flavour labelling update

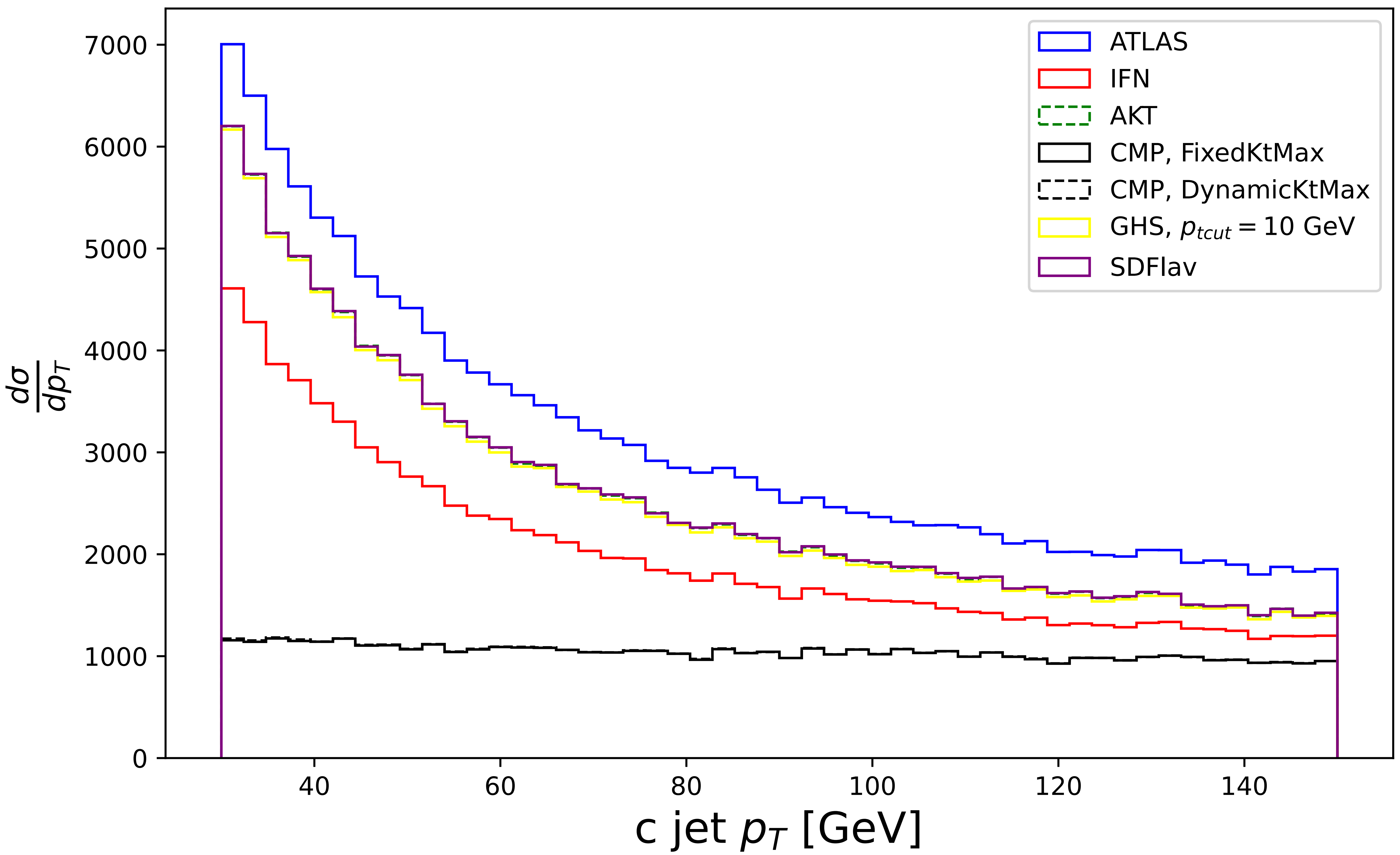
Radosław Grabarczyk

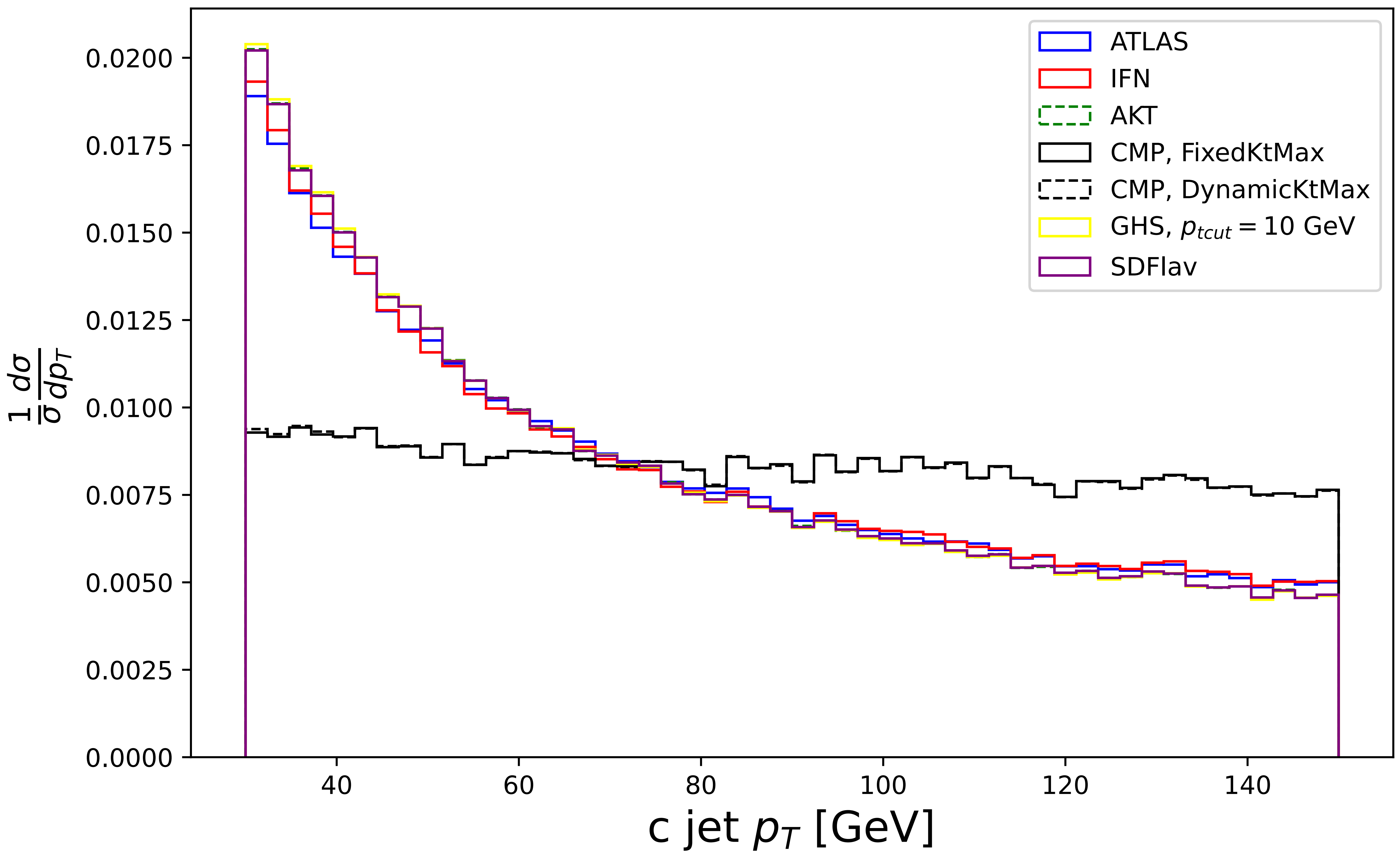
PhysTev 2025, Les Houches

The dataset

Decided to take the Z' sample ($m_{Z'} = 4$ TeV) and look at 40 GeV $< p_T < 150$ GeV c jets i.e. "QCD junk" *not* from the hard event

