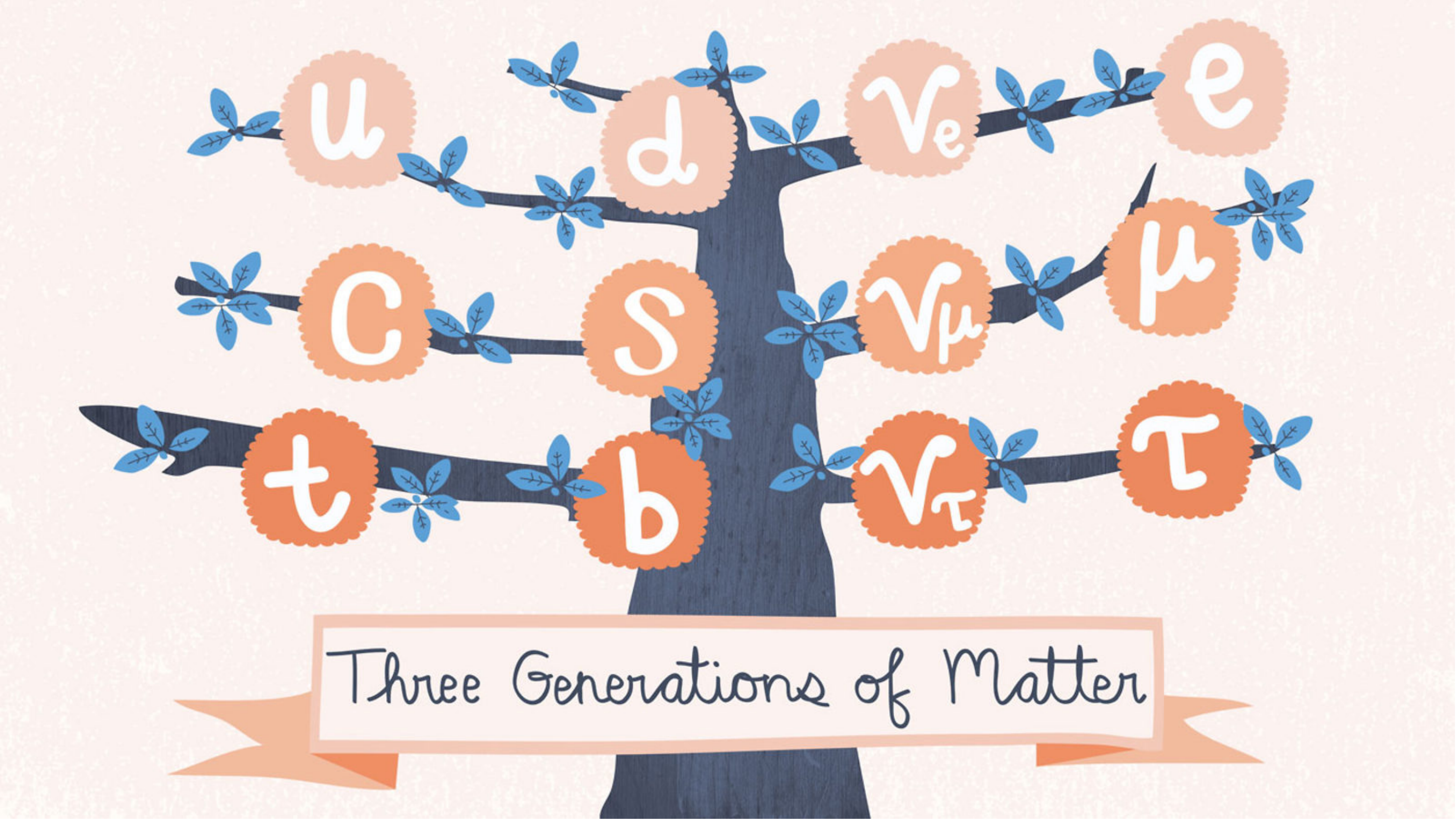
Sensitivity estimates of the VH, $H\to c\bar{c}$ process in the Lorentz-boosted regime

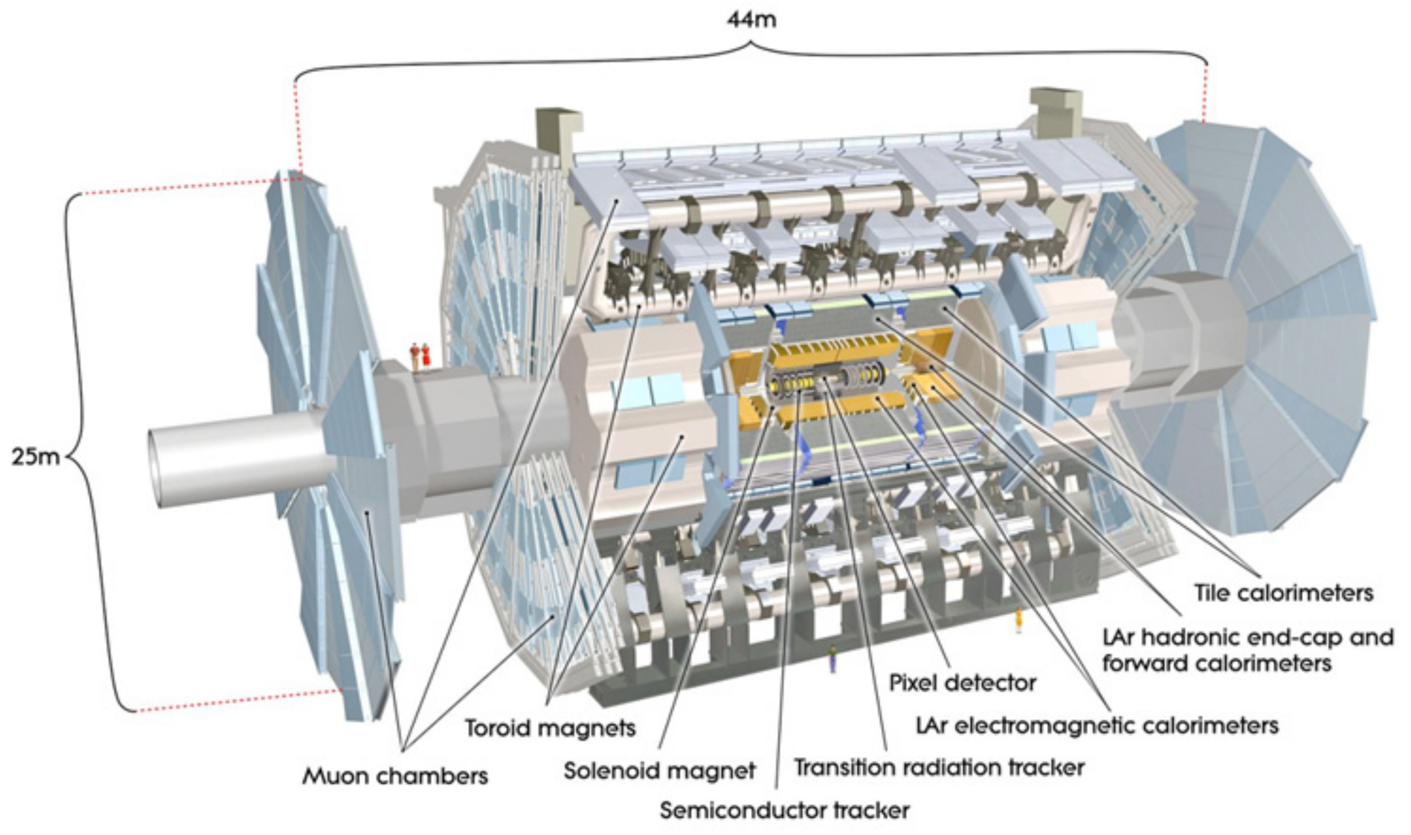
Jurjan Bootsma | 07-11-2025 | NNV Lunteren

Open questions in the Higgs sector

[Artwork by Sandbox Studio. Chicago]

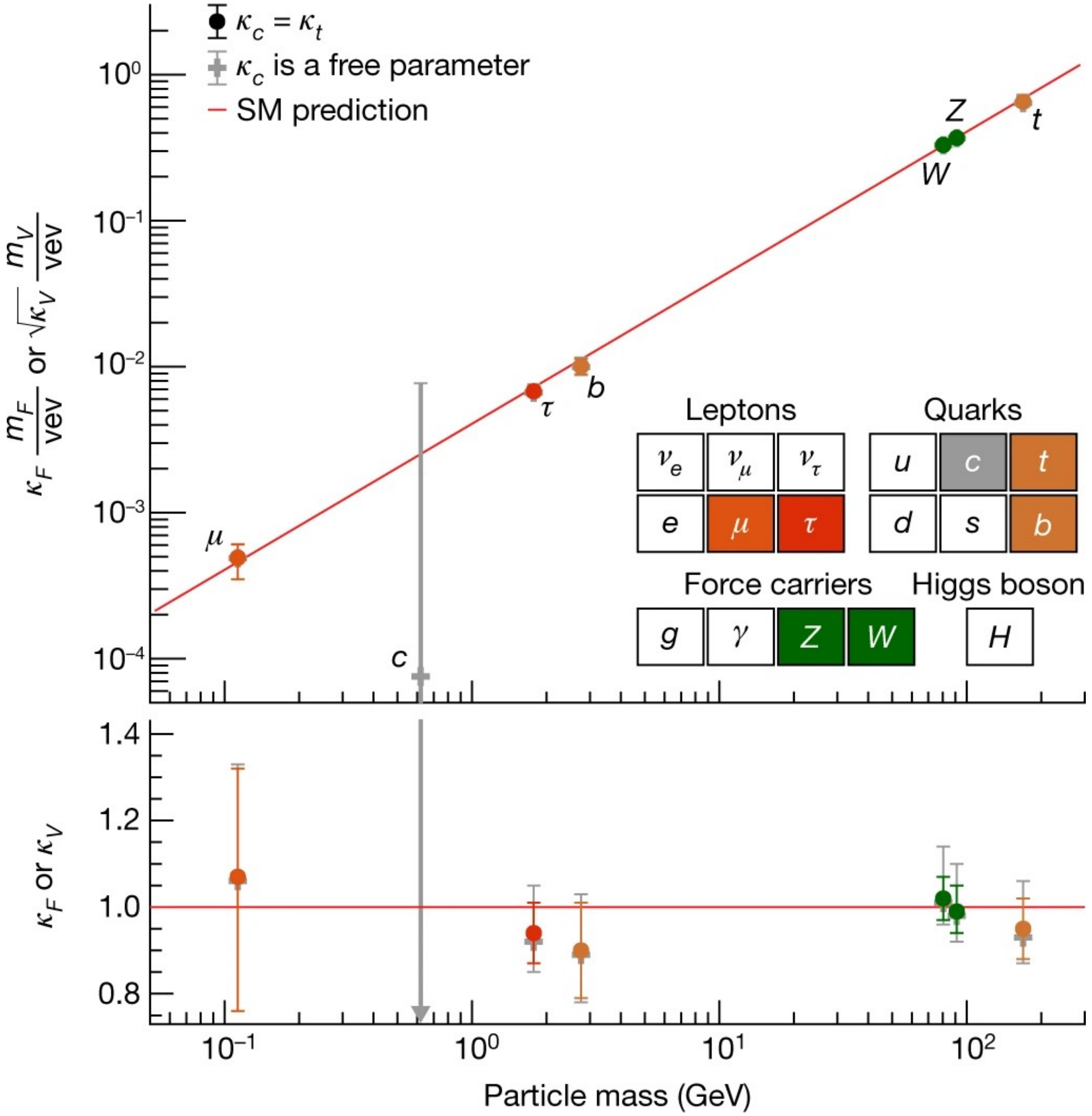


- Standard Model incomplete
- 12/19 free parameters of the SM relate to particle masses
  - Might point at deeper physics related to Higgs
- Experimentally probe the mass-generation mechanism
  - Using the ATLAS detector at the LHC



Higgs couplings and particle masses

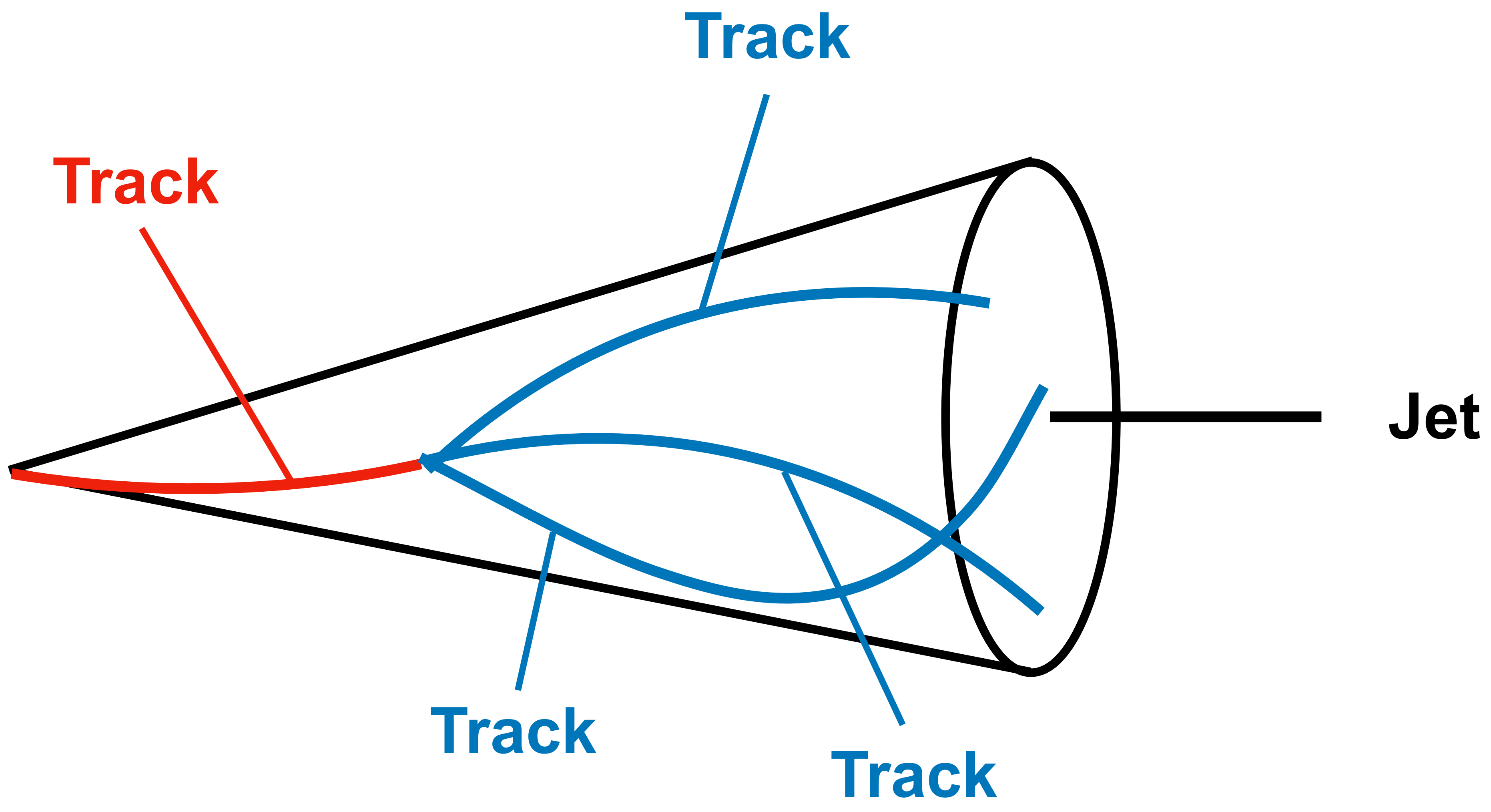
- Studying Higgs boson key research goal
  - Mass, spin, parity, natural width
- Couplings to vector bosons
- Couplings to third-generation fermions
- Next goal: couplings to second-generation fermions
  - Same as for third-generation fermions?
  - Affected by several BSM theories
  - Evidence for Higgs-muon coupling
  - This study: Higgs-charm coupling

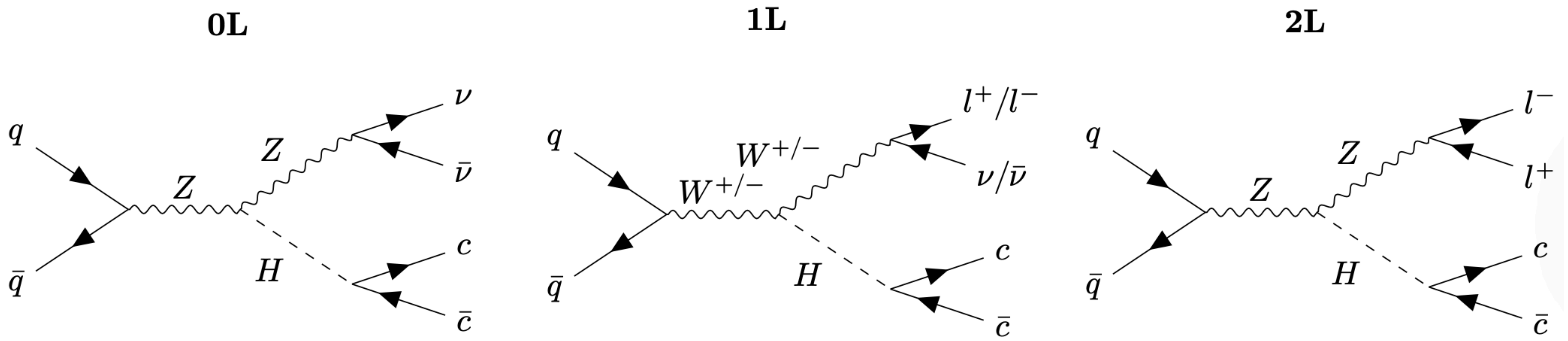


[A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery. Nature, 2022, 607.7917: 52-59.]

$VH(\rightarrow c\bar{c})$ process

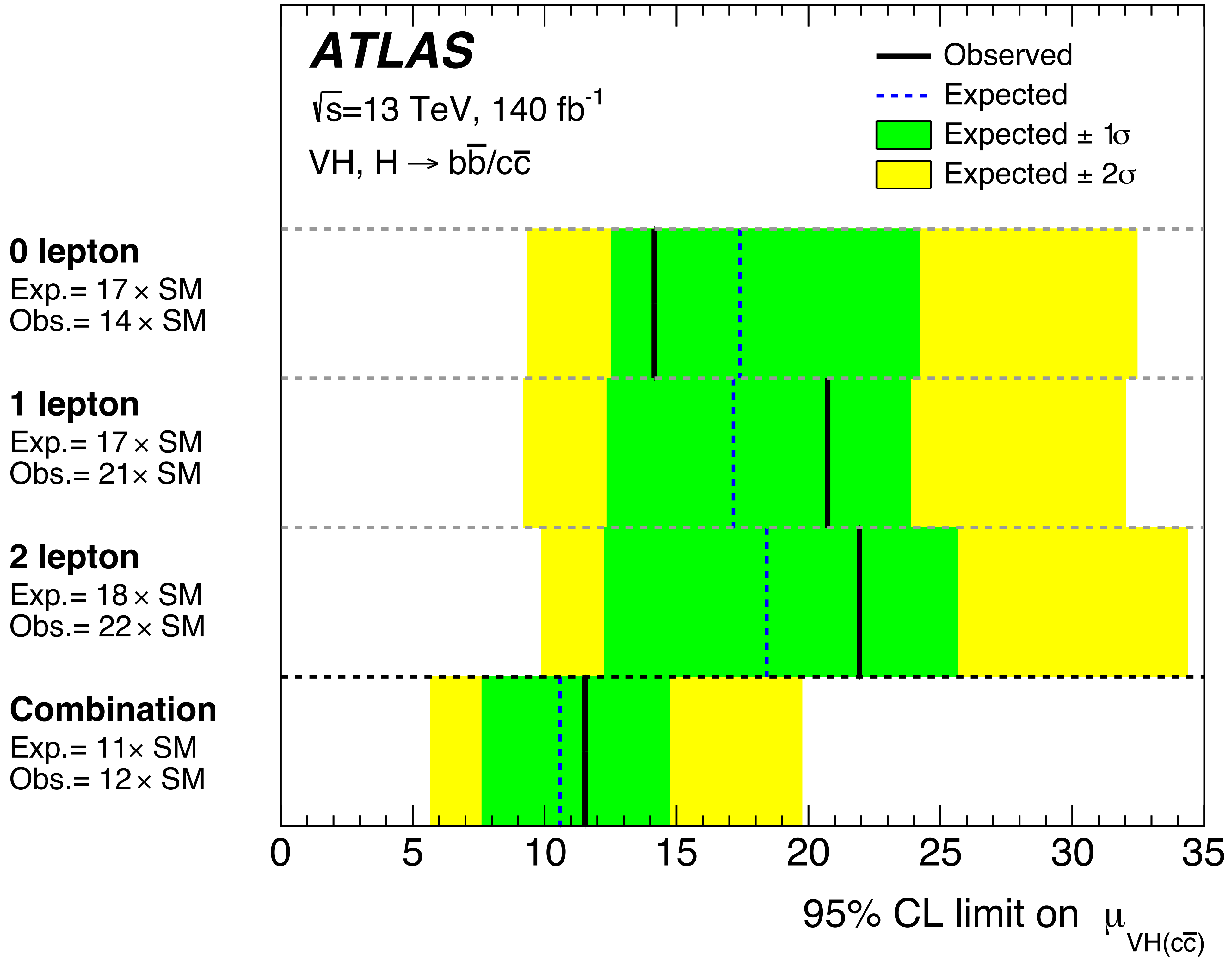
- Measure Yukawa coupling to *c*-quarks
- $H \rightarrow c\bar{c}$ decay rare: ~3%
- *VH* production mechanism
  - Significant QCD multijet rejection
  - Split in channels: 0L, 1L and 2L
- Reconstruct jet
  - Spray of hadrons





Previous ATLAS results: legacy analysis

- $VH(\to c\bar{c})$ legacy analysis
  - Used Run-2 data
- Constraints (95% confidence interval)
  - $\mu_{VHcc} < 11.5$
    - World-leading observed $VH(\to c\bar{c})$ limit
  - $|\kappa_c| < 4.2$

