(Heavy) Flavour Tagging in ATLAS and CMS

Alexander Froch on behalf of the ATLAS and CMS Collaboration

05.06.2023 - HADRON Conference 2023

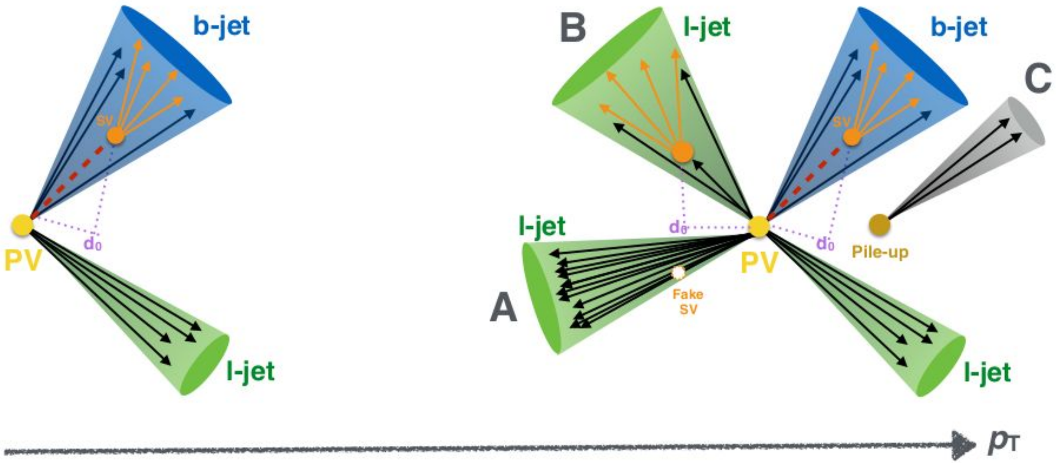
What is Flavour Tagging and how does it work?

- Identify jets originating from heavy flavour (*b*, *c*) quarks and separate from other sources (e.g. light quarks)
    - Mainly *b*-tagging

What is Flavour Tagging and how does it work?

- Identify jets originating from heavy flavour (*b*, *c*) quarks and separate from other sources (e.g. light quarks)
    - Mainly *b*-tagging

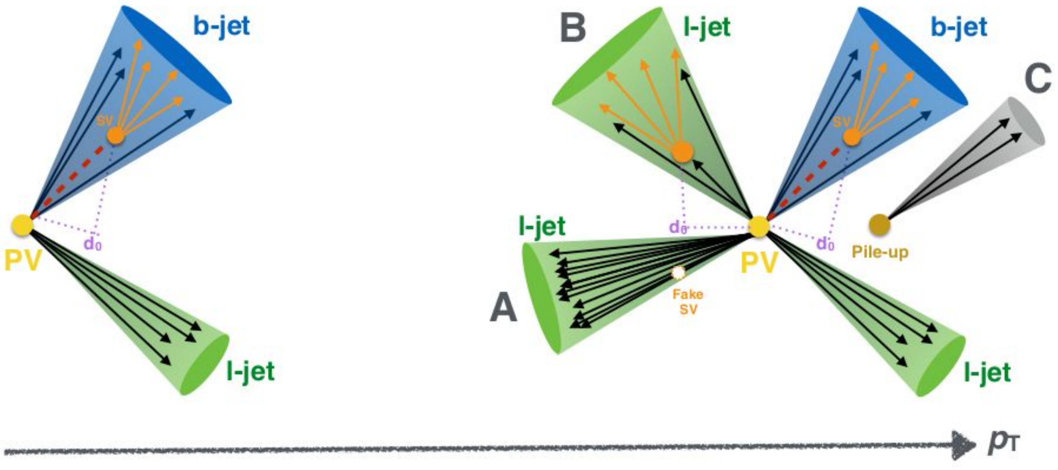
- Using the topology of heavy-flavour jets



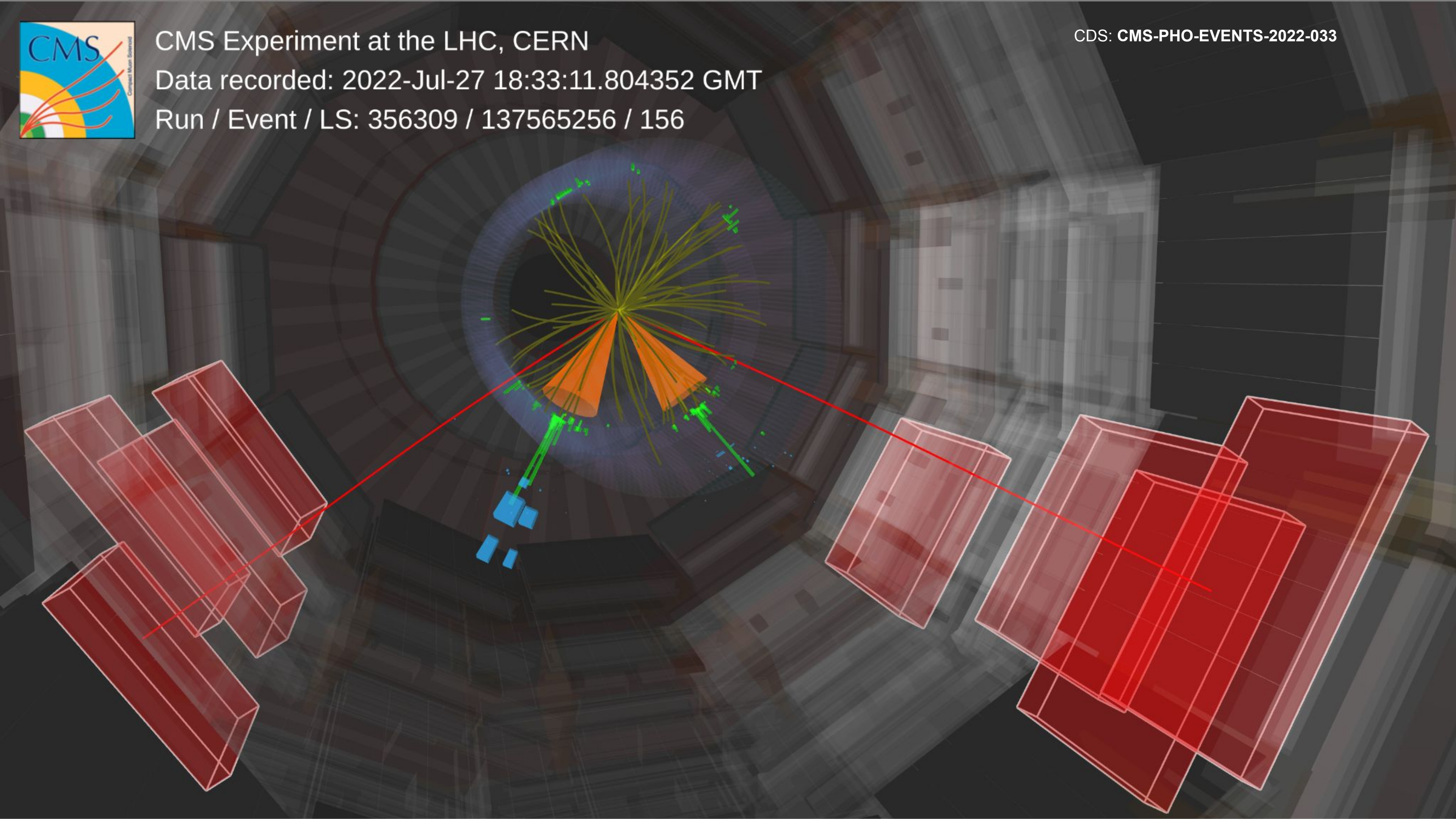
Becomes more complicated in high p_T

What is Flavour Tagging and how does it work?

- Identify jets originating from heavy flavour (*b*, *c*) quarks and separate from other sources (e.g. light quarks)
  - Mainly *b*-tagging
- Using the topology of heavy-flavour jets
- Lifetime of the *b*-hadrons (1.5ps) gives unique properties to the jets
  - Hard fragmentation
  - Displaced secondary and tertiary vertices
  - Large impact parameters (d_0)



Becomes more complicated in high p_T


CMS
CMS Experiment at the LHC, CERN
Data recorded: 2022-Jul-27 18:33:11.804352 GMT
Run / Event / LS: 356309 / 137565256 / 156
CDS: CMS-PHO-EVENTS-2022-033

CMS - DeepCSV Tagger

- Using a Deep Neural Network (DNN)

- Using charged constituents, secondary vertices and global variables of the jet

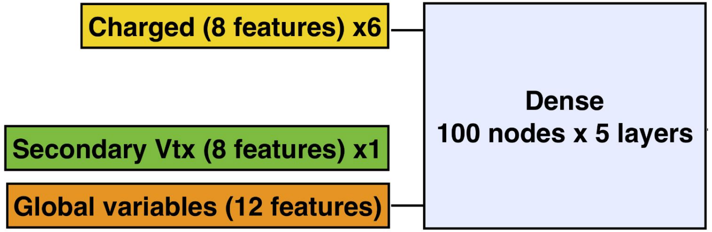
Charged (8 features) x6

Secondary Vtx (8 features) x1

Global variables (12 features)

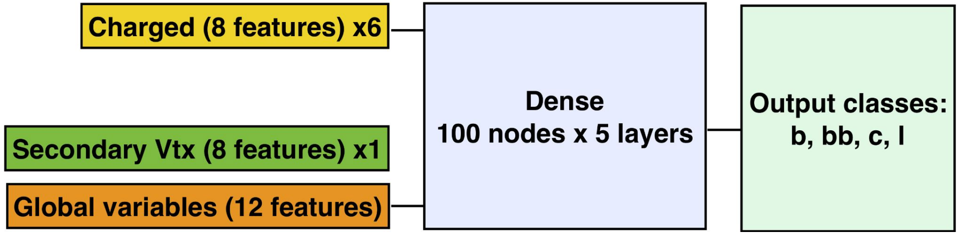
CMS - DeepCSV Tagger

- Using a Deep Neural Network (DNN)
- Using charged constituents, secondary vertices and global variables of the jet
- Using a deep neural network block with 5x 100 node dense layers



CMS - DeepCSV Tagger

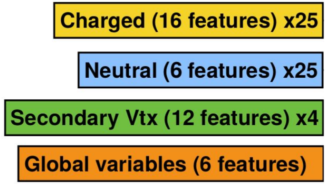
- Using a Deep Neural Network (DNN)
- Using charged constituents, secondary vertices and global variables of the jet
- Using a deep neural network block with 5x 100 node dense layers
- Outputs probabilities for jet originating from a certain source



CMS - DeepJet Tagger

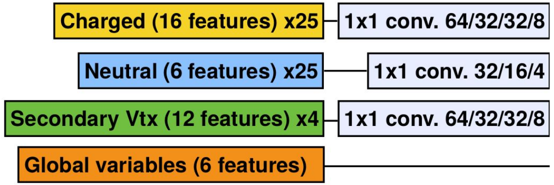
- Using a Deep Neural Network (DNN)

- Using charged and neutral constituents, secondary vertices and global variables of the jet



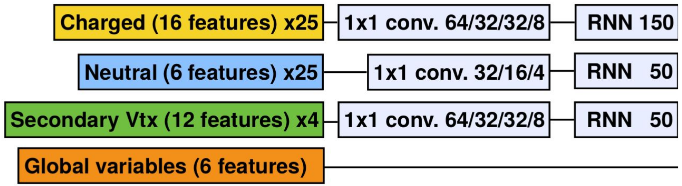
CMS - DeepJet Tagger

- Using a Deep Neural Network (DNN)
- Using charged and neutral constituents, secondary vertices and global variables of the jet
- Charged and neutral constituents and secondary vertices variables are automatic feature engineered using 1x1 convolutional layers (CNNs)



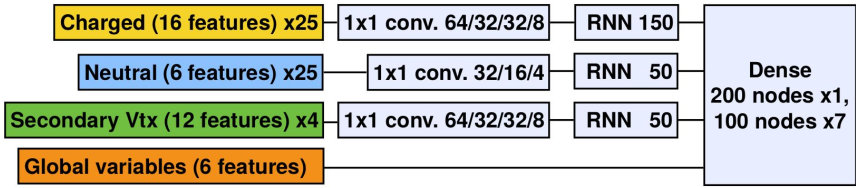
CMS - DeepJet Tagger

- Using a Deep Neural Network (DNN)
- Using charged and neutral constituents, secondary vertices and global variables of the jet
- Charged and neutral constituents and secondary vertices variables are automatic feature engineered using 1x1 convolutional layers (CNNs)
- Using Recurrent Neural Networks (RNNs) to further process the information



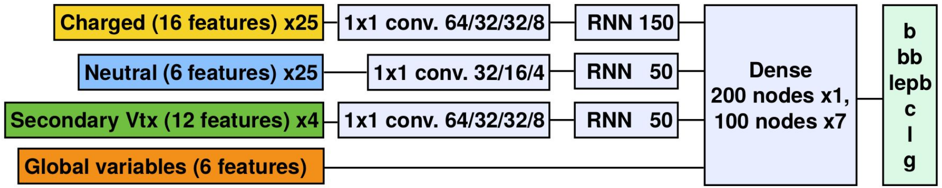
CMS - DeepJet Tagger

- Using a Deep Neural Network (DNN)
- Using charged and neutral constituents, secondary vertices and global variables of the jet
- Charged and neutral constituents and secondary vertices variables are automatic feature engineered using 1x1 convolutional layers (CNNs)
- Using Recurrent Neural Networks (RNNs) to further process the information
- Concatenating RNN outputs and global features and feed it into a multi-classifier DNN



CMS - DeepJet Tagger

- Using a Deep Neural Network (DNN)
- Using charged and neutral constituents, secondary vertices and global variables of the jet
- Charged and neutral constituents and secondary vertices variables are automatic feature engineered using 1x1 convolutional layers (CNNs)
- Using Recurrent Neural Networks (RNNs) to further process the information
- Concatenating RNN outputs and global features and feed it into a multi-classifier DNN
- Outputs probabilities for jet originating from a certain source



Inputs, Outputs & Preprocessing

- Inputs:
  - Jet p_T and Jet η (global)
  - Charged constituent variables (e.g impact parameters)
  - Neutral constituent variables (e.g. p_T^{frac}, ΔR to jet axis)
  - Using up to 25 charged and 25 neutral constituents and 4 secondary vertices

Inputs, Outputs & Preprocessing

- Inputs:
  - Jet p_T and Jet η (global)
  - Charged constituent variables (e.g impact parameters)
  - Neutral constituent variables (e.g. p_T^{frac}, ΔR to jet axis)
  - Using up to 25 charged and 25 neutral constituents and 4 secondary vertices

