Heavy-flavour jet tagging in ATLAS

… and how we use Machine Learning

Philipp Windischhofer
University of Oxford
philipp.windischhofer@cern.ch

on behalf of the ATLAS Collaboration

ML4Jets, January 15-17, 2020, NYU

Why flavour tagging?

quarks:
"elementary"

Quantum numbers:
symmetries

momentum p^{μ}

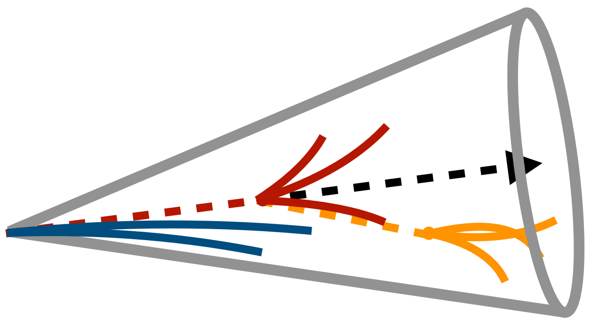
charge, flavour, … σ

Why flavour tagging?

quarks:
"elementary"

TeV **QCD** GeV

jets:
observables

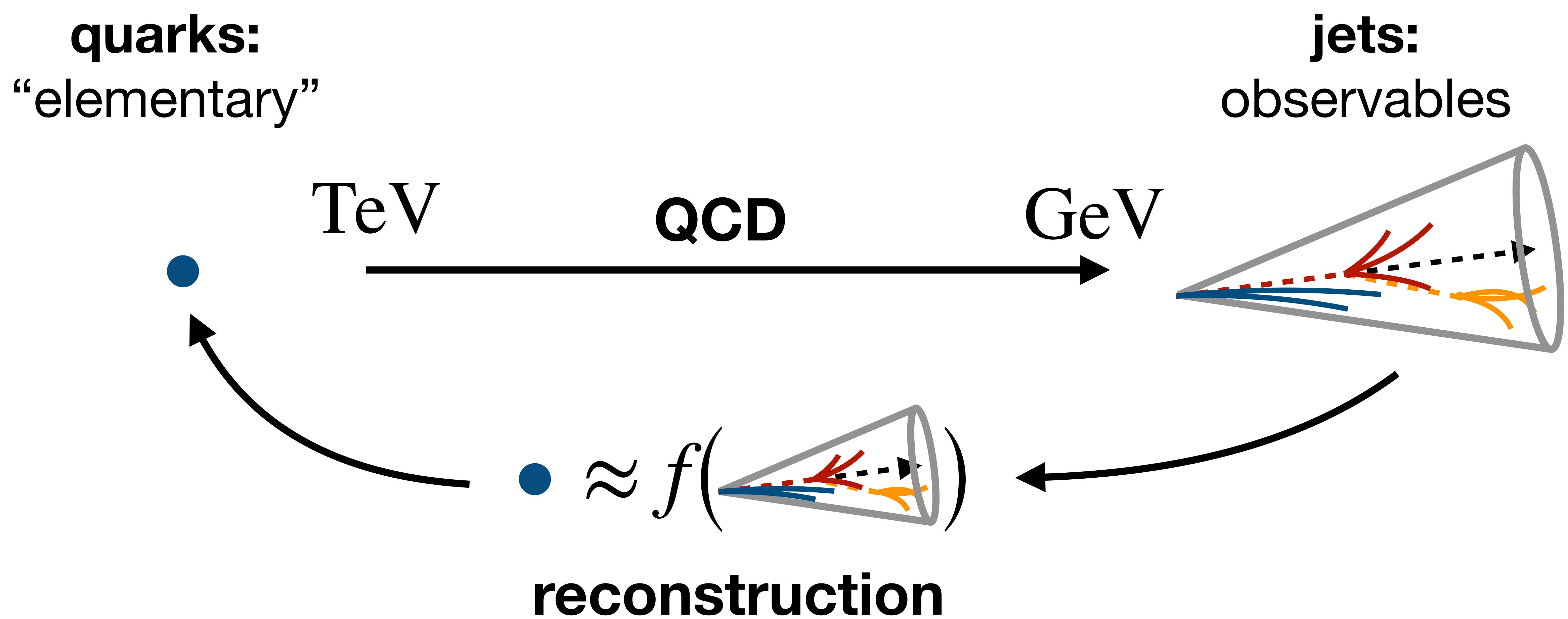


Quantum numbers:
symmetries

momentum p^μ

charge, flavour, … σ

Why flavour tagging?



Quantum numbers:
symmetries

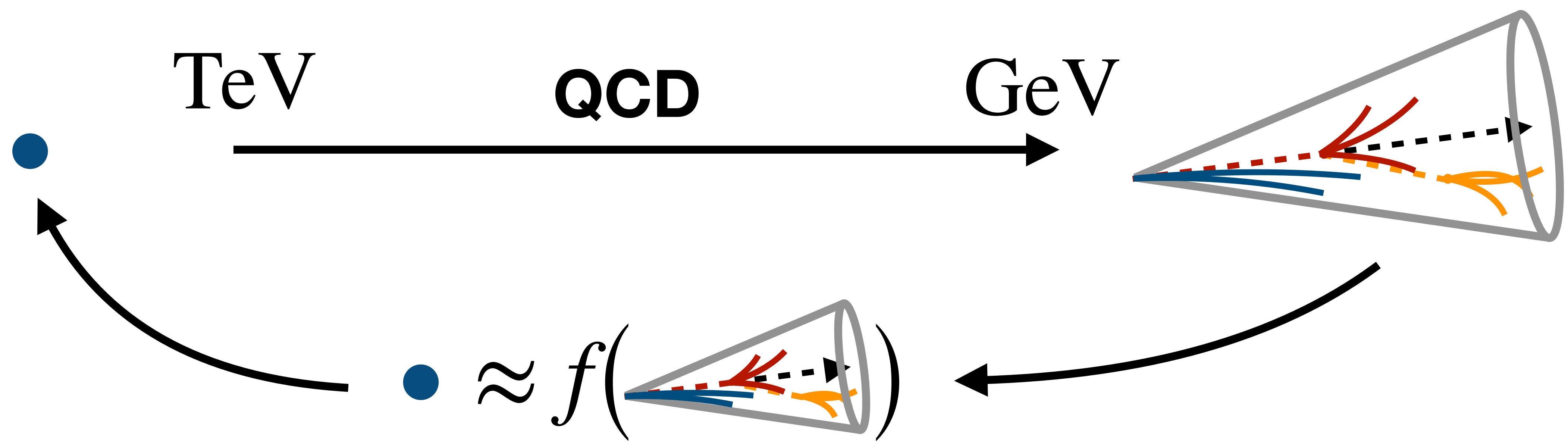
momentum p^μ

charge, flavour, … σ

Why flavour tagging?

quarks:
"elementary"

jets:
observables



reconstruction

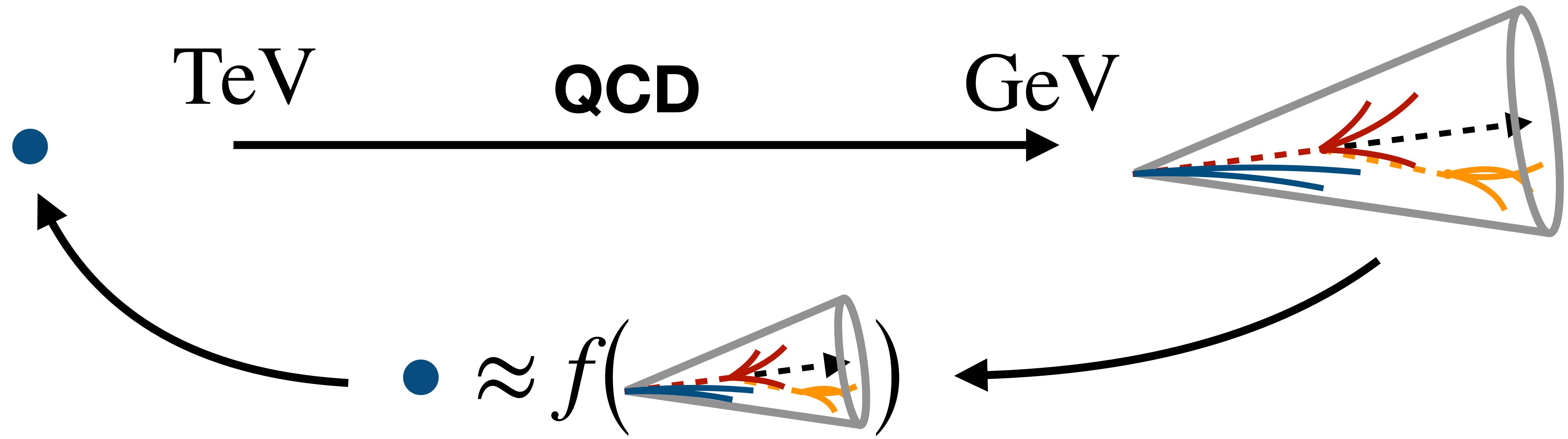
Quantum numbers:
symmetries

momentum	p^μ	$\longrightarrow$	$f =$ jet clustering algorithm
charge, flavour, …	σ	$\longrightarrow$	$f =$ jet tagger

Why flavour tagging?

quarks:
"elementary"

jets:
observables



reconstruction

Quantum numbers:
symmetries

momentum p^μ $\longrightarrow$ $f =$ jet clustering algorithm

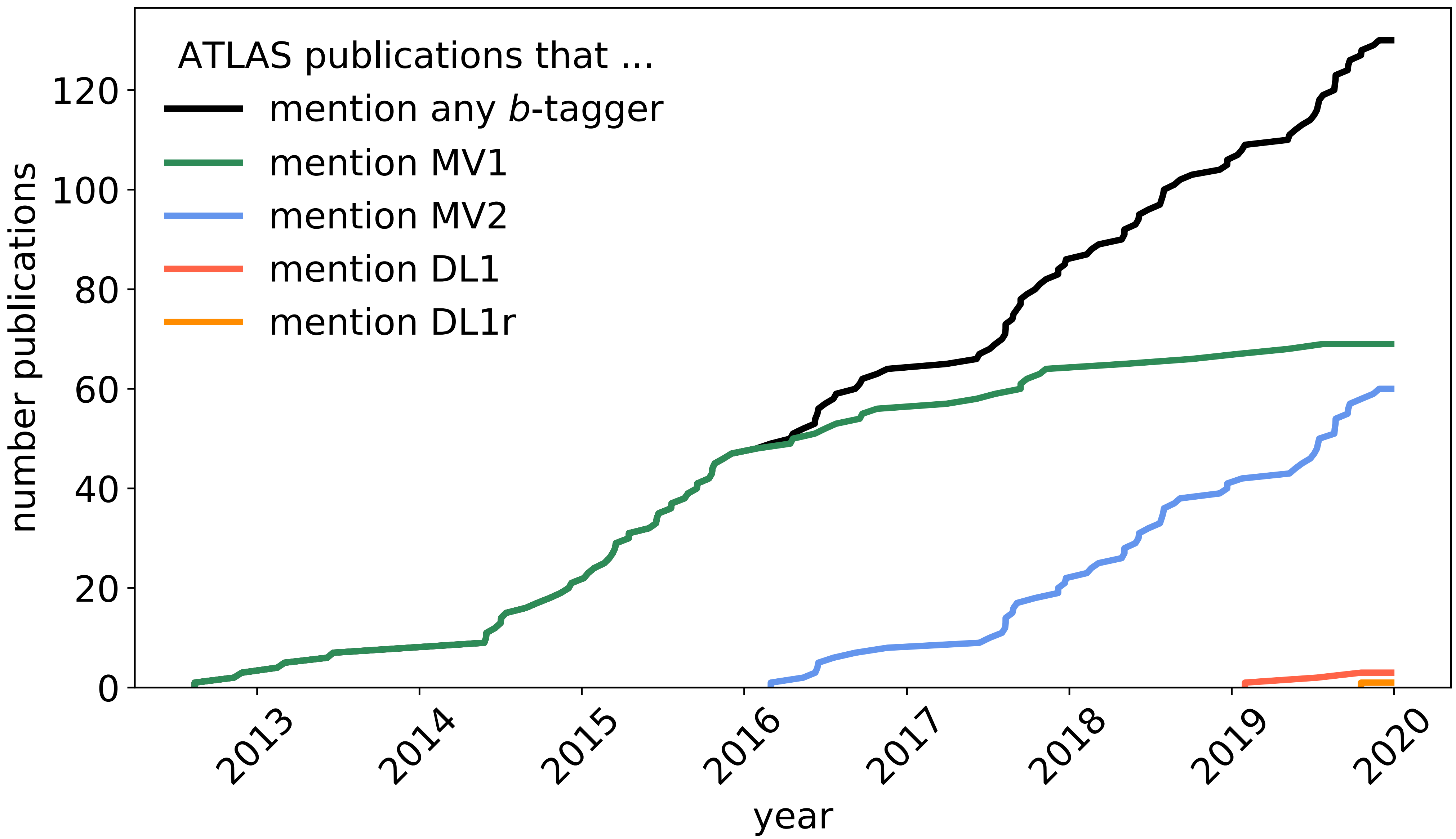
charge, **flavour**, … σ $\longrightarrow$ $f =$ jet flavour tagger

What can you expect?

By the end of this talk, you will know …

- … how flavour tagging helps ATLAS
- … how the ATLAS flavour tagging tools work
- … how machine learning helped to improve them
- … what their performance is on data and simulation

The market for b-tagging



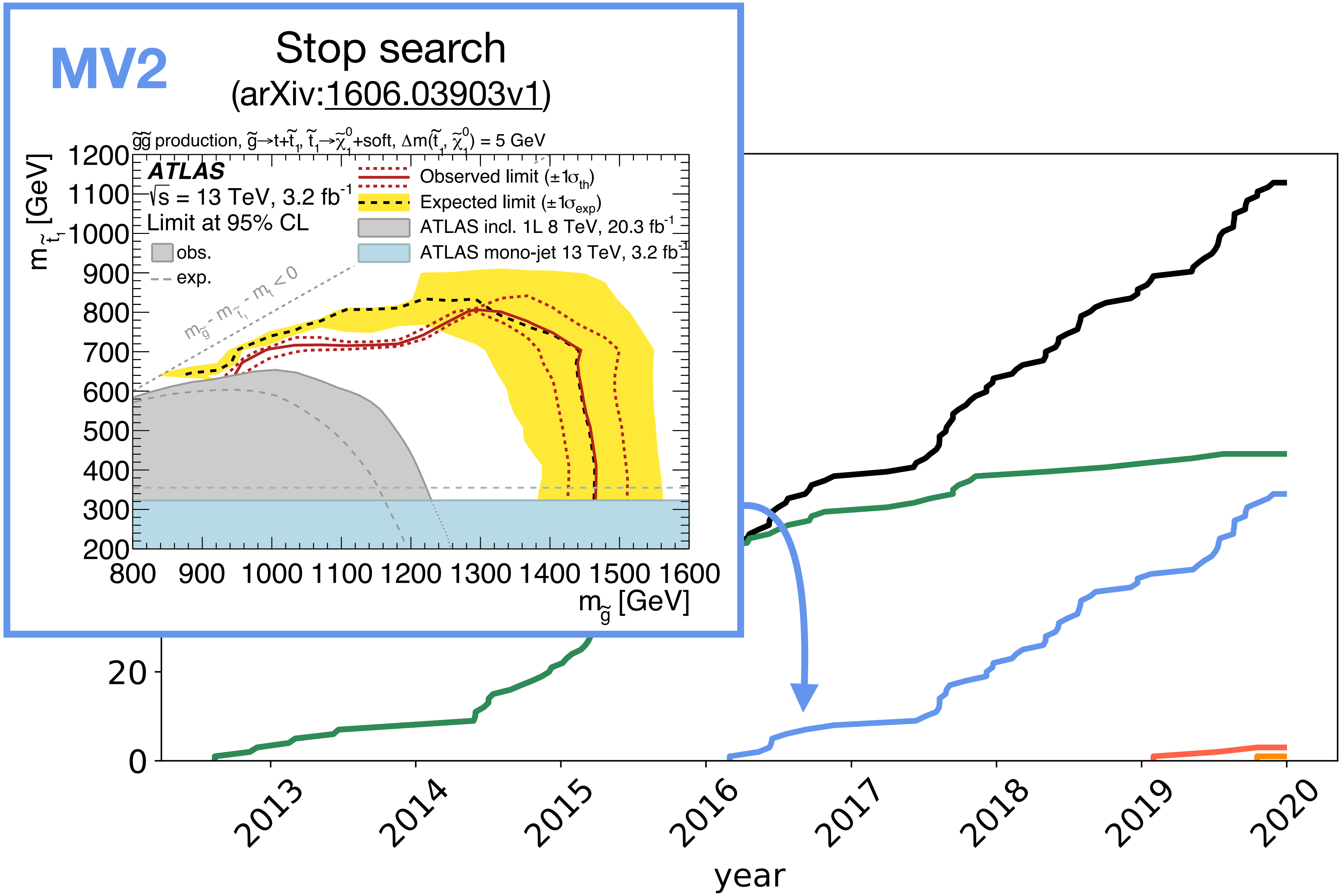
MV1: Run 1 workhorse

MV2: Run 2 workhorse

DL1, **DL1r**: the future!

https://gitlab.cern.ch/phwindis/arxivscraper

The market for b-tagging



MV1: Run 1 workhorse

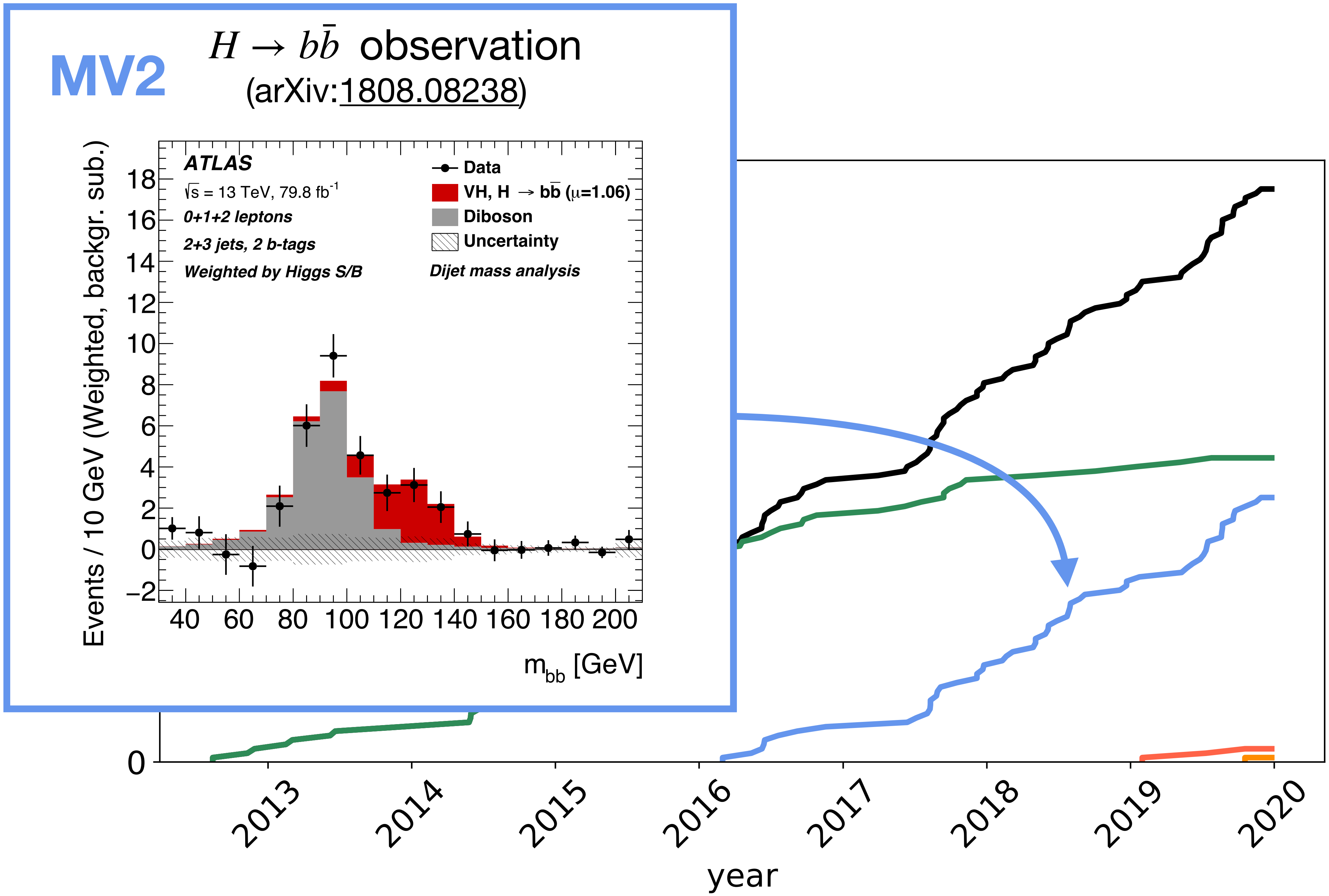
MV2: Run 2 workhorse

DL1, **DL1r**: the future!

MV2

https://gitlab.cern.ch/phwindis/arxivscraper

The market for b-tagging



MV1:
Run 1 workhorse

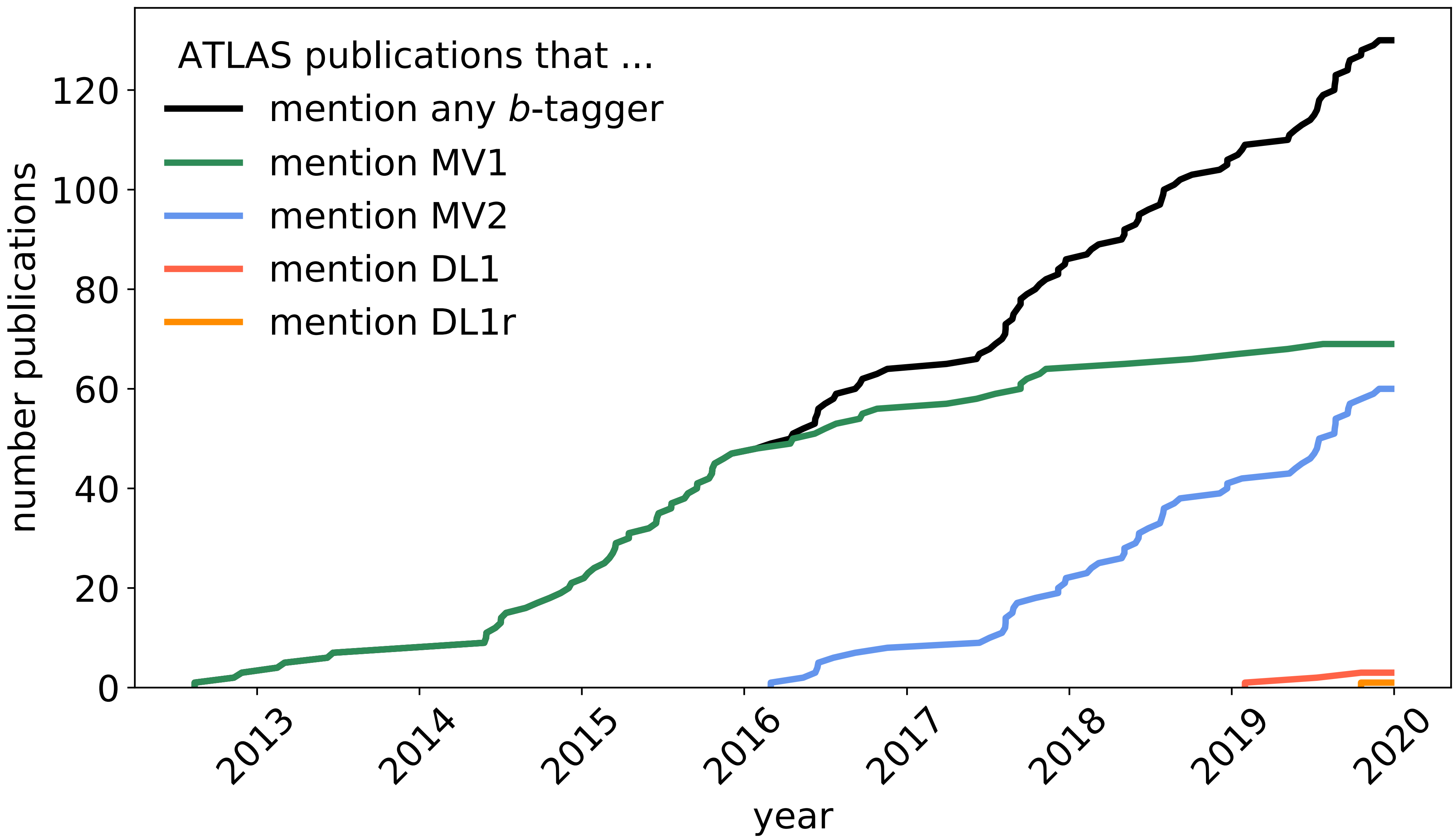
MV2:
Run 2 workhorse

DL1, **DL1r**:
the future!

