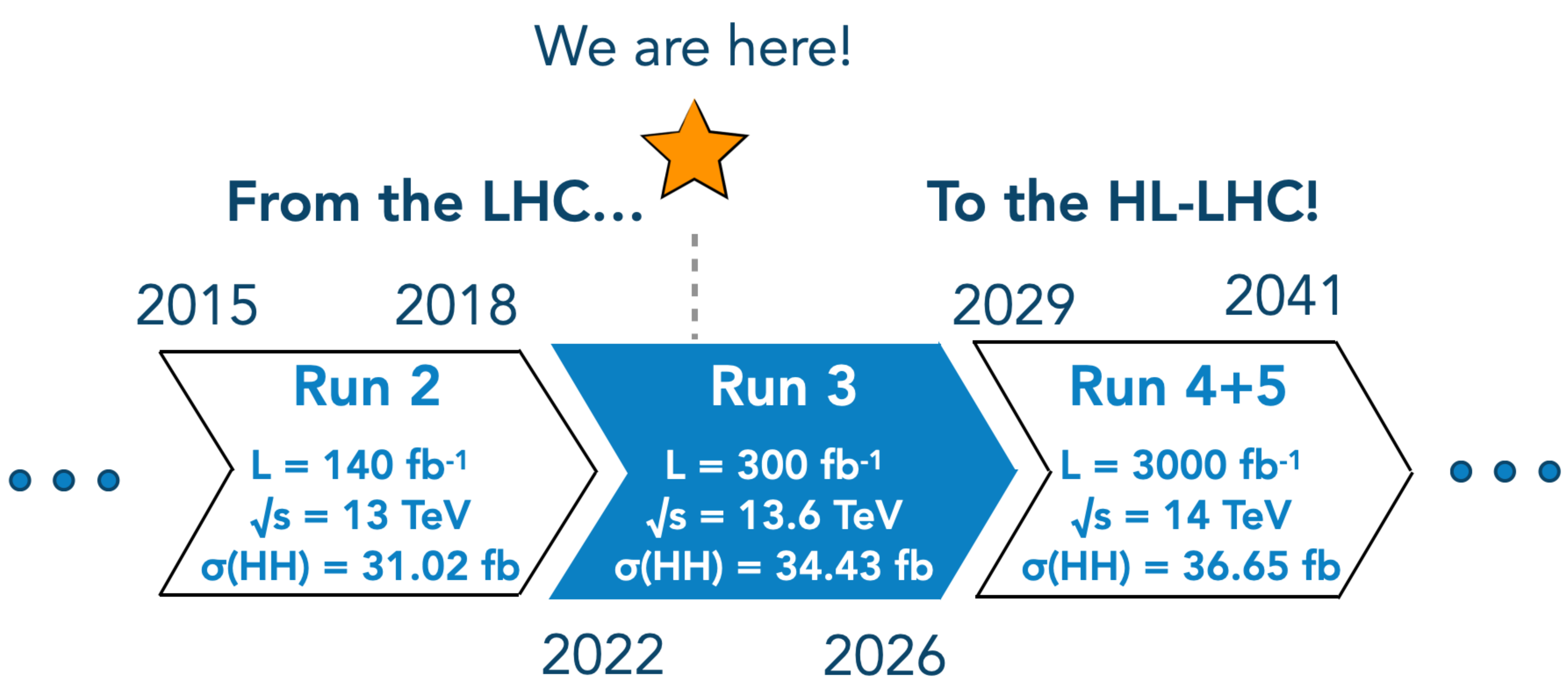
Milano, Meeting di gruppo, 27 Giugno 2024

b-jet tagging: efficiency measurement and impact on $HH \to b\bar{b}\gamma\gamma$ analysis

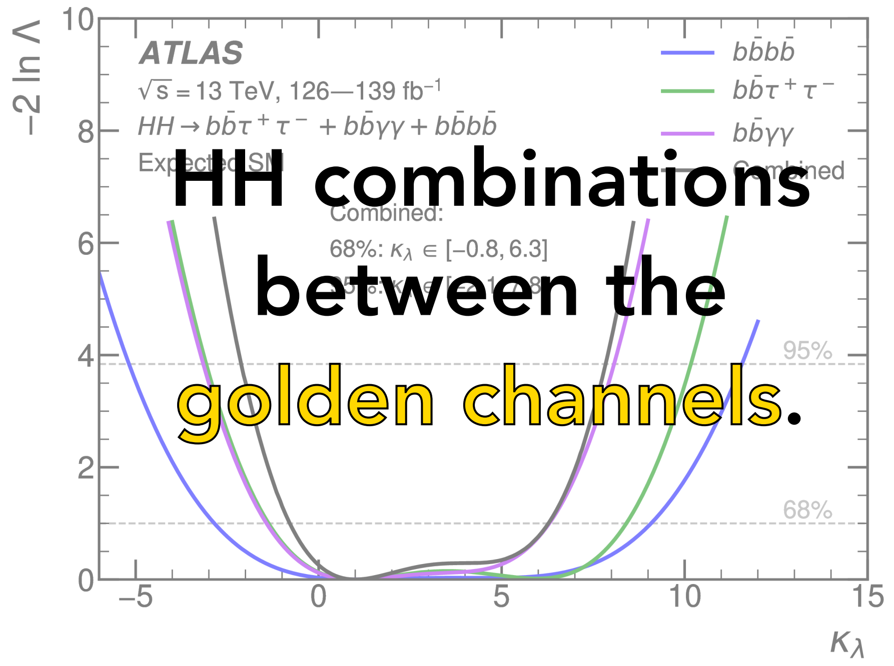
Elena Mazzeo, Leonardo Carminati,
Stefano Manzoni, Ruggero Turra

UNIVERSITÀ DEGLI STUDI DI MILANO
INFN
ATLAS EXPERIMENT

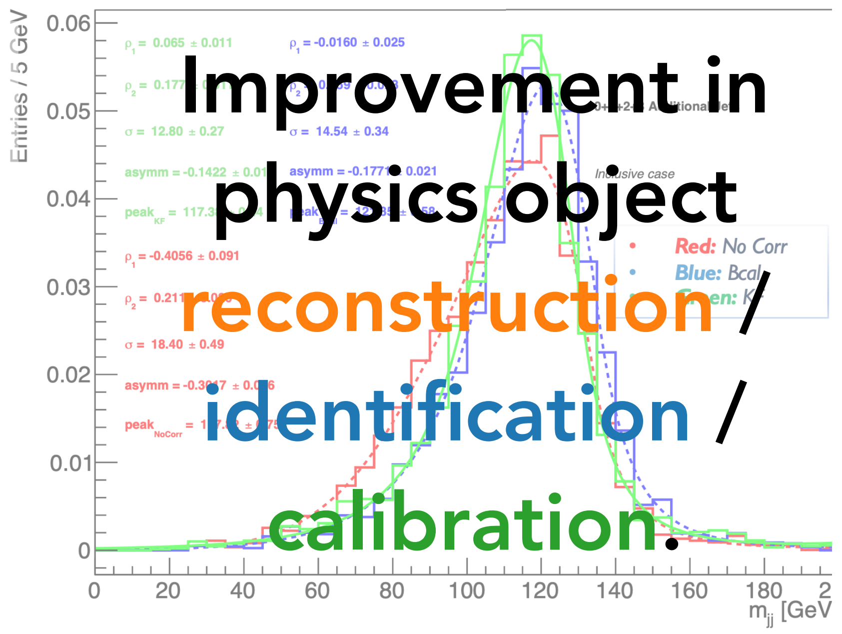
Outline: outlook for HH searches in Run 3



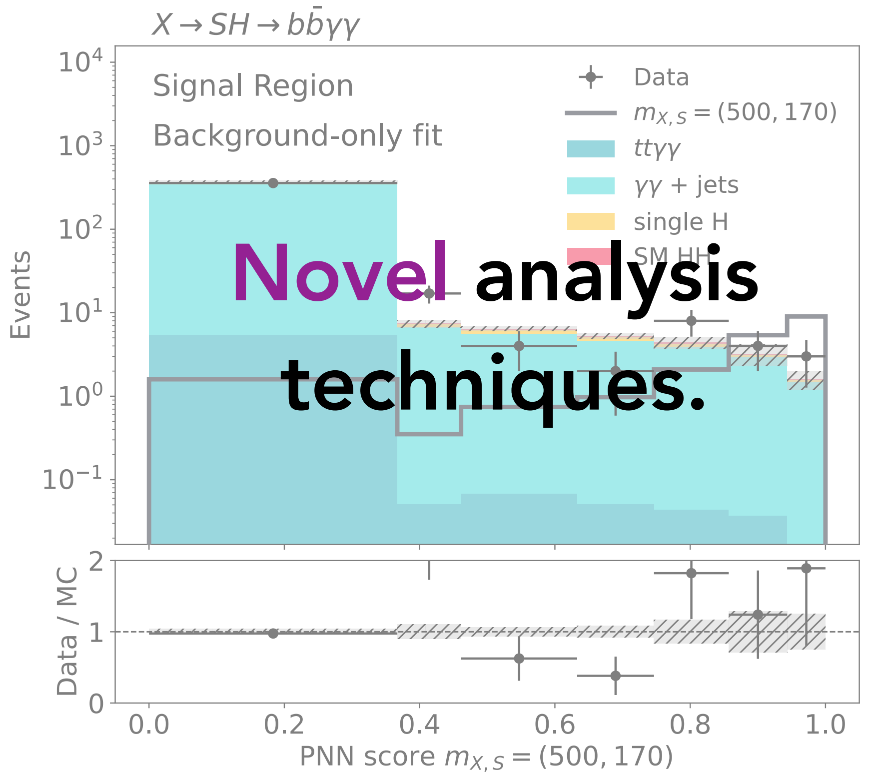
Expected SM HH: 4000 | ~10^4 | ~10^5!



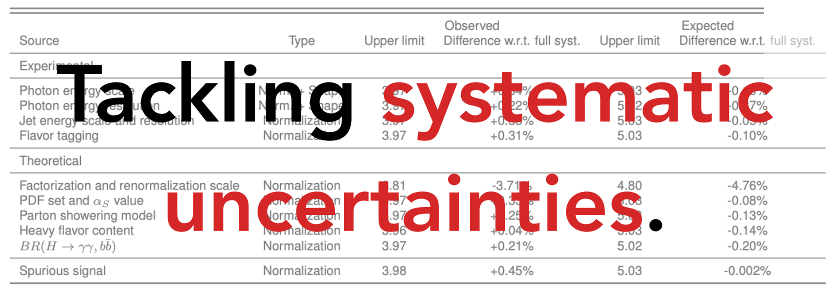
HH combinations between the golden channels.



Improvement in physics object reconstruction / identification / calibration.



Novel analysis techniques.



Tackling systematic uncertainties.

- **Impressive sensitivity to HH production** from the **new batch** of **Legacy Run 2 HH analyses**, that was **finalized** in the past months!
  - $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, and $b\bar{b}b\bar{b}$ (main channels), + $b\bar{b}ll$ and multi leptons + **combination**!

HH combination

	36.1 fb^{-1}	140 fb^{-1}
Expected limit on μ_{HH}	10	2.4

Improvement of a **factor > 4 w.r.t. partial Run 2**!

- **Run 3** represents a **unique opportunity** for **HH searches**!
  - **High momentum** and **increasing interest** from many teams / institutions!
- There are **many aspects** that we can (and should) **explore** to try and **reach** a **first hint of observation of HH production** already with the **full Run 3 dataset**.
  - One of them is the improvement in object identification, especially b-tagging!

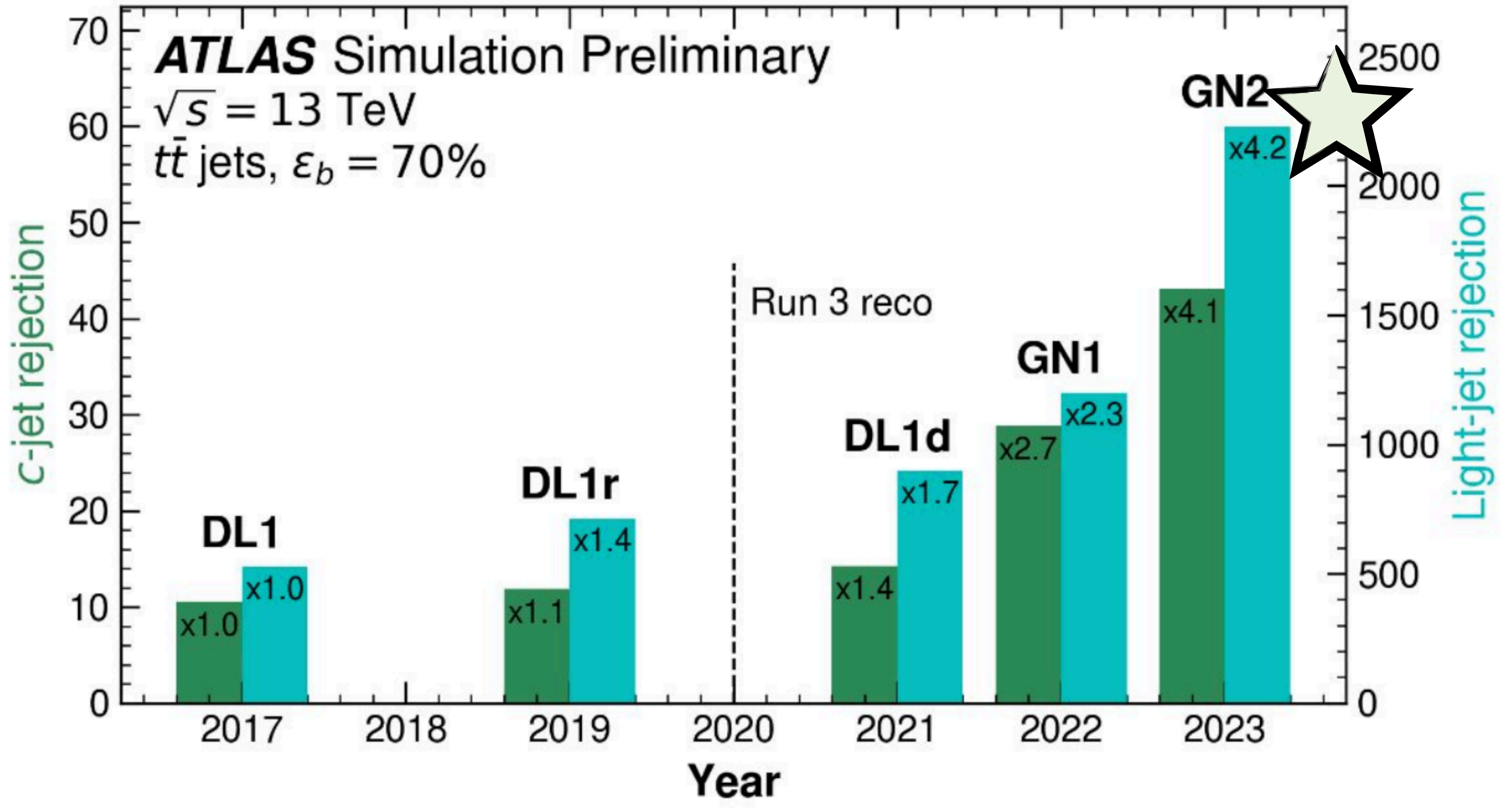
Outline: state of the art b-tagging in ATLAS and HH searches

- A **new generation** of **GNN-based b-tagging algorithms** was developed in ATLAS.

  ➡ **Boost** in **performances** of the new b-tagging algorithms w.r.t. the **older RNN-based generation** used in Run 2 (e.g. DL1r).

- Expected to **sharpen the sensitivity** of the new di-Higgs analyses!

  ➡
  - The **most sensitive channels** (= $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, and $b\bar{b}b\bar{b}$) have **at least 2 b-jets** in the **final state**.
  - Impressive performances in the $HH \rightarrow 4b$ analysis by CMS after the **adoption** of their new **ParticleNet-based tagger already in Run 2**!



GN2 is the current **state of the art** in ATLAS!

	35.9 fb^{-1}	138 fb^{-1}
Expected limit on μ_{HH}	114	5.1

➡ × **11 (!) improvement**, after factoring out the factor 2 improvement from the increase in luminosity.

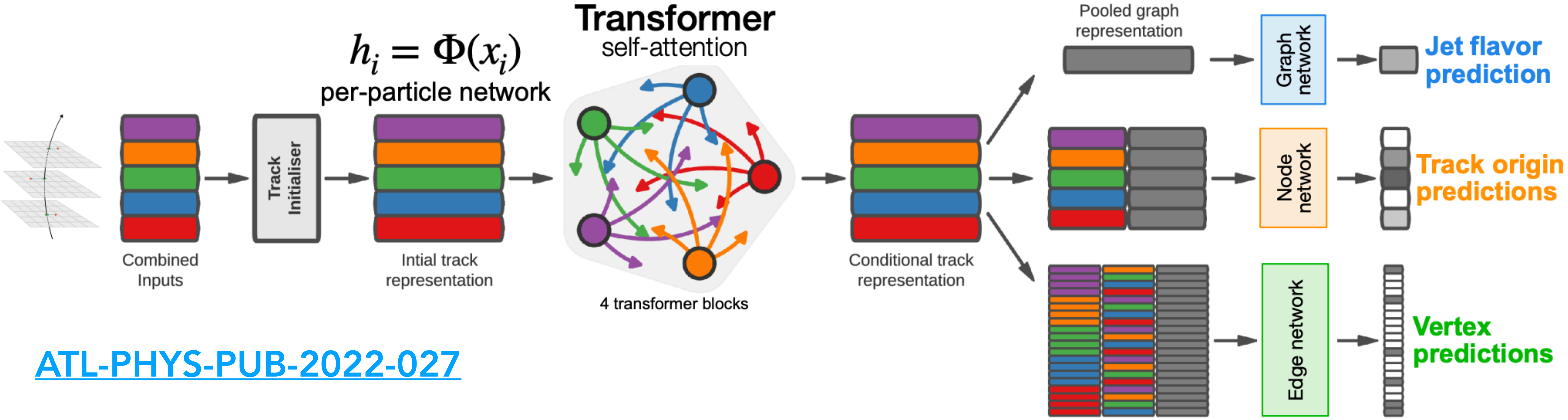
- Before being able to employ the new GN2 tagger in ATLAS physics analysis, we need to calibrate these algorithms (i.e. measuring their efficiency on **real pp collision data**)!

  ➡
  - In this presentation I will try to give an overview of the new GN2 tagger and present the measurement of the b-tagging efficiency on data.
  - Finally, I will try to estimate the impact of adopting the new GN2 tagger, using, as baseline, the Run 2 Legacy $HH \rightarrow b\bar{b}\gamma\gamma$ analysis.

GN2: state of the art b-tagging in ATLAS

b-tagging with GN2: inputs & training

- **GN2** sees a **jet** as collection of **tracks** + some **global information** (= such as its η and ϕ **coordinates**)!
- Trained to **understand** the **internal structure** of the jets, thanks to relying on **low-level information** about **tracks**, **impact parameters**, and **hits**.



