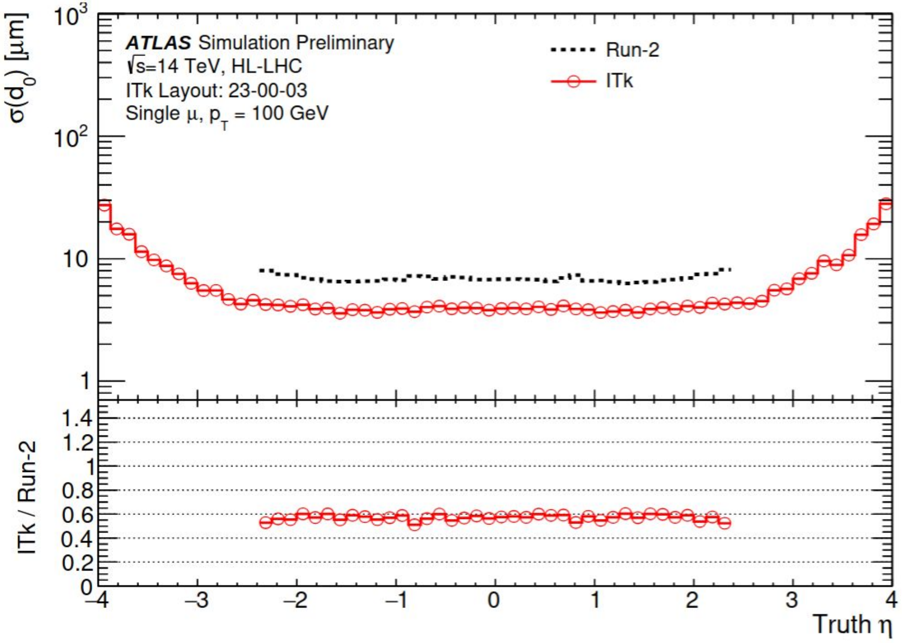
Plots of the b-efficiency as functions of p_T.

Efficiency vs p_T plots highlight the performance improvements compared to Run 3.

At fixed rejection, Run 4 training shown to outperform Run 3 one for p_T > 80 GeV.

Improvement at high pT



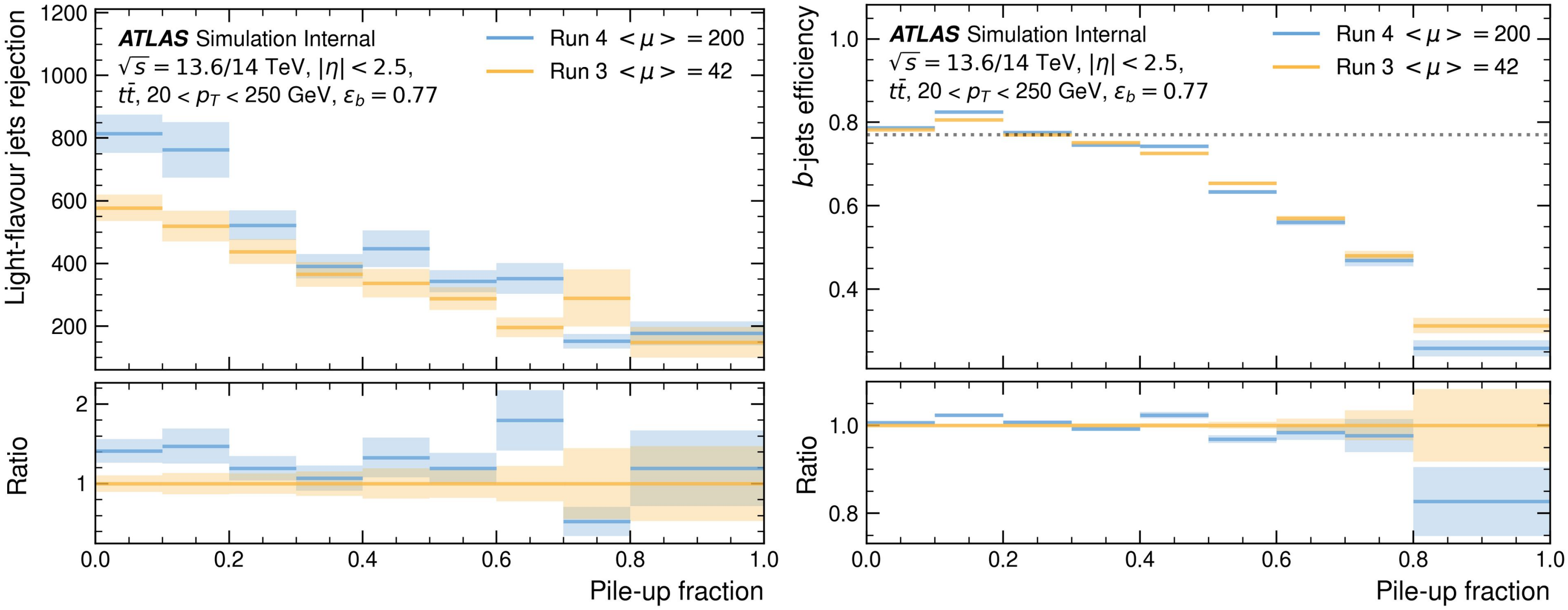


Large improvement on the Z' samples at high p_T not surprising.

