Search for $t\bar{t}H$ with $H \rightarrow b\bar{b}$ and the Development of Flavour Tagging for Run 3

Alexander Froch

Supervised by Manuel Guth, Andrea Knue

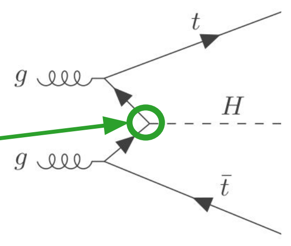
29.09.2022 - GRK Workshop

Motivation

- Why is the $t\bar{t}H$ process important?
  - Directly measure the Top-Higgs Yukawa coupling

    $$y_{\text{Fermion}} = \sqrt{2}\frac{M_{\text{Fermion}}}{246\text{ GeV}}$$

  - Top has strongest coupling to Higgs: $y_t \approx 1$



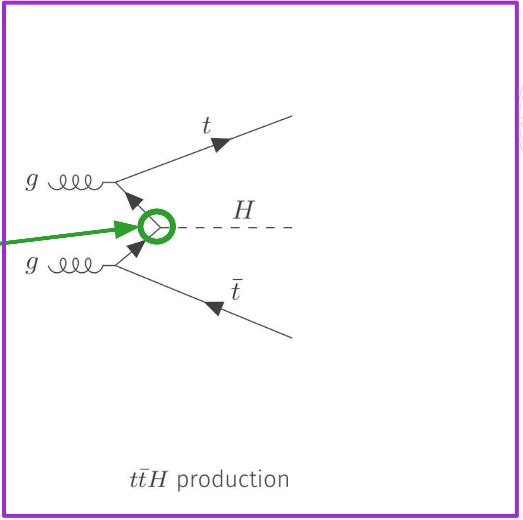
$t\bar{t}H$ production

Motivation

- Why is the $t\bar{t}H$ process important?
    - Directly measure the Top-Higgs Yukawa coupling

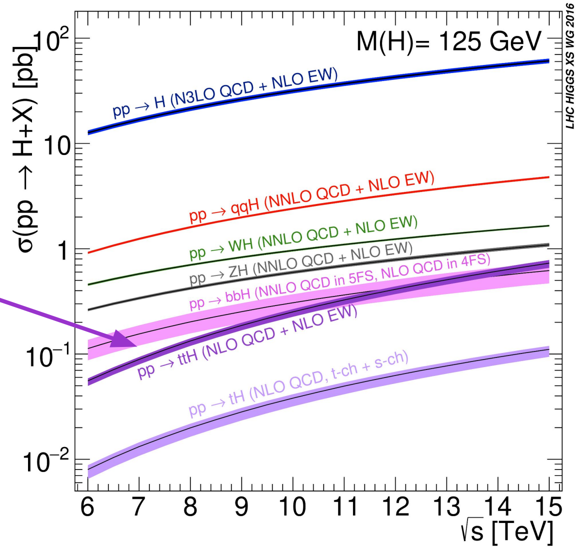
      $$y_{\text{Fermion}} = \sqrt{2}\frac{M_{\text{Fermion}}}{246\ \text{GeV}}$$

    - Top has strongest coupling to Higgs: $y_t \approx 1$



$t\bar{t}H$ production

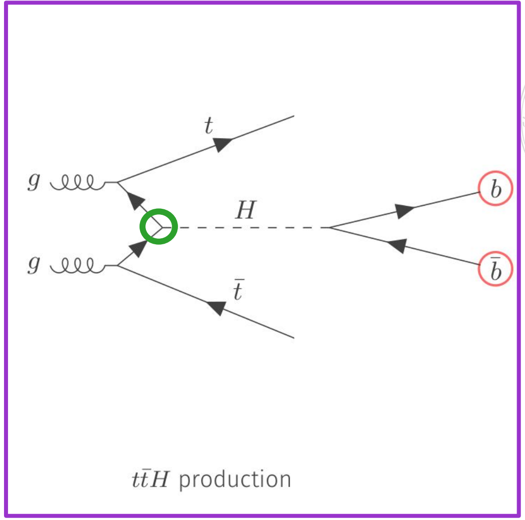
- Still: $t\bar{t}H$ process is rare!



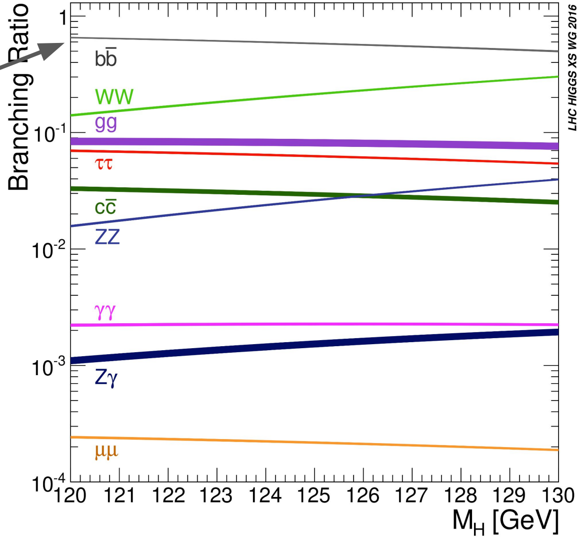
Motivation

- Why is the $t\bar{t}H$ process important?
  - Directly measure the Top-Higgs Yukawa coupling

    $$y_{\text{Fermion}} = \sqrt{2}\frac{M_{\text{Fermion}}}{246\text{ GeV}}$$
  - Top has strongest coupling to Higgs: $y_t \approx 1$



- Still: $t\bar{t}H$ process is rare!
  - Use the most dominant decay channel with $H \rightarrow b\bar{b}$
  - Here: Single lepton channel

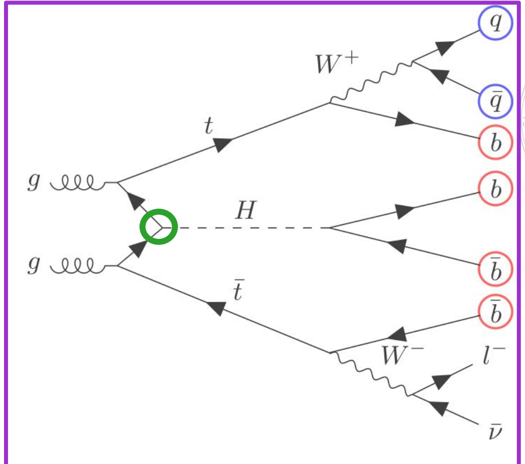


Motivation

- Why is the $t\bar{t}H$ process important?
  - Directly measure the Top-Higgs Yukawa coupling

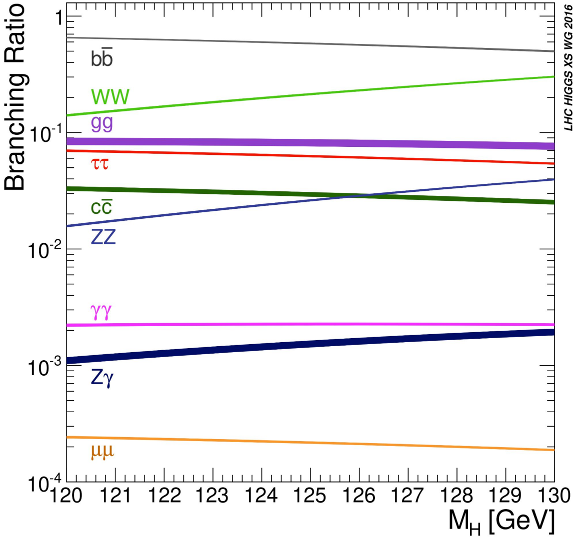
    $$y_{\text{Fermion}} = \sqrt{2}\frac{M_{\text{Fermion}}}{246\ \text{GeV}}$$

  - Top has strongest coupling to Higgs: $y_t \approx 1$



$t\bar{t}H$ production and decay

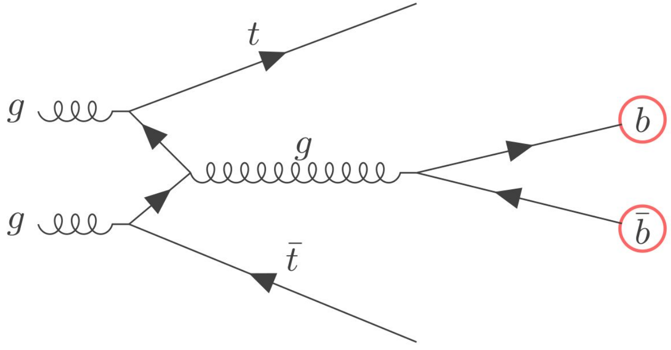
- Still: $t\bar{t}H$ process is rare!
  - Use the most dominant decay channel with $H \rightarrow b\bar{b}$
  - Here: Single lepton channel



- Single lepton channel final state:
  - 6 jets
  - 4 *b*-tagged jets, 2 non-*b*-tagged jets
  - 1 lepton (electron/muon)
  - 1 neutrino

Irreducible Background - $t\bar{t} + b\bar{b}$

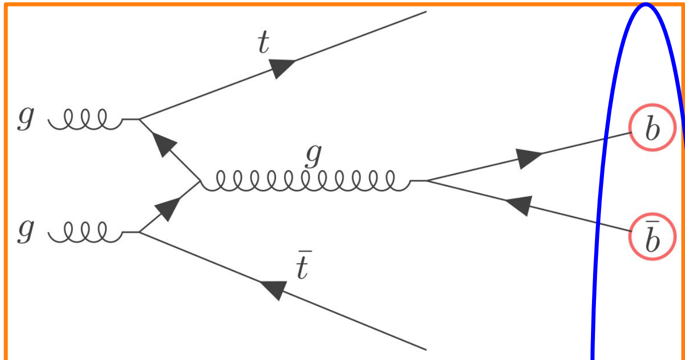
- Most dominant background → $t\bar{t} + b\bar{b}$



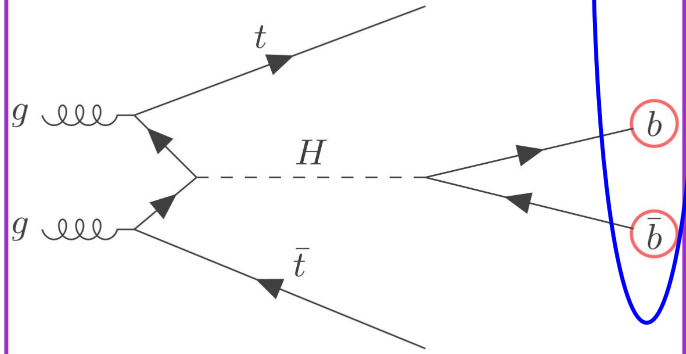
$t\bar{t} + b\bar{b}$ production and decay

Irreducible Background - $t\bar{t} + b\bar{b}$

- Most dominant background → $t\bar{t} + b\bar{b}$
- Irreducible background for the $t\bar{t}H$ process
  - Same final state particles
  - Difference in kinematics



$t\bar{t} + b\bar{b}$ production and decay



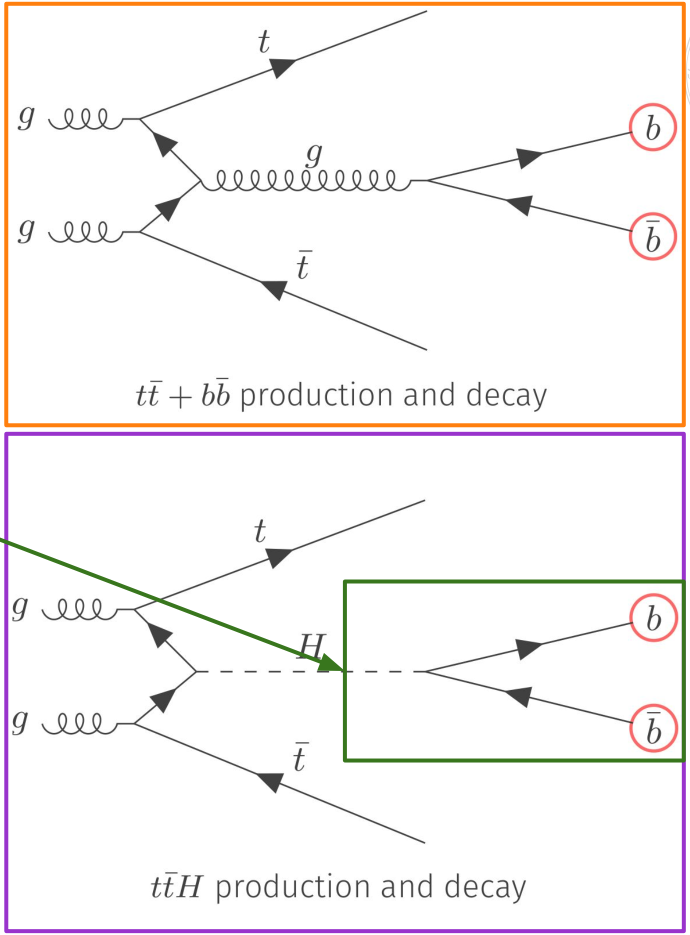
$t\bar{t}H$ production and decay

Irreducible Background - $t\bar{t} + b\bar{b}$

- Most dominant background → $t\bar{t} + b\bar{b}$

- Irreducible background for the $t\bar{t}H$ process
  - Same final state particles
  - Difference in kinematics

- Can be separated by reconstruction of the Higgs
  - Problem: Not all jets from Higgs and top quarks survive jet selection
  - Also: Jet/parton assignment difficult due to large jet multiplicity
  - Good *b*-tagging important for jet/parton assignment



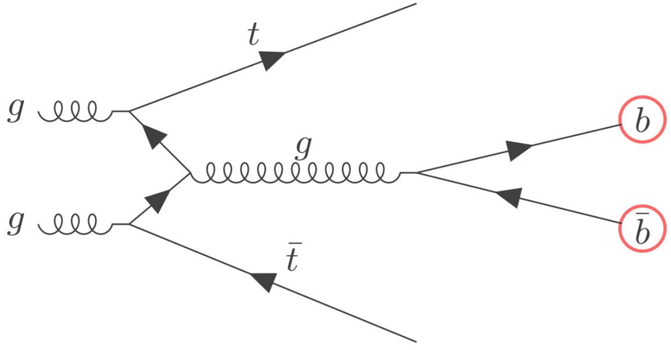
Irreducible Background - $t\bar{t} + b\bar{b}$

- Most dominant background → $t\bar{t} + b\bar{b}$

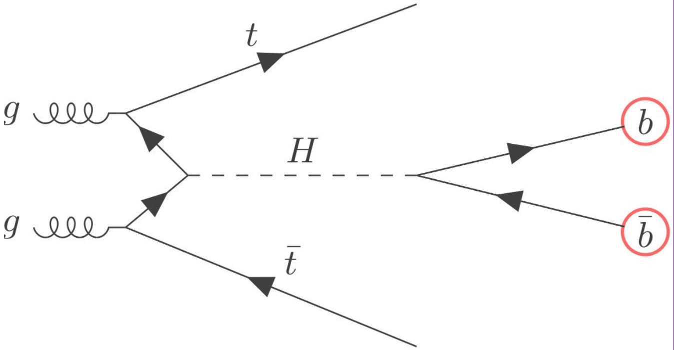
- Irreducible background for the $t\bar{t}H$ process
    - Same final state particles
    - Difference in kinematics

- Can be separated by reconstruction of the Higgs
    - Problem: Not all jets from Higgs and top quarks survive jet selection
    - Also: Jet/parton assignment difficult due to large jet multiplicity
    - Good *b*-tagging important for jet/parton assignment

- Also: $t\bar{t} + b\bar{b}$ → Large theory uncertainties



$t\bar{t} + b\bar{b}$ production and decay

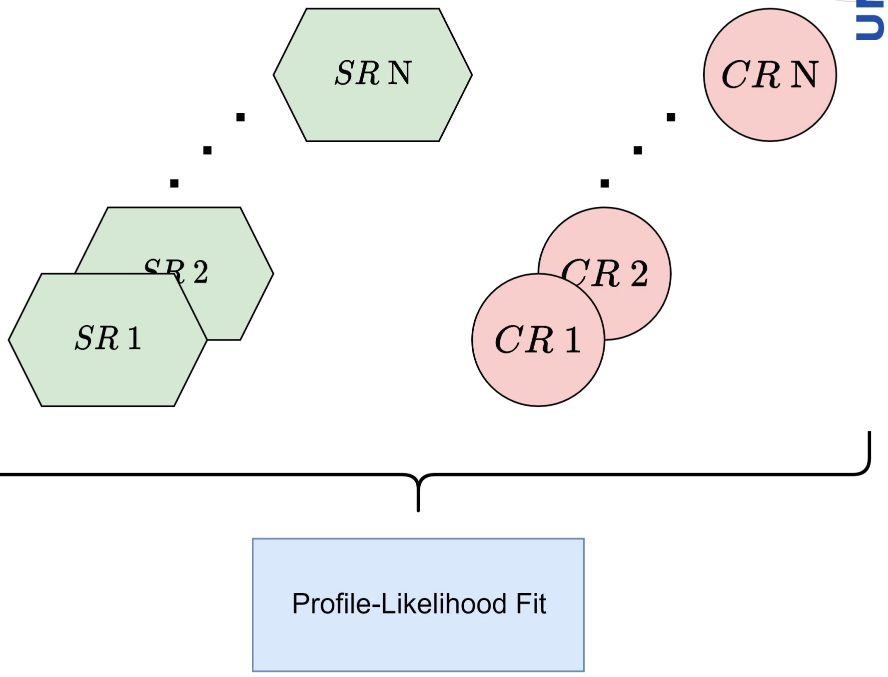


$t\bar{t}H$ production and decay

Results from Last Round Analysis

- Signal and control region defined on jet and *b*-tag multiplicity
  - Signal → nJets ≥ 6, nBTags@85 ≥ 4
  - Control → nJets = 5, nBTags@85 ≥ 4