- Outputs probabilities for jet initiating source:
  - p_b: Hadronically decaying b
  - p_{bb}: Double b
  - p_{lepb}: Leptonically decaying b
  - p_c: c-quark
  - p_l: Light-flavour quark (u-, d-, s-quark)
  - p_g: Gluon

Inputs, Outputs & Preprocessing

- Inputs:
  - Jet p_T and Jet η (global)
  - Charged constituent variables (e.g impact parameters)
  - Neutral constituent variables (e.g. p_T^{frac}, ΔR to jet axis)
  - Using up to 25 charged and 25 neutral constituents and 4 secondary vertices

- Outputs probabilities for jet initiating source:
  - p_b: Hadronically decaying b
  - p_{bb}: Double b
  - p_{lepb}: Leptonically decaying b
  - p_c: c-quark
  - p_l: Light-flavour quark (u-, d-, s-quark)
  - p_g: Gluon

- Preprocessing:
  - Using different samples for training
    - Jets from $t\bar{t}$ events (both hadronically decaying)
    - Jets from QCD events

Inputs, Outputs & Preprocessing

- Inputs:
  - Jet p_T and Jet η (global)
  - Charged constituent variables (e.g impact parameters)
  - Neutral constituent variables (e.g. p_T^{frac}, ΔR to jet axis)
  - Using up to 25 charged and 25 neutral constituents and 4 secondary vertices

- Outputs probabilities for jet initiating source:
  - p_b: Hadronically decaying *b*
  - p_{bb}: Double *b*
  - p_{lepb}: Leptonically decaying *b*
  - p_c: *c*-quark
  - p_l: Light-flavour quark (*u*-, *d*-, *s*-quark)
  - p_g: Gluon

- Preprocessing:
  - Using different samples for training
    - Jets from $t\bar{t}$ events (both hadronically decaying)
    - Jets from QCD events
  - Downsample each flavour in jet p_T and η to match *b*-jet distributions
  - Ensure kinematic independence
  - Apply scaling and shifting

Inputs, Outputs & Preprocessing

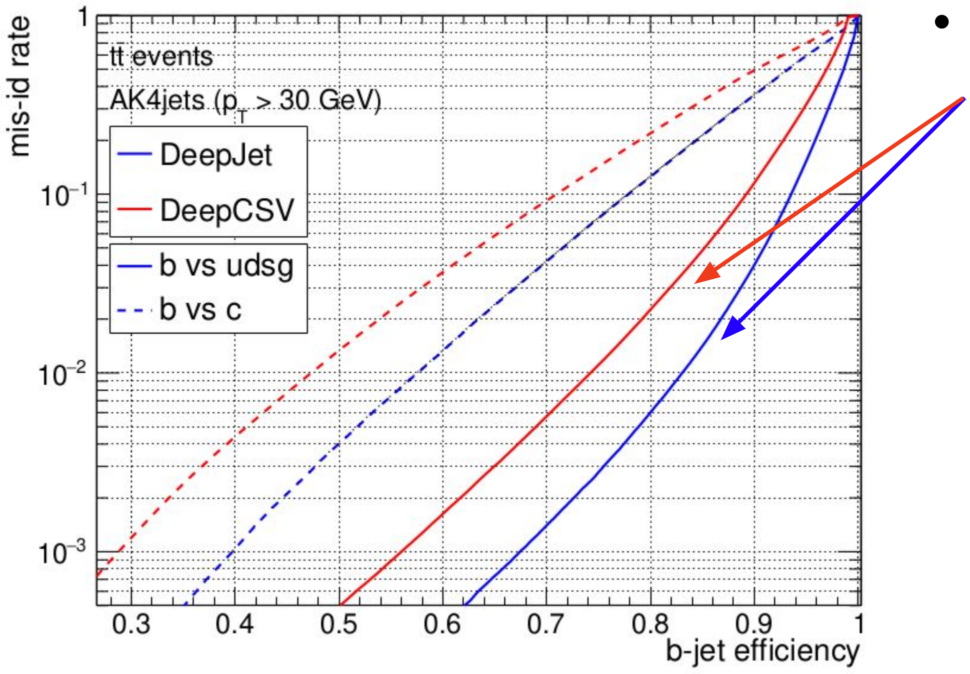
- Inputs:
  - Jet p_T and Jet η (global)
  - Charged constituent variables (e.g impact parameters)
  - Neutral constituent variables (e.g. p_T^{frac}, ΔR to jet axis)
  - Using up to 25 charged and 25 neutral constituents and 4 secondary vertices

- Outputs probabilities for jet initiating source:
  - p_b: Hadronically decaying *b*
  - p_{bb}: Double *b*
  - p_{lepb}: Leptonically decaying *b*
  - p_c: *c*-quark
  - p_l: Light-flavour quark (*u*-, *d*-, *s*-quark)
  - p_g: Gluon

- Preprocessing:
  - Using different samples for training
    - Jets from $t\bar{t}$ events (both hadronically decaying)
    - Jets from QCD events
  - Downsample each flavour in jet p_T and η to match *b*-jet distributions
  - Ensure kinematic independence
  - Apply scaling and shifting

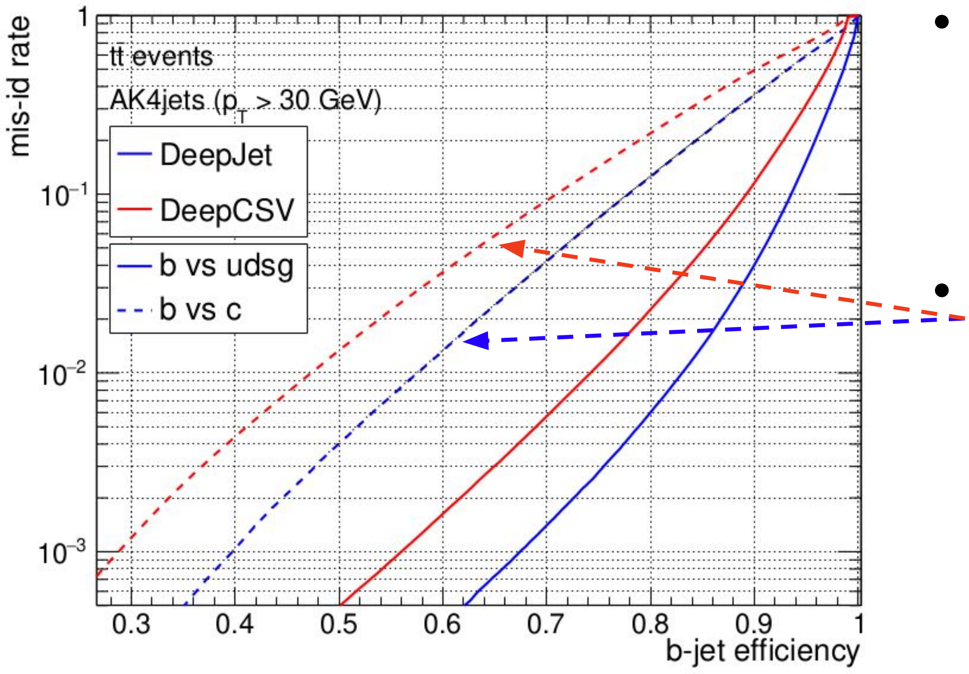
- Resulting in 130M jets which are splitted into training, validation and testing

DeepJet - Results



- Large performance gains for DeepJet over current default DeepCSV for light- and gluon jet rejection

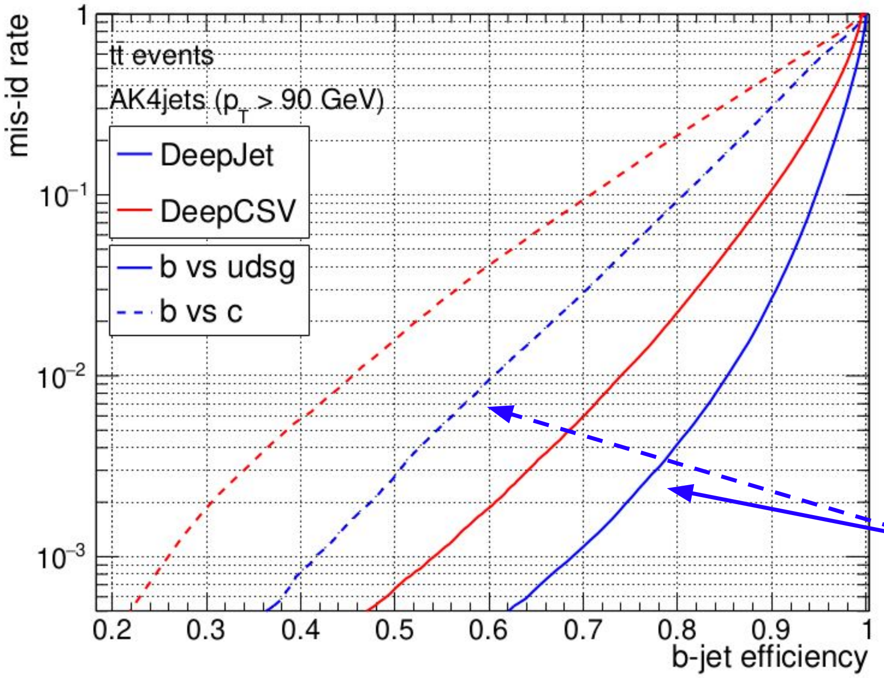
DeepJet - Results



- Large performance gains for DeepJet over current default DeepCSV for light- and gluon jet rejection

- Also: Large performance gains for *c*-jet rejection!

DeepJet - Results



- Large performance gains for DeepJet over current default DeepCSV for light- and gluon jet rejection
- Also: Large performance gains for *c*-jet rejection!
- Performance gains in higher p_T regions also significant

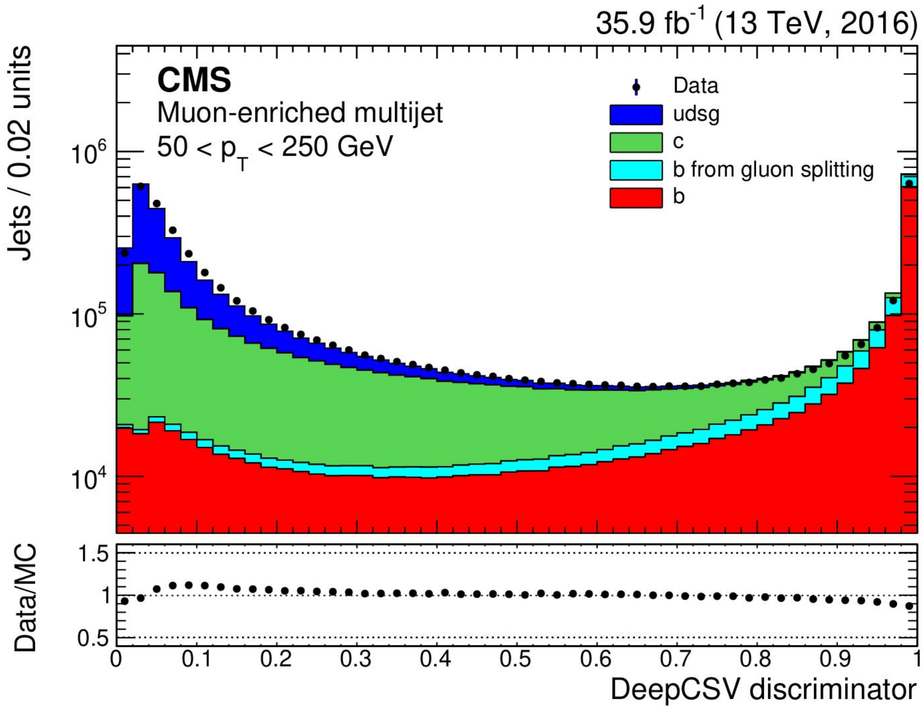
DeepJet - Calibration

- As known, no Monte-Carlo (MC) generator is perfect!

- Need to calibrate the trained taggers also to data!

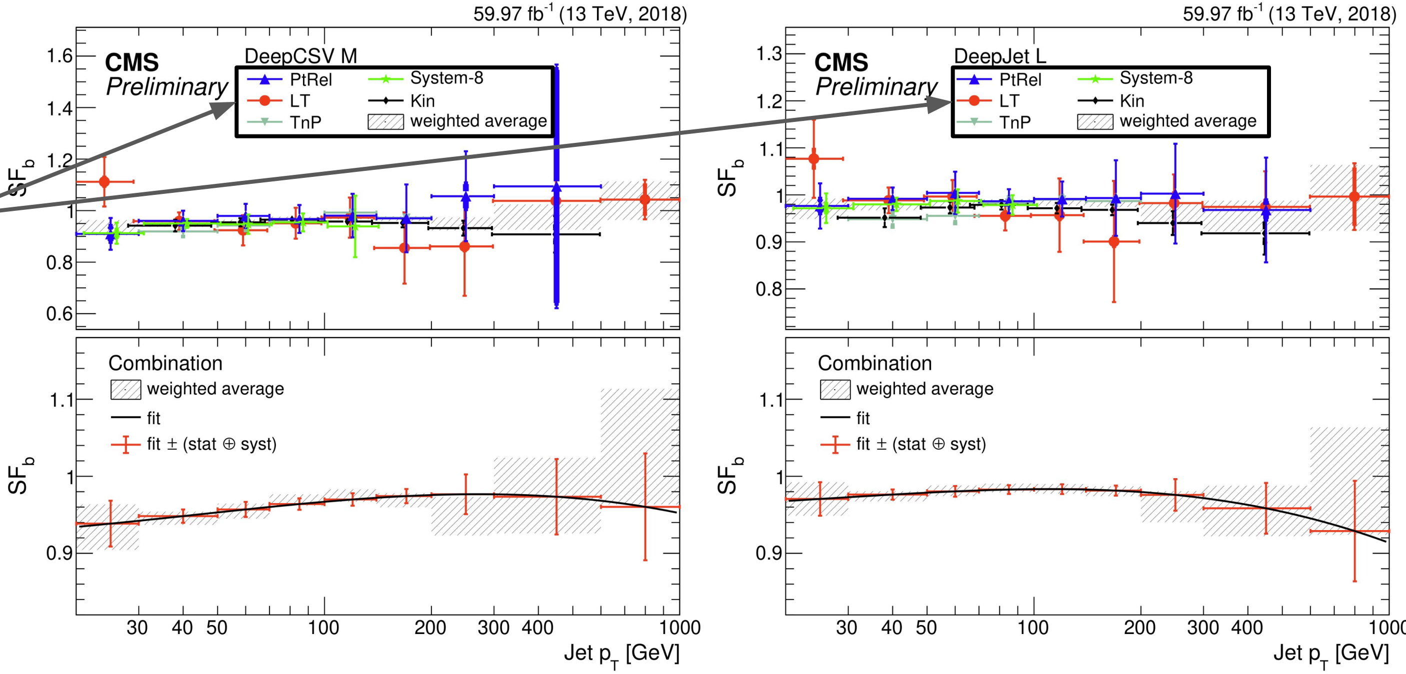
DeepJet - Calibration

- As known, no Monte-Carlo (MC) generator is perfect!
- Need to calibrate the trained taggers also to data!
- Check performance on data and apply scale factors (SFs) to MC to correct for disagreement
- Using different methods to distinguish scale factors