ITk offers a significant **improvement** in **tracking efficiency** and **impact parameter resolution** compared to Run 3. Finer granularity improving tracking for dense jets. Naturally translates to a higher flavour tagging efficiency.

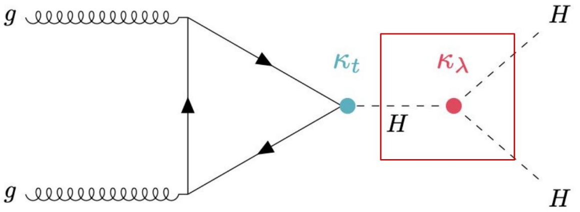
Pile-up robustness

b-tagging efficiency and *light-jet* rejection as functions of ratio between number pile-up tracks and total tracks in jet. Working point set at 77% b-tagging efficiency.



Both Run 3 and HL-LHC b-jet efficiency strongly impacted by PU fraction.
Overall **similar response to PU** between the two taggers.

DiHiggs analyses: motivation



Higgs self-coupling constrains the shape of Higgs potential.

Currently very unconstrained.

Studying processes involving H → HH will improve these constraints.

Interesting for innumerable physics topics.

Di-Higgs search channels include (but are not limited to):

- HH → bbττ
- HH → bb$\gamma\gamma$
- HH → 4b
- …

Di-Higgs analyses involving b-jets have the highest statistics.

	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

[doi.org/10.3390/sym14020260]

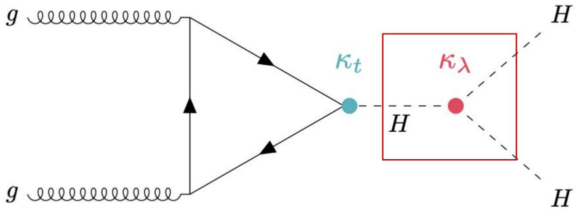
DiHiggs analyses: motivation

Higgs self-coupling constrains the shape of Higgs potential.

Currently very unconstrained.

Studying processes involving H → HH will improve these constraints.

Interesting for innumerable physics topics.



Di-Higgs search channels include (but are not limited to):

- HH → bbττ
- HH → bb$\gamma\gamma$
- HH → 4b
- …

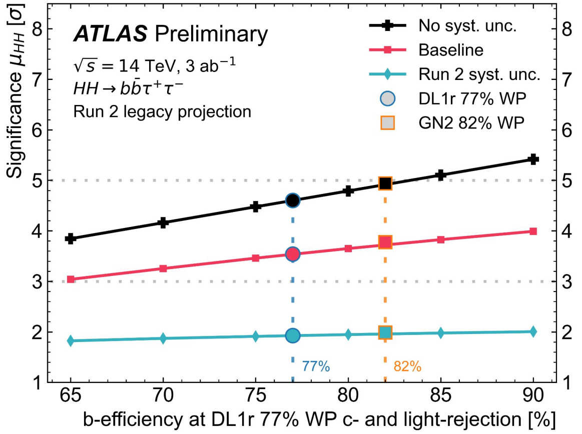
That is a lot of b-jets.

Di-Higgs analyses involving b-jets have the highest statistics.

	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

[doi.org/10.3390/sym14020260]

DiHiggs analyses: impact



[ATL-PHYS-PUB-2024-016]

Algorithmic improvement	Uncertainty scenario	Significance [σ]	68% CI μ_{HH} [%]	68% CI κ_λ
b-tagging improved by 5%	Run 2 syst. unc.	2.85	+40/−34	[0.46, 1.66]
	Theory unc. halved	3.47	+31/−29	[0.55, 1.55]
	Baseline	4.44	+26/−24	[0.60, 1.45]
	No syst. unc.	6.33	+17/−17	[0.73, 1.31]
Nominal	Run 2 syst. unc.	2.73	+47/−36	[0.39, 1.71]
	Theory unc. halved	3.32	+33/−31	[0.52, 1.59]
	Baseline	4.26	+28/−25	[0.58, 1.48]
	No syst. unc.	5.98	+18/−18	[0.71, 1.33]

[ATL-PHYS-PUB-2025-006]

Combined HH expected significance as function of b-tagging efficiency.

Projections for HL-LHC. Higher luminosity vital to diHiggs program.

Baseline: theoretical uncertainties halved, statistical ones scaled down with expected luminosity. Experimental ones scaled down according to expected improvements from larger datasets.

Improvements in b-tagging impact the projections.

Nominal baseline scenario and 5% improved b-tagging well within reach.