- Each region is than used in a Profile Likelihood fit

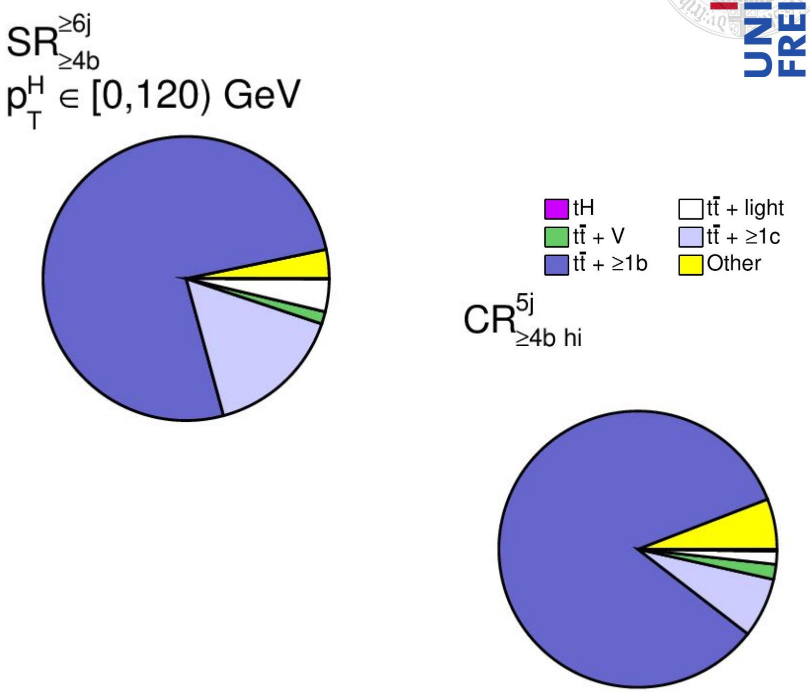


Results from Last Round Analysis

- Signal and control region defined on jet and *b*-tag multiplicity
    - Signal → nJets ≥ 6, nBTags@85 ≥ 4
    - Control → nJets = 5, nBTags@85 ≥ 4

- Each region is than used in a Profile Likelihood fit

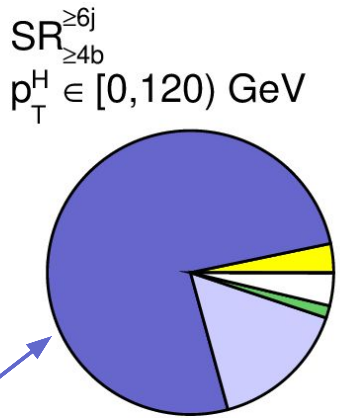
- Looking at the background-only compositions

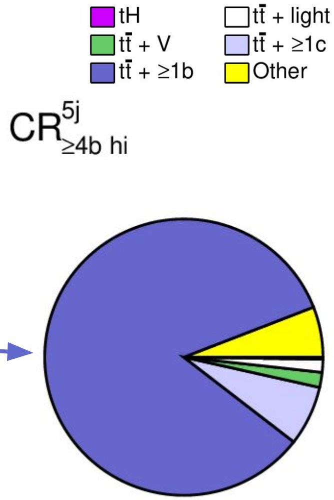


HIGG-2020-23

Results from Last Round Analysis

- Signal and control region defined on jet and *b*-tag multiplicity
  - Signal → nJets ≥ 6, nBTags@85 ≥ 4
  - Control → nJets = 5, nBTags@85 ≥ 4
- Each region is than used in a Profile Likelihood fit
- Looking at the background-only compositions
- Signal- and control regions dominated by $t\bar{t} + b\bar{b}$
- No clean control regions
  - Used to constrain systematic uncertainties
  - Is there a way to define cleaner signal- and control regions?





Background composition of the different signal- and control regions
From HIGG-2020-23

Results from Last Round Analysis

- Results of the last analysis are shown here

- Public paper results with full Run 2 data

$$\boxed{\mu_{t\bar{t}H} = \frac{\sigma^{t\bar{t}H}}{\sigma^{t\bar{t}H}_{\mathrm{SM}}} = 0.35^{+0.36}_{-0.34}\left({}^{+0.20}_{-0.20}\,\mathrm{stat.}\ {}^{+0.30}_{-0.28}\,\mathrm{sys.}\right)}$$

HIGG-2020-23

Results from Last Round Analysis

- Results of the last analysis are shown here
- Public paper results with full Run 2 data
- Was dominated by systematic uncertainties

$$\mu_{t\bar{t}H} = \frac{\sigma^{t\bar{t}H}}{\sigma^{t\bar{t}H}_{\mathrm{SM}}} = 0.35^{+0.36}_{-0.34} \left(^{+0.20}_{-0.20}\ \mathrm{stat.}\ \boxed{^{+0.30}_{-0.28}\ \mathrm{sys.}}\right)$$

HIGG-2020-23

Results from Last Round Analysis

- Results of the last analysis are shown here

$$\mu_{t\bar{t}H} = \frac{\sigma^{t\bar{t}H}}{\sigma^{t\bar{t}H}_{\mathrm{SM}}} = 0.35^{+0.36}_{-0.34}\,\left(^{+0.20}_{-0.20}\,\mathrm{stat.}\ \boxed{^{+0.30}_{-0.28}\,\mathrm{sys.}}\right)$$

HIGG-2020-23

- Public paper results with full Run 2 data

- Was dominated by systematic uncertainties

- Largest uncertainty: $t\bar{t}$ + bb modelling

- Due to systematic limitations: Redo the analysis with multiple improvements to reduce systematic uncertainty

Total statistical/systematic uncertainties with the largest systematic contributions (above 0.1)

Uncertainty source	$\Delta\mu$	
$t\bar{t}H$ modelling	$+0.13$	-0.05
$t\bar{t}+b\bar{b}$ NLO matching	$+0.21$	-0.20
$t\bar{t}+b\bar{b}$ fractions	$+0.12$	-0.12
$t\bar{t}+b\bar{b}$ FSR	$+0.10$	-0.11
Total systematic uncertainty	$+0.30$	-0.28
Total statistical uncertainty	$+0.20$	-0.20
Total uncertainty	$+0.36$	-0.34

Based on HIGG-2020-23

$N_{e/\mu} = 1$

Object- and Region Definitions

- Lepton definitions:
  - Single-lepton: One lepton with p_T >= 27 GeV

$$N_{e/\mu} = 1 \quad\downarrow\quad N_{\text{jets}} \geq 5$$

Object- and Region Definitions

- Lepton definitions:
  - Single-lepton: One lepton with p_{T} >= 27 GeV

- Jet definition:
  - Using the anti-k_t jet cluster algorithm with ΔR = 0.4
  - Jet p_{T} > 25 GeV, Jet $|\eta|$ < 2.5

Object- and Region Definitions

$N_{e/\mu} = 1$

↓

$N_{\text{jets}} \geq 5$

↓

$N_{b\text{-jets}}^{@70\%} \geq 3$

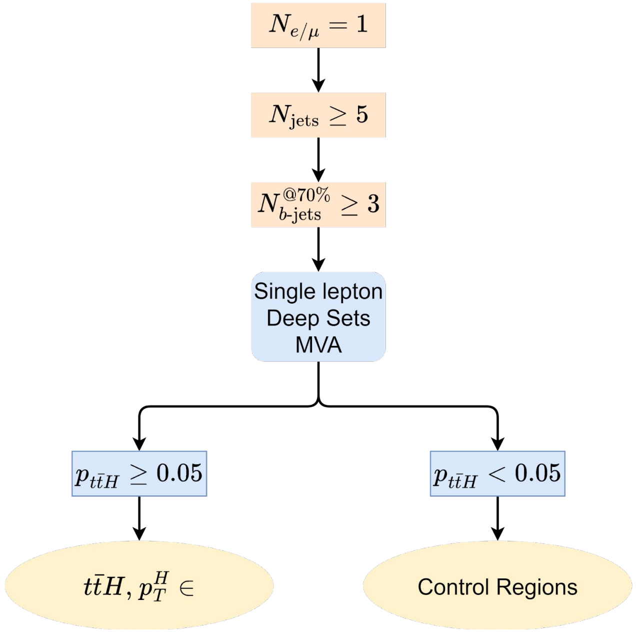
- Lepton definitions:
    - Single-lepton: One lepton with p_T >= 27 GeV

- Jet definition:
    - Using the anti-k_t jet cluster algorithm with $\Delta R = 0.4$
    - Jet p_T > 25 GeV, Jet $|\eta| < 2.5$

- *b*-tagged definition:
    - Jets passing the corresponding DL1r working point (WP) cuts

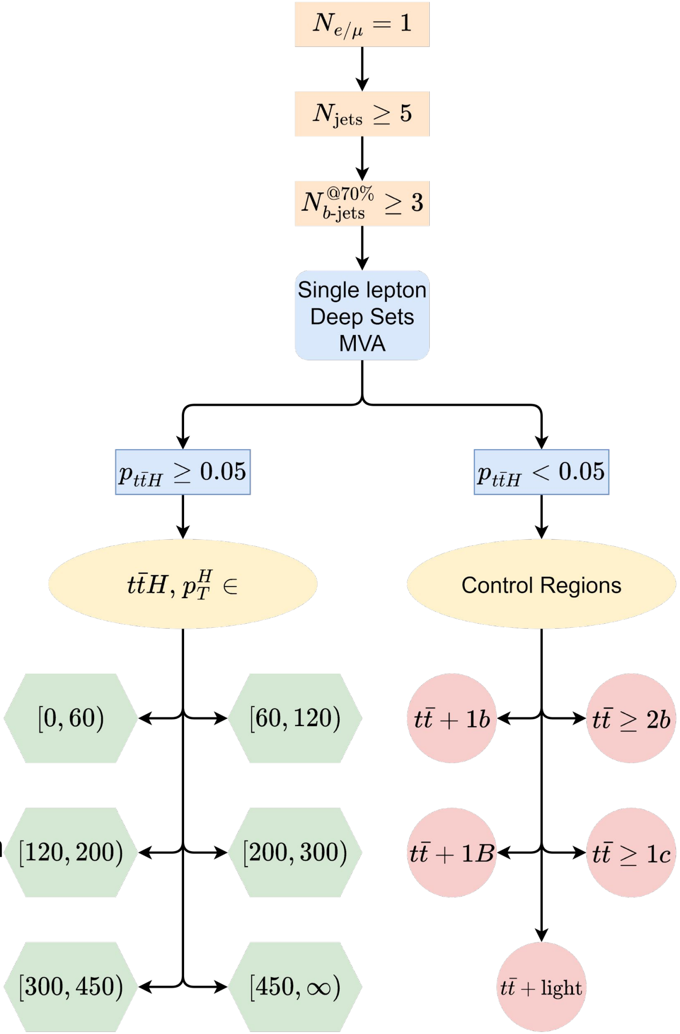
Object- and Region Definitions

- Lepton definitions:
  - Single-lepton: One lepton with p_T >= 27 GeV
- Jet definition:
  - Using the anti-k_t jet cluster algorithm with ΔR = 0.4
  - Jet p_T > 25 GeV, Jet $|\eta|$ < 2.5
- *b*-tagged definition:
  - Jets passing the corresponding DL1r working point (WP) cuts
- Using Deep-Sets based deep neural networks (DNNs)



Object- and Region Definitions

- Lepton definitions:
  - Single-lepton: One lepton with p_T >= 27 GeV
- Jet definition:
  - Using the anti-k_t jet cluster algorithm with ΔR = 0.4
  - Jet p_T > 25 GeV, Jet $|\eta|$ < 2.5
- *b*-tagged definition:
  - Jets passing the corresponding DL1r working point (WP) cuts
- Using Deep-Sets based deep neural networks (DNNs)
  - Multiclass classifier
  - Outputs probability for event belonging to one region
  - Developed by students from DESY
  - No need for building jet-parton permutations →Deep-Sets are permutation-invariant!



Fake Estimation

- Very good reconstruction algorithms for objects in ATLAS
  - But: Nobody's perfect!
    Soft leptons or non-leptonic particles can pass the requirements
  - Most backgrounds are modelled with Monte Carlo (MC)
  - Lepton fakes contribution → Data-driven with Matrix method

Fake Estimation

- Very good reconstruction algorithms for objects in ATLAS
  - But: Nobody's perfect!
    Soft leptons or non-leptonic particles can pass the requirements
  - Most backgrounds are modelled with Monte Carlo (MC)
  - Lepton fakes contribution → Data-driven with Matrix method

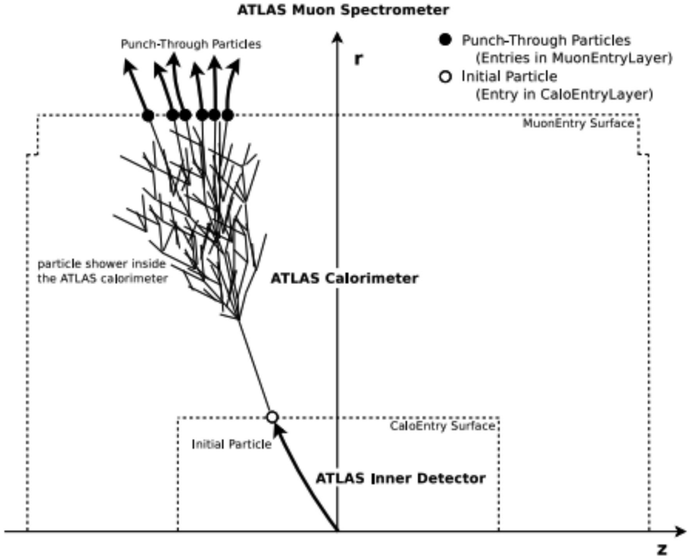
- Sources for fake electrons
  - Light- or gluon jets
  - Deposit most energy in the electromagnetic calorimeter (ECAL)
  - γ conversion in the ECAL

Fake Estimation

- Very good reconstruction algorithms for objects in ATLAS
  - But: Nobody's perfect!
    Soft leptons or non-leptonic particles can pass the requirements
  - Most backgrounds are modelled with Monte Carlo (MC)
  - Lepton fakes contribution → Data-driven with Matrix method

- Sources for fake electrons
  - Light- or gluon jets
  - Deposit most energy in the electromagnetic calorimeter (ECAL)
  - γ conversion in the ECAL

- Sources for fake muons
  - High energy particles with elongated shower shapes (Punch-Through)
  - Decay of a charged meson (i.e. K^{+}) producing a muon

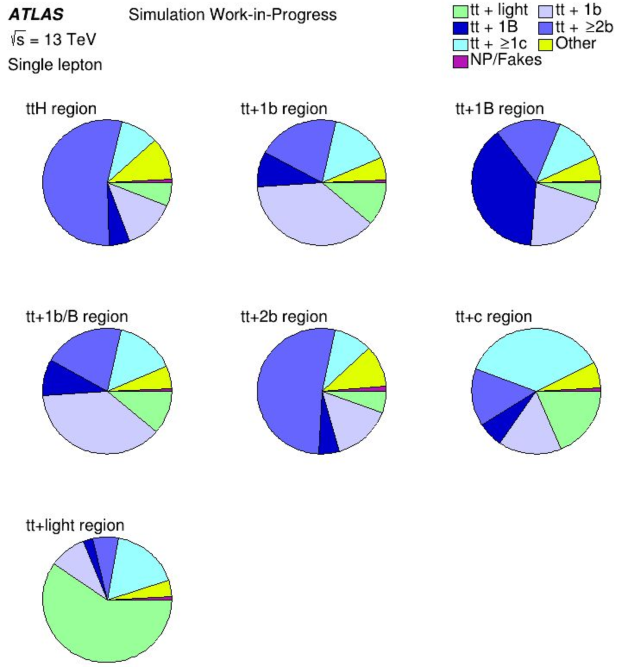


Fake Estimation

- Currently studying the effect of lepton fakes for the analysis
  - No large contribution in SR expected but maybe in CR

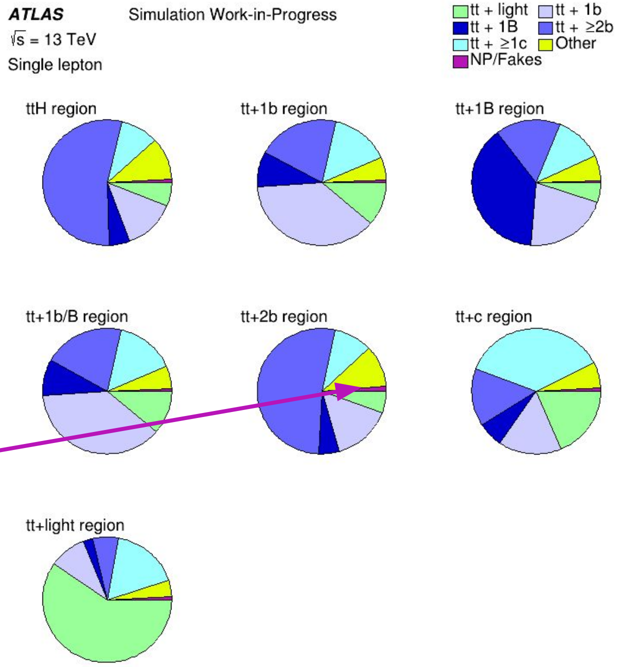
- Splitting $t\bar{t} + b\bar{b}$ in three different processes
  - $t\bar{t} + 1b$
  - $t\bar{t} + 1B$
  - $t\bar{t} + \geq 2b$

- Control regions now defined by process



Fake Estimation

- Currently studying the effect of lepton fakes for the analysis
  - No large contribution in SR expected but maybe in CR
- Splitting $t\bar{t} + b\bar{b}$ in three different processes
  - $t\bar{t} + 1b$
  - $t\bar{t} + 1B$
  - $t\bar{t} + \geq 2b$
- Control regions now defined by process
- First results show no big impact for the background-only composition pie charts for the signal and control regions