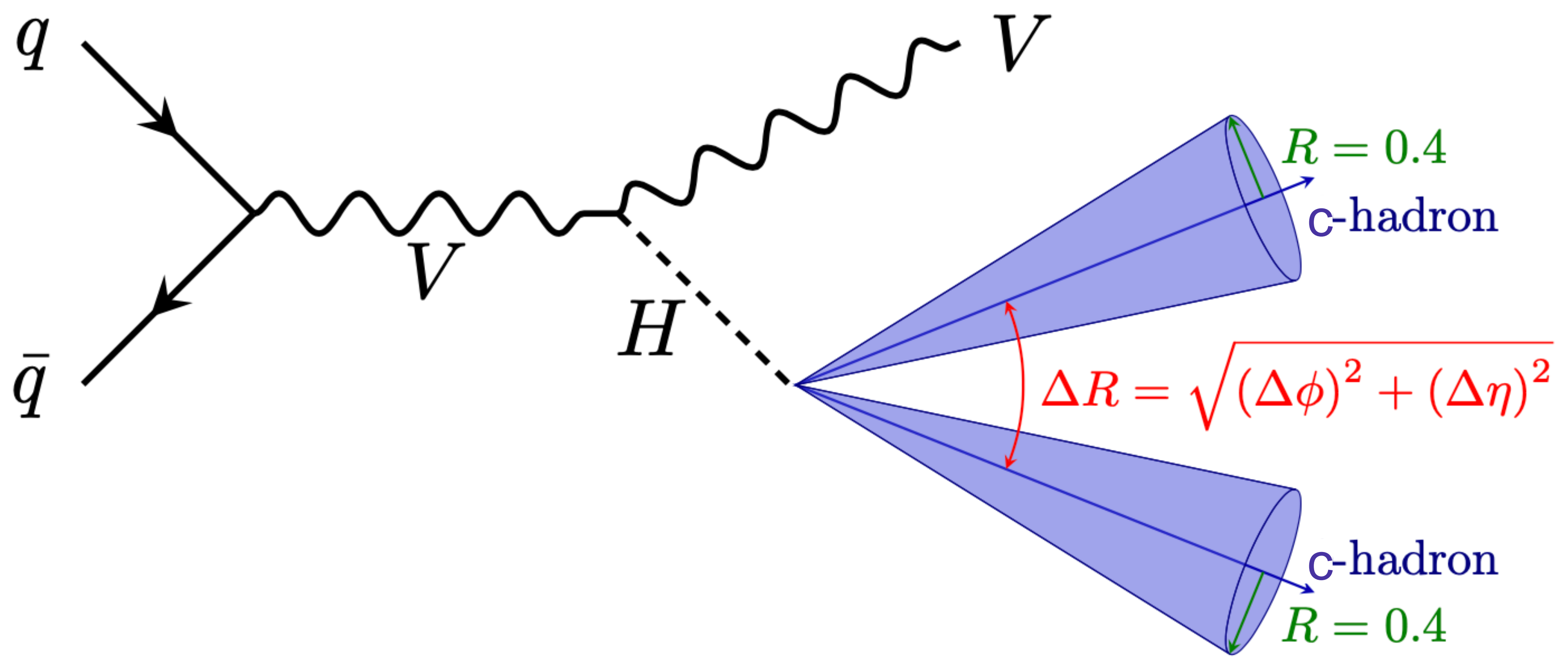


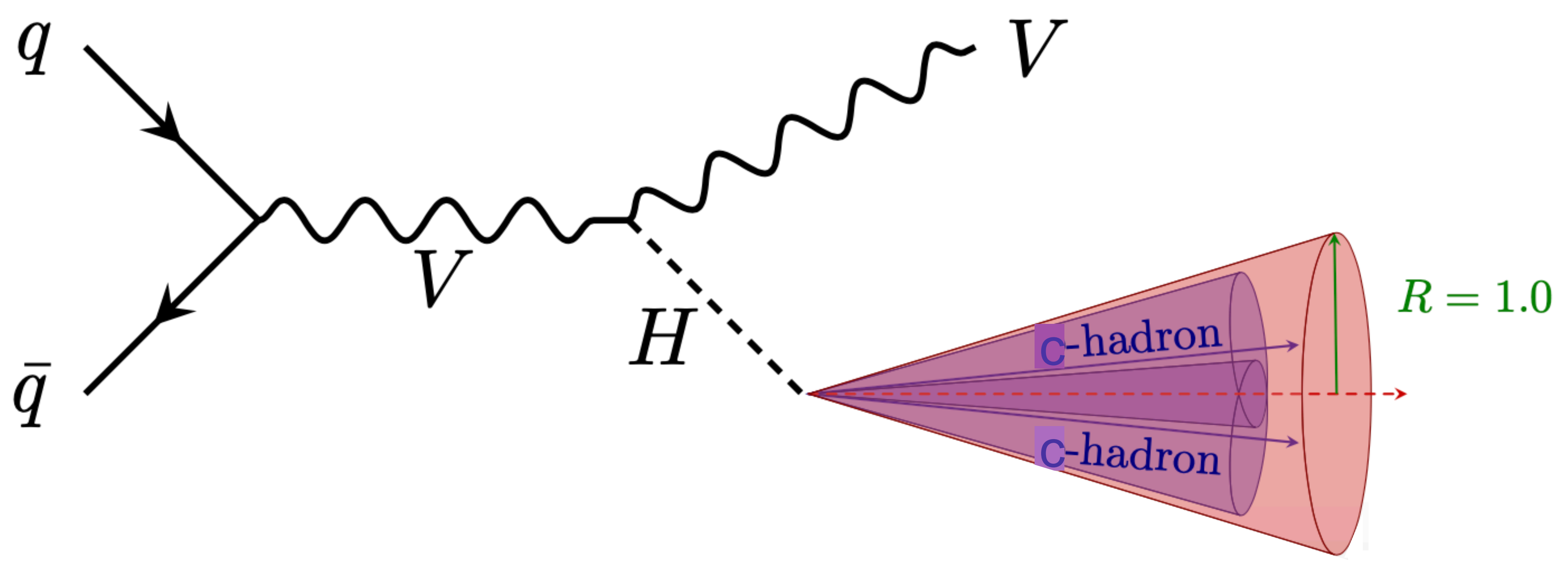
[ATLAS Collaboration. (2024). Measurements of WH and ZH production with Higgs boson decays into bottom quarks and direct constraints on the charm Yukawa coupling in 13 TeV pp collisions with the ATLAS detector. arXiv preprint arXiv:2410.19611.]

Boosted VH(cc)

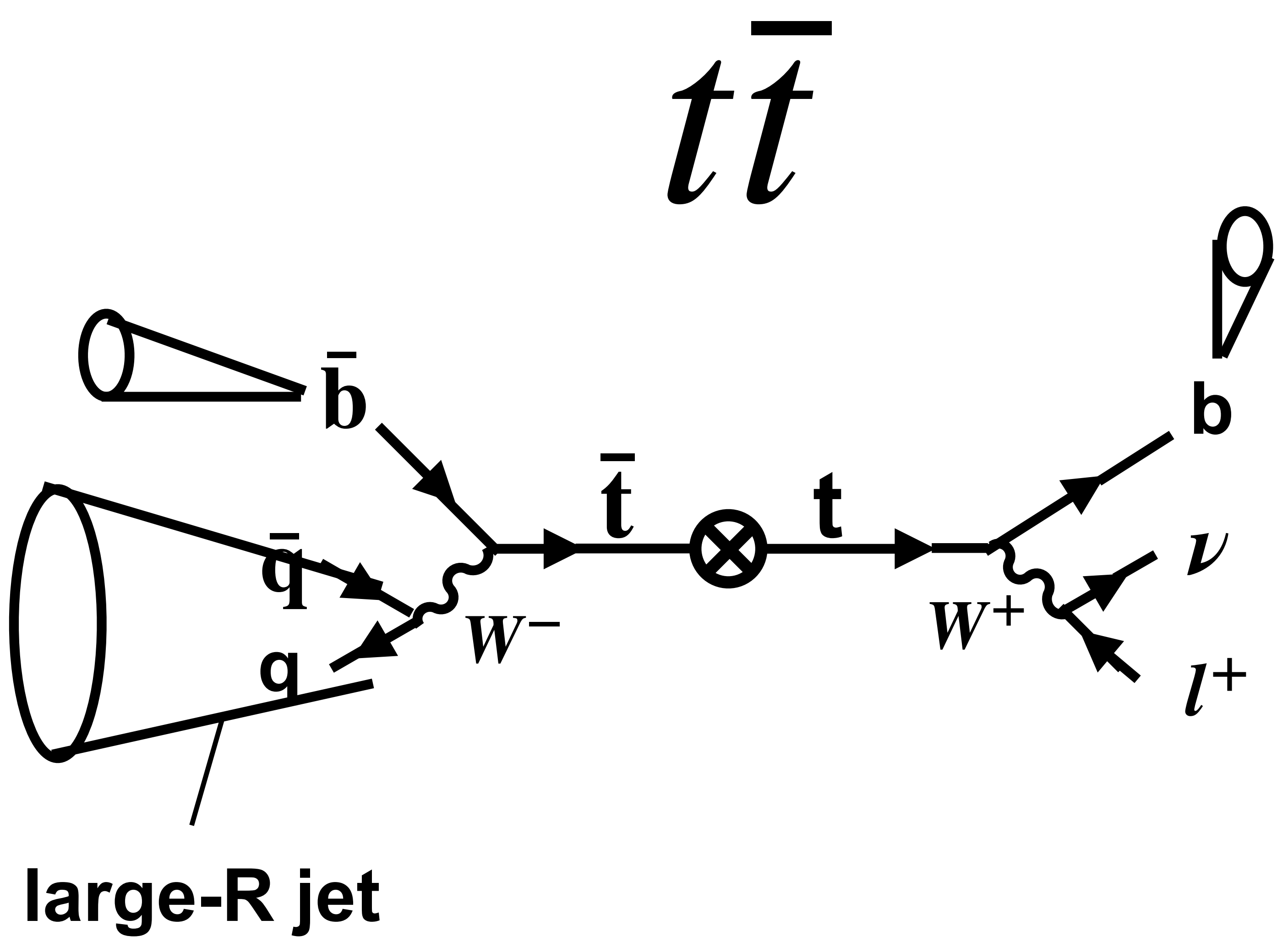
$$\Delta R(jet1, jet2) \approx \frac{2m_H}{p_T^H}$$

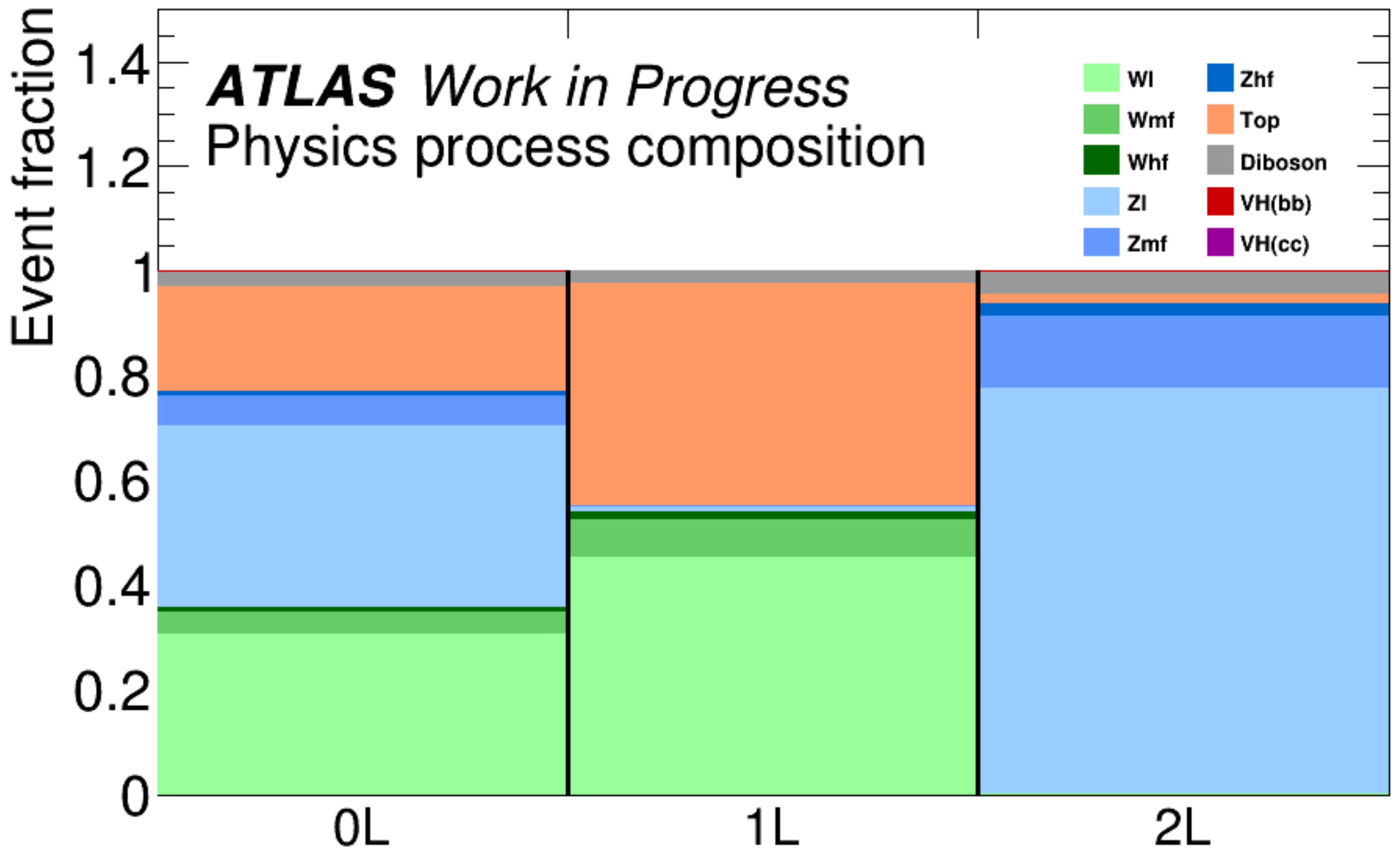
- Higgs boson with low transverse momentum (p_T)
  - Decays into 2 resolved charm-jets
  - Reconstruct as small-R jets
- Higgs boson with high p_T
  - Jets become merged (large-R jet)
  - Requires different reconstruction
  - Unexplored region in VH(cc) analysis

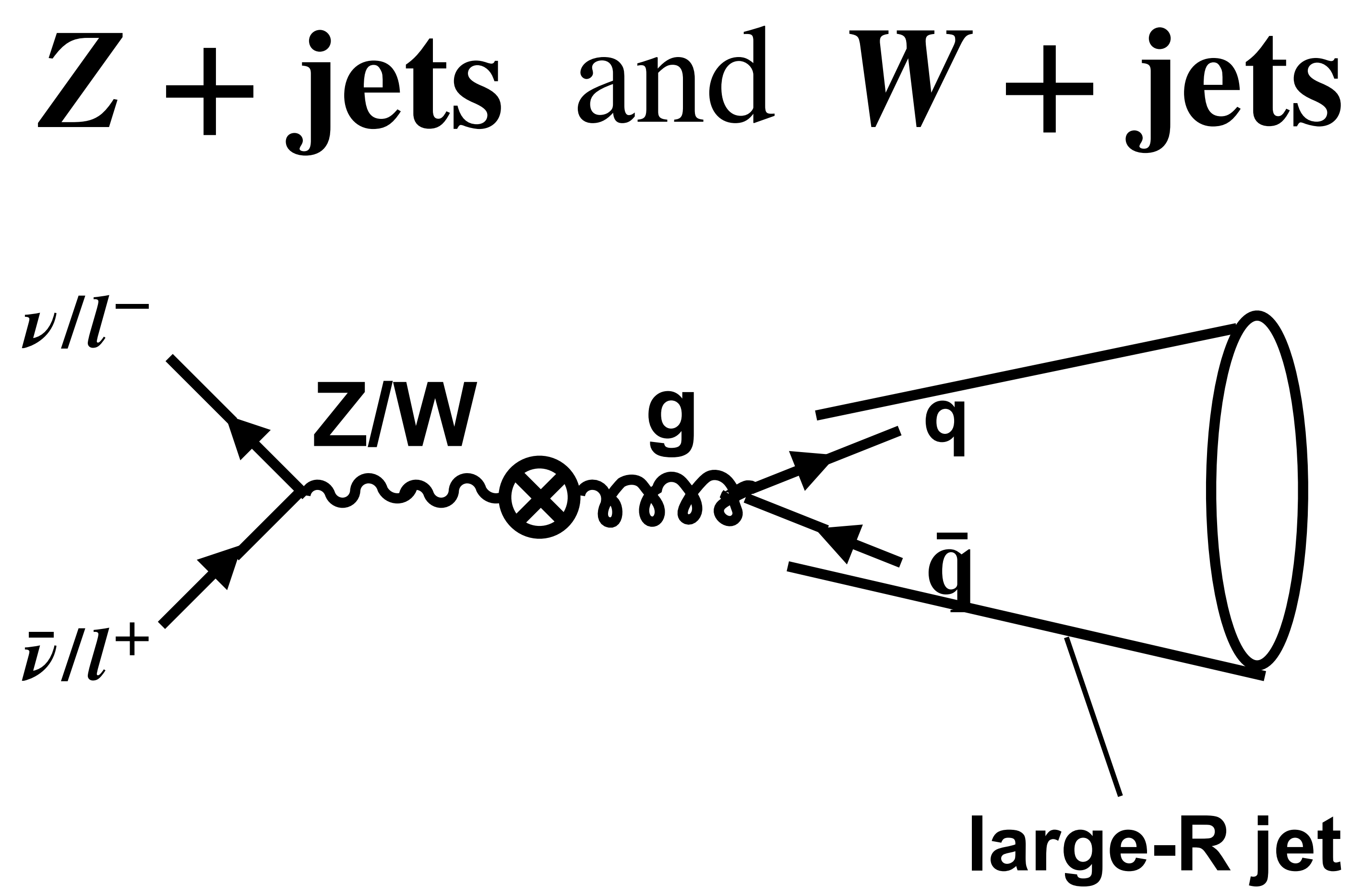




Background contributions



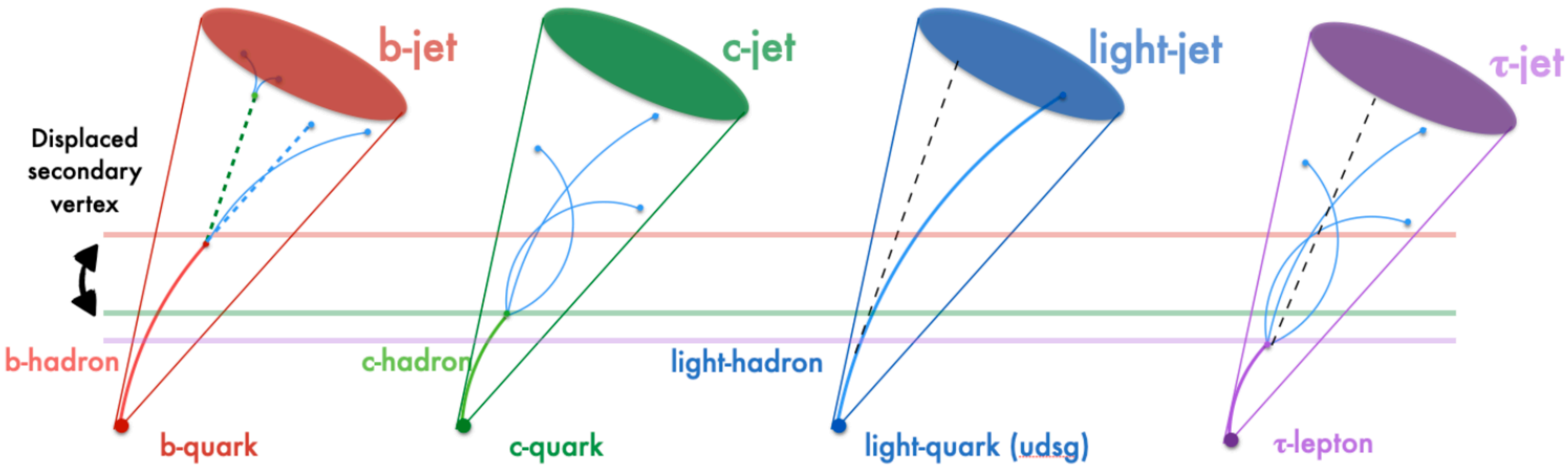




- Dominant backgrounds before additional event selection
  - Top events (mostly $t\bar{t}$)
  - *W/Z* in association with jets

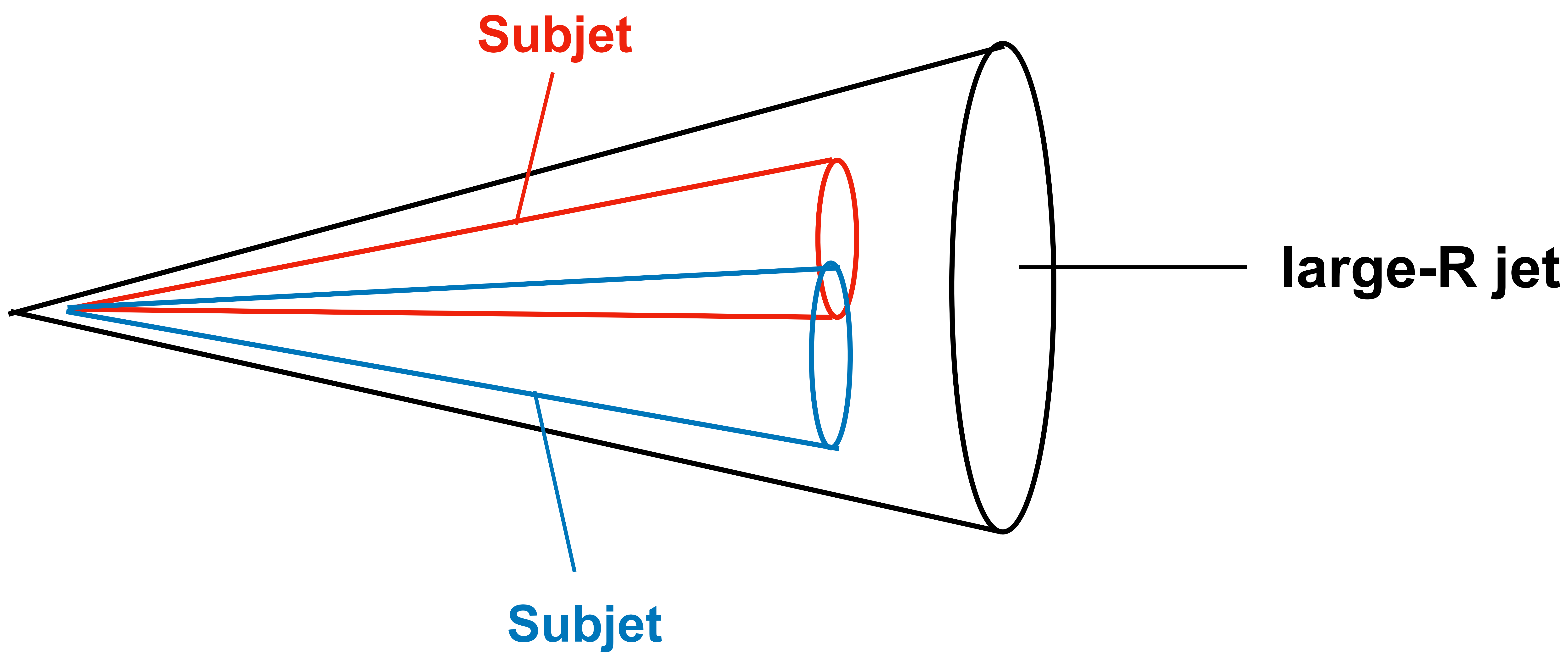
Flavour tagging

[Stamenkovic, M. PhD thesis: Charming Higgs bosons.]