MV2

https://gitlab.cern.ch/phwindis/arxivscraper

The market for b-tagging



MV1:
Run 1 workhorse

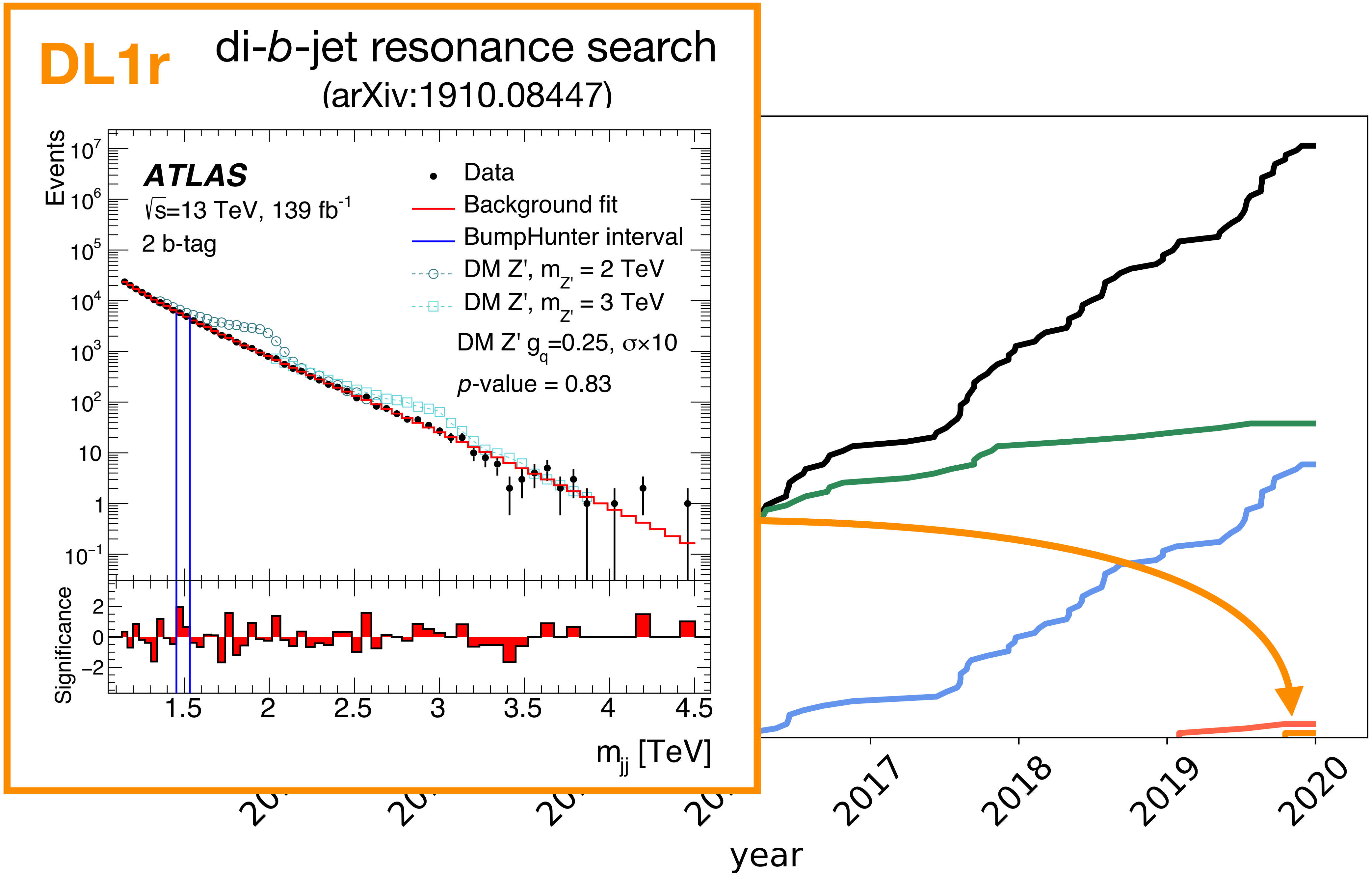
MV2:
Run 2 workhorse

DL1, **DL1r**:
the future!

MV2

https://gitlab.cern.ch/phwindis/arxivscraper

The market for b-tagging



MV1:
Run 1 workhorse

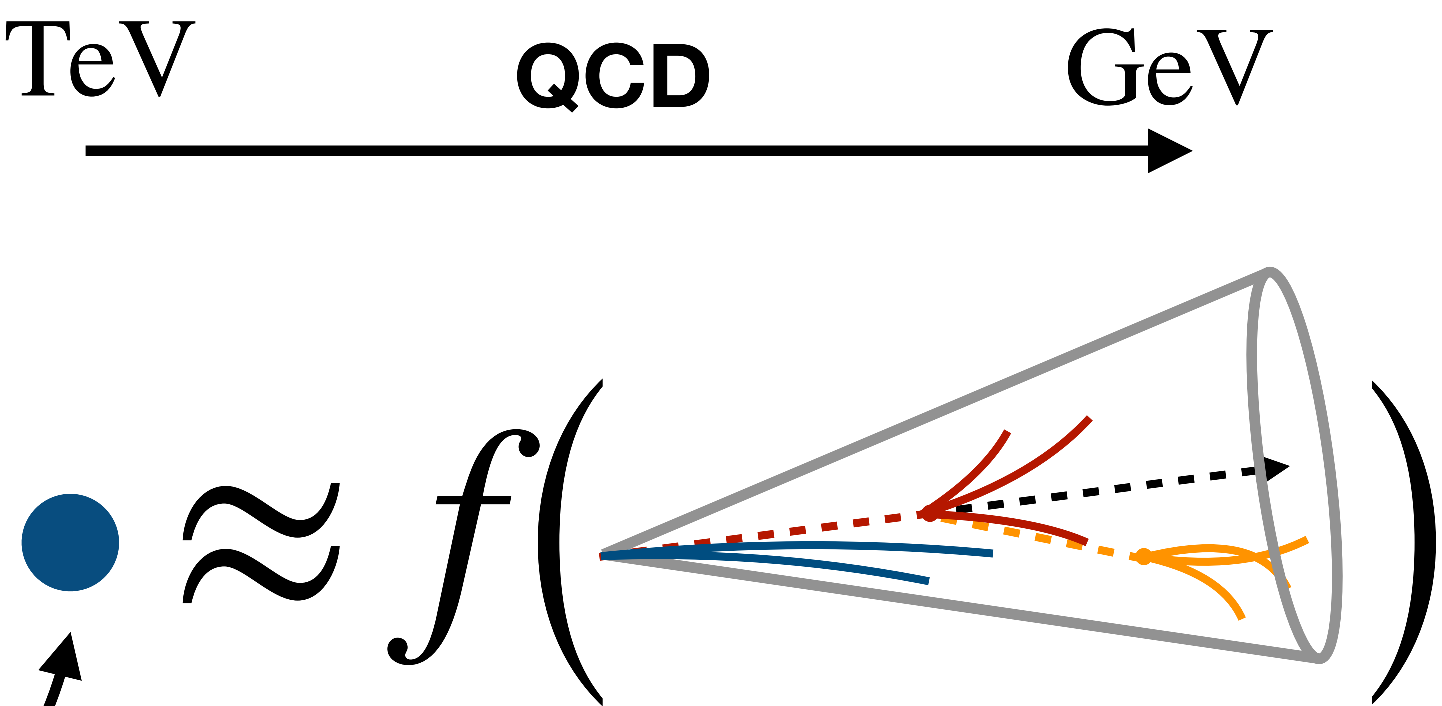
MV2:
Run 2 workhorse

DL1, **DL1r**:
the future!

MV2, **DL1**, **DL1r**

https://gitlab.cern.ch/phwindis/arxivscraper

The ATLAS strategy for b-tagging



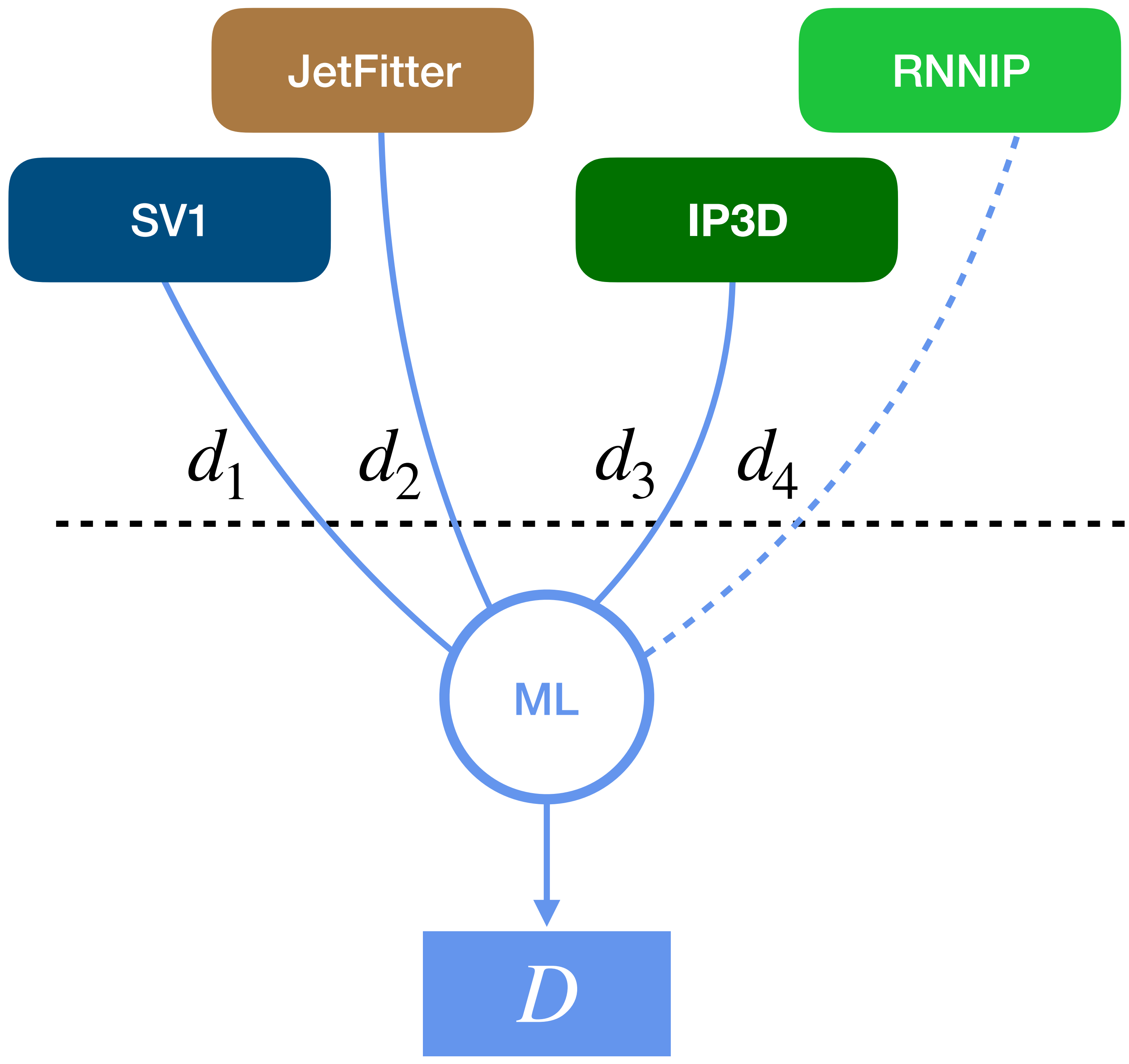
Quark
Low-dimensional
properties

Jet
High-dimensional
features

momentum p^{μ}

charge, **flavour**, … σ

The ATLAS strategy for b-tagging



MV2, DL1, DL1r

Low-level taggers

Physics-based
feature extraction

Detector-specific

High-level taggers

Likelihood ratio estimation

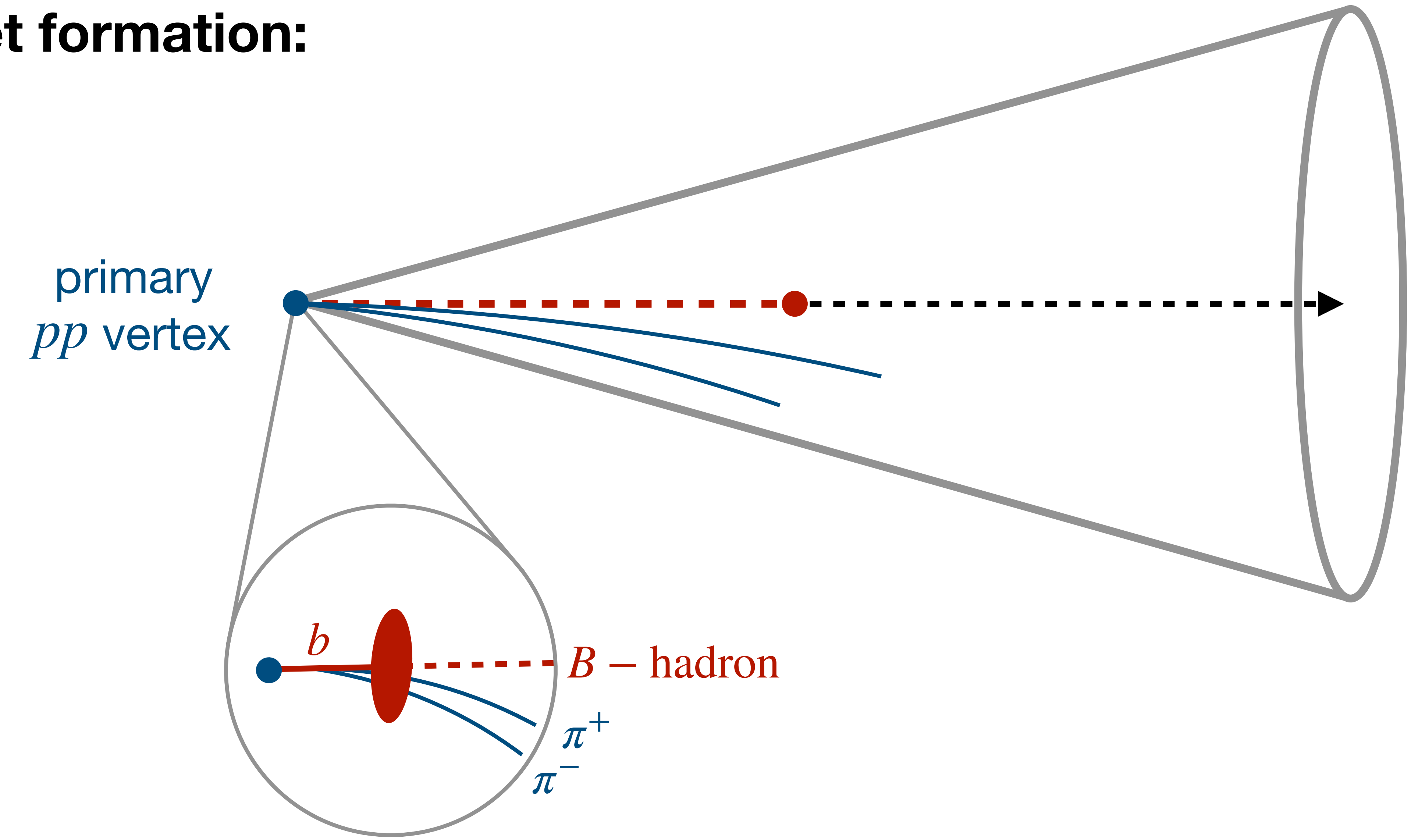
$$D \sim \log \frac{p(d_i \,|\, b - \text{jet})}{p(d_i \,|\, \neg b - \text{jet})}$$

Detector agnostic

Fast turnaround

Low-level taggers

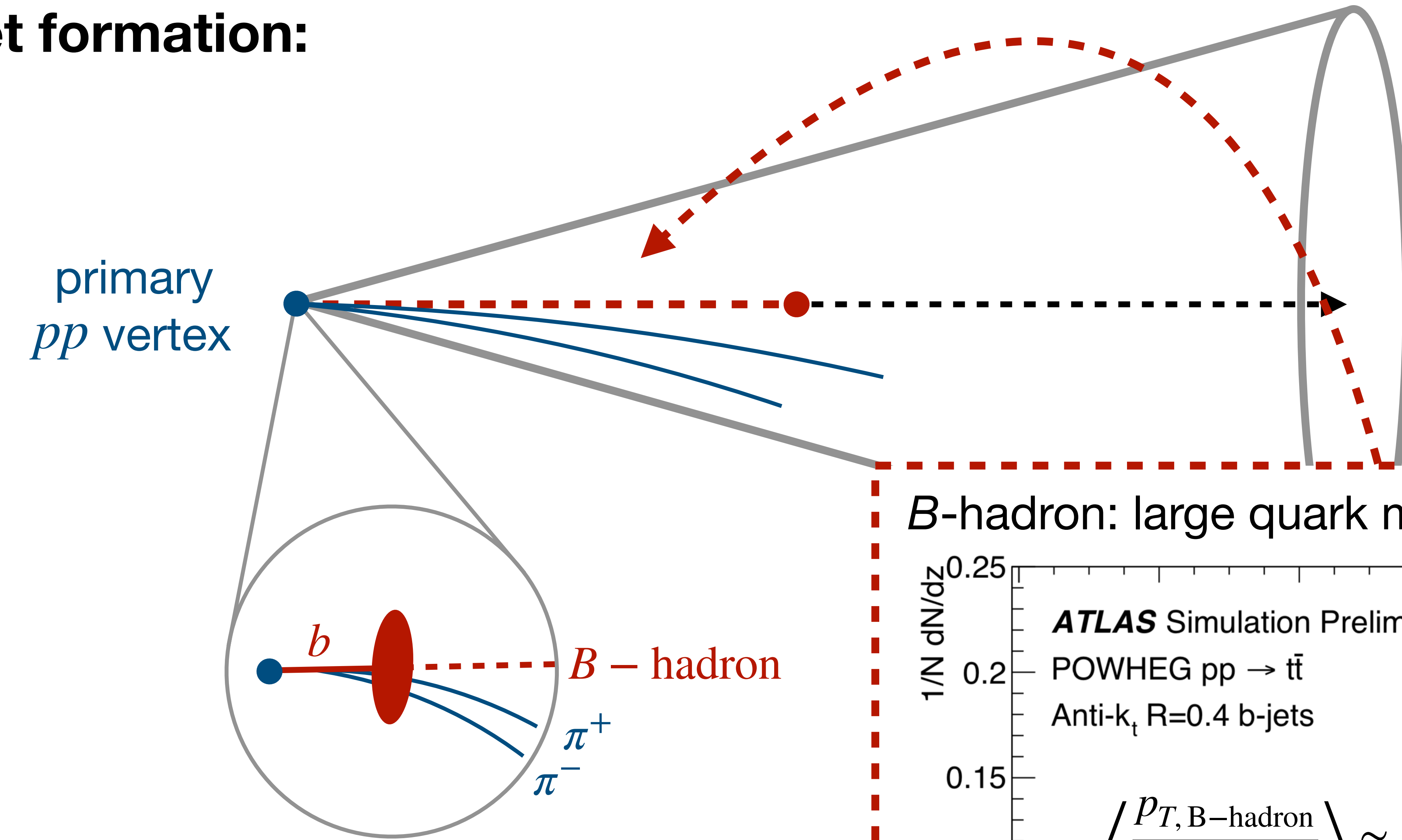
b-jet formation:



$$\tau_{\text{had.}} \sim \frac{1}{1\ \text{GeV}} \sim 10^{-24} s$$

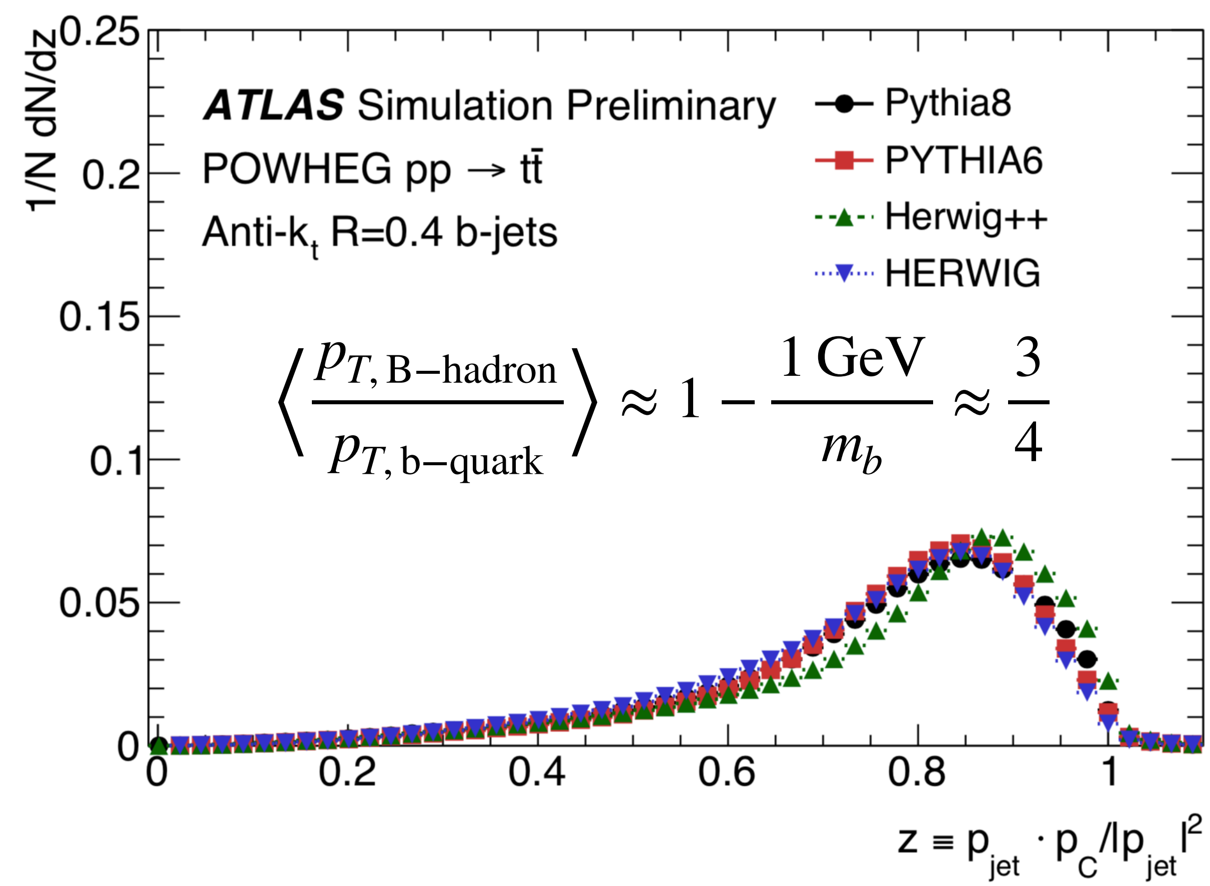
Low-level taggers

b-jet formation:



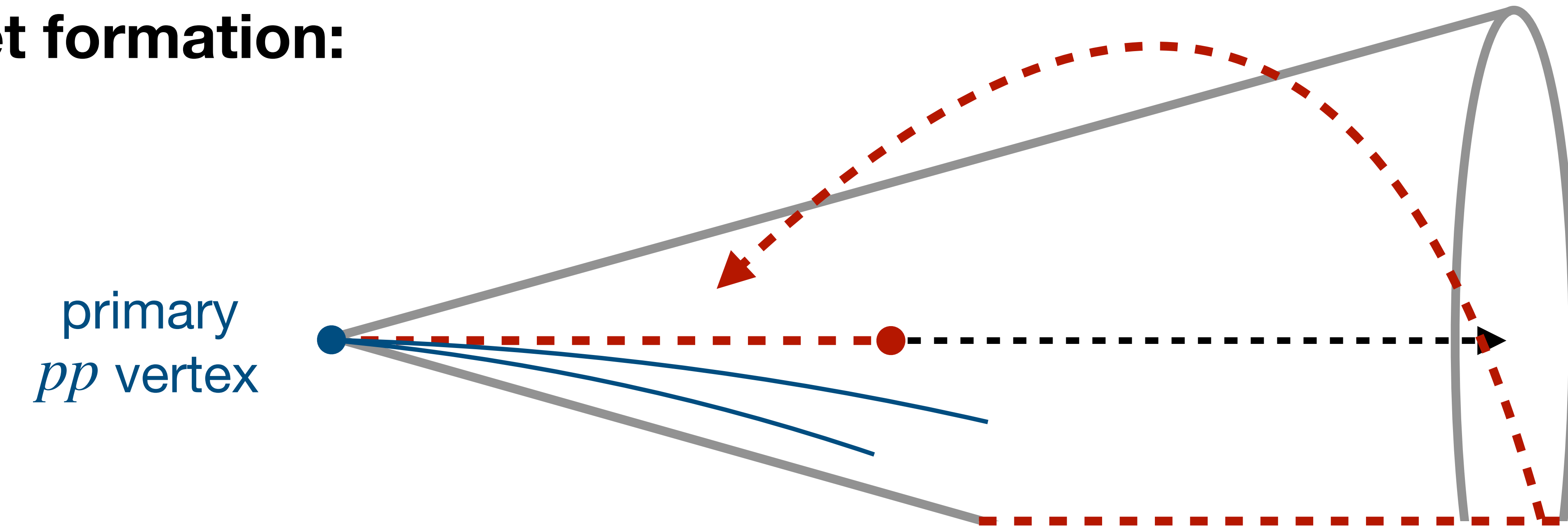
$$\tau_{\text{had.}} \sim \frac{1}{1\,\text{GeV}} \sim 10^{-24}s$$