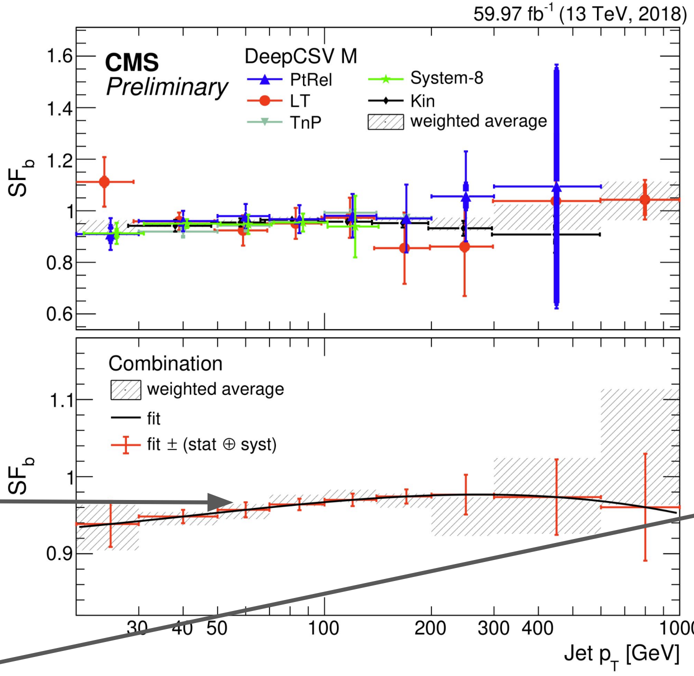
DeepJet - Calibration

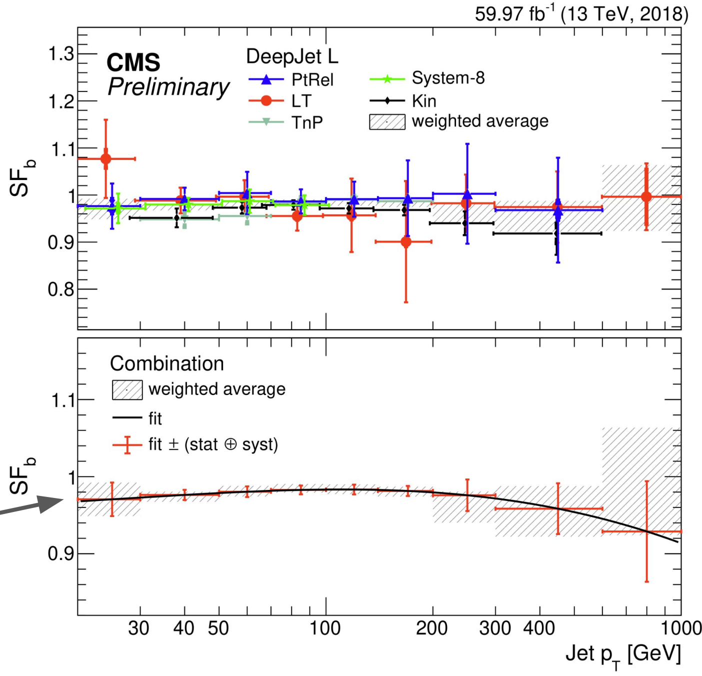
- Different calibration methods used to distinguish *b*-tagging efficiency
- Methods use different selections and techniques

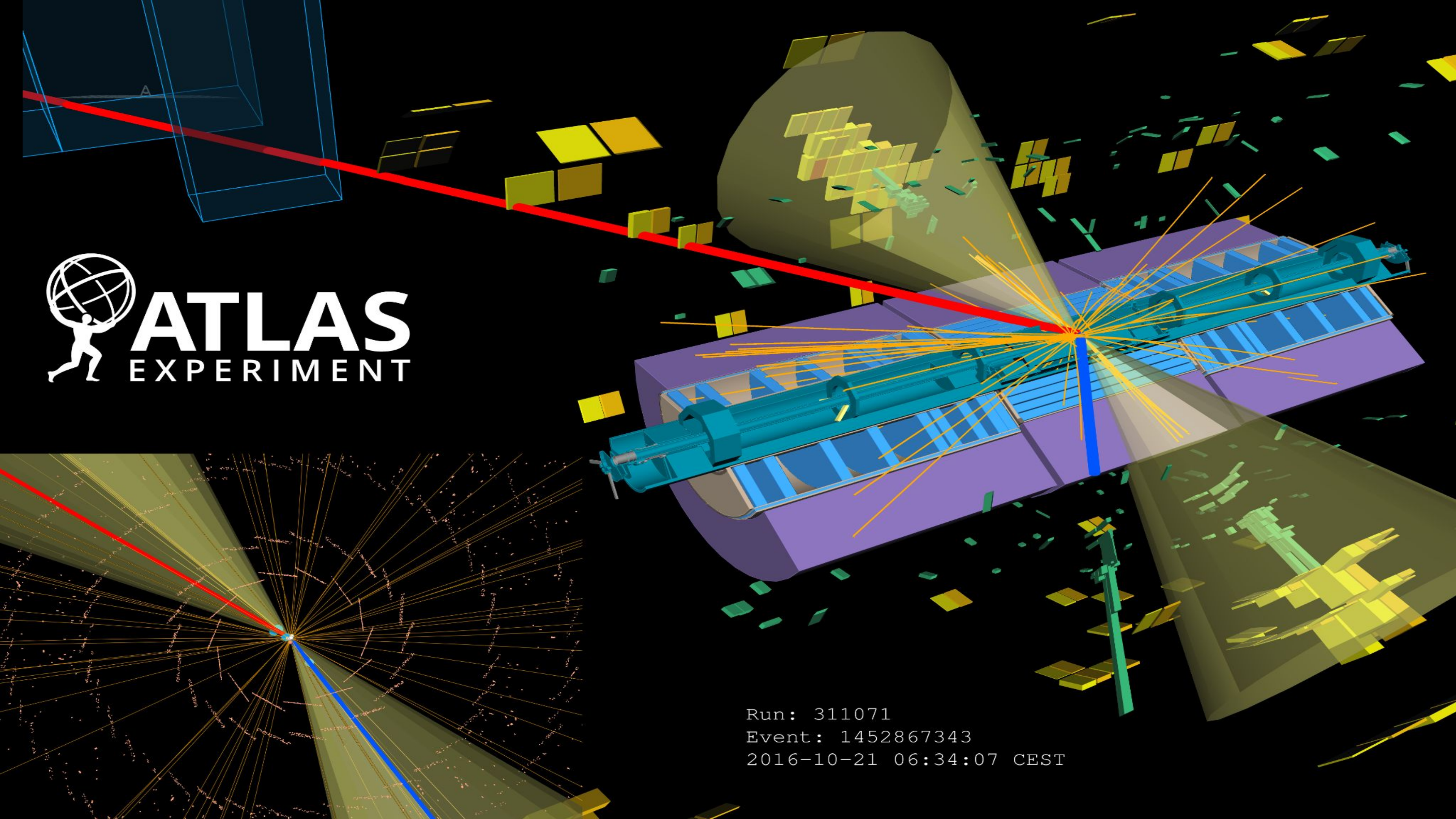


DeepJet - Calibration

- Different calibration methods used to distinguish *b*-tagging efficiency
- Methods use different selections and techniques
- SFs close to/slightly under one
- SFs of DeepJet are closer to 1 with smaller uncertainties!

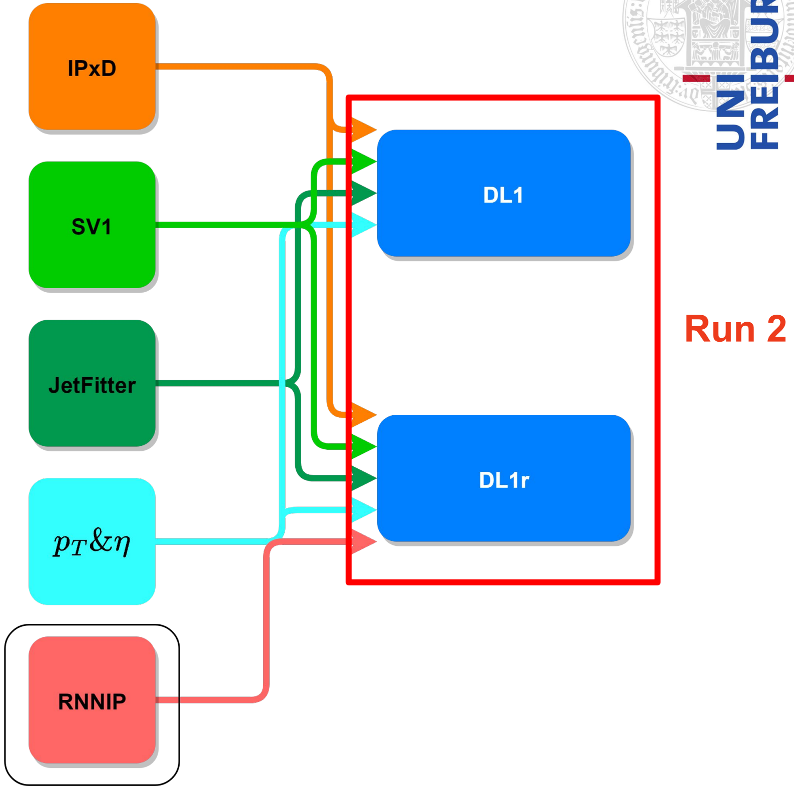





ATLAS
EXPERIMENT
Run: 311071
Event: 1452867343
2016-10-21 06:34:07 CEST

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

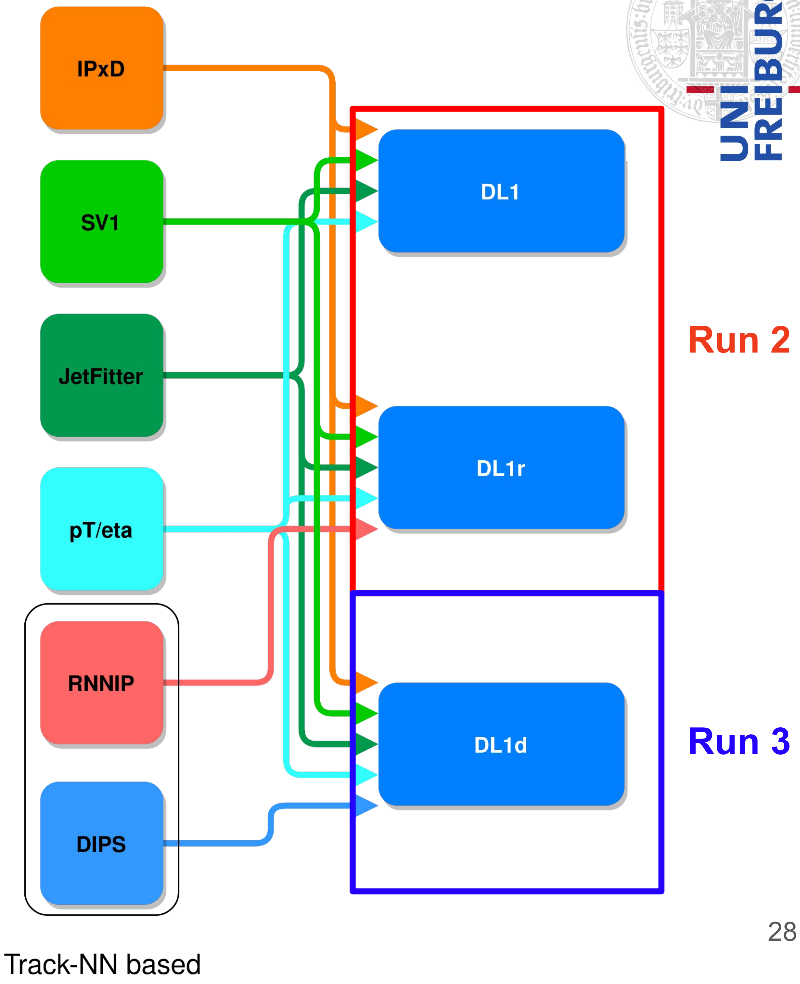


Track-based Neural Network

ATLAS High-Level *b*-Tagging Algorithms

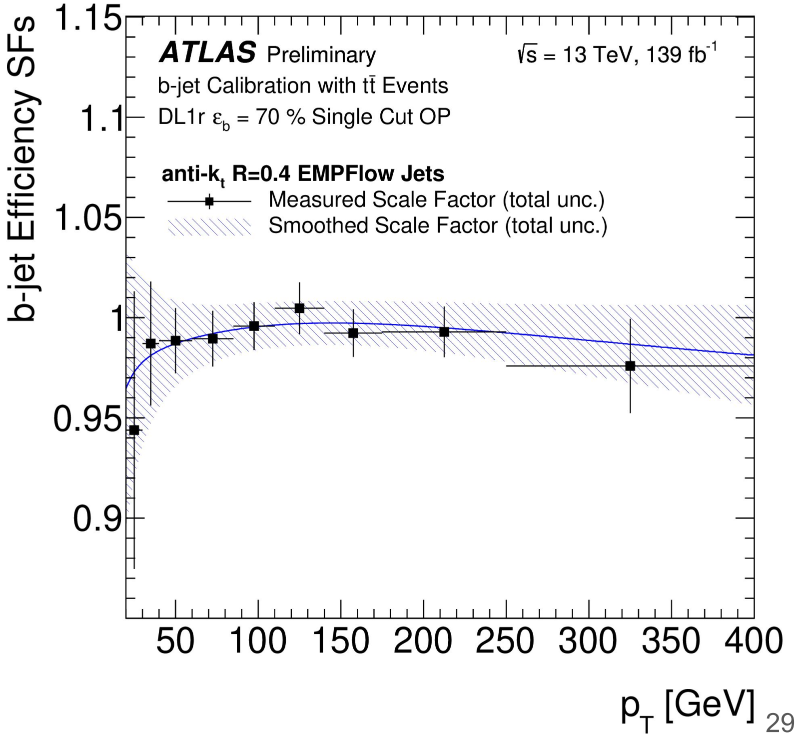
- Default tagger in Run 2 was DL1r (ATL-PHYS-PUB-2017-013)
- Uses jet-level variables and many different low-level algorithms (i.e. IPxD, SV1, JetFitter)
- For track information, DL1r uses the Recurrent Neural Network Impact Parameter (RNNIP) tagger
- Many improvements were implemented for Run 3
- RNNIP 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 pre-recommended high level tagger for Run 3</u>



Where are We?