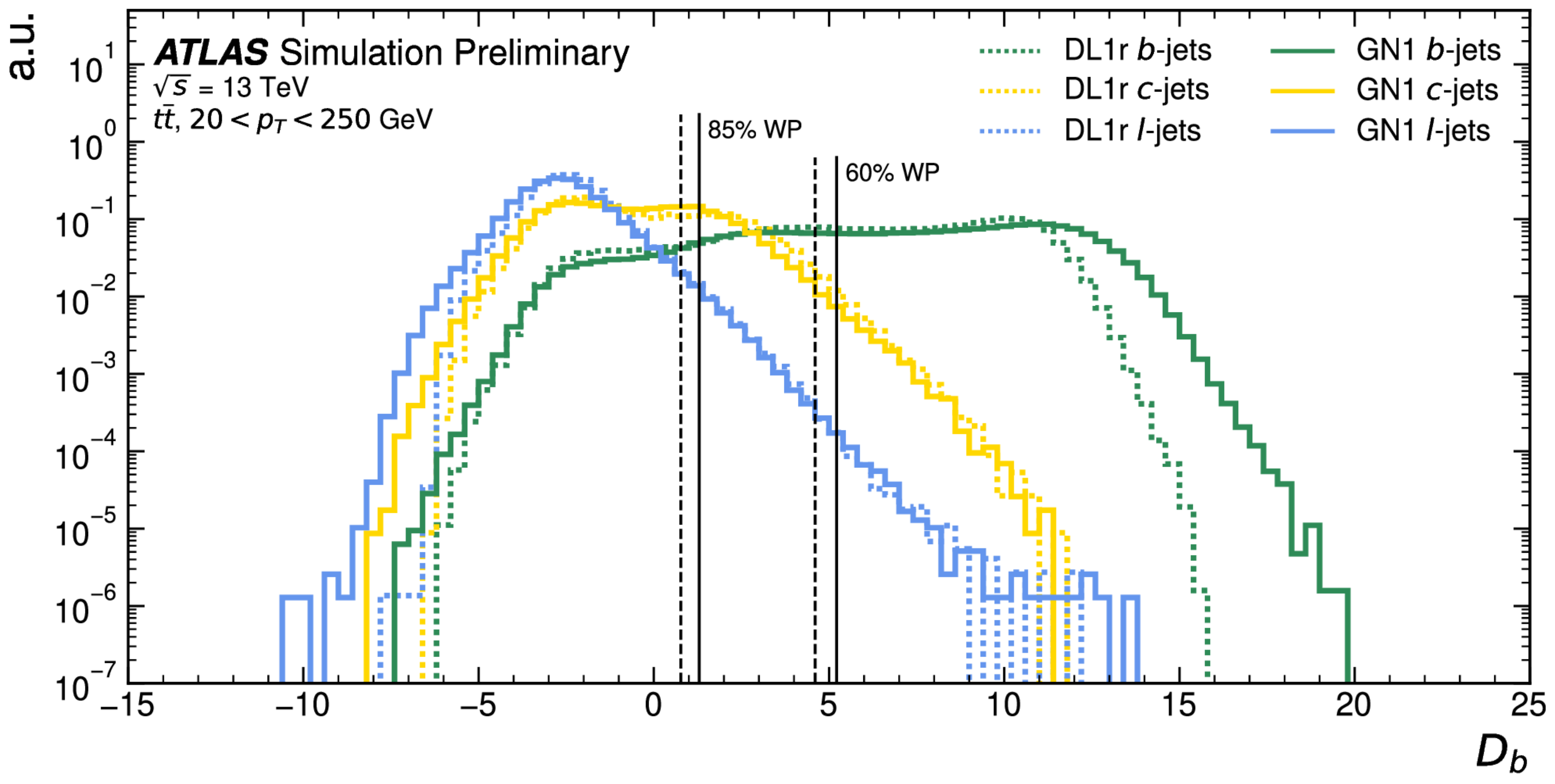
ATL-PHYS-PUB-2022-027

➡ Provides a prediction of the jet's flavour.

➡ 4 output classes: $\begin{pmatrix} p_b \\ p_c \\ p_u \\ p_\tau \end{pmatrix}$ ☆ **New!**

Combined together to define a **discriminant for b-tagging**!

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



By cutting on the D_b discriminant we can define WPs characterized by a given b-tagging efficiency.

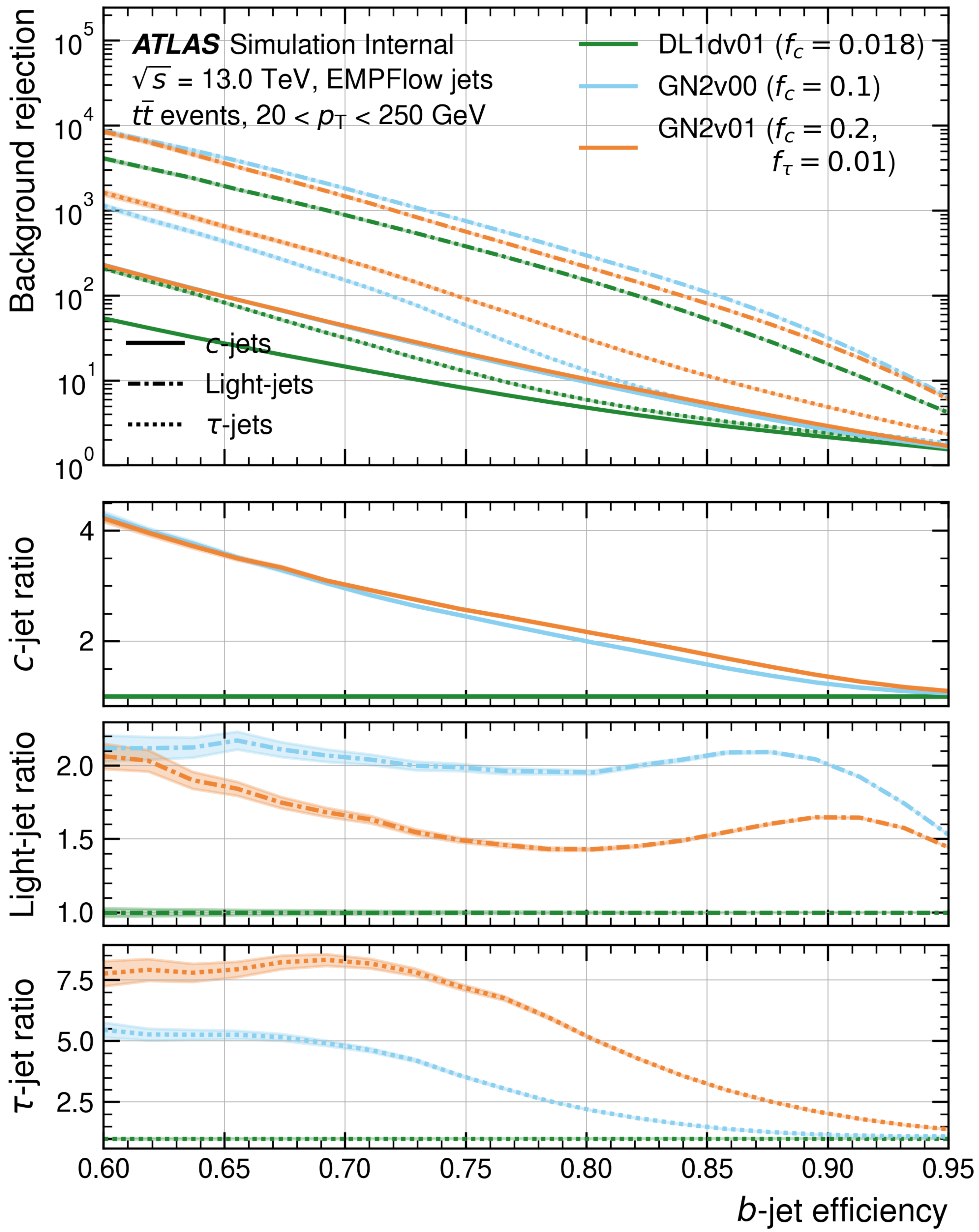
- Measured inclusively on a $t\bar{t}$ MC sample.
- **5 WPs**! ➡ 90%, 85%, 77%, 70%, 65%.

➕ Two **auxiliary tasks**:

1. **Prediction** of the **physics process** initiating the tracks.
2. **Grouping the tracks into vertices**!

➡ Help convergence of the main task.

b-tagging with GN2: performances on MC



The performances of GN2 can be compared with those from the previous taggers by checking the **light** and c**-jet rejection**, given a determined b-jet efficiency.

➡ The new version of GN2 (= GN2v01) is also optimized to reject hadronic τs!

Expected performances at a low p_T regime

➡ Jets from $t\bar{t}$ sample, 20 GeV < p_T < 250 GeV

- **Light jet rejection**: up to a factor × 2 improvement w.r.t. DL1d!
- c**-jet rejection**: up to a factor × 4 improvement w.r.t. DL1d!
- τ**-jet rejection**: up to a factor × 7.5 improvement w.r.t. DL1d!

Calibration of the GN2-based b-tagging efficiency

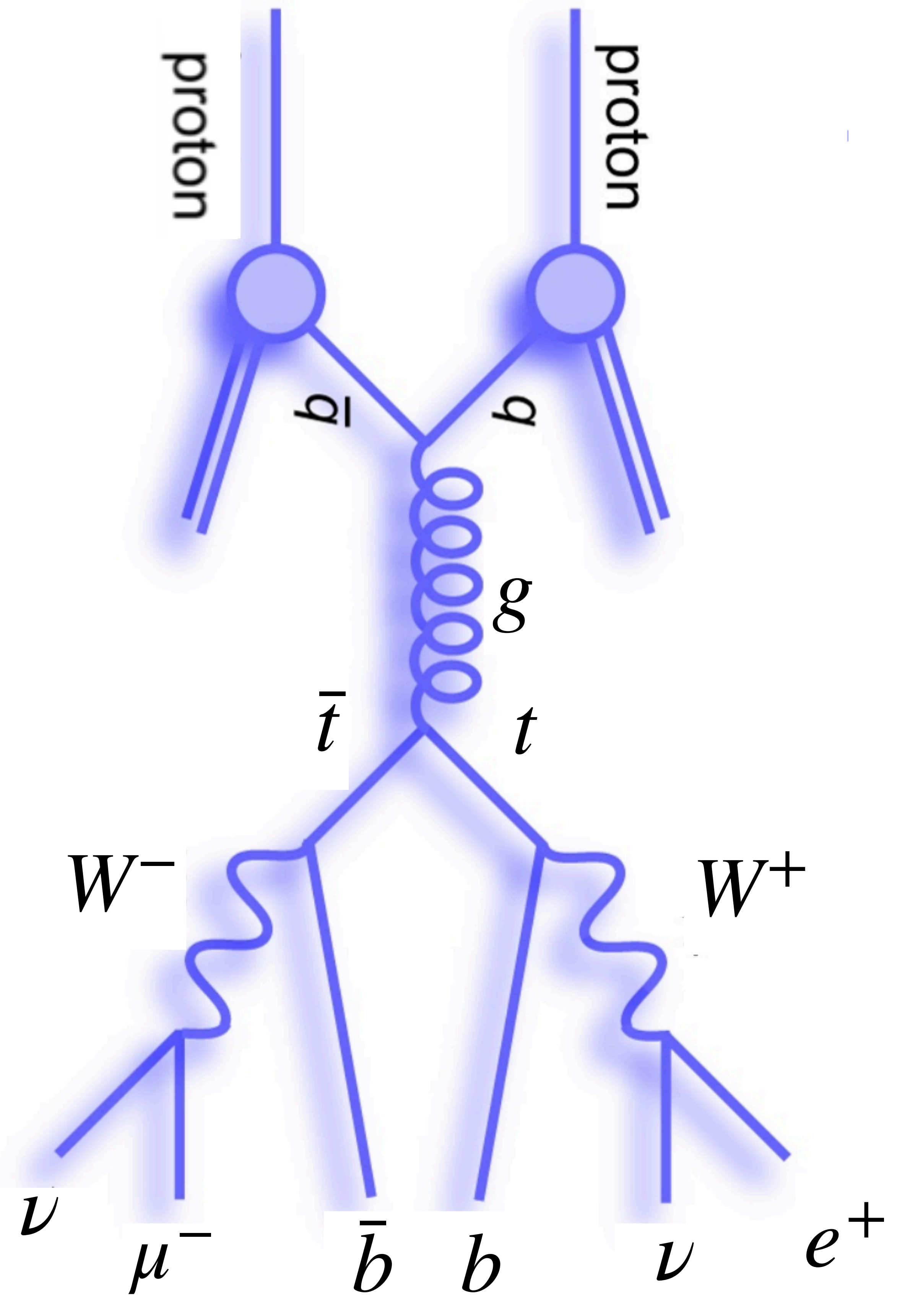
b-tagging efficiency measurement: outline

- Our **taggers (including GN2)** are optimized on **MC events**.
- However, MC samples may be affected by **imperfections** in the **modelling** of the physics process, or in describing the **detector response**.

➡ The **efficiency** has to be measured on **real pp collision data**!

➡ **How?**

1. Select a **pure sample of $t\bar{t}$ events** in data, where **both top** quarks **decay leptonically**.
   ➡ - Provides a sample **enriched** in true b**-jets**, thanks to the $t \rightarrow Wb$ decay.
   - Can be **easily triggered upon** when considering **leptonic W decay**.
2. Identify **backgrounds** and estimate **systematic uncertainties**.
   ➡ - Our leptonic $t\bar{t}$ signal includes **two b-jets** and **two leptons** in the final state.
   - The **fake b-jet backgrounds** are estimated with the help of **CRs**, enriched in ll, bl, lb events.
3. Extract b**-tagging efficiency** (= selected b-jets / all b-jets) via a **maximum-likelihood fit** of the discriminant distribution **to data**, as a function of the leading and subleading jet p_T and b**-tag efficiency WP**.
   ➡ **Scale factors** = b-tagging efficiency in data / b-tagging efficiency in MC.

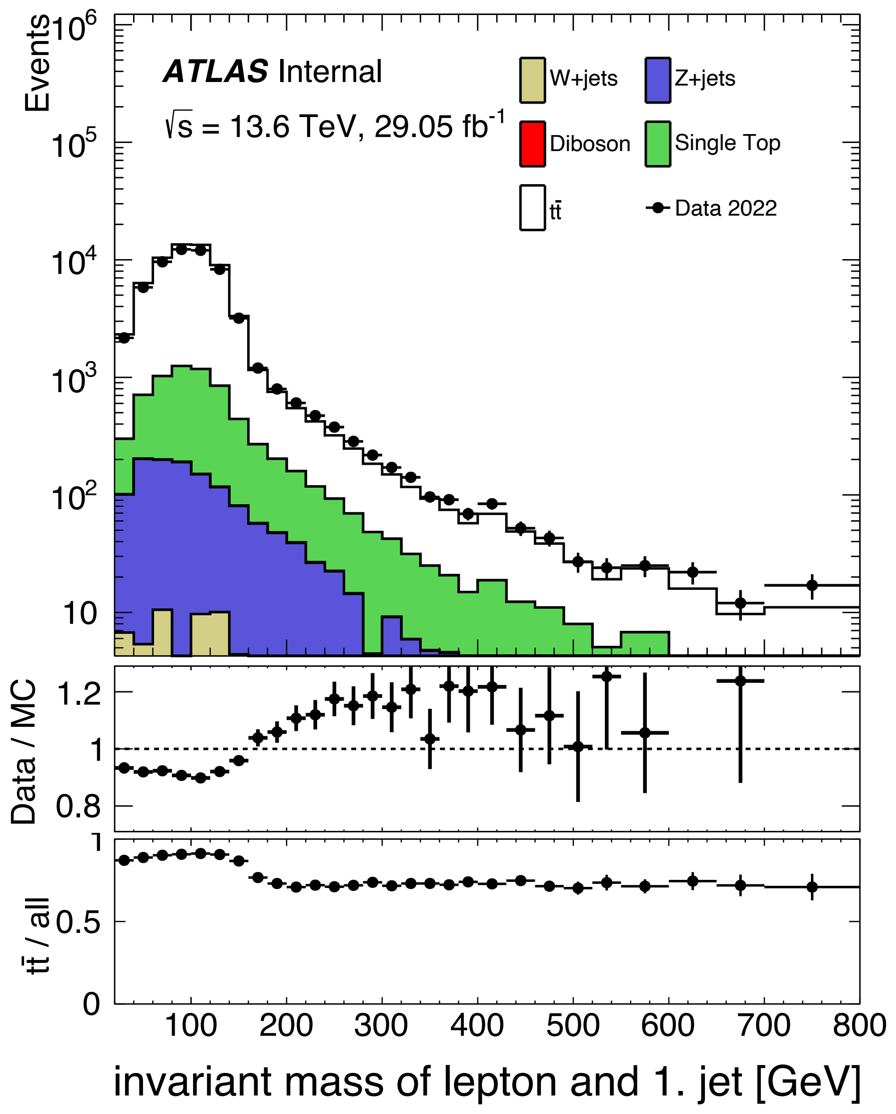


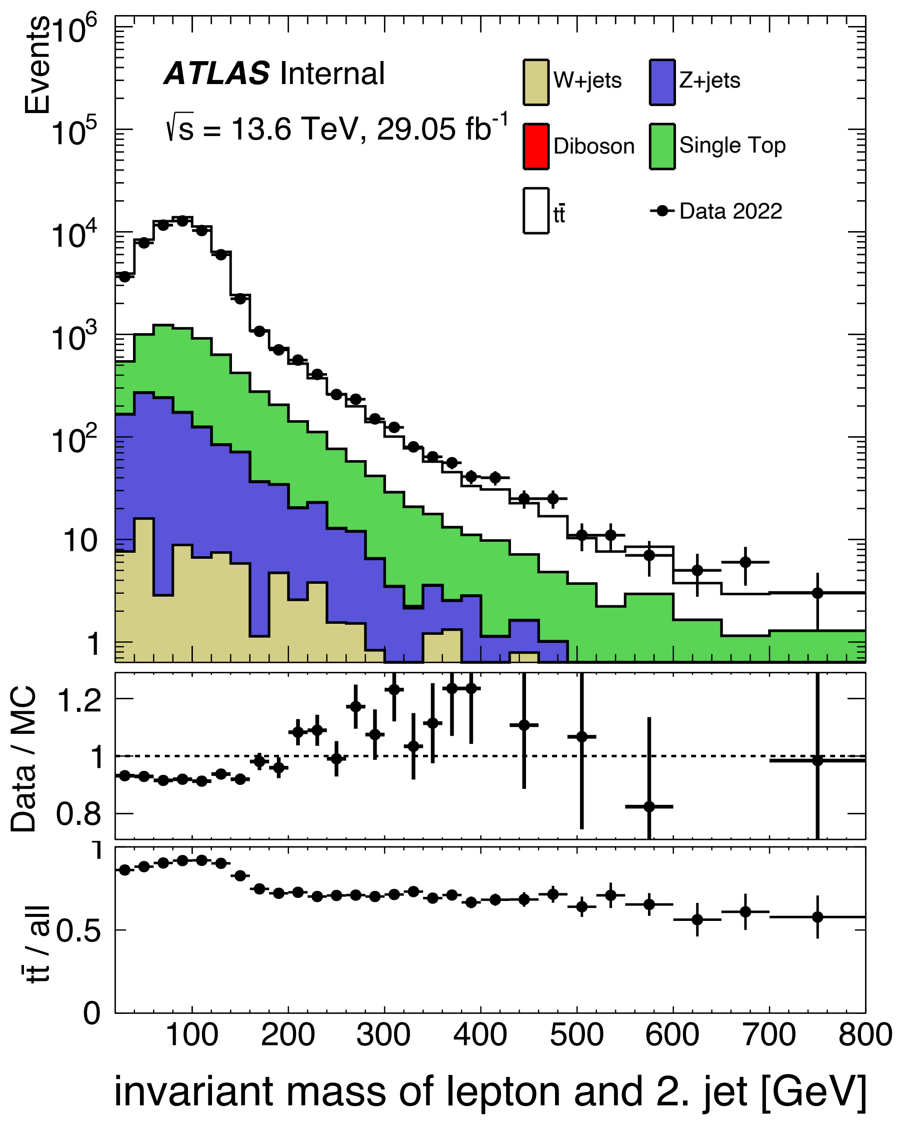
Event selection

Main backgrounds: W+jets, Z+jets, Single top production, and di-boson production.

- **OR** of **single-lepton** (electrons or muons) **triggers**.
- **Exactly two** reconstructed **opposite charge** and **different flavor leptons** (e and μ) with $p_T > 27$ GeV. → Reduces background from $Z \to \ell\ell$ **events**.
- $m_{e\mu} >$ 50 GeV. → Reduces background from $Z \to \tau^{lep}\tau^{lep}$ **events**.
- Exactly **two jets** with $p_T > 20$ GeV. → Reduces background from **single W production** and light jet production from ISR and FSR.