Fake Estimation

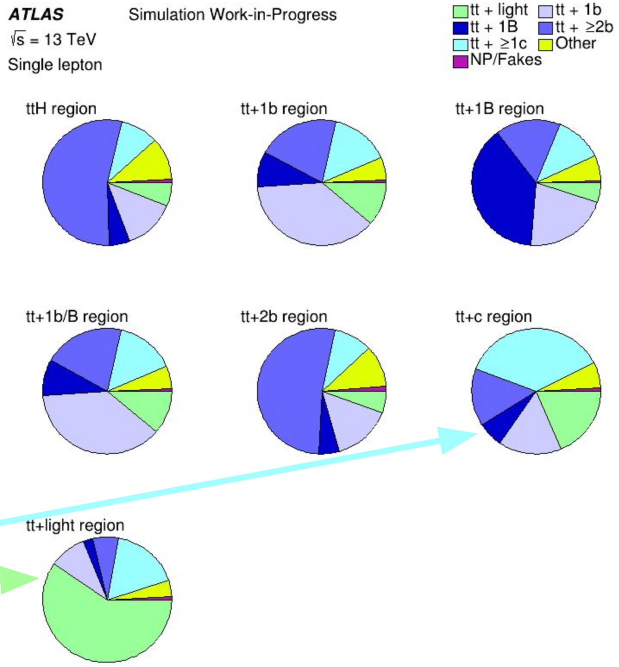
- Currently studying the effect of lepton fakes for the analysis
  - No large contribution in SR expected but maybe in CR

- Splitting $t\bar{t} + b\bar{b}$ in three different processes
  - $t\bar{t} + 1b$
  - $t\bar{t} + 1B$
  - $t\bar{t} + \geq 2b$

- Control regions now defined by process

- First results show no big impact for the background-only composition pie charts for the signal and control regions

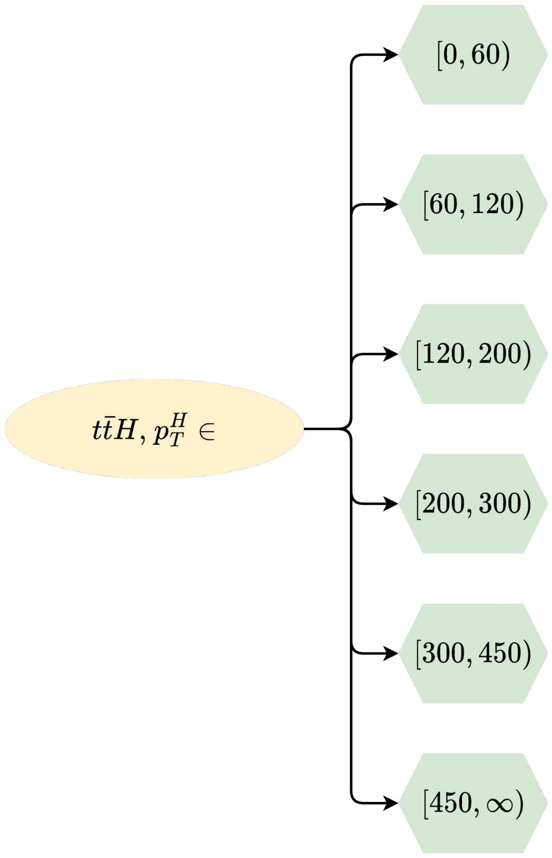
- But: Redefined control regions are much cleaner in comparison to last analysis (i.e $t\bar{t}$ + light and $t\bar{t} + c$ regions)



Simplified Template Cross-Section (STXS)

- Comparable differential Higgs kinematic measurements between LHC experiments and Higgs decay channels

- Performed in exclusive kinematic phase space regions (STXS bins)

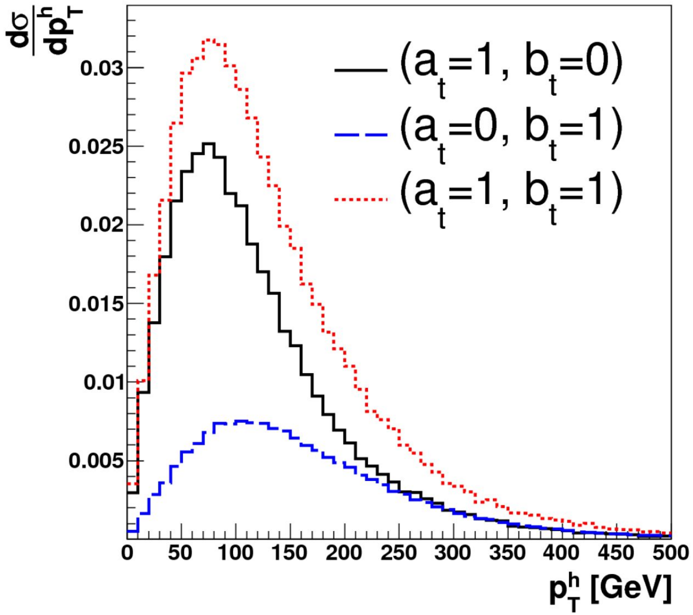


Simplified Template Cross-Section (STXS)

- Comparable differential Higgs kinematic measurements between LHC experiments and Higgs decay channels

- Performed in exclusive kinematic phase space regions (STXS bins)

- Higgs p_T binning optimised to
  - Reduce theory uncertainties
  - Maximise sensitivity of $H \rightarrow b\bar{b}$ channel
  - Also: Higgs p_T is sensitive to CP structure of the Higgs



arXiv:1501.03157

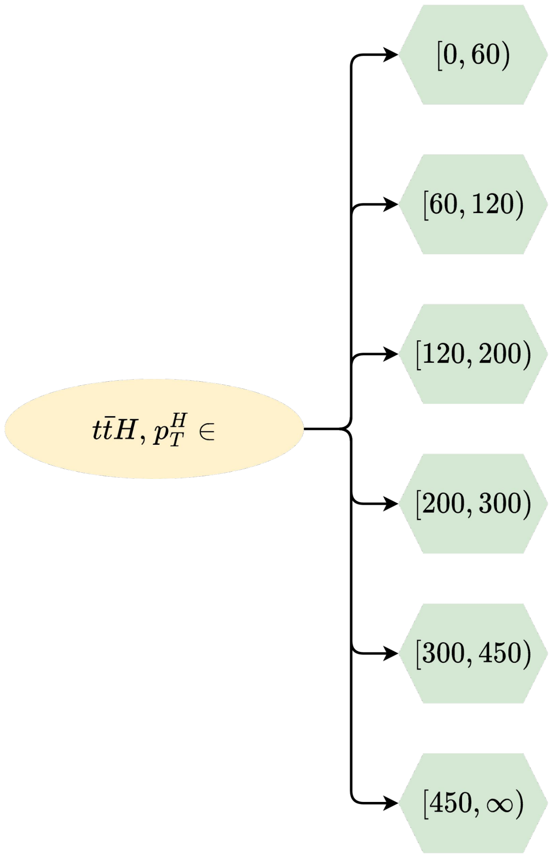
Simplified Template Cross-Section (STXS)

- Comparable differential Higgs kinematic measurements between LHC experiments and Higgs decay channels

- Performed in exclusive kinematic phase space regions (STXS bins)

- Higgs p_T binning optimised to
  - Reduce theory uncertainties
  - Maximise sensitivity of $H \to b\bar{b}$ channel
  - Also: Higgs p_T is sensitive to CP structure of the Higgs

- For each STXS bin a separate signal template is defined
  - Based on the signal truth MC prediction
  - All templates are used for the final Profile Likelihood fit



Simplified Template Cross-Section (STXS)

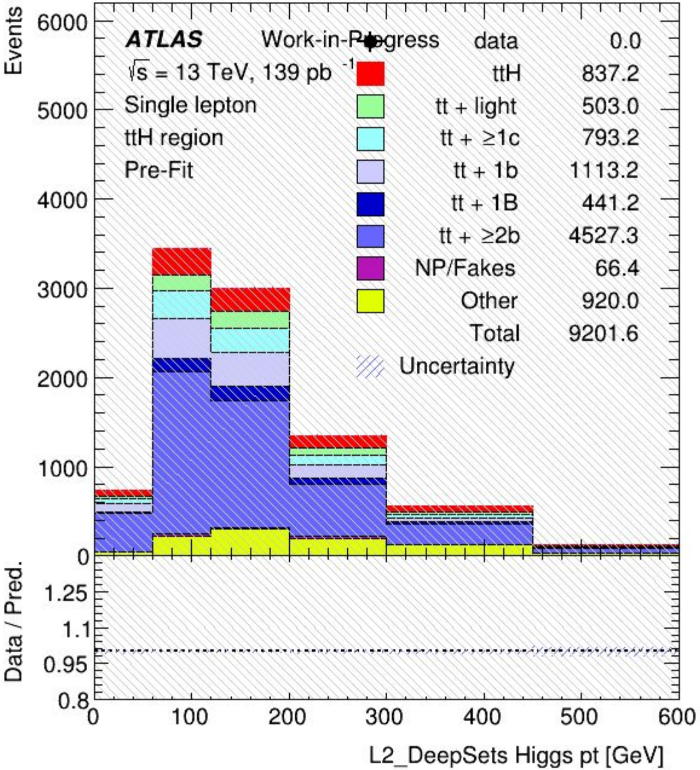
- Comparable differential Higgs kinematic measurements between LHC experiments and Higgs decay channels

- Performed in exclusive kinematic phase space regions (STXS bins)

- Higgs p_T binning optimised to
  - Reduce theory uncertainties
  - Maximise sensitivity of $H \rightarrow b\bar{b}$ channel
  - Also: Higgs p_T is sensitive to CP structure of the Higgs

- For each STXS bin a separate signal template is defined
  - Based on the signal truth MC prediction
  - All templates are used for the final Profile Likelihood fit

- MC composition of the different STXS bins
  - Fully blinded due to blinding policy
  - But: $t\bar{t}H$ contribution clearly visible



$t\bar{t}H$ Summary

- $t\bar{t}H$ with $H \rightarrow b\bar{b}$ analysis presented

- Goal: Measuring the $t\bar{t}H$ cross-section using the STXS method in Higgs p_T bins

- Biggest challenge: Irreducible $t\bar{t} + b\bar{b}$ with large modelling uncertainties

$t\bar{t}H$ Summary

- $t\bar{t}H$ with $H \rightarrow b\bar{b}$ analysis presented
- Goal: Measuring the $t\bar{t}H$ cross-section using the STXS method in Higgs p_T bins
- Biggest challenge: Irreducible $t\bar{t} + b\bar{b}$ with large modelling uncertainties
- Changes to last analysis round (already full Run 2):
  - New signal- and control region definition by novel deep-sets neural network
  - Improved $t\bar{t} + b\bar{b}$ modelling

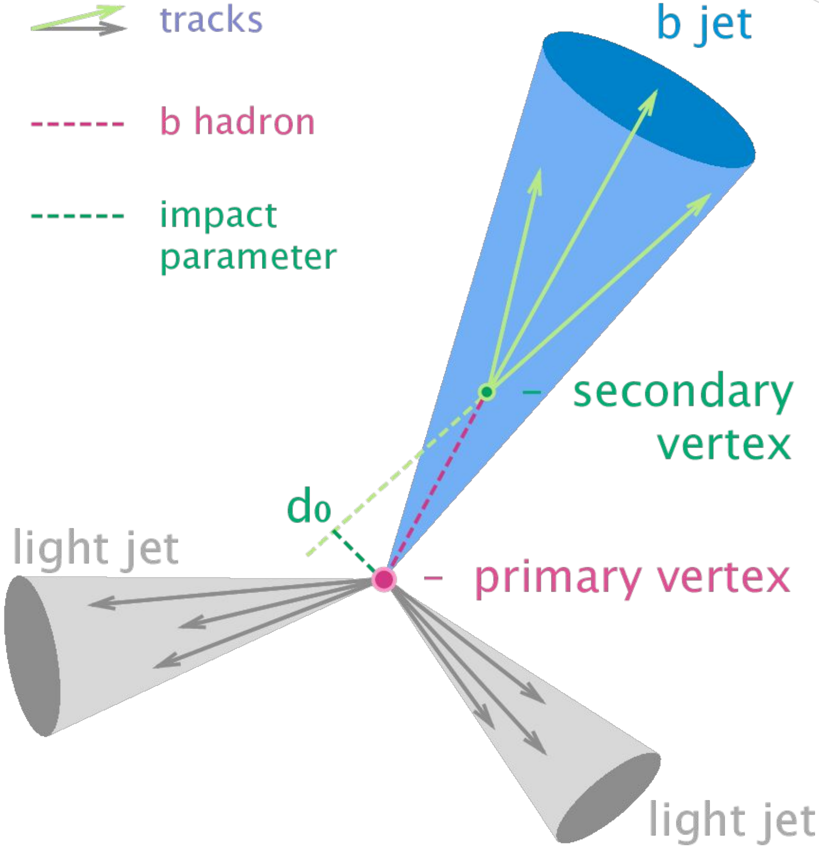
$t\bar{t}H$ Summary

- $t\bar{t}H$ with $H \rightarrow b\bar{b}$ analysis presented
- Goal: Measuring the $t\bar{t}H$ cross-section using the STXS method in Higgs p_T bins
- Biggest challenge: Irreducible $t\bar{t} + b\bar{b}$ with large modelling uncertainties
- Changes to last analysis round (already full Run 2):
  - New signal- and control region definition by novel deep-sets neural network
  - Improved $t\bar{t} + b\bar{b}$ modelling
- First fake studies ongoing → First results show no significant impact of fakes!
- Next up:
  - Finish fake studies
  - Looking into first fits with full systematics
  - Also: Contribute to the $t\bar{t}H$ with $H \rightarrow b\bar{b}$ CP analysis

Heavy-Flavour Tagging

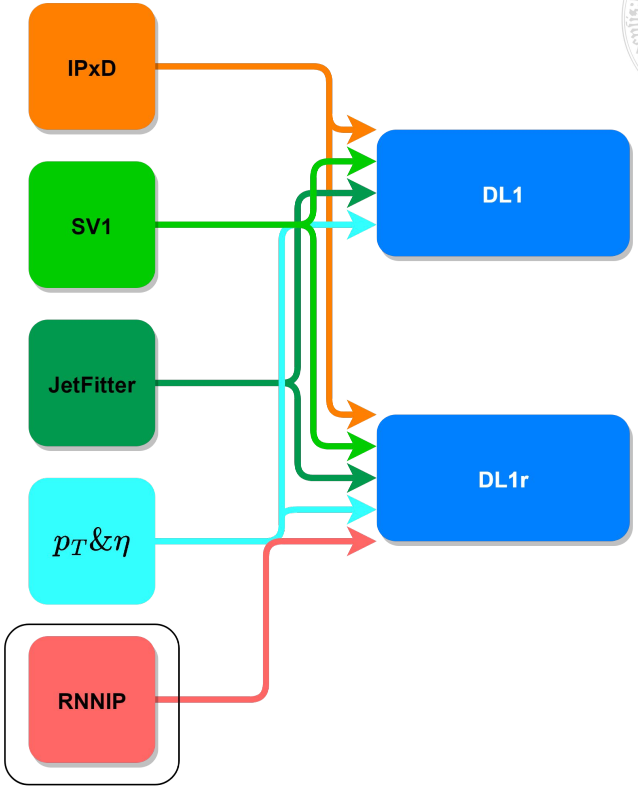
How b-Tagging works

- Using the topology of heavy-flavour jets
  - Lifetime of the b-hadrons ($c \cdot \tau \approx 5\text{mm}$ at $p_T = 50$ GeV)
- Different track- and jet variables are used
- Track variables:
  - e.g number of inner detector hits, $\Delta R(\text{track, jet})$
  - $p_T^{\text{frac}} = \frac{\text{track } p_T}{\text{jet } p_T}$
- Jet variables:
  - e.g. p_T, η
- Also: Information provided by low-level algorithms (i.e. JetFitter, Secondary Vertex Finder (SV1))



ATLAS High-Level *b*-Tagging Algorithms

- Default tagger in Run 2 was DL1r (ATL-PHYS-PUB-2017-013)
- Uses jet-level variables and many different low-level algorithms (i.e. IPxD, SV1, JetFitter)
- For track information, DL1r uses the Recurrent Neural Network Impact Parameter (RNNIP) tagger

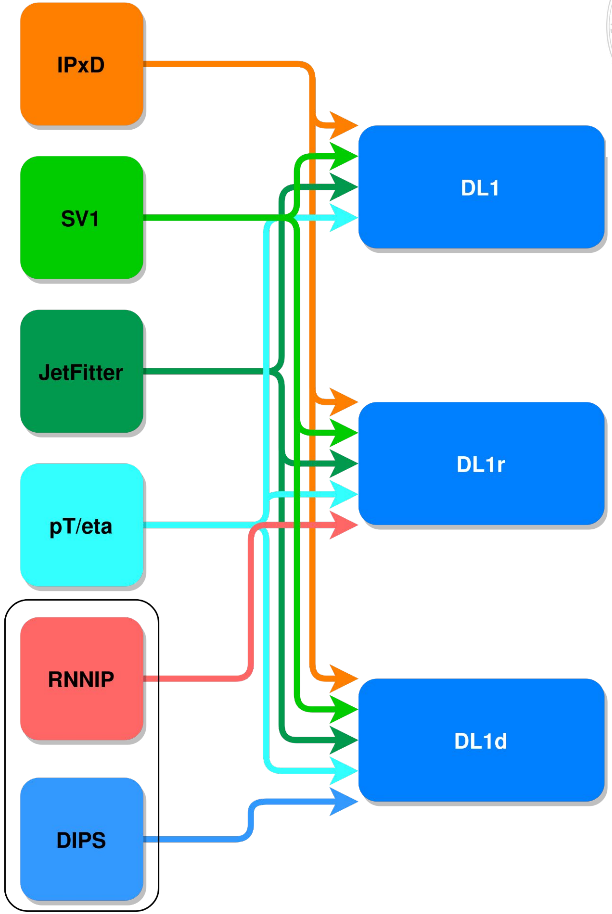


Track-based Neural Network

ATLAS High-Level *b*-Tagging Algorithms

- Default tagger in Run 2 was DL1r (ATL-PHYS-PUB-2017-013)
- Uses jet-level variables and many different low-level algorithms (i.e. IPxD, SV1, JetFitter)
- For track information, DL1r uses the Recurrent Neural Network Impact Parameter (RNNIP) tagger
- Many improvements were implemented for Run 3
- RNNIP was replaced with the Deep-Impact-Parameter-Sets (DIPS) tagger
- DIPS: Deep neural network based on the Deep Sets architecture
- DL1r (r = RNNIP) ➔ DL1d (d = DIPS)
- Biggest change in DL1d w.r.t DL1r → DIPS