B-hadron: large quark momentum fraction

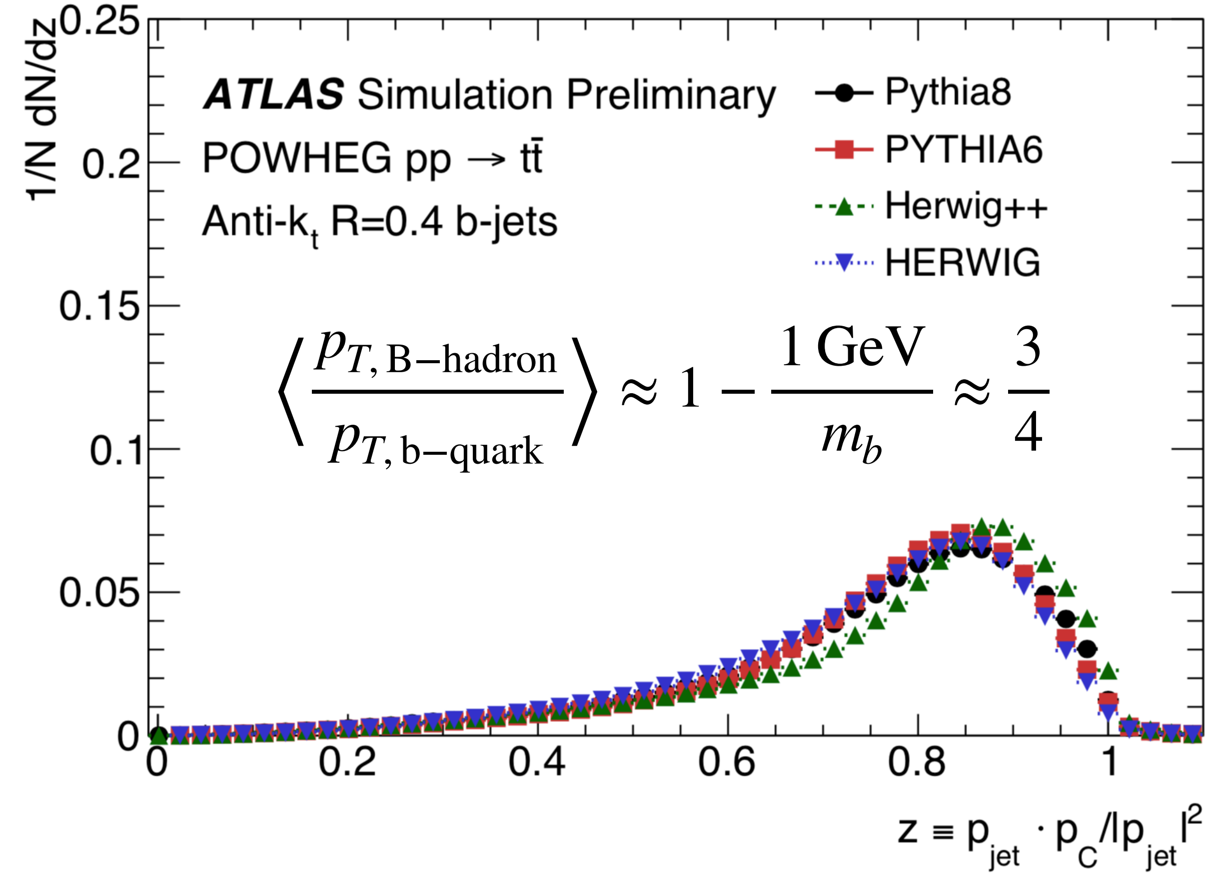


Low-level taggers

***b*-jet formation:**



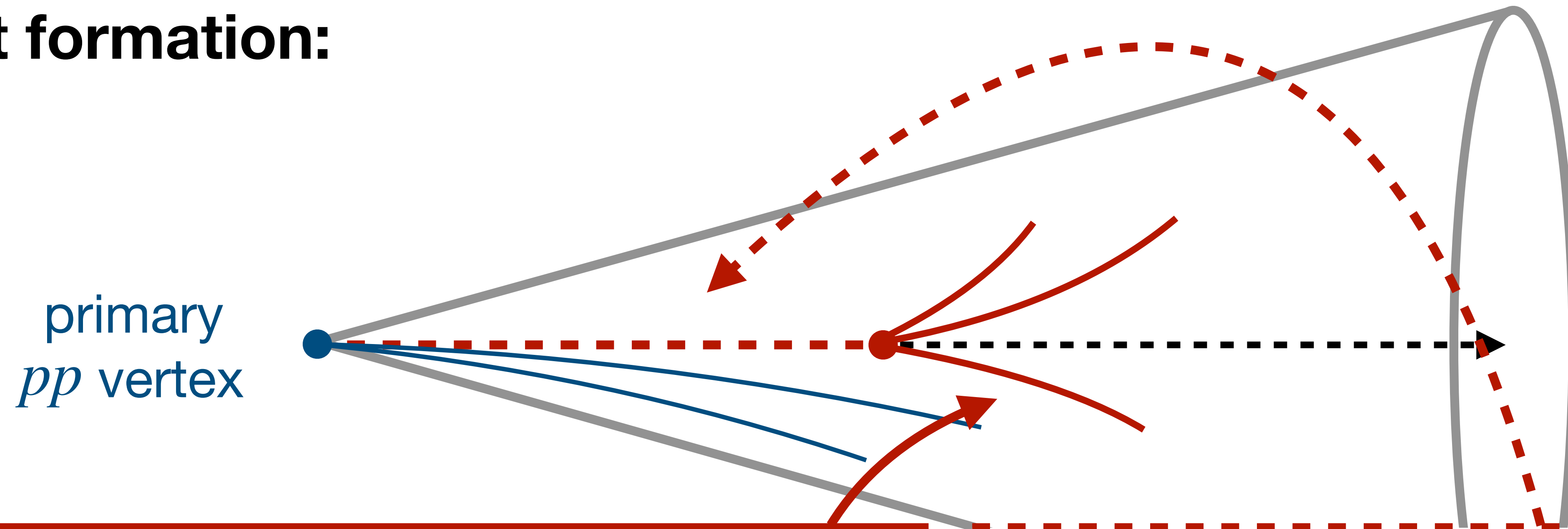
B-hadron: large quark momentum fraction



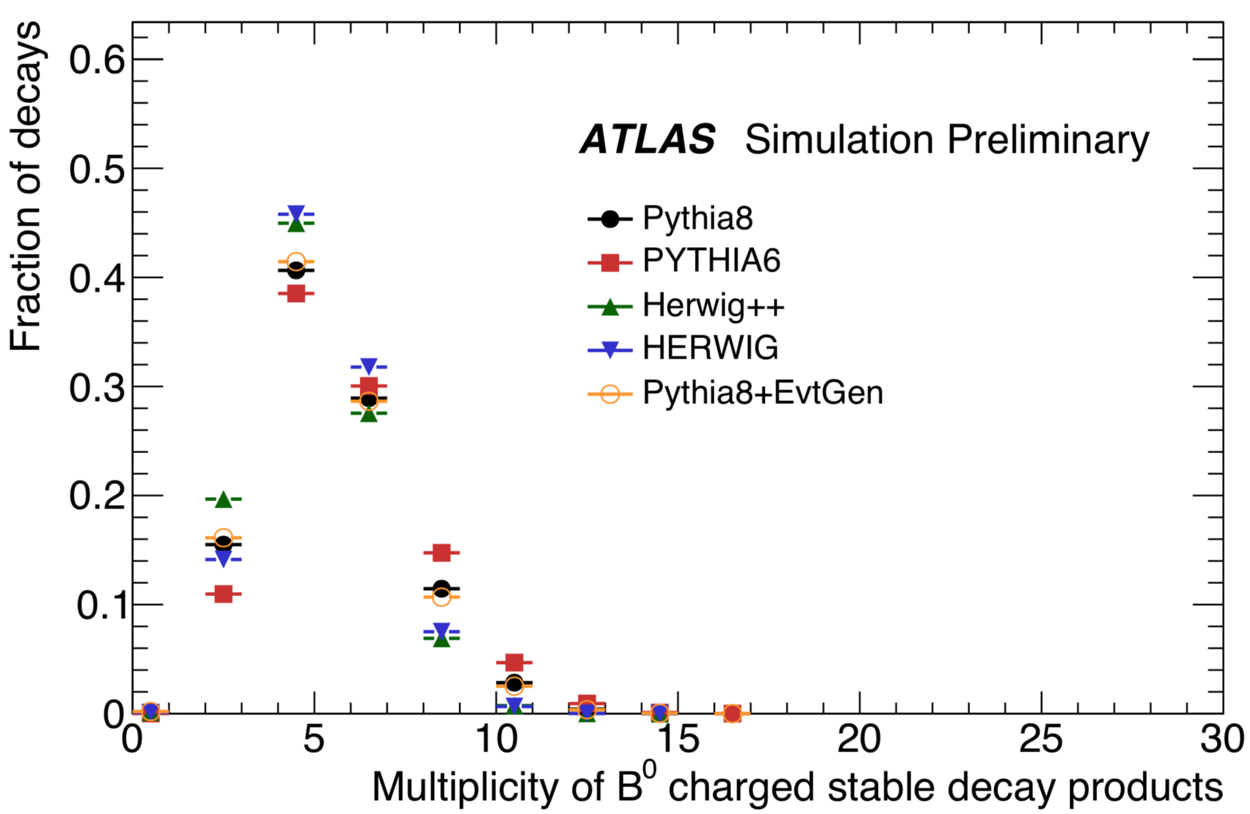
$$\left\langle \frac{p_{T,\,\mathrm{B-hadron}}}{p_{T,\,\mathrm{b-quark}}} \right\rangle \approx 1 - \frac{1\,\mathrm{GeV}}{m_b} \approx \frac{3}{4}$$

Low-level taggers

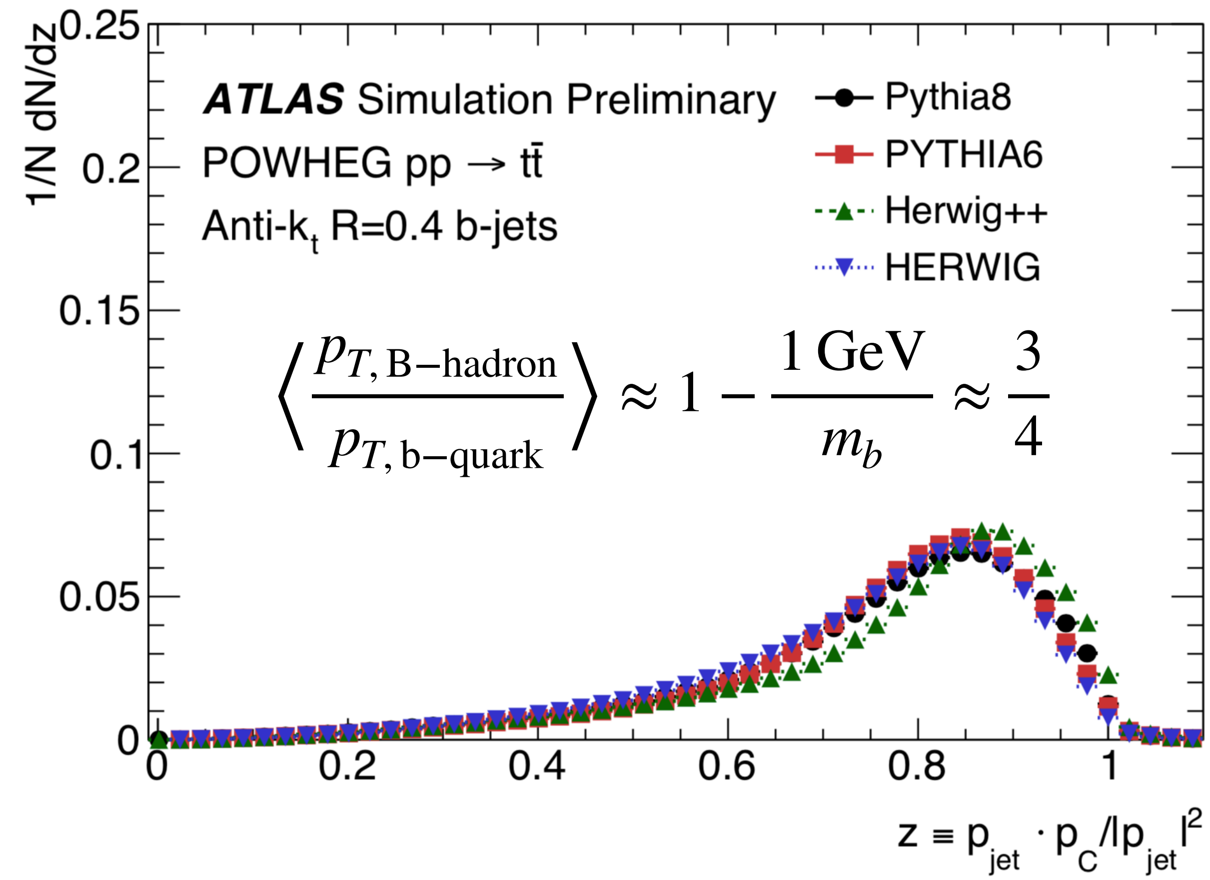
***b*-jet formation:**



B-hadron: high decay multiplicity

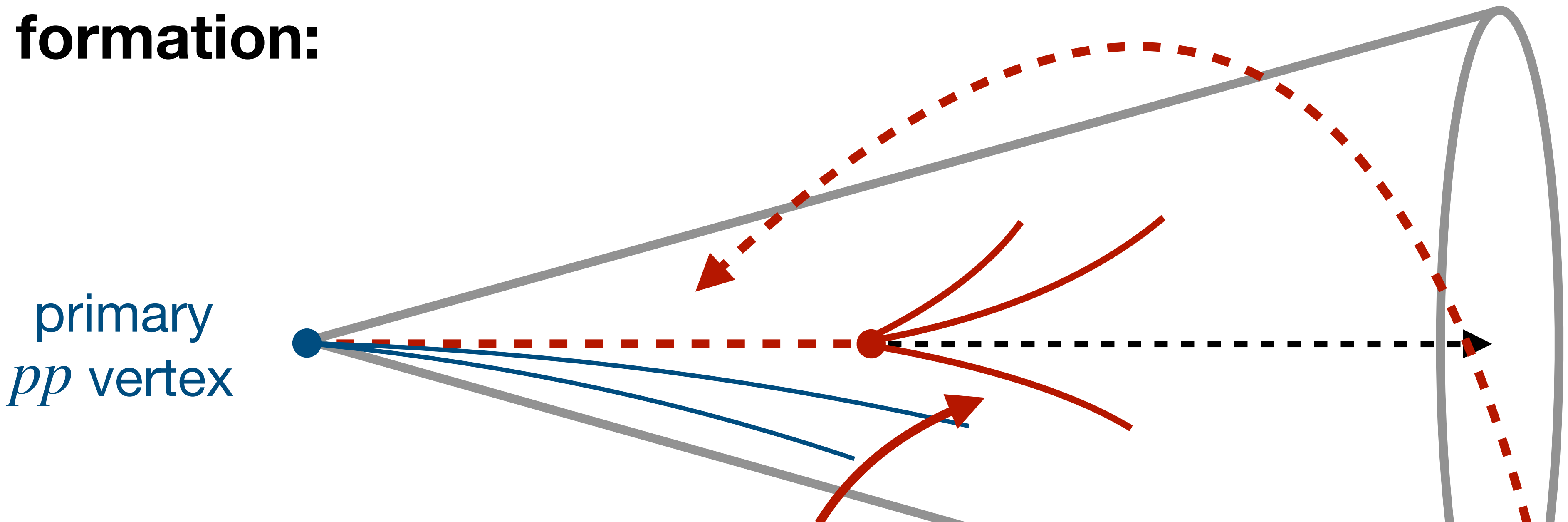


B-hadron: large quark momentum fraction

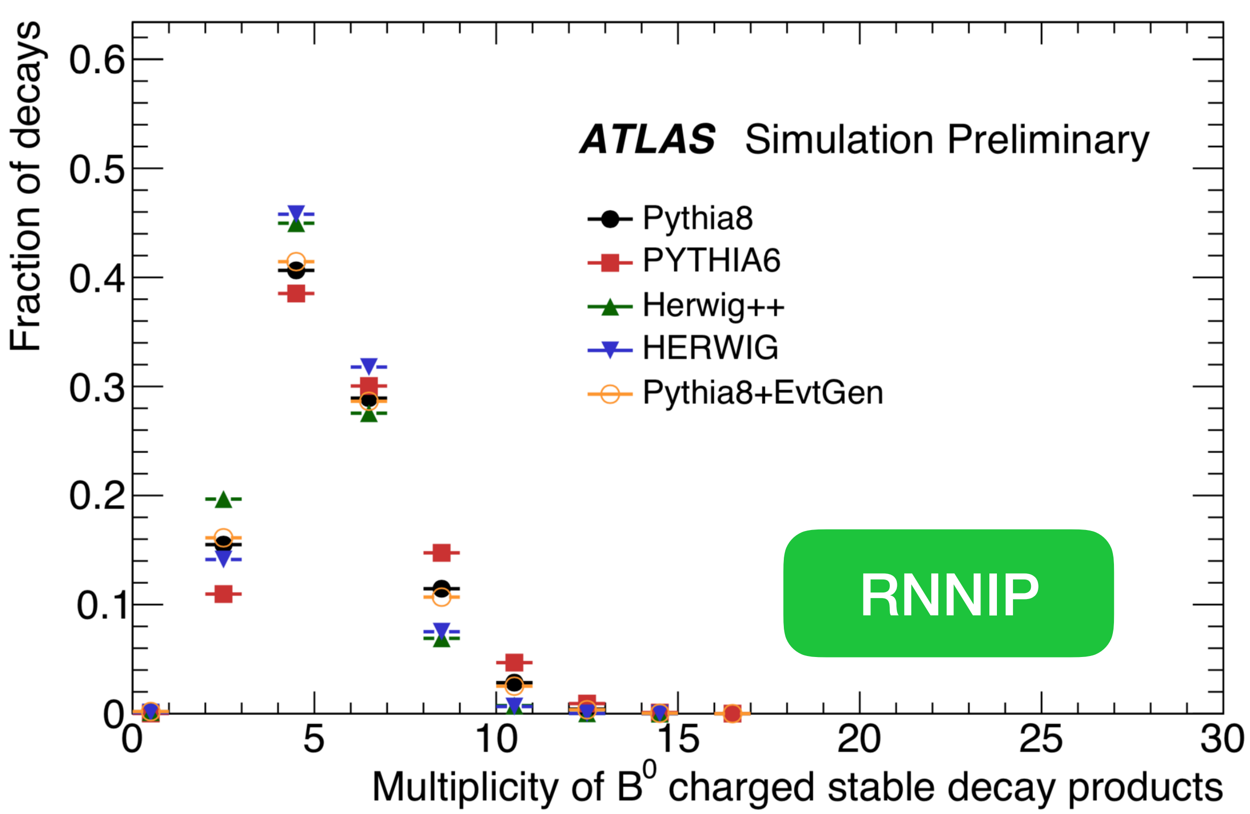


Low-level taggers

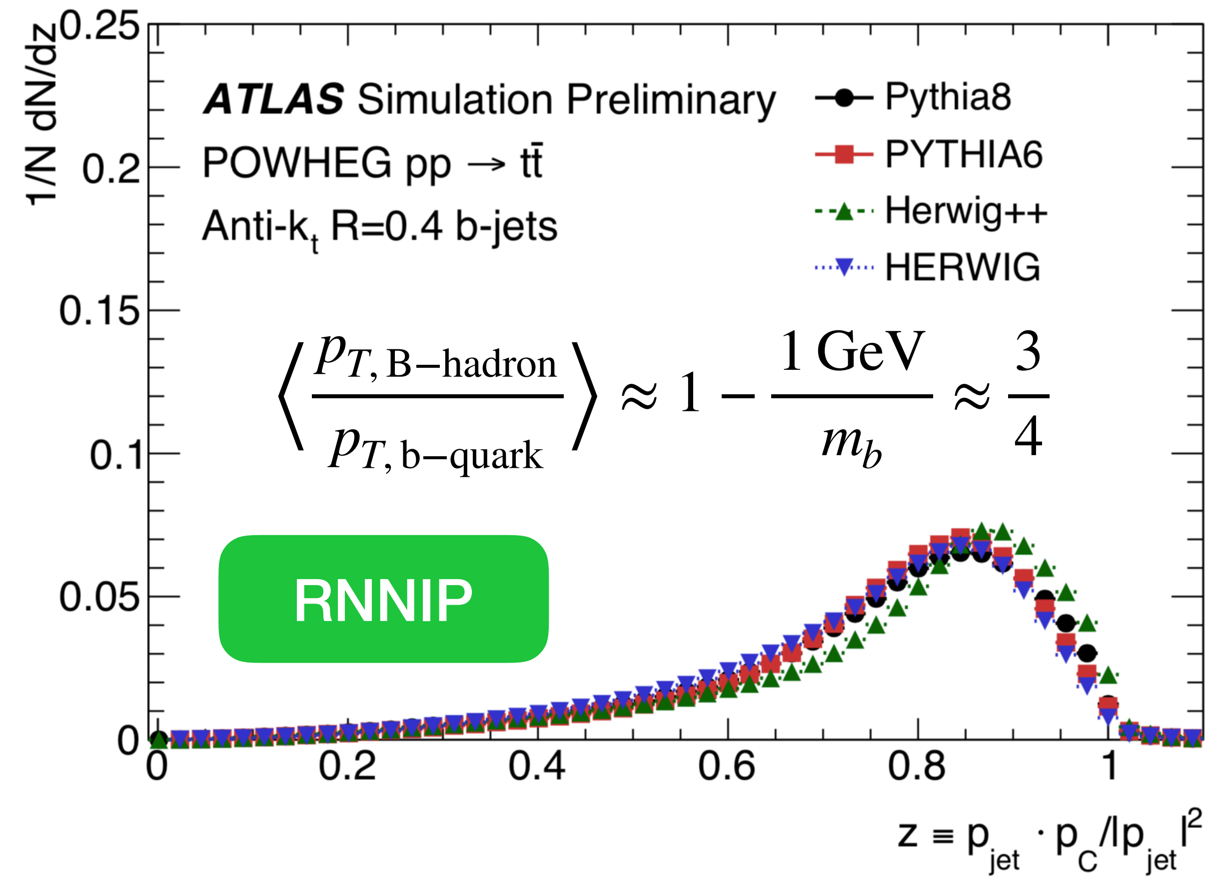
_b_-jet formation:



B-hadron: high decay multiplicity

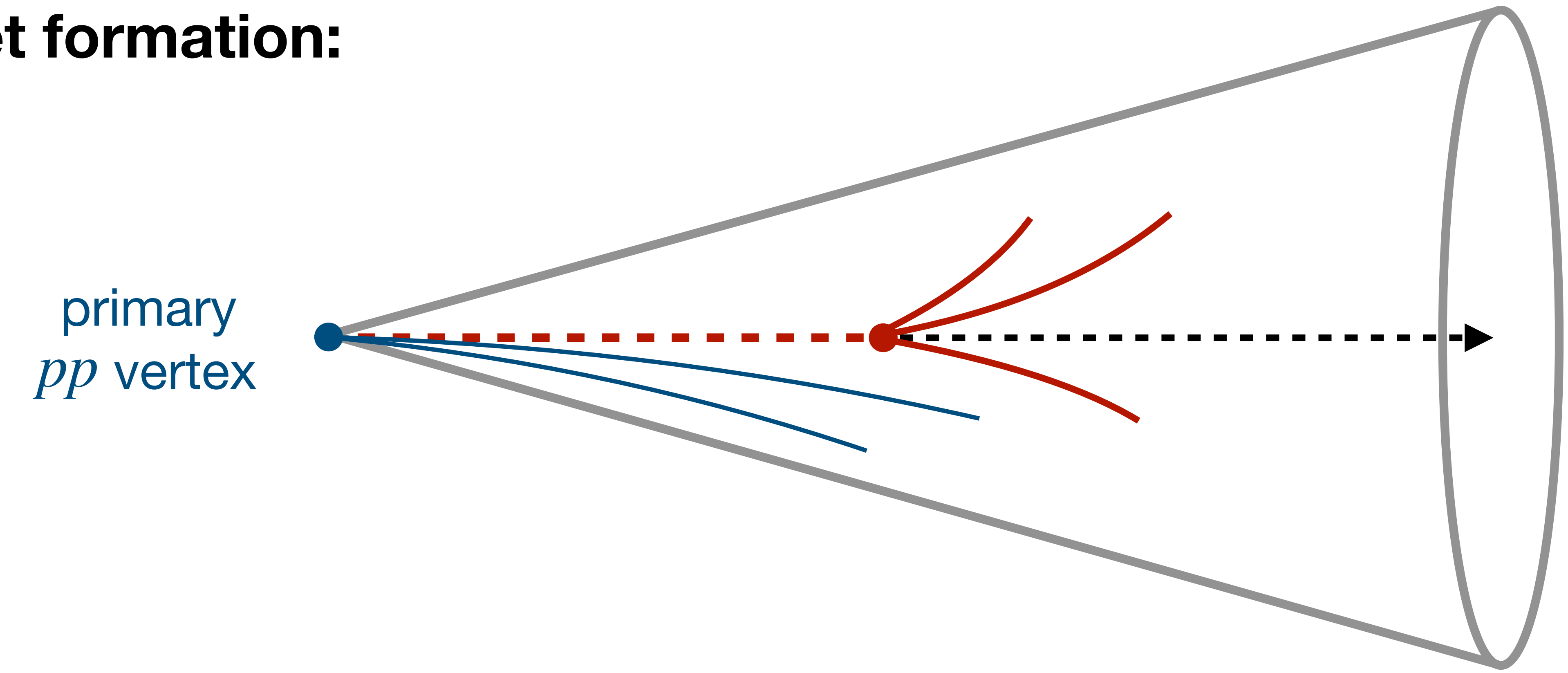


B-hadron: large quark momentum fraction



Low-level taggers

_b_-jet formation:

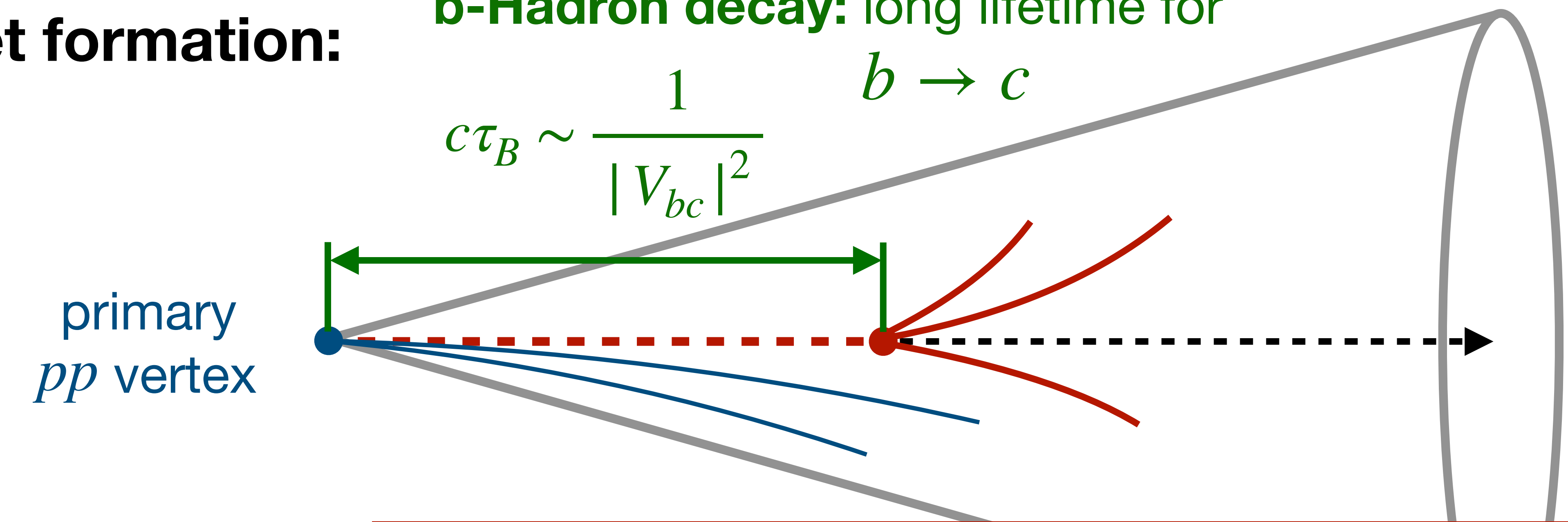


Low-level taggers

***b*-jet formation:**

b-Hadron decay: long lifetime for $b \to c$

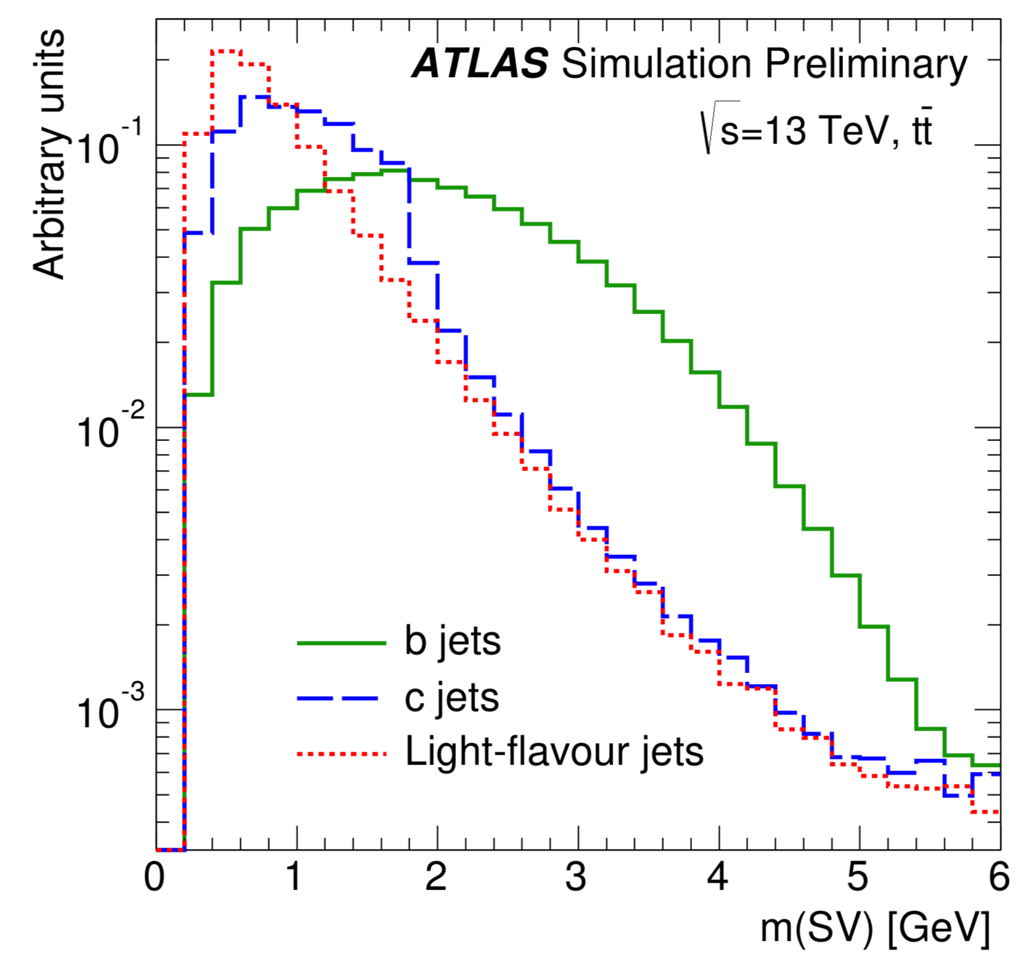
$$c\tau_B \sim \frac{1}{|V_{bc}|^2}$$



SV1

Compute properties of **S**econdary **V**ertex:

- $\boxed{\text{SV mass } m(\text{SV}) \sim m_B}$
- SV decay length significance $S_{xyz} \sim c\tau_B$
- …



Low-level taggers

***b*-jet formation:**

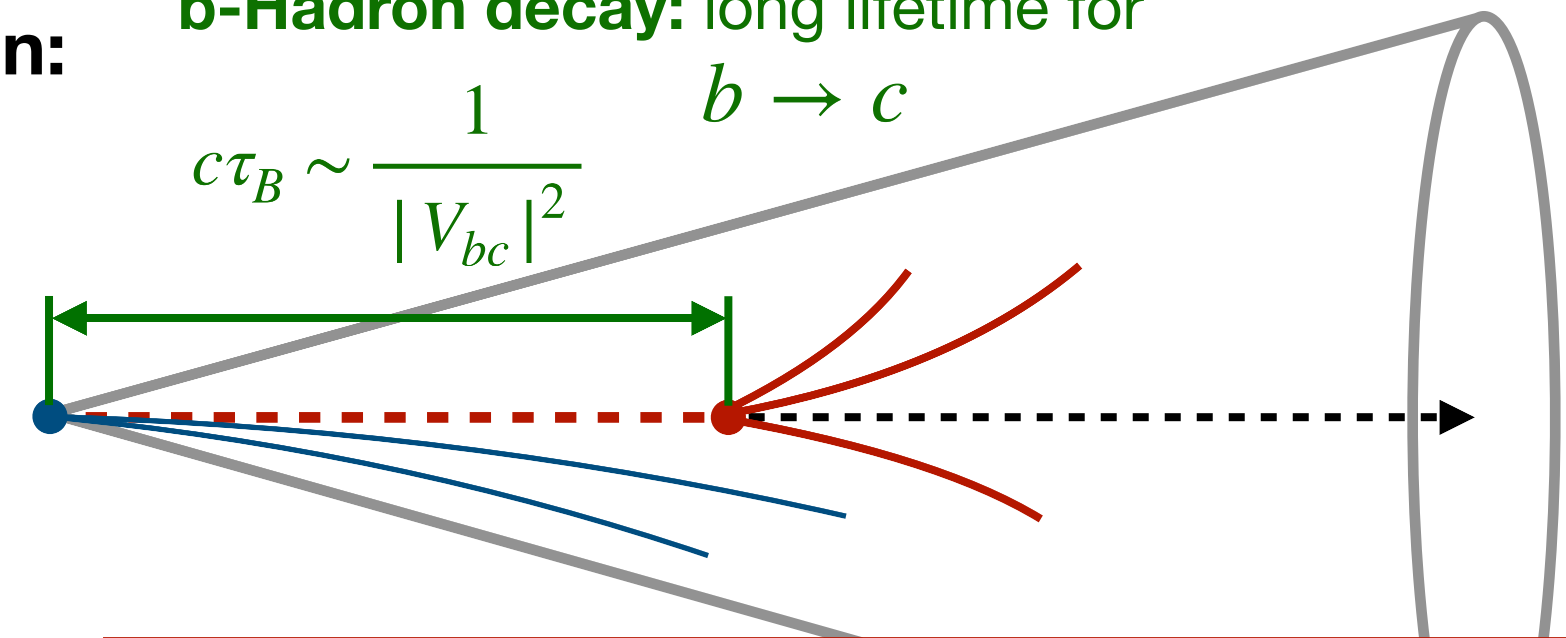
b-Hadron decay: long lifetime for

$b \to c$

$c\tau_B \sim \frac{1}{|V_{bc}|^2}$

primary

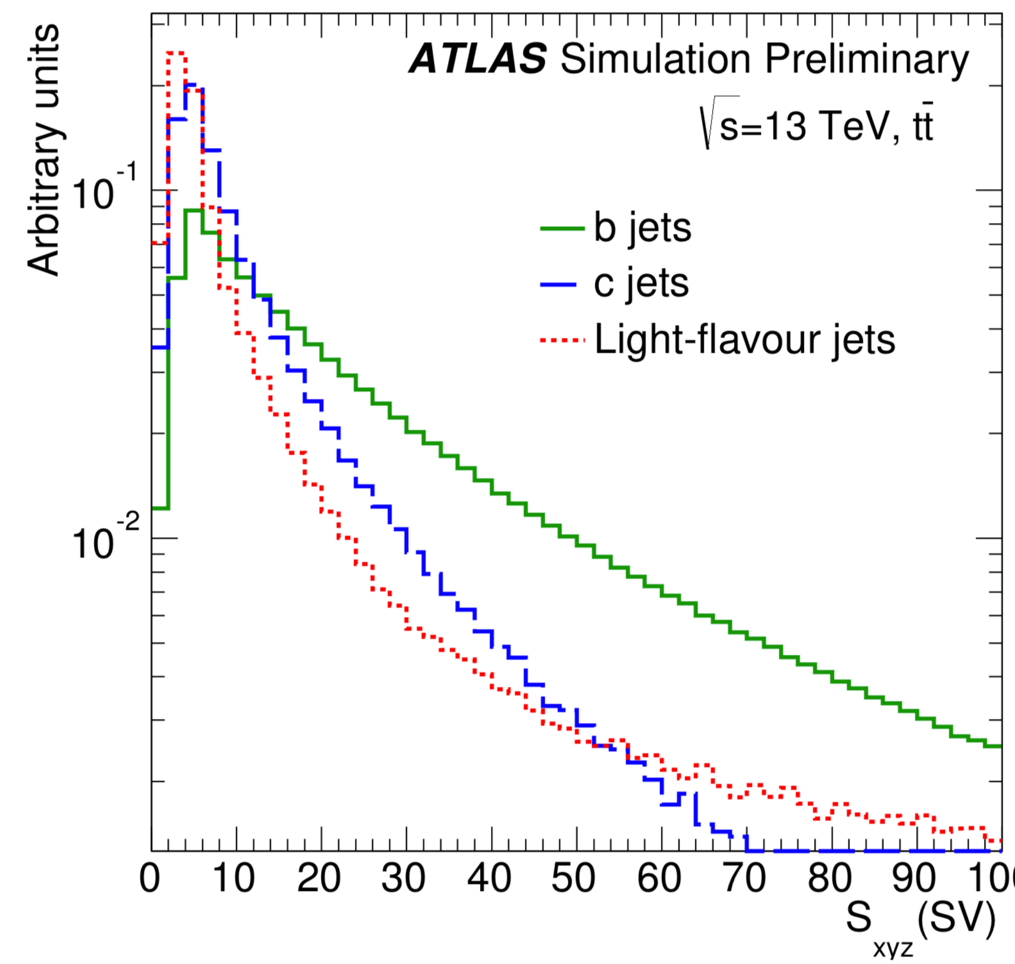
pp vertex



SV1

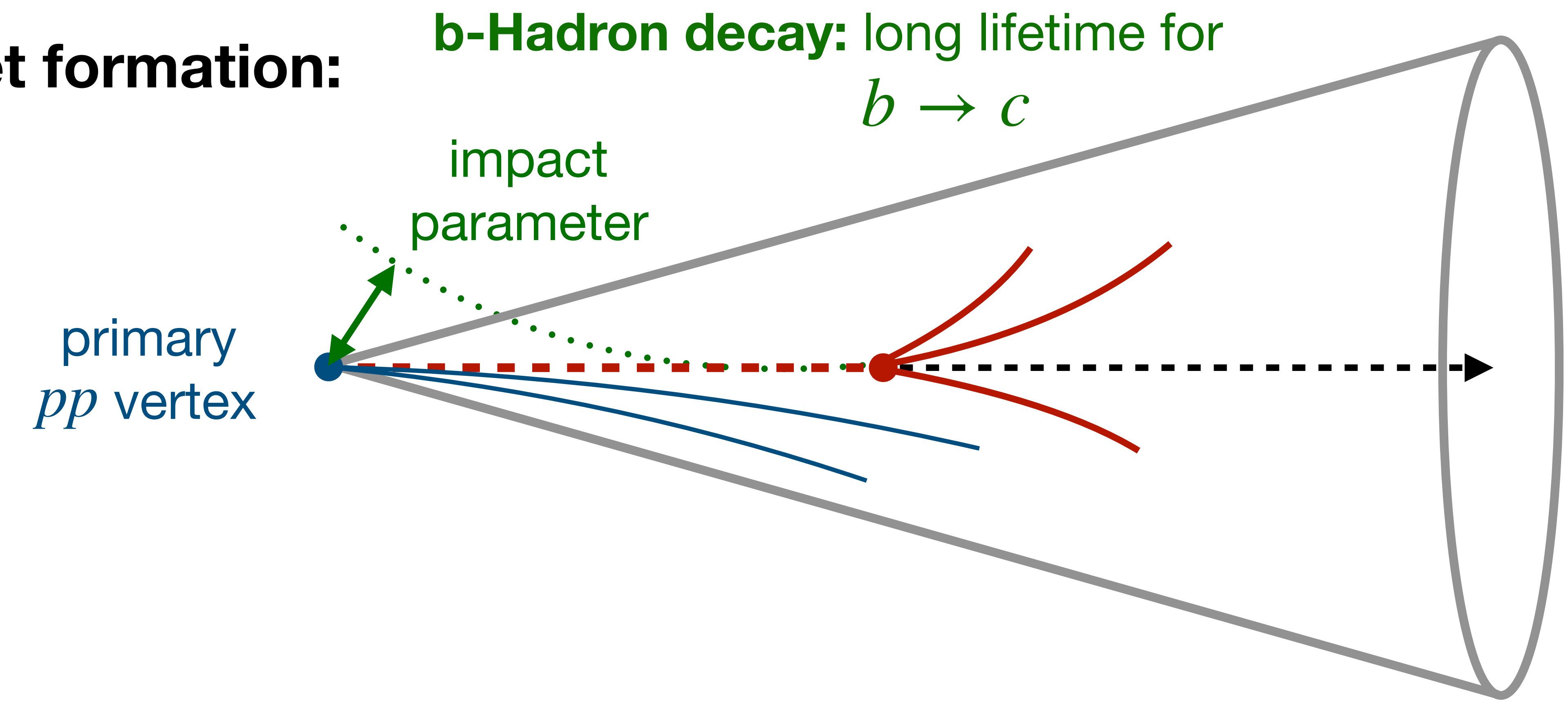
Compute properties of **S**econdary **V**ertex:

- SV mass $m(\mathrm{SV}) \sim m_B$
- SV decay length significance $S_{xyz} \sim c\tau_B$
- …