Selected events are then split in **bins** of the **leading** and **subleading jet** p_T, and then further in a **SR** (enriched in events with two **true b-jets**) and in three **CRs** (enriched on bl, lb, ll events), based on the variables $m_{j_1\ell}$ ad $m_{j_2\ell}$.





- The **two jets** are paired with the **two leptons**, in the configuration that minimizes $m^2_{j_1\ell} + m^2_{j_2\ell}$ (= penalizes configurations where $m_{j_1\ell}$ and $m_{j_2\ell}$ are asymmetric).

  m_{jl} is a proxy of the **top mass**! → For true b-jets, $m_{j\ell} \leq m_t \approx$ 172.5 GeV.

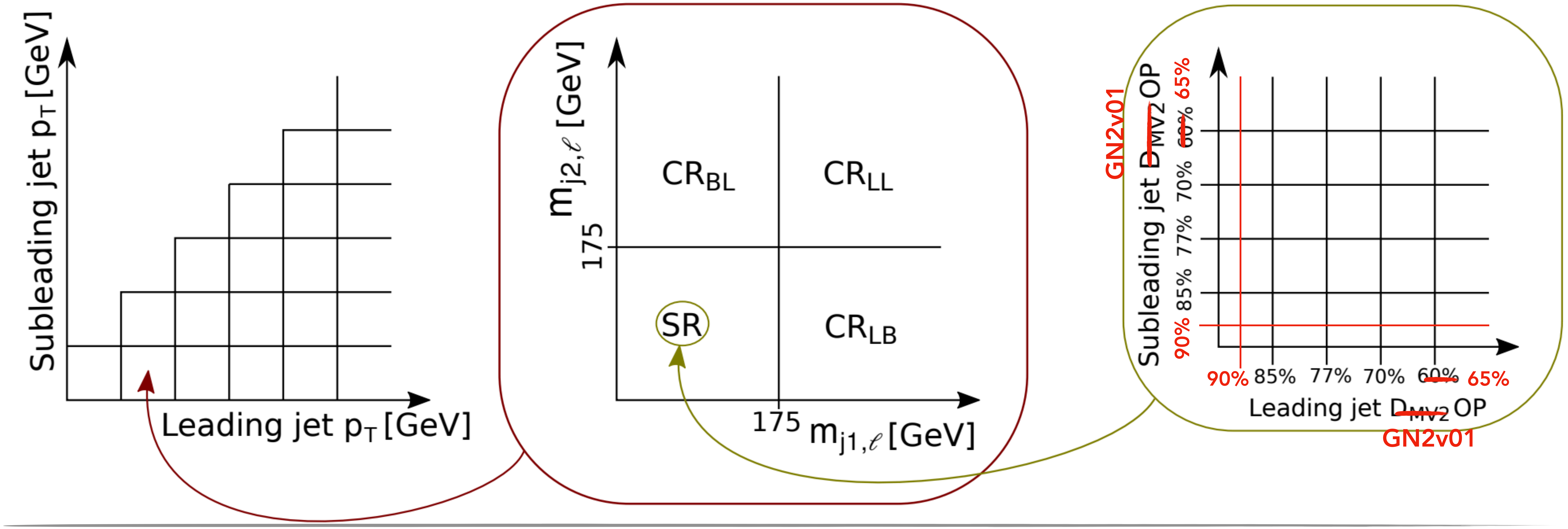
- In all regions, $m_{j_1\ell} > 20$ GeV and $m_{j_2\ell} > 20$ GeV.
- SR = $m_{j_1\ell} < 175$ GeV and $m_{j_2\ell} < 175$ GeV.
- CR$_{BL}$ = $m_{j_1\ell} < 175$ GeV and $m_{j_2\ell} \geq 175$ GeV.
- CR$_{LB}$ = $m_{j_1\ell} \geq 175$ GeV and $m_{j_2\ell} < 175$ GeV.
- CR_{LL} = $m_{j_1\ell} \geq 175$ GeV and $m_{j_2\ell} \geq 175$ GeV.

b-tagging efficiency measurement: fitting strategy

The b-tagging efficiency is extracted from a **maximum likelihood fit**, performed simultaneously in the SR and in the three CRs.

	Run 2	Run 3
Bins in pT(j1) and pT(j2)	[20, 30, 40, 60, 85, 110, 140, 175, 250, 400] GeV	[20, 40, 60, 140, 250, 400] GeV
GN2v01 eff. WPs	[90%, 85%, 77%, 70%, 65%]	[90%, 85%, 77%, 70%, 65%]
Total number of bins	(9 x 10) / 2 × (3 + 1 × 6 × 6) = 1755 total bins. (9 x 10) / 2 × (6 × 6) = 1620 SR bins.	(5 x 6) / 2 × (3 + 1 × 6 × 6) = 585 total bins. (5 x 6) / 2 × (6 × 6) = 540 SR bins.

= 4 regions, 5 bins in GN2 score, and 10 or 6 bins in p_T × 2 jets.



b-tagging efficiency measurement: fitting strategy

Extended binned likelihood:

$$\mathscr{L}(\nu(\theta)) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}$$

- N = total number of bins.
- ν_i (n_i) = observed (expected) number of events in the i-th bin.

- In the **CRs**, the expected number of events depend from the leading jet p_T bin T^m and the subleading jet p_T bin T^n, and receives contributions from each of the four flavor fractions bb, bl, lb, and ll.

$$\nu_{CR}(T^m, T^n) = c_{bb}^{m,n}\nu_{CR,bb}^{m,n} + c_{bl}^{m,n}\nu_{CR,bl}^{m,n} + c_{lb}^{m,n}\nu_{CR,lb}^{m,n} + c_{ll}^{m,n}\nu_{CR,ll}^{m,n}$$

- In the **SRs**, the expected number of events depend also from the b-tagger eff. WP O^k (O^p) of the (sub) leading jet!

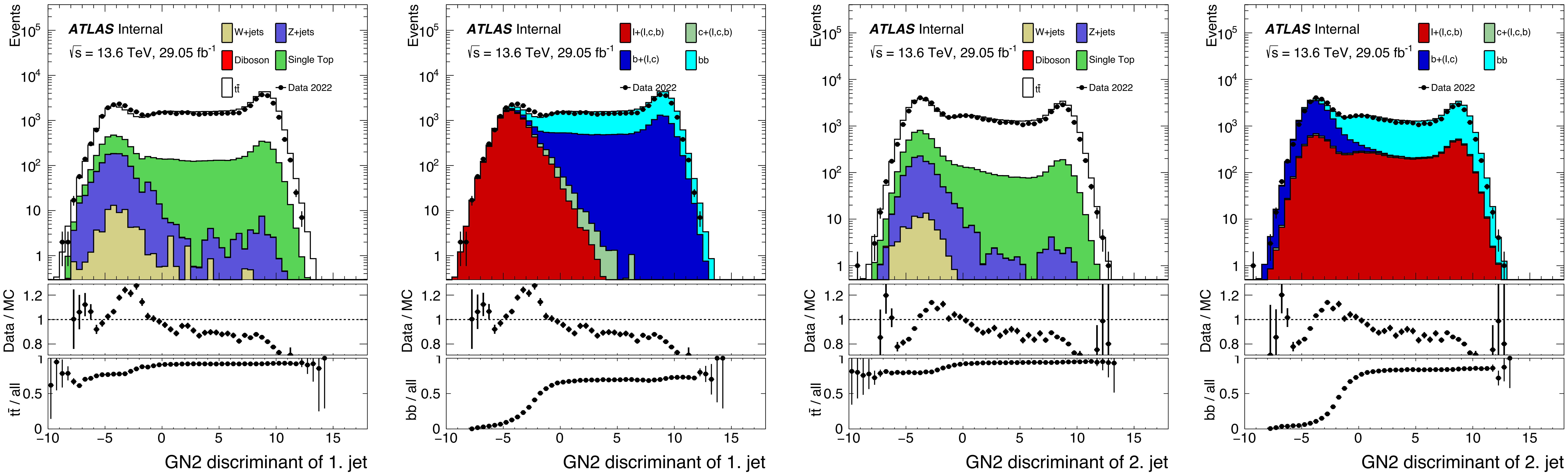
$$\begin{aligned}\nu_{SR}(T^m, T^n, O^k, O^p) = & c_{bb}^{m,n}\nu_{SR,bb}^{m,n} \cdot \mathscr{P}_b(O^k | T^m) \cdot \mathscr{P}_b(O^p | T^n) \\ & + c_{bl}^{m,n}\nu_{SR,bl}^{m.n} \cdot \mathscr{P}_b(O^k | T^m) \cdot \mathscr{P}_l(O^p | T^n) \\ & + c_{lb}^{m,n}\nu_{SR,lb}^{m.n} \cdot \mathscr{P}_l(O^k | T^m) \cdot \mathscr{P}_b(O^p | T^n) \\ & + c_{ll}^{m,n}\nu_{SR,ll}^{m.n} \cdot \mathscr{P}_l(O^k | T^m) \cdot \mathscr{P}_l(O^p | T^n)\end{aligned}$$

- $\nu_{CR(SR)ij}^{m,n}$ = constant param. → Yields of events with true jet flavors i and j in each p_T bin (from MC).
- $c_{ij}^{m,n}$ = unconstrained NPs. → Correction factor for the yields of the different flavor fractions in each p_T bin.
- $\mathscr{P}_b(O^k | T^m)$ = POIs. → Probability that a true b-jet with p_T bin = T^m is selected by the O^k eff. WP.
- $\mathscr{P}_l(O^k | T^m)$ = constant param. Probability that a true light / c-jet with p_T bin = T^m is selected by the O^k eff. WP.

b-tagging efficiency measurement: data / MC agreement

- As a first step, we had a look at the **data / MC agreement**, for some interesting variables in our analysis (= leading and subleading jet's η and p_T, the m_{jl} variable, and the GN2 discriminant).
- For the MC, we tried to look at the different contributions from both the **physics process decomposition** (= $t\bar{t}$, single top, Z+jets, W+jets, and di-boson samples), and the **jet's flavor decomposition**.

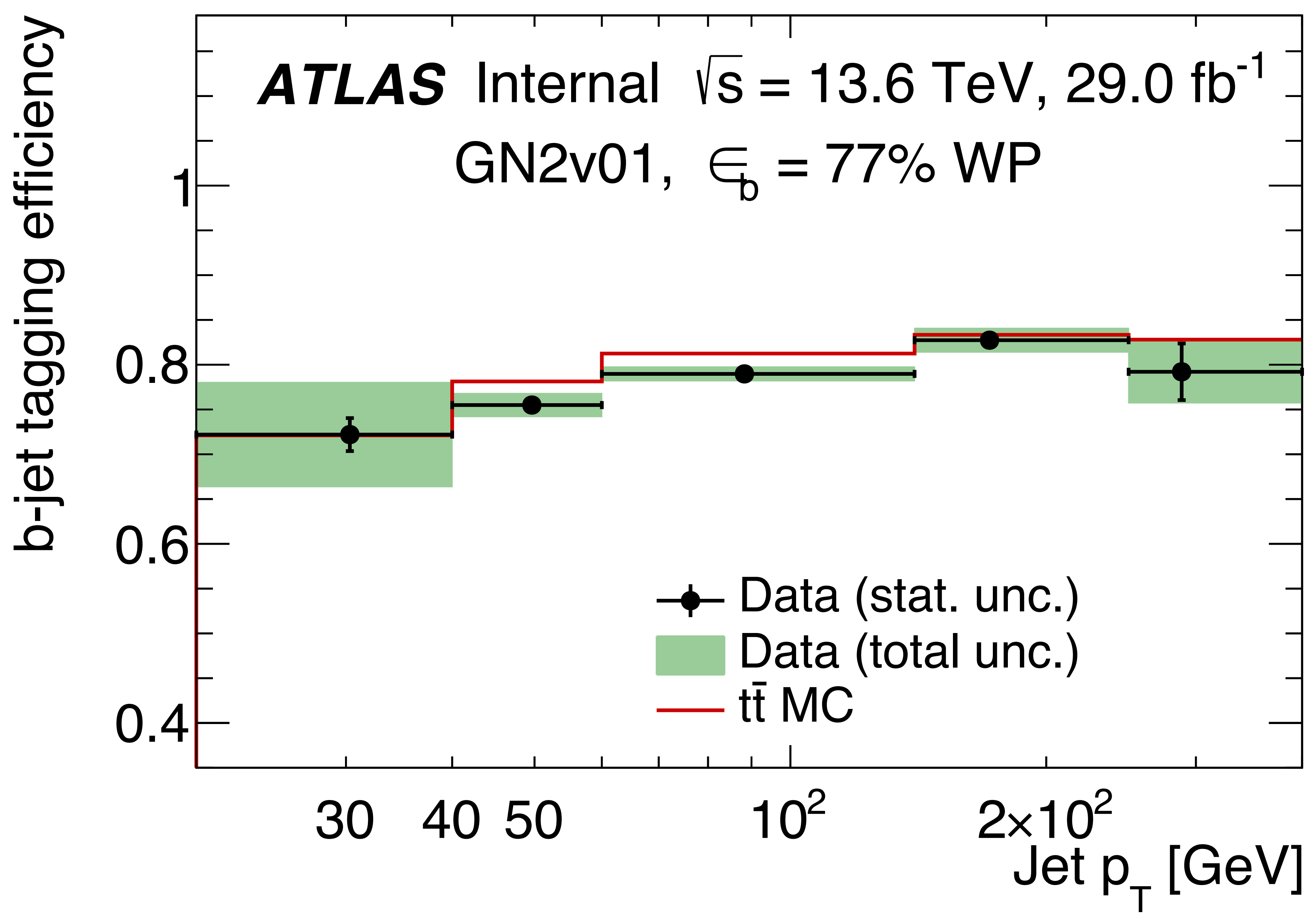
Other variables are shown in backup slides!

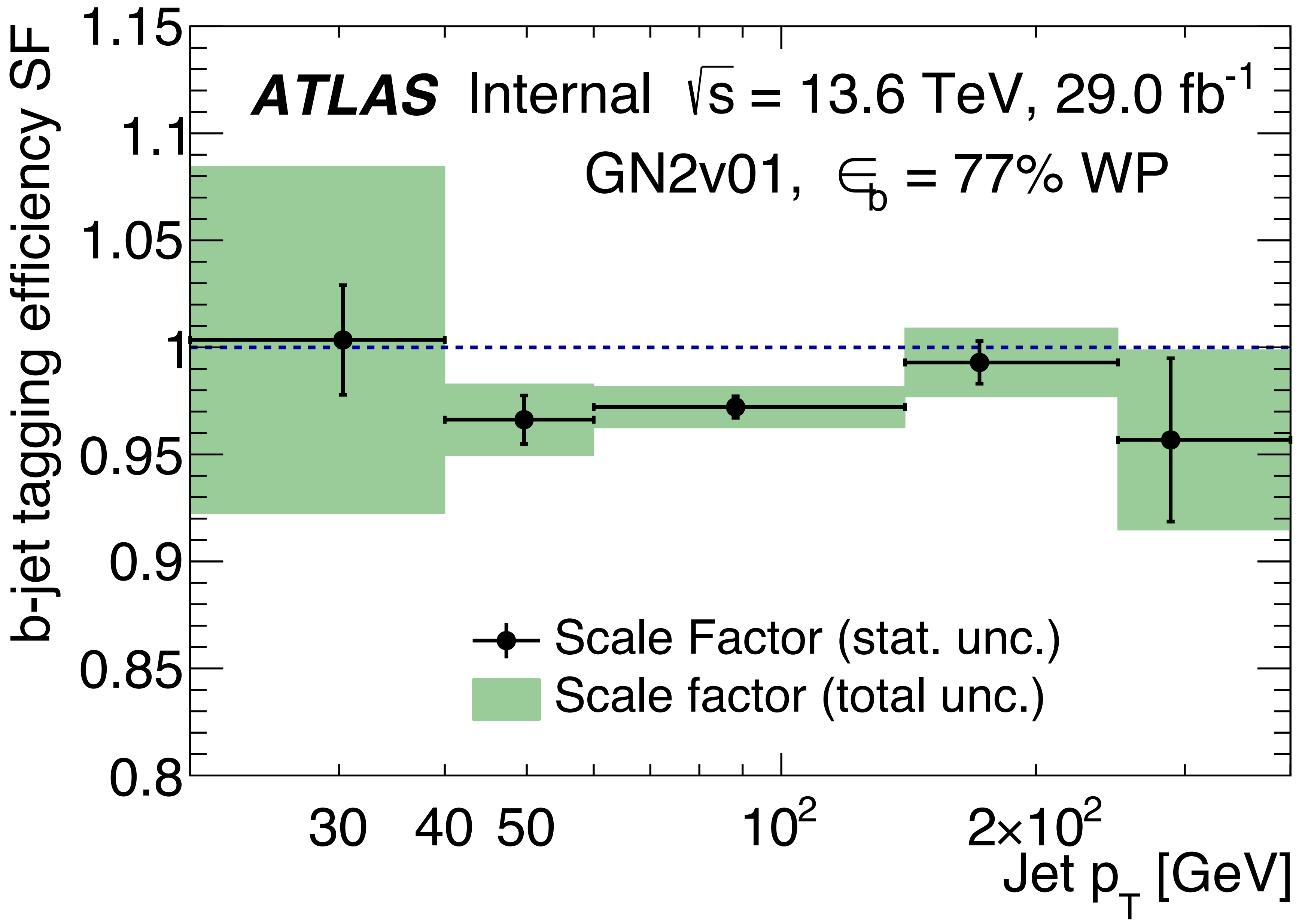


b-tagging efficiency measurement: Run 3 (2022 data)

- The shape of the GN2 discriminant, split between the contribution to each true jet flavor fraction for the leading and subleading jet (= bb, bl, lb, ll), is the final observable in our fit!
- After extracting the histograms of the GN2 discriminant in the analysis regions, we are ready to look at the b-tagging efficiency, and to the Scale Factors.

b-tagging efficiency and SFs as a function of the jet p_T, for the 77% eff. WP of GN2





b-tagging efficiency measurement: systematic uncertainties

The following systematic uncertainties (that may affect our results) are considered in the fit.

➡ Almost all included for the Run 3 results!

Theory	**Modelling uncertainties**	• Alternative PDF set correlated tt-bar and single top (difference between `NNPDF30_nlo_as_0118` and `PDF4LHC`). • Matrix element unc. for tt-bar (hdamp and pthard parameters). • Parton shower for tt-bar (difference between Pythia8 and Herwig7). • Modelling unc. for single top difference between DS and DR).
	Scale variations, PDF variations, ISR & FSR	• Scale and PDF + α_S for tt-bar and single top. • ISR and FSR for tt-bar and single top.
Exp.	**Physics objects, triggers, pileup.**	• Pile-up modelling; • Electron and muon trigger efficiency; • Electron energy scale and resolution; • Jet energy scale and resolution; • Jet vertex tagger efficiency; → For now excluded, since negative weights variations cause problems in the fit stability (= under investigation). • Muon energy calibration and identification; • Missing energy reconstruction.
	Custom	• Fake lepton correction. • **Light flavor jet mis-tagging efficiency**.