Summary & Conclusions

- Flavour tagging algorithms need to be re-optimized for HL-LHC.
- Flavour tagging extended up to $|\eta| < 4$.
- Large improvement in background rejection at high p_T.
- Pile-up robustness consistent with Run 3.
- Improvements in b-tagging of high interest to future analyses.

Thank you for following

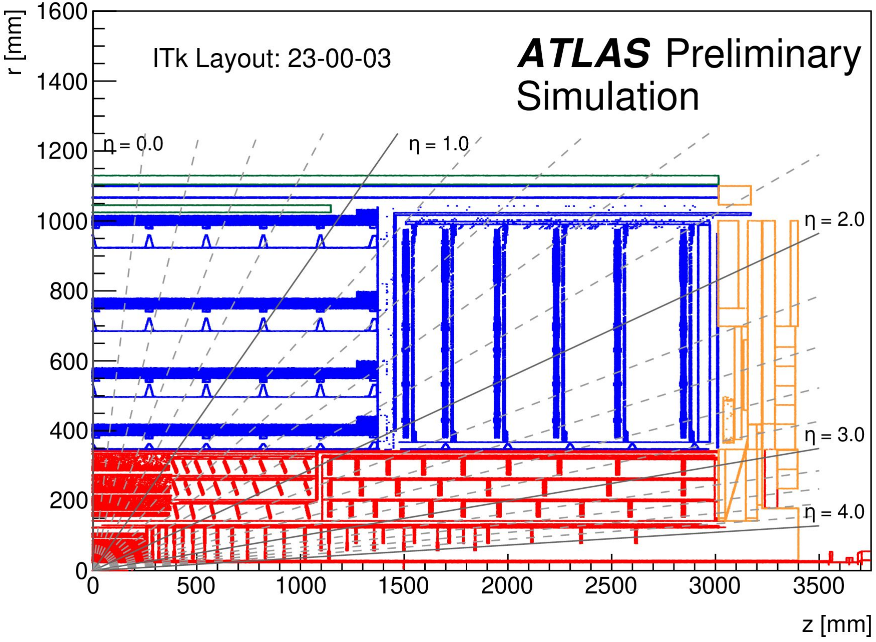
Backup

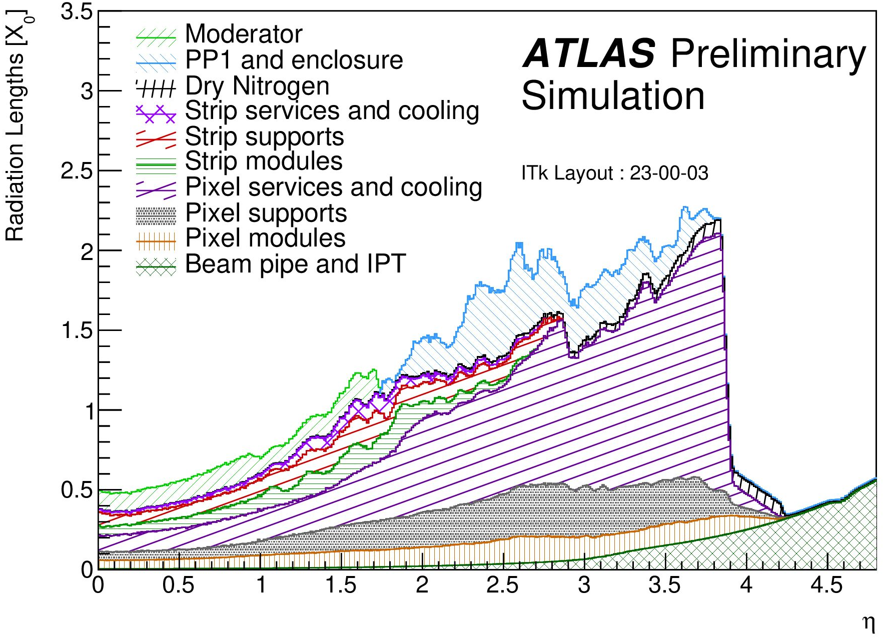
Backup: Track selections

Different cuts on the tracks between Run 3 and Run 4 samples.

Requirements	Run 3	Run 4		
	$\|\eta\| < 2.5$	$\|\eta\| < 2.0$	$2.0 < \|\eta\| < 2.6$	$2.6 < \|\eta\| < 4.0$
silicon hits	≥ 8	≥ 9	≥ 8	≥ 7
pixel hits	≥ 1	≥ 1	≥ 1	≥ 1
holes	≤ 2	≤ 2	≤ 2	≤ 2
p_T [MeV]	> 500	> 900	> 500	> 500
$\|d_0\|$ [mm]	< 3.5	< 2.0	< 2.0	< 3.5
$\|z_0 \sin\theta\|$ [mm]	< 5.0	< 5.0	< 5.0	< 5.0

Backup: ITk Material Budget





[ATL-PHYS-PUB-2021-024]