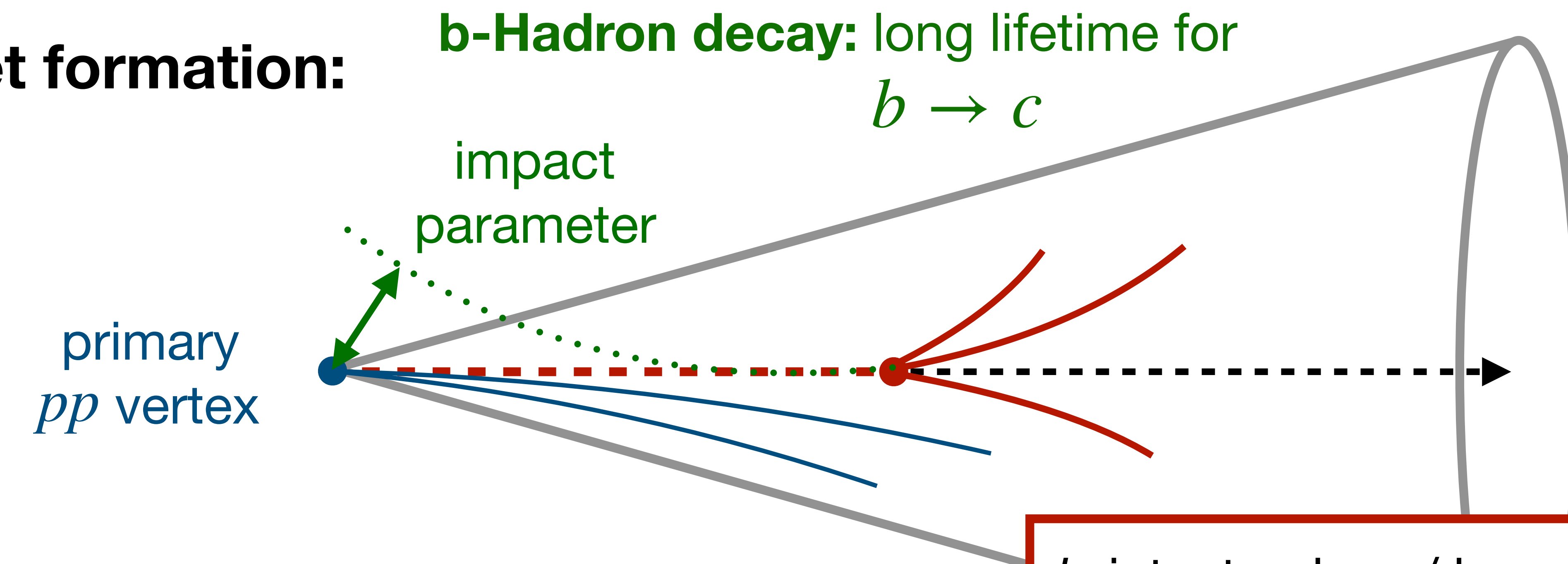
Low-level taggers

***b*-jet formation:**

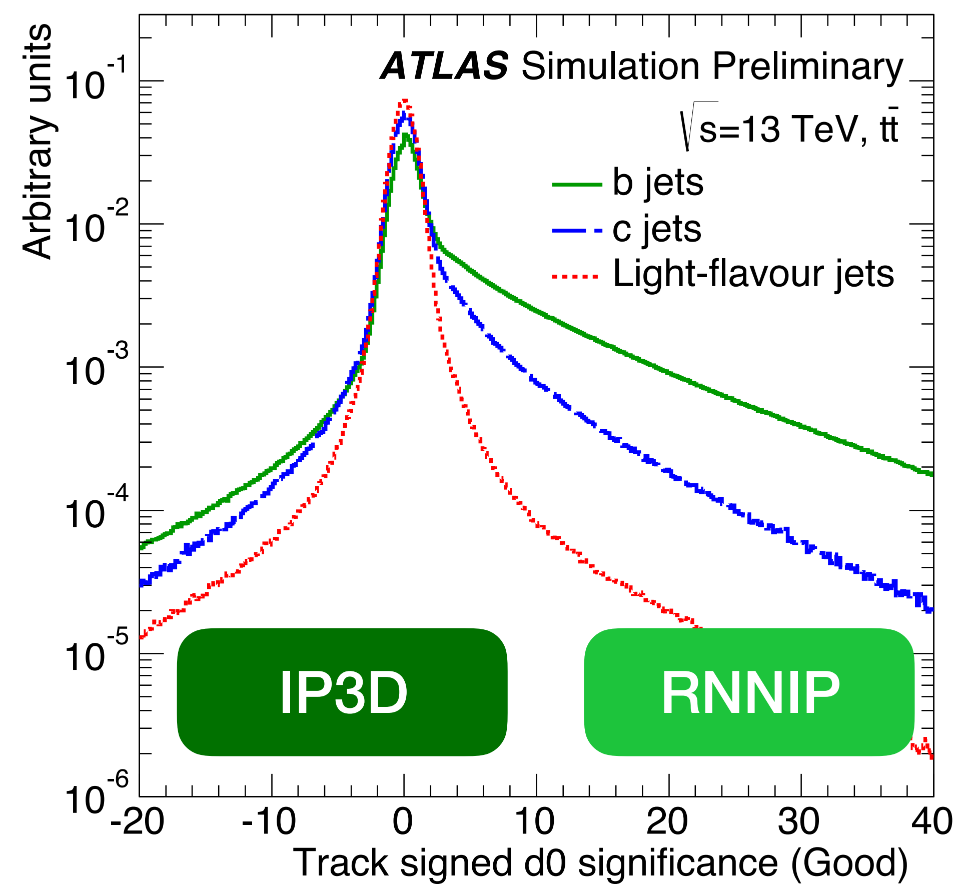


Low-level taggers

***b*-jet formation:**

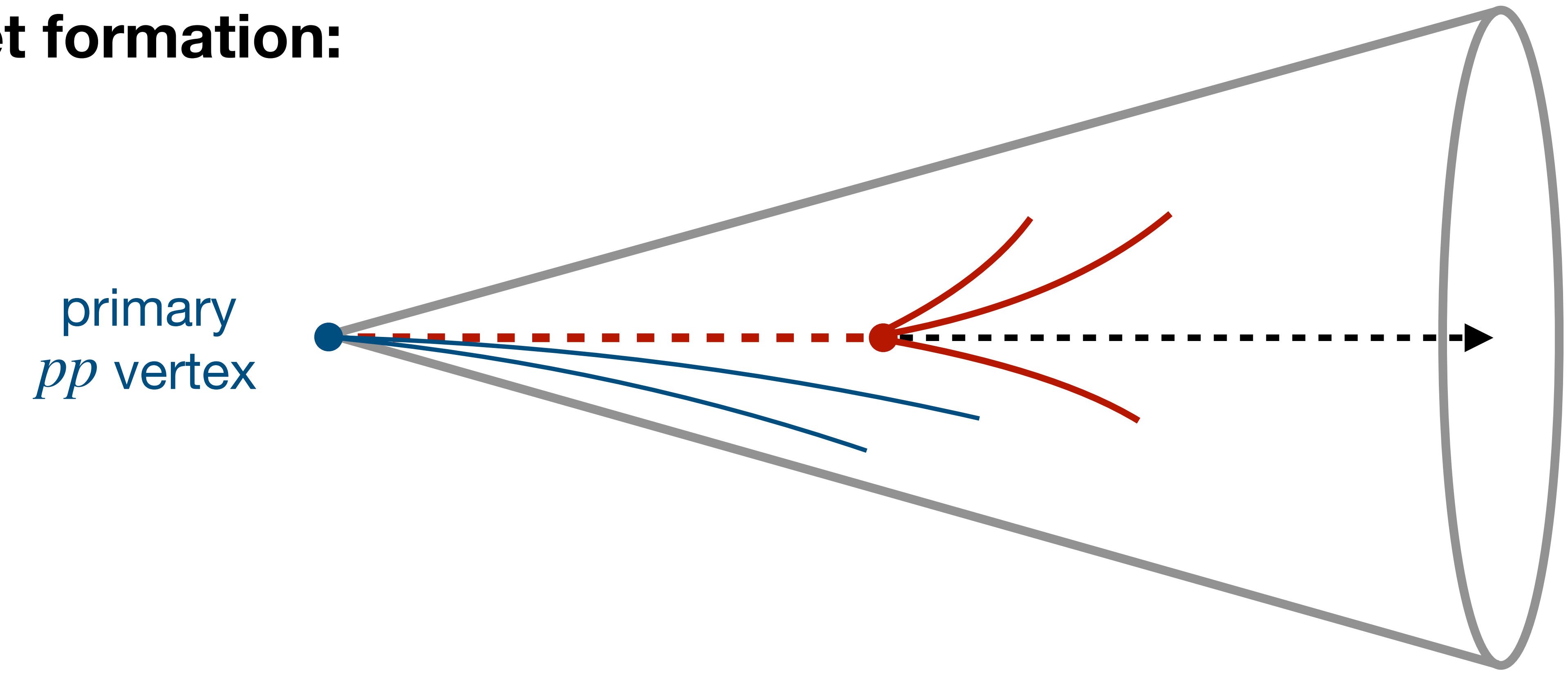


b-jets: tracks w/ large impact parameter



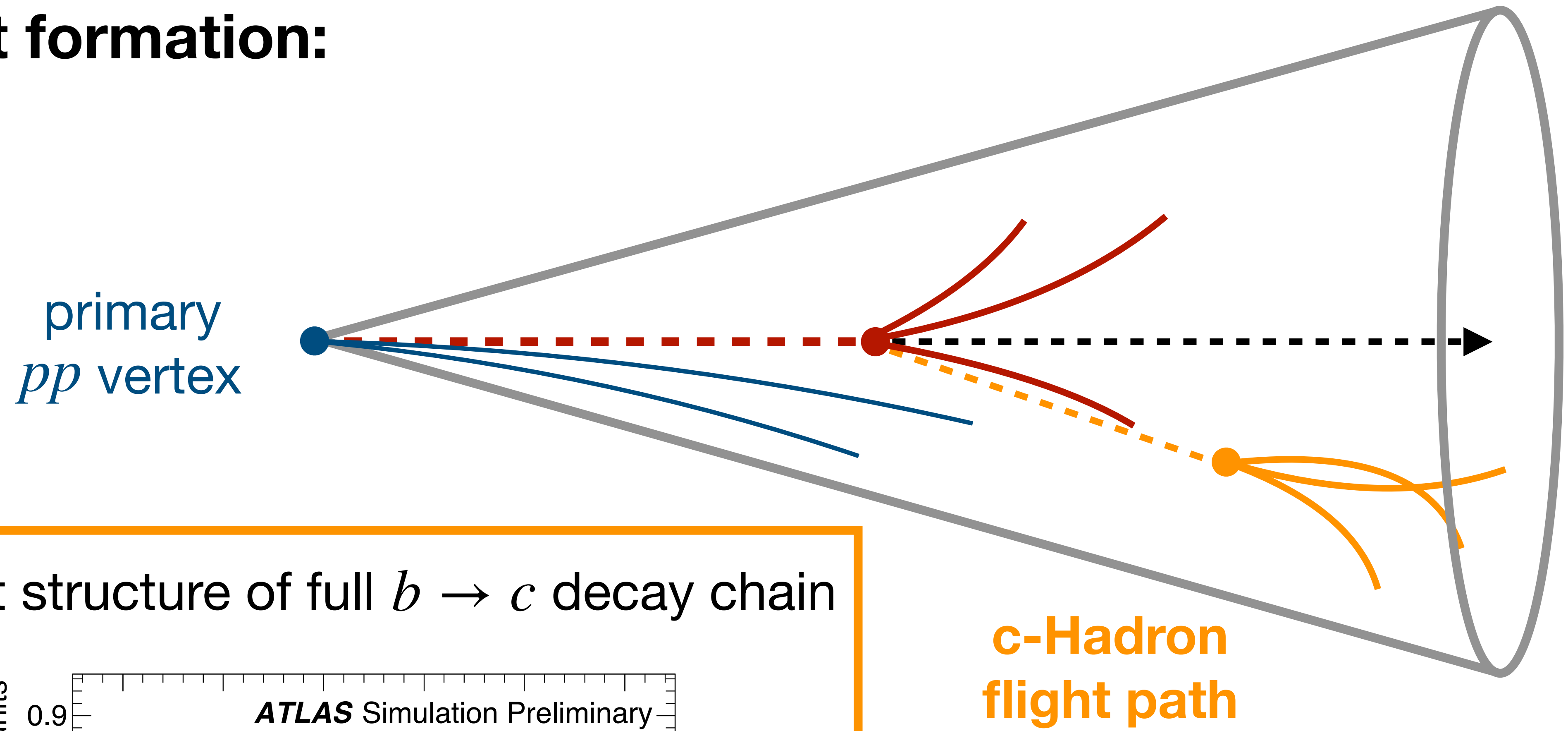
Low-level taggers

***b*-jet formation:**

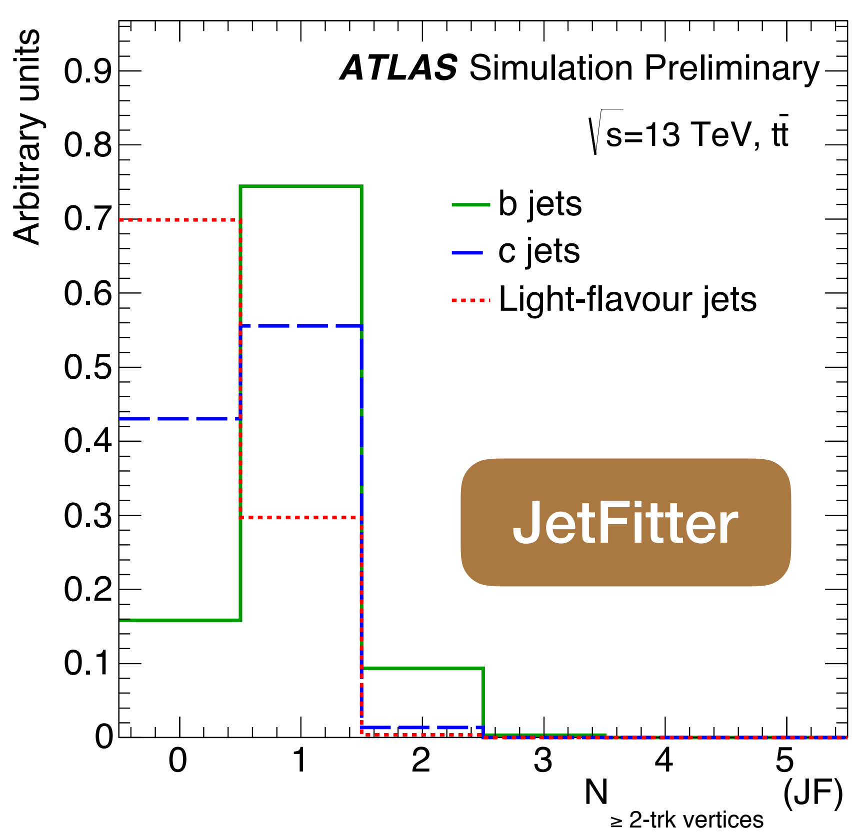


Low-level taggers

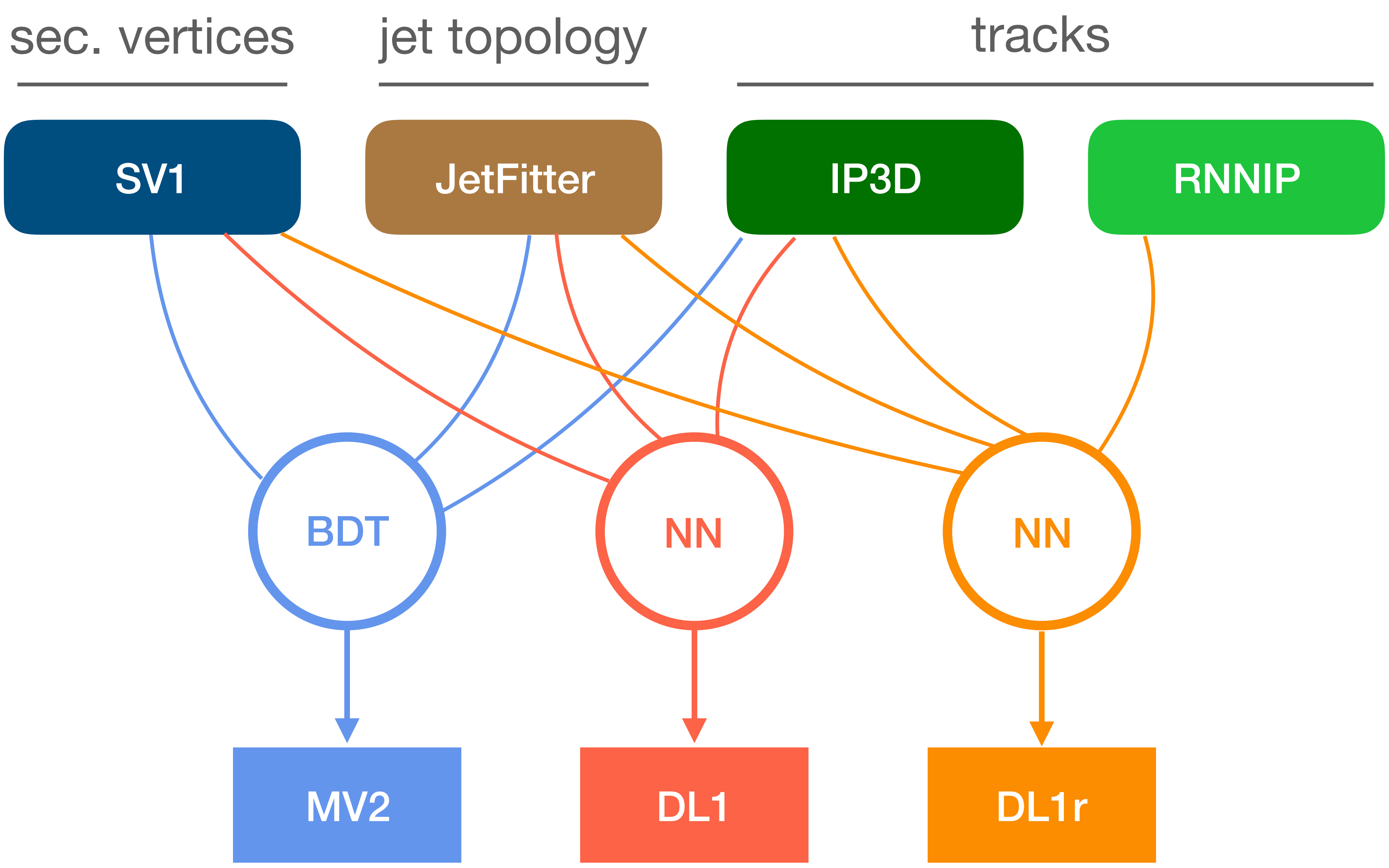
_b_-jet formation:



Exploit structure of full $b \to c$ decay chain



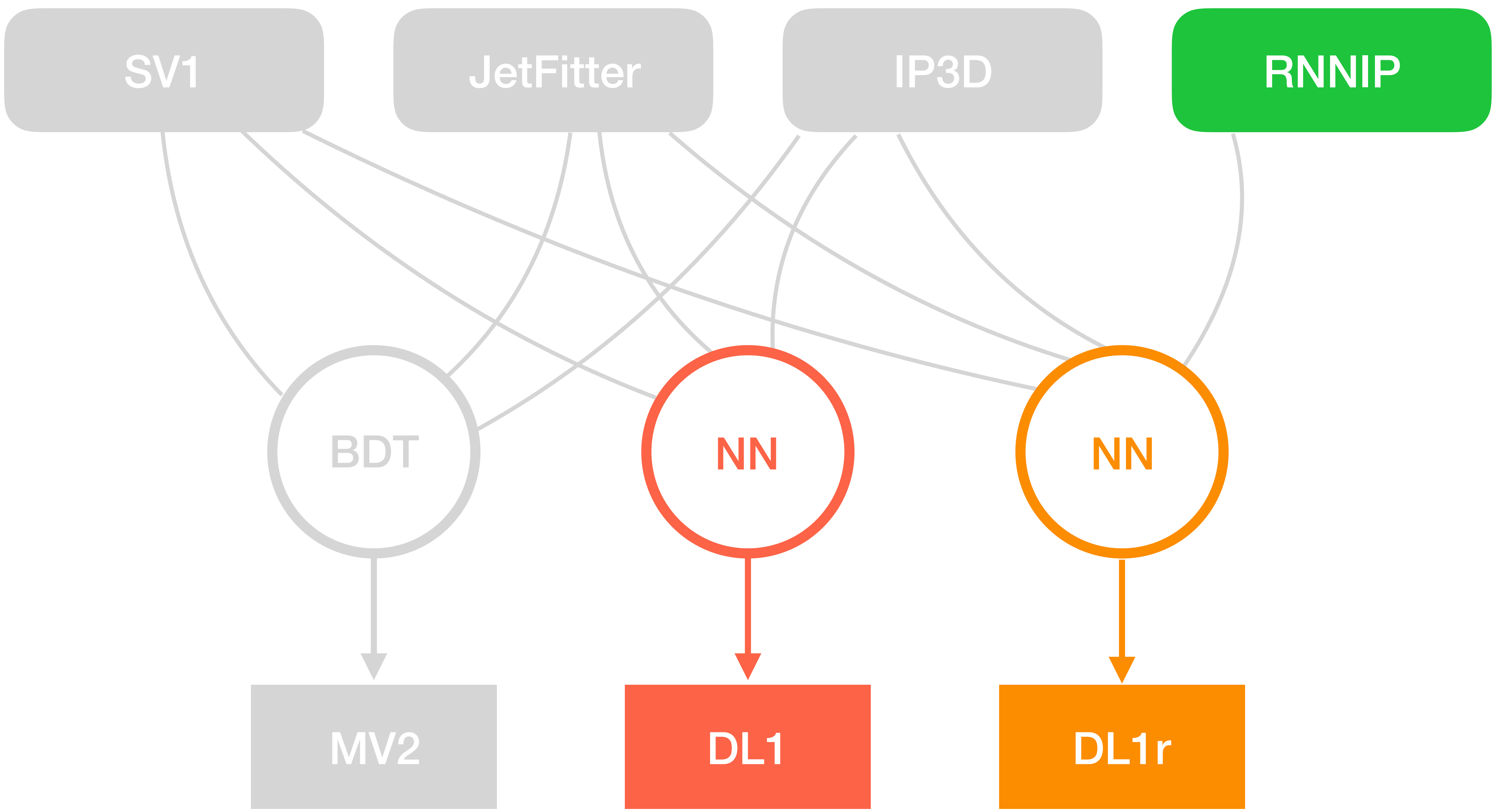
The ATLAS strategy for b-tagging



- **MV2** vs. **DL1**: different architecture, same inputs
- **DL1r**: also add RNNIP

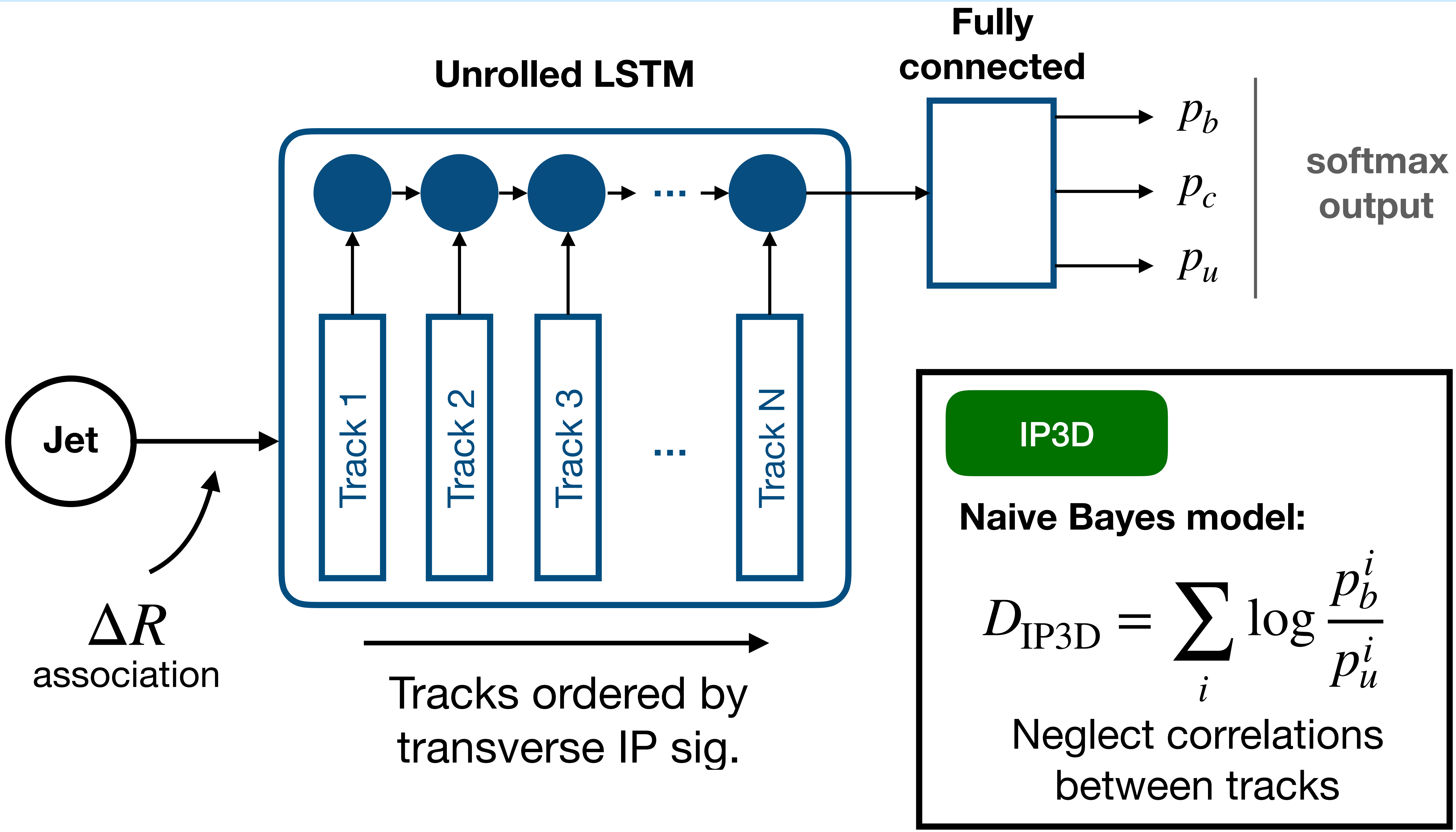
The ATLAS strategy for b-tagging

Focus on taggers with strong ML component:



- **MV2** vs. **DL1**: different architecture, same inputs
- **DL1r**: also add RNNIP

RNNIP



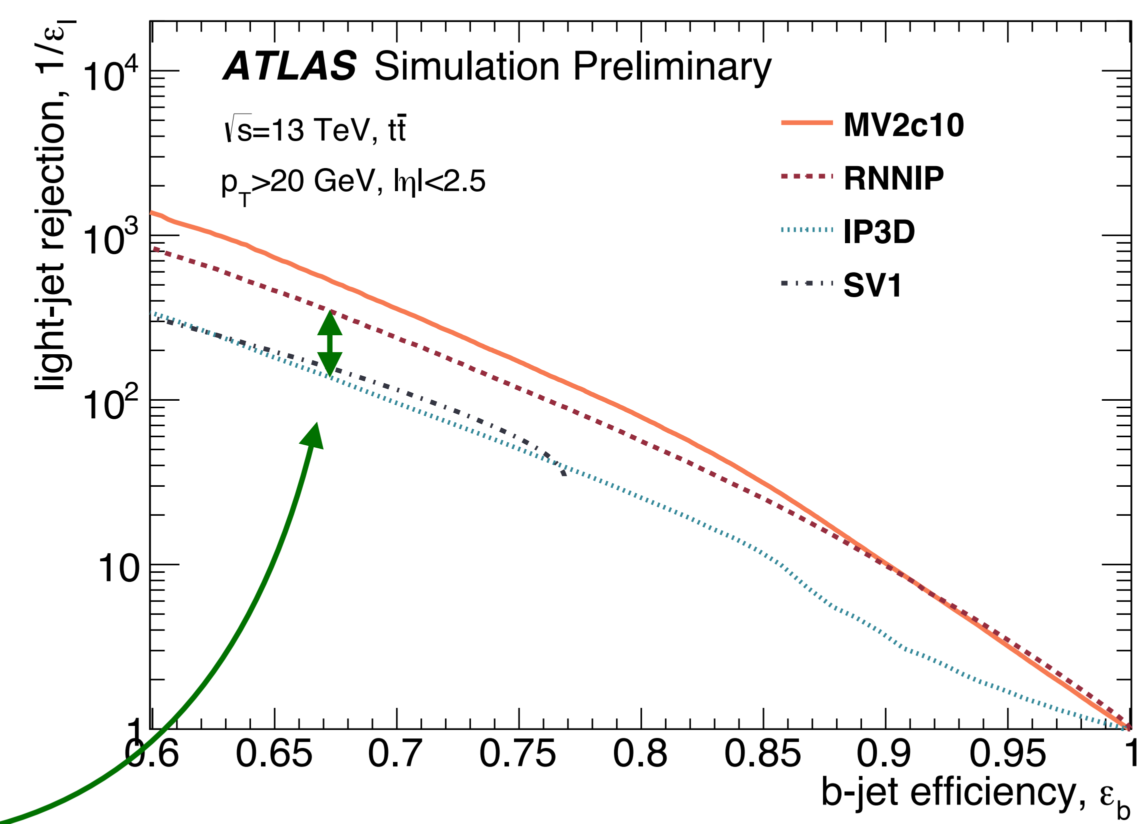
- LSTM vs. Naive Bayes: **correlations are important**!
  - … but hard to model and exploit

RNIP

Tracks contain a lot of information:

- Jet flavour discrimination:
  - Impact parameter, track momentum, …
- Track quality:
  - Number of (shared) pixel hits

(more information in backup)



Provides higher light (and charm-) rejection compared to other low-level taggers!

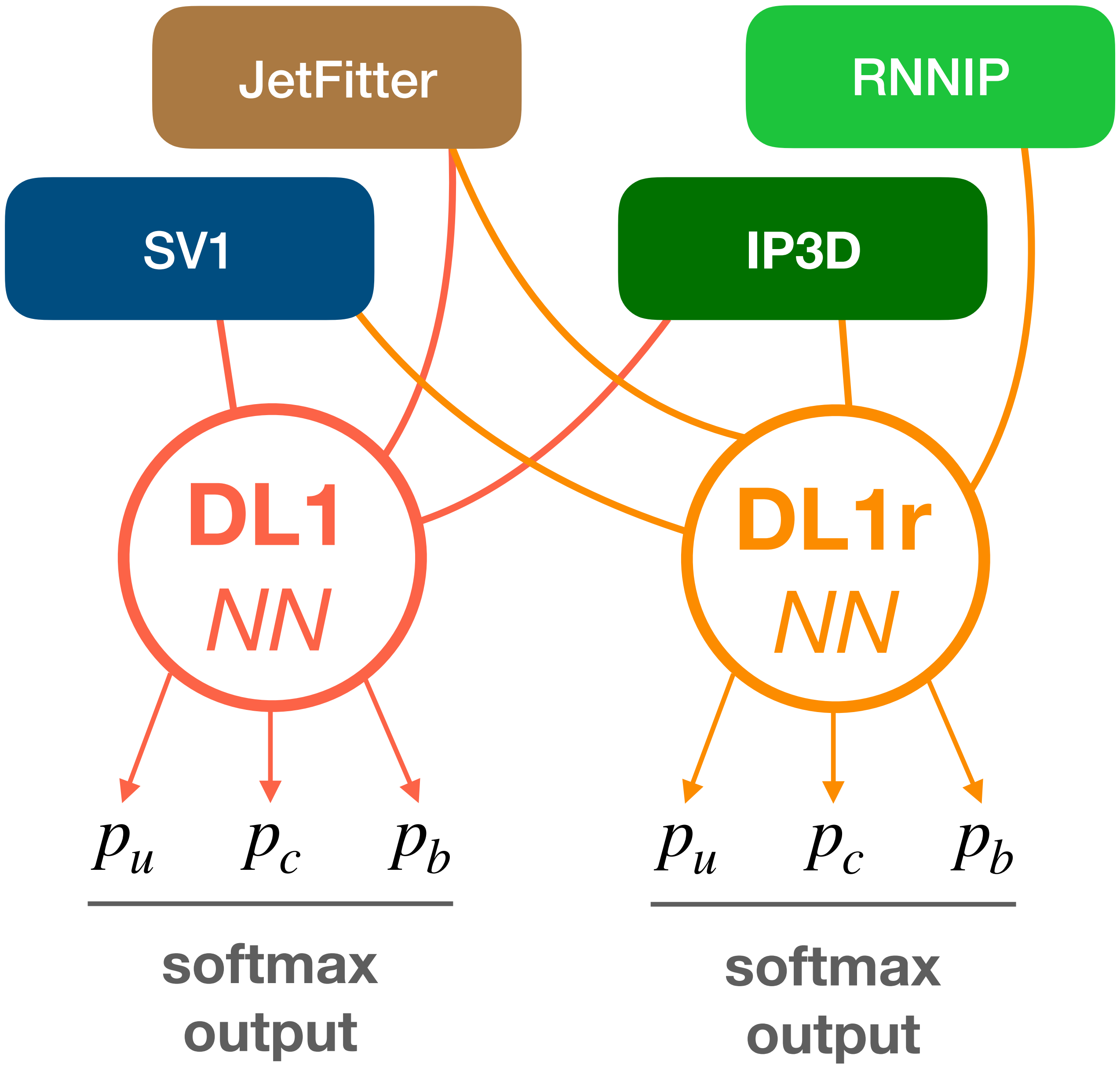
RNNIP now part of officially recommended taggers:

- Ensure that data/simulation scale factors can be reliably measured
  - Transform tagger to have same light-rejection, worse b-efficiency
  - Calibrate transformed tagger, then extrapolate to original tagger

- Fully connected (deep) neural network
- Estimate likelihood ratio of low-level tagger outputs

$$D_{\mathrm{DL1}}^{b\mathrm{-tag}} \sim \log \frac{p_b}{p_c + p_u}$$

- Supports *b*- and *c*-tagging



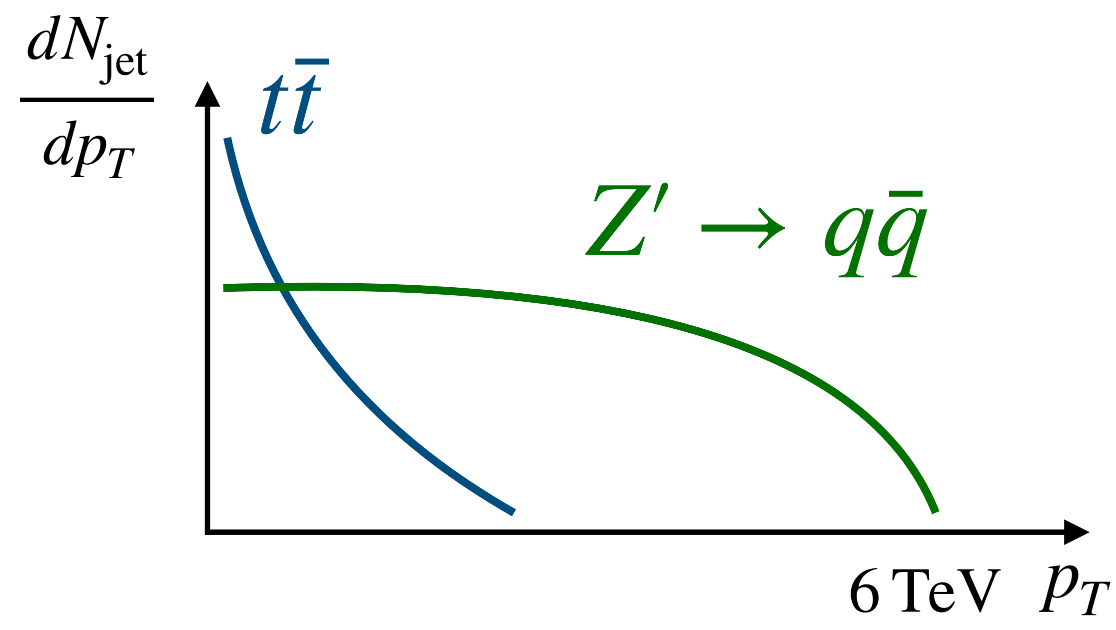
Separate trainings for both supported jet collections:
- Particle-flow jets: new ATLAS baseline
- Variable-radius track jets: invaluable for boosted decays

DL1 DL1r

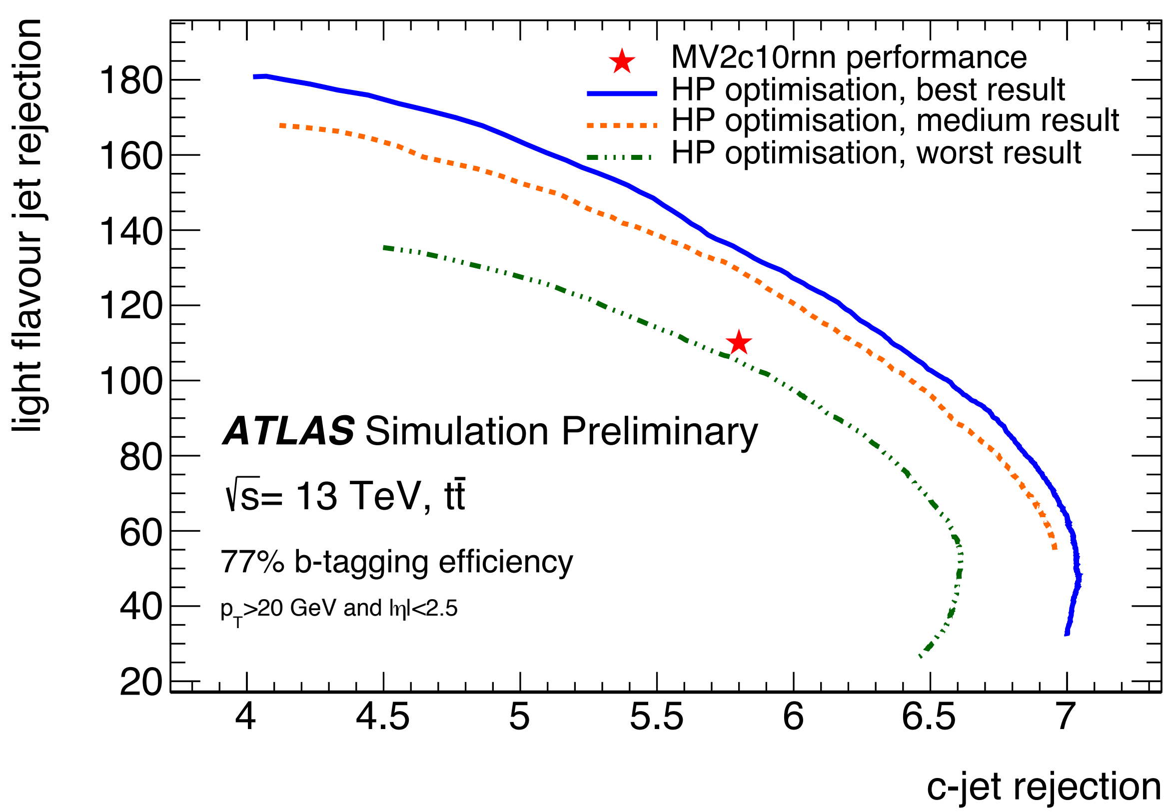
Training dataset: "hybrid" sample

- simulated $t\bar{t}$ for $p_T < 250\,\mathrm{GeV}$
- $Z' \to q\bar{q}$ for $p_T > 250\,\mathrm{GeV}$

Taggers well-behaved even for multi-TeV jets!