<u>DL1d is the recommended high level tagger for Run 3</u>

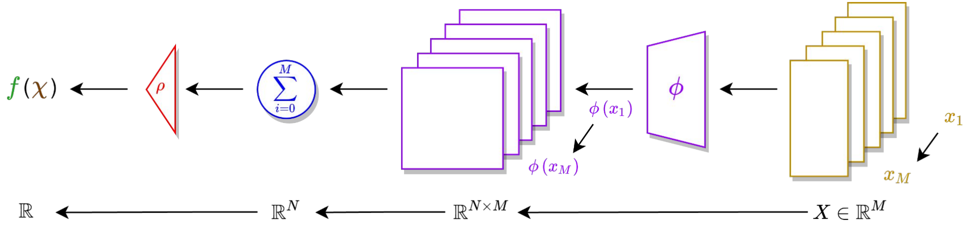


Track-NN based

Deep Sets

- First use in HEP: [arXiv:1810.05165]
- Set function f on set of tracks χ can be decomposed

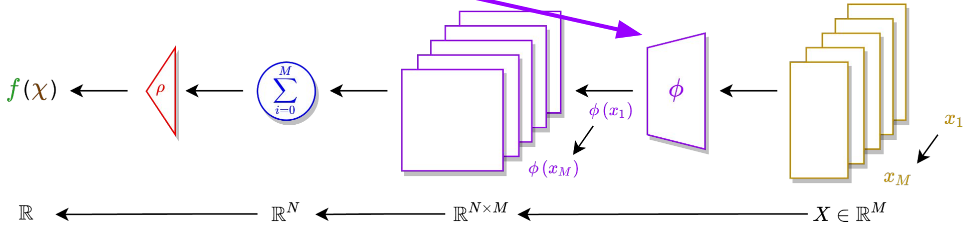
$$f(\chi) = \rho\left(\sum_{x\in\chi}\phi(x)\right)$$



Deep Sets

- First use in HEP: [arXiv:1810.05165]
- Set function ***f*** on set of tracks $\boldsymbol{\chi}$ can be decomposed
- Process each element of the set with mapping function $\boldsymbol{\phi}$

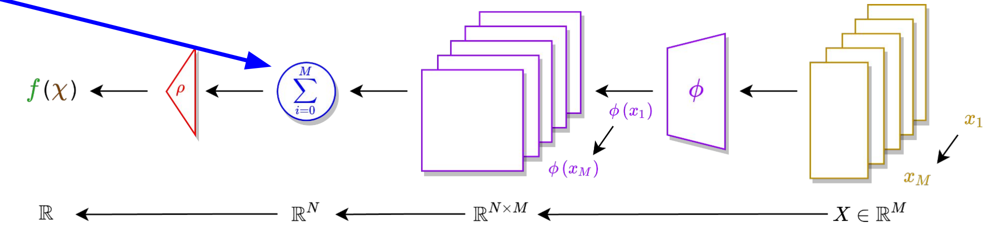
$$f\left(\chi\right)=\rho\left(\sum_{x\in\chi}\phi\left(x\right)\right)$$



Deep Sets

- First use in HEP: arXiv:1810.05165
- Set function ***f*** on set of tracks ***χ*** can be decomposed
- Process each element of the set with mapping function **ϕ**
- Aggregate processed elements into invariant description with aggregation function (here: summation)

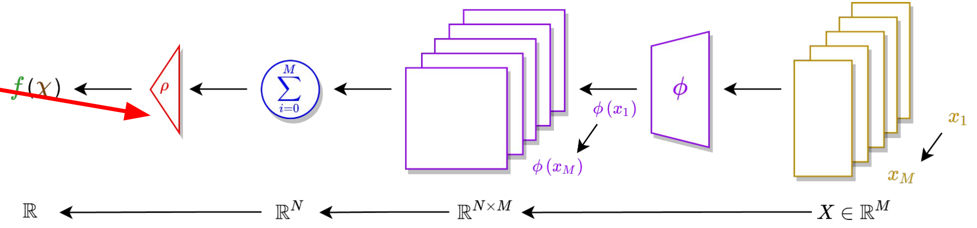
$$f(\chi) = \rho\left(\sum_{x \in \chi} \phi(x)\right)$$



Deep Sets

- First use in HEP: [arXiv:1810.05165]
- Set function ***f*** on set of tracks ***χ*** can be decomposed
- Process each element of the set with mapping function **ϕ**
- Aggregate processed elements into invariant description with aggregation function (here: summation)
- Process the aggregated description with ρ

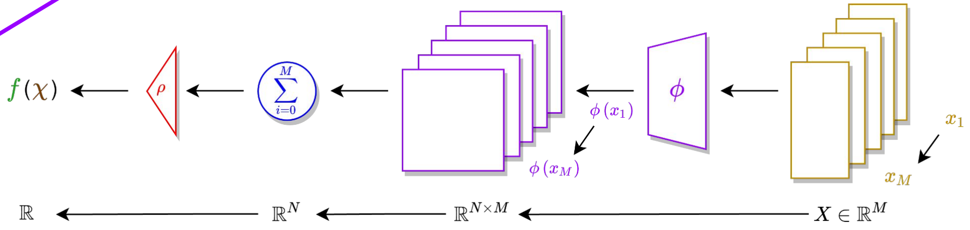
$$f(\chi) = \rho\left(\sum_{x \in \chi} \phi(x)\right)$$



Deep Sets

- First use in HEP: [arXiv:1810.05165]
- Set function ***f*** on set of tracks ***χ*** can be decomposed
- Process each element of the set with mapping function **ϕ**
- Aggregate processed elements into invariant description with aggregation function (here: summation)
- Process the aggregated description with ρ
- **ϕ** and ρ don't operate on set of tracks!
  - **ϕ** works on one track at the time
  - ρ works on the aggregated description
  - Plug in neural network for that!

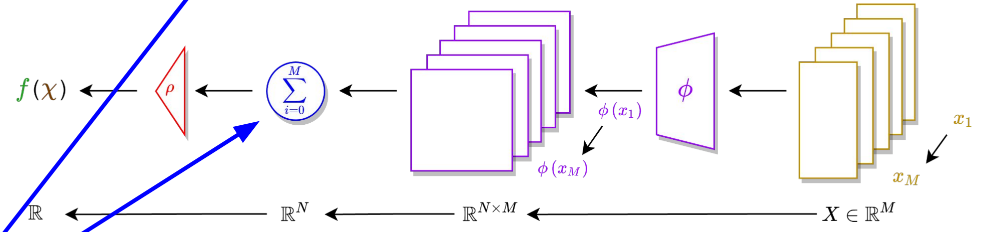
$$f(\chi) = \rho\left(\sum_{x \in \chi} \phi(x)\right)$$



Deep Sets

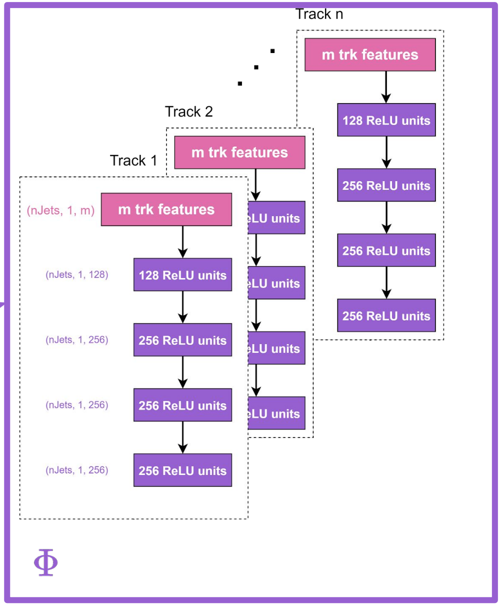
- First use in HEP: arXiv:1810.05165
- Set function f on set of tracks χ can be decomposed
- Process each element of the set with mapping function ϕ
- Aggregate processed elements into invariant description with aggregation function (here: summation)
- Process the aggregated description with ρ
- ϕ and ρ don't operate on set of tracks!
  - ϕ works on one track at the time
  - ρ works on the aggregated description
  - Plug in neural network for that!
- Aggregation negates the order dependency of the set!

$$f(\chi) = \rho\left(\sum_{x \in \chi} \phi(x)\right)$$



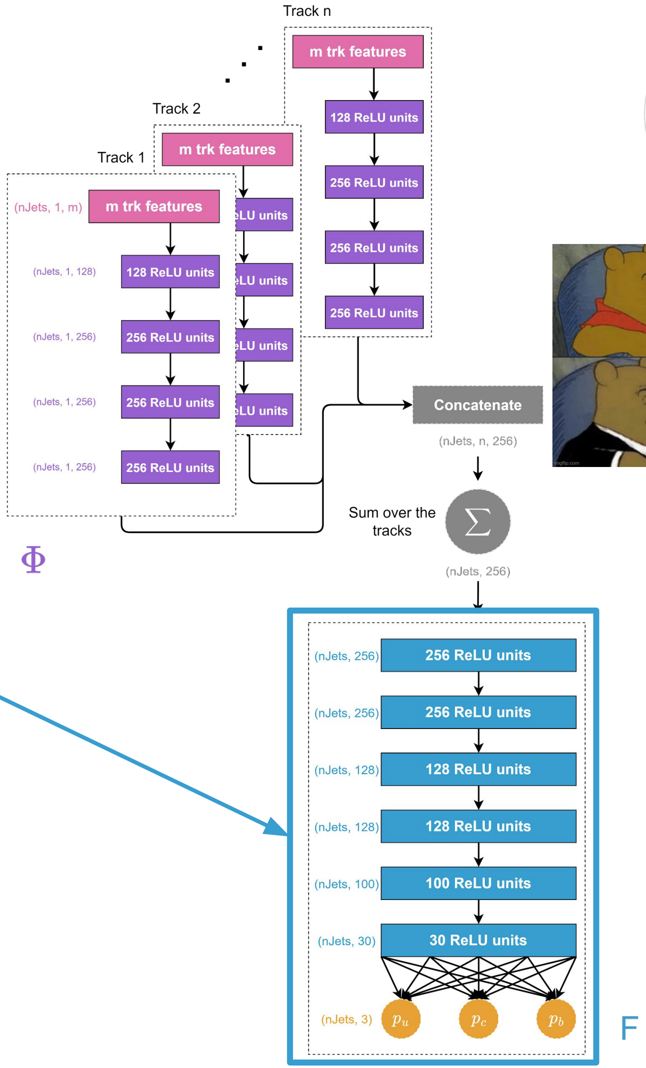
DIPS - Deep Impact Parameter Sets

- First studies by Nicole Hartmann
- Consists of two sub-networks:
  - ϕ: Works on the track input features



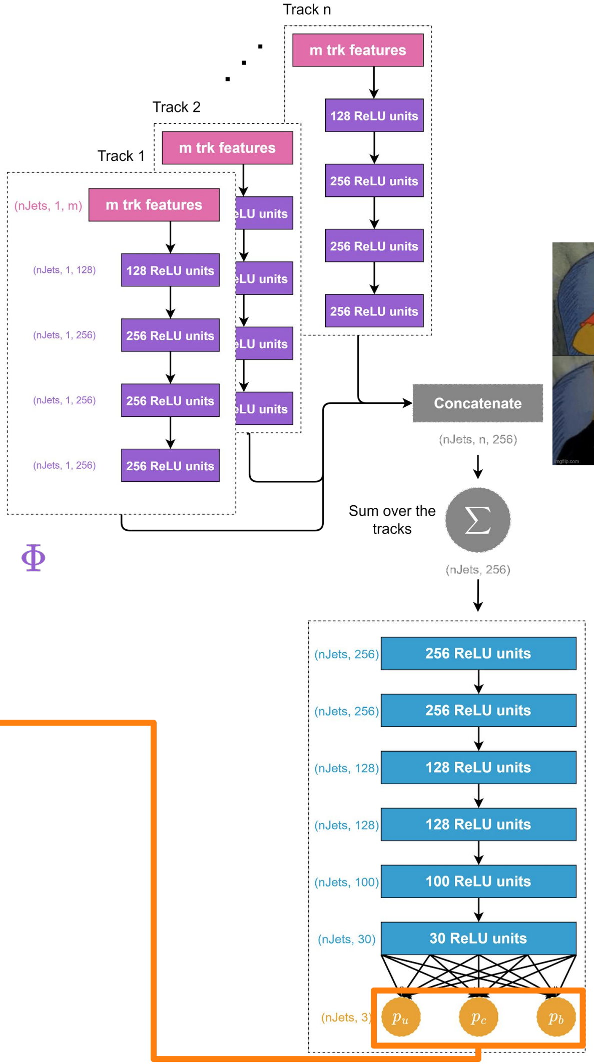
DIPS - Deep Impact Parameter Sets

- First studies by [Nicole Hartmann]
- Consists of two sub-networks:
  - ϕ: Works on the track input features
  - F: Works on the aggregated output of the ϕ networks



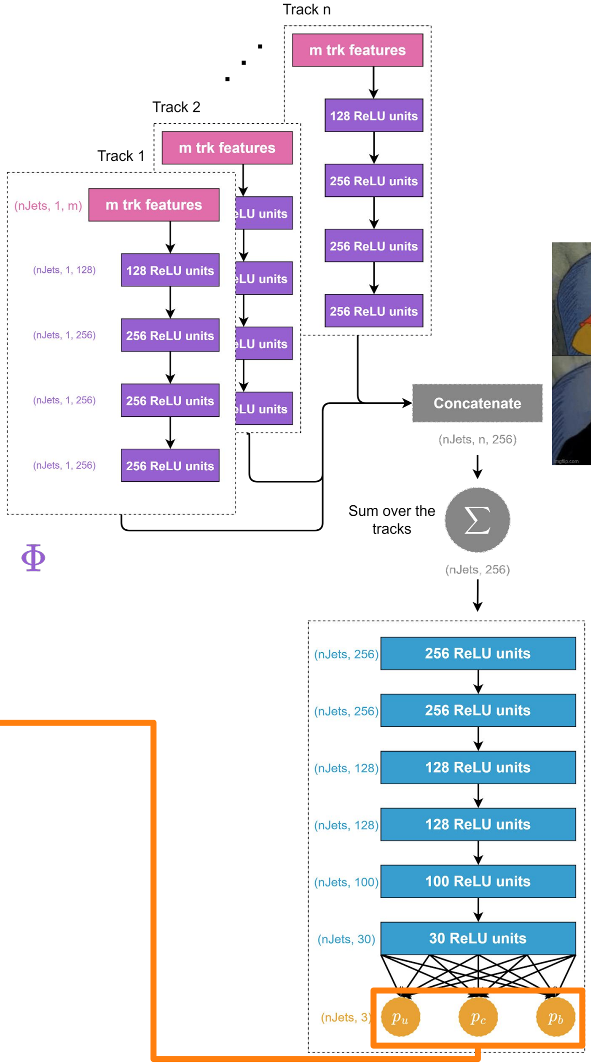
DIPS - Deep Impact Parameter Sets

- First studies by [Nicole Hartmann]
- Consists of two sub-networks:
  - ϕ: Works on the track input features
  - F: Works on the aggregated output of the ϕ networks
- DIPS uses softmax function as last layer activation
→ Outputs can be interpreted as probabilities:
  - p_b: Probability the jet originates from a *b*-quark
  - p_c: Probability the jet originates from a c-quark
  - p_u: Probability the jet originates from a light-flavour quark (up, down, strange)



DIPS - Deep Impact Parameter Sets

- First studies by [Nicole Hartmann]
- Consists of two sub-networks:
  - ϕ: Works on the track input features
  - F: Works on the aggregated output of the ϕ networks
- DIPS uses softmax function as last layer activation
  → Outputs can be interpreted as probabilities:
  - p_b: Probability the jet originates from a *b*-quark
  - p_c: Probability the jet originates from a c-quark
  - p_u: Probability the jet originates from a light-flavour quark (up, down, strange)
- Advantages of the new architecture:
  - Parallelizability of track processing
  - Much faster training time (able to use GPUs)
  - Can go to looser track selection!