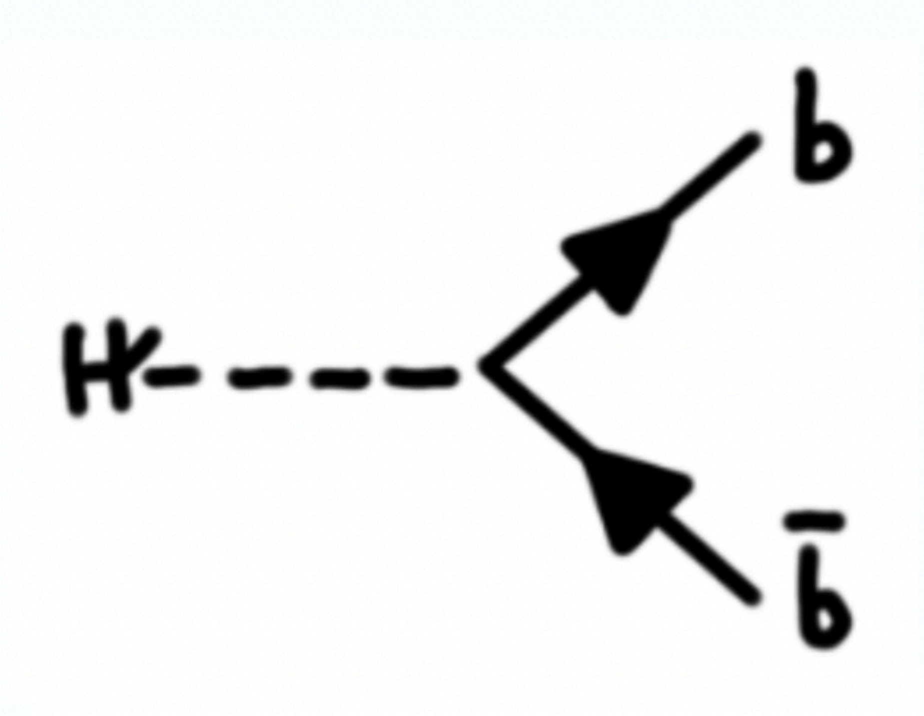
- We assigned a **preliminary 100%** uncertainty on the **light jet efficiency** (= used as input in our fit).

  ➡ **Strong impact** at low p_T in **high efficiency bins** (e.g. 90% WP).

- **Will change**, once the **light jet mis-tagging efficiency measurement** is **ready**!

Estimating the impact of GN2-based b-tagging for $HH \to b\bar{b}\gamma\gamma$ analysis

b-jet requirements for the $HH \to b\bar{b}\gamma\gamma$



- Run 2: exactly 2 b-jets using the **77% efficiency WP** of **DL1r**.
- Run 2 + partial Run 3: exactly 2 b-jets using the **77% efficiency WP** of **GN2**.

Our **signal** includes **two true b-jets**!

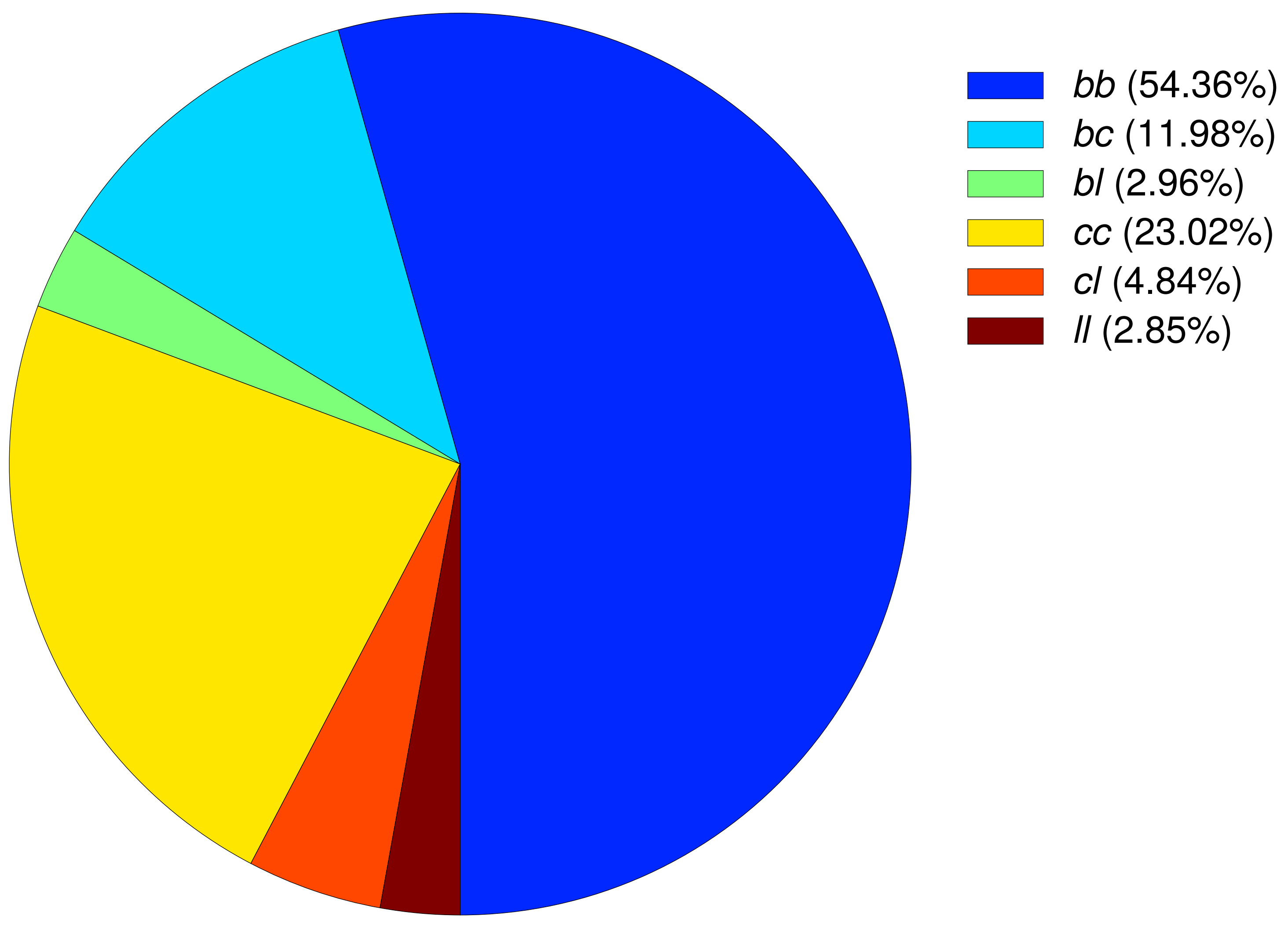
Same signal efficiency but **better bkg. rejection** with new **b-tagging**.

Improved **rejection** of **bkg. events** with **light and c-jets misidentified** as **b-jets**.

Question: how do we quantify the impact of **improved fake b-jet rejection** in the $HH \to b\bar{b}\gamma\gamma$ analysis?

1. Estimate the **contribution of bkg. events** with **true light** or **c-flavor jets** misidentified as b-jets.
   - **Replacing** the **DL1r 77% WP** with the **GN2 77% WP** will help to **suppress** the **fake b-jets backgrounds**!
2. Given the 77% signal efficiency, **rescale** the **contributions** of **each true jet flavors** with the ratio between the DL1r efficiency and the GN2 efficiency.
   - Estimated using ROC curves of the two taggers!
3. Estimate the **new bkg. yields**, and evaluate the **impact** on the **upper limits on the HH signal strength**.

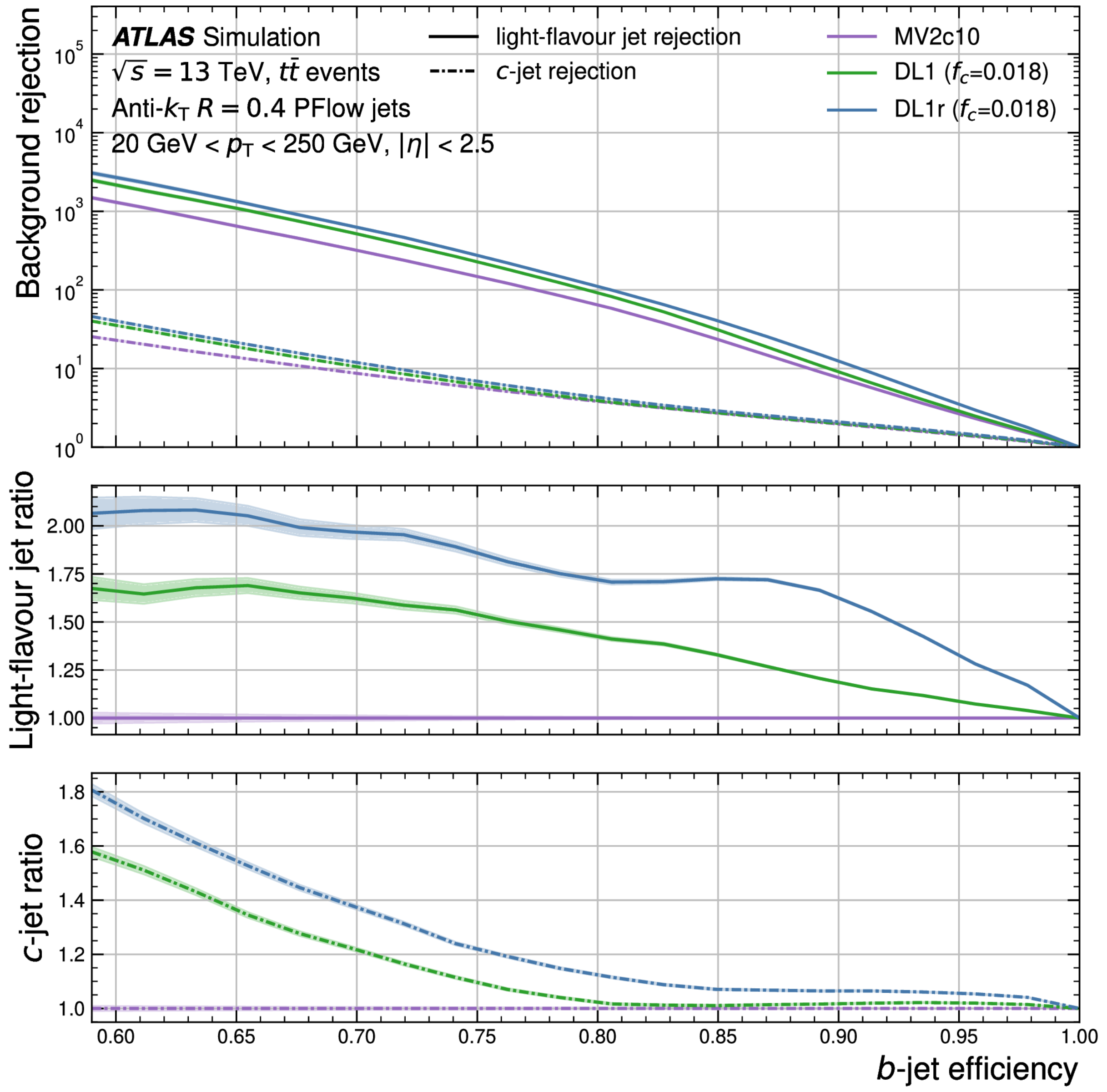
$\gamma\gamma$+jets true jet flavor fractions.

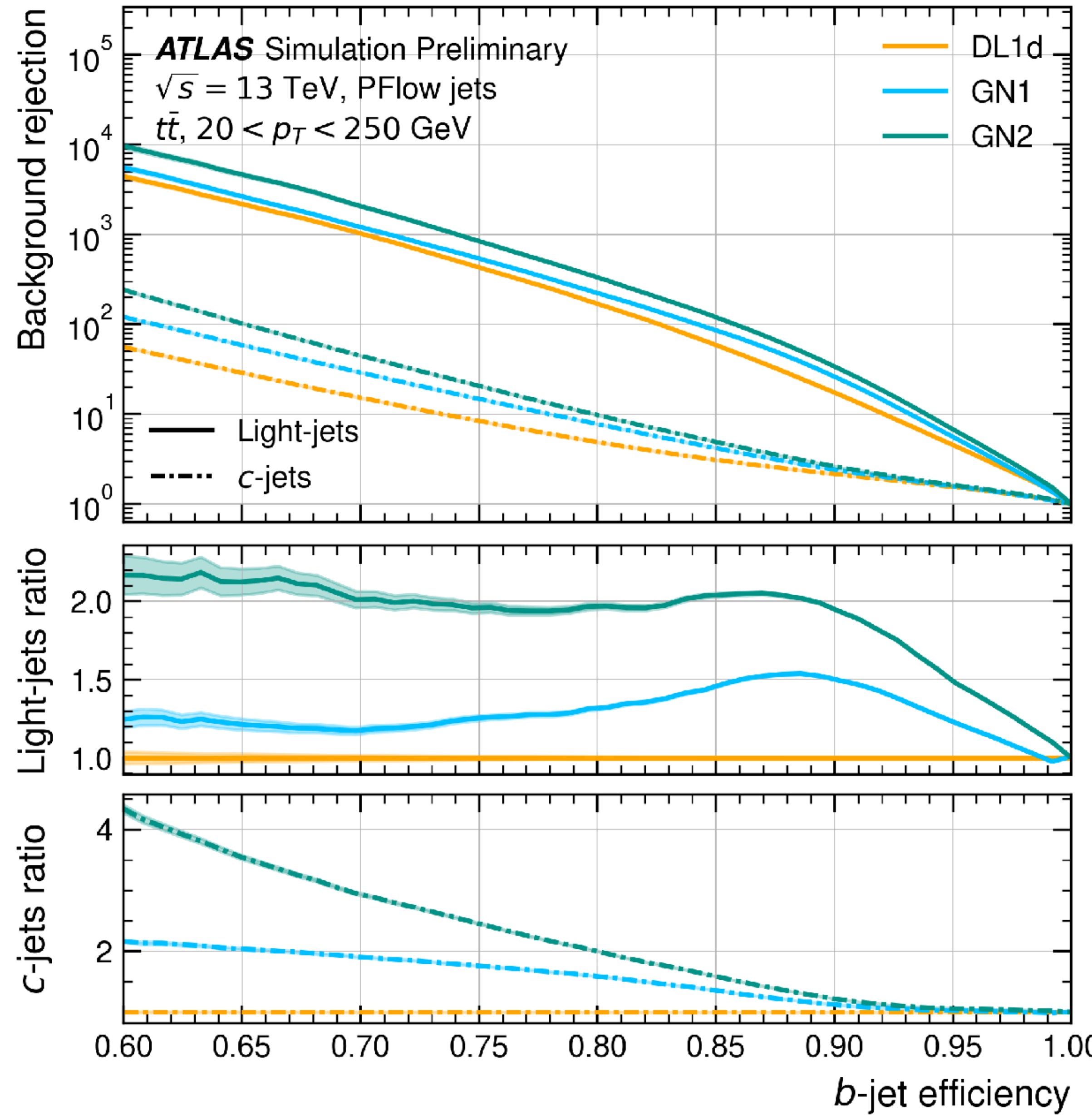


c- and light jet rejection with DL1r and GN2

Tagger bkg. efficiency @ 77% b-jet eff. WP

	c-jet efficiency	Light jet efficiency
DL1r	1/6 = 16.6%	1/300 = 0.33%
GN2	1/12 = 8.3%	1/700 = 0.14%



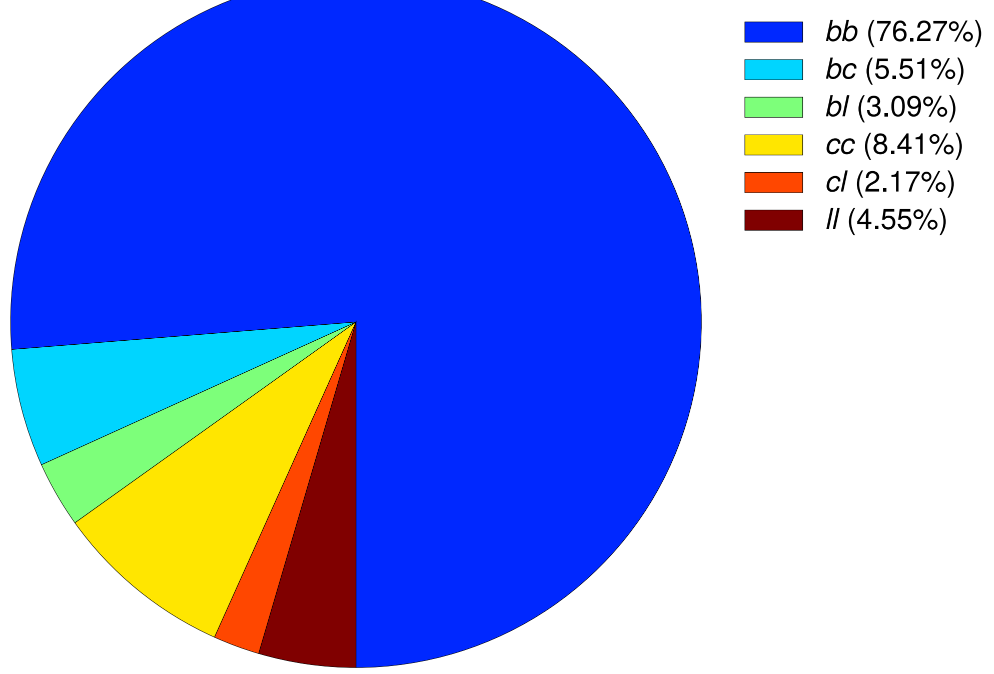


- Given the selected events with **two reco b-jets**, we can evaluate the ratio of the efficiency between DL1r and GN2 WPs for each true flavor fraction.
- When rescaling the true jet flavor contributions with the GN2 / DL1r efficiency ratios, both the true jet flavor composition of the selected (bkg.) events will change, and the overall yields in each analysis category! → Will have an impact on the HH significance!

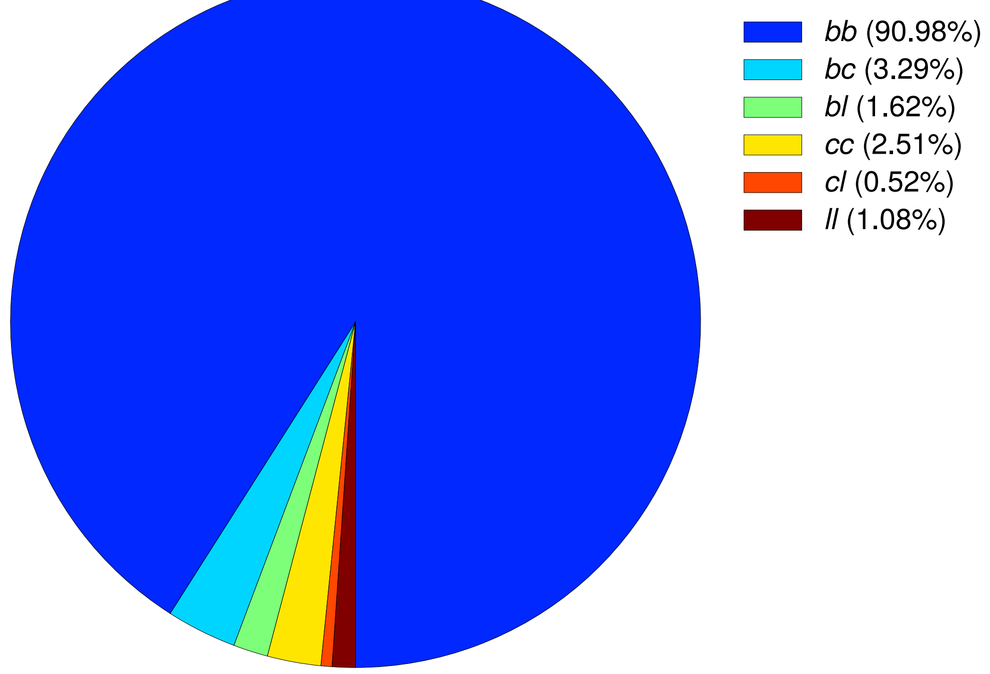
⇨ The **True bb / All purities** for each background samples in each analysis categories are available in backup slides!