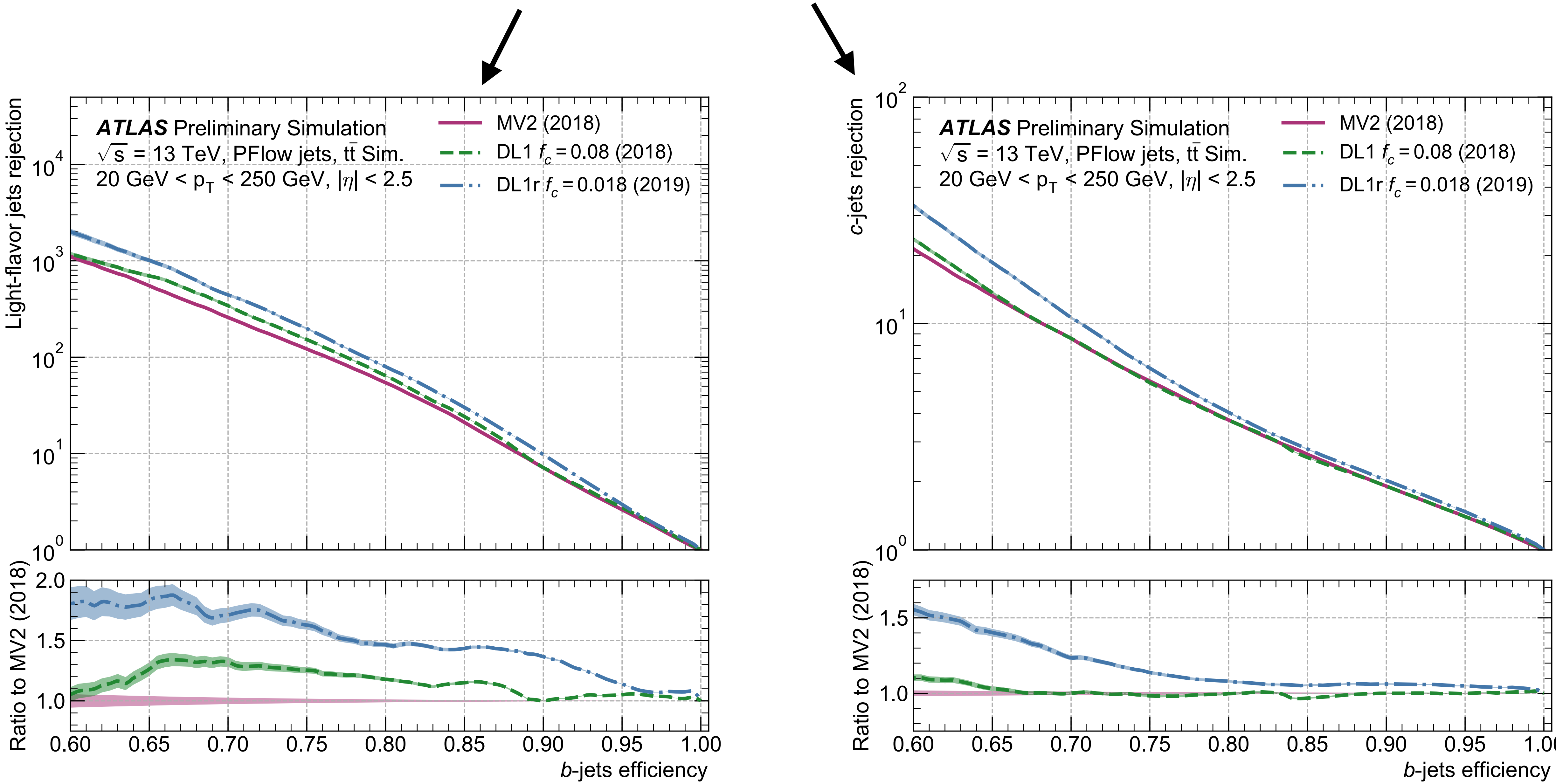
$$\left(\frac{dN_{\text{jet}}}{dp_T}\right)_{t\bar{t}} \sim \frac{1}{p_T^5} \qquad \left(\frac{dN_{\text{jet}}}{dp_T}\right)_{Z'} \sim \text{const}\,.$$



- Models implemented in Keras + Tensorflow
- More efficient training and optimisation pipeline, heavily containerised

Tagger performance in simulation

- **MV2** ⟶ **DL1**: very similar performance
- **DL1** ⟶ **DL1r**: adding RNNIP (+ optimising network architecture) significantly improves light-and charm rejection

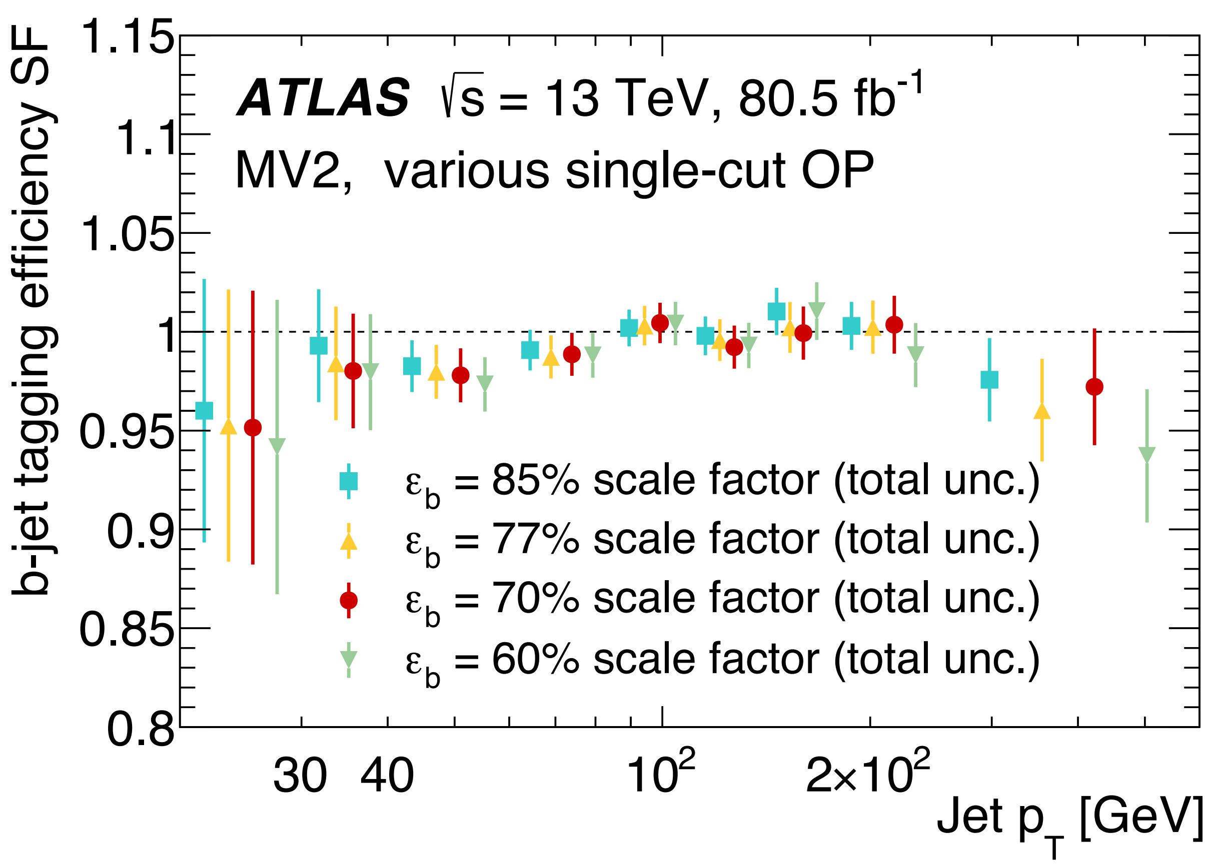


Tagger performance on data

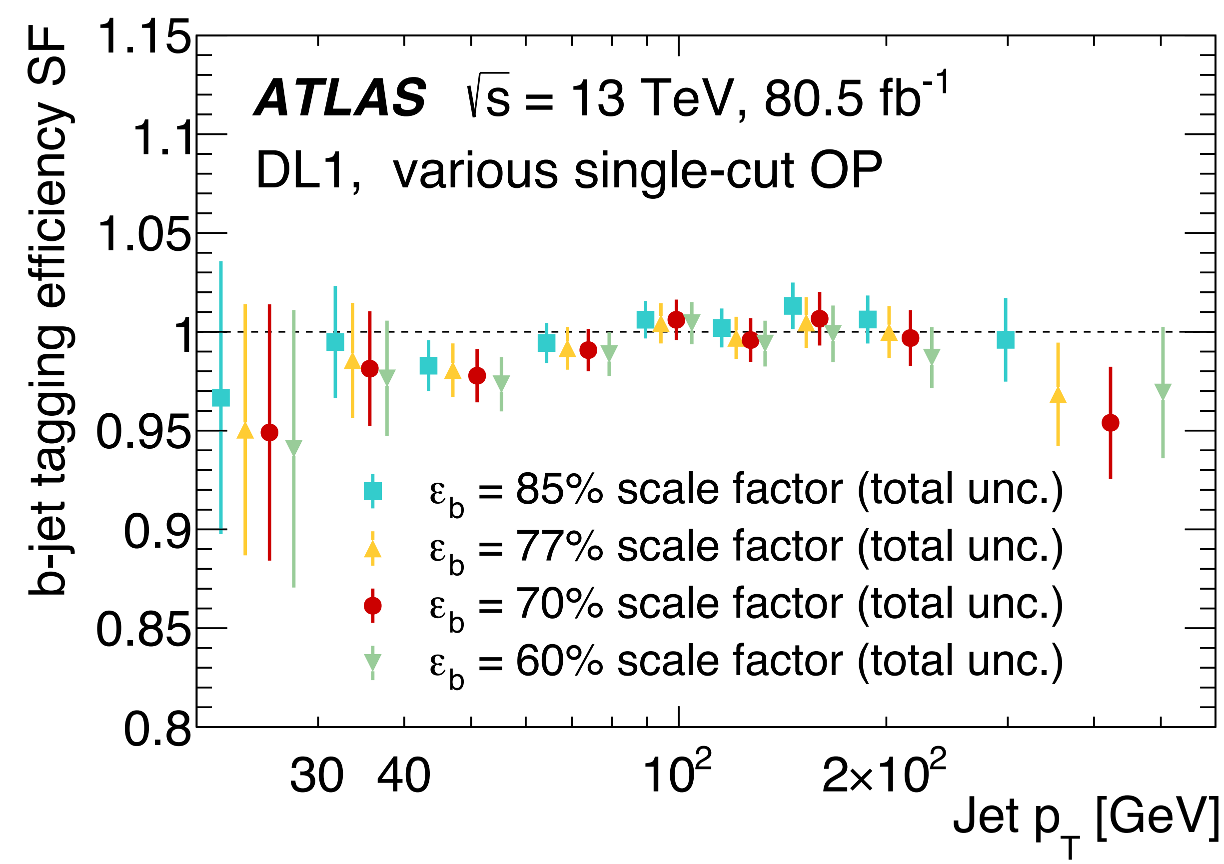
- Good modelling is *essential* for training
- Scale factor determination very complex
  - Measure efficiency in data for b, c, light jets
  - b-SF: precision top measurement!!

$$\mathrm{SF}_b = \frac{\epsilon^{\mathrm{data}}_{b-\mathrm{jets}}}{\epsilon^{\mathrm{sim.}}_{b-\mathrm{jets}}}$$

(more plots in backup)



MV2



DL1 (2018)

Summary

Two-stage approach:

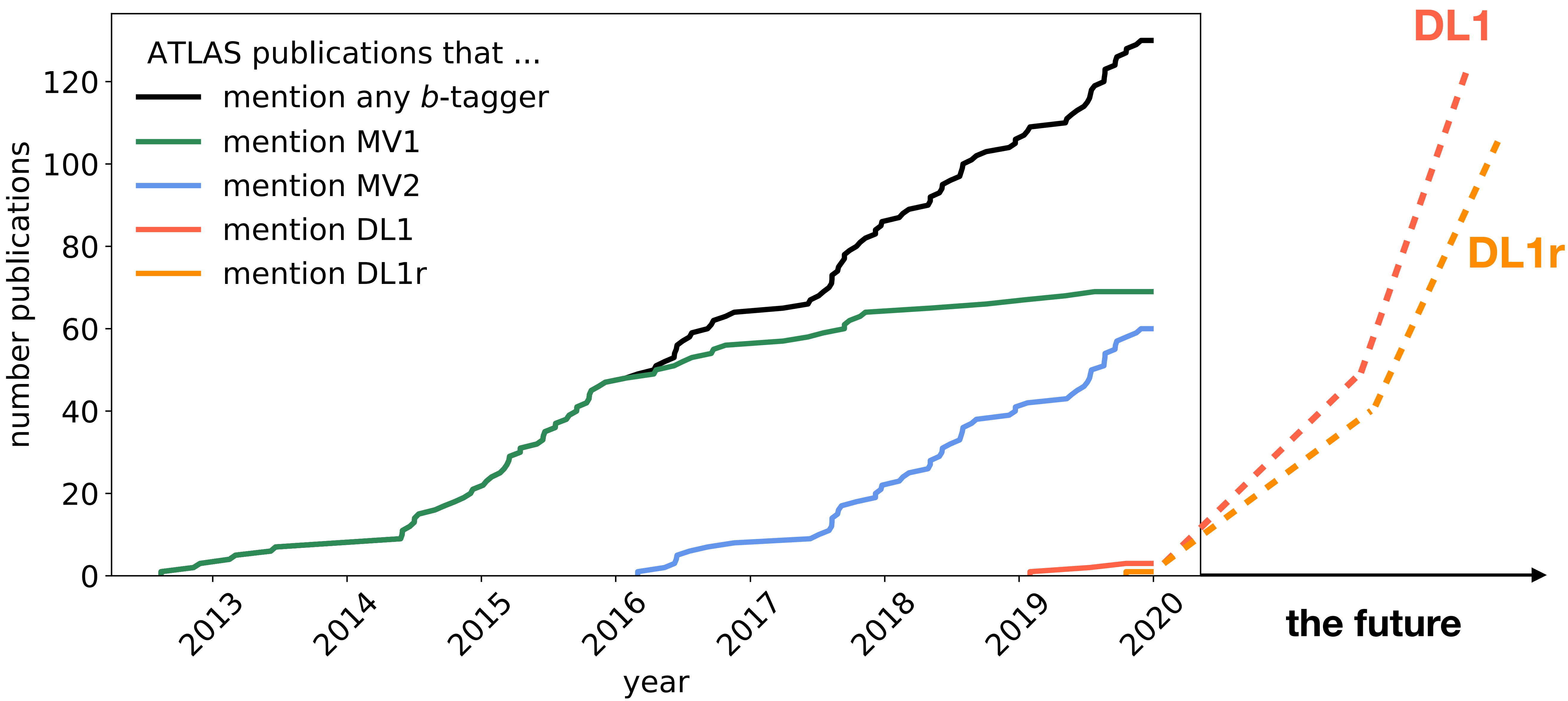
- Robust, physics-driven low-level taggers
- Detector-agnostic, ML-based high-level taggers

New LSTM-based low-level tagger

Improved training pipeline for high-level taggers

Significant performance gain compared to Run-2 baseline tagger

The future



https://gitlab.cern.ch/phwindis/arxivscraper

Backup

References

(1) Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high pT jets, ATL-PHYS-PUB-2014-008

(2) Expected performance of the ATLAS b-tagging algorithms in Run-2, ATL-PHYS-PUB-2015-022

(3) ATLAS b-jet identification performance and efficiency measurement with ttbar events in pp collisions at 13 TeV, Eur. Phys. J. C 79 (2019) 970

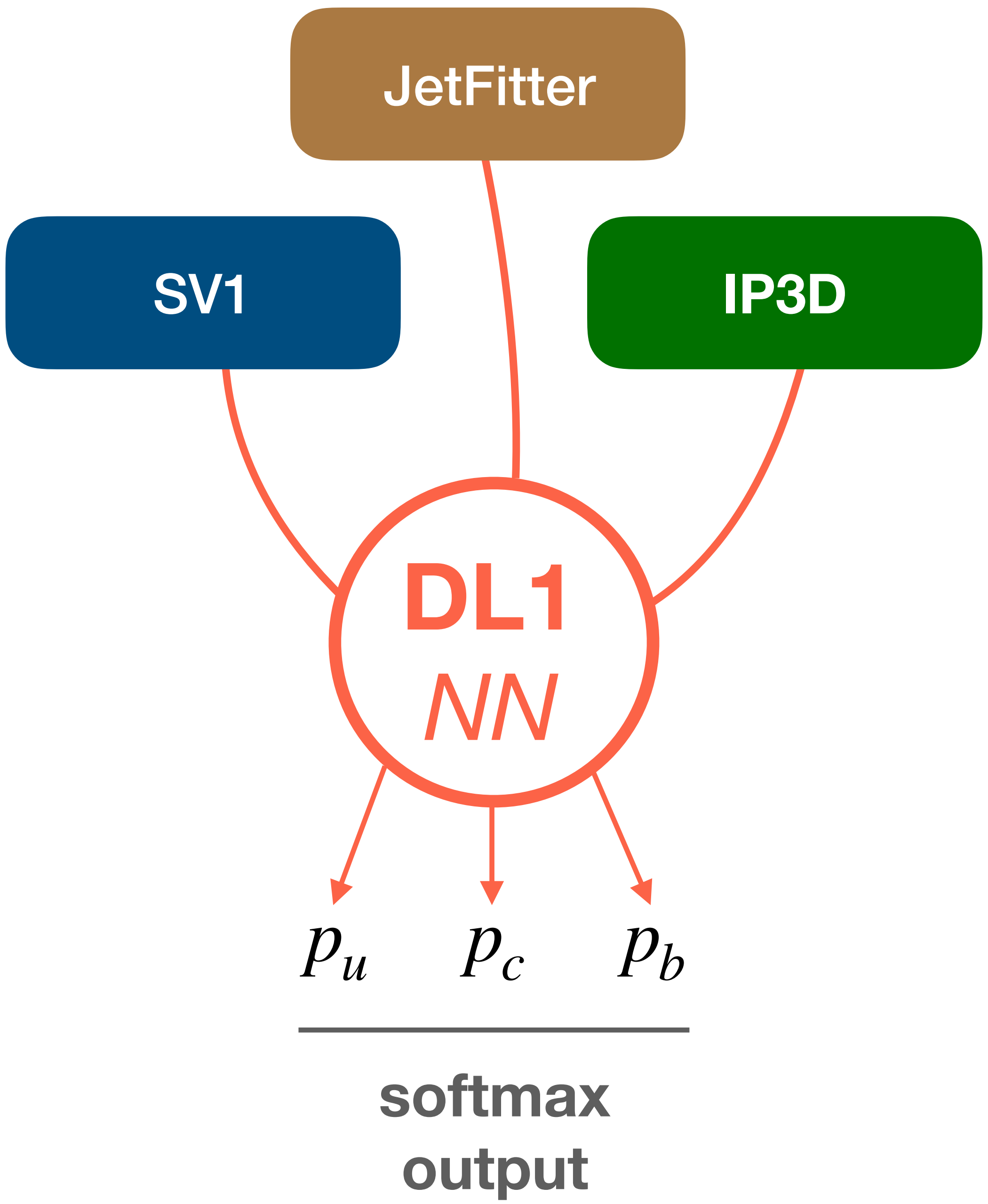
(4) Expected performance of the 2019 ATLAS b-taggers, ATL-FTAG-2019-005

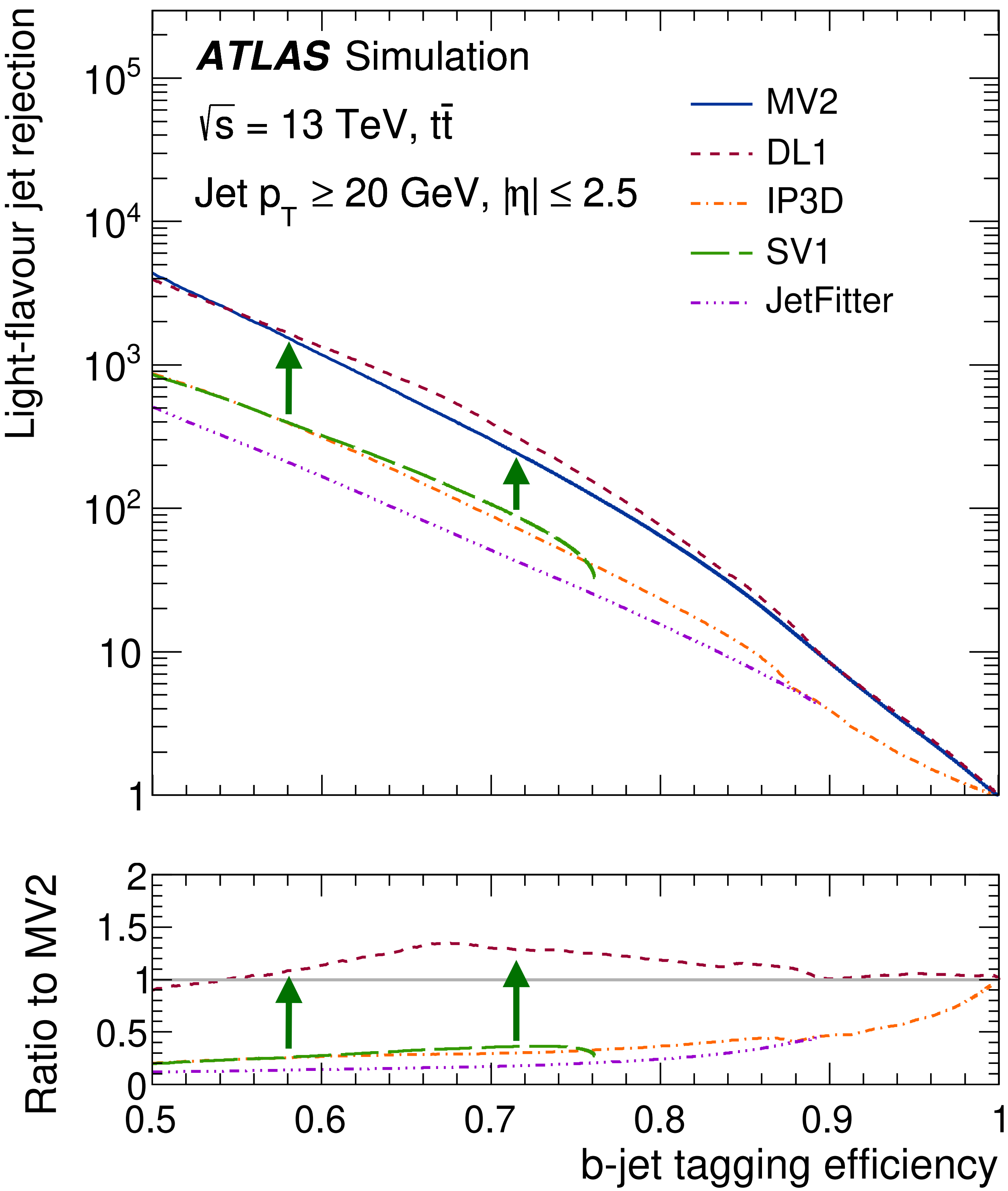
(5) ATLAS flavour-tagging calibration results with 139 ifb, ATL-FTAG-2019-004

(6) Hyper-parameter scan with the Deep Learning heavy-flavour tagger (DL1), ATL-FTAG-2019-001

(7) Machine learning algorithms for *b*-jet tagging at the ATLAS experiment, ATL-PHYS-PROC-2017-211