Φ

F

Using a basic MLP

Using Deep sets

Training Sample

- Training sample consists of:
  - 70% $t\bar{t}$, 30% Z'
  - $t\bar{t}$: 20-250 GeV, Z': 250-6000 GeV

Training Sample

- Training sample consists of:
  - 70% $t\bar{t}$, 30% Z'
  - $t\bar{t}$: 20-250 GeV, Z': 250-6000 GeV
  - 120M jets in total (40M *b*-, *c*- and light-flavour)
  - 2D-resampling in p_T and $|\eta|$ bins to achieve kinematic independent training
  - Using mixture of over- and undersampling (Importance sampling with replacement)

Training Sample

- Training sample consists of:
  - 70% $t\bar{t}$, 30% *Z'*
  - $t\bar{t}$: 20-250 GeV, *Z'*: 250-6000 GeV
  - 120M jets in total (40M *b*-, *c*- and light-flavour)
  - 2D-resampling in p_T and $|\eta|$ bins to achieve kinematic independent training
  - Using mixture of over- and undersampling (Importance sampling with replacement)
  - Training time per epoch:
    - DIPS: ~31 min (120M jets)
    - RNNIP: ~40 min (6M jets)

-

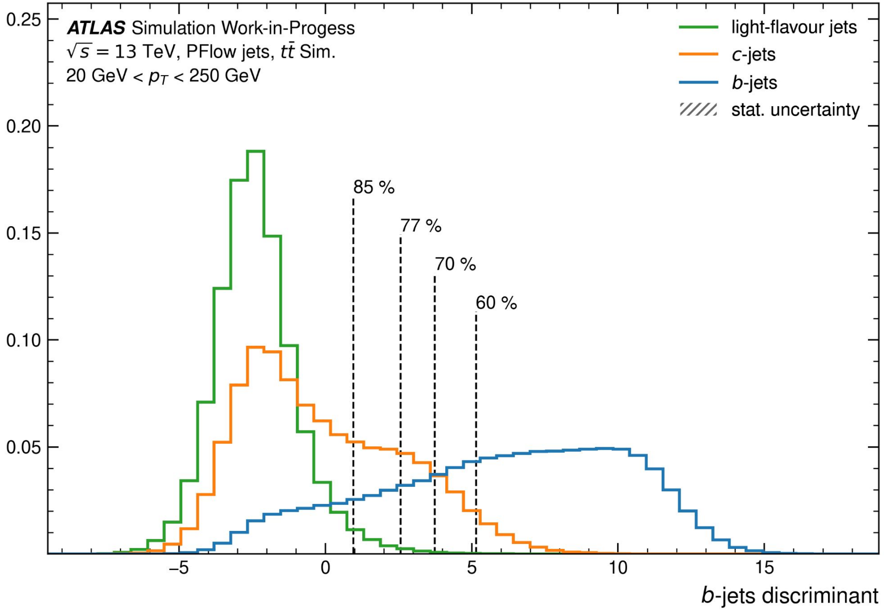
Tagger	recomm. RNNIP	DIPS Default	DIPS Loose
Max N_{Tracks}	25		40
p_T	$> 1\,\text{GeV}$		$> 0.5\,\text{GeV}$
$\|d_0\|$	$< 1\,\text{mm}$		$< 3.5\,\text{mm}$
$\|z_0 \sin(\theta)\|$	$< 1.5\,\text{mm}$		$< 5\,\text{mm}$
$\|\eta\|$	< 2.5		
$N_{\text{Pixel holes}}$	< 2		
$N_{\text{Silicon hits}}$	≥ 7		
$N_{\text{Silicon shared hits}}$	< 2		
$N_{\text{Silicon holes}}$	< 3		

DIPS Results - Discriminant Scores

- Probability outputs of the network is used to calculate the *b*-tagging discriminant D_b

$$D_b = \log(\frac{p_b}{f_c p_c + f_u p_u})$$

- Fraction values can be adapted to balance the two background class rejections



DIPS Results - Discriminant Scores

- Probability outputs of the network is used to calculate the *b*-tagging discriminant D_b

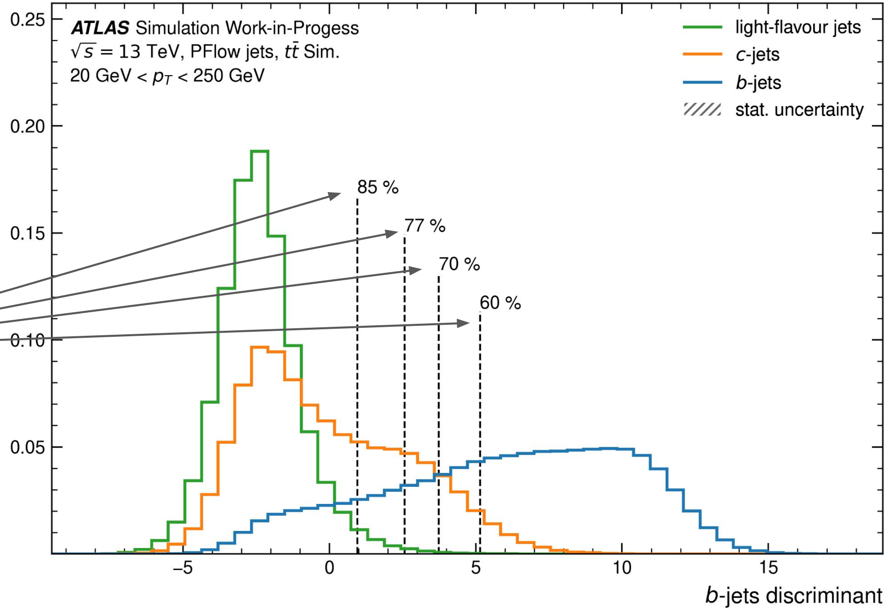
$$D_b = \log\left(\frac{p_b}{f_c p_c + f_u p_u}\right)$$

- Fraction values can be adapted to balance the two background class rejections

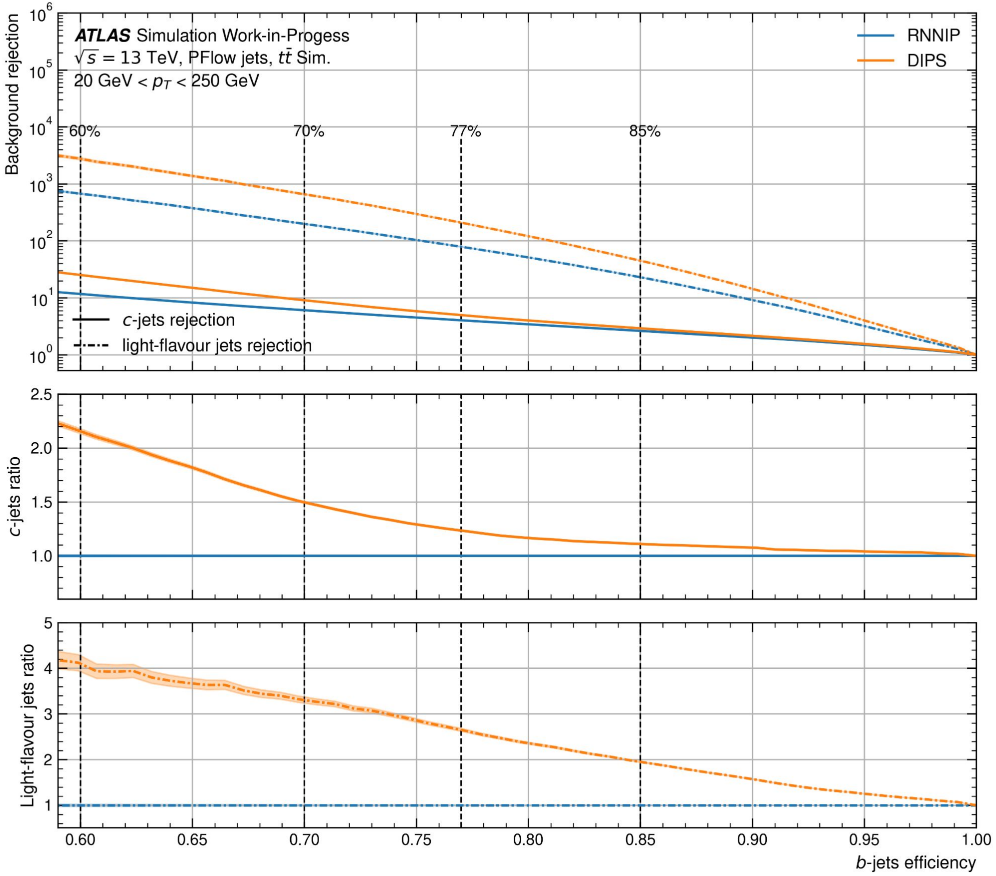
- *b*-tagging working point (WP) is defined such that x% of all *b*-jets are above this value (i.e. 70% WP)

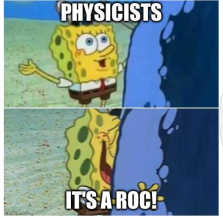
- Cut values for the WPs are calculated by integrating over *b*-distribution from right to left.

- WPs are marked here by vertical lines



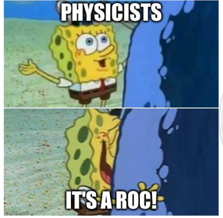
DIPS Results - ROC Curve

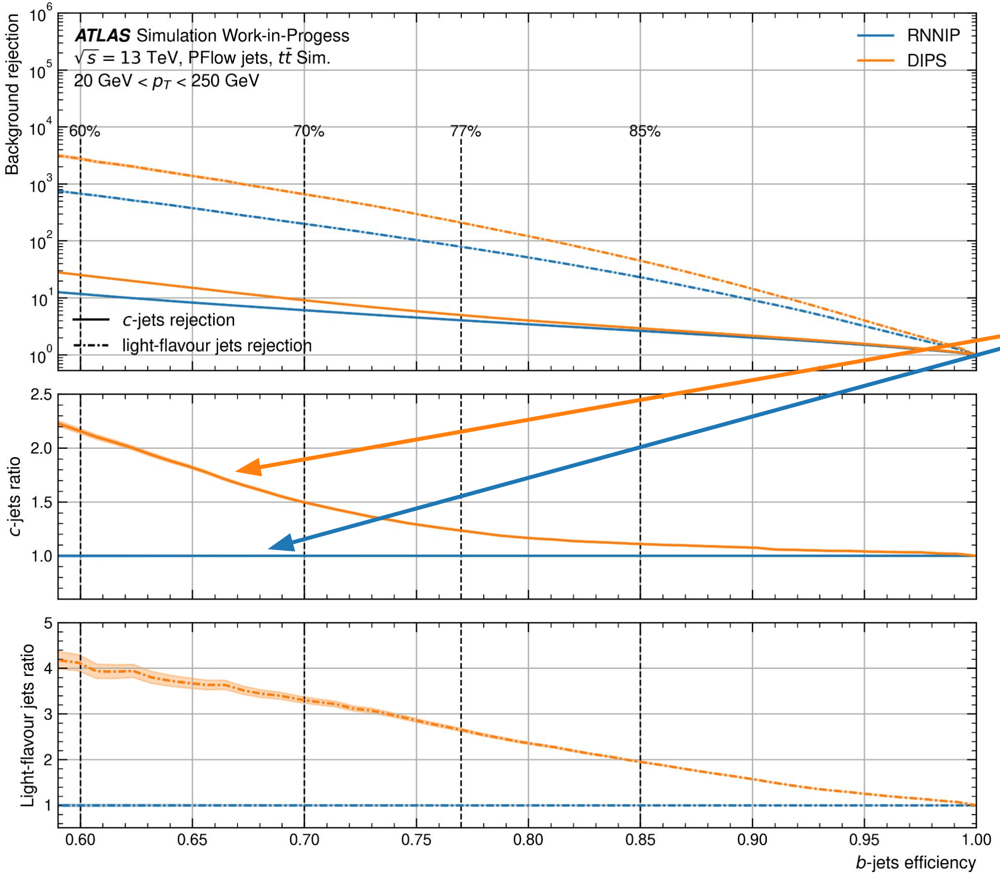




- Comparing background rejections of the two models vs signal efficiency

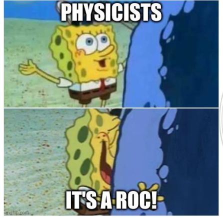
DIPS Results - ROC Curve

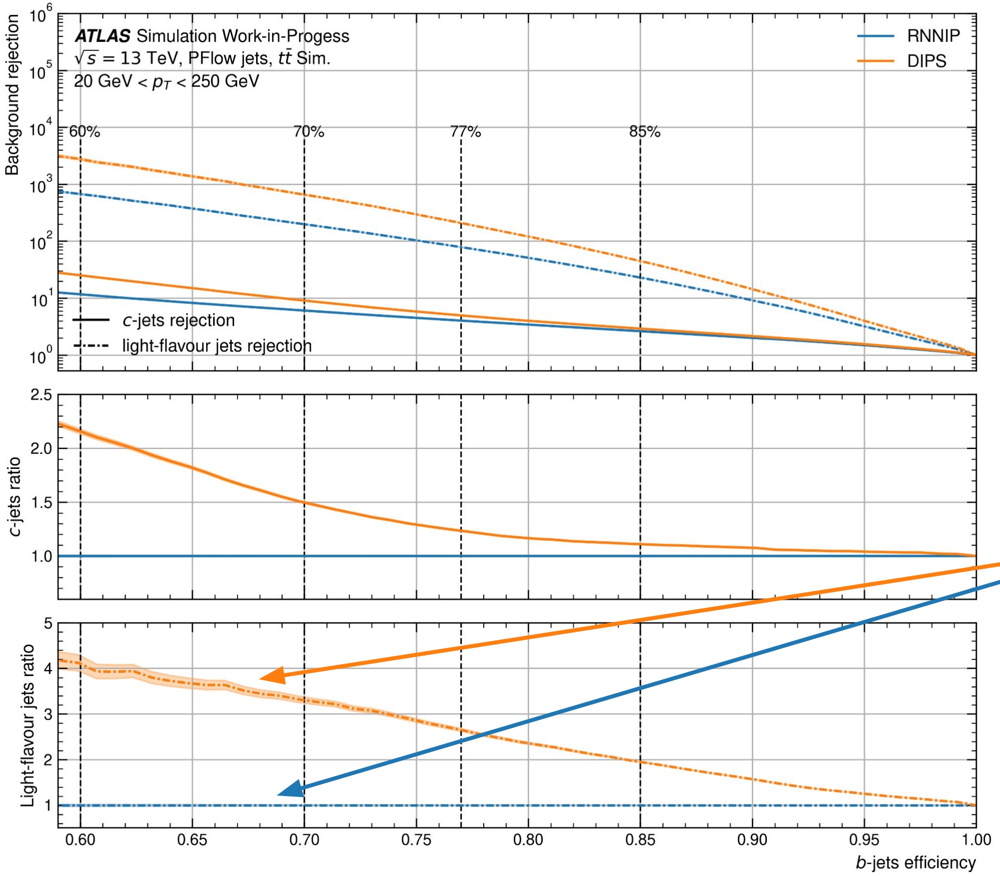




- Comparing background rejections of the two models vs signal efficiency
- Clearly better *c*-rejection for DIPS in comparison RNNIP (~2.15x better at 60% WP)

DIPS Results - ROC Curve



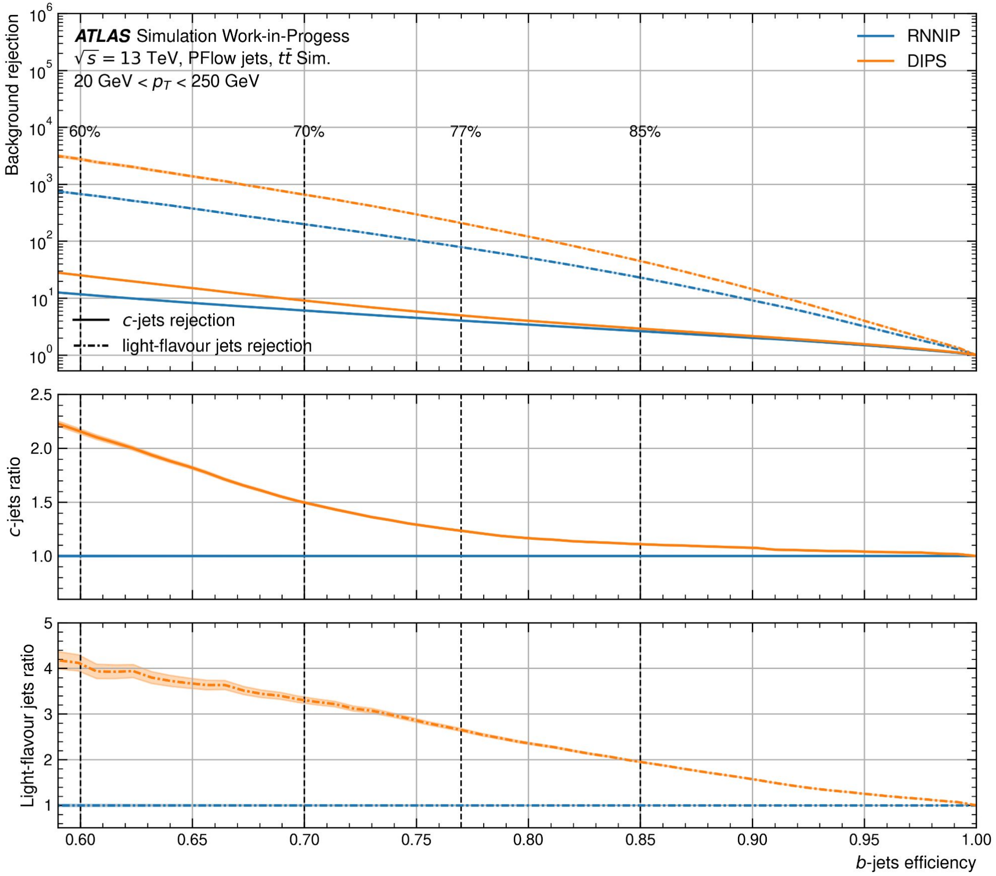


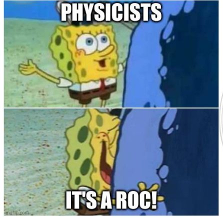
- Comparing background rejections of the two models vs signal efficiency

- Clearly better *c*-rejection for DIPS in comparison RNNIP (~2.15x better at 60% WP)

- Also: Huge improvement in light-flavour rejection for DIPS in comparison RNNIP (~4.05x better at 60% WP)