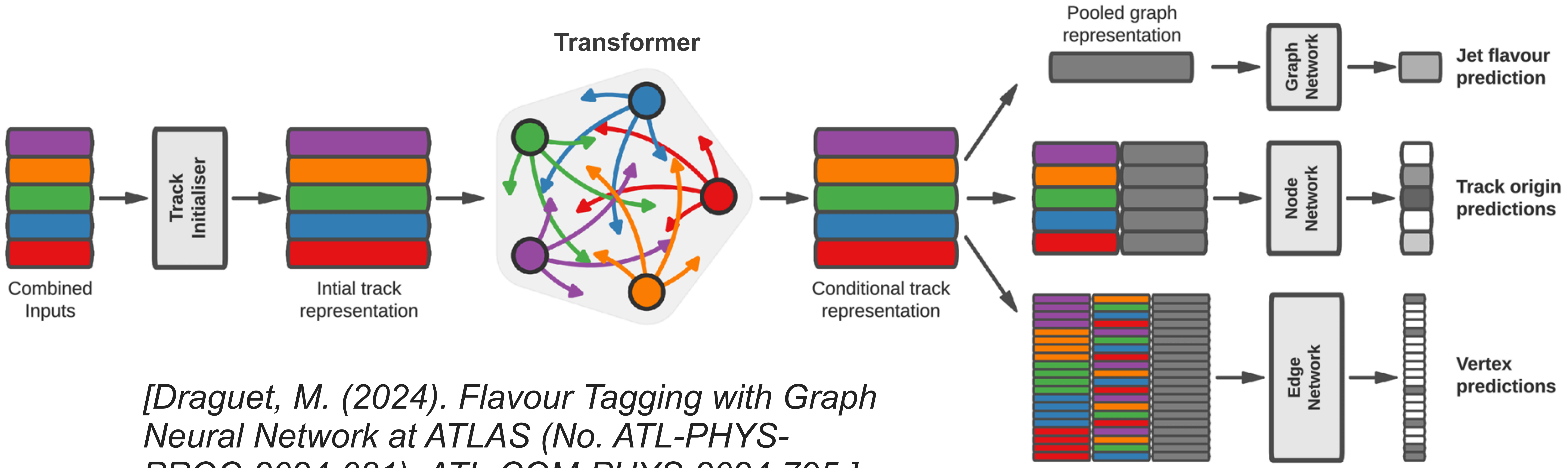
- Further background suppression → flavour tagging
  - Identification of original quark flavour (*b*, *c*, light)
- Use different track structures within jet
  - Use machine learning for optimal identification
- *c*-jets hard to distinguish

Boosted flavour-tagging



- We need to tag the large-R jet directly!

Boosted flavour-tagger: GN2X



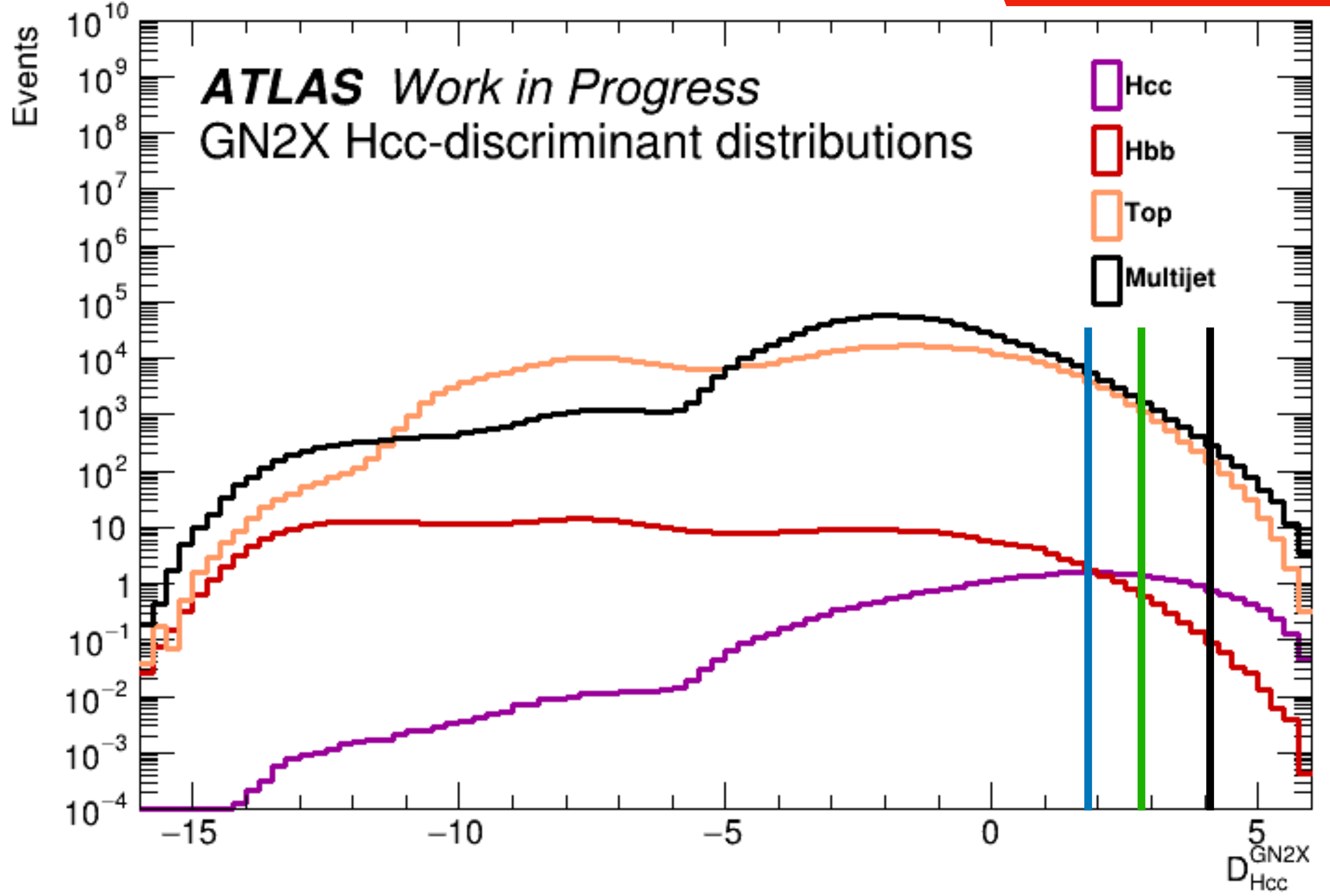
[Draguet, M. (2024). Flavour Tagging with Graph Neural Network at ATLAS (No. ATL-PHYS-PROC-2024-081). ATL-COM-PHYS-2024-795.]

- New boosted flavour-tagger
  - GN2X
  - Transformer-based architecture
  - Track-level input
- Expect VH(cc) improvements
  - Higher background rejection
  - Higher VH(cc) efficiency
  - Motivates boosted analysis

Event selection

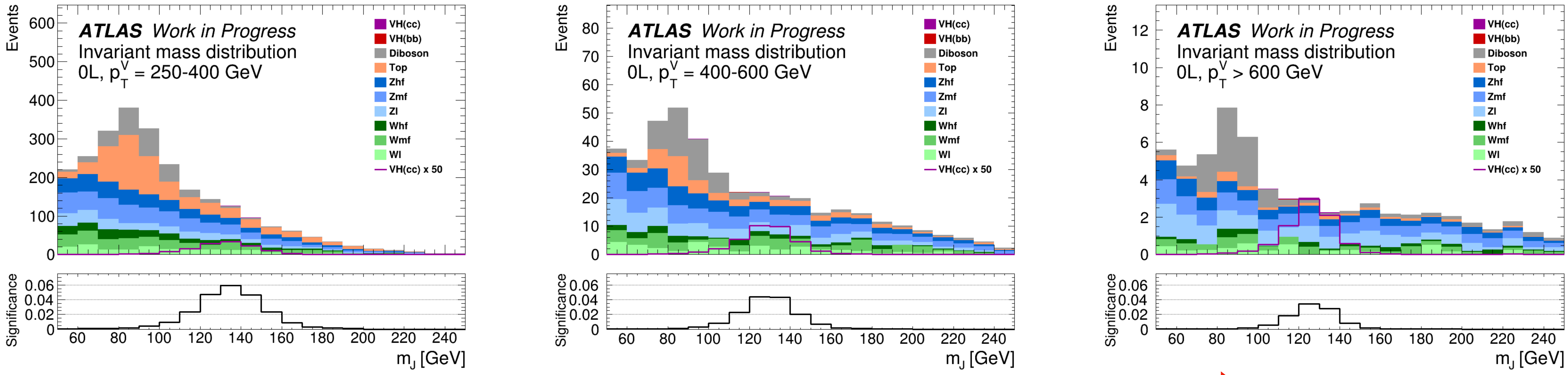
- GN2X provides probability scores
  - $p_{Hcc}, p_{Hbb}, p_{top}, p_{QCD}$
- Construct 1D discriminant
- Cut on samples using discriminant
  - Working point (WP) corresponding to ϵ_{Hcc}
  - Tight WP: high background rejection
  - Loose WP: high signal efficiency
- 30% was found to be optimal

$$D_{GN2X}^{Hcc} = log(\frac{p_{Hcc}}{f_{Hbb}p_{Hbb} + f_{top}p_{top} + (1 - f_{Hbb} - f_{top})p_{QCD}})$$



Working point	Hcc	Hbb	Top	Multijet
10%	10%	0.06%	0.09%	0.10%
30%	30%	0.57%	0.95%	0.82%
50%	50%	2.3%	4.2%	3.3%

GN2X-tagged invariant mass distributions



- Higher VH(cc) purity
- Lower statistical power
- Lower sensitivity

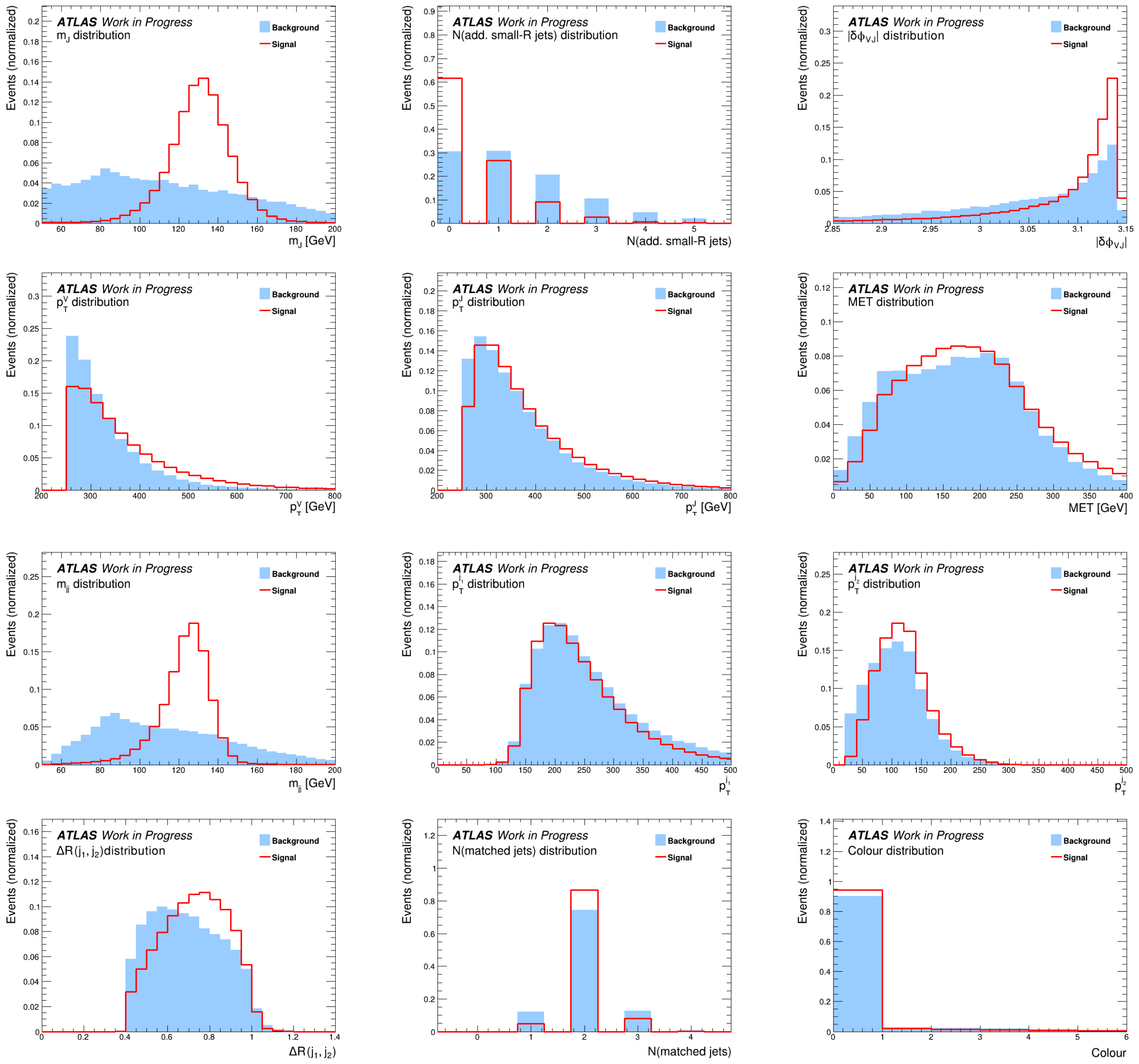
BDT variable selection studies

- Use Boosted Decision Tree (BDT)
  - Enhance signal-to-background separation
- Apply optimized event selection
- Input variable selection
  - 6 event-level variables
    - Up to 40% improvement
  - 6 subjet-level variables
    - Additional 2%-14% improvement

Event-level variables
m_J
p_T^V
p_T^J
MET
$N(\mathrm{add.jets})$
$\delta\phi(V,J)$

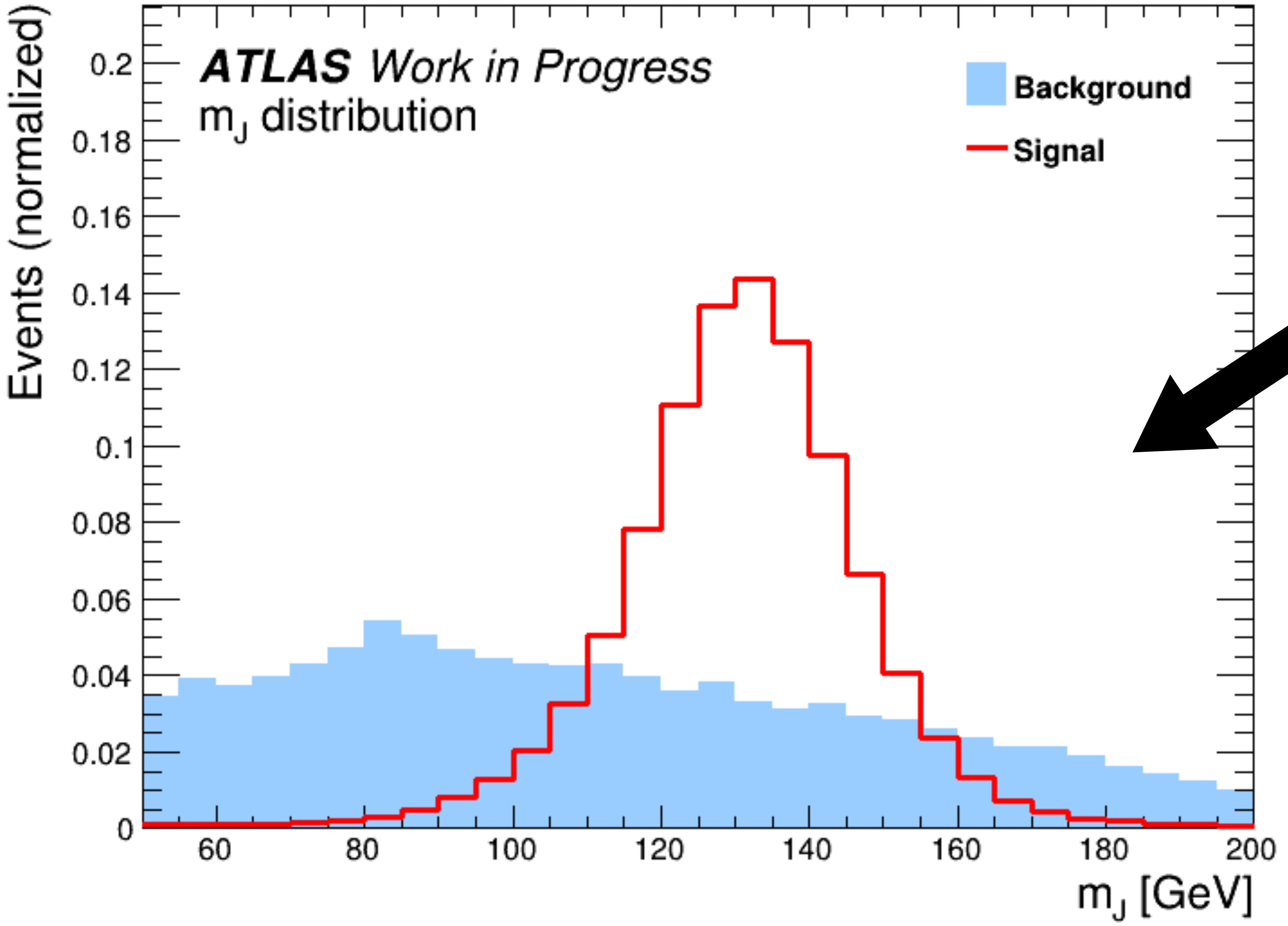
Subjet-level variables
m_{jj}
$p_T^{j_1}$
$p_T^{j_2}$
$\Delta R(j_1, j_2)$
$N(\text{matched jets})$
Colour