Improvement from high-level taggers

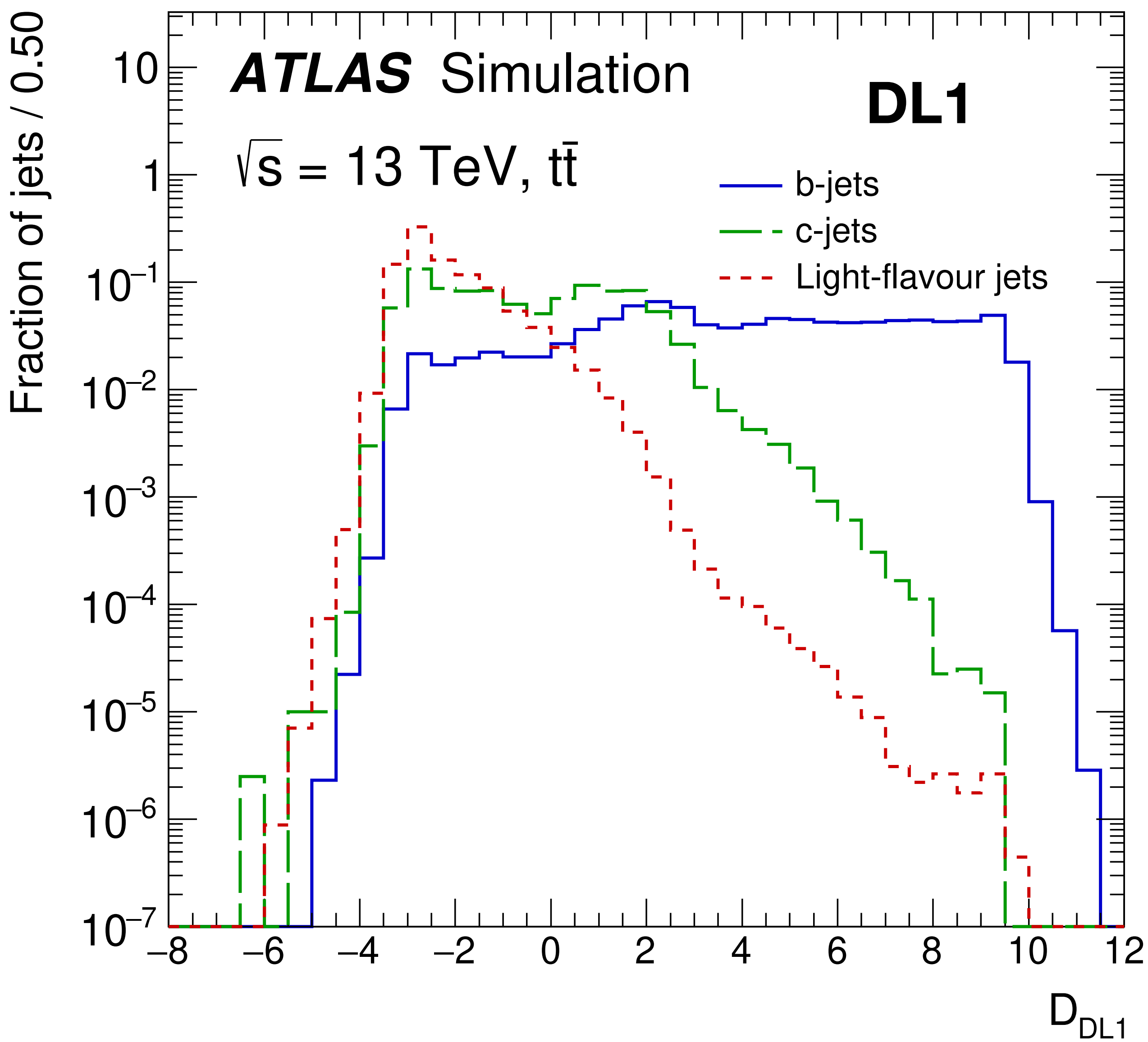




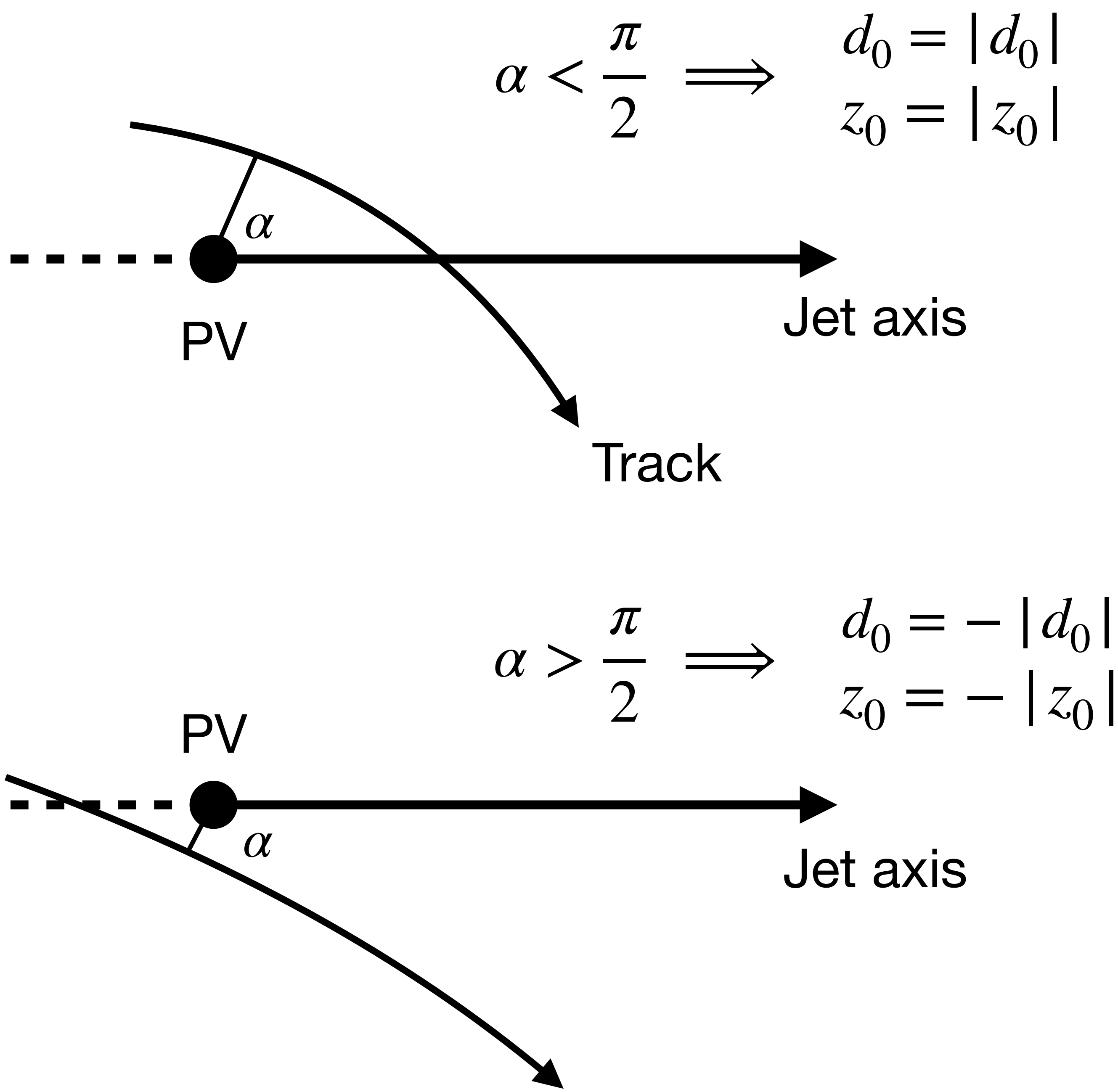
Combination of low-level taggers leads to huge performance gain!

Tagger performance in simulation

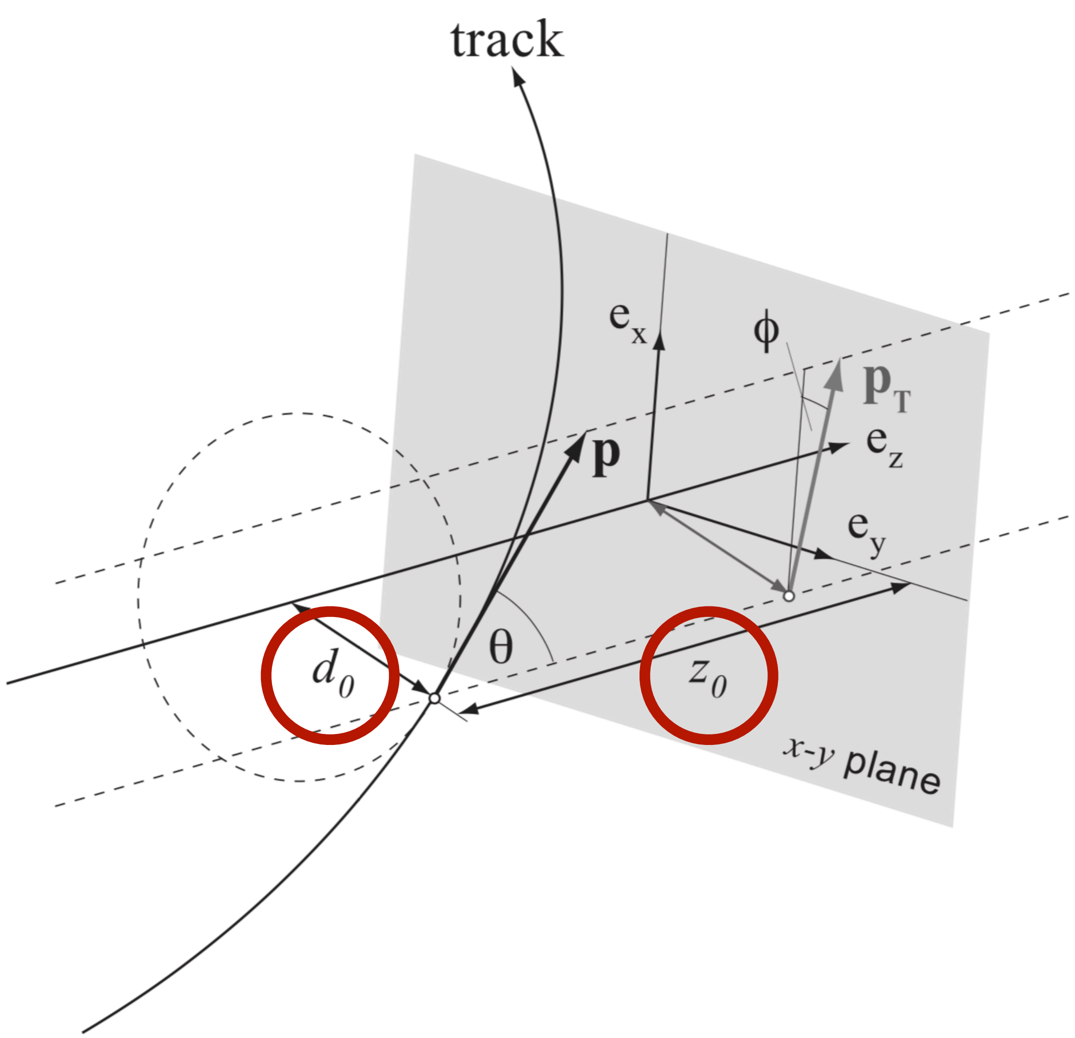
- **MV2** ⟶ **DL1**: very similar performance
- **DL1** ⟶ **DL1r**: adding RNNIP (+ optimising network architecture) significantly improves light-and charm rejection



IP sign convention



Lifetime sign convention

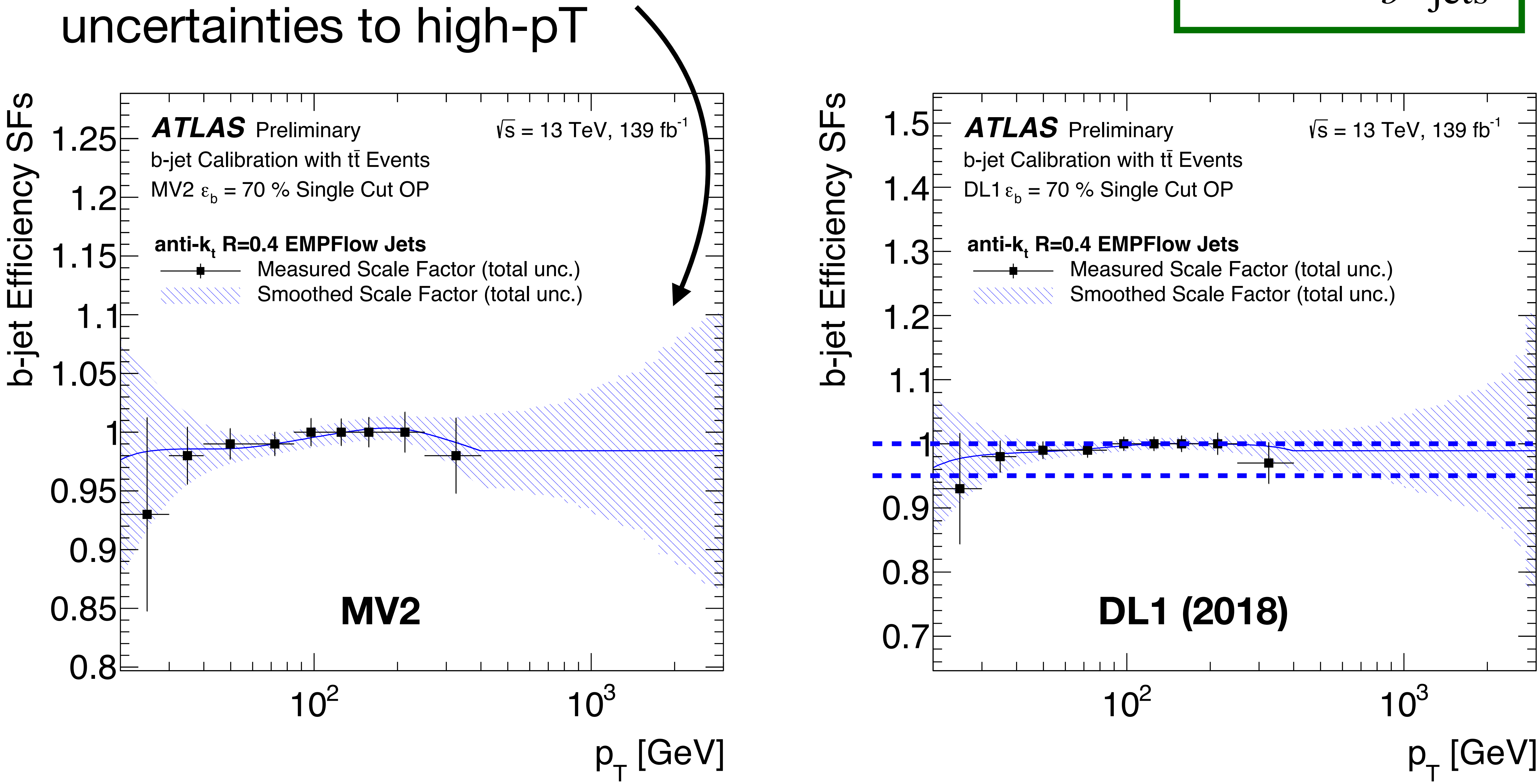


Track parametrisation around interaction point

Tagger performance in data

- b-SF: top-precision measurement!!
  - Uncertainties O(1%)
- Simulation-based extrapolation of uncertainties to high-pT

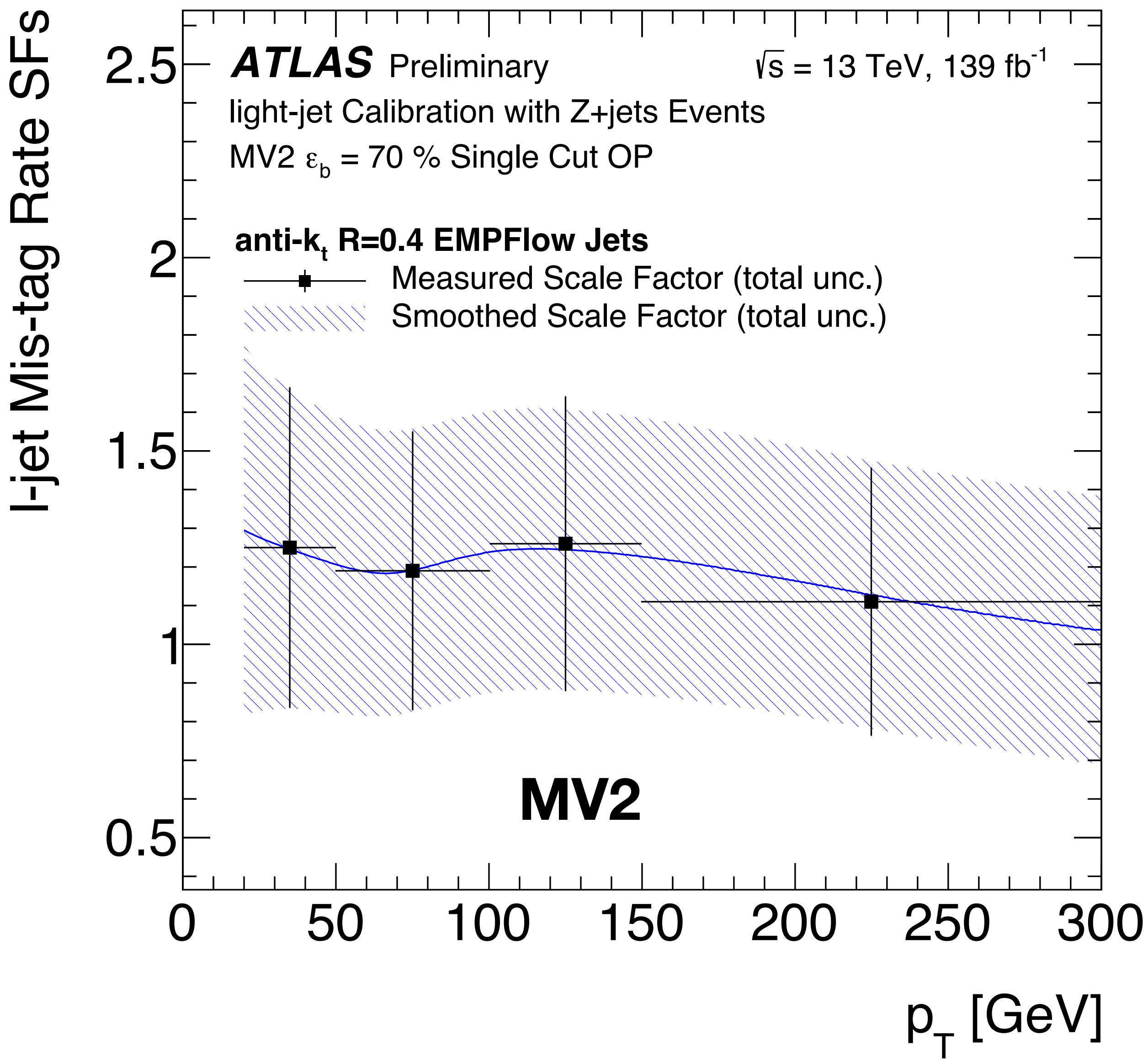
$$\mathrm{SF}_b = \frac{\epsilon^{\mathrm{data}}_{b\text{-jets}}}{\epsilon^{\mathrm{sim.}}_{b\text{-jets}}}$$

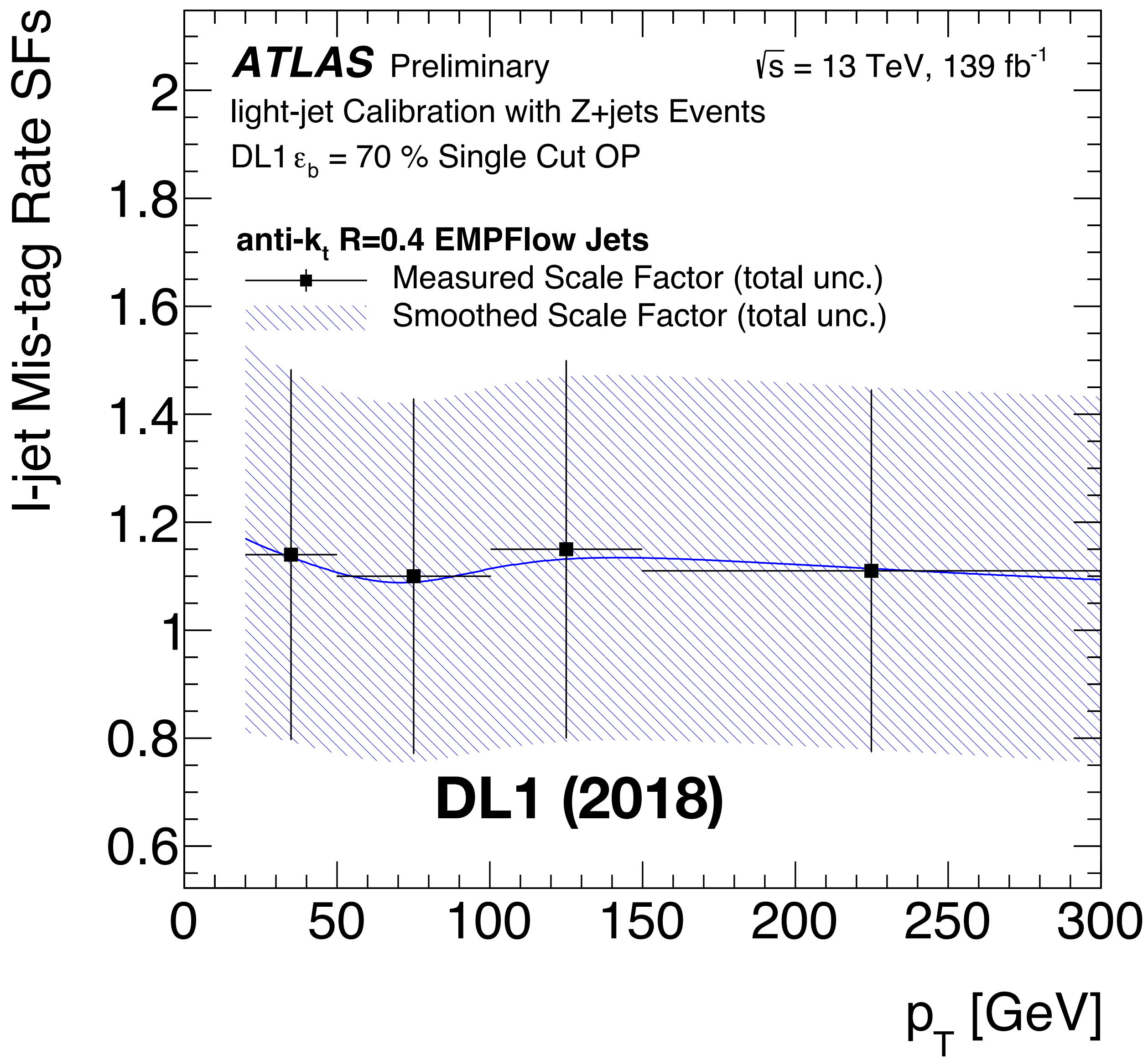


Tagger performance in data

- light-SF measured in Z + jets
- Calibrate “flipped” tagger, then extrapolate to nominal tagger

$$\boxed{SF_b = \frac{\epsilon^{\text{data}}_{b-\text{jets}}}{\epsilon^{\text{sim.}}_{b-\text{jets}}}}$$