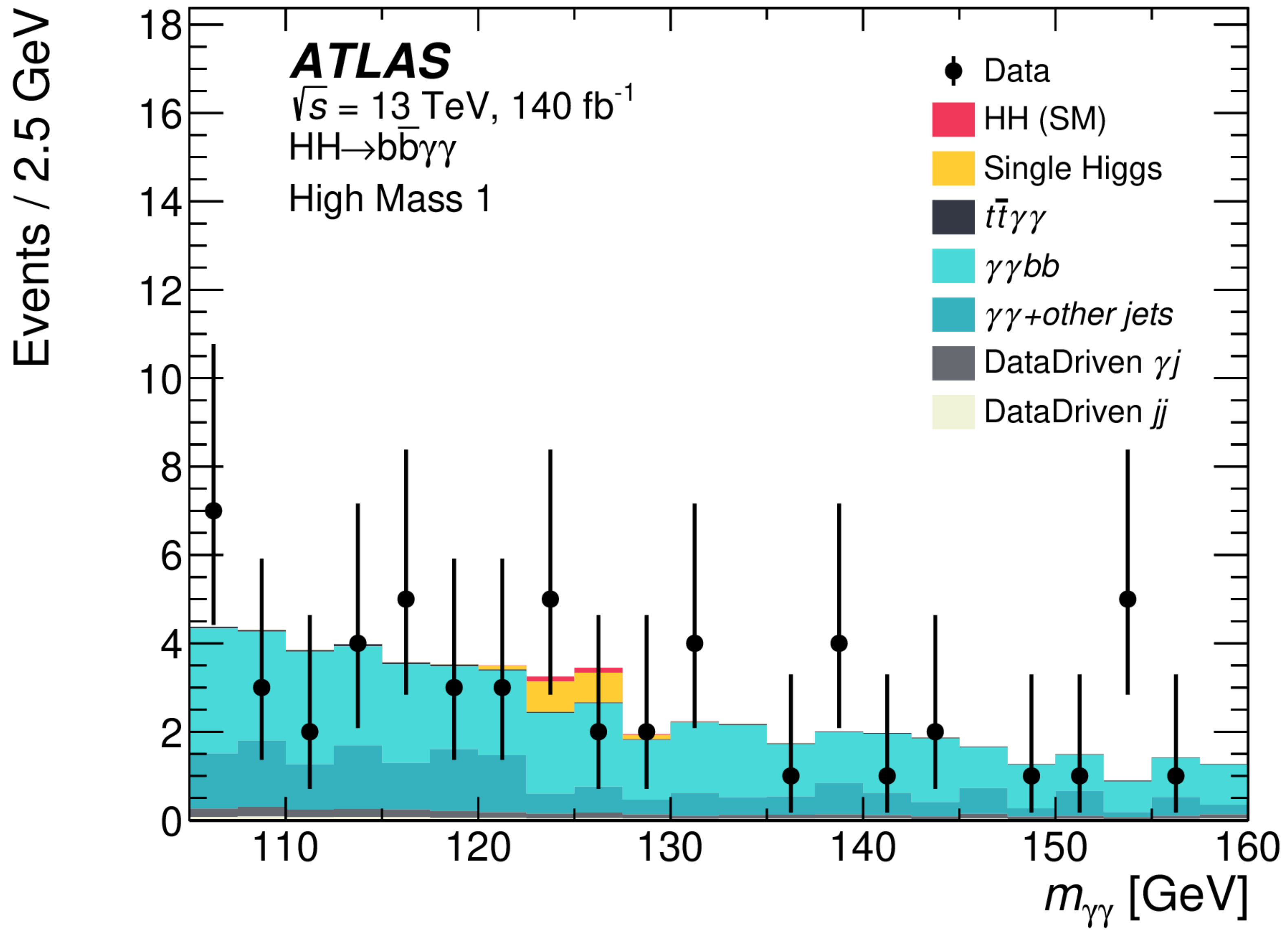
True jet flavor	ϵ (2 reco b | true ij): GN2	ϵ (2 reco b | true ij): DL1r	GN2 / DL1r
bb	77% × 77% = 59%	77% × 77% = 59%	1
bc	77% × 8.3% = 6.4%	77% × 16.6% = 12.8%	0.5
bl	77% × 0.14% = 0.11%	77% × 0.33% = 0.25%	0.44
cc	8.3% × 8.3% = 0.69%	16.6% × 16.6% = 2.76%	0.25
cl	8.3% × 0.14% = 0.01%	16.6% × 0.33% = 0.05%	0.2
ll	0.14% × 0.14% = 0.0002%	0.33% × 0.33% = 0.001%	0.2

ggH sample (DL1r 77% WP)



ggH sample (GN2 77% WP)





Two main **backgrounds**:

1. **Resonant** background → Estimated using MC.
   - Single Higgs production + jets, where $H \to \gamma\gamma$.
   - Main contribution from ggH, ttH and ZH production modes.
2. **Continuum** background → Estimated using data (mostly from $m_{\gamma\gamma}$ sidebands region).
   - Main contribution from di-photon production + additional jets.
   - Single photon + jets and multijet background (where 1 or 2 jets are misidentified as photons) account for < 15%.

Ratio of the expected bkg. yields for each analysis category after replacing the DL1r WP with GN2 WP.

Samples	High Mass 1	High Mass 2	High Mass 3	Low Mass 1	Low Mass 2	Low Mass 3	Low Mass 4	$b\bar{b}\gamma\gamma$ preselection
ggH	0.899244	0.914420	0.959389	0.767223	0.909515	0.958215	0.902911	0.838291
ttH	0.867658	0.864079	0.863263	0.887443	0.900553	0.906607	0.900554	0.912129
ZH	0.968675	0.969956	0.978217	0.959593	0.969878	0.970003	0.978440	0.943449
Other H	0.862848	0.836129	0.856764	0.853913	0.852318	0.917792	0.970914	0.839883
$\gamma\gamma$+jets	0.833301	0.880967	0.907797	0.756044	0.812988	0.842126	0.817748	0.689437

Caveat:
- The GN2 / DL1r expected yield ratio for the continuum bkg. was estimated relying on the true flavor fractions extracted from the $\gamma\gamma$+jets sample.
- This ratio is however used to rescale the continuum bkg. yields evaluated in data (following the recipe of the $HH \to b\bar{b}\gamma\gamma$ analysis).

Bkg. yields reduced of a factor of up to 25% (depending on the samples and the category)!

Expected improvement on the statistical results

- Expected stat. only exclusion limits are set on the **di-Higgs signal strength** at **95% CL**.

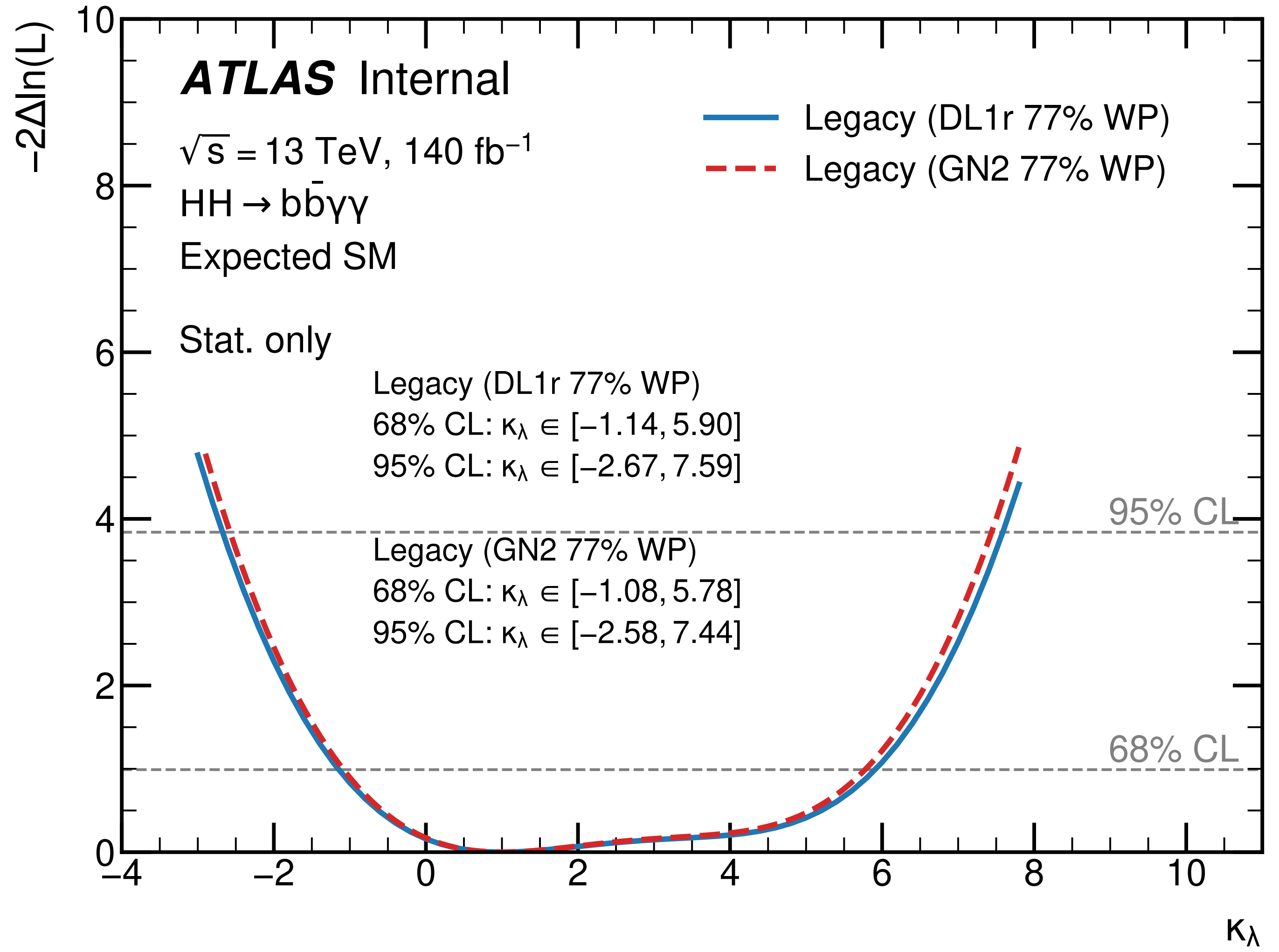
Expected upper limits on $\mu(\mathbf{HH})$ @ 95% CL.

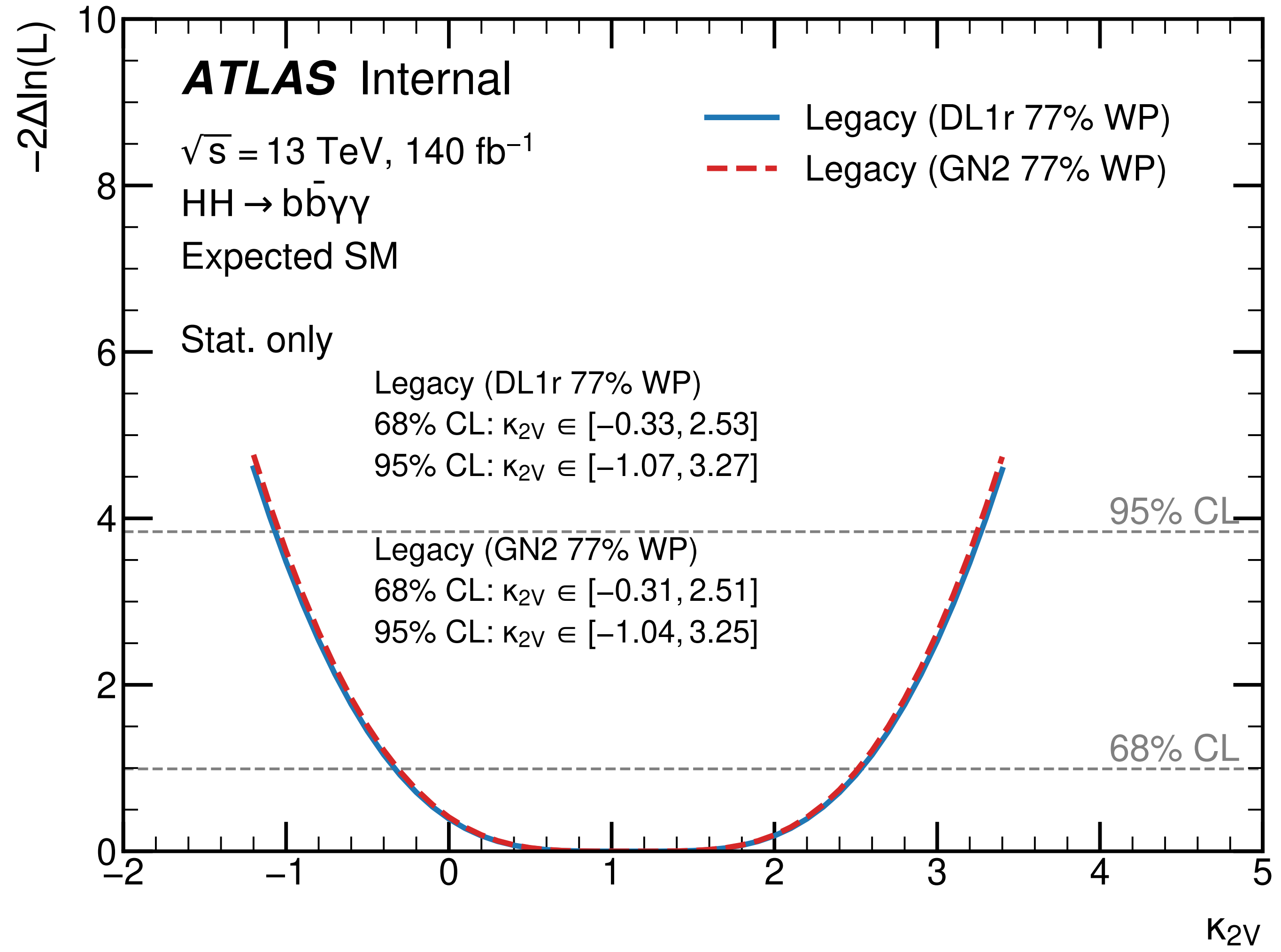
	Nominal (DL1r)	Improved (GN2)
μ(HH)	4.86	4.70

We observe a **3.3% improvement** in the expected stat. only limit on $\mu(\mathbf{HH})$, coming **from** the **new bkg. rejections only**!

- No retraining of the Legacy BDT or redefinition of the categories involved here. → This improvement is very conservative!
- We also made assumption on the true flavor fraction composition from the MC, and on the GN2 / DL1r bkg. rejection estimated from MC.
  → Reasonable as long as Scale Factors for DL1r and GN2 are similar.

- We can also derive expected **68%** and **95% confidence intervals** for κ_λ and κ_{2V} via a **profile log-likelihood ($-2\Delta\ln(\mathbf{L})$) scan**.





- Replacing the bkg. rejection efficiency of DL1r 77% WP with those of GN2 77% WP seems to **improve** the **expected constraints** on κ_λ by a relative factor of **2.3%**.
- The improvement for the κ_{2V} constraint is a bit **smaller** (1.1%).

Summary

- A **new generation** of **GNN-based b-tagging algorithms** was developed in ATLAS.
  - The **GN2** algorithm is the **current state of the art**!
  - We expect **at least a factor $\times$ 2 of improvement** in **fake b-jets background rejection** w.r.t. the older RNN-based taggers adopted in Run 2 (e.g. DL1r).
- **We look forward** to adopting the new tagger in our **di-Higgs analyses** (= **especially in the** $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, and $b\bar{b}b\bar{b}$ channels)!
- Before being able to apply the new algorithms in physics analysis, we need to **calibrate** it!

Calibration of the **GN2 WPs** = **measuring** the **(mis-)tagging efficiency for b-jets, c-jets and light jets** in **real data**!

Three different analyses (proceeding in parallel), using each other's results as inputs.

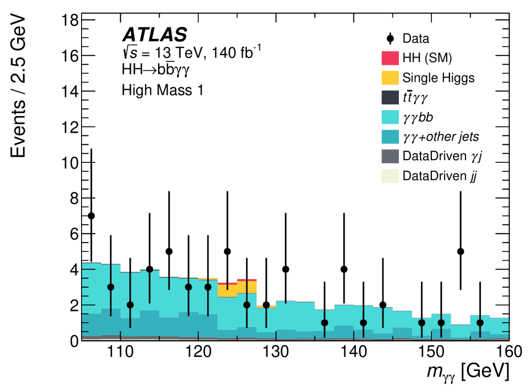
- I have been working on the **measurement** of the **b-jet efficiency**, using **di-leptonic $t\bar{t}$ events!**
  - Efficiency measurement with 2022 data ~ ready!
    - A first version of the efficiency and the corresponding SFs is ready, together with the systematic uncertainties!
  - Still working through our to-do list for Run 2 and 2023 data.
    - We have nominal results for Run 2 (no systematic yet), and finalizing sample production for 2023!

Timescale for releasing the recommendations: ~ 1 month!

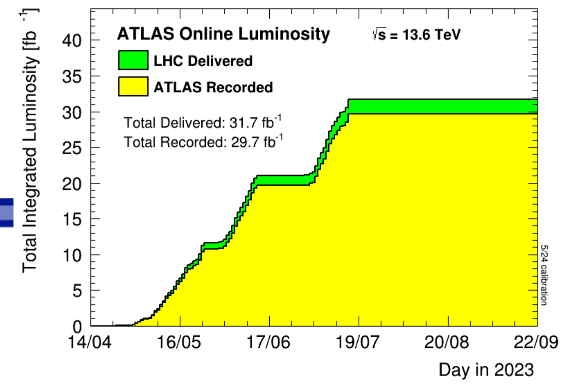
Summary

- Finally, I have tried to perform a qualitative estimate of the improvement from adopting the GN2 tagger in the $HH \to b\bar{b}\gamma\gamma$ analysis.
    - The analysis requires **exactly 2 b-jets** selected by the **77% WP** of the **DL1r tagger**.
    - **The pseudo-continuous b-tagging DL1r scores** for the jets are used as **input** for defining the **analysis categories**!
- We tried to adjust the expected background yields in the Legacy $HH \to b\bar{b}\gamma\gamma$ analysis, according to the expected fake b-jet rejection power of the GN2 77% WP compared with the DL1r 77% WP, and we repeated the statistical analysis to extract the expected results.
    - We observed a $\sim 3\%$ improvement, coming from the new bkg. rejection only!
    - **Caveat**: we are being very conservative in this estimation.
        - No re-definition of the category is involved here (i.e. the categories are still relying on the DL1r scores).
        - We are only studying the effect of the improved bkg. rejection!
            Since our analysis is statistically limited and our signal includes 2 true b-jets, using a higher eff. WP (e.g. 90%) will bring a much stronger improvement!
- New Run 2 + partial Run 3 $HH \to b\bar{b}\gamma\gamma$ analysis is already testing the GN2 algorithm!

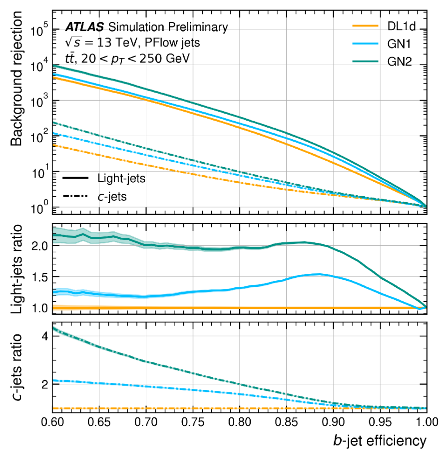
Stay tuned for **new di-Higgs results**!



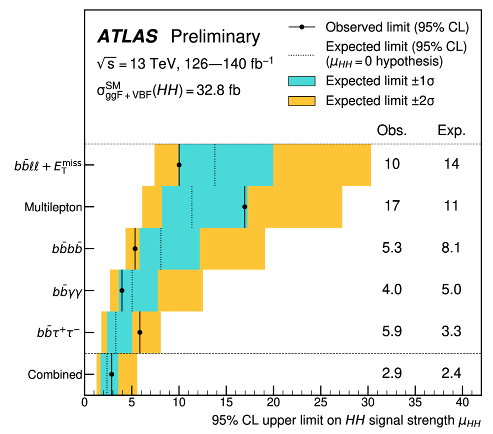
$HH \to b\bar{b}\gamma\gamma$ analysis

+ 

Increase in statistics with 2022 and 2023 data (+ 66 fb^{-1})!

+ 

New b-tagging algorithm!

+ 

More analysis improvements and combination!