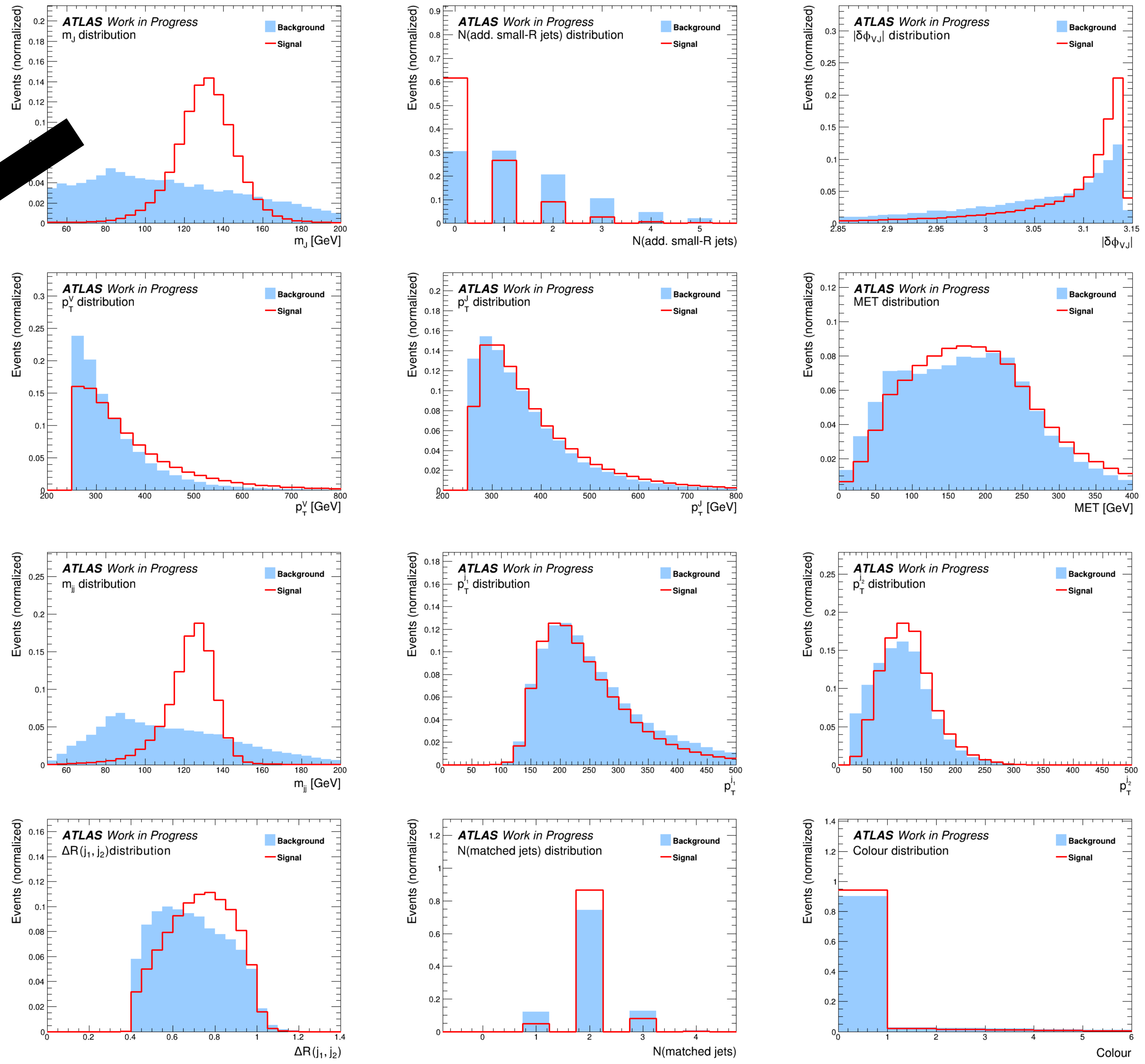
Discriminating power of variables



Discriminating power of variables

Invariant mass

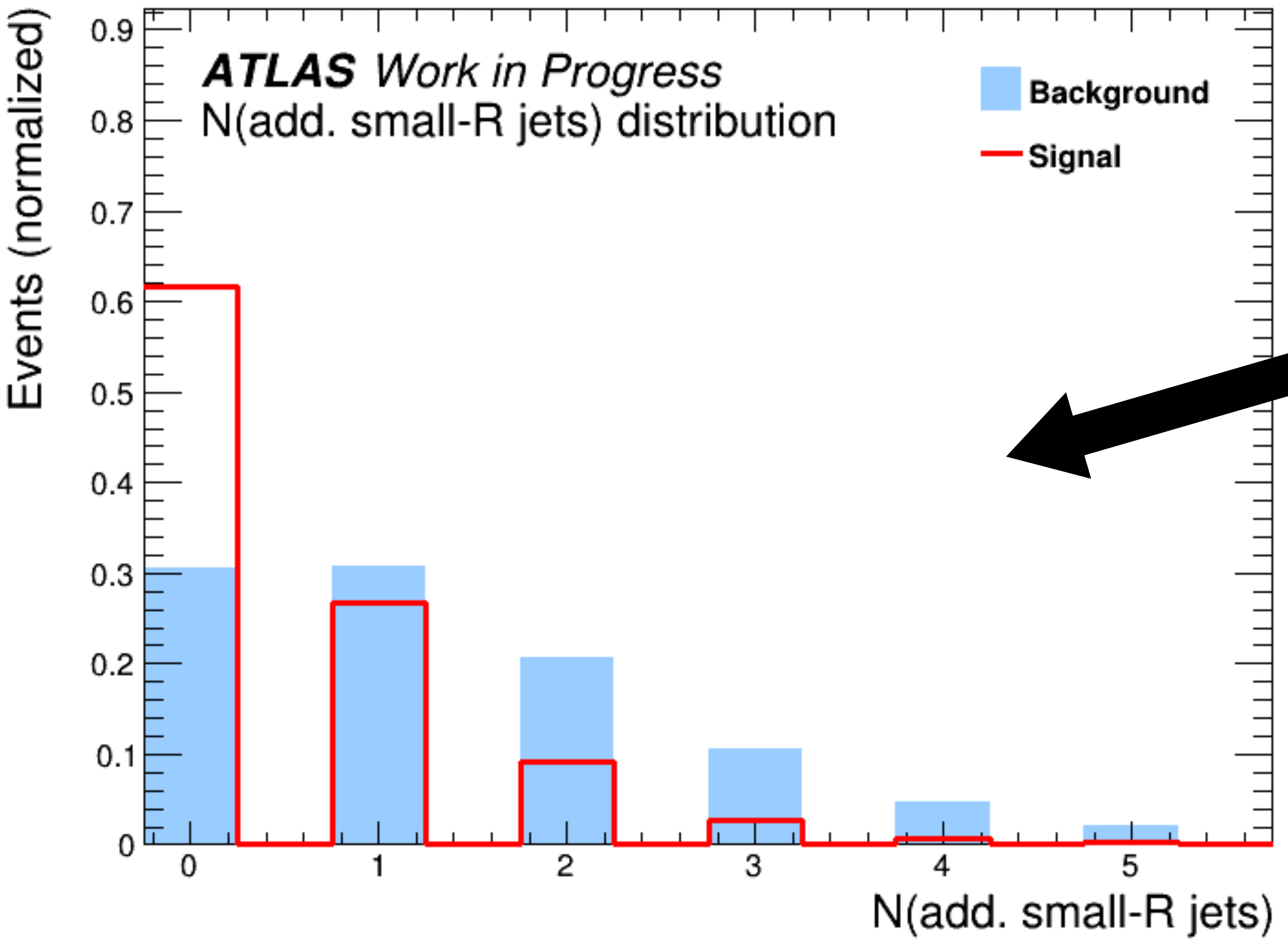


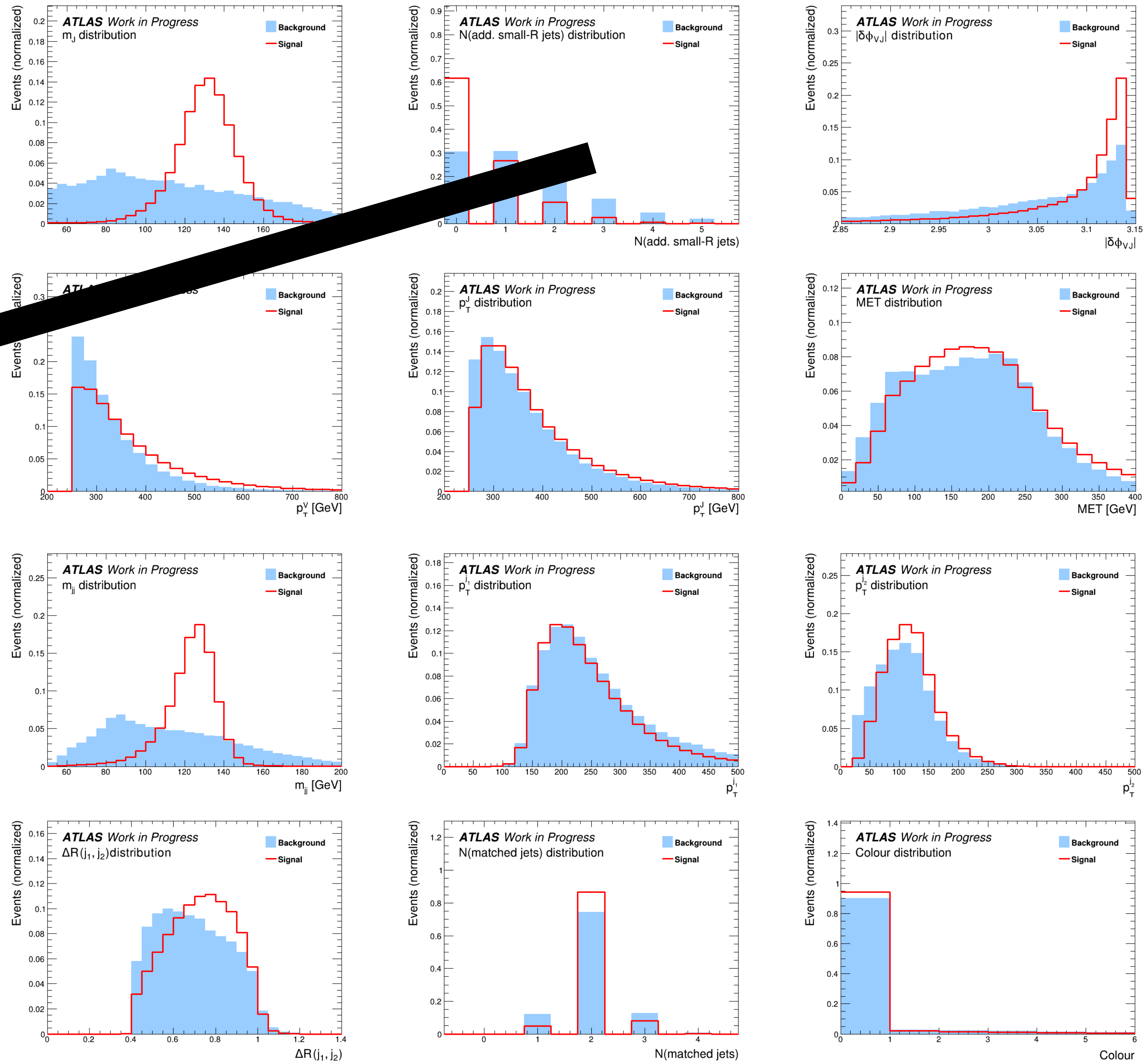


- Most important separating variable
  - Due to Higgs resonance

Discriminating power of variables

Number of additional small-R jets

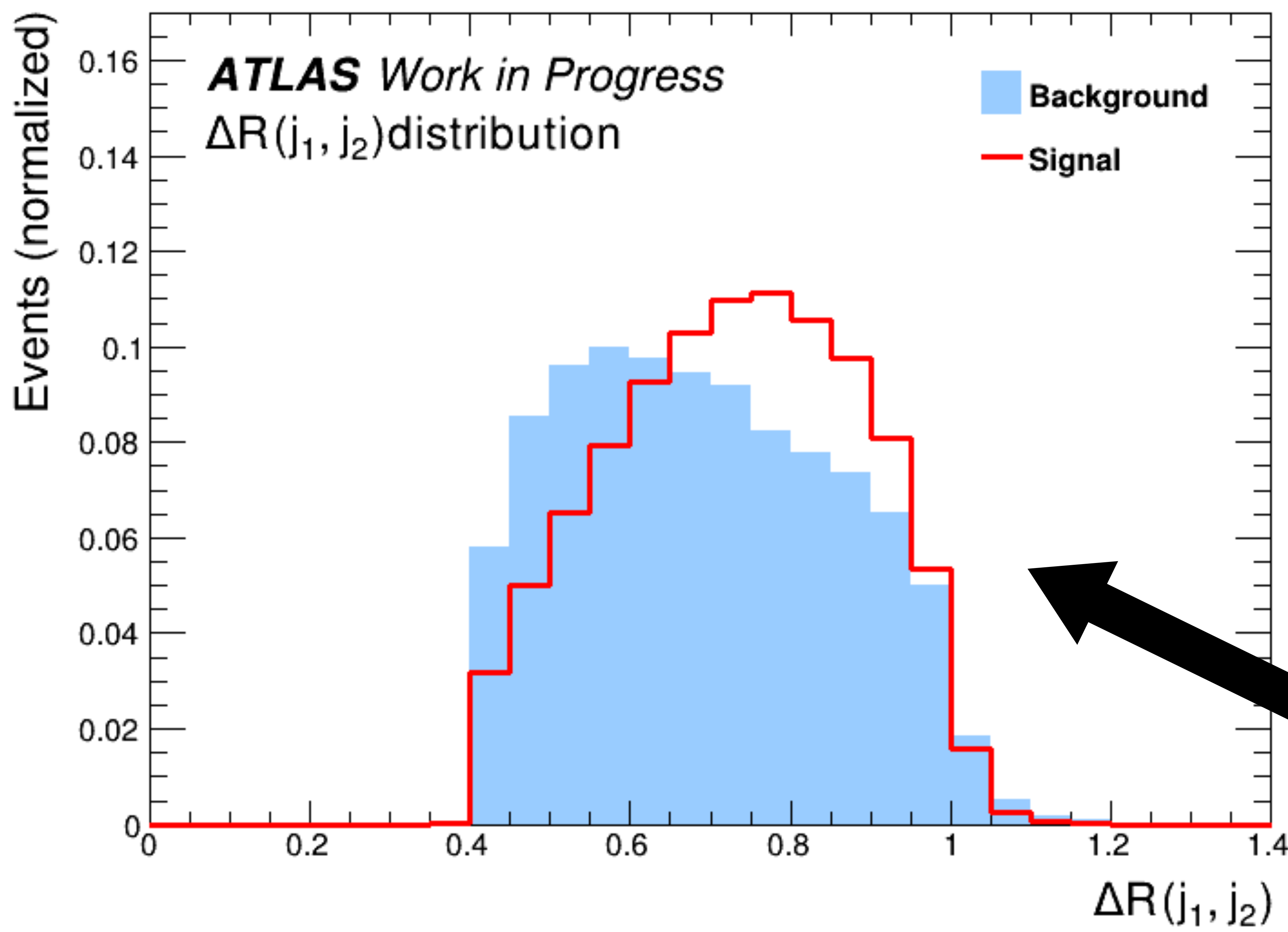


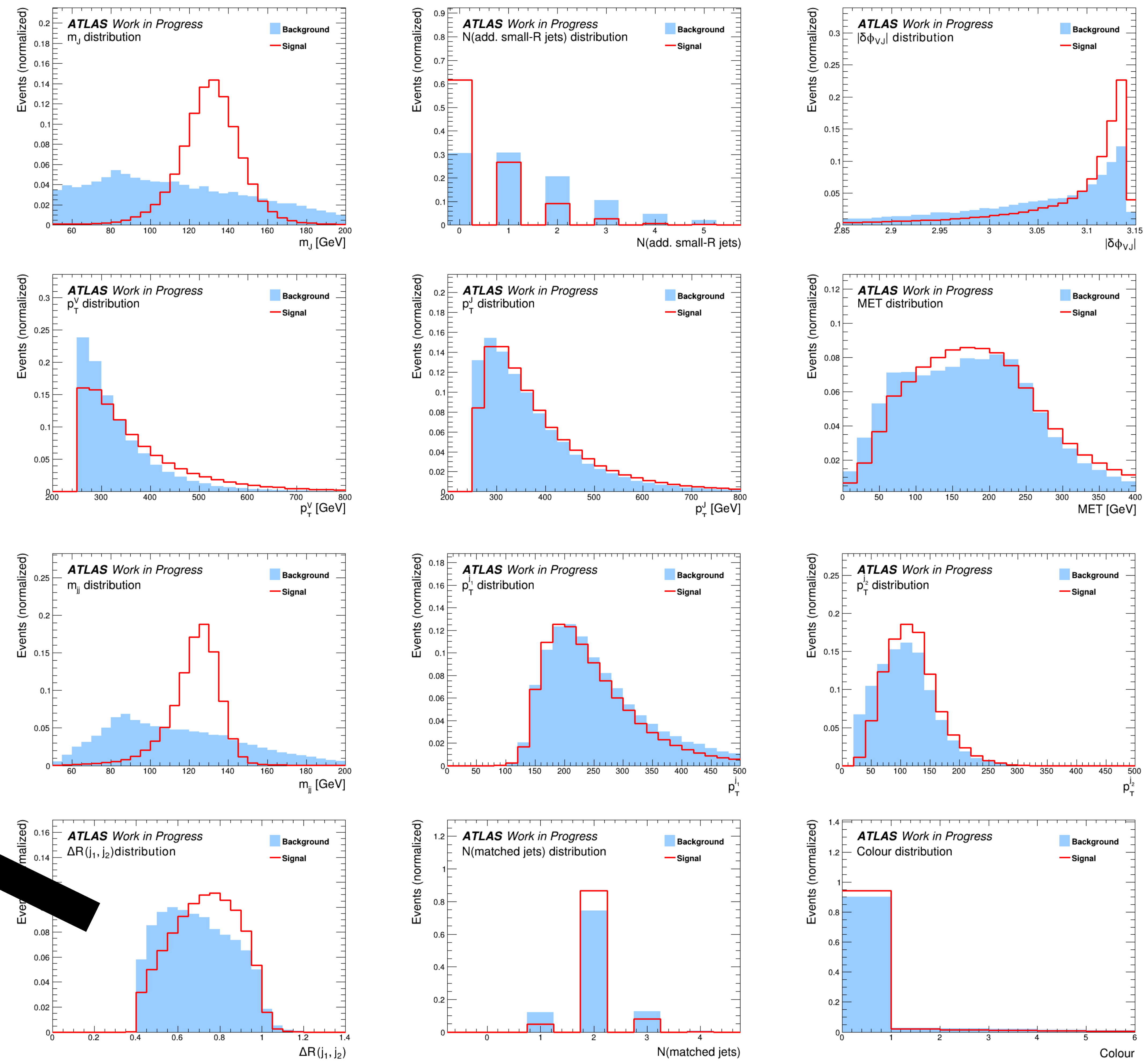


- VH(cc) dominant at 0 add. jets
  - Mainly expect only two charm-jets

Discriminating power of variables

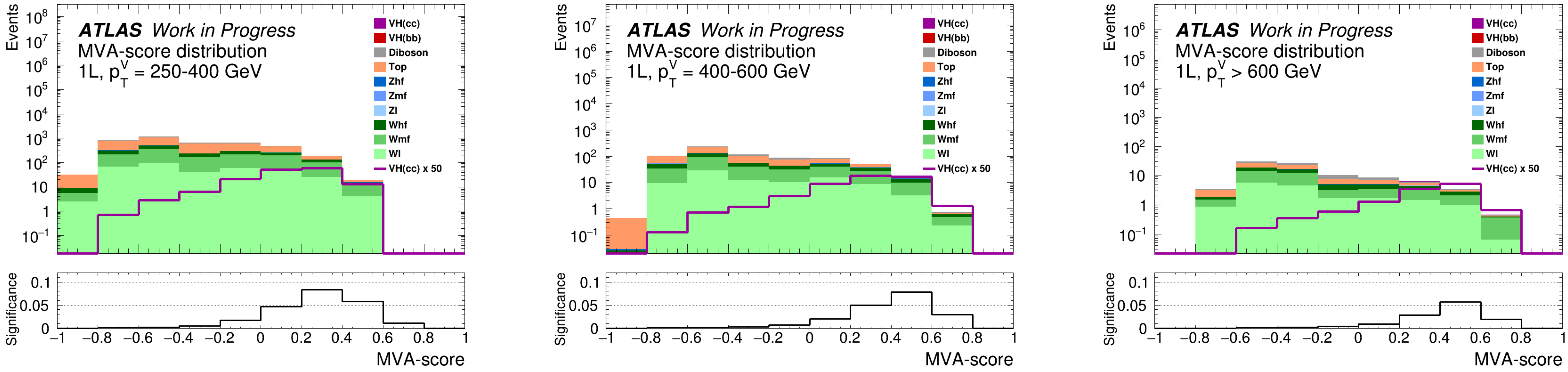
Angular separation of the 2 subjets





- VH(cc) generally has larger separation
  - Reflects correlation with m_J

BDT distributions



- Output BDT: MVA-score
- Demonstrates enhanced separation
- Up to 50% sensitivity improvement

Comparison to Run-2 resolved $VH(\rightarrow c\bar{c})$

- Improvements across lepton channels
  - Significant improvements in 0L and 1L
  - Sensitivity decrease in 2L
    - Limited training sample for BDT
- Larger p_T shows larger improvements
  - Larger abundance of merged jets
  - More useful for higher merged jets
  - Shows great potential in high-p_T

Lepton channel	p_T^V [GeV]	Resolved	Boosted	Difference
0L	250-400	0.101σ	0.136σ	+35%
0L	> 400	0.058σ	0.122σ	+111%
1L	250-400	0.098σ	0.118σ	+20%
1L	> 400	0.061σ	0.126σ	+107%
2L	250-400	0.083σ	0.080σ	-3.6%
2L	> 400	0.047σ	0.068σ	+44%

Conclusions

Measuring the charm Yukawa coupling important test of SM

First sensitivity estimates of $VH(\rightarrow c\bar{c})$ in the Lorentz-boosted regime

Conclusions

Measuring the charm Yukawa coupling important test of SM

First sensitivity estimates of $VH(\rightarrow c\bar{c})$ in the Lorentz-boosted regime

Previous Resolved Analysis

- Small-R jet flavour-tagging
- Optimal in low-p_T regime

Conclusions

Measuring the charm Yukawa coupling important test of SM

First sensitivity estimates of $VH(\to c\bar{c})$ in the Lorentz-boosted regime

Previous Resolved Analysis

$p_T > 400\,\mathrm{GeV}$

+100%

New Boosted Analysis

Previous Resolved Analysis:

- Small-R jet flavour-tagging
- Optimal in low-p_T regime

New Boosted Analysis:

- Large-R jet flavour-tagging
- Optimal in high-p_T regime

Conclusions

Measuring the charm Yukawa coupling important test of SM

First sensitivity estimates of *VH*($\to c\bar{c}$) in the Lorentz-boosted regime

Previous Resolved Analysis — $p_T > 400\,\text{GeV}$ → **+100%** → **New Boosted Analysis**

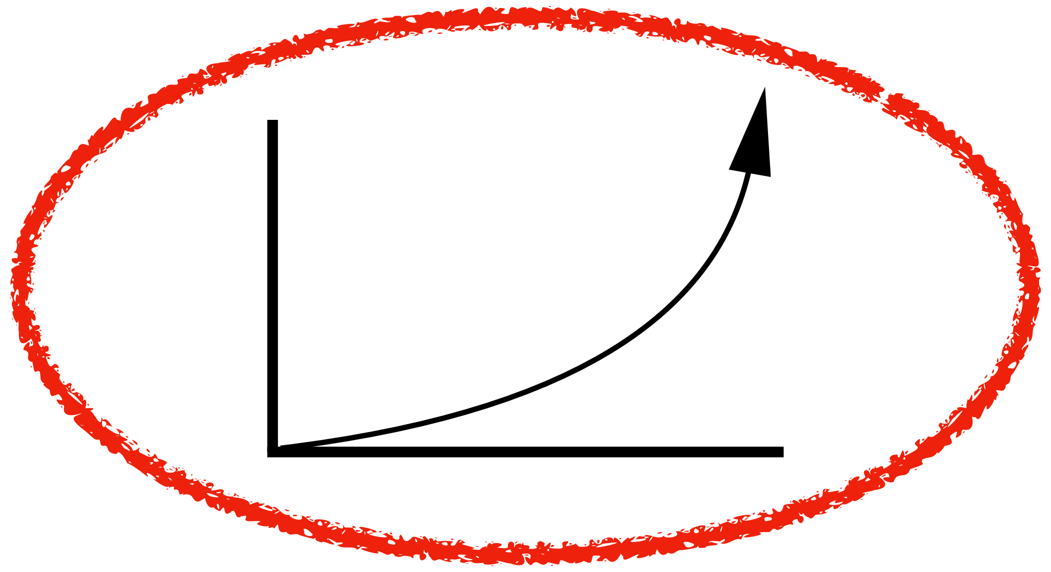
- Small-R jet flavour-tagging
- Optimal in low-p_T regime

- Large-R jet flavour-tagging
- Optimal in high-p_T regime

Using boosted jets in the high-$\mathrm{p_T}$ regime is crucial for probing the Yukawa charm coupling!