- Tagger development for DL1d is done!
- Expecting similar calibration results as for DL1r
- Full Run 2 *b*-efficiency calibration for DL1r
- Very good agreement with unity, although we see some differences which we take into account
- Calibration for DL1d ongoing!

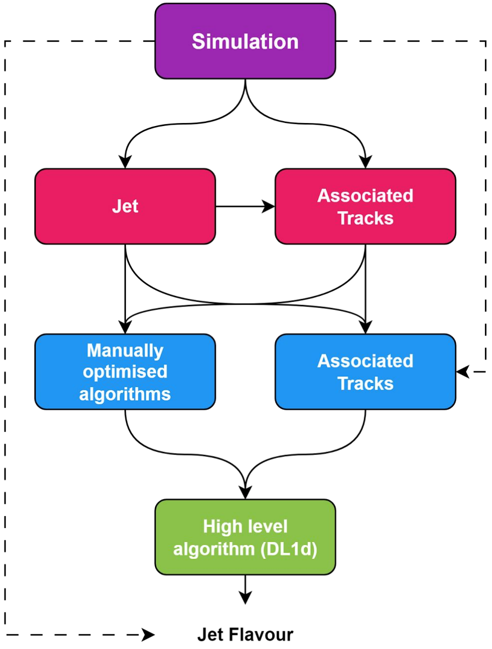


A New Hope .. Approach for ATLAS - GN1/GN2

- Previous taggers used two-stage approach

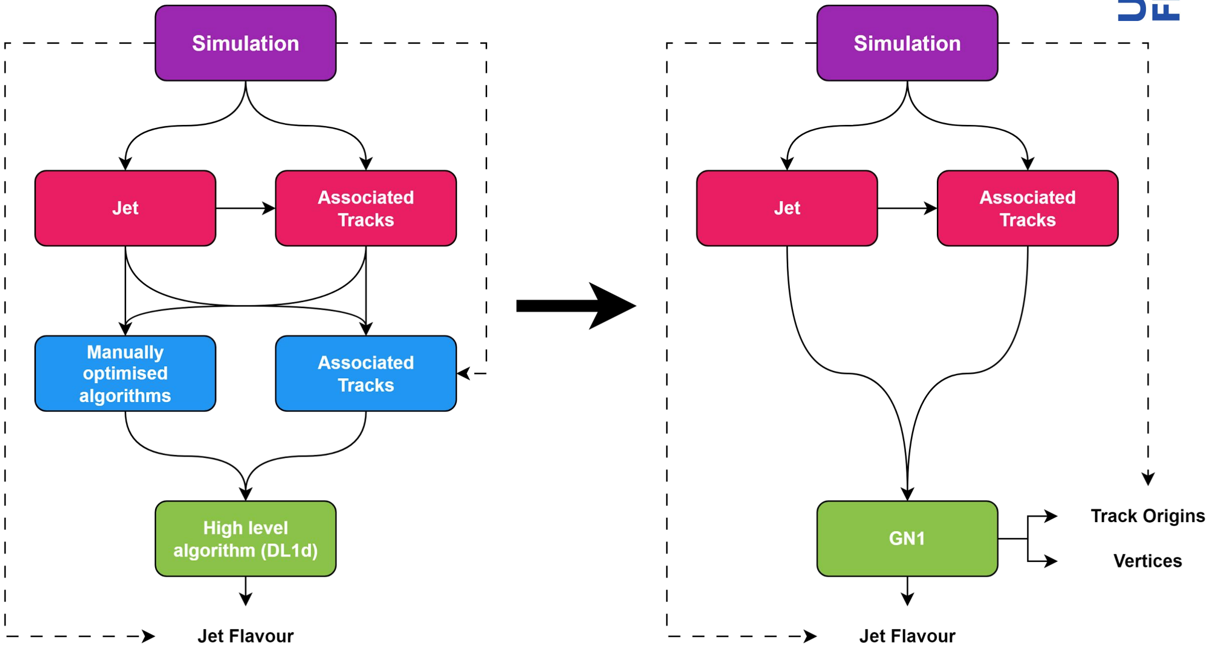
- Manually optimised algorithms → Low level

- Final neural network which uses low level algorithms as input → High level



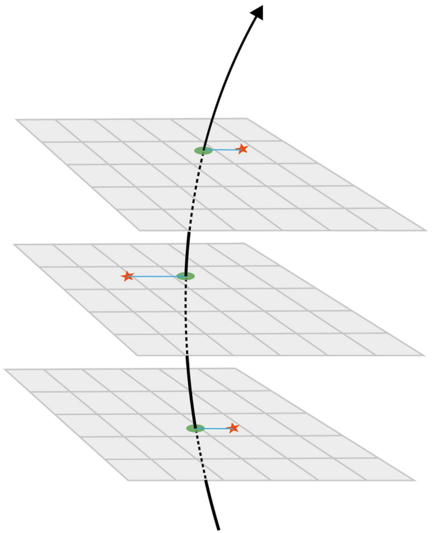
A New Hope .. Approach for ATLAS - GN1/GN2

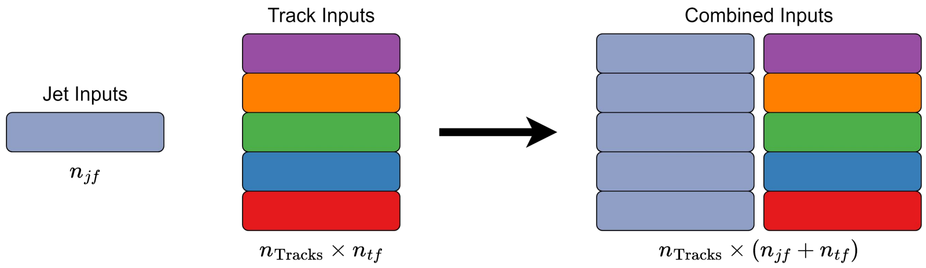
- Previous taggers used two-stage approach
- Manually optimised algorithms → Low level
- Final neural network which uses low level algorithms as input → High level
- New tagger (GN1/GN2) uses-one stage approach
- Easier to handle → Less manual optimisation!
- Improved performance



Inputs, Outputs & Preprocessing

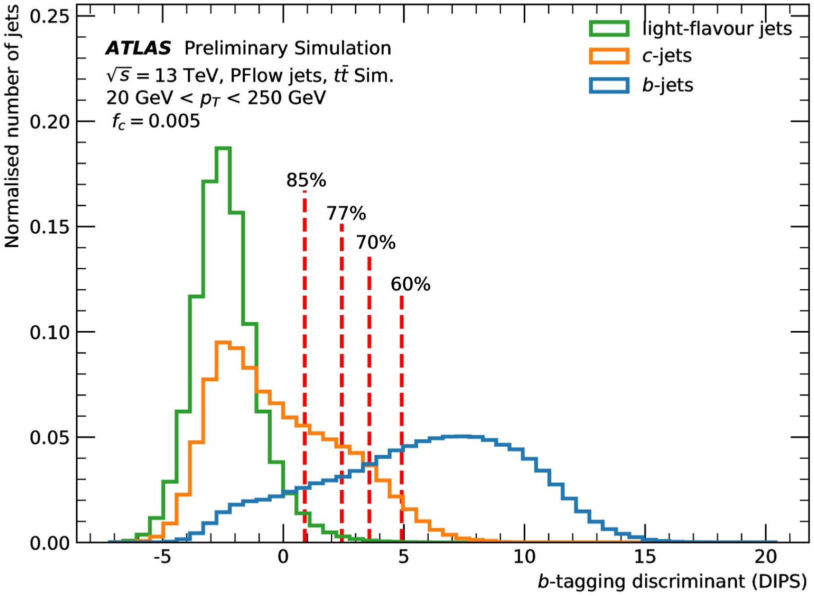
- Inputs:
  - Jet p_T and Jet η
  - Track variables (e.g Parameters, uncertainties)
  - Hit information (e.g. Number of pixel hits)
  - Concatenate jet and track variables
  - Using up to 40 tracks per jet (charged constituents)





Inputs, Outputs & Preprocessing

- Inputs:
  - Jet p_T and Jet η
  - Track variables (e.g Parameters, uncertainties)
  - Hit information (e.g. Number of pixel hits)
  - Concatenate jet and track variables
  - Using up to 40 tracks per jet (charged constituents)



- Outputs:
  - p_b, p_c and p_u
  - Probability of jet originating from a *b*-, *c*- or light quark
  - Discriminant using probabilities: D_b and D_c

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

Inputs, Outputs & Preprocessing

- Inputs:
  - Jet p_T and Jet η
  - Track variables (e.g Parameters, uncertainties)
  - Hit information (e.g. Number of pixel hits)
  - Concatenate jet and track variables
  - Using up to 40 tracks per jet
    (charged constituents)

- Outputs:
  - p_b, p_c and p_u
  - Probability of jet originating from a *b*-, *c*- or light quark
  - Discriminant using probabilities: D_b and D_c

- Preprocessing:
  - Using different samples for training
    - Low p_T jets from $t\bar{t}$
      (non-all hadronic decays only)
    - High p_T jets from Z'
      ($Z' \to q\bar{q}$)

Inputs, Outputs & Preprocessing

- Inputs:
  - Jet p_T and Jet η
  - Track variables (e.g Parameters, uncertainties)
  - Hit information (e.g. Number of pixel hits)
  - Concatenate jet and track variables
  - Using up to 40 tracks per jet (charged constituents)

- Outputs:
  - p_b, p_c and p_u
  - Probability of jet originating from a *b*-, *c*- or light quark
  - Discriminant using probabilities: D_b and D_c

- Preprocessing:
  - Using different samples for training
    - Low p_T jets from $t\bar{t}$ (non-all hadronic decays only)
    - High p_T jets from Z' ($Z' \rightarrow q\bar{q}$)
  - Hybrid resampling each flavour in jet p_T and η to match *b*-jet distributions
  - Ensure kinematic independence
  - Apply scaling and shifting

Inputs, Outputs & Preprocessing

- Inputs:
  - Jet p_T and Jet η
  - Track variables (e.g Parameters, uncertainties)
  - Hit information (e.g. Number of pixel hits)
  - Concatenate jet and track variables
  - Using up to 40 tracks per jet (charged constituents)

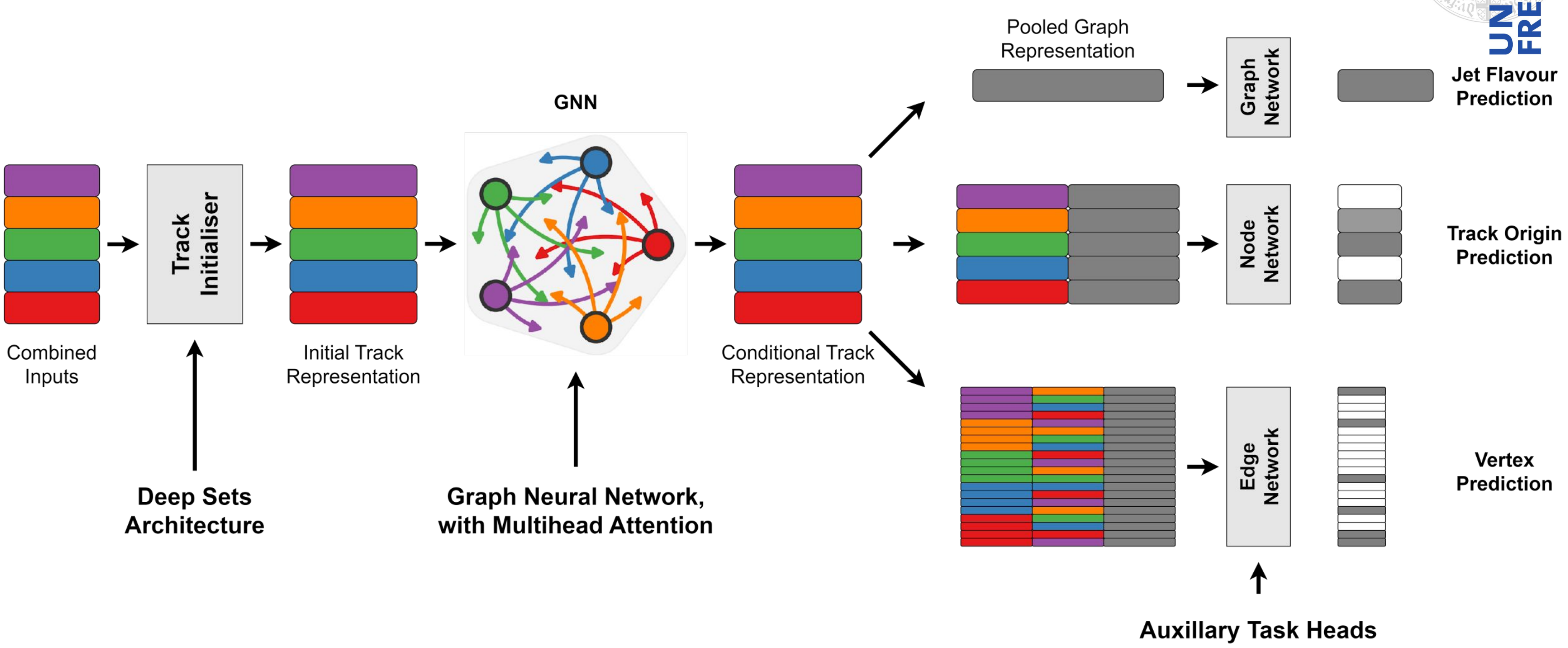
- Outputs:
  - p_b, p_c and p_u
  - Probability of jet originating from a *b*-, *c*- or light quark
  - Discriminant using probabilities: D_b and D_c

- Preprocessing:
  - Using different samples for training
    - Low p_T jets from $t\bar{t}$ (non-all hadronic decays only)
    - High p_T jets from Z' ($Z' \rightarrow q\bar{q}$)
  - Hybrid resampling each flavour in jet p_T and η to match *b*-jet distributions
  - Ensure kinematic independence
  - Apply scaling and shifting