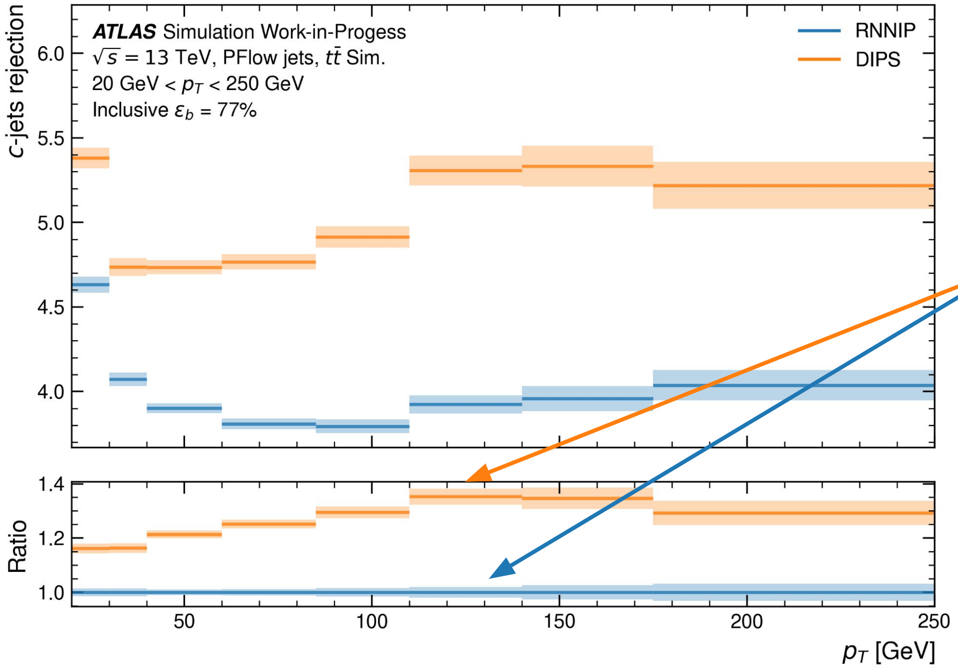
DIPS Results - ROC Curve





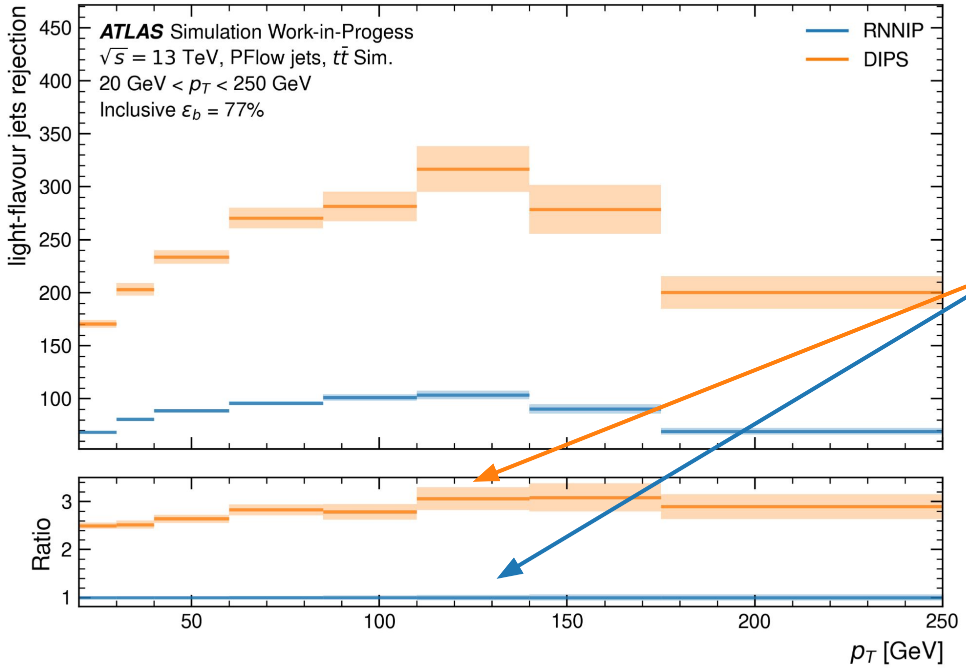
- Comparing background rejections of the two models vs signal efficiency

DIPS Results - Inclusive c-Rejection



- Comparing the inclusive c-rejection per p_T bin for the 77% WP
- Constant improvement for DIPS in comparison to RNNIP (35% improvement for 110-140 GeV bin)

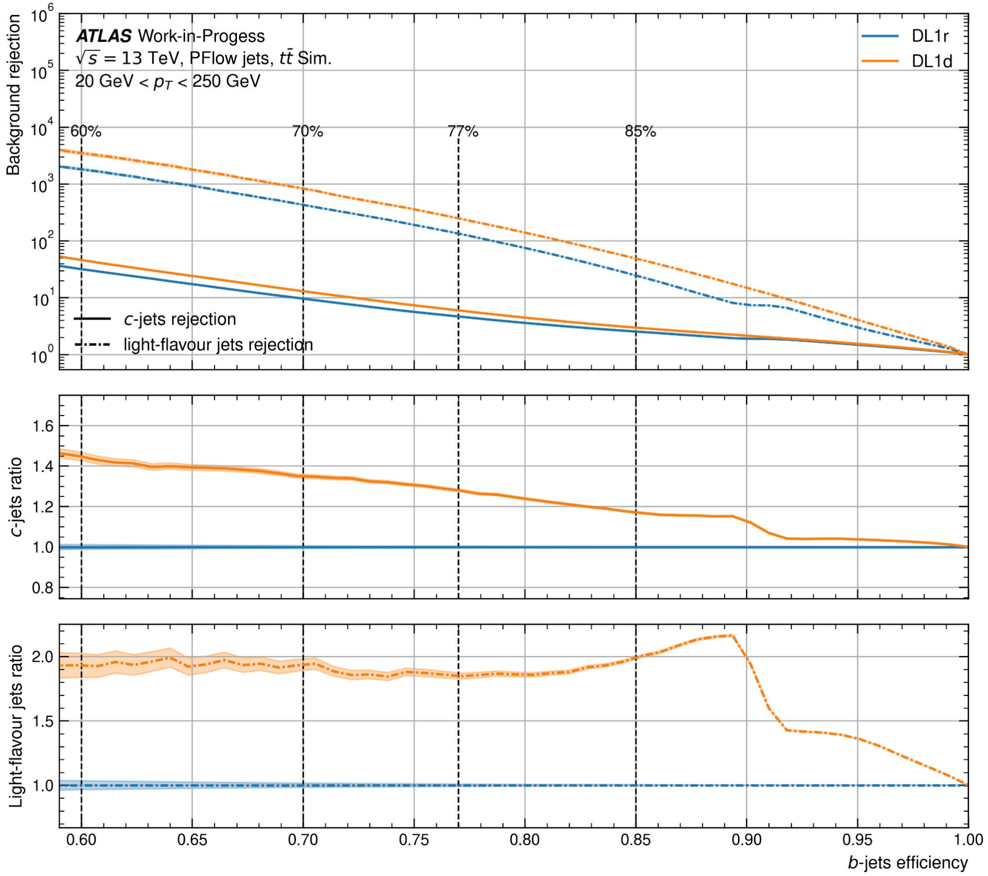
DIPS Results - Inclusive Light-flavour Rejection



- Comparing the inclusive light-flavour rejection per p_T bin for the 77% WP

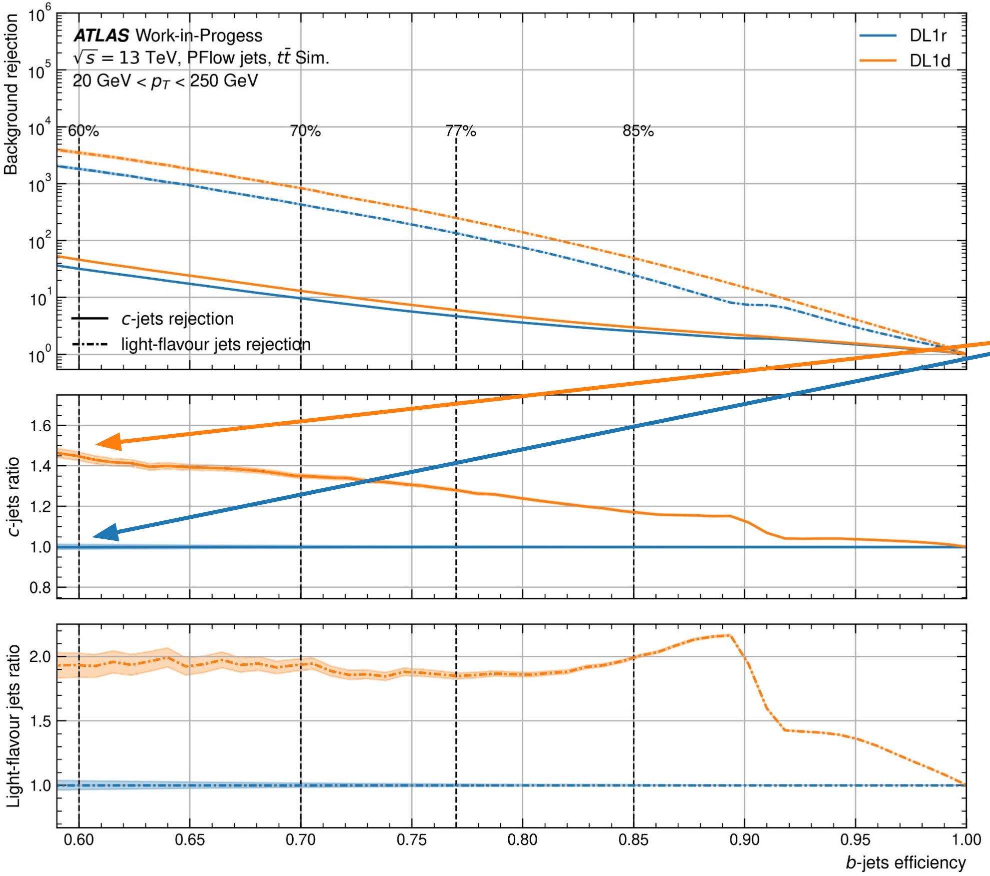
- Constant improvement for DIPS in comparison to RNNIP (200% improvement for 110-140 GeV bin)

DL1d Results - ROC Curve



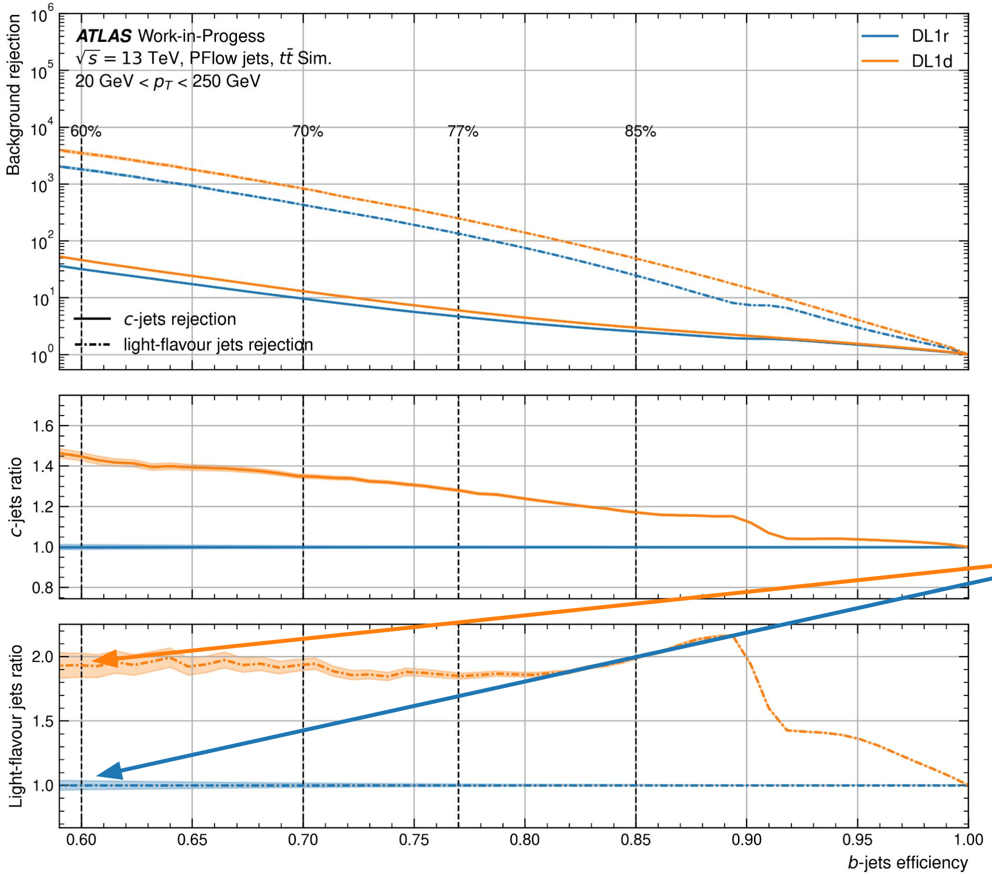
- Comparing background rejections vs signal efficiency for both DL1r and DL1d

DL1d Results - ROC Curve



- Comparing background rejections vs signal efficiency for both DL1r and DL1d
- Clearly better *c*-rejection for DL1d in comparison DL1r
(~1.45x better at 60% WP)

DL1d Results - ROC Curve

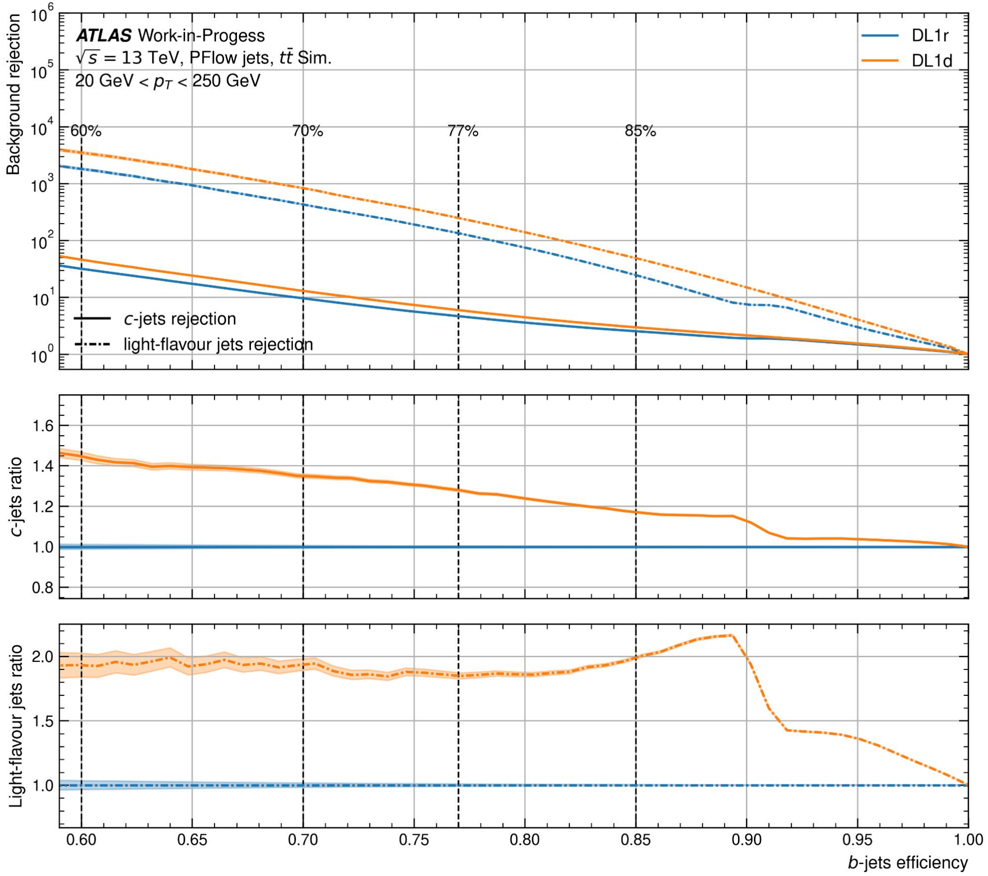


- Comparing background rejections vs signal efficiency for both DL1r and DL1d

- Clearly better *c*-rejection for DL1d in comparison DL1r
(~1.45x better at 60% WP)

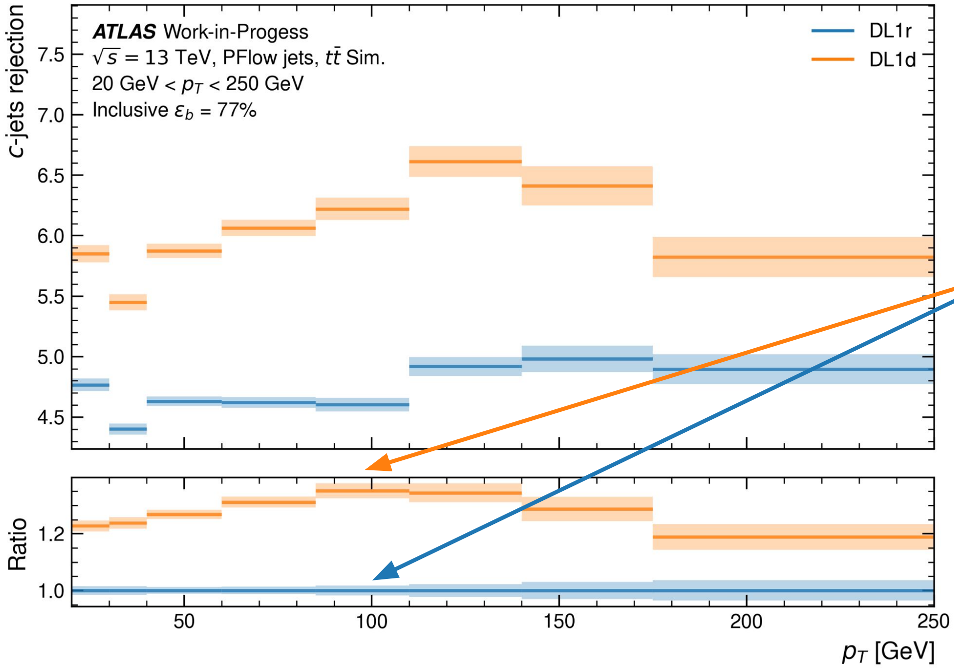
- Also: Huge improvement in light-flavour rejection for DL1d in comparison DL1r
(~1.92x better at 60% WP)