Outlook

New data



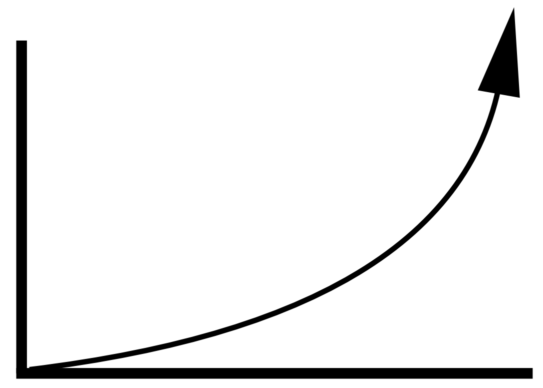
HL-LHC

Significance: 1.2σ

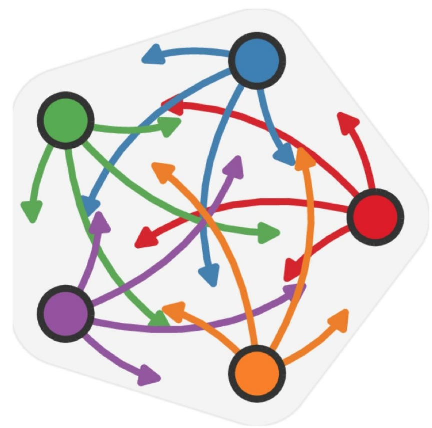
Expected sensitivity of the ATLAS experiment to $H \to bb$ and $H \to cc$ decays in the VH production mode at the High Luminosity LHC. Technical report, CERN, Geneva, 2025.

Outlook

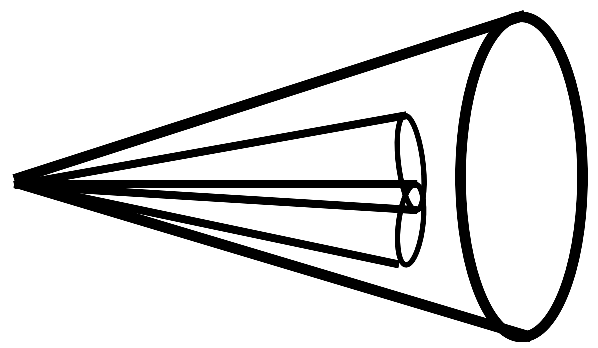
New data



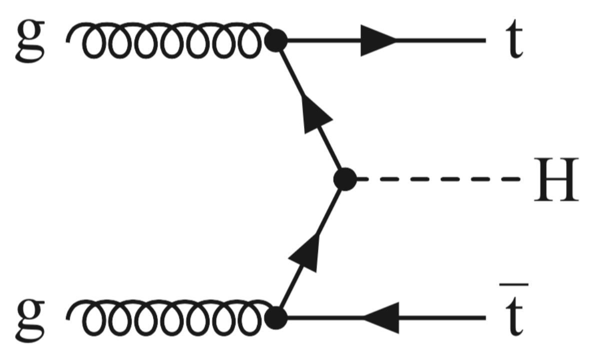
Flavour tagging



Boosted topologies



Additional productions



HL-LHC

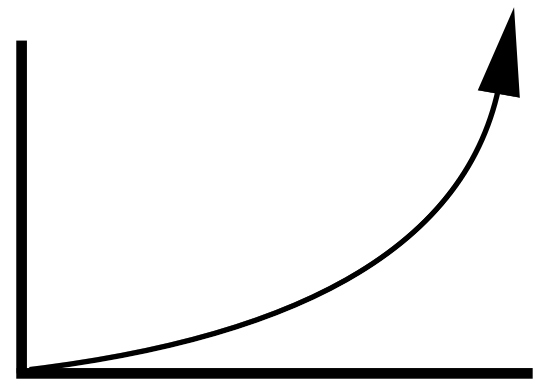
Significance: ~~1.2σ~~

$> 1.2\sigma$

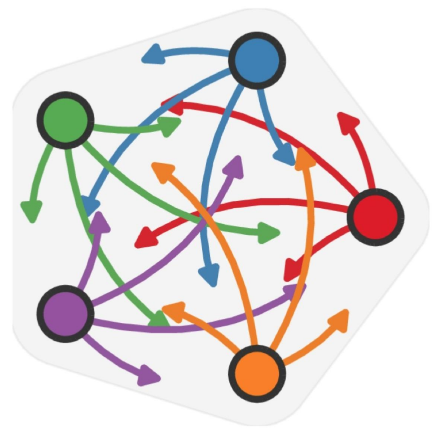
Expected sensitivity of the ATLAS experiment to $H \to bb$ and $H \to cc$ decays in the VH production mode at the High Luminosity LHC. Technical report, CERN, Geneva, 2025.

Outlook

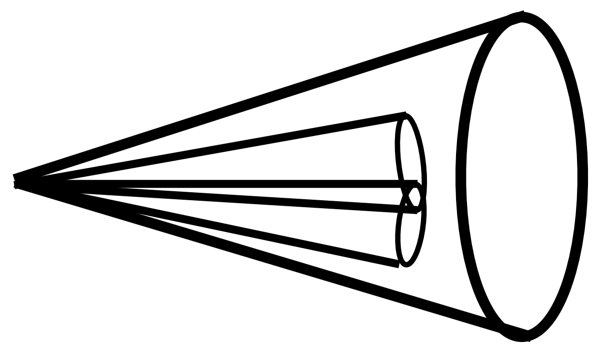
New data



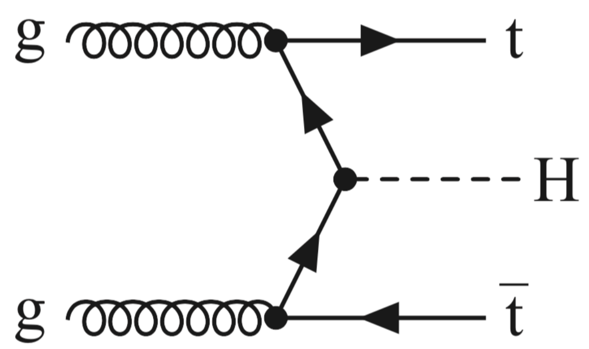
Flavour tagging



Boosted topologies



Additional productions



HL-LHC

Significance: $\cancel{1.2}\sigma$
$> 1.2\sigma$

Expected sensitivity of the ATLAS experiment to H → bb and H → cc decays in the VH production mode at the High Luminosity LHC. Technical report, CERN, Geneva, 2025.

FCC-ee

Precision: 1.6%

M Selvaggi, A Blondel, and J Eysermans. Prospects in electroweak, Higgs and top physics at FCC. doi 10:2025-2026, 2025.

Backup

BSM impact on Yukawa couplings

2HDM model $\longrightarrow$ $A \quad H^- \; H^+$
$h \quad H$

Potential: $V(\phi, \phi')$

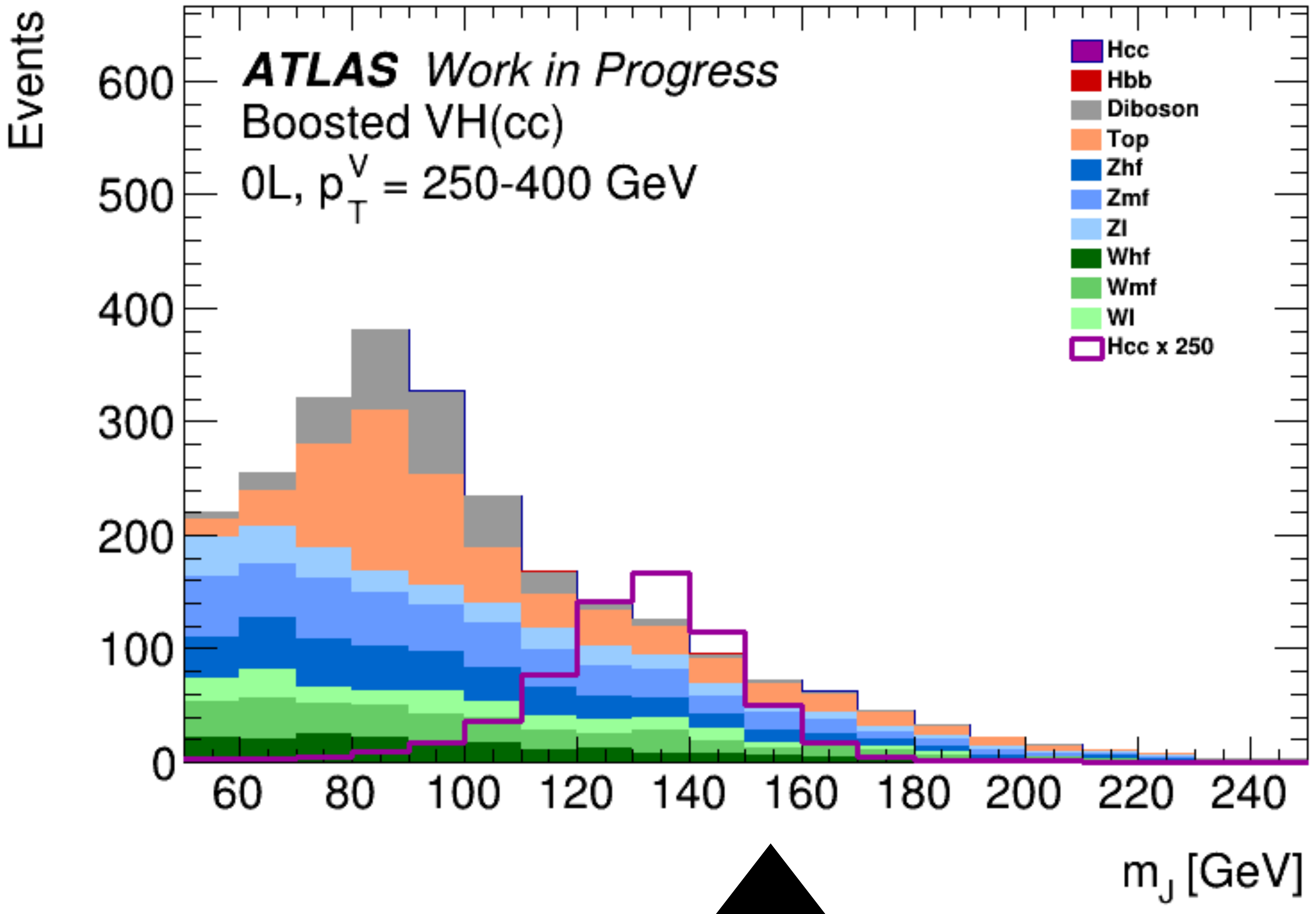
Example impact:
Particles couple to both scalar fields

$$\lambda^h_{q_u,\bar{q}_u} = y^{\mathrm{SM}}_{q_u} \sin(\beta - \alpha) + y'_{q_u} \cos(\beta - \alpha)$$

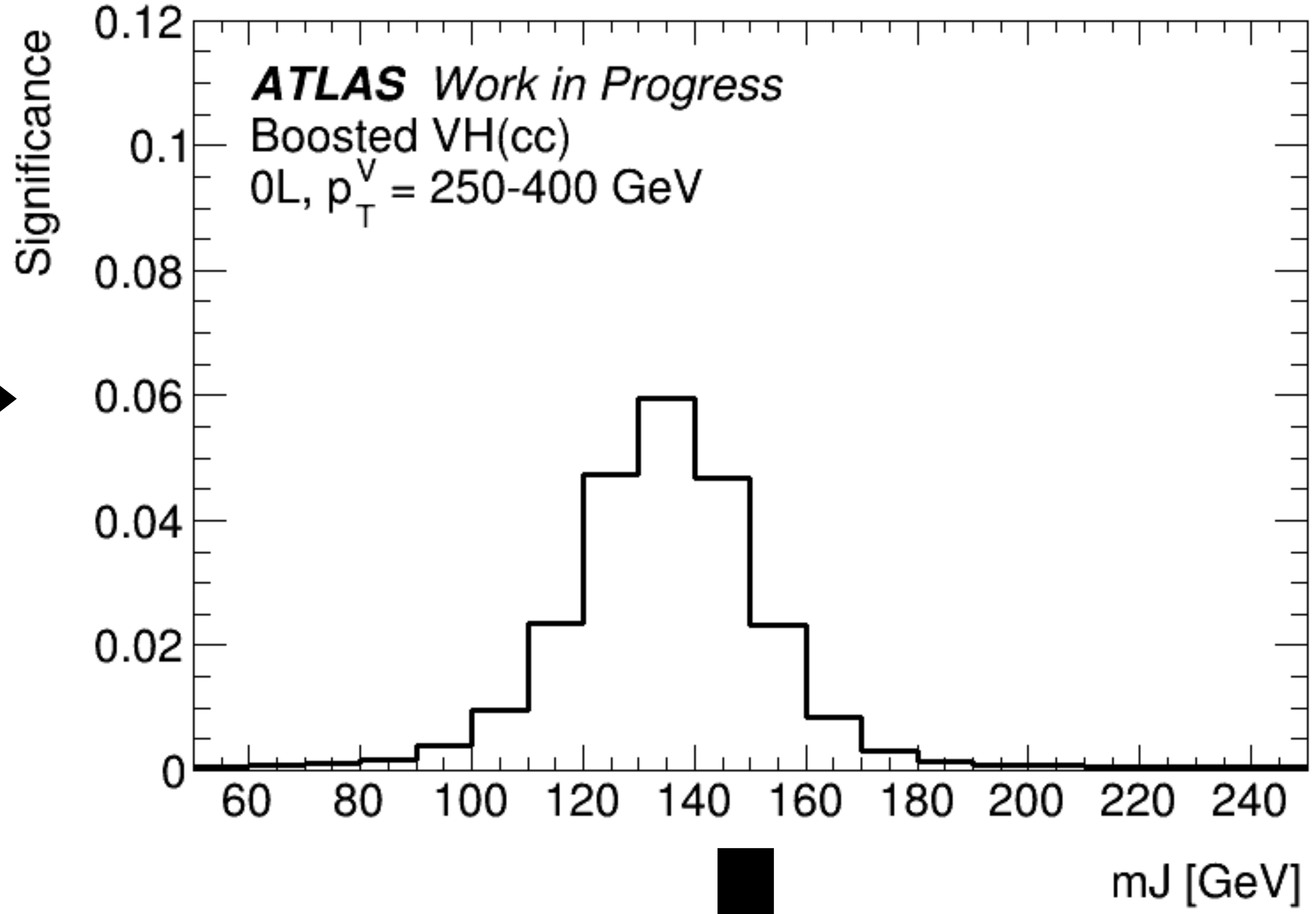
$\Downarrow$

Enhanced coupling possible for certain mixing and y

Estimating sensitivity using the invariant mass



$$\sigma_{\text{bin}} = \frac{S}{\sqrt{S+B}}$$

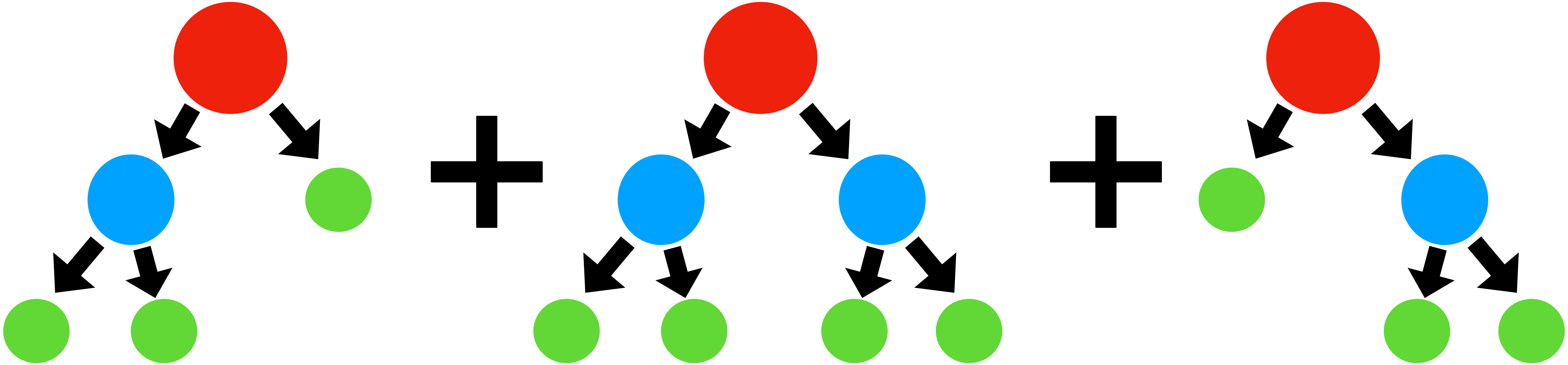


$$m_J^2 = E^2 - |\vec{p}\,|^2$$

$$\sigma_{\text{tot}}^2 = \sum_{\text{bin}=1}^{N_{\text{bins}}} \sigma_{\text{bin}}^2$$

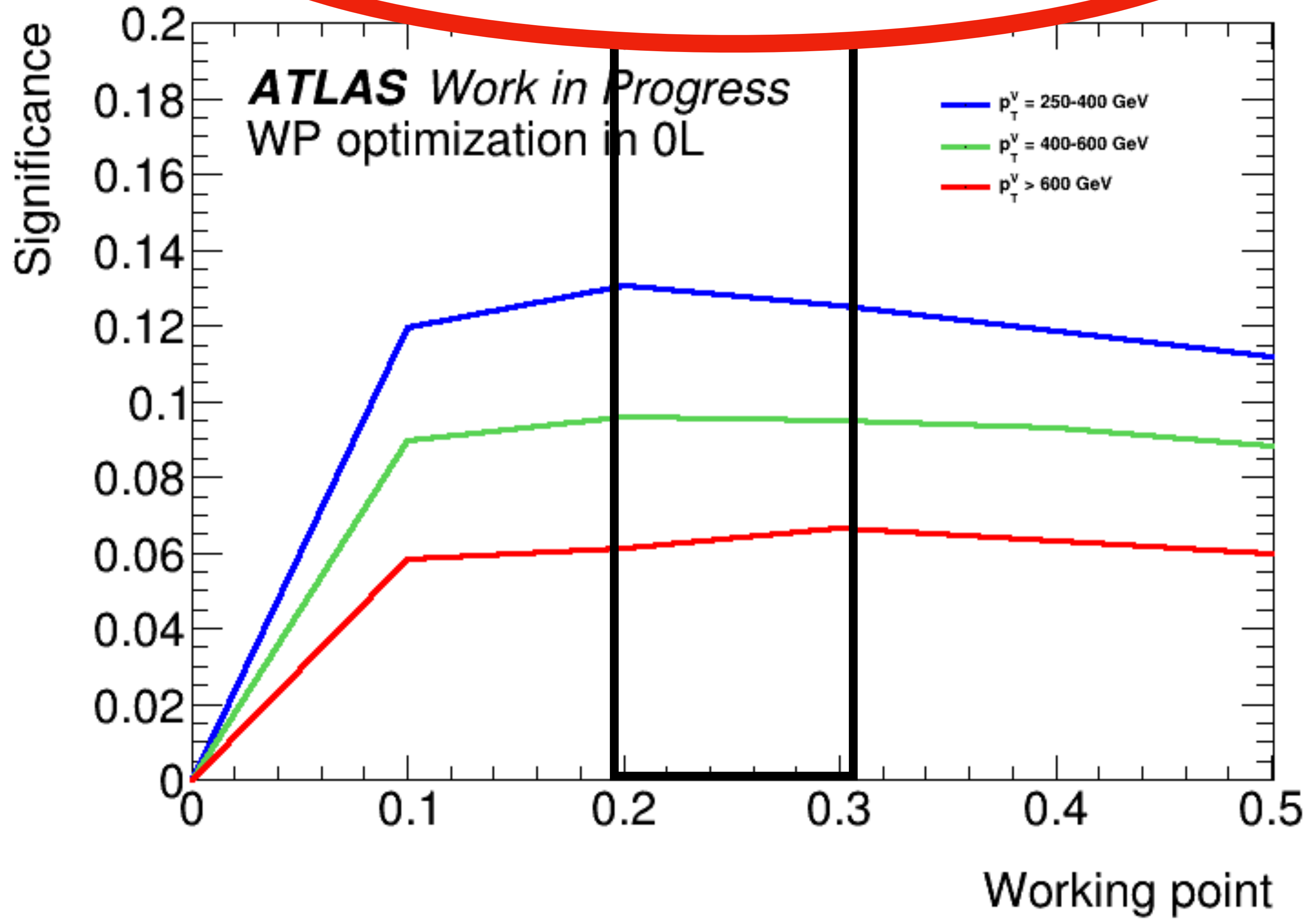
Estimating sensitivity using a BDT

- Exploit additional discriminating variables
  - Multivariate Approach (MVA) → improve signal-to-background separation
  - Use MVA-score as discriminant (instead of m_J)
  - Use Boosted Decision Tree (BDT)