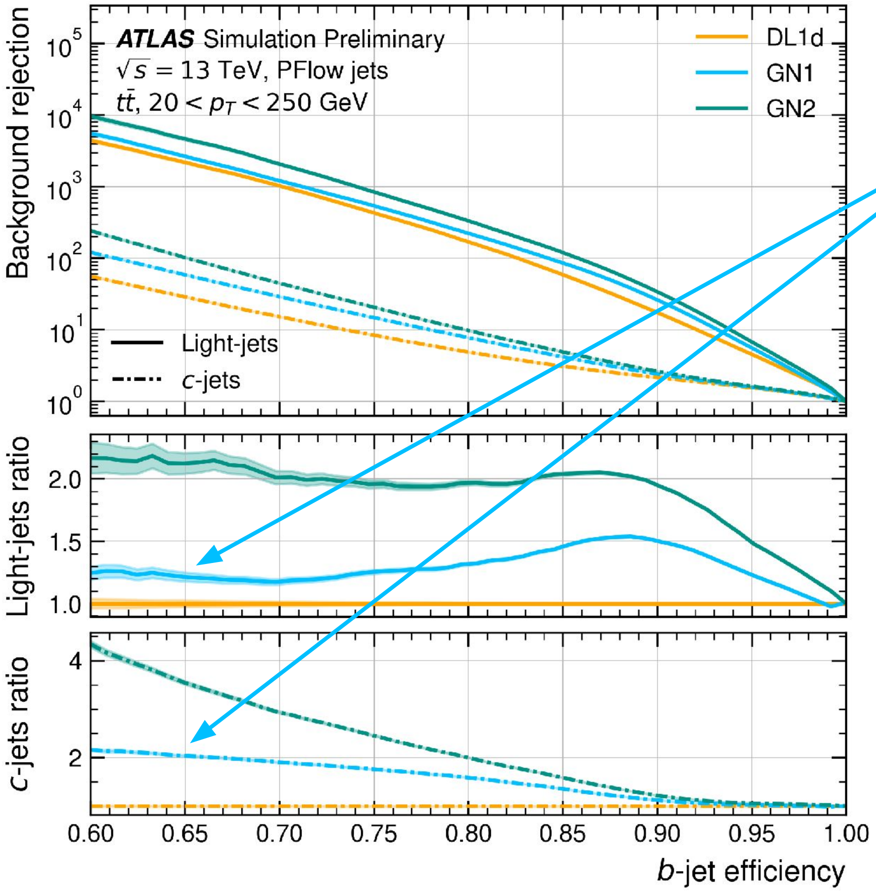
- Resulting in **30M** training jets, 500k each for validation and testing for GN1

- Resulting in **192M** training jets, 500k each for validation and testing for GN2

GN1/GN2 - Architecture

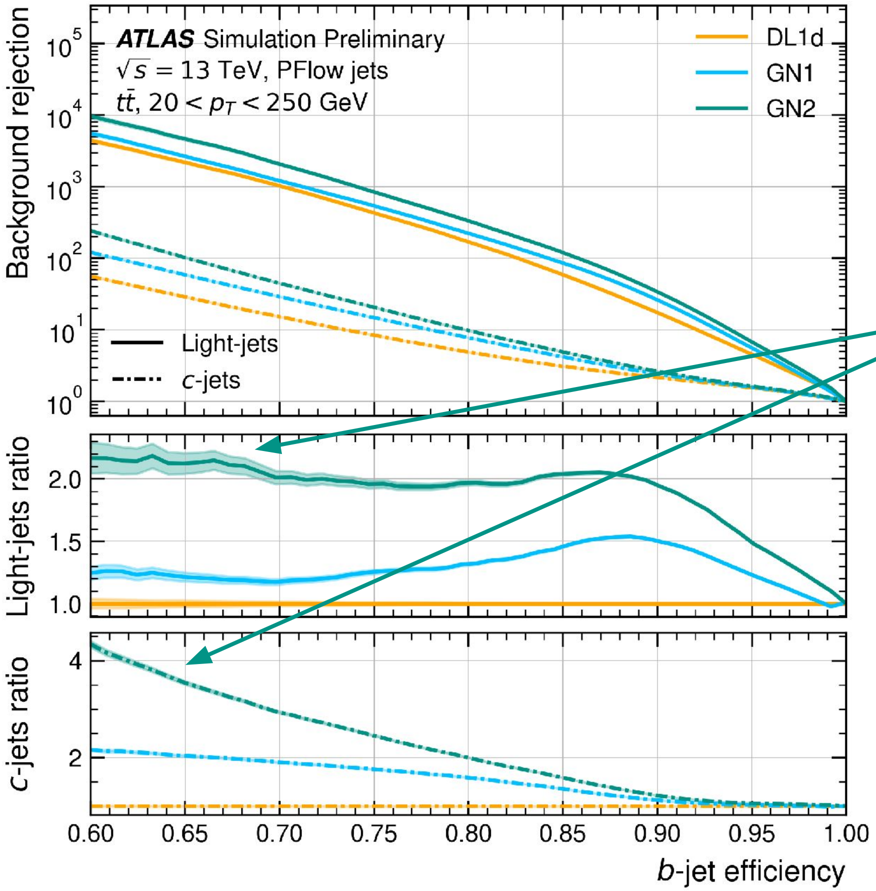


GN1/GN2 - Results



- Large performance gains on low p_T jets in background rejection for GN1 over current ATLAS default tagger (DL1d)

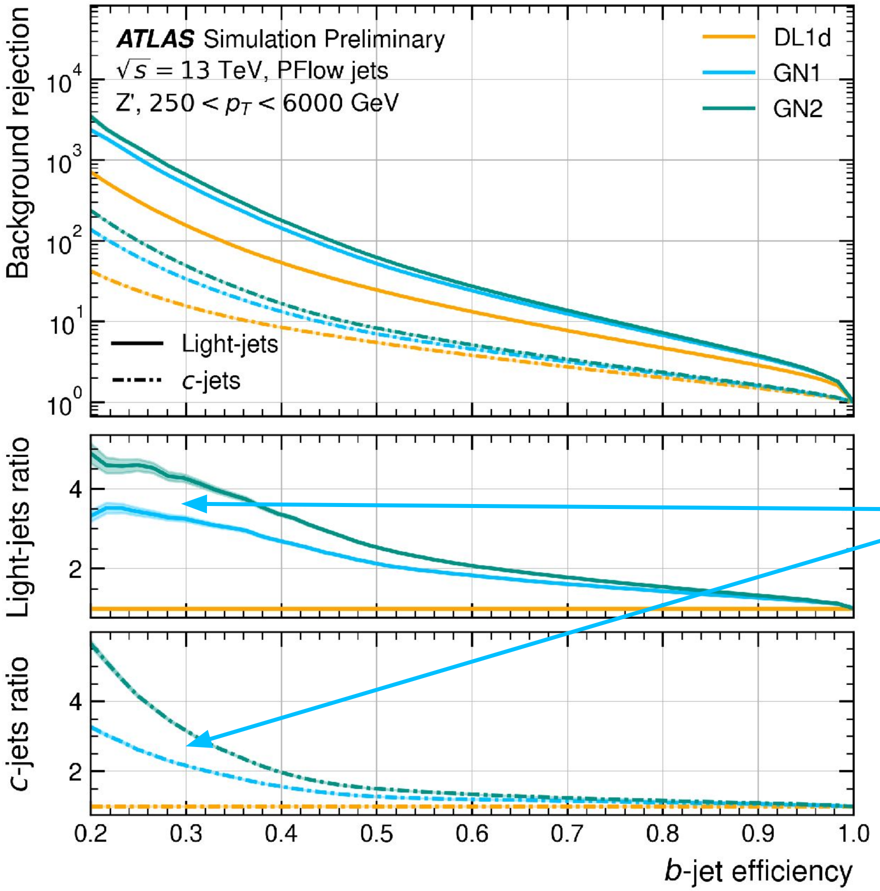
GN1/GN2 - Results



- Large performance gains on low p_T jets in background rejection for GN1 over current ATLAS default tagger (DL1d)

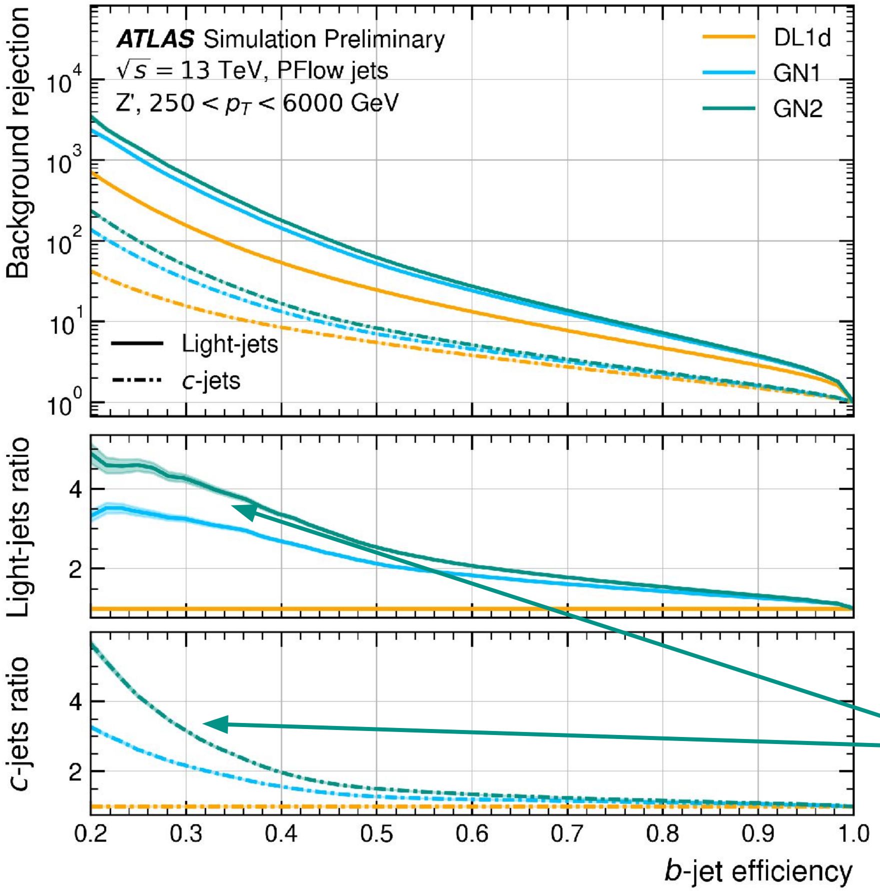
- Even larger performance gains for new GN2 version over both GN1 and DL1d

GN1/GN2 - Results



- Large performance gains on low p_T jets in background rejection for GN1 over current ATLAS default tagger (DL1d)
- Even larger performance gains for new GN2 version over both GN1 and DL1d
- Also:
  Large performance gain on high p_T jets for GN1 over DL1d

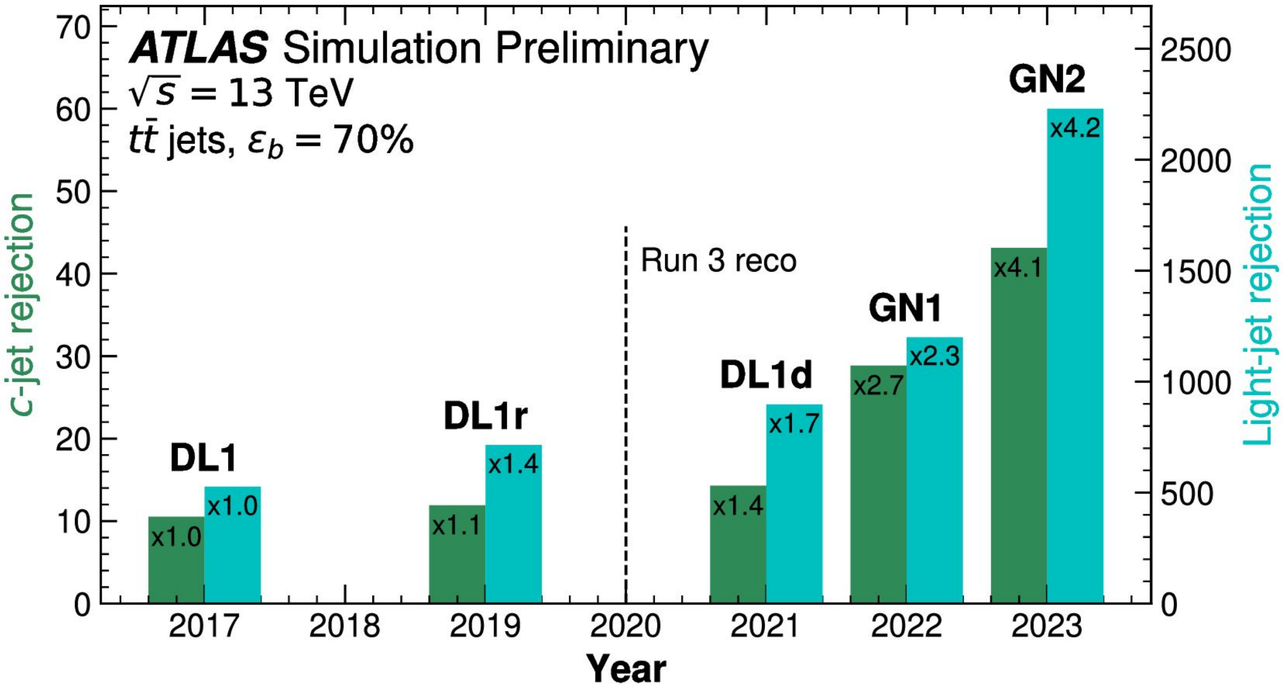
GN1/GN2 - Results



- Large performance gains on low p_T jets in background rejection for GN1 over current ATLAS default tagger (DL1d)
- Even larger performance gains for new GN2 version over both GN1 and DL1d
- Also:
  Large performance gain on high p_T jets for GN1 over DL1d
- Similar:
  GN2 performs even better than GN1

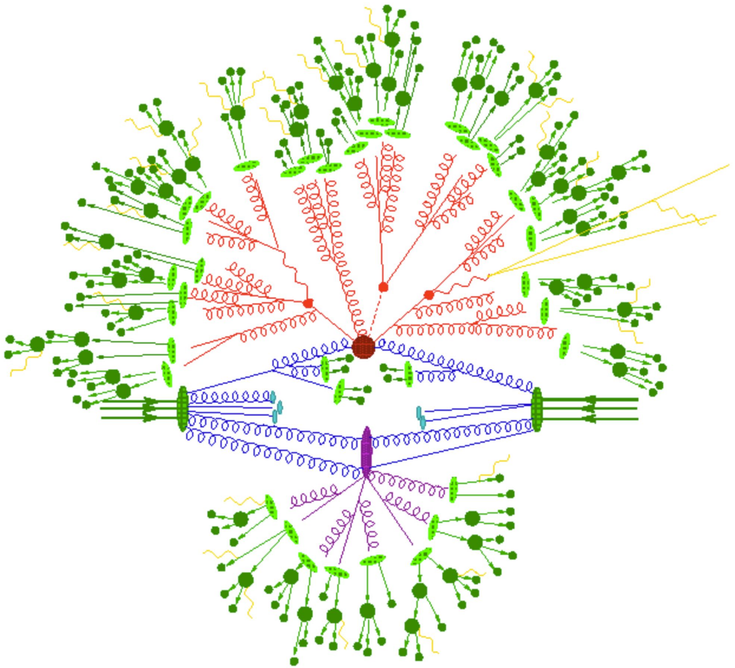
GN1/GN2 - Results

x4 improvement w.r.t the first DL1!