= ? **First hint** of **SM HH** at the **end of Run 3** ?

Thank you for your attention!

Backup

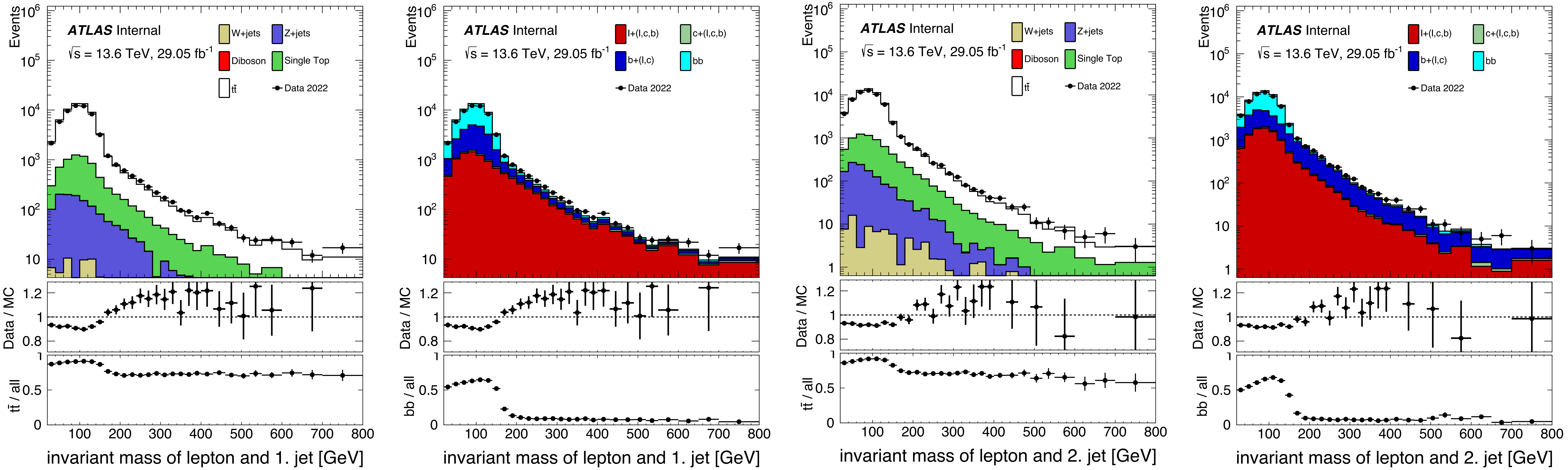
b-tagging efficiency measurement: data / MC agreement

η and p_T of the leading and subleading jets, showing the contribution from each physics process

b-tagging efficiency measurement: data / MC agreement

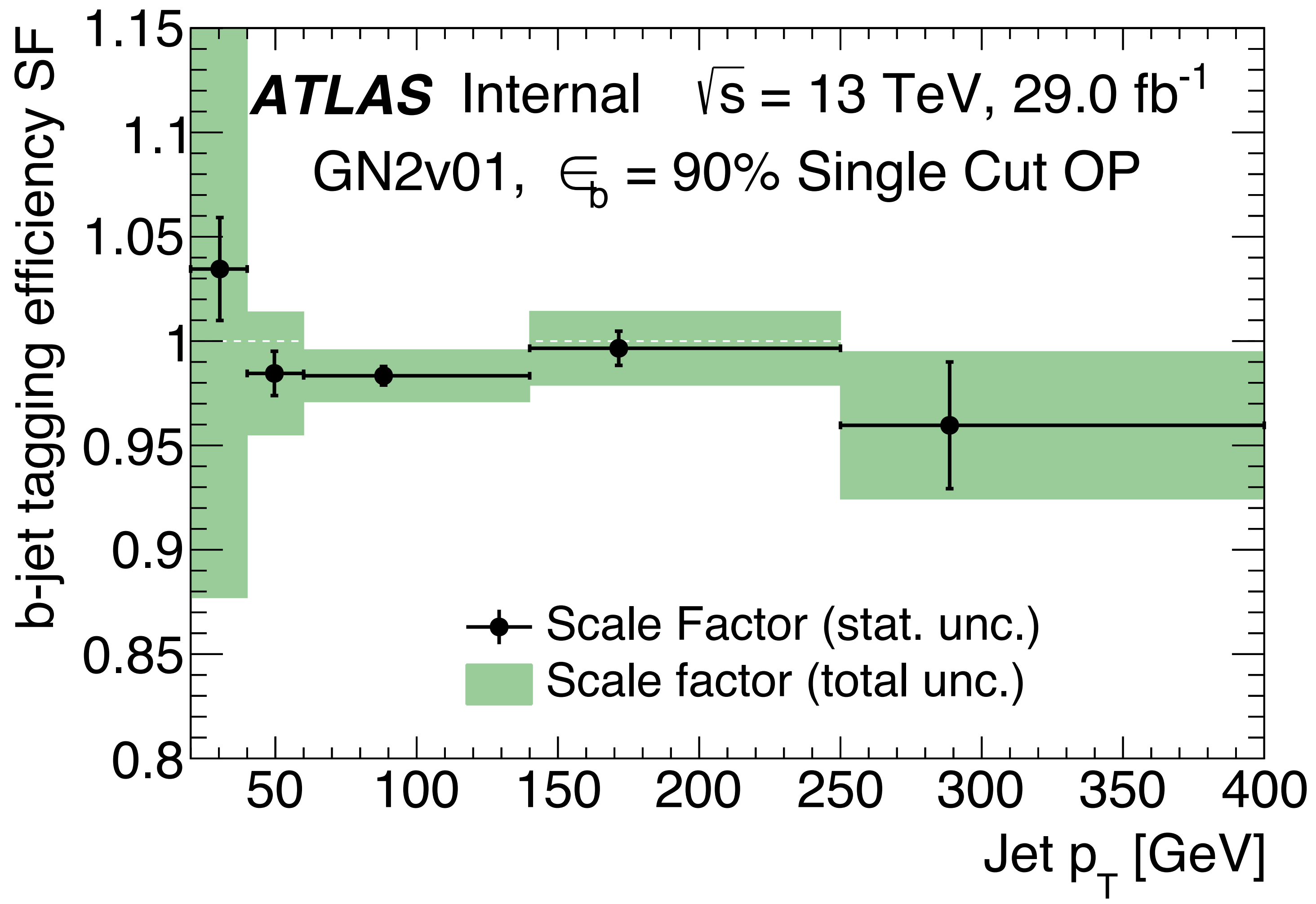
η and p_T of the leading and subleading jets, showing the contribution from the true jet flavor fractions

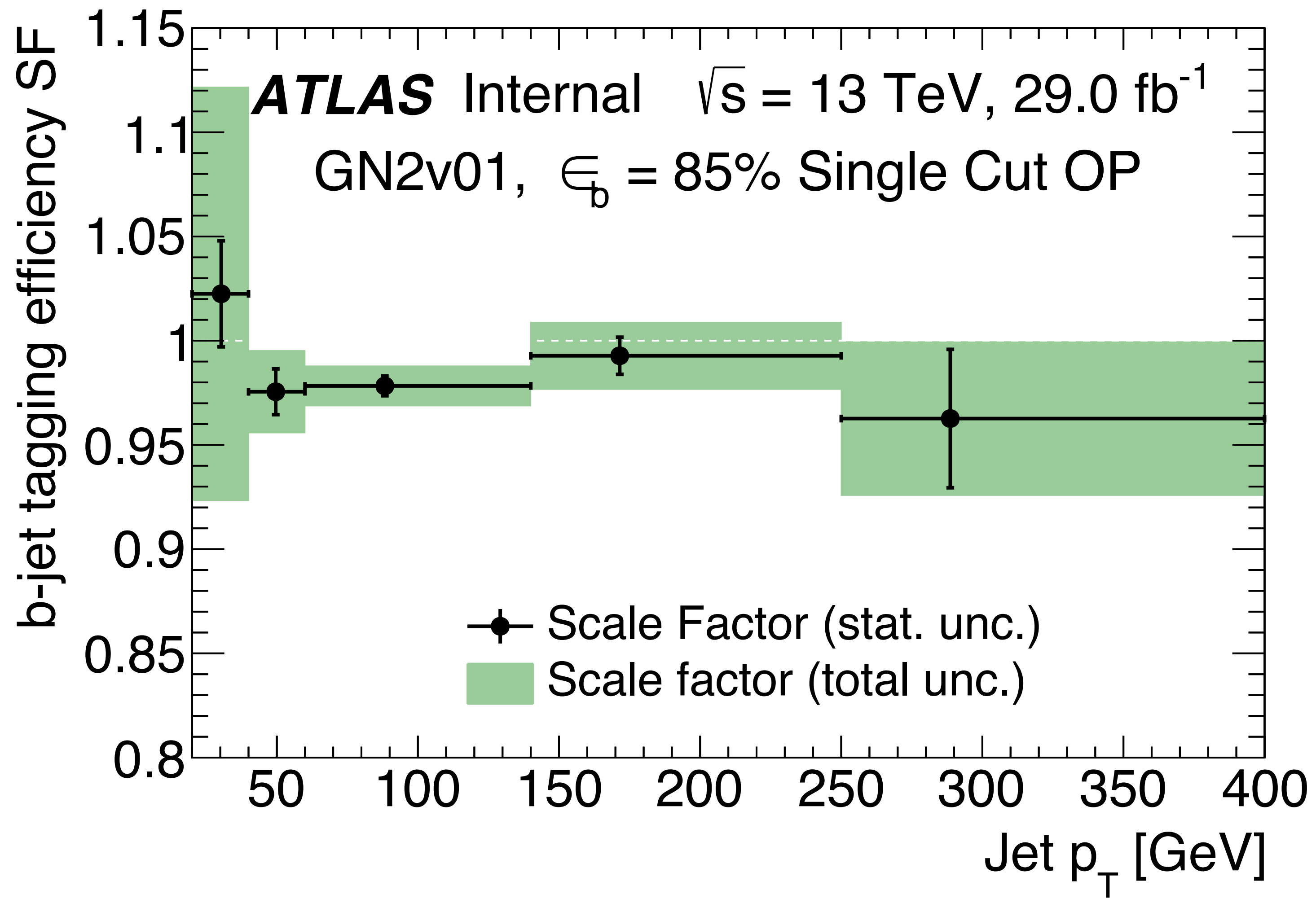
Lepton + jet invariant mass for leading and subleading jets

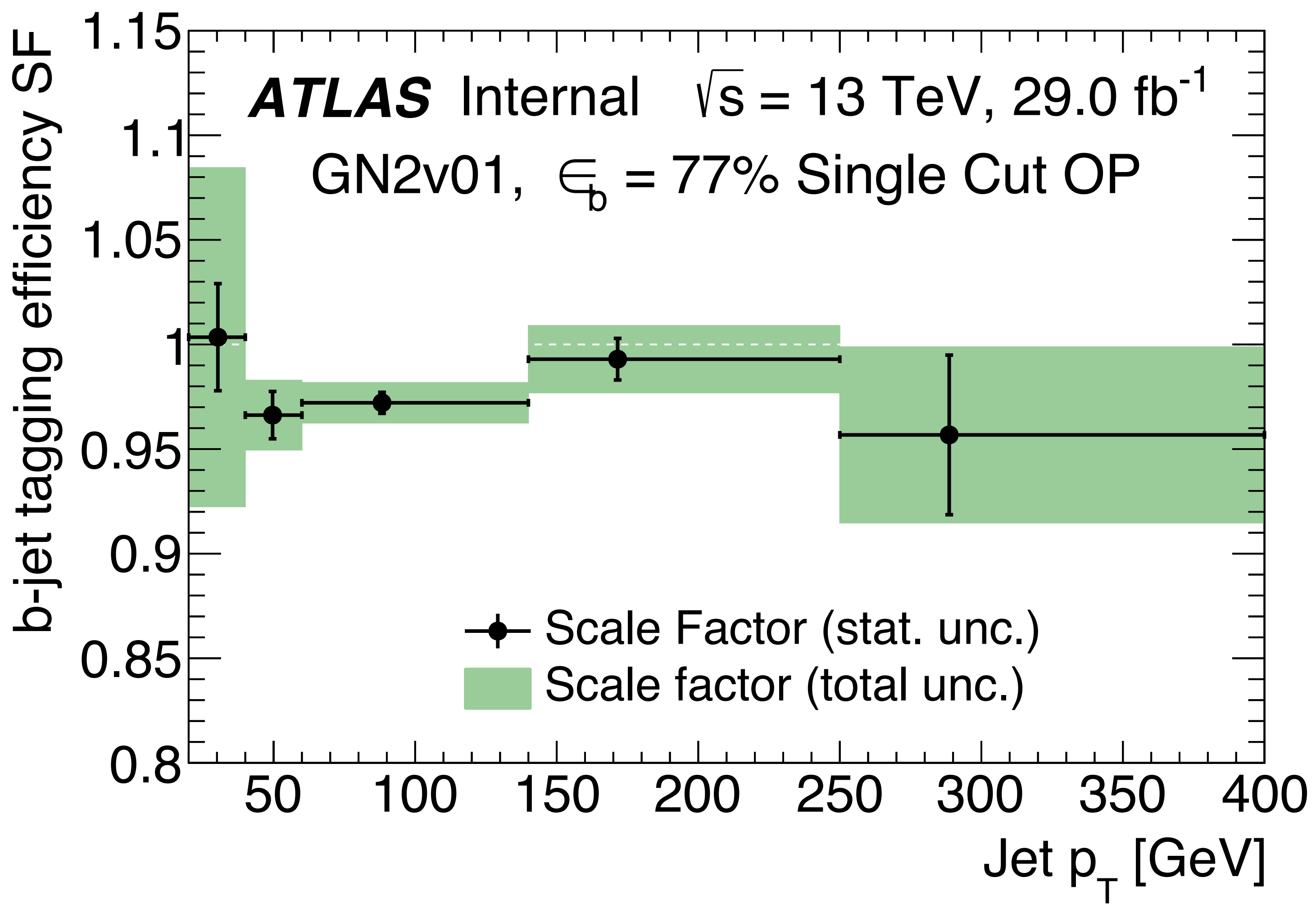


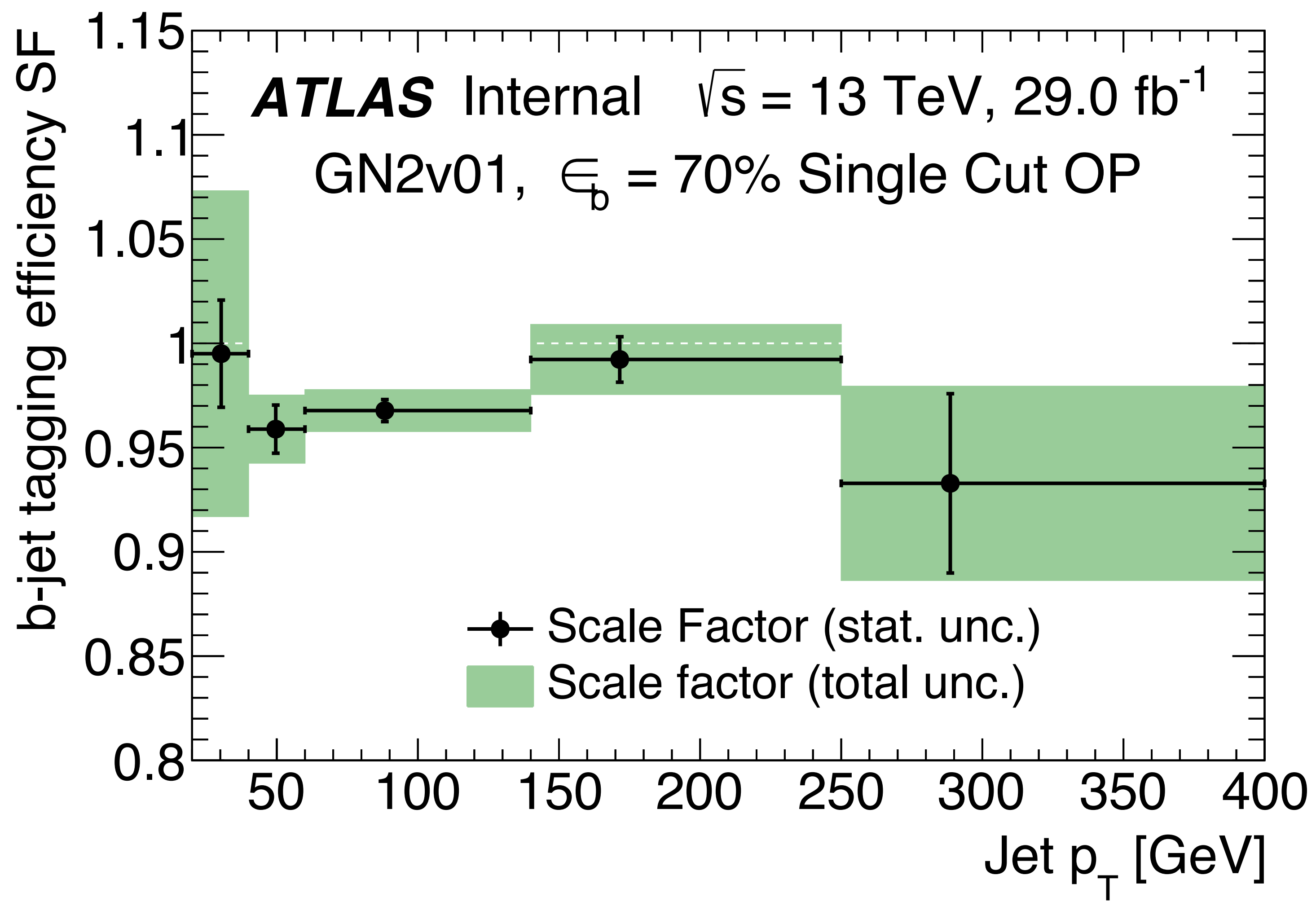
b-tagging efficiency measurement: Run 3

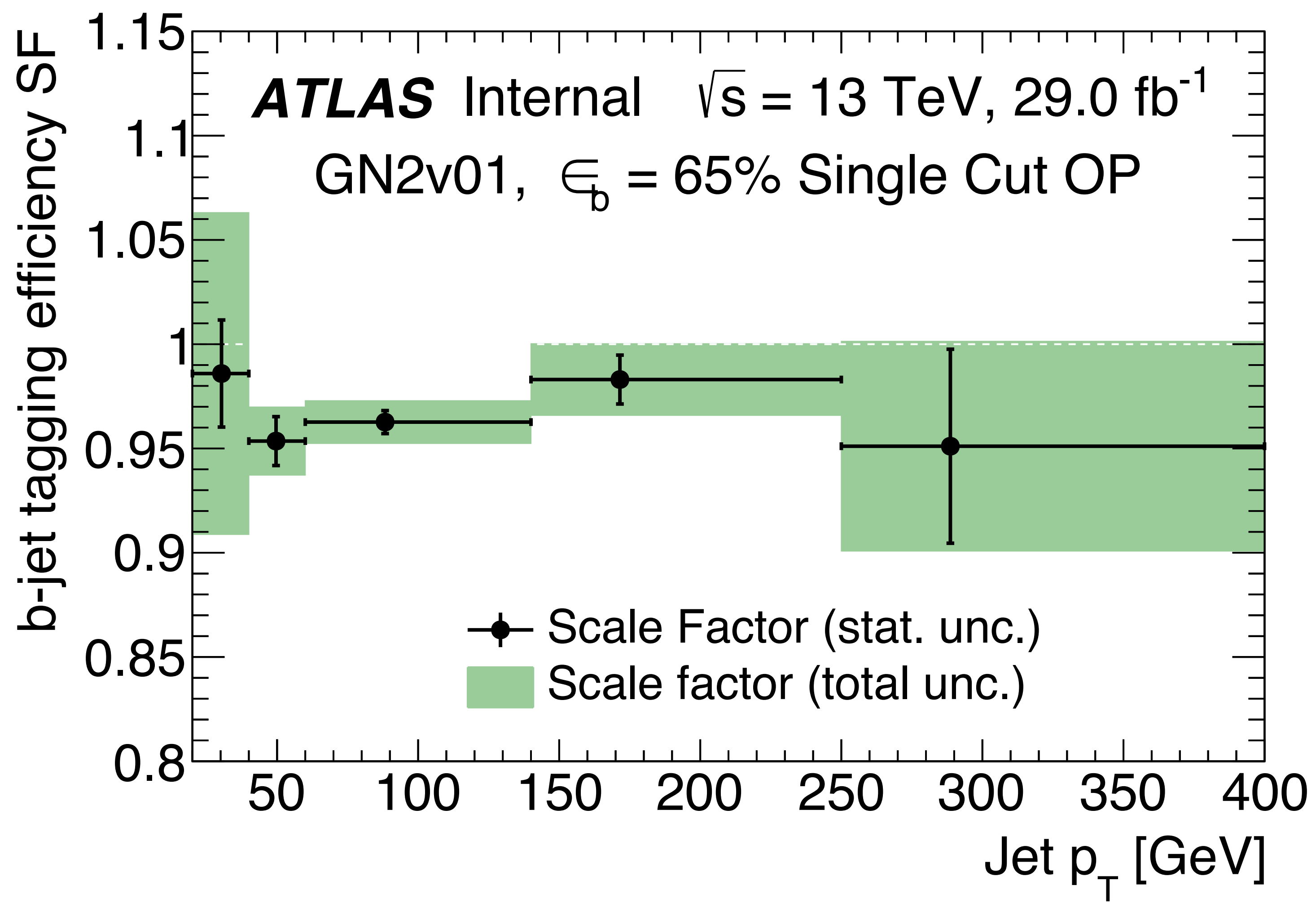
b-tagging SFs as a function of the jet p_T, for the 65, 70, 77, 85, and 90% eff. WP of GN2