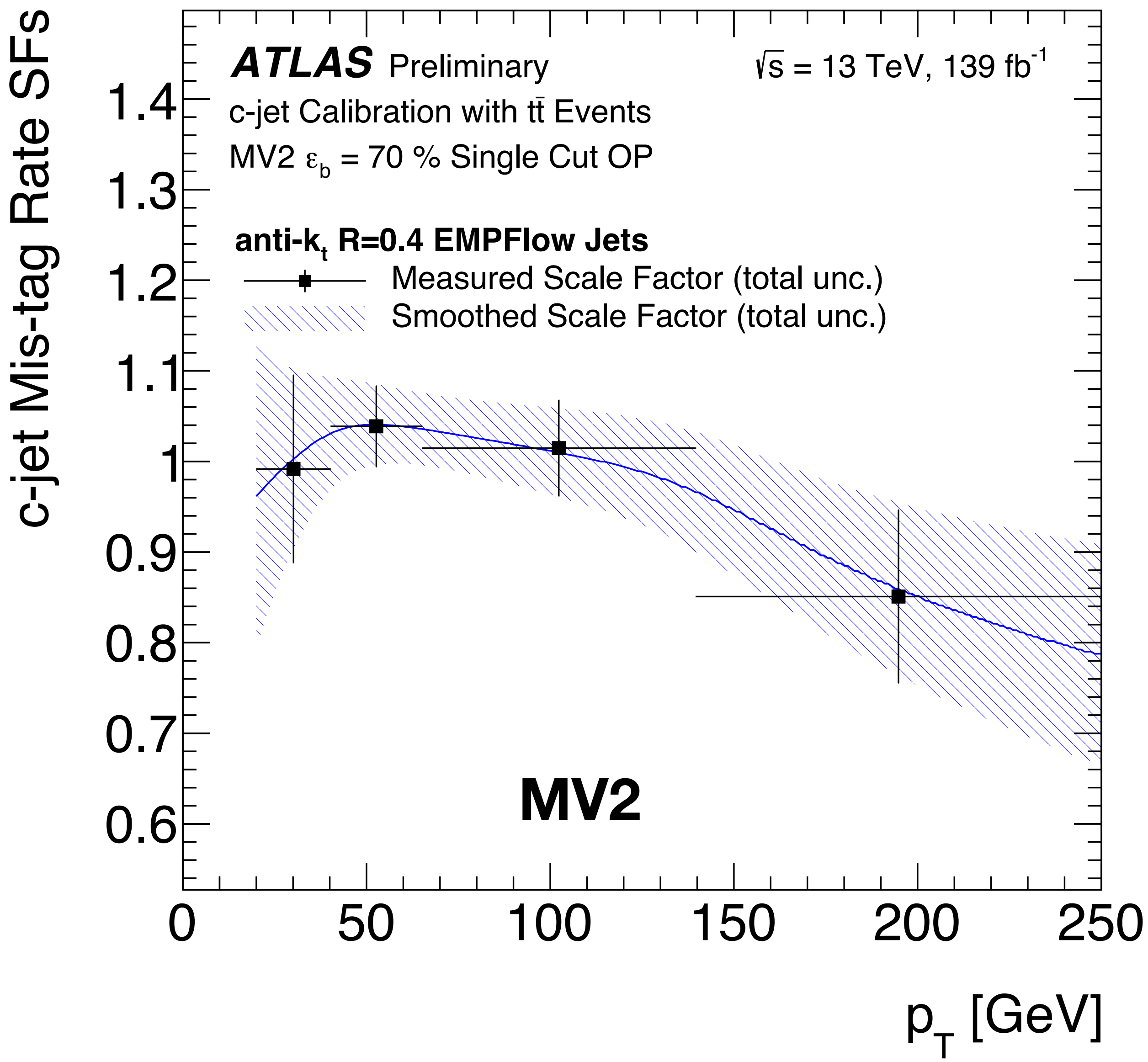


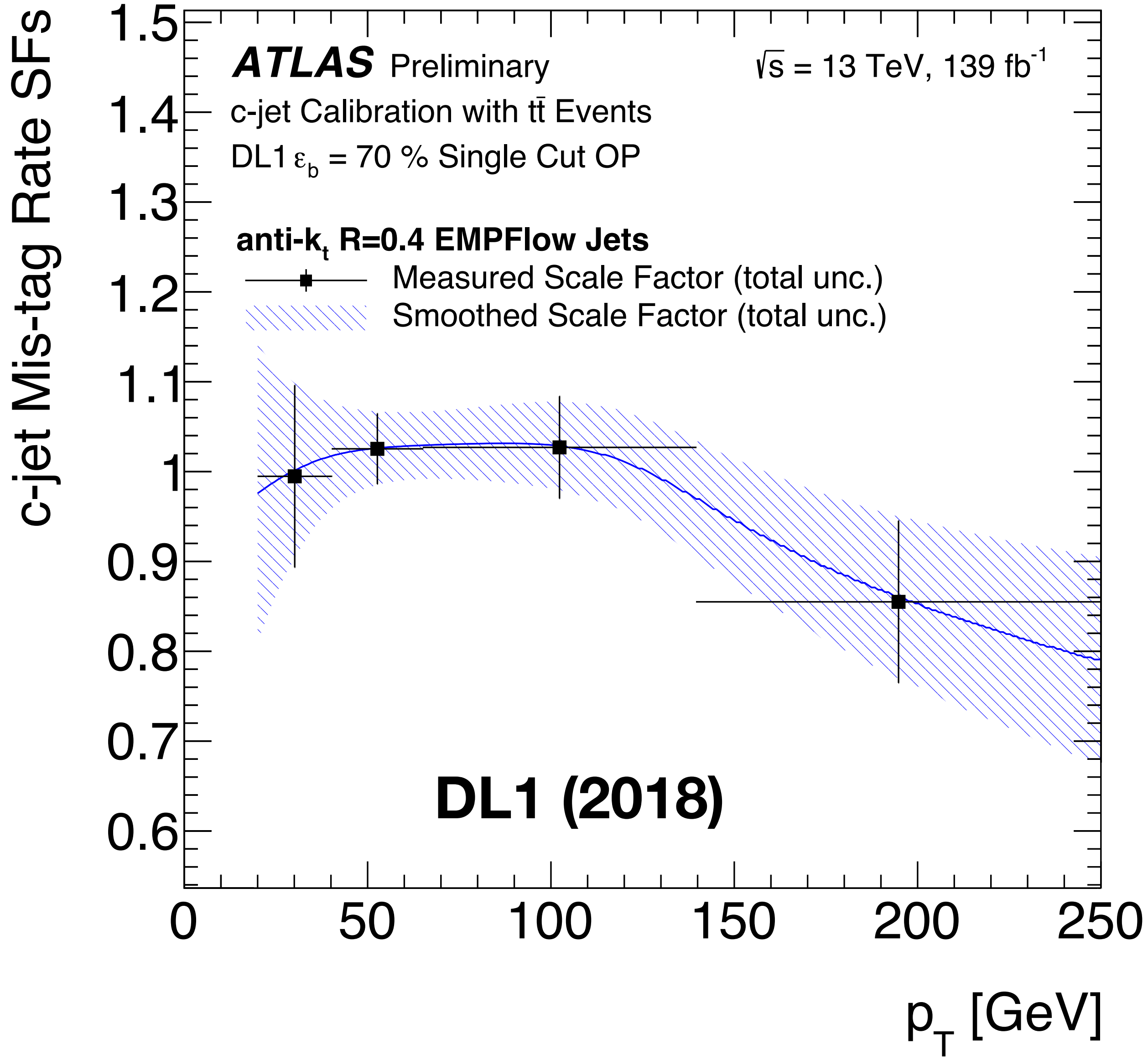


Tagger performance in data

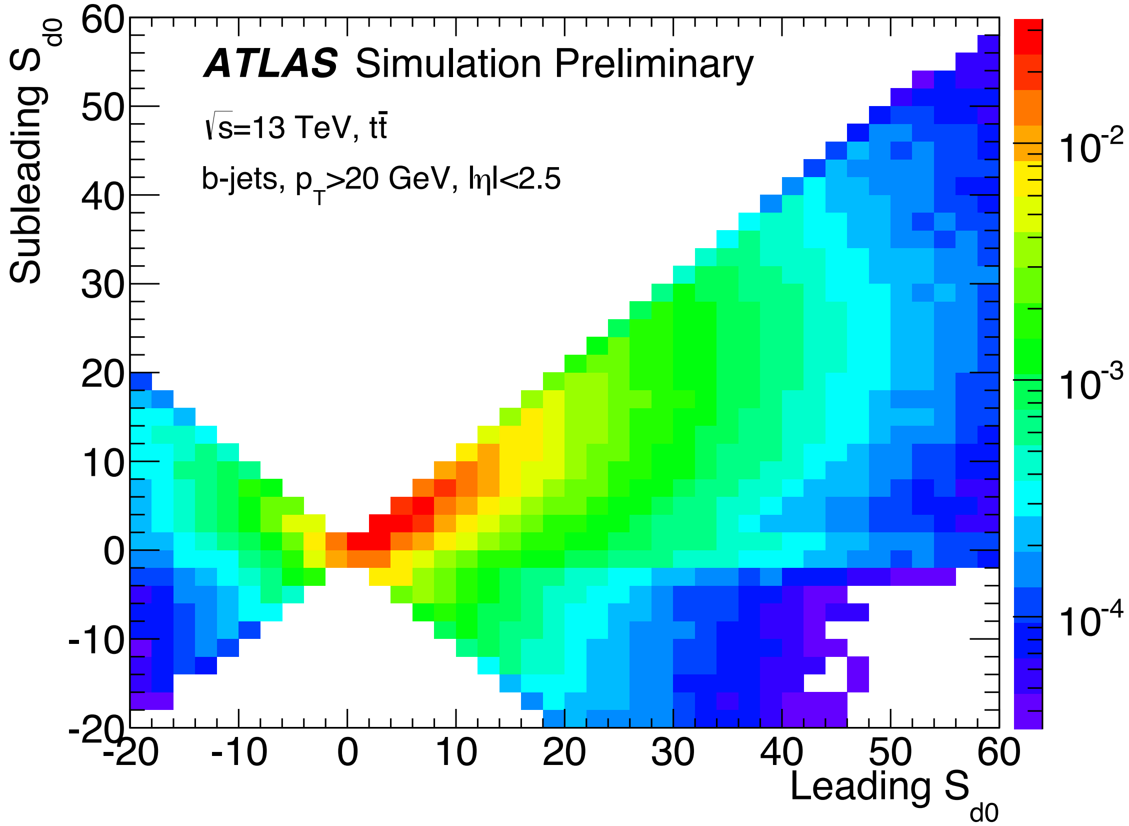
- charm-SF measured in semileptonic ttbar

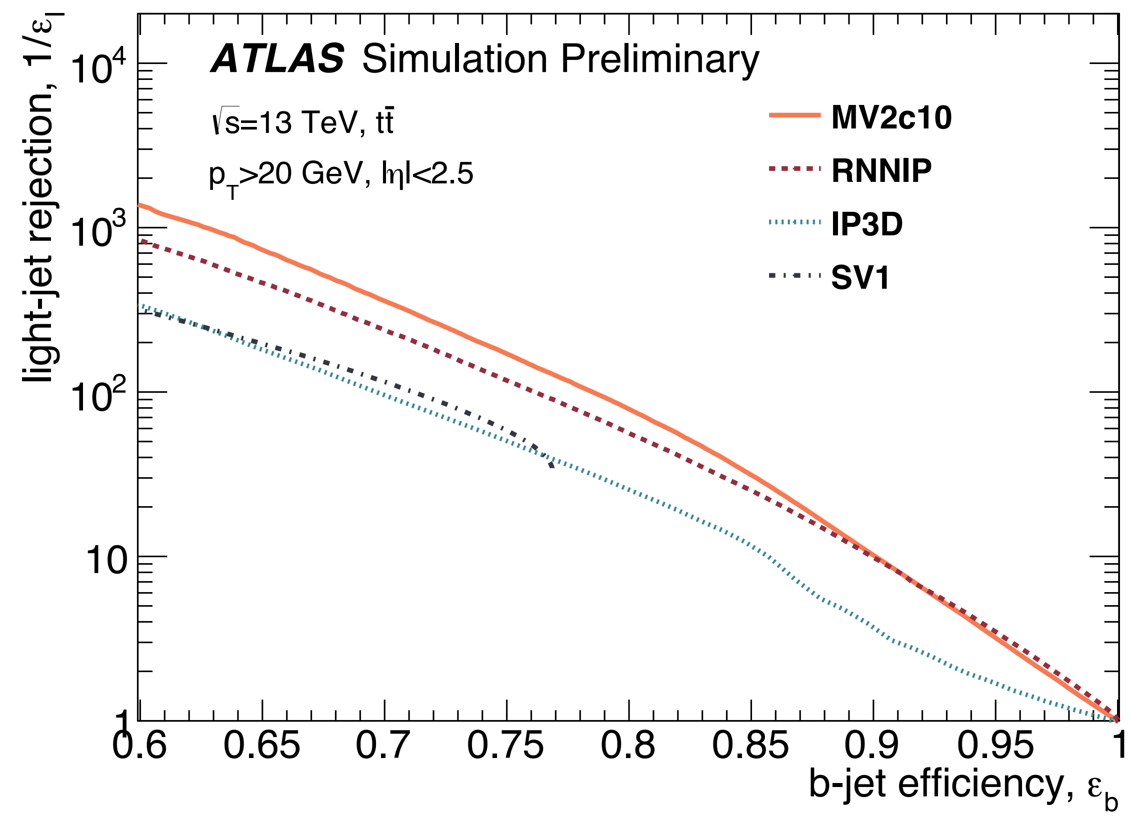
$$\mathrm{SF}_b = \frac{\epsilon^{\mathrm{data}}_{b\mathrm{-jets}}}{\epsilon^{\mathrm{sim.}}_{b\mathrm{-jets}}}$$

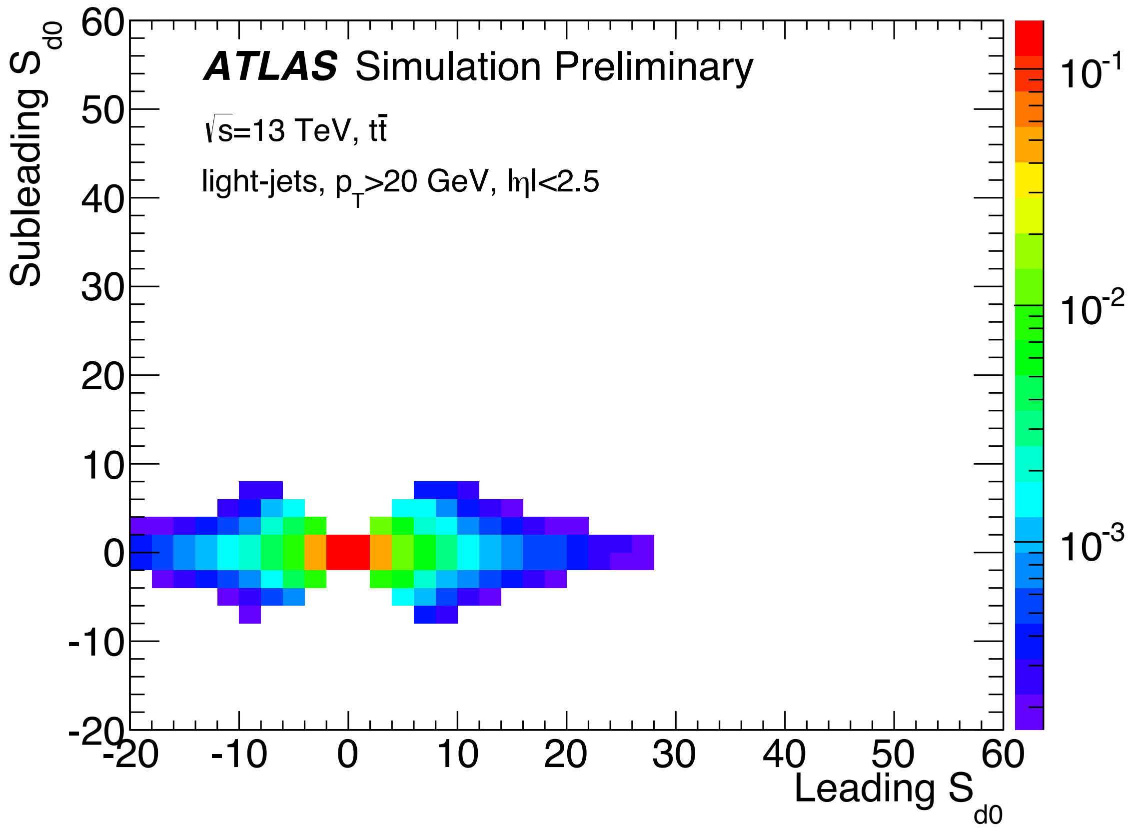


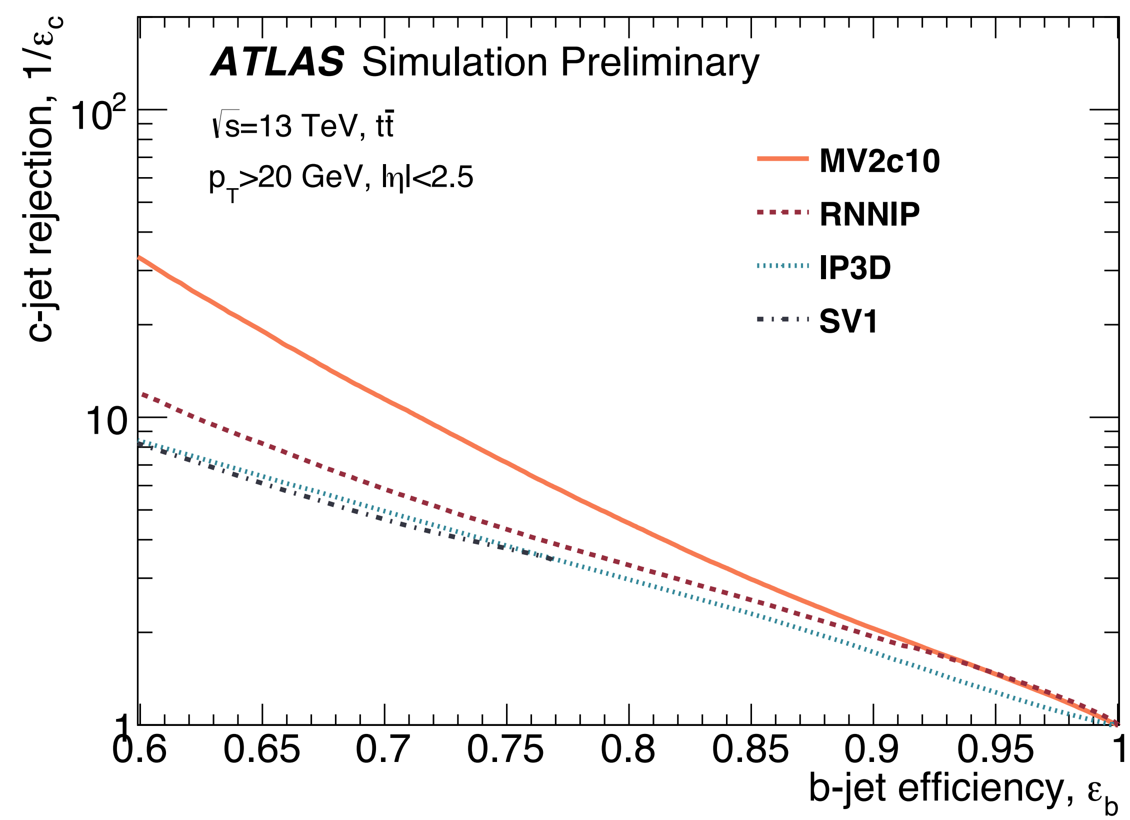


RNNIP


ATLAS Simulation Preliminary
√s=13 TeV, tt̄
b-jets, p_T>20 GeV, |η|<2.5
Subleading S_{d0}
Leading S_{d0}


ATLAS Simulation Preliminary
√s=13 TeV, tt̄
p_T>20 GeV, |η|<2.5
MV2c10
RNNIP
IP3D
SV1
light-jet rejection, $1/\varepsilon_l$
b-jet efficiency, ε_b


ATLAS Simulation Preliminary
√s=13 TeV, tt̄
light-jets, p_T>20 GeV, |η|<2.5
Subleading S_{d0}
Leading S_{d0}


ATLAS Simulation Preliminary
√s=13 TeV, tt̄
p_T>20 GeV, |η|<2.5
MV2c10
RNNIP
IP3D
SV1
c-jet rejection, $1/\varepsilon_c$
b-jet efficiency, ε_b

RNNIPFlip

- RNNIP leads to significant enhancement of light-jet rejection

How to measure light-jet rejection in data?

- Flavour composition after tagging dominated by heavy-flavour jets
- Cannot establish pure enough sample of light jets to perform calibration

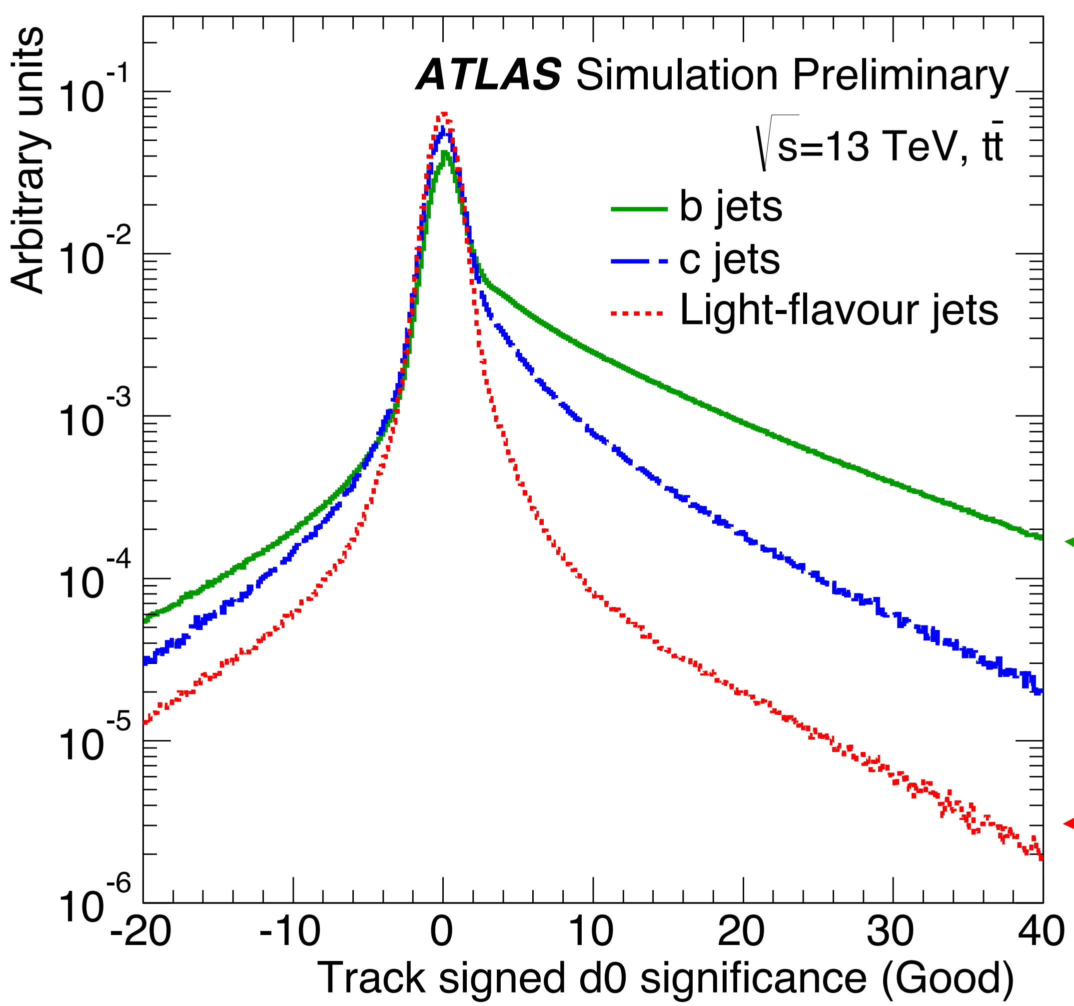
Instead:

- Define another tagger “RNNIPFlip”, with the same light rejection, but much worse b-efficiency
- Easy to calibrate in data, then extrapolate to actual RNNIP

RNNIPFlip

RNNIPFlip: flip sign of transverse and longitudinal impact parameters, then evaluate RNNIP network.

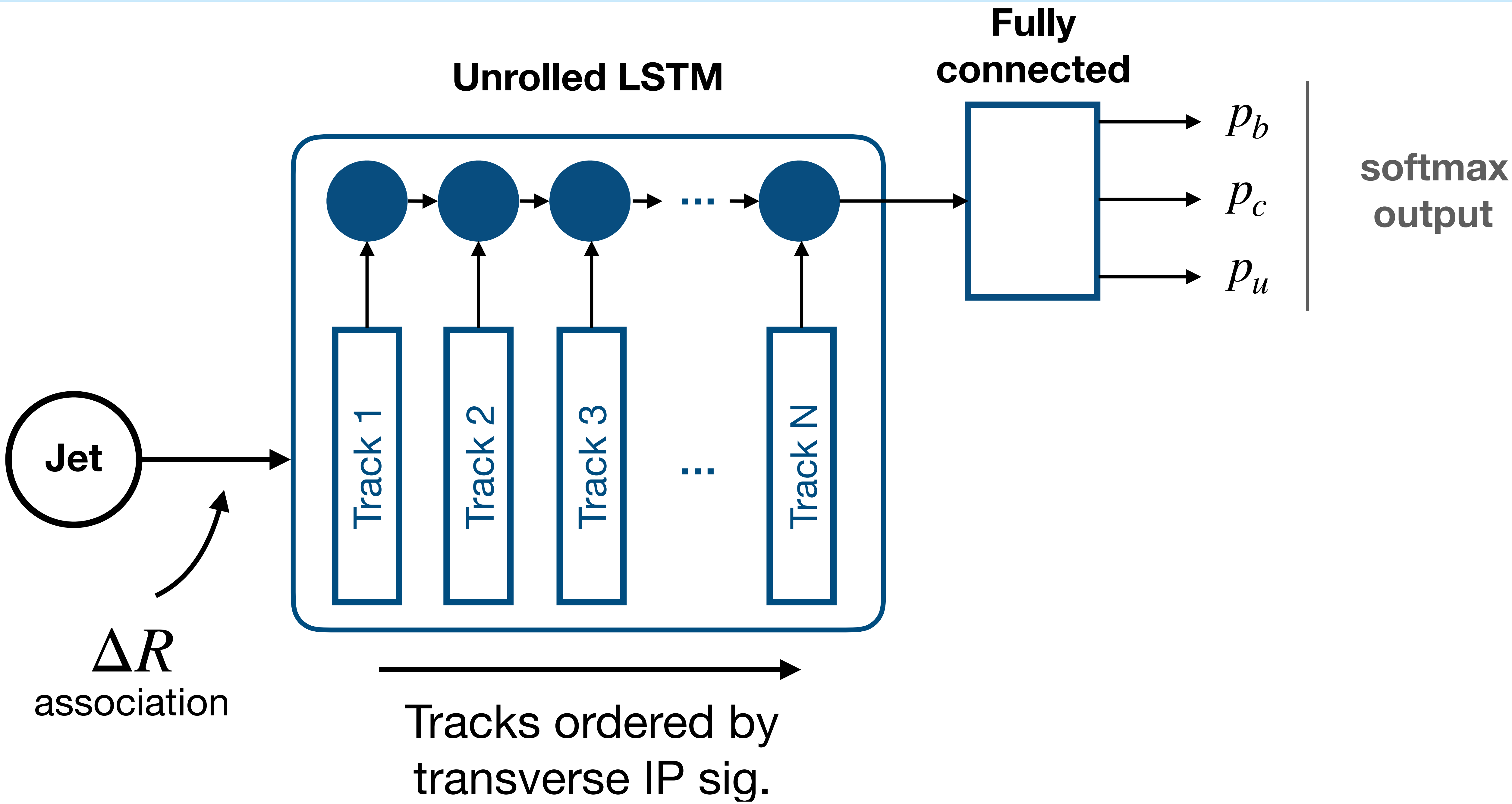
I.e. sequence order is reversed as well.



Highly asymmetric for tracks in heavy-flavour jets

Symmetric around zero for tracks in light-jets

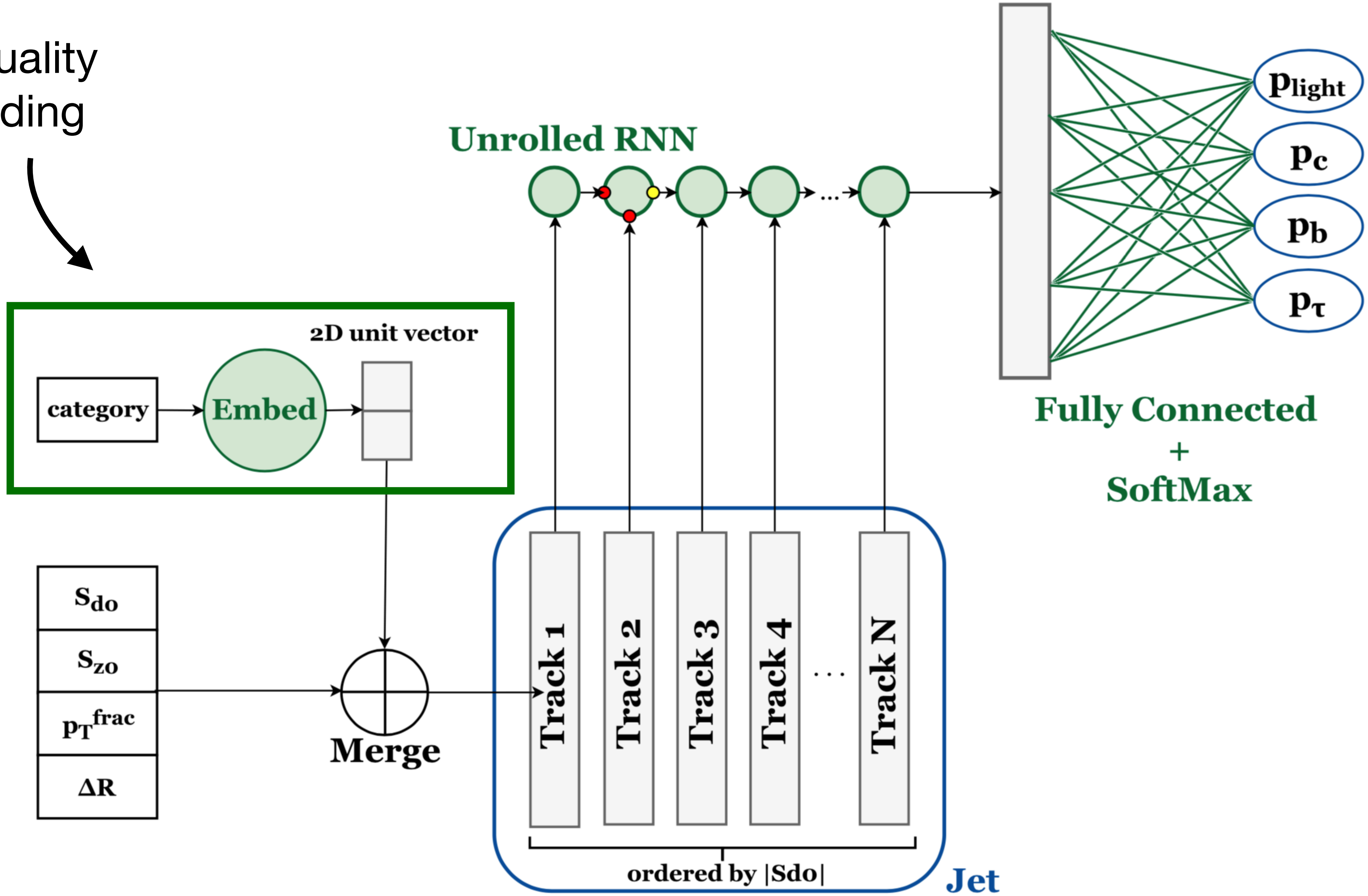
RNNIP network architecture



- LSTM with 50 hidden units
- Dense layer with 10 hidden units before softmax

Old RNNIP architecture

Track quality embedding



Ordered by *magnitude* of transverse IP significance

RNNIP training details

Training technicalities:

- Train on same “hybrid” sample as DL1, DL1r
- Sort delta-R associated tracks by transverse IP sig.
- Use first 15 tracks for training, zero-pad if shorter

Track features:

- 4 continuous features per track:
  - Transverse & longitudinal IP sig.
  - $\Delta R(\text{track}, \text{jet})$
  - $p_{T,\,\text{track}} / p_{T,\,\text{jet}}$
- Additional track quality features:
  - Number of (shared) hits in (innermost pixel layer | pixel | Si-strip tracker)

RNNIP training details

What does the network "learn"?

- The track multiplicity of the B-hadron decay, large impact parameters for these tracks
- Tracks with large IPs tend to be *harder* and *wider* for b-jets