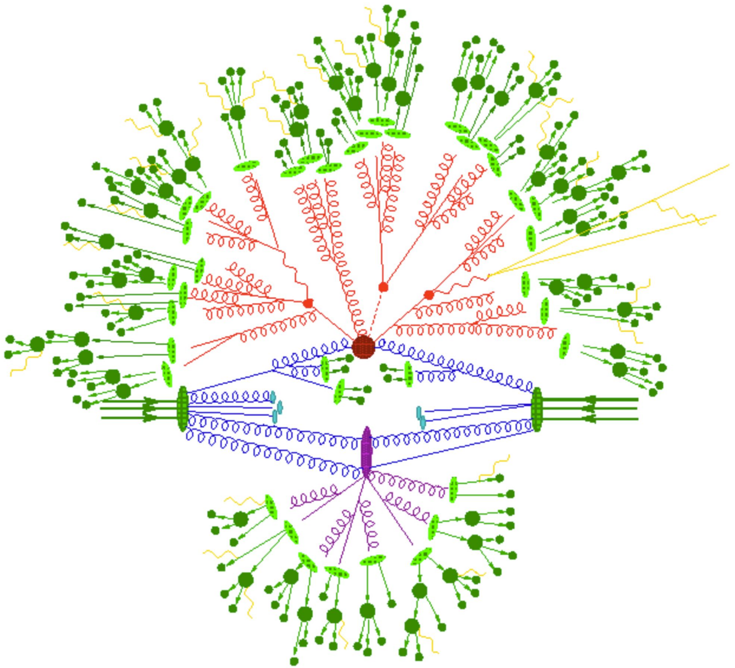
DL1d/GN1 - Generator Dependency

- Models are trained with samples from certain generators
  - Low p_T jets from $t\bar{t}$: PowHEG + Pythia8 + EVTGen
  - High p_T jets from Z': Pythia8 + EVTGen

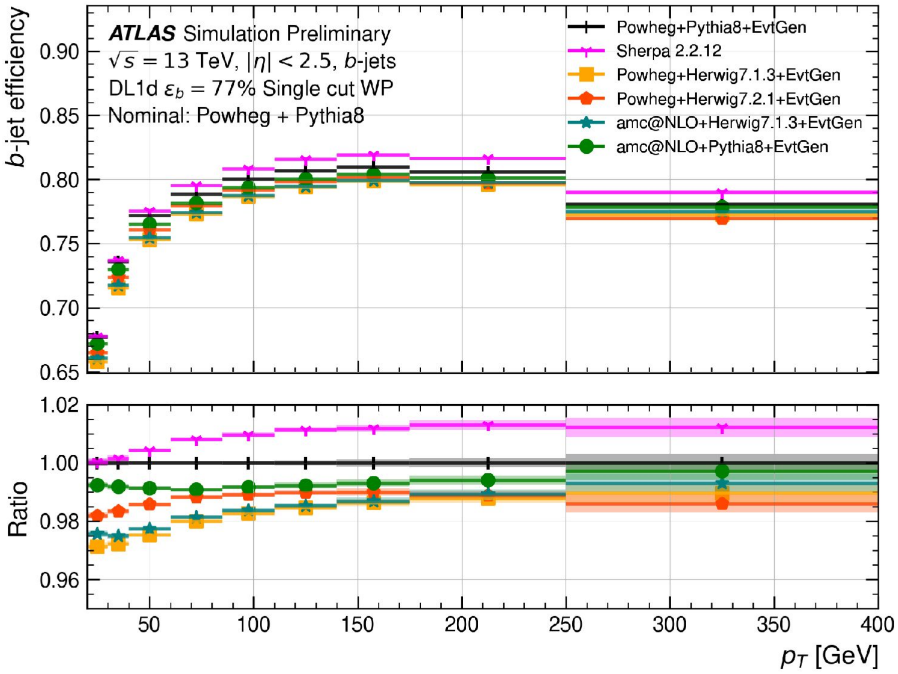


DL1d/GN1 - Generator Dependency

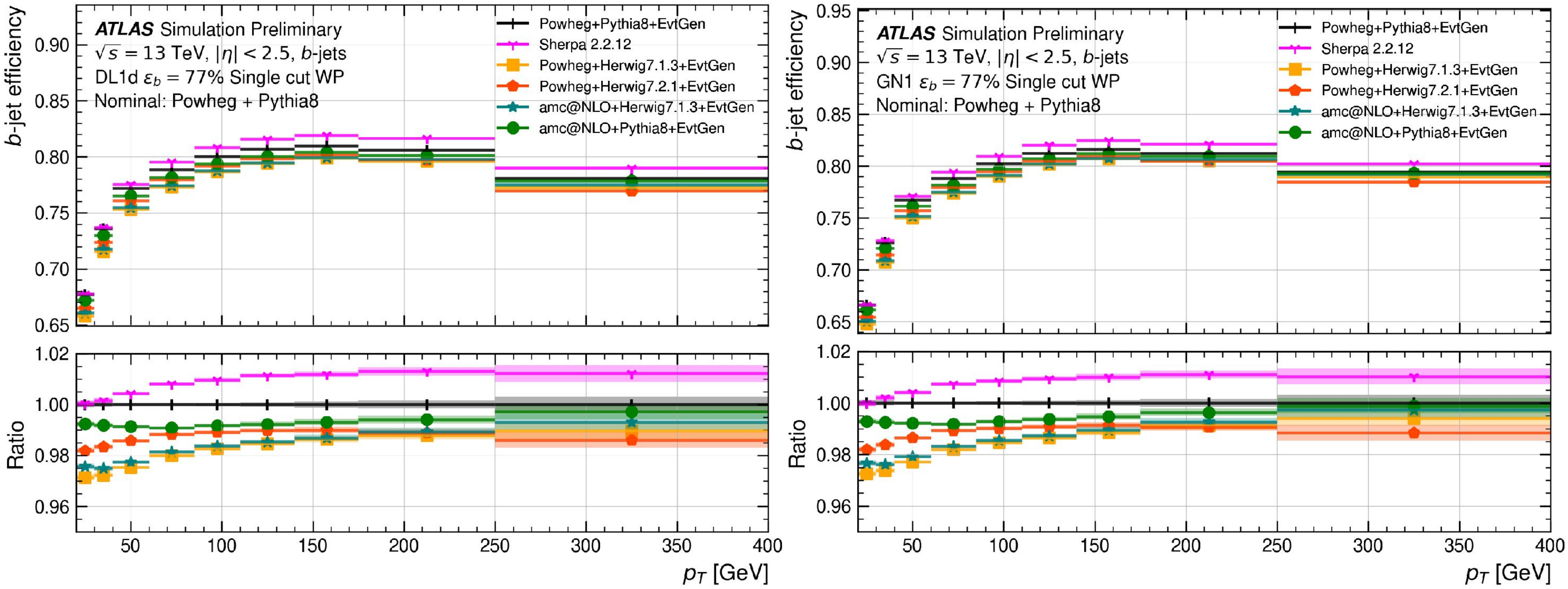
- Models are trained with samples from certain generators
  - Low p_T jets from $t\bar{t}$: PowHEG + Pythia8 + EVTGen
  - High p_T jets from Z': Pythia8 + EVTGen
- All generators have some caveats which might be exploited by different algorithms
- Need to check for these using samples from different generators
- Models should **NOT** show high dependency on generator



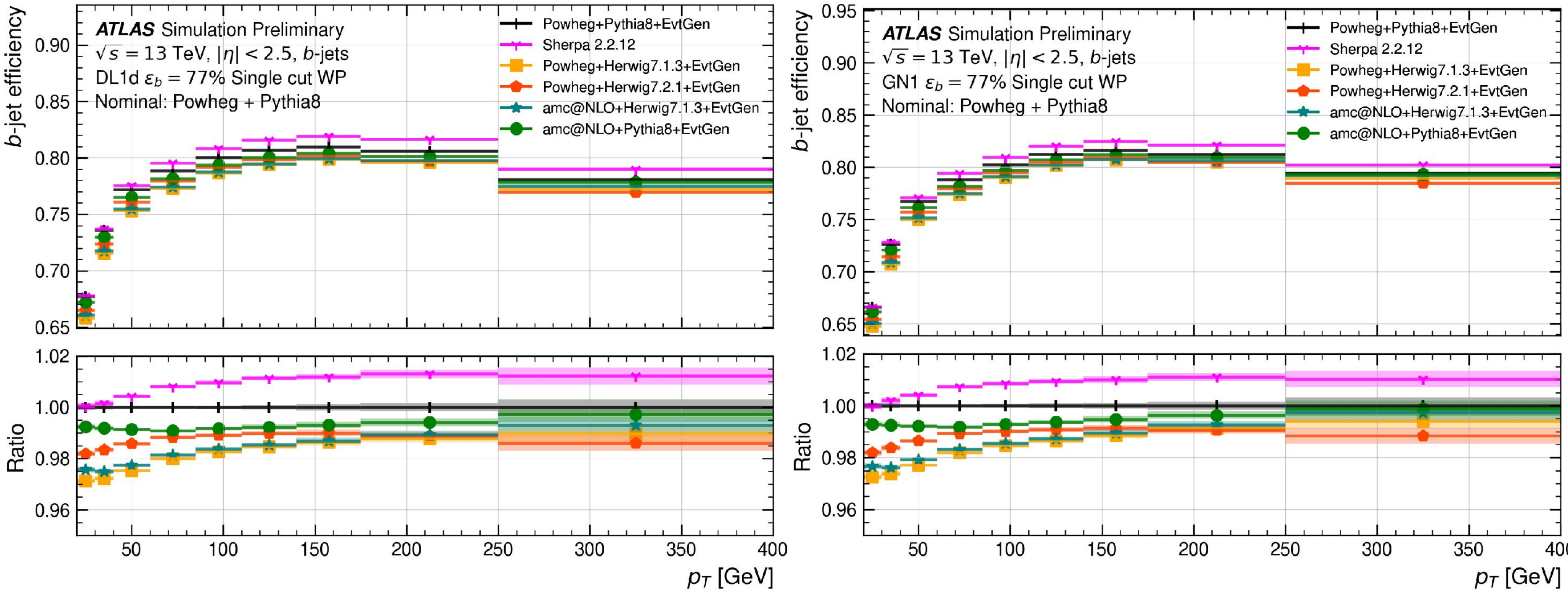
DL1d/GN1 - Generator Dependency



DL1d/GN1 - Generator Dependency



DL1d/GN1 - Generator Dependency



- Overall good agreement between generators for both DL1d and GN1
- Indicating that both models are not exploiting generator specific information

DL1d/GN1 - Data/MC Agreement

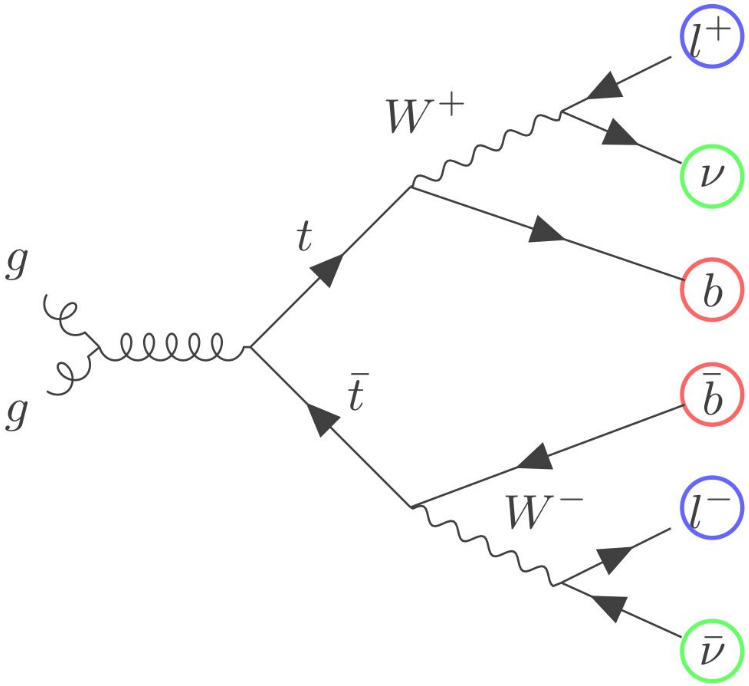
- As already known, generators are not perfect!

- Need to check performance on data

- Derive efficiencies for the different flavours on data and correct MC via scale factors

DL1d/GN1 - Data/MC Agreement

- As already known, generators are not perfect!
- Need to check performance on data
- Derive efficiencies for the different flavours on data and correct MC via scale factors
- Using a variety of different, easy to select, processes to calibrate the taggers (mainly $t\bar{t}$ processes)
  - e.g. dilepton $t\bar{t}$ events
- After full calibration, taggers can be used in analysis!