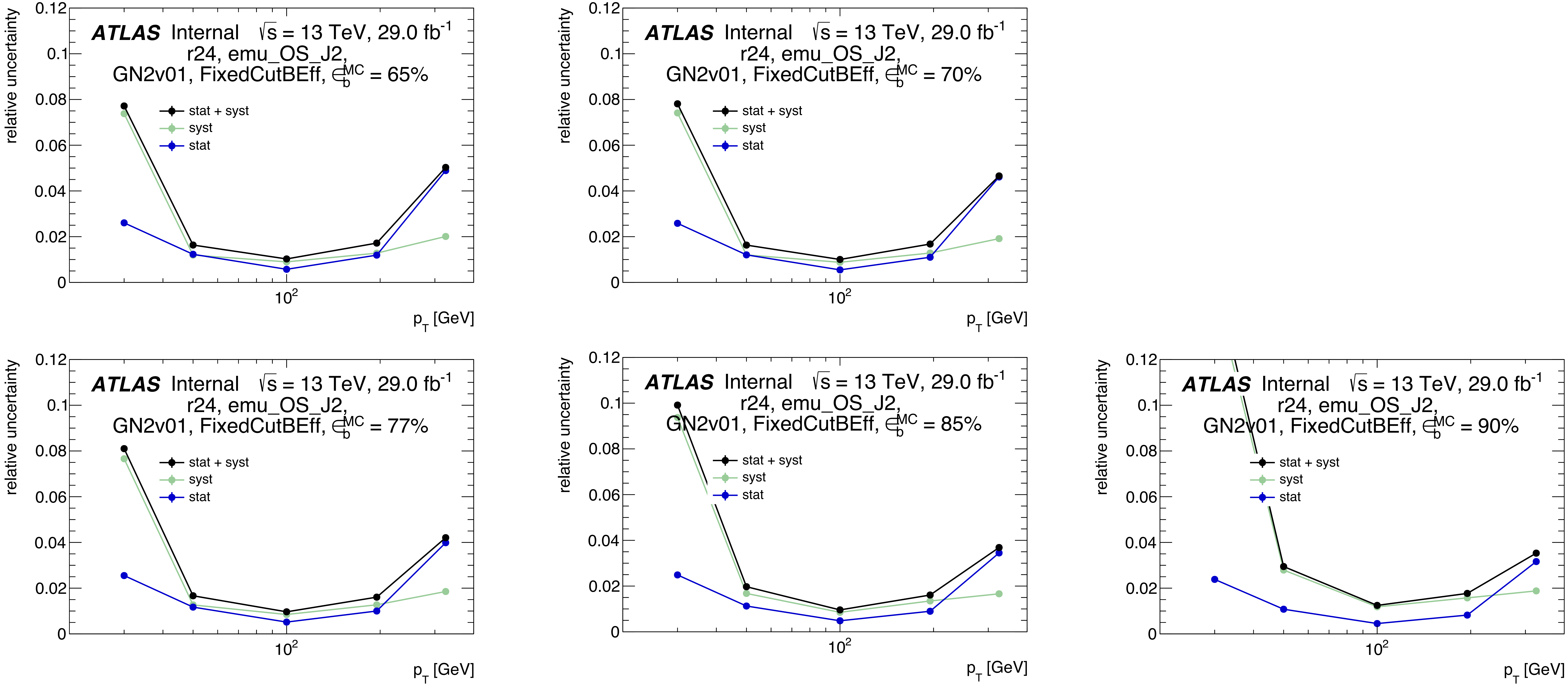






b-tagging efficiency measurement: Run 3

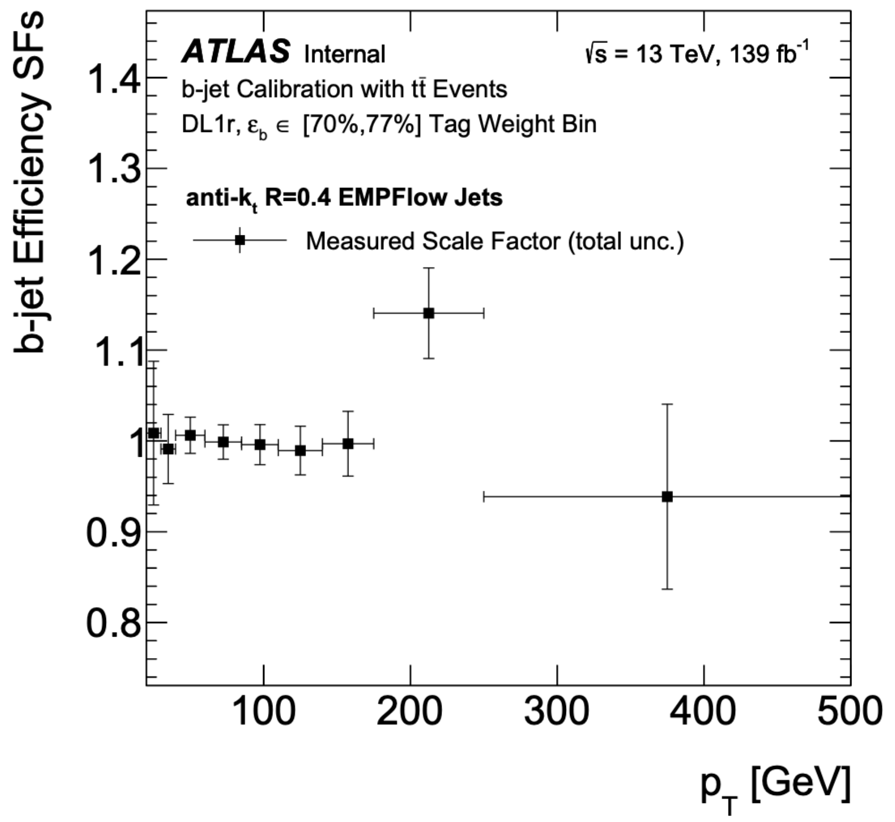
Relative uncertainty to the b-tagging SFs as a function of the jet p_T, for the 65, 70, 77, 85, and 90% eff. WP of GN2

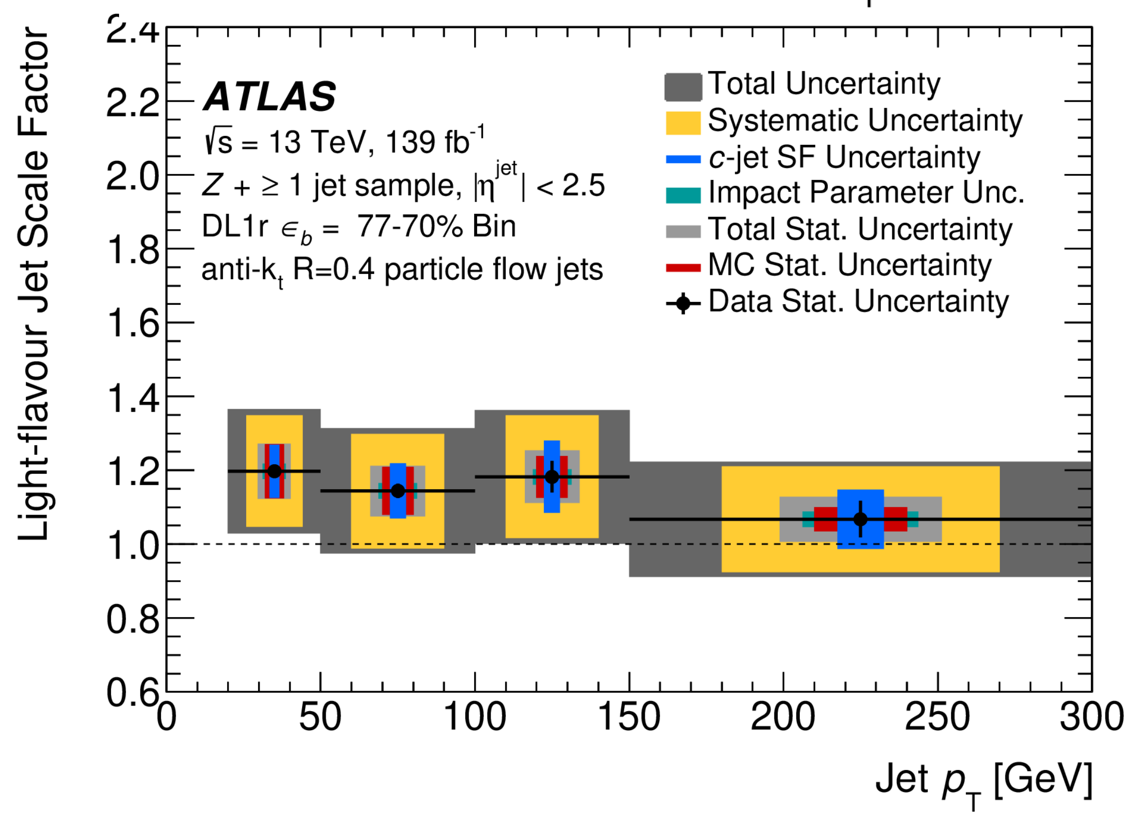


b-, c, and light jet scale factors for DL1r and GN2

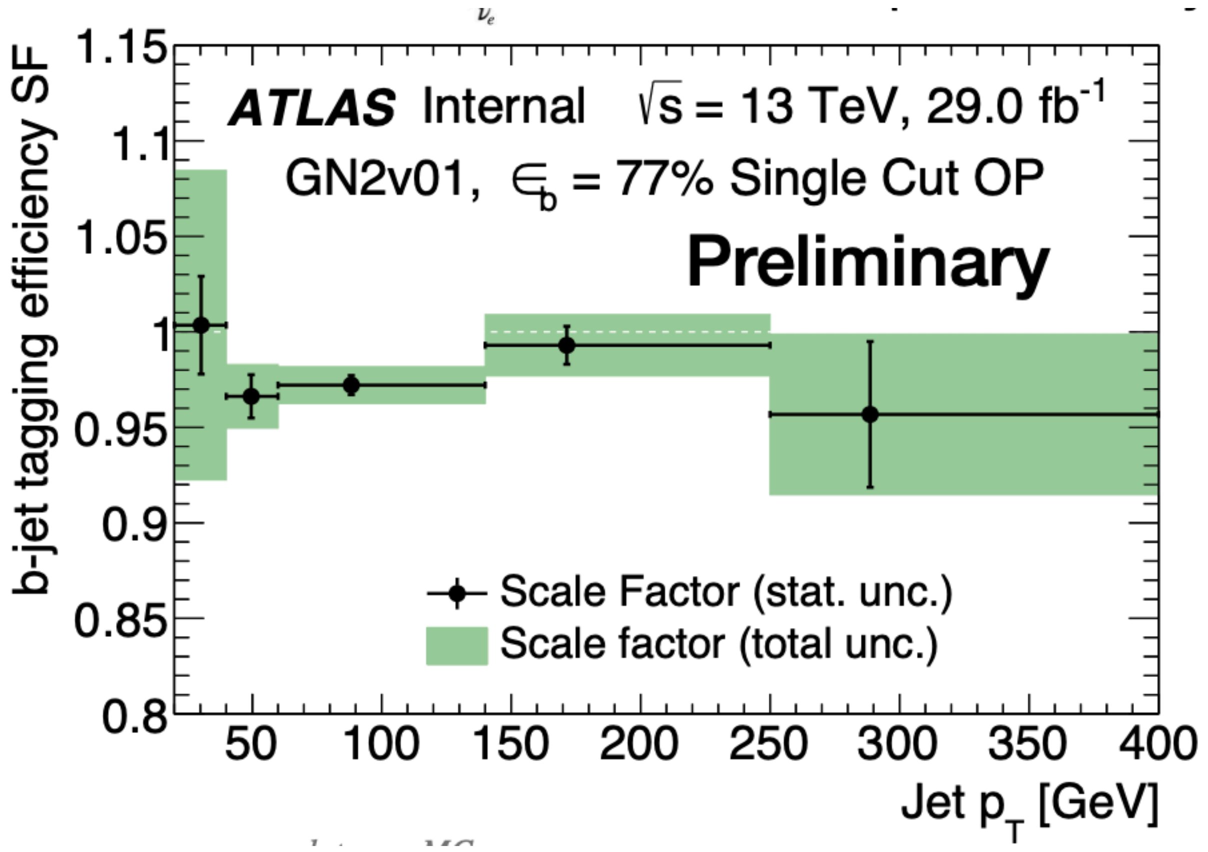
- In order to estimate the ratio between the expected yields from the DL1r and GN2 77% eff. WPs, we relied on the c- and light-jet rejection curves (as a function of the b-jet efficiency) evaluated using MC samples.
- Trusting the ratios that we used to rescale the expected yields is a reasonable assumption, as long as the MC → Data efficiency scale factors for b-, c, and light jets are similar, between DL1r and GN2!

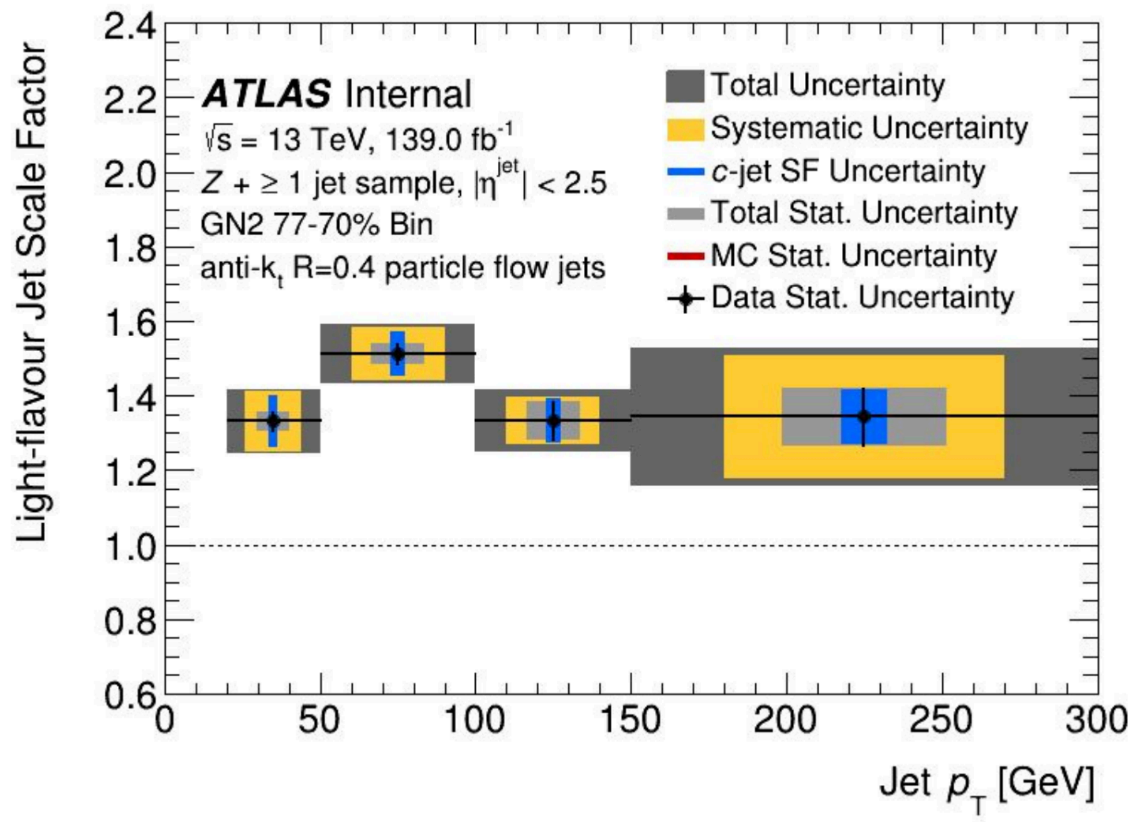
DL1r (Rel. 21)





GN2 (Rel. 25)





- MC → Data efficiency scale factors for b-jets are similar, between DL1r and GN2!
  - Very close to 1 in both cases!
- For light jets, the scale factors are close to 1.3 for GN2, while they are closer to 1 in the DL1r case.
  - The two sets of Scale Factors differ of a factor between 10 and 20%, but they are compatible within their systematic uncertainties!

True bb / All purities in bkg. samples for the $HH \to b\bar{b}\gamma\gamma$ analysis

True bb / All purities (with DL1r-based b-tagging)

Samples	High Mass 1	High Mass 2	High Mass 3	Low Mass 1	Low Mass 2	Low Mass 3	Low Mass 4	$b\bar{b}\gamma\gamma$ preselection
ggH	0.843846	0.854195	0.929171	0.661551	0.848940	0.925240	0.845122	0.762660
ttH	0.741365	0.733432	0.734020	0.781049	0.805715	0.817540	0.805060	0.829875
ZH	0.954499	0.956188	0.968055	0.938304	0.951592	0.952504	0.962574	0.915358
Other H	0.757329	0.705959	0.736176	0.742982	0.734976	0.862540	0.957203	0.736247
$\gamma\gamma$+jets	0.732071	0.801465	0.854433	0.625333	0.710732	0.754355	0.711985	0.543612

- The background samples for the $HH \to b\bar{b}\gamma\gamma$ analysis include both true b-jets and c and light jets mistagged as b-jets.
- Given the 77% efficiency WP, the true bb purity in each sample strongly depends from the analysis category (= the pseudo-continuous DL1r score is used as input for the category definition).
  - **Resonant** background
    - Between 60% and 95%, depending on the category and the production mode.
    - ZH has the highest purity (above 90% in each category!).
  - **Continuum** background
    - Between 60% and 85%, depending on the category.