DL1d Results - ROC Curve



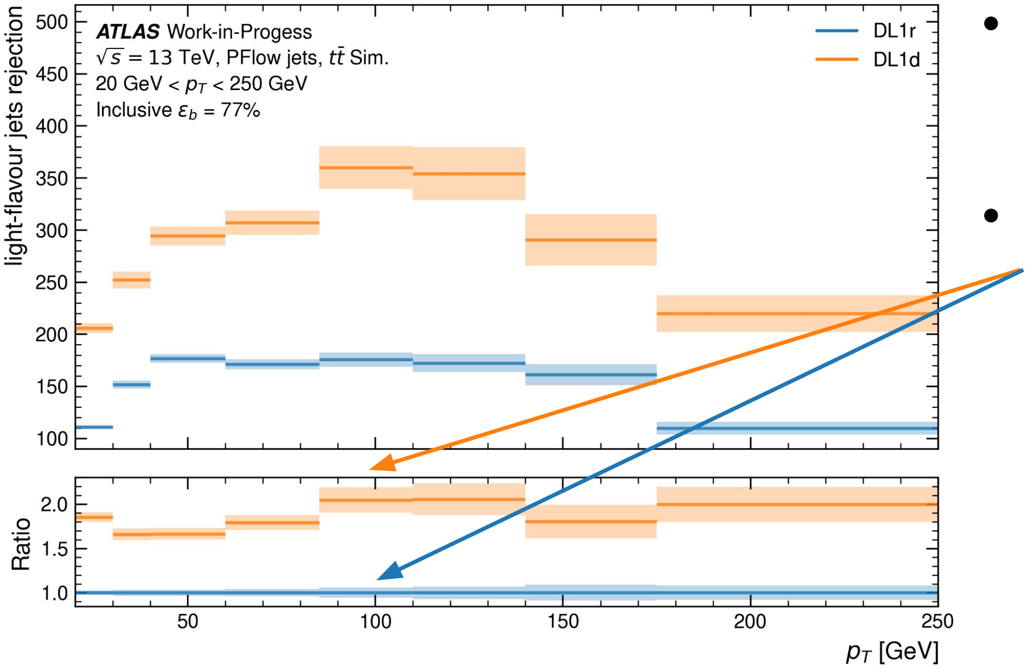
- Comparing background rejections vs signal efficiency for both DL1r and DL1d

DL1d Results - Inclusive *c*-Rejection



- Comparing the inclusive *c*-rejection per p_T bin for the 77% WP
- Constant improvement for DL1d in comparison to DL1r (40% improvement for 85-110 GeV bin)

DL1d Results - Inclusive Light-flavour Rejection



- Comparing the inclusive light-flavour rejection per p_T bin for the 77% WP
- Constant improvement for DL1d in comparison to DL1r (105% improvement for 85-110 GeV bin)

Summary

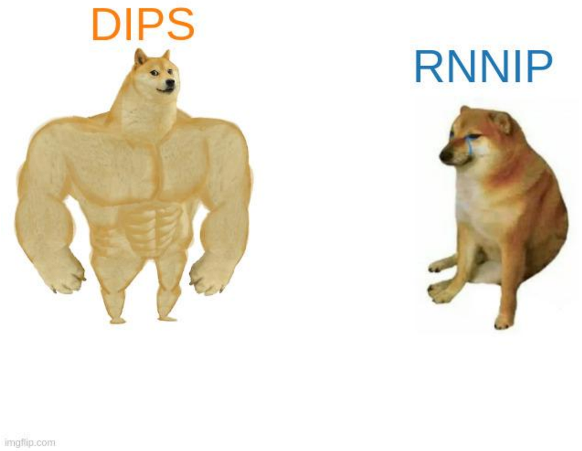
- Presented the new Deep-Impact-Parameter-Sets (DIPS) tagger

Summary

- Presented the new Deep-Impact-Parameter-Sets (DIPS) tagger
- New architecture allows to train with a looser track selection and much higher number of training jets
- Developed a new resampling method, combining over- and undersampling to make full use of the available statistics

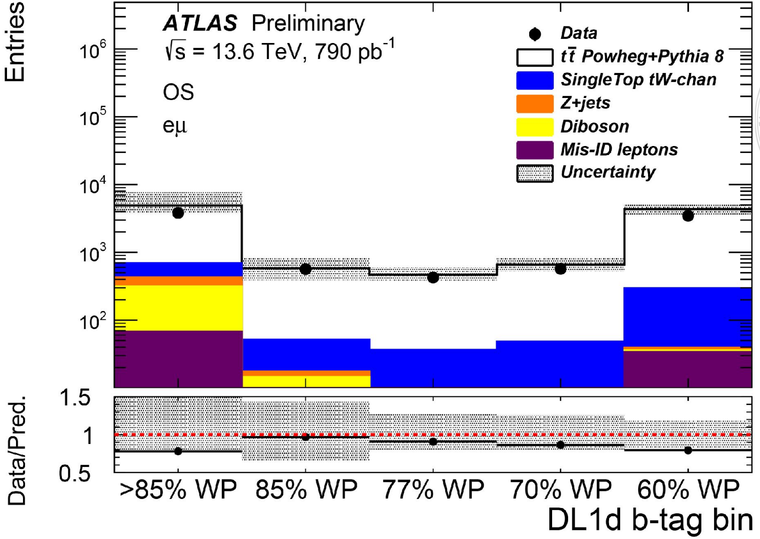
Summary

- Presented the new Deep-Impact-Parameter-Sets (DIPS) tagger
- New architecture allows to train with a looser track selection and much higher number of training jets
- Developed a new resampling method, combining over- and undersampling to make full use of the available statistics
- DIPS clearly outperforms the for Run 2 recommended track-based Recurrent Neural Network Impact Parameter (RNNIP) tagger
- DIPS is part of the new recommended high-level *b*-tagging algorithm DL1d

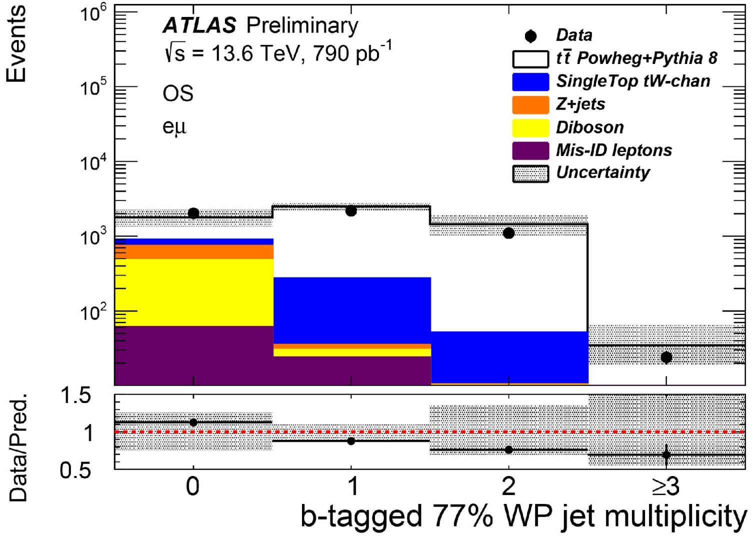


Summary

- Presented the new Deep-Impact-Parameter-Sets (DIPS) tagger
- New architecture allows to train with a looser track selection and much higher number of training jets
- Developed a new resampling method, combining over- and undersampling to make full use of the available statistics
- DIPS clearly outperforms the for Run 2 recommended track-based Recurrent Neural Network Impact Parameter (RNNIP) tagger
- DIPS is part of the new recommended high-level *b*-tagging algorithm DL1d
- First Run 3 public Data/MC plots for DL1d



FTAG-2022-003



Summary

- Presented the new Deep-Impact-Parameter-Sets (DIPS) tagger
- New architecture allows to train with a looser track selection and much higher number of training jets
- Developed a new resampling method, combining over- and undersampling to make full use of the available statistics
- DIPS clearly outperforms the for Run 2 recommended track-based Recurrent Neural Network Impact Parameter (RNNIP) tagger
- DIPS is part of the new recommended high-level *b*-tagging algorithm DL1d
- First Run 3 public Data/MC plots for DL1d

- Presented the current status and improvements of the $t\bar{t}H$ with $H \rightarrow b\bar{b}$ legacy analysis
- Changes to last analysis round:
  - New signal- and control region definition by novel deep-sets neural network
  - Improved $t\bar{t} + b\bar{b}$ modelling

Summary

- Presented the new Deep-Impact-Parameter-Sets (DIPS) tagger
- New architecture allows to train with a looser track selection and much higher number of training jets
- Developed a new resampling method, combining over- and undersampling to make full use of the available statistics
- DIPS clearly outperforms the for Run 2 recommended track-based Recurrent Neural Network Impact Parameter (RNNIP) tagger
- DIPS is part of the new recommended high-level *b*-tagging algorithm DL1d
- First Run 3 public Data/MC plots for DL1d

- Presented the current status and improvements of the $t\bar{t}H$ with $H \rightarrow b\bar{b}$ legacy analysis
- Changes to last analysis round:
  - New signal- and control region definition by novel deep-sets neural network
  - Improved $t\bar{t} + b\bar{b}$ modelling
- Next up:
  - Finish fake studies
  - Looking into first fits with full systematics
  - Also: Contribute to the $t\bar{t}H$ with $H \rightarrow b\bar{b}$ CP analysis

Thanks!
Questions?


BRACE YOURSELF
BACKUP SLIDES ARE COMING
memegenerator.net

Back-Up

Results of the Full Run 2 CP Analysis

- Similar Preselection and signal and control region (SR and CR) defined as the "main" analysis

- Defining CP sensitive observables

$$b_2 = \frac{(\vec{p_1} \times \hat{n}) \cdot (\vec{p_2} \times \hat{n})}{|\vec{p_1}||\vec{p_2}|}$$

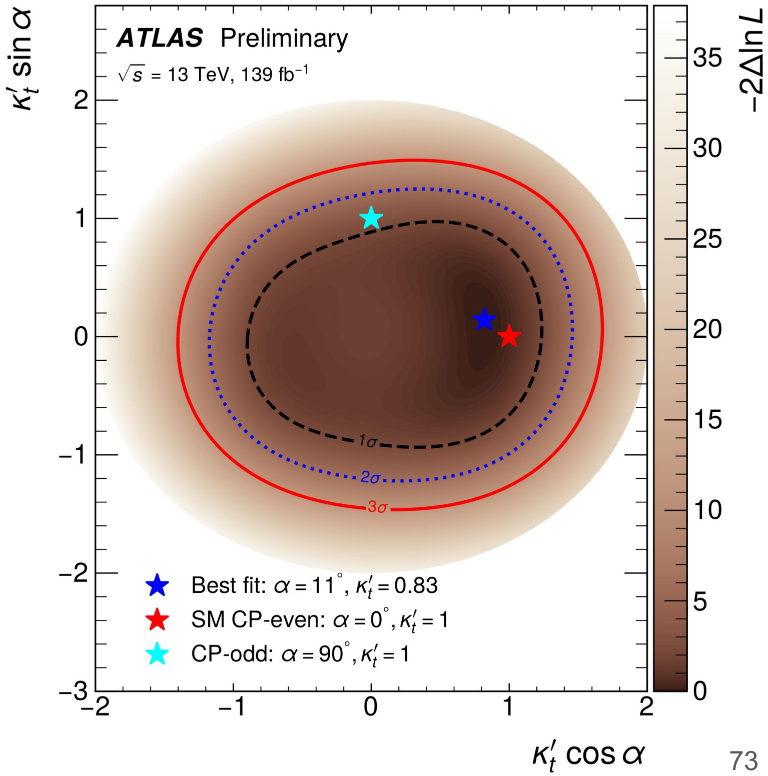
$$b_4 = \frac{p_1^z p_2^z}{|\vec{p_1}||\vec{p_2}|}$$

- Fitting both κ_t and α at the same time with binned profile likelihood fit

- Best fit values: $\kappa_t' = 0.83^{+0.30}_{-0.46}$ $\alpha = 11^{\circ}{}^{+55^{\circ}}_{-77^{\circ}}$

- Expected values: $\alpha_{\text{even}} = 0.0^{\circ}{}^{+49^{\circ}}_{-50^{\circ}}, \ \kappa'_{t,\text{ even}} = 1.00^{+0.25}_{-0.27}$

  $\alpha_{\text{odd}} = 90^{\circ}{}^{+49^{\circ}}_{-43^{\circ}}, \ \kappa'_{t,\text{ odd}} = 1.00^{+0.23}_{-0.33}$



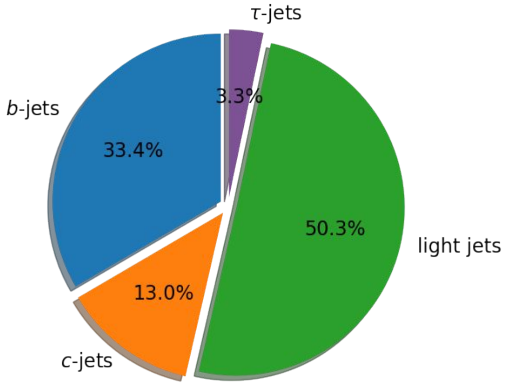
DIPS - Architecture

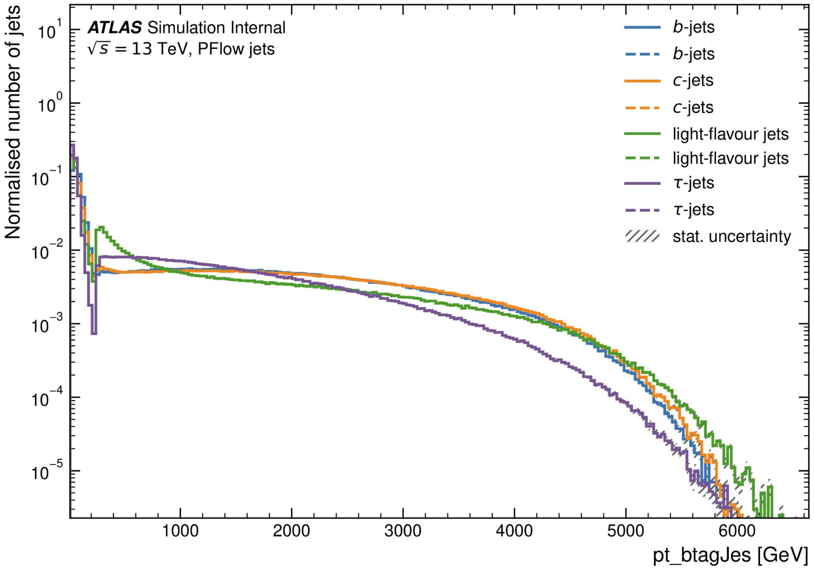
Table 4: Hyperparameters of the different DIPS models

Hyperparameter	PUB Note DIPS	DIPS Loose R21	DIPS Loose R22
Aggregation function	Summation		
Loss function	Categorical Crossentropy		
Optimiser	ADAM (Adaptive Moment Estimation)		
Activation function	ReLU (Rectified Linear Unit)		
Output activation function	Softmax		
Regularisation	Batch Normalisation		
Training sample composition	$t\bar{t}$	70 % $t\bar{t}$, 30 % Z'	
Batch size	256	15000	
ϕ $N_{\text{Hidden layer}}$	3		4
ϕ $N_{\text{Nodes/layer}}$	[100, 100, 128]		[128, 256, 256, 256]
F $N_{\text{Hidden layer}}$	2	4	6
F $N_{\text{Nodes/layer}}$	[100, 100]	[100, 100, 100, 30]	[256, 256, 128, 128, 100, 30]
Number of training jets	3 M	22.8 M	120 M
Free (trainable) parameters	48987	62167	367259
Fixed parameter	1056	1316	3588

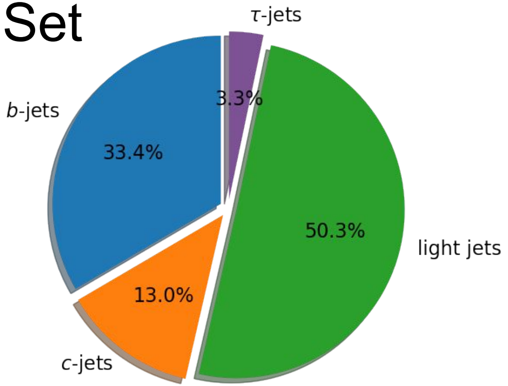
Training Set

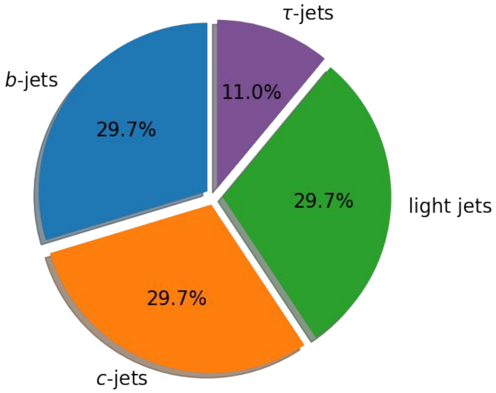
- Hybrid training set:
  - ➢ Low p_T jets: $t\bar{t}$
  - ➢ High p_T jets: Z'
- Different kinematic shapes of the flavours
- For kinematic independent training $p_T - |\eta|$ bin
  - ➢ Resample the different flavours
- Undersampling the jet flavours p_T and $|\eta|$ value
  - ➢ Same number of jets in each
- Ensure tagging independence from kinematics!
- Stitching the two samples to one hybrid sample