DL1d/GN1 - Data/MC Agreement

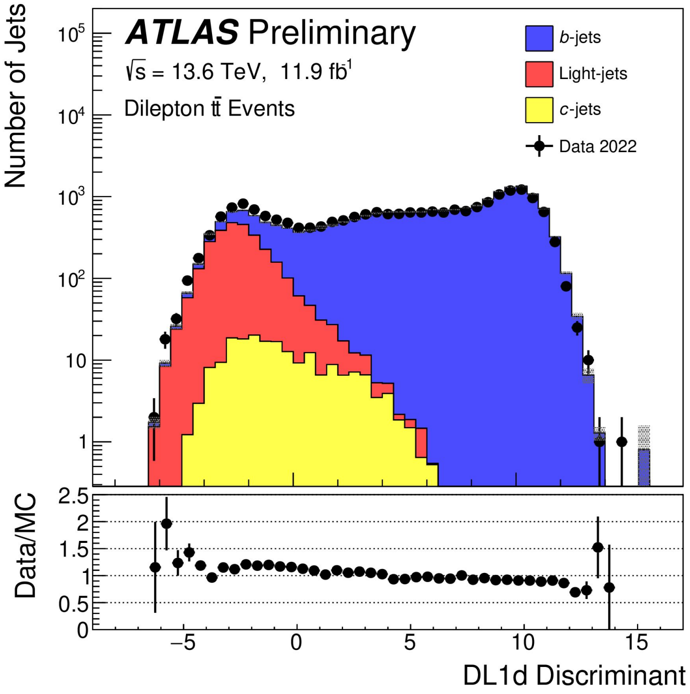
Event selection:

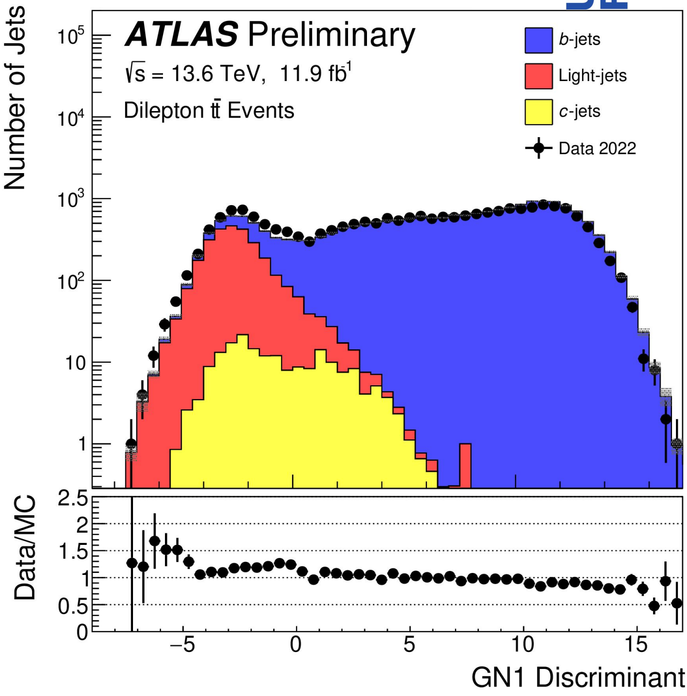
- Exactly two leptons and two jets
- Opposite sign muon and electron
- Invariant mass of each jet-lepton pair below 175 GeV
- Plotting tagger discriminant for leading jet p_T

DL1d/GN1 - Data/MC Agreement

Event selection:

- Exactly two leptons and two jets
- Opposite sign muon and electron
- Invariant mass of each jet-lepton pair below 175 GeV
- Plotting tagger discriminant for leading jet p_T

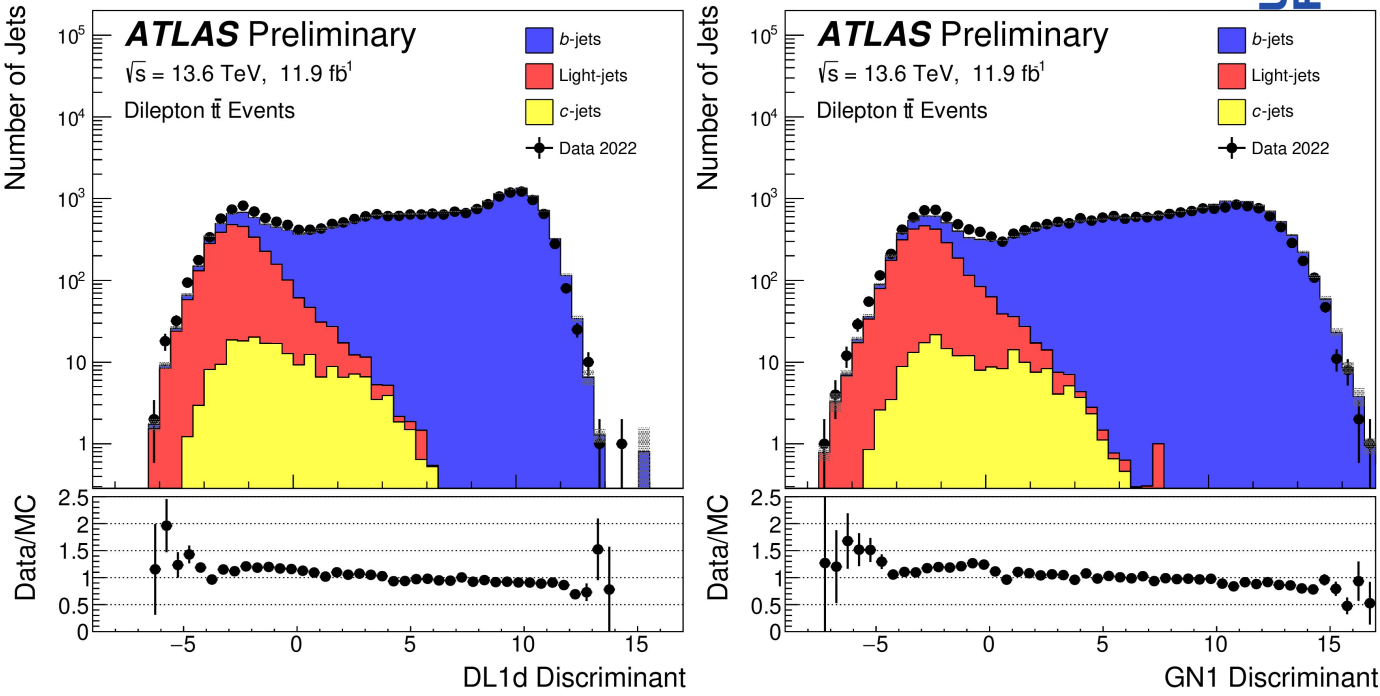




DL1d/GN1 - Data/MC Agreement

Event selection:

- Exactly two leptons and two jets
- Opposite sign muon and electron
- Invariant mass of each jet-lepton pair below 175 GeV
- Plotting tagger discriminant for leading jet p_T



Very good Data/MC agreement for both DL1d and GN1!
Calibration for GN1 is ongoing!

Summary

- CMS:
    - New DeepJet tagger is a new network capable of identifying different jet flavours
    - Using charged and neutral constituents as well as secondary vertices and global variables
    - Trained with 130M jets (from hadronically decaying $t\bar{t}$ pairs and QCD mutlijet events)
    - DeepJet is a combination of different layer architectures like CNNs, RNNs and DNNs
    - Significant performance boost for *c*- and light-flavour rejection for DeepJet over DeepCSV
    - Strong benefit for CMS physics programme!

Summary

- CMS:
  - New DeepJet tagger is a new network capable of identifying different jet flavours
  - Using charged and neutral constituents as well as secondary vertices and global variables
  - Trained with 130M jets (from hadronically decaying $t\bar{t}$ pairs and QCD mutlijet events)
  - DeepJet is a combination of different layer architectures like CNNs, RNNs and DNNs
  - Significant performance boost for *c*- and light-flavour rejection for DeepJet over DeepCSV
  - Strong benefit for CMS physics programme!

- ATLAS:
  - Next generation tagger GN1 and his successor GN2 are showing very promising results
  - Graph neural network architecture (GN1) in addition with Self-Attention (GN2)
  - Trained with 30M (GN1) and 192M (GN2) jets using jets from $t\bar{t}$ and *Z'* events
  - Significant boost in rejection power for both GN1 and GN2 over current default tagger DL1d
  - First look into generator dependency and Data/MC show very promising performance of GN1
  - If you want to know more about GN1/GN2, CERN EP-IT Data Science Seminar by S. van Stroud
  - Also: Great improvement of *b*-tagging for the ATLAS physics programme!

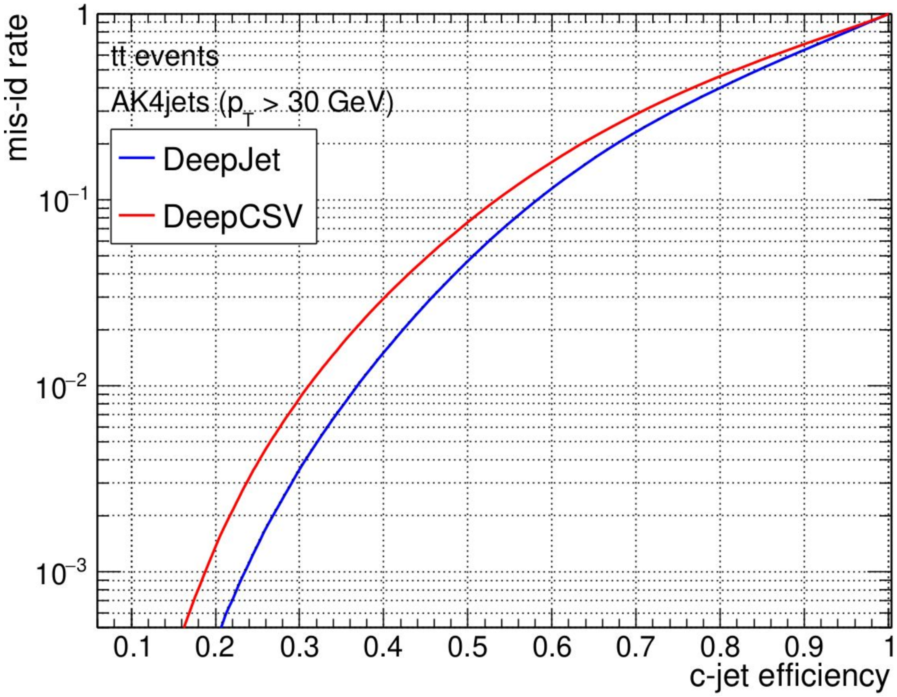
Thanks for your Attention!

Questions?


BRACE YOURSELF
BACKUP SLIDES ARE COMING
memegenerator.net

Backup

DeepJet c-Tagging



- Using binary discriminant

$$D_c = \frac{P(c)}{P(c) + P(uds) + P(g)}$$

- Great improvement of c-tagging for DeepJet over DeepCSV

- Improvement over the full range of c-jet efficiency

DeepJet Constituent Variables

A.2 List of charged candidate variables

- Charged track η relative to the jet axis
- Charged track p_t relative to the jet axis
- Dot product of the jet and track momentum
- Dot product of the jet and track momentum divided by the magnitude of the jet momentum
- ΔR between the jet axis and the track
- The track 2D impact parameter value
- The track 2D impact parameter significance
- The track 3D impact parameter value
- The track 3D impact parameter significance
- The track distance to the jet axis
- Fraction of the jet momentum carried by the track.
- ΔR between the track and the closest secondary vertex
- An integer flag that indicate whether the track was used in the primary vertex fit.
- The charged candidates PUPPI weight
- χ^2 of the charged track fit.
- A integer flag which indicate the quality of the fitted track, based on number of detector hits used for the reconstruction as well as the overall χ^2 of the charged track fit.

A.3 List of neutral candidate variables

- Fraction of the jet momentum carried by the neutral candidate
- ΔR between the jet axis and the neutral candidate
- A integer flag indicating whether the neutral candidate is a photon.
- Fraction of the neutral candidate energy deposited in the hadronic calorimeter.
- ΔR between the neutral candidate and the closest secondary vertex
- The neutral candidates PUPPI weight

DeepJet Global/Secondary Vertices Variables

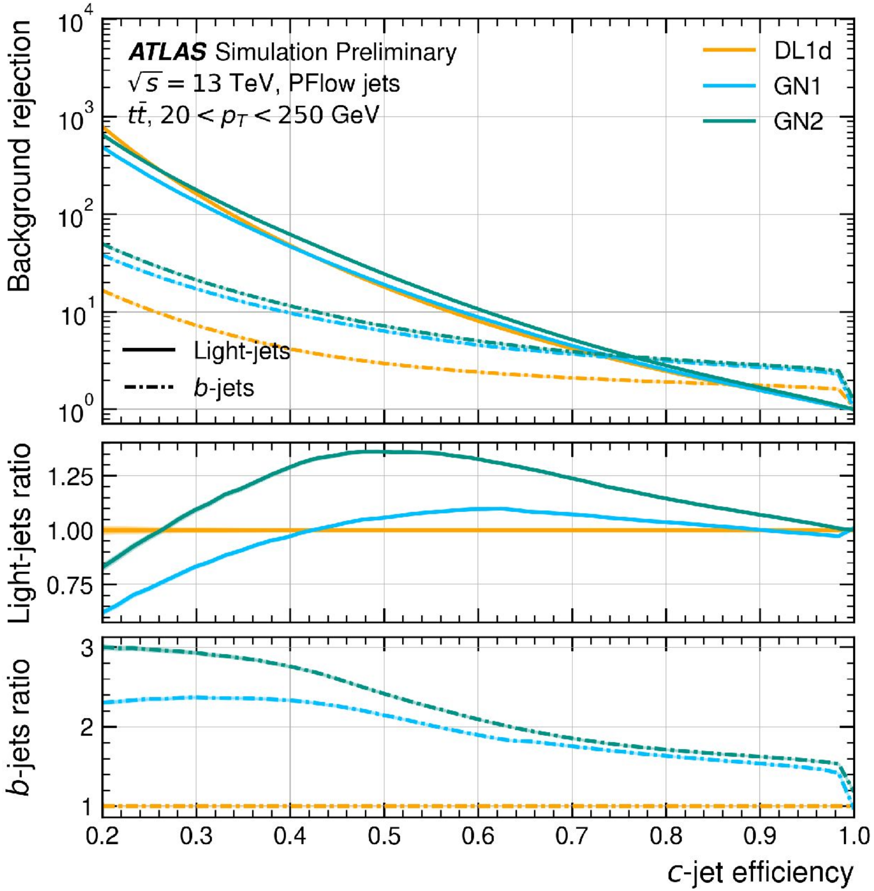
A.1 List of global variables

- Jet p_t
- Jet η
- The number of charged particle flow candidates in the jet
- The number of neutral particle flow candidates in the jet
- The number of secondary vertices in the jet
- The number of primary vertices in the event

A.4 List of secondary vertex variables

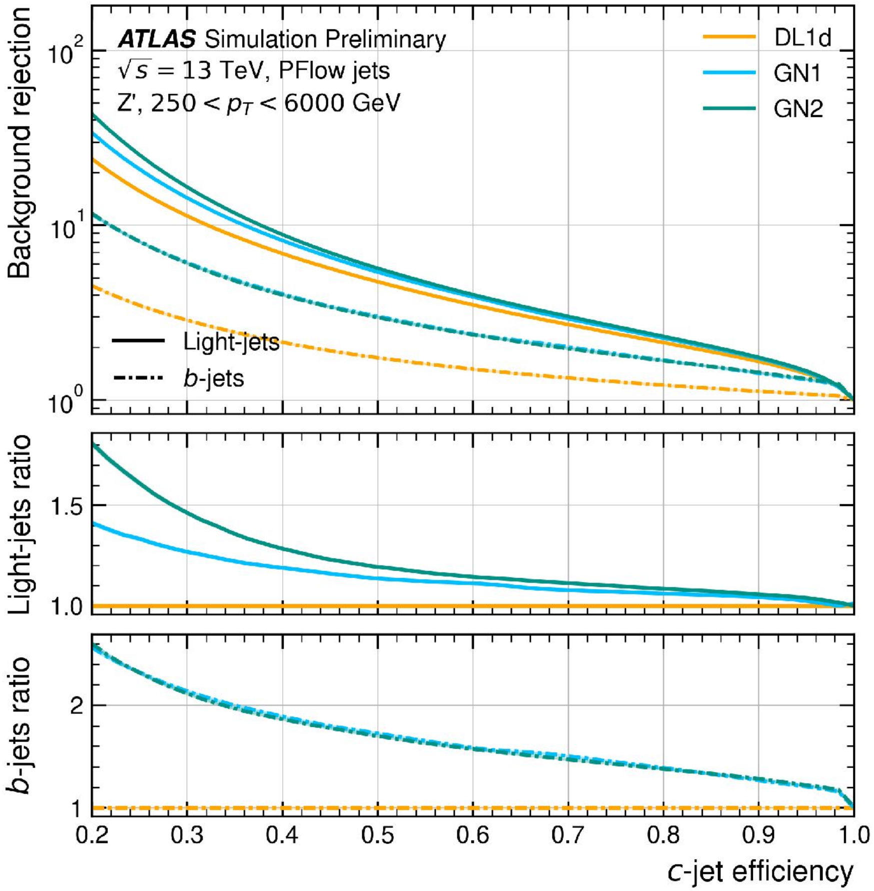
- Secondary vertex p_t
- ΔR between the jet axis and the secondary vertex
- Secondary vertex mass
- Number of tracks in the secondary vertex
- χ^2 of the secondary vertex fit
- Reduced χ^2 of the secondary vertex fit
- The secondary vertex 2D impact parameter value
- The secondary vertex 2D impact parameter significance
- The secondary vertex 3D impact parameter value
- The secondary vertex 3D impact parameter significance
- Cosine of the angle between the secondary vertex flight direction and the direction of the secondary vertex momentum.
- Ratio of the secondary vertex energy to the jet energy