Jets clustered with decayed hadrons, labelled by matching with jets with undecayed hadrons with algos ran on them



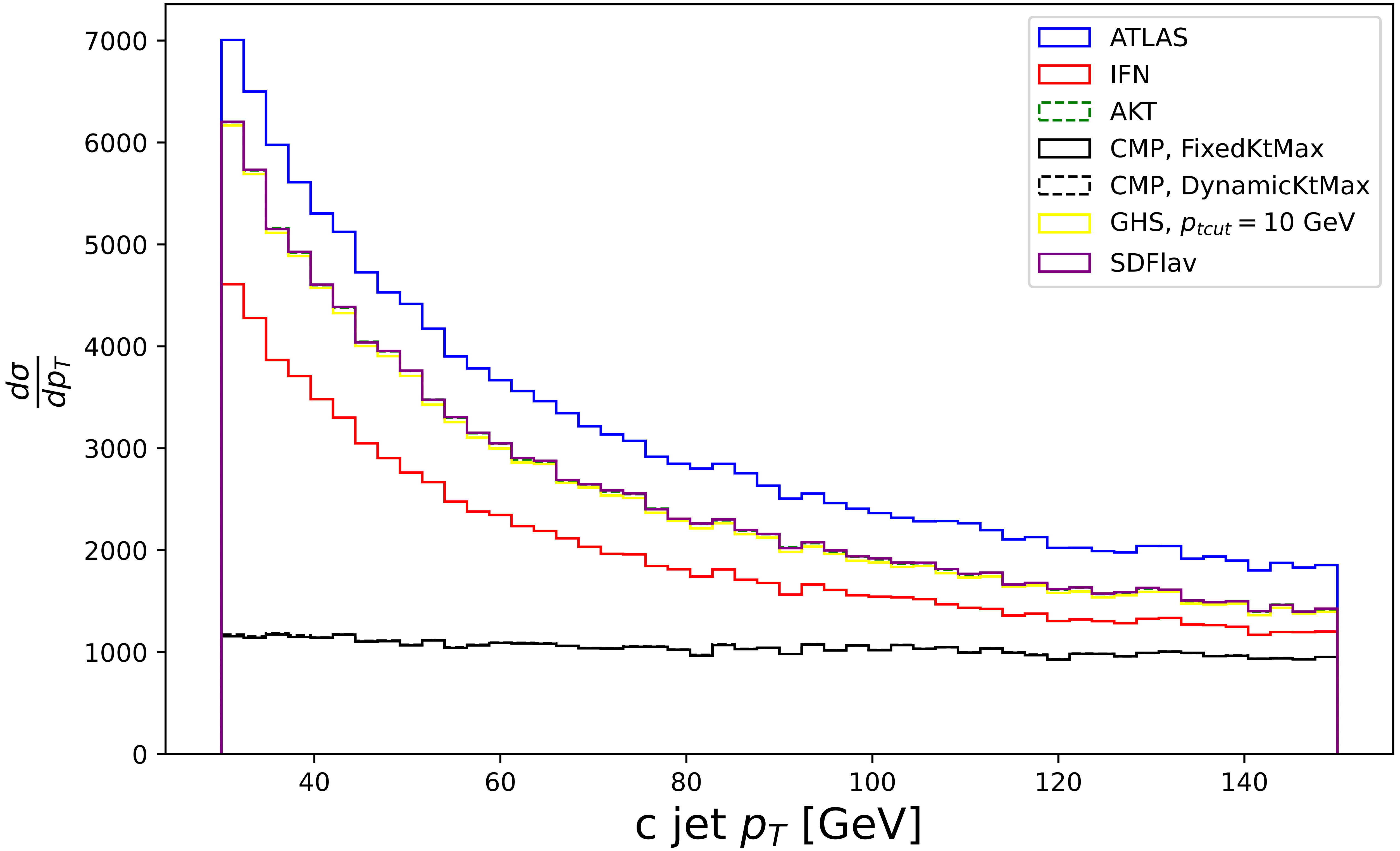