Tagger parameter optimization

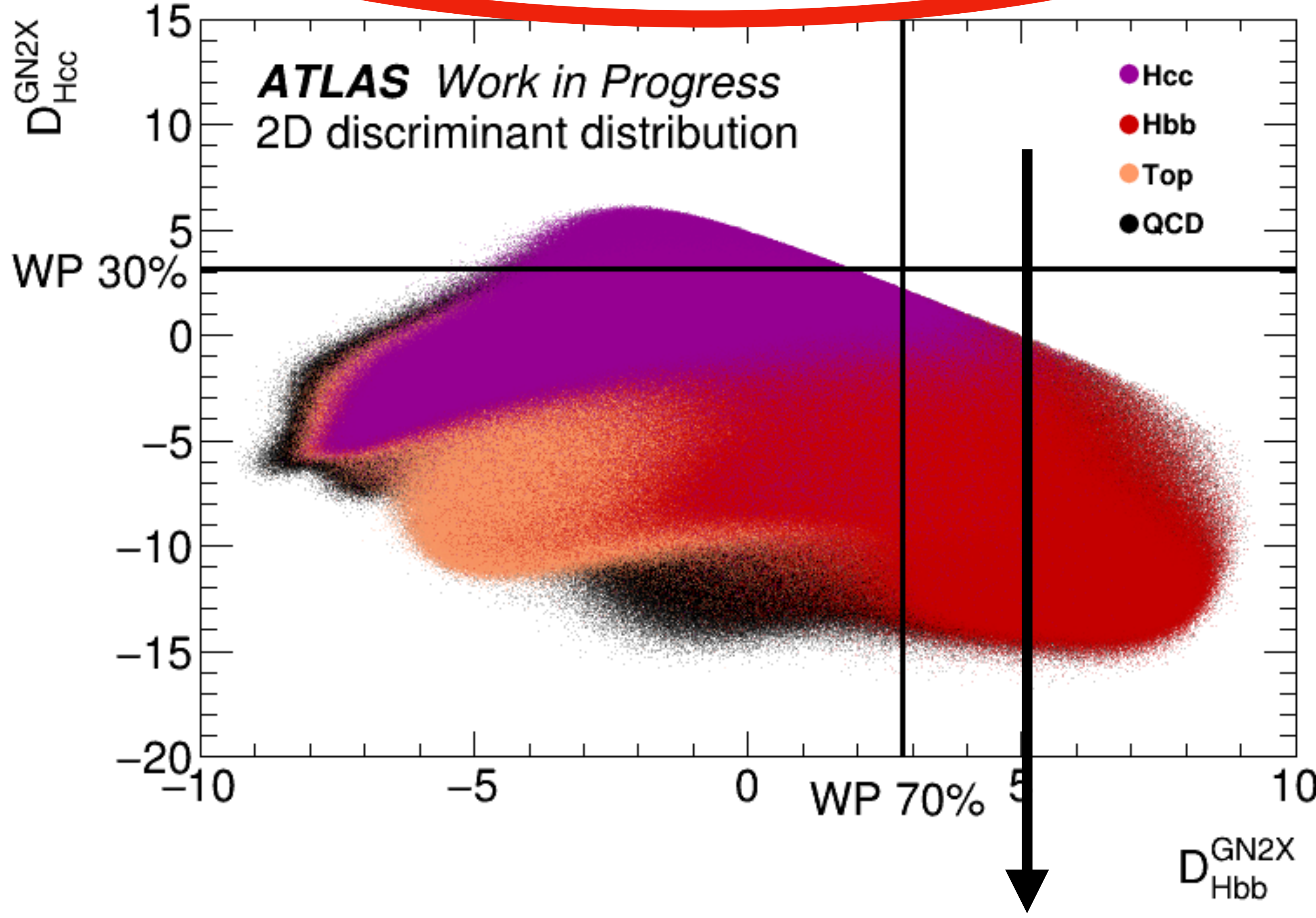
Optimal WP: 20%-30%



f_{Hbb} = 0.0 increases sensitivity

f_{Hbb} → / f_{top} ↓	0.0	0.1	0.2	0.3	0.4
0.45	+6.3%	+3.6%	-1.2%	-6.8%	-14.5%
0.35	+7.6%	+5.8%	+1.5%	-2.5%	-8.2%
0.25	+8.6%	+7.7%	+3.3%	Standard	-4.2%
0.15	+9.0%	+8.2%	+5.2%	+1.1%	-2.3%
0.05	+6.0%	+7.8%	+5.1%	+2.1%	-0.9%

No impact from Hbb-veto

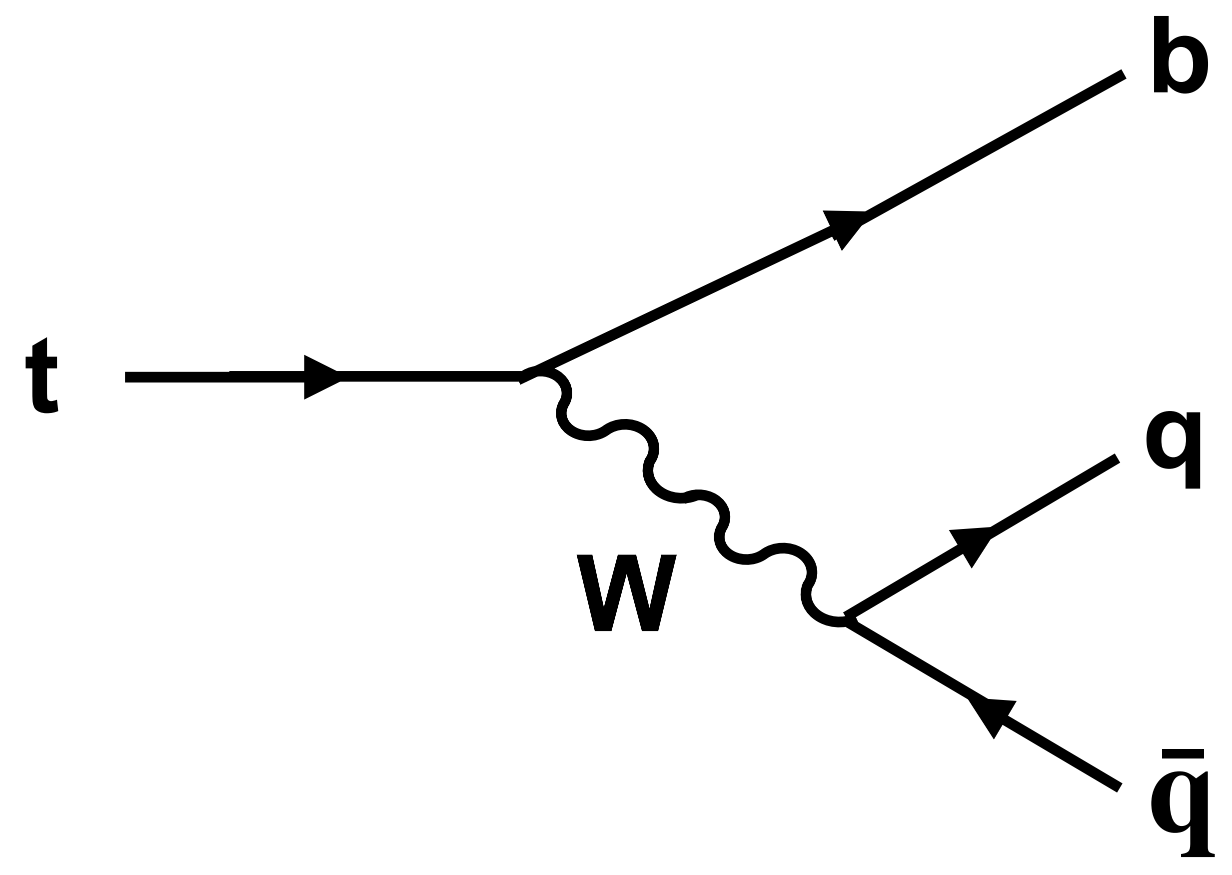


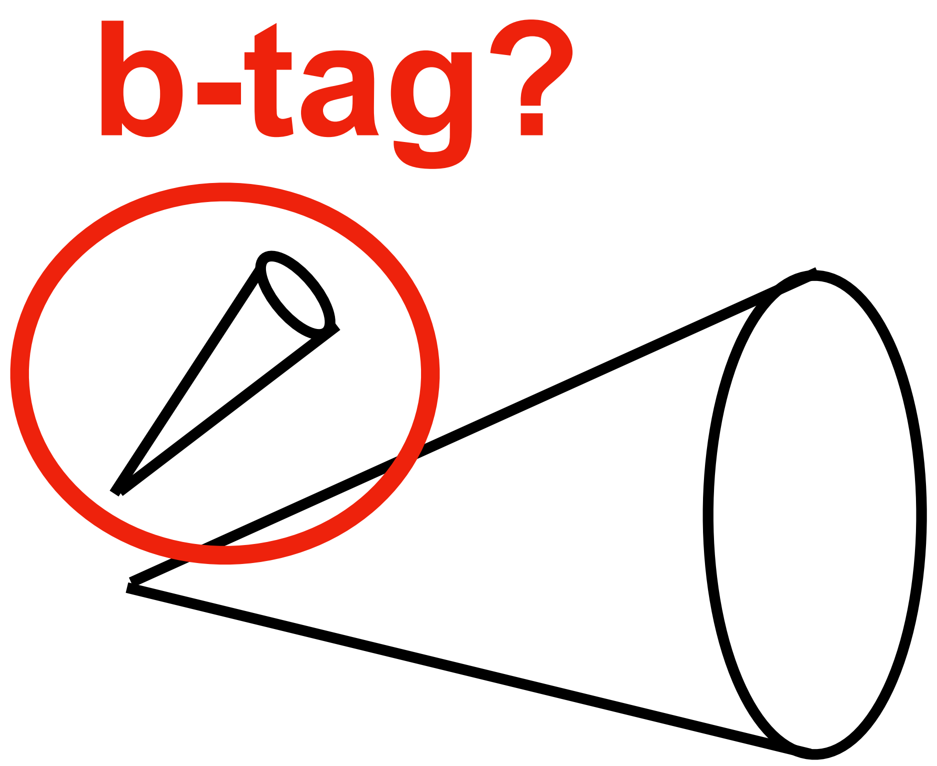
Empty veto region

$$D_{GN2X}^{Hcc} = log\left(\frac{p_{Hcc}}{f_{Hbb}p_{Hbb} + f_{top}p_{top} + (1 - f_{Hbb} - f_{top})p_{QCD}}\right) > \quad threshold$$

Top control region

Top-quark decay



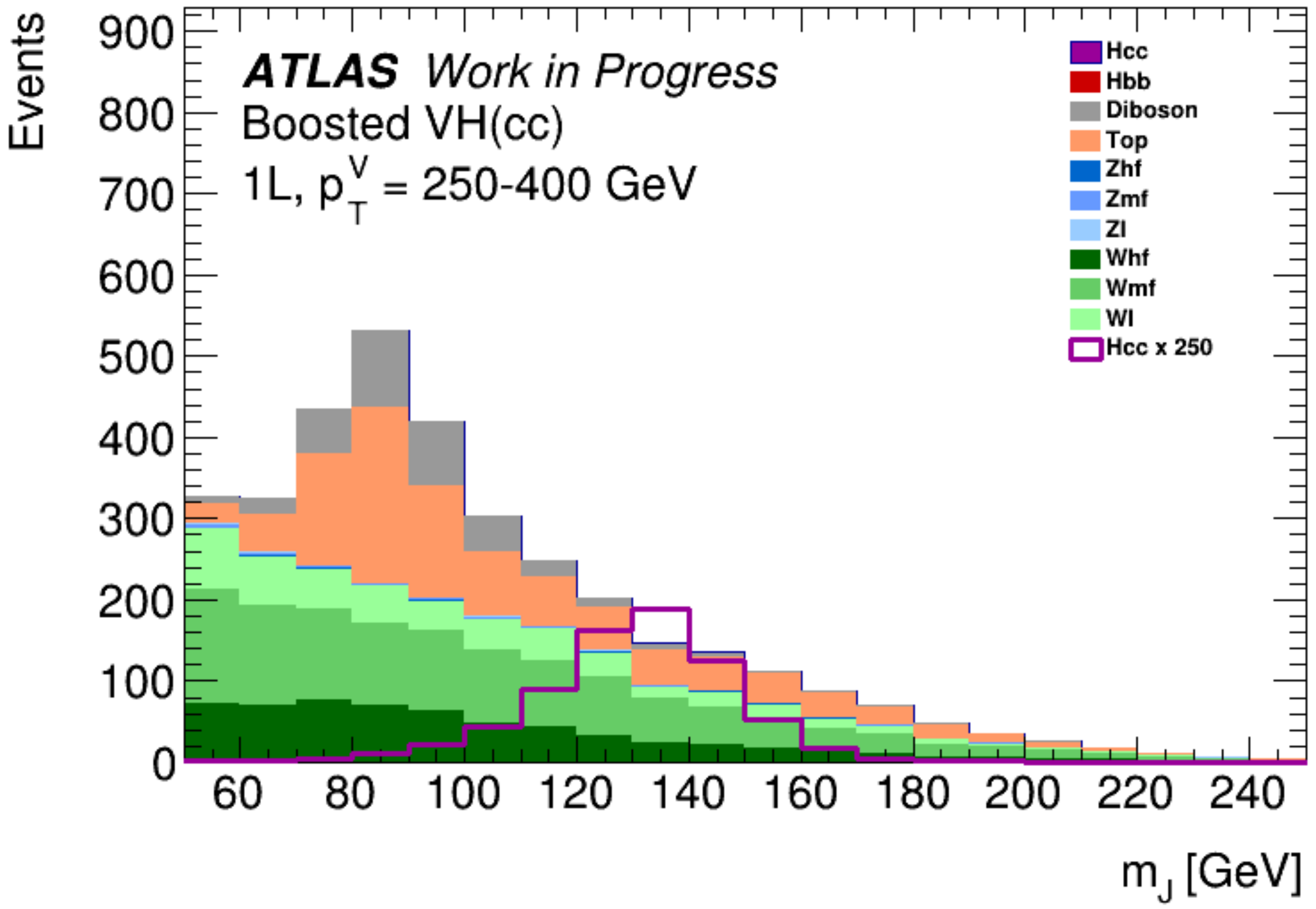


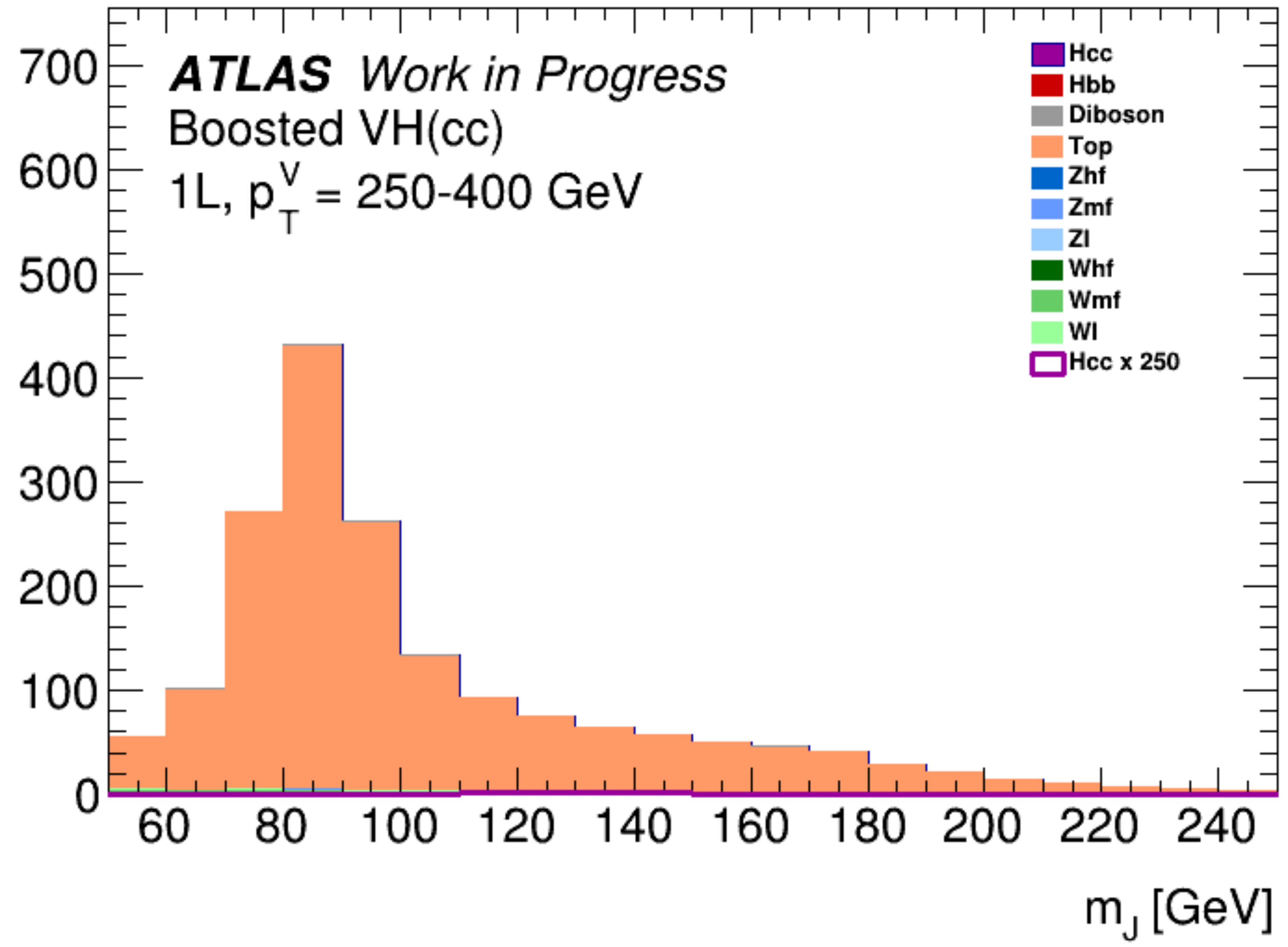
No

Yes

Signal Region

Control Region

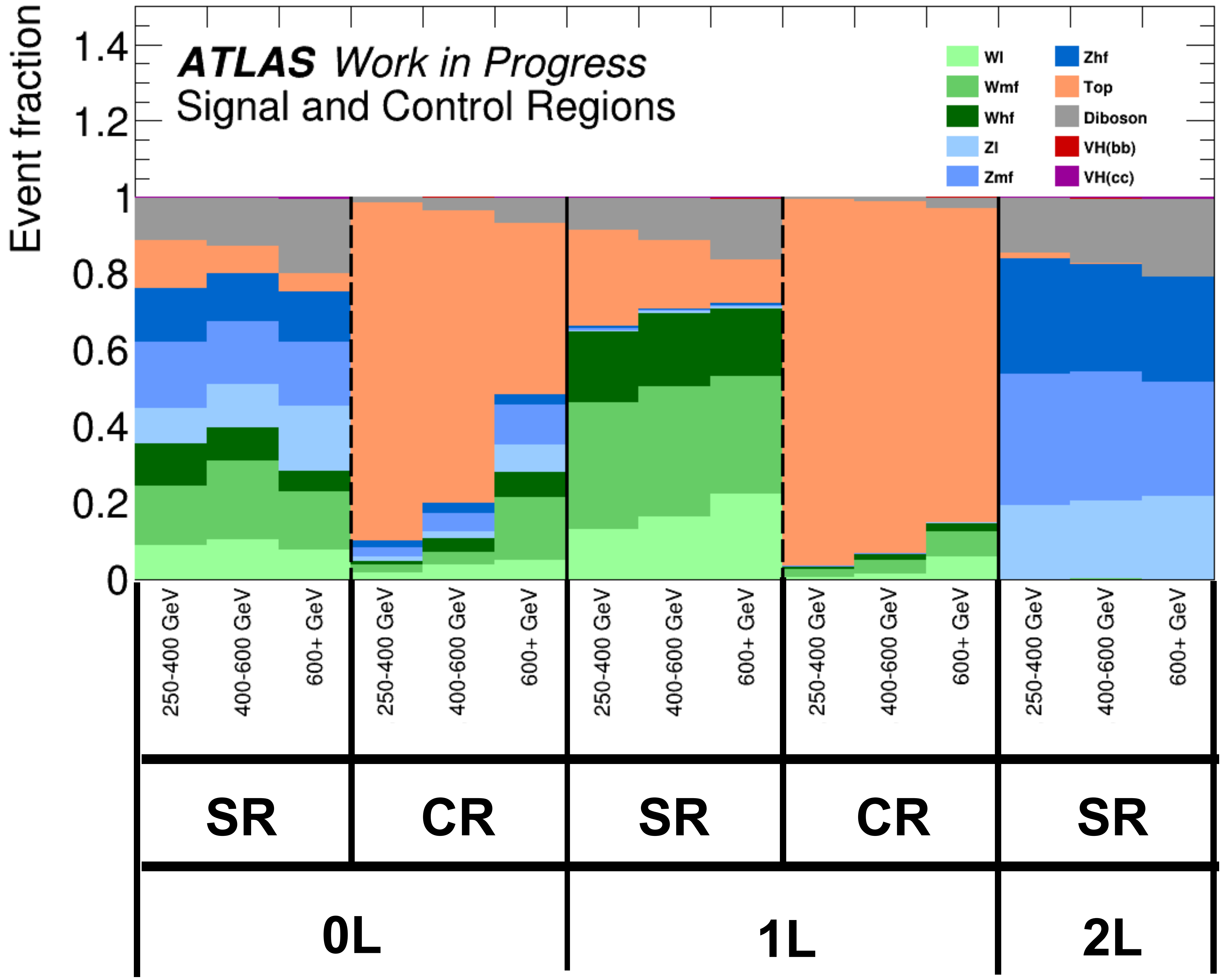




- Constrain top backgrounc
- Create top-rich region
  - Look for add. *b*-jets

Signal and control regions

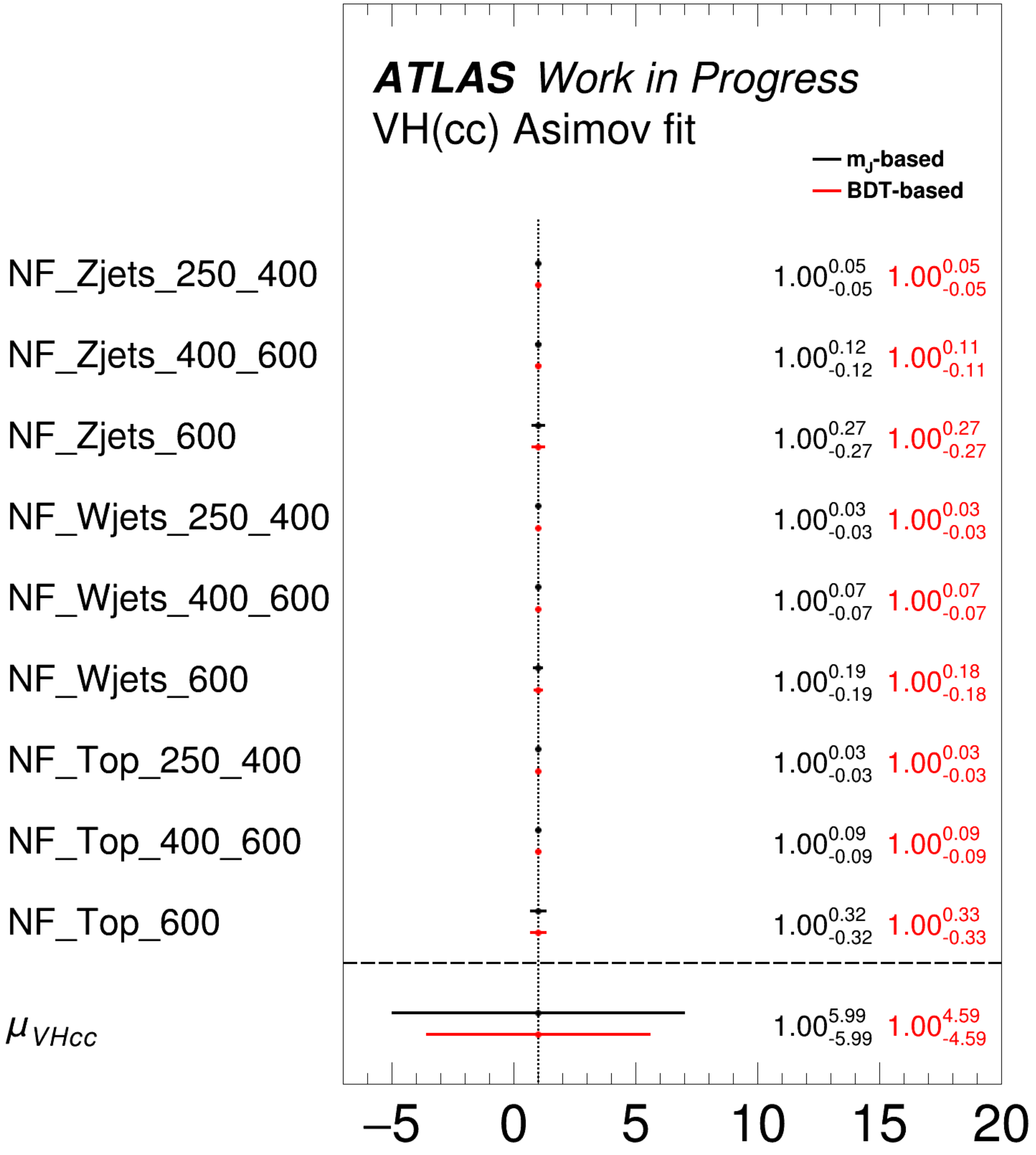
- 9 signal regions
  - Added as $\mathrm{m_J}$ distributions
  - Added as MVA-score distributions
- 6 control regions
  - Top background control
  - Shows high top purity (orange)