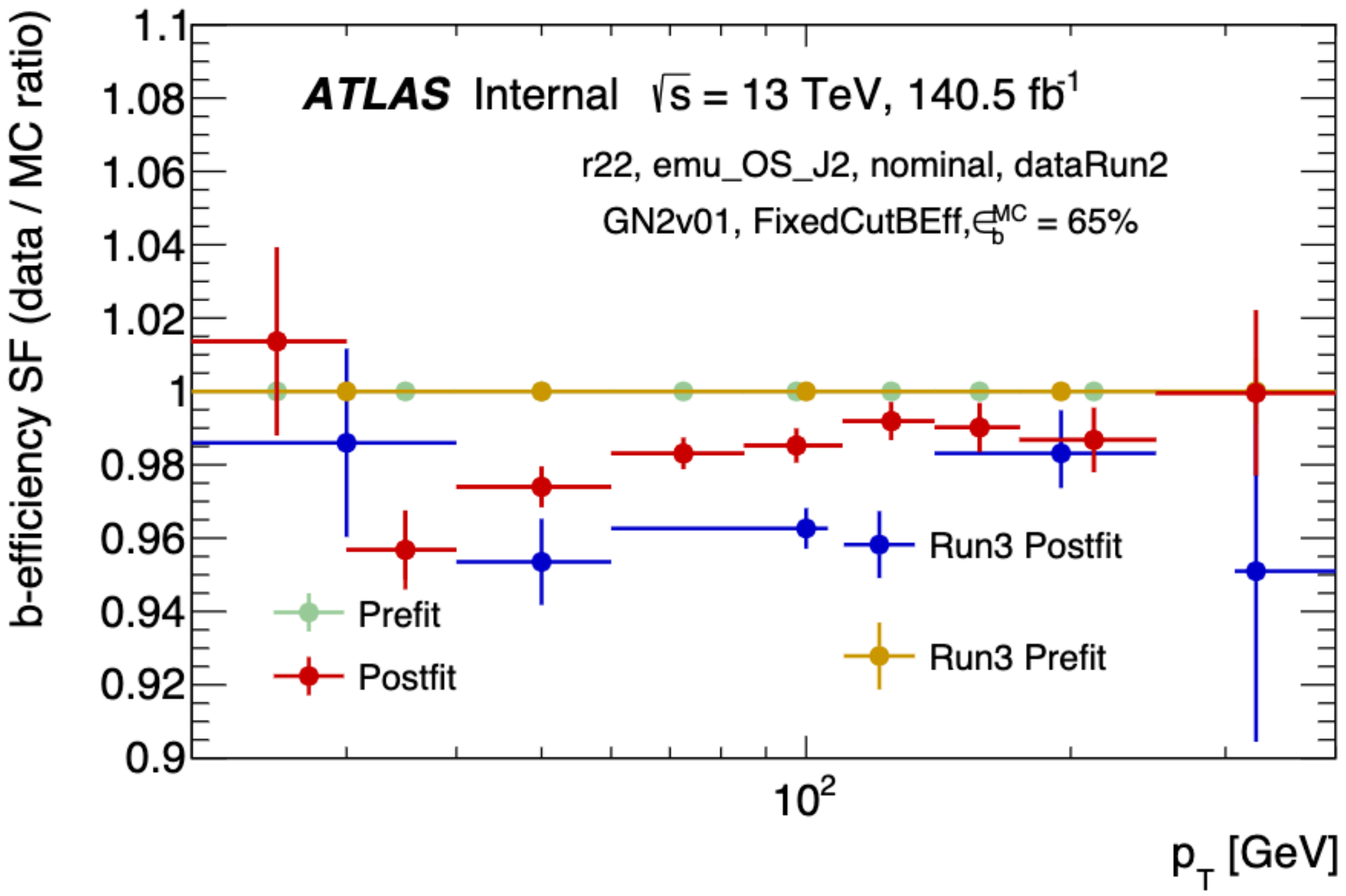
True bb / All purities in bkg. samples for the $HH \to b\bar{b}\gamma\gamma$ analysis

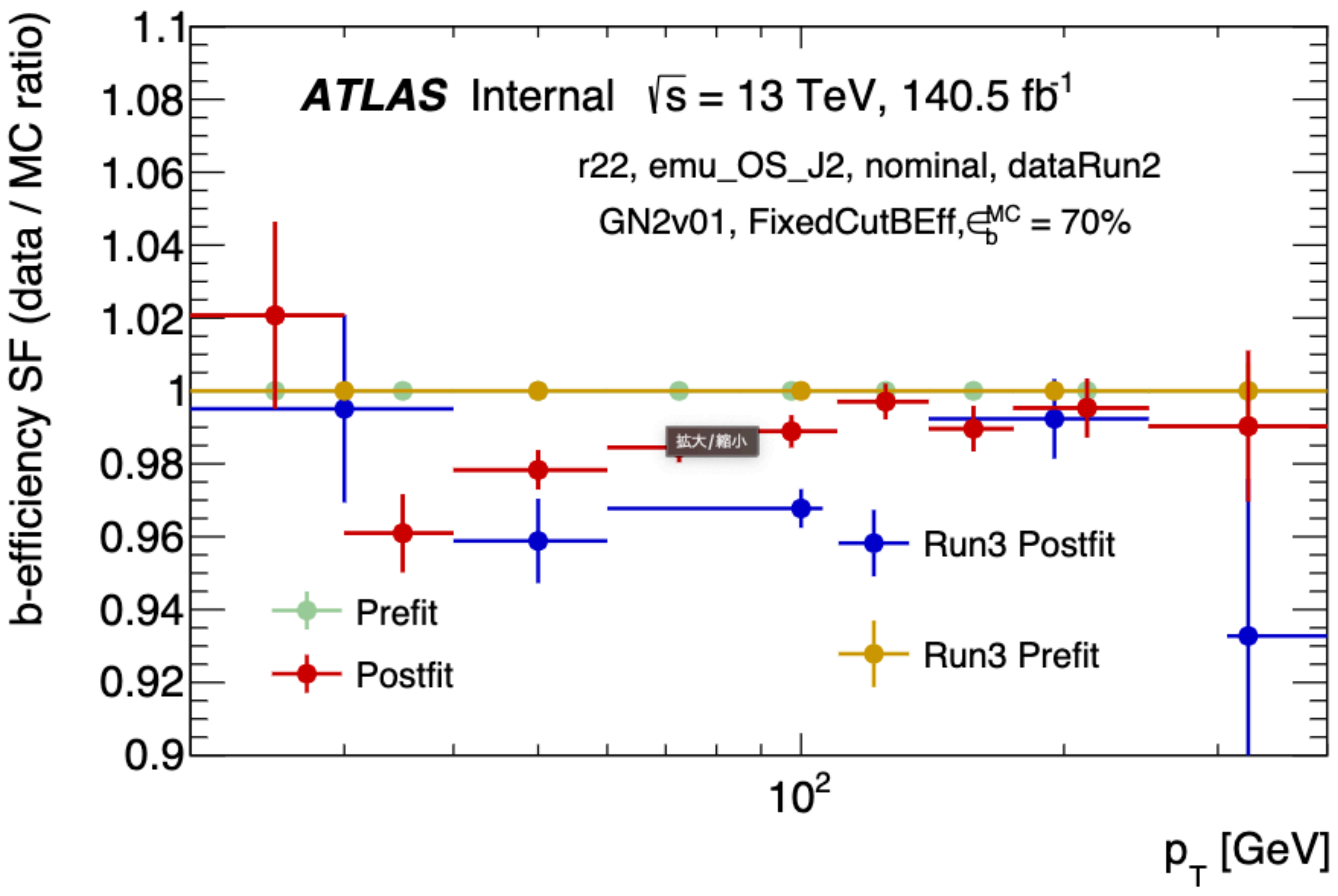
True bb / All purities (with GN2-based b-tagging)

Samples	High Mass 1	High Mass 2	High Mass 3	Low Mass 1	Low Mass 2	Low Mass 3	Low Mass 4	$b\bar{b}\gamma\gamma$ preselection
ggH	0.938394	0.934139	0.968502	0.862267	0.933399	0.965587	0.935997	0.909779
ttH	0.854443	0.848802	0.850285	0.880112	0.894690	0.901759	0.893961	0.909822
ZH	0.985365	0.985805	0.989611	0.977814	0.981146	0.981960	0.983784	0.970225
Other H	0.877708	0.844318	0.859252	0.870091	0.862325	0.939799	0.985878	0.876607
$\gamma\gamma$+jets	0.878519	0.909757	0.941216	0.827112	0.874222	0.895774	0.870666	0.788487

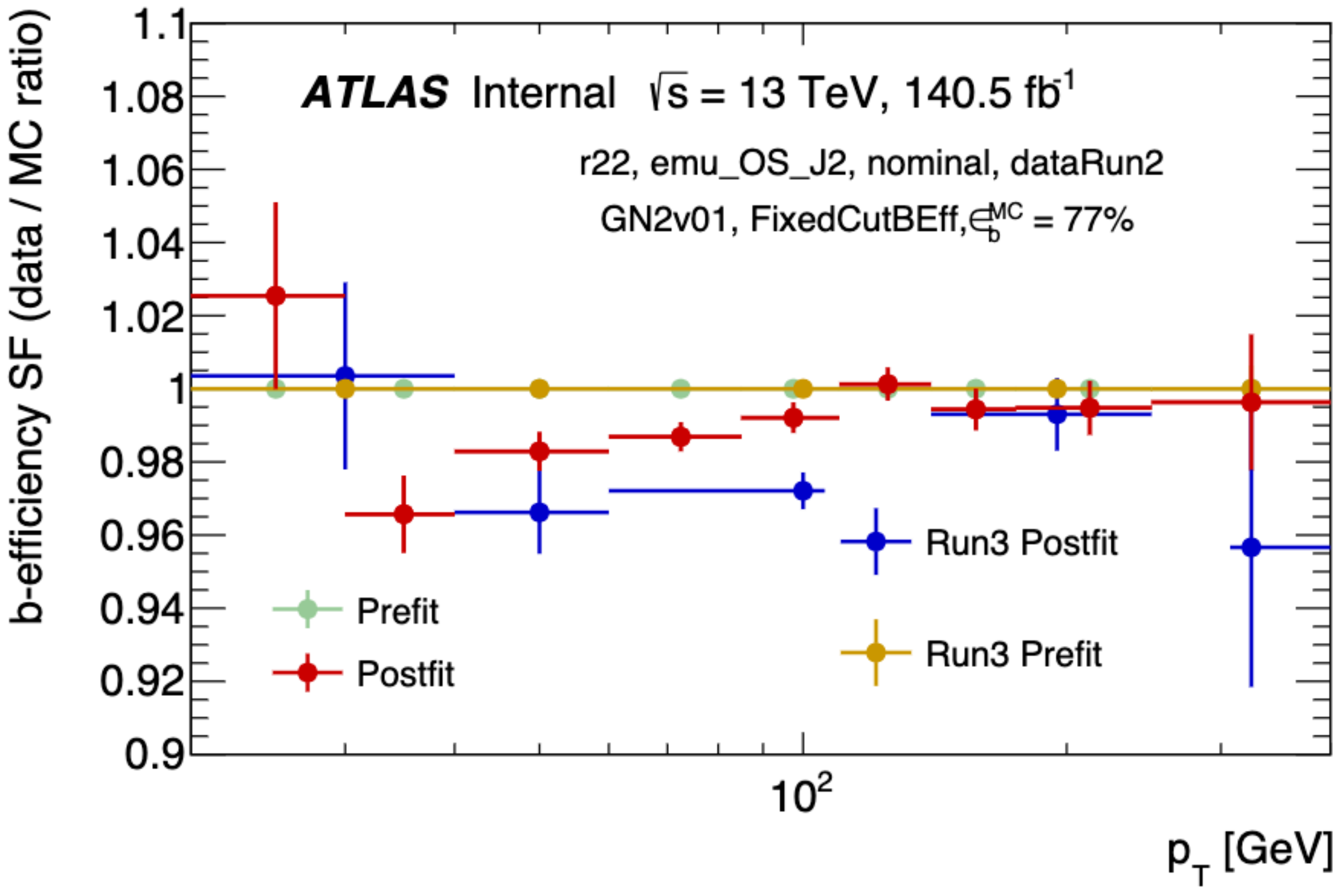
- After replacing the bkg. efficiency for DL1r 77% WP with the estimated bkg. efficiency for the GN2 WP, we could improve the overall True bb purities in each analysis category, and for each sample!
- The margin of improvement depends from both the analysis category and the particular bkg. process.
  - **Resonant** background
    - Improvement within 10% and ~ few %.
    - The new purity is always above 85% for each analysis category.
  - **Continuum** background
    - Improvement between 10% and 20%.
    - The new purity is always above 80%!

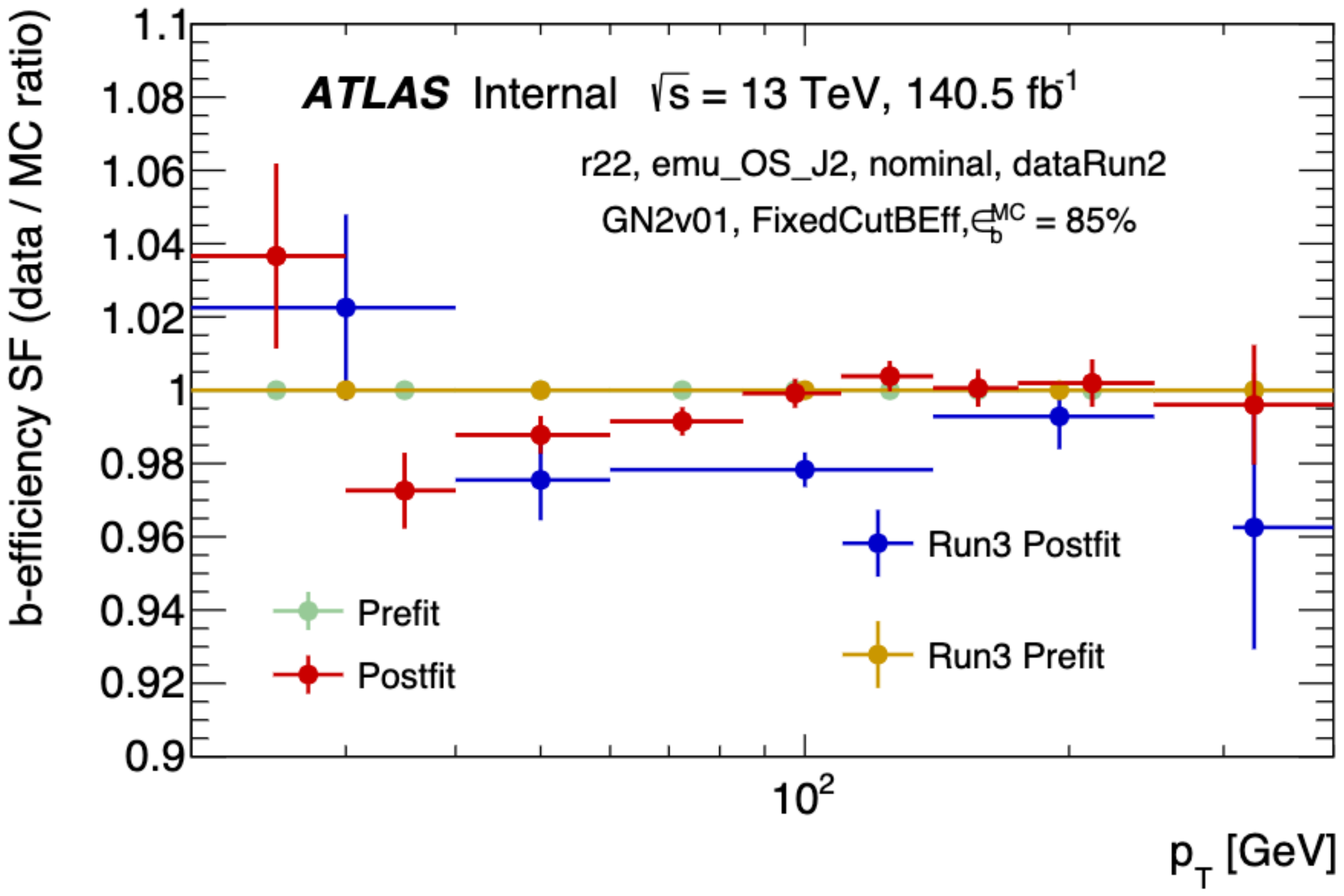
b-tagging efficiency measurement: Run 2 VS Run 3

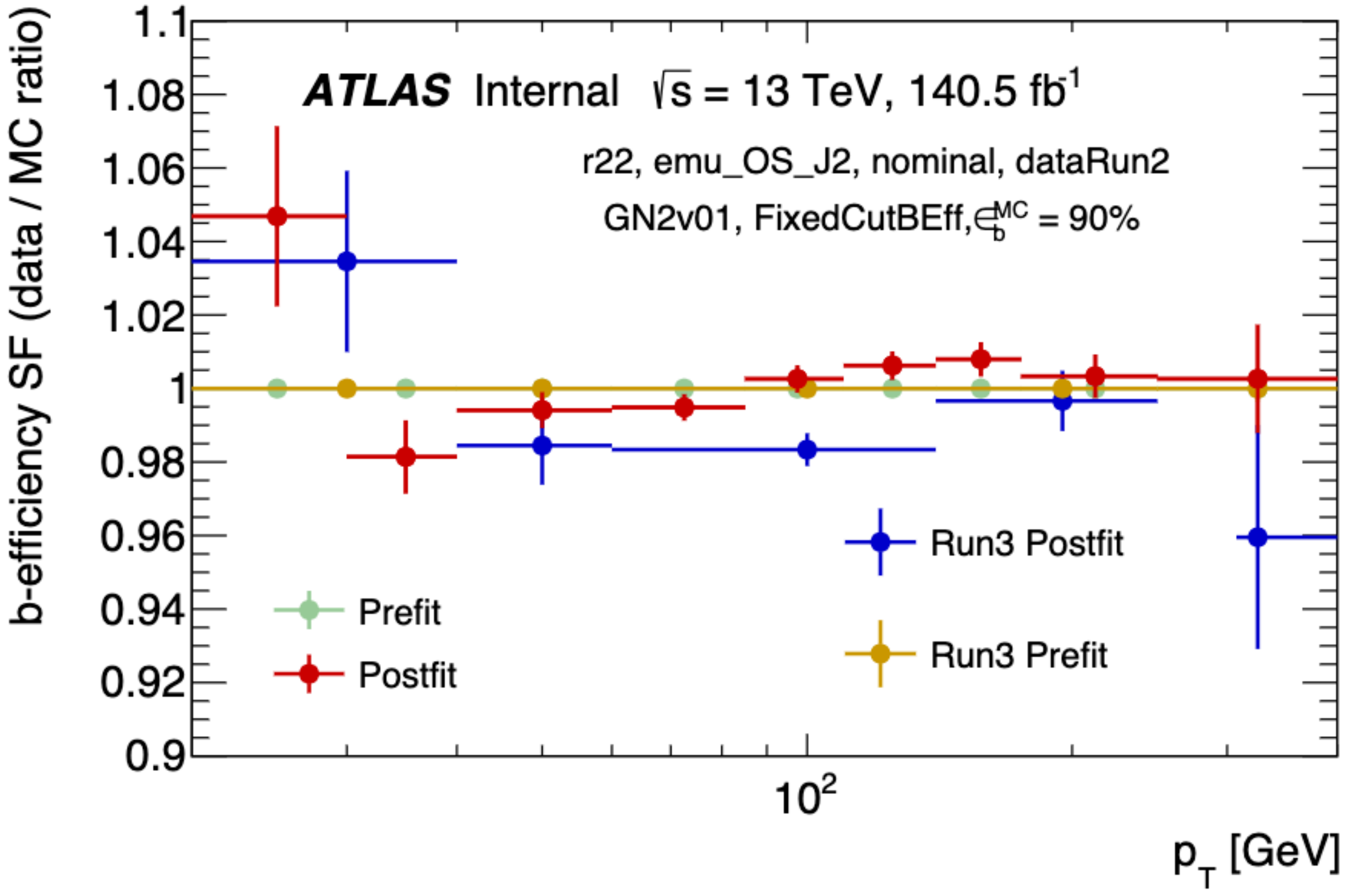




- We compared the nominal b-tagging efficiency SFs (no. systematic uncertainties yet), between full Run 2 and Run 3 (2022 data and mc23a only).
- For Run 3, SFs seems to be systematically larger w.r.t. Run 2 (except for low p_T).
- Similar trend as a function of p_T w.r.t. observed in the past, for DL1d (see backup).







b-tagging efficiency measurement: DL1d tagger, Run 2