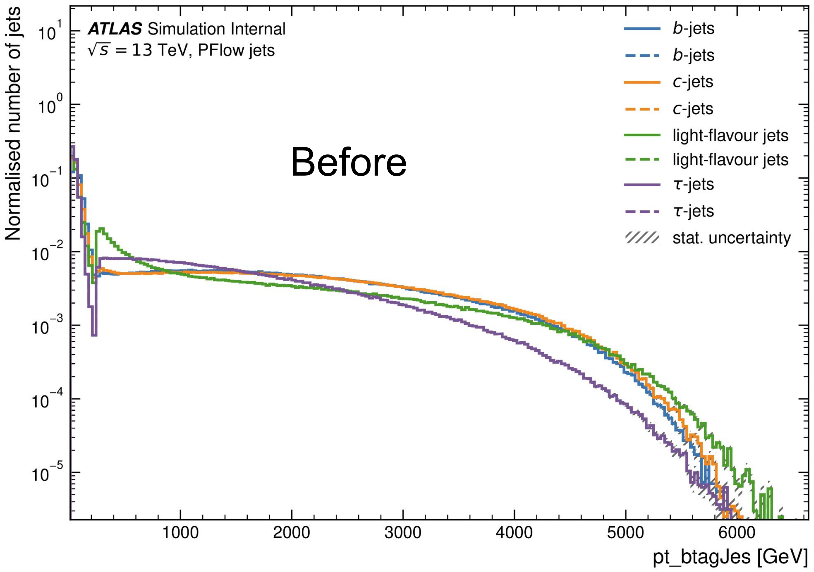


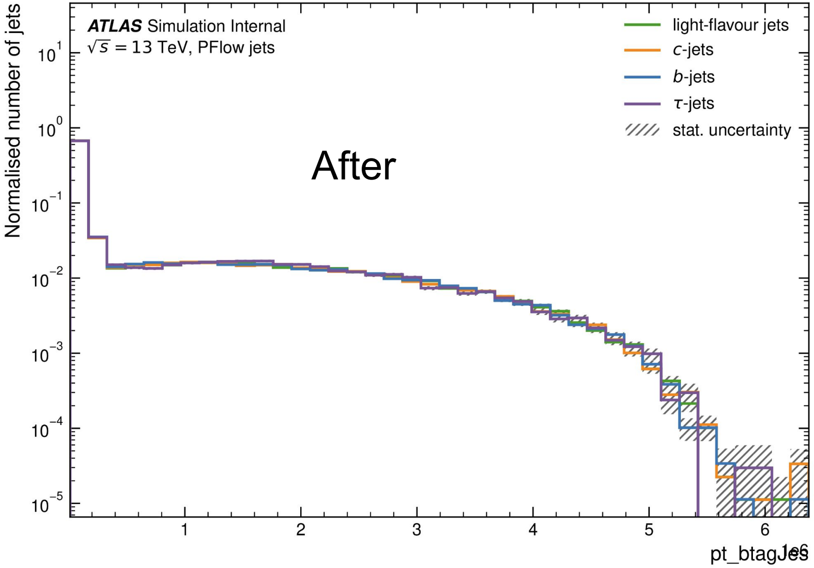


Training Set

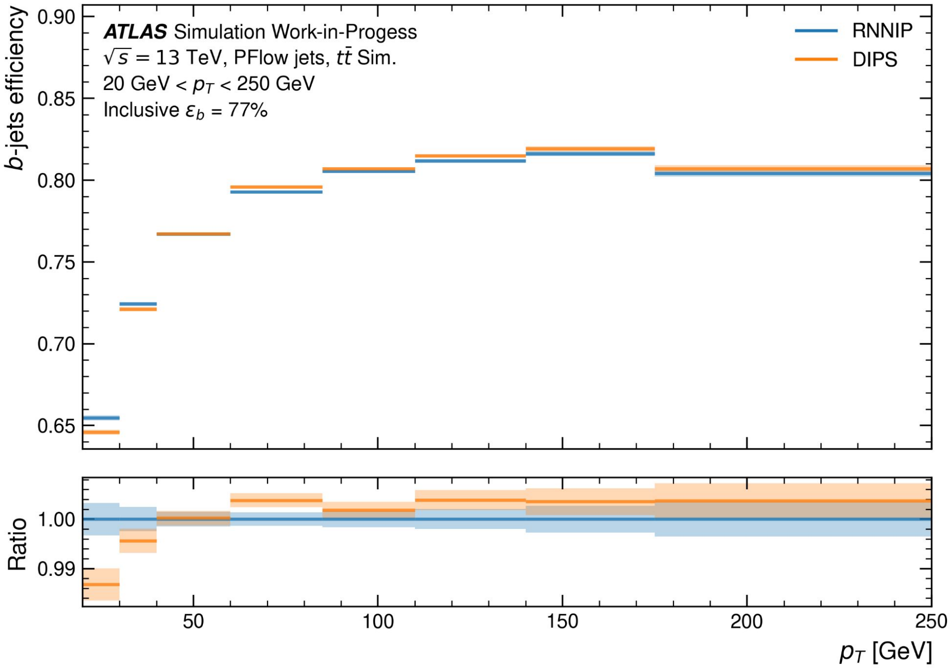






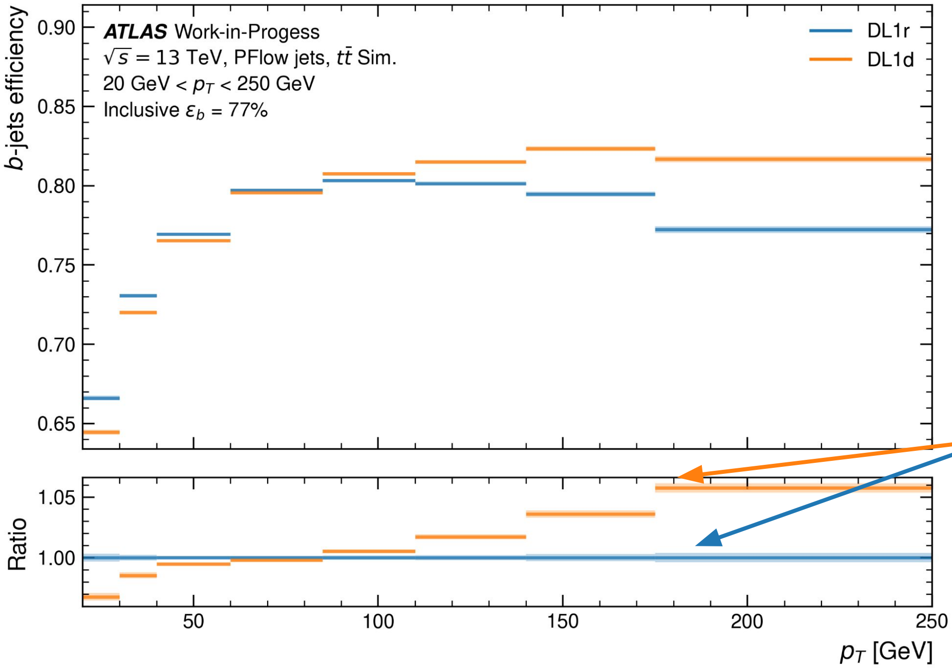


DIPS Results - Inclusive *b*-Efficiency



- Comparing the inclusive *b*-efficiency per p_T bin for the 77% WP
- Similar *b*-efficiency for DIPS and RNNIP
- No significant shift of performance between the bins

DL1d Results - Inclusive b-Efficiency



- Comparing the inclusive b-efficiency per p_T bin for the 77% WP
- Slight shift in performance to higher p_T values
- DL1d has better performance in high p_T than DL1r
(6% improvement for 175-250 GeV bin)
- Shift of around 6%

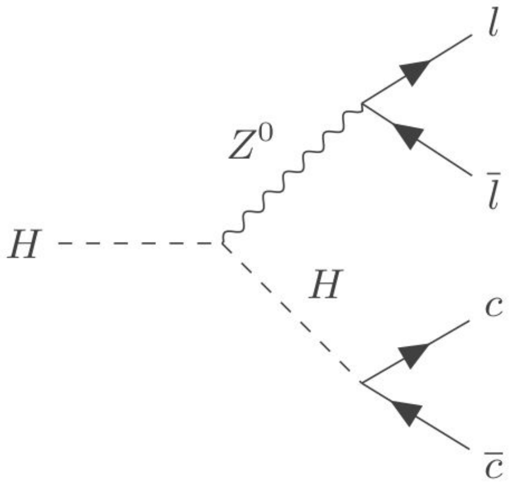
DIPS Tau - Motivation

- Heavy flavour tagging (*b*- and *c*-tagging) are crucial parts for most of ATLAS analyses

- Some analyses are very dependent on *c*-tagging, like the measurement of the *c*-quark Yukawa coupling

- VH(cc) tried to use the DL1 tagger as *c*-tagger using a redefined discriminant

$$D_c = \log\left(\frac{p_c}{f_b p_b + f_u p_u}\right)$$



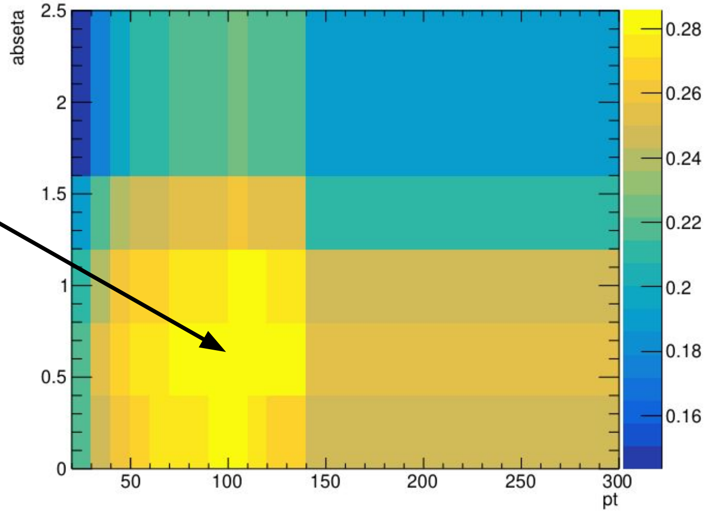
One of the signal processes of the VH(cc) analysis

DIPS Tau - Motivation

$$D_c = \log\left(\frac{p_c}{f_b p_b + f_u p_u}\right)$$

- Good performance was achieved, but a big τ-jet contamination was observed

- For the 27% c-tagging working point (WP), the τ-jet efficiency was at around 28%



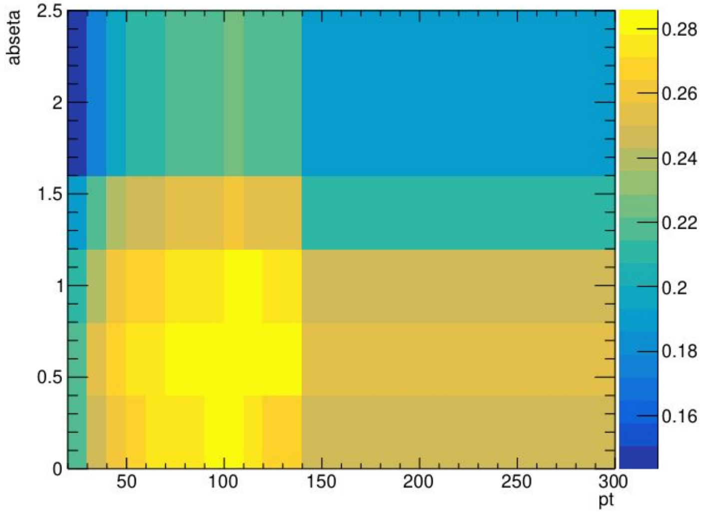
τ-jets efficiency map for the c-tagging working point

ANA-HIGG-2018-41

DIPS Tau - Motivation

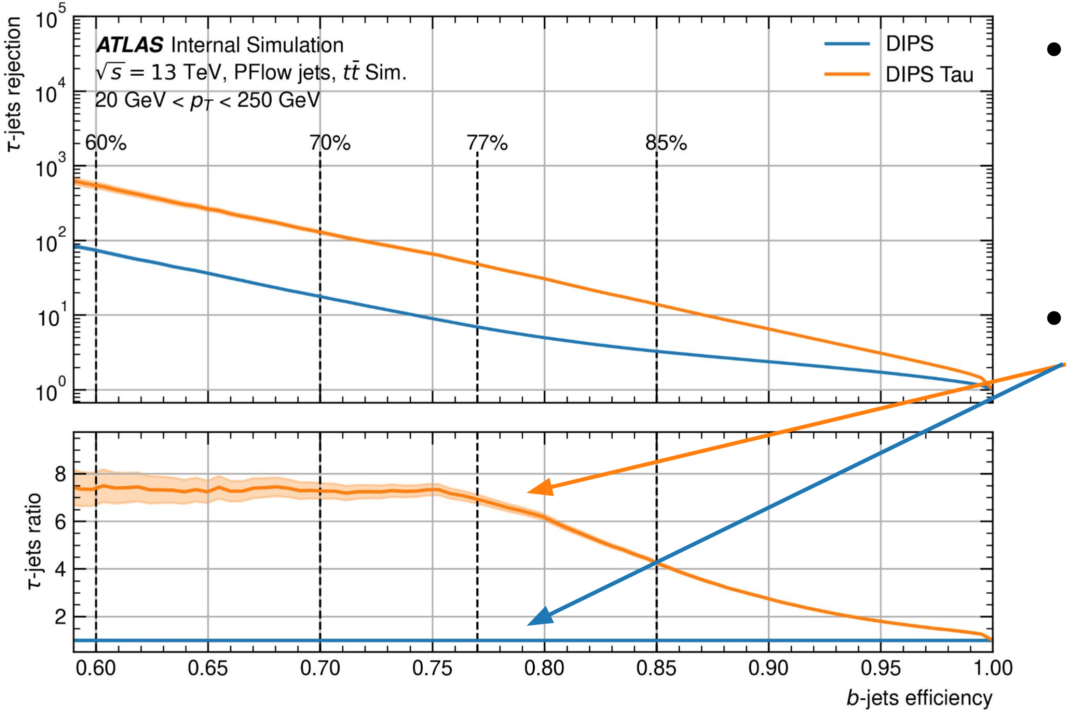
$$D_c = \log\left(\frac{p_c}{f_b p_b + f_u p_u}\right)$$

- Good performance was achieved, but a big τ-jet contamination was observed
- For the 27% *c*-tagging working point (WP), the τ-jet efficiency was at around 28%
- DL1 and DL1r are not trained on τ-jets!
- Can we achieve similar performance for *b*-tagging while adding the τ-jets to the training?



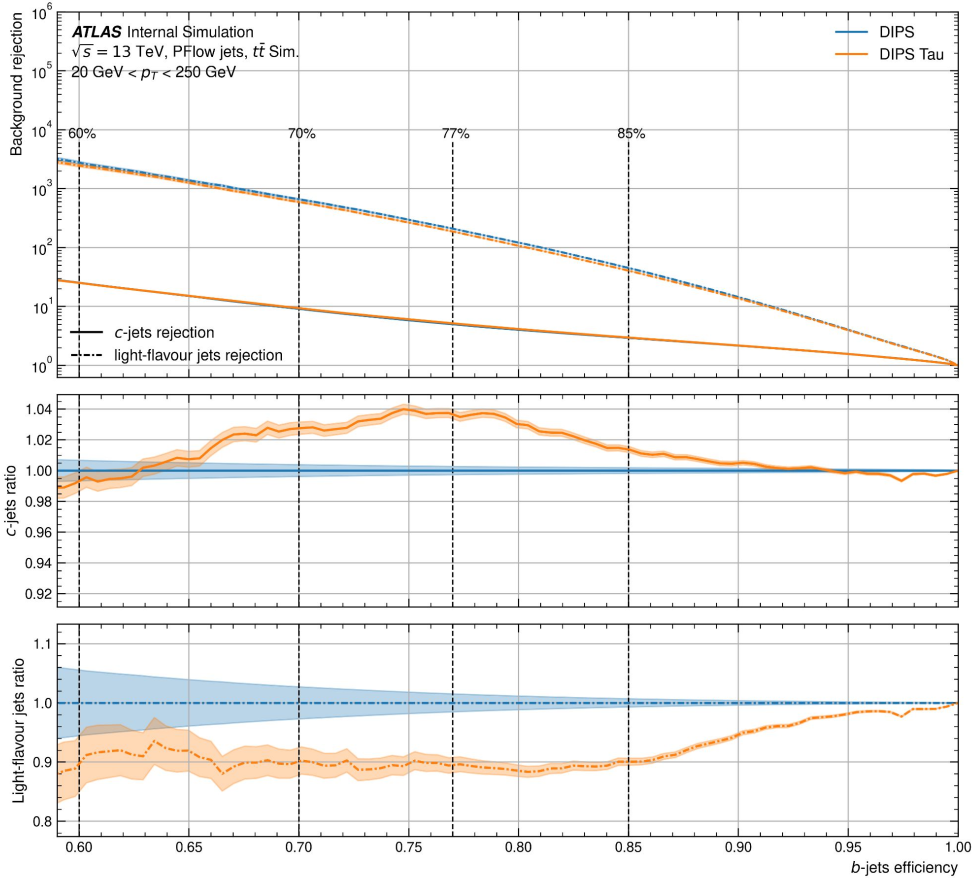
τ-jets efficiency map for the 27% *c*-tagging working point

DIPS Tau - ROC Curve



- Comparing the τ-rejection for the non-τ-trained DIPS and a new four-classes τ-trained DIPS
- Clear improvement over whole *b*-efficiency range for DIPS Tau in comparison to DIPS

DIPS Tau - ROC Curve



- Comparing the *c*- and light-flavour rejection for the non-τ-trained DIPS and a new four-classes τ-trained DIPS
- Slightly better *c*- and slightly worse light-flavour rejection for DIPS Tau in comparison to DIPS
- Light-flavour rejection can be recovered!
- Adapting the fraction values in the *b*-tagging discriminant calculation
  - Shifting performance from *c*-rejection to light-flavour rejection