Nikhef

Asimov fit: mJ vs BDT

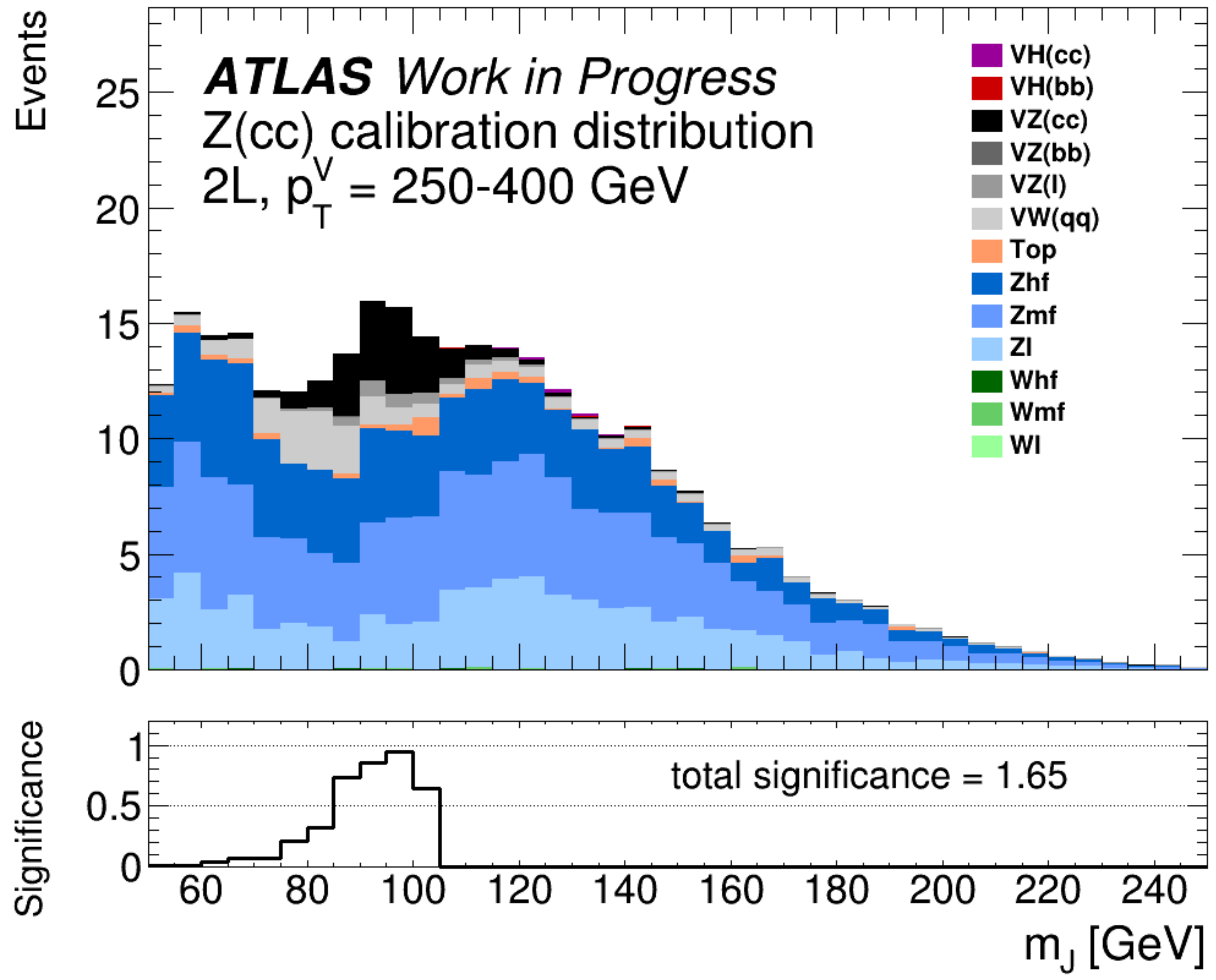
- Performed $VH(\rightarrow c\bar{c})$ Asimov fit
  - Fit to simulation samples
  - Enables combination of analysis regions
  - No systematic uncertainties included
- $VH(\rightarrow c\bar{c})$ signal strength
  - m_J-based: 1.00 ± 5.99
  - BDT-based: 1.00 ± 4.59
    - 23% uncertainty reduction



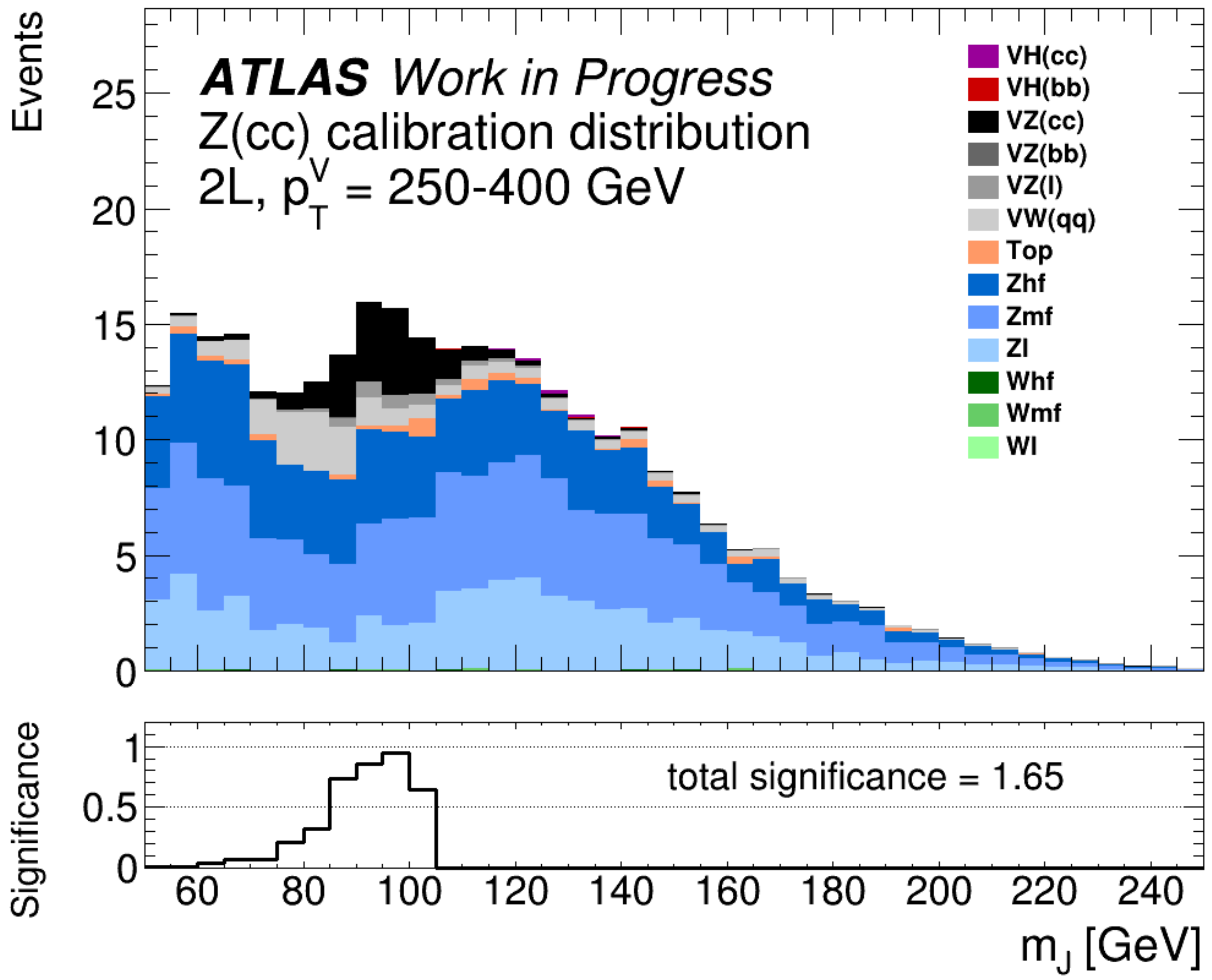
Nikhef

In-situ GN2X Hcc calibration

- No calibration of *Hcc*-tagger yet
  - Correct for different tagging performance on simulations and collision data
- In-situ calibration
  - Use calibration region in our own sample
    - m_J < 105 GeV to not affect $VH(\to c\bar{c})$
  - Fit $Z(\to c\bar{c})$
    - Derive flavour-tag scale factor
- Explored feasibility across lepton channels
  - 2L channel optimal

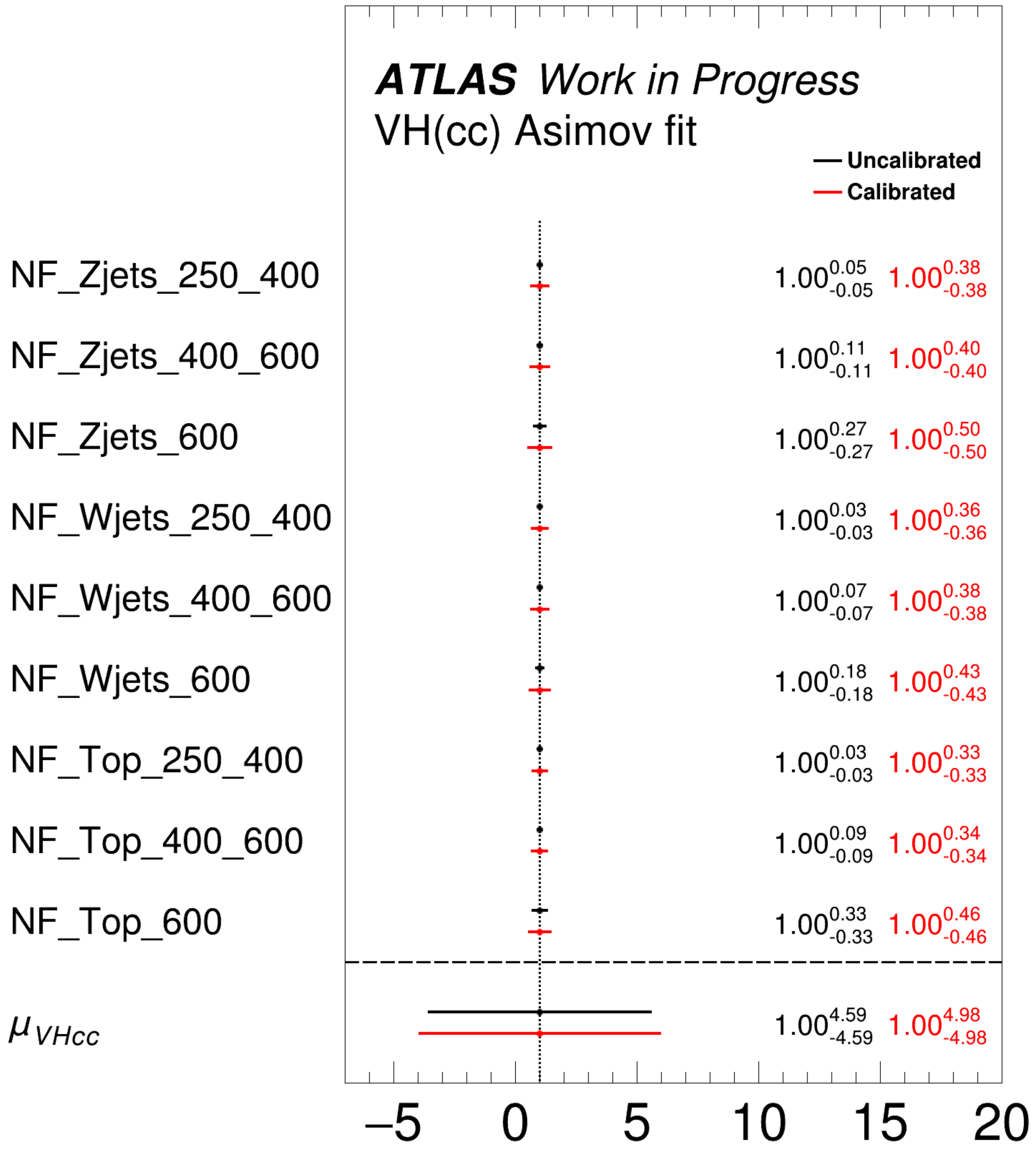


Calibration impact on *VH*($\rightarrow c\bar{c}$) fit



Fit Z($\rightarrow c\bar{c}$)

SF: 1.00±0.69



- Uncertainty from flavour-tagging: 69%
- *VH*($\rightarrow c\bar{c}$) uncertainty increase: 8.5%

Comparison to resolved $VH(\rightarrow c\bar{c})$ GN2

- High-p_T region (>400 GeV)
  - Significant improvements across lepton channels
- Low-p_T region (250-400 GeV)
  - Comparable significances
  - Need more investigation in this transition regime!

m_J-based

Lepton channel	p_T^V [GeV]	Resolved GN2	Boosted GN2X	Difference
0L	**250-400**	0.104	0.103	-1.2%
0L	**> 400**	0.066	0.098	+49%
1L	**250-400**	0.099	0.080	-19%
1L	**> 400**	0.076	0.094	+23%
2L	**250-400**	0.064	0.066	+3.9%
2L	**> 400**	0.035	0.055	+56%

Comparison to boosted $VH(\to c\bar{c})$ DL1r

- Improvements in both regions
  - ~100%
- Demonstrates GN2X advantages in Lorentz-boosted regime
  - Algorithm: DNN → Transformer
  - Input: More low-level information

BDT-based

Lepton channel	p_T^V [GeV]	Boosted DL1r	Boosted GN2X	Difference
0L	**250-400**	0.070	0.136	+94%
0L	**> 400**	0.059	0.122	+107%

Sensitivity improvements at ϵ_{Hcc} = 30%

Backup

- (Pseudo-)continuous scheme leads to significant sensitivity improvements
  - Double-WP: +11%
  - Continuous: +19%
- However: calibration difficult
  - Statistics too low for traditional $Z(\rightarrow c\bar{c})$ calibration
  - Requires new methods

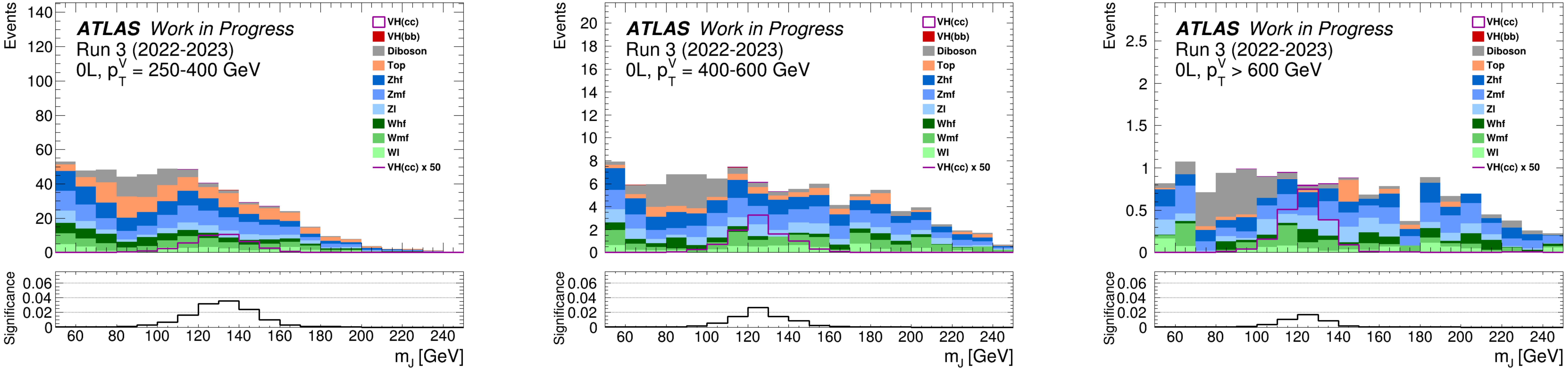
Lepton channel	pTV [GeV]	Single-WP	Double-WP	Continuous
0L	250-400	0.136	+11%	+20%
0L	400-600	0.103	+11%	+22%
0L	>600	0.066	+6.7%	+14%
1L	250-400	0.118	+13%	+21%
1L	400-600	0.105	+12%	+19%
1L	>600	0.070	+7.4%	+10%
2L	250-400	0.080	+7.9%	+16%
2L	400-600	0.058	+8.1%	+16%
2L	>600	0.035	+0.59%	+11%

Advanced MVA approaches

- This study: BDT
- Exploring advanced MVA approaches
  - DNN: +1%
  - Transformer: +5%
- Developments ongoing
- Need fit with full systematics to choose optimal MVA

Run-3 data

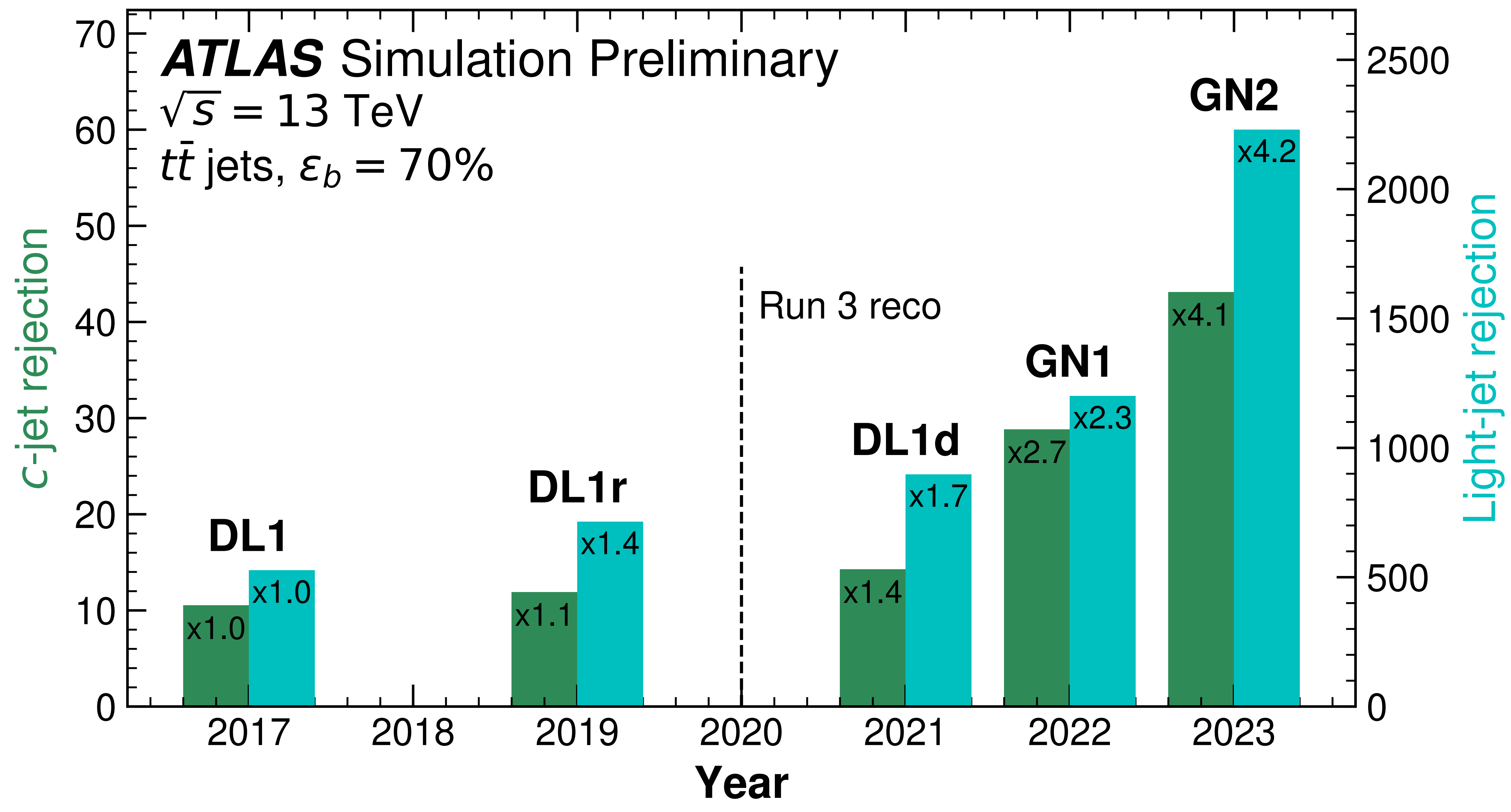
Backup

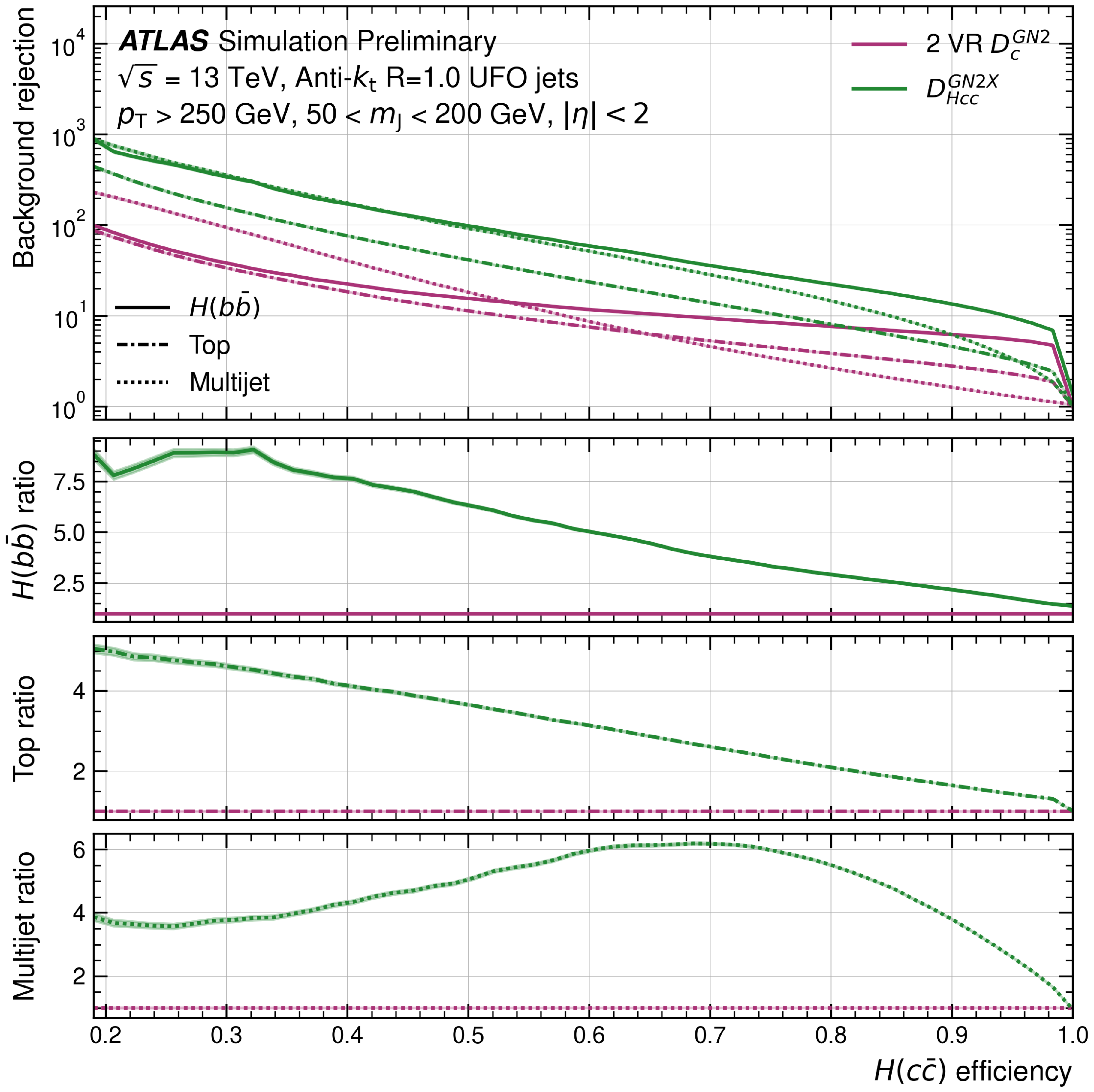


- Study focused on Run-2 (140 fb^{-1})
- Including partial Run-3 data (52 fb^{-1})
  - Increase statistics → increase sensitivity
  - Increase from 2022-2023: +20%