GN1/GN2 - Results - *c*-Tagging



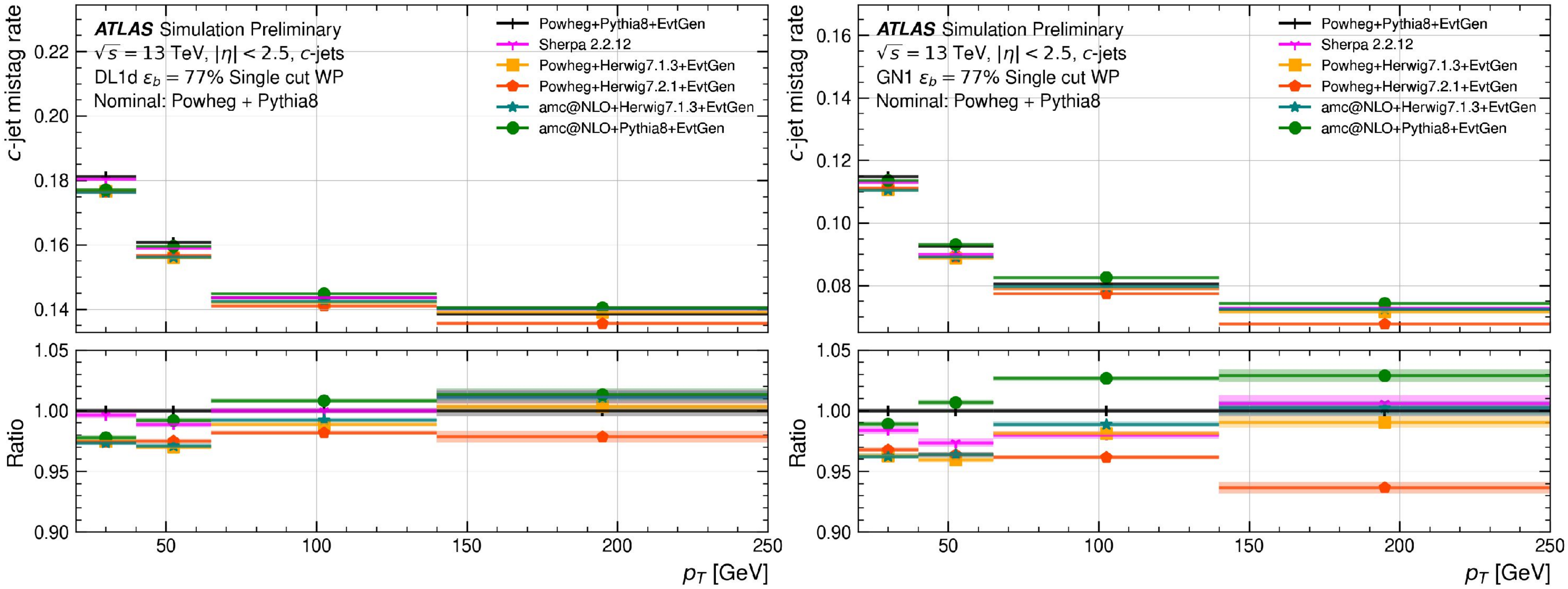
- Large performance gains on low p_T jets in *b*-jet rejection for GN1 over current ATLAS default tagger (DL1d)
- Slight performance gains in light jets rejection for *c*-jet efficiencies over 40%
- Slight performance decrease in lower *c*-jet efficiencies area
- Larger performance gains in both light- and *b*-jet rejections for GN2

GN1/GN2 - Results - c-Tagging



- Large performance gains on high p_T jets in background rejection for GN1 over current ATLAS default tagger (DL1d)
- Larger performance gains for light jet rejection for GN2
- Similar performance for b-jet rejection for GN1 and GN2

GN1/GN2 - Generator Dependency - *c*-Jets



- Overall good agreement between generators for both DL1d and GN1 (Slightly larger disagreement for GN1)
- Indicating that both models are not exploiting generator specific information

GN1 Variables

Jet Input	Description
p_T	Jet transverse momentum
η	Signed jet pseudorapidity

Track Input	Description
q/p	Track charge divided by momentum (measure of curvature)
$\mathrm{d}\eta$	Pseudorapidity of the track, relative to the jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin\theta$	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used in the reconstruction of an electron or muon (only for GN1 Lep)

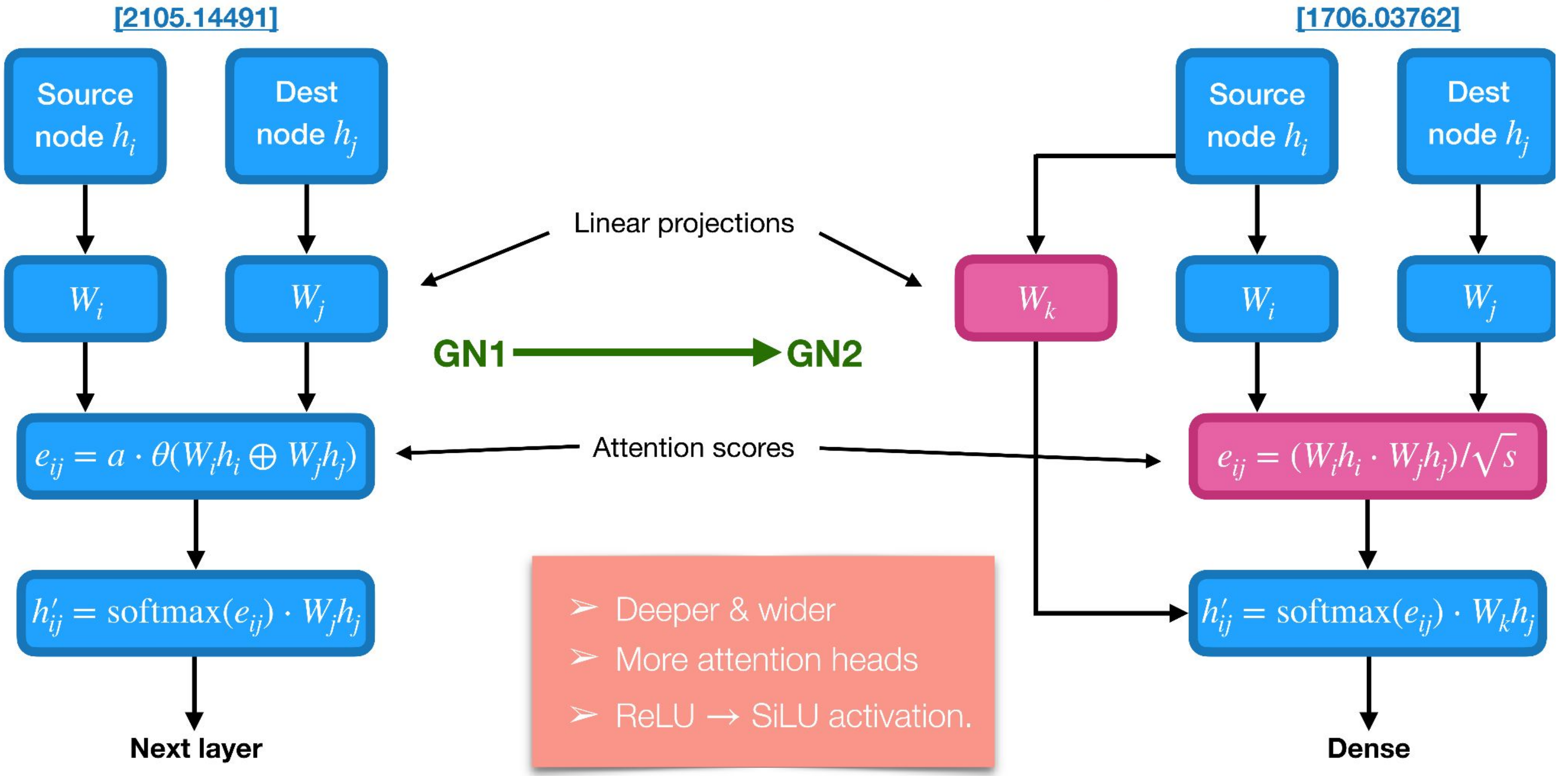
GN1 and GN2 - Main Differences

- Learning rate optimisation
  - Using OneCycle Learning Rate Scheduler
- Added layer normalisation and Dropout
  - Stablises training
  - Allows larger model size
- New framework
  - More efficient training
  - Enables high stat. training
  - 30M → 192M training jets
- Updated attention mechanism

Type	Name	GN1	GN2
Hyperparameter	Trainable parameters	0.8M	1.5M
Hyperparameter	Learning rate	$1e{-}3$	OneCycle LRS (max LR $4e{-}5$)
Hyperparameter	GNN Layers	3	6
Hyperparameter	Attention Heads	2	8
Hyperparameter	Embed. dim	128	192
Architectural	Attention type	GATv2	ScaledDotProduct
Architectural	Dense update	No	Yes (dim 256)
Architectural	Separate value projection	No	Yes
Architectural	LayerNorm + Dropout	No	Yes
Inputs	Num. training jets	30M	192M

GN1 and GN2 - Updated Attention Mechanism

GN2 follows more closely the *transformer* architecture [1706.03762]



- Deeper & wider
- More attention heads
- ReLU → SiLU activation.

GN1 Hyperparameters

Track Truth Origins

Truth Origin	Description
Pileup	From a pp collision other than the primary interaction
Fake	Created from the hits of multiple particles
Primary	Does not originate from any secondary decay
fromB	From the decay of a b-hadron
fromBC	From a c-hadron decay, which itself is from the decay of a b-hadron
fromC	From the decay of a c-hadron
OtherSecondary	From other secondary interactions and decays

Input Track Quality Selections

Parameter	Selection
p_{T}	> 500 MeV
$\lvert d_0 \rvert$	< 3.5 mm
$\lvert z_0 \sin\theta \rvert$	< 5 mm
Silicon hits	≥ 8
Shared silicon hits	< 2
Silicon holes	< 3
Pixel holes	< 2

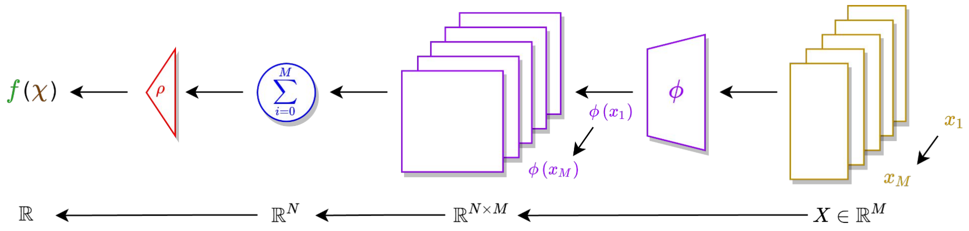
Auxiliary Task Heads

Network	Hidden layers	Output size
Node classification network	128, 64, 32	7
Edge classification network	128, 64, 32	1
Graph classification network	128, 64, 32, 16	3

Deep Sets

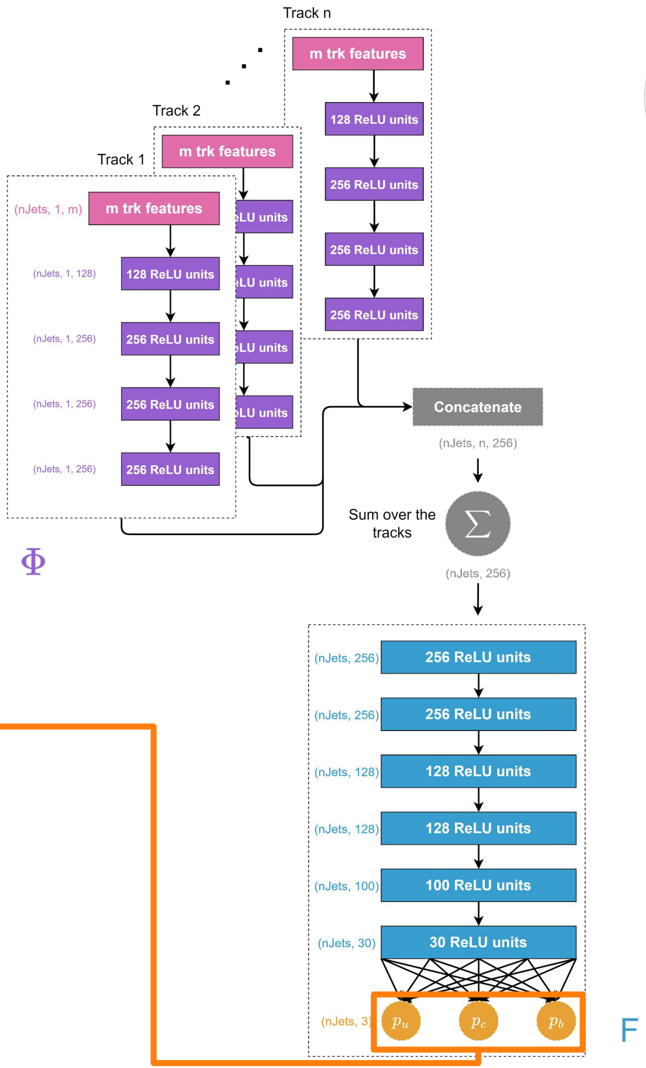
- First use in HEP: [arXiv:1810.05165]
- Set function f on set of tracks χ can be decomposed
- Process each element of the set with mapping function ϕ
- Aggregate processed elements into invariant description with aggregation function (here: summation)
- Process the aggregated description with ρ
- ϕ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



Training Sample DIPS/DL1d

- 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)