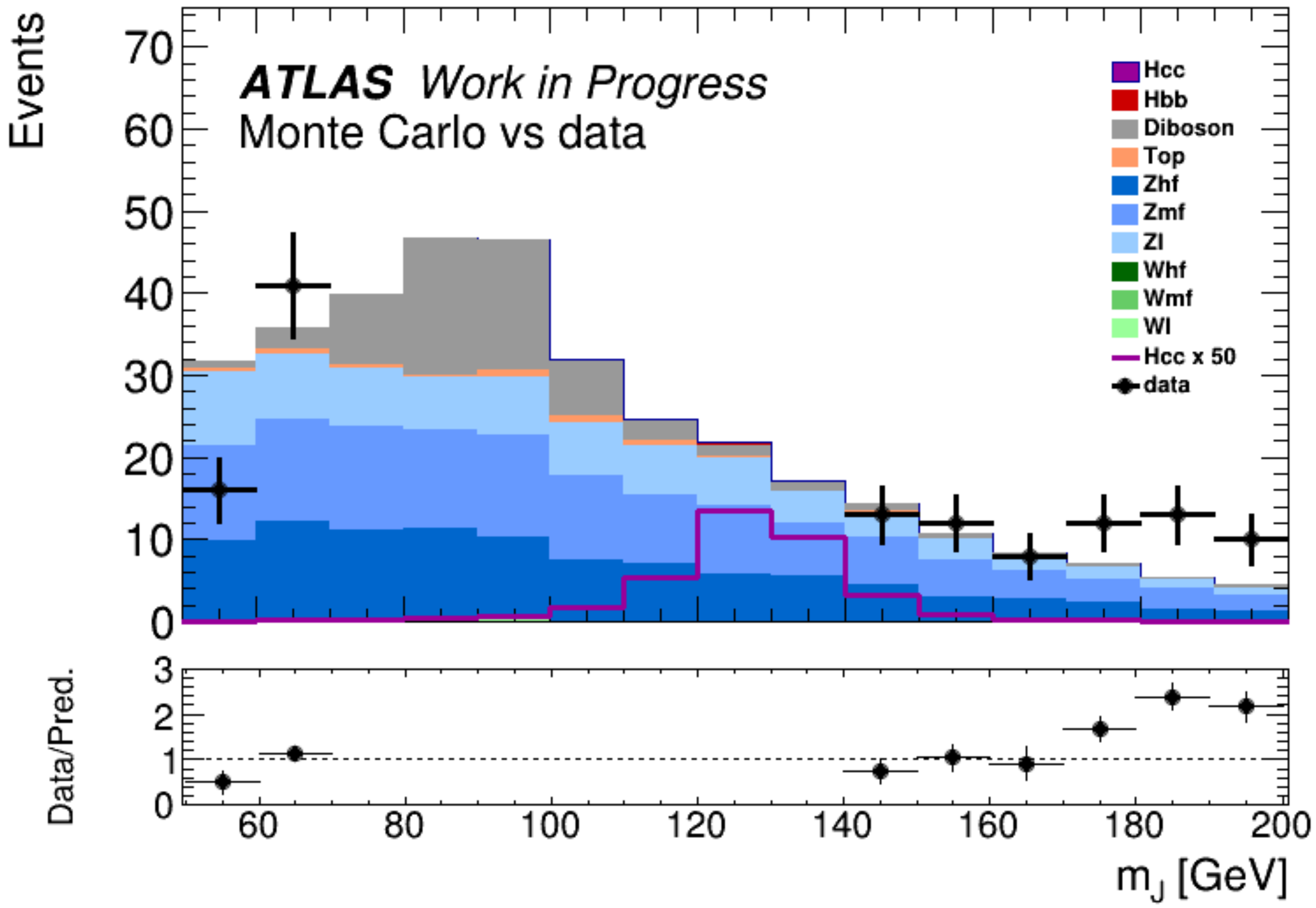
GN2X flavour-tagging advancements

Backup



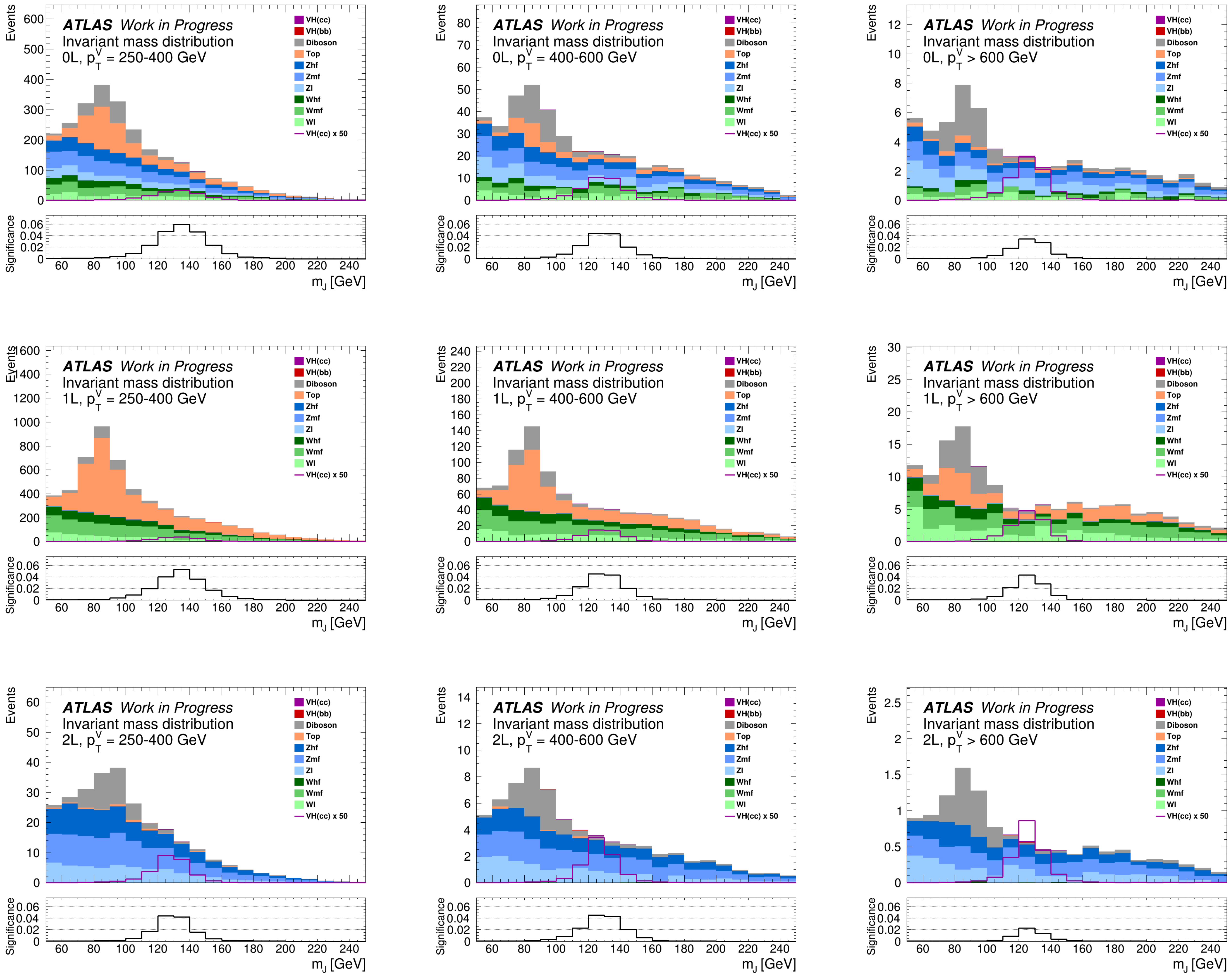


Comparison MC simulations with data



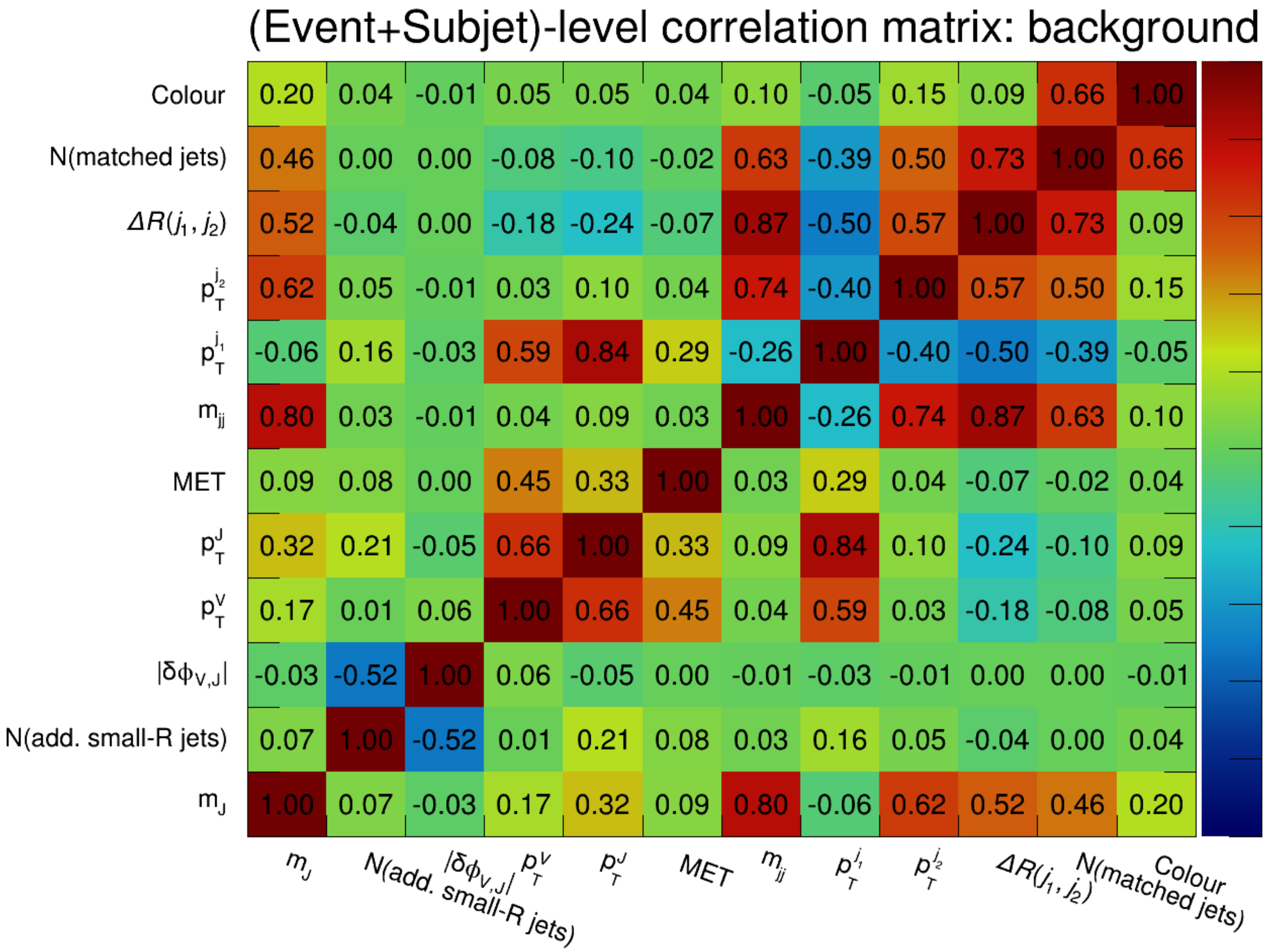
Backup

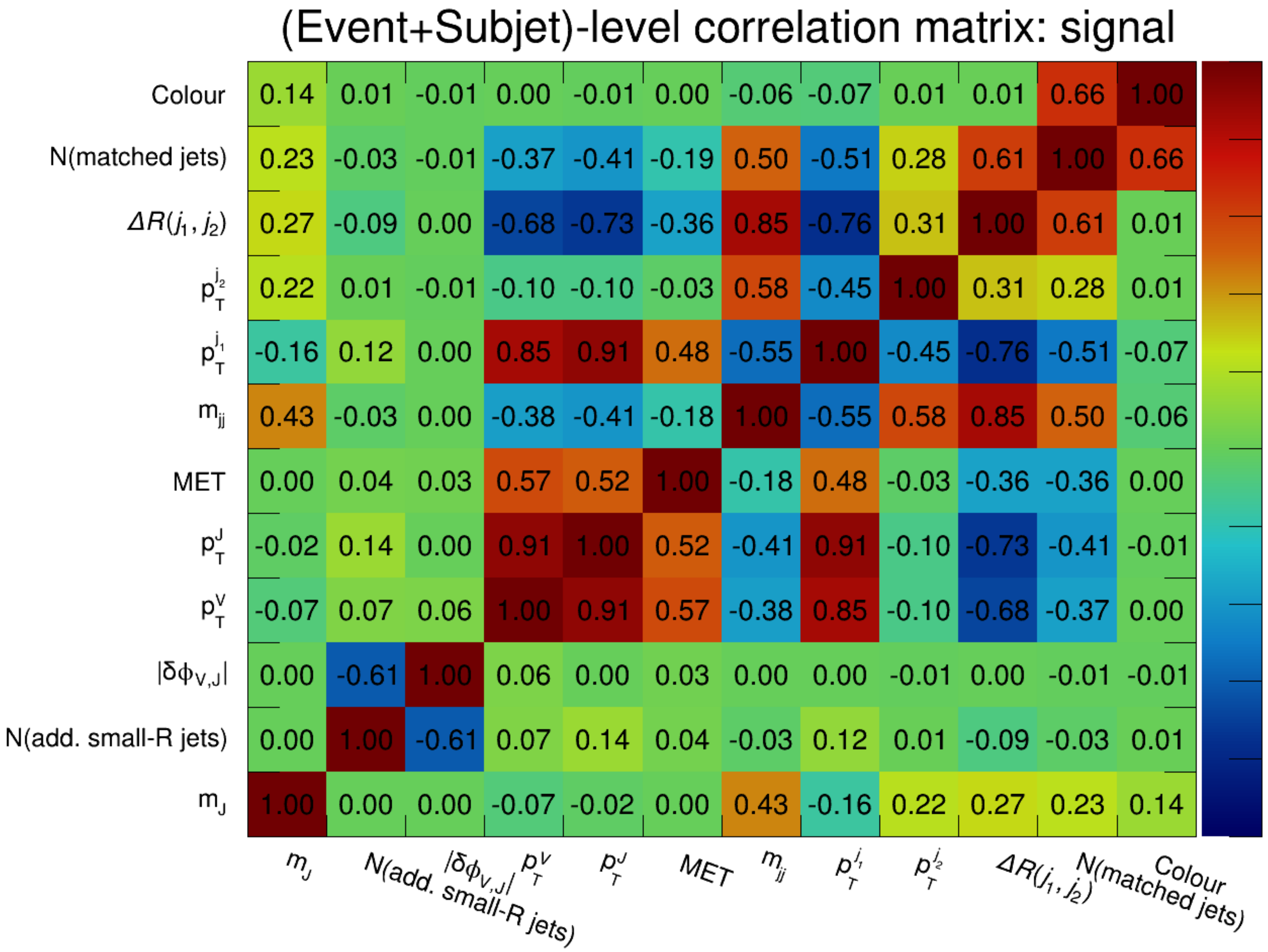
All invariant mass distributions



Correlation matrices BDT in 1L ((Event+subjet)-level)

Backup





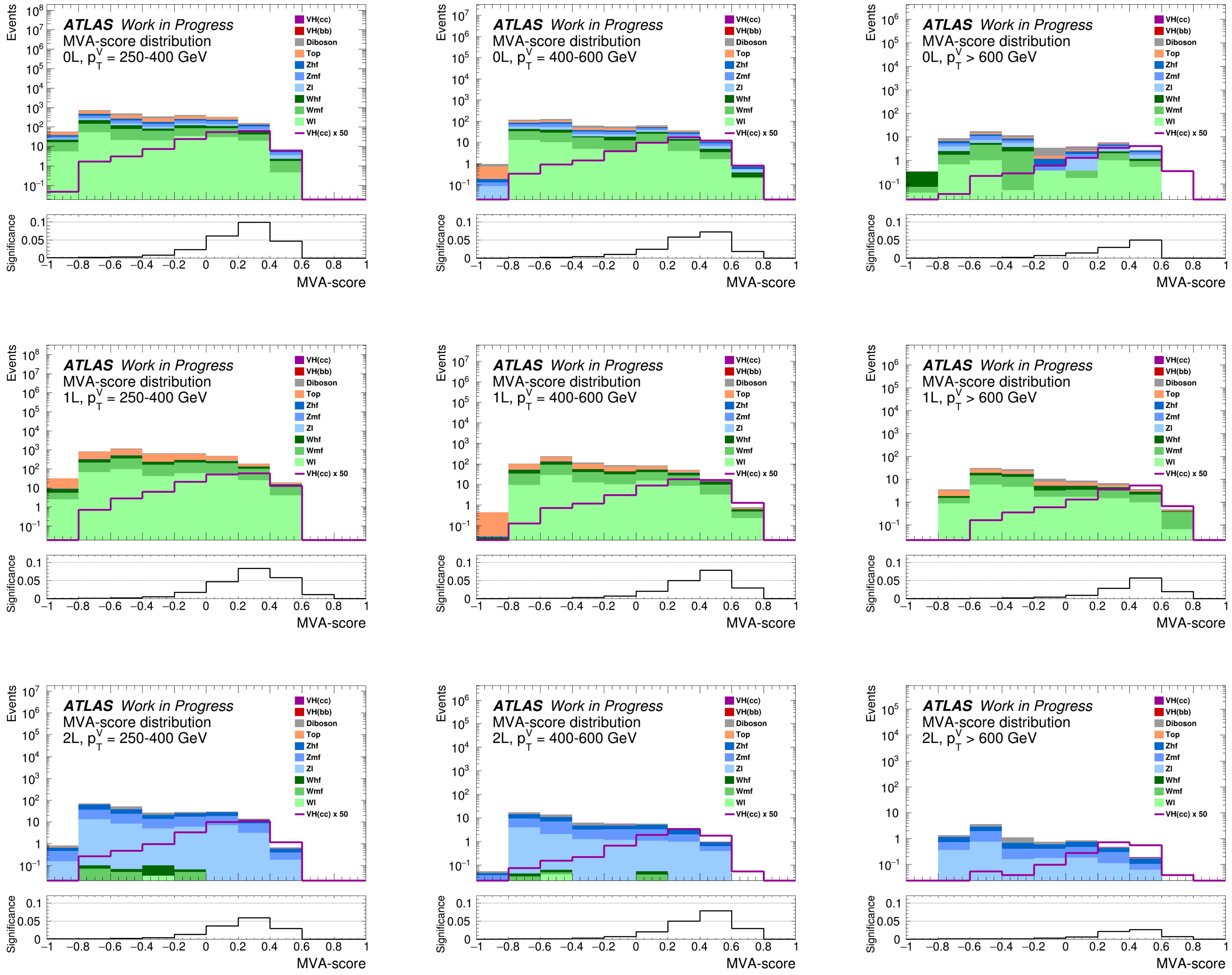
Colour

- Colour (= colour ring)
  - Exploit difference in color-flow between gluons splittings and decays from QCD singlet states

$$Colour = \frac{\theta^2_{j_1 j_3} + \theta^2_{j_2 j_3}}{\theta^2_{j_1 j_2}}$$

All BDT-score distributions

Backup



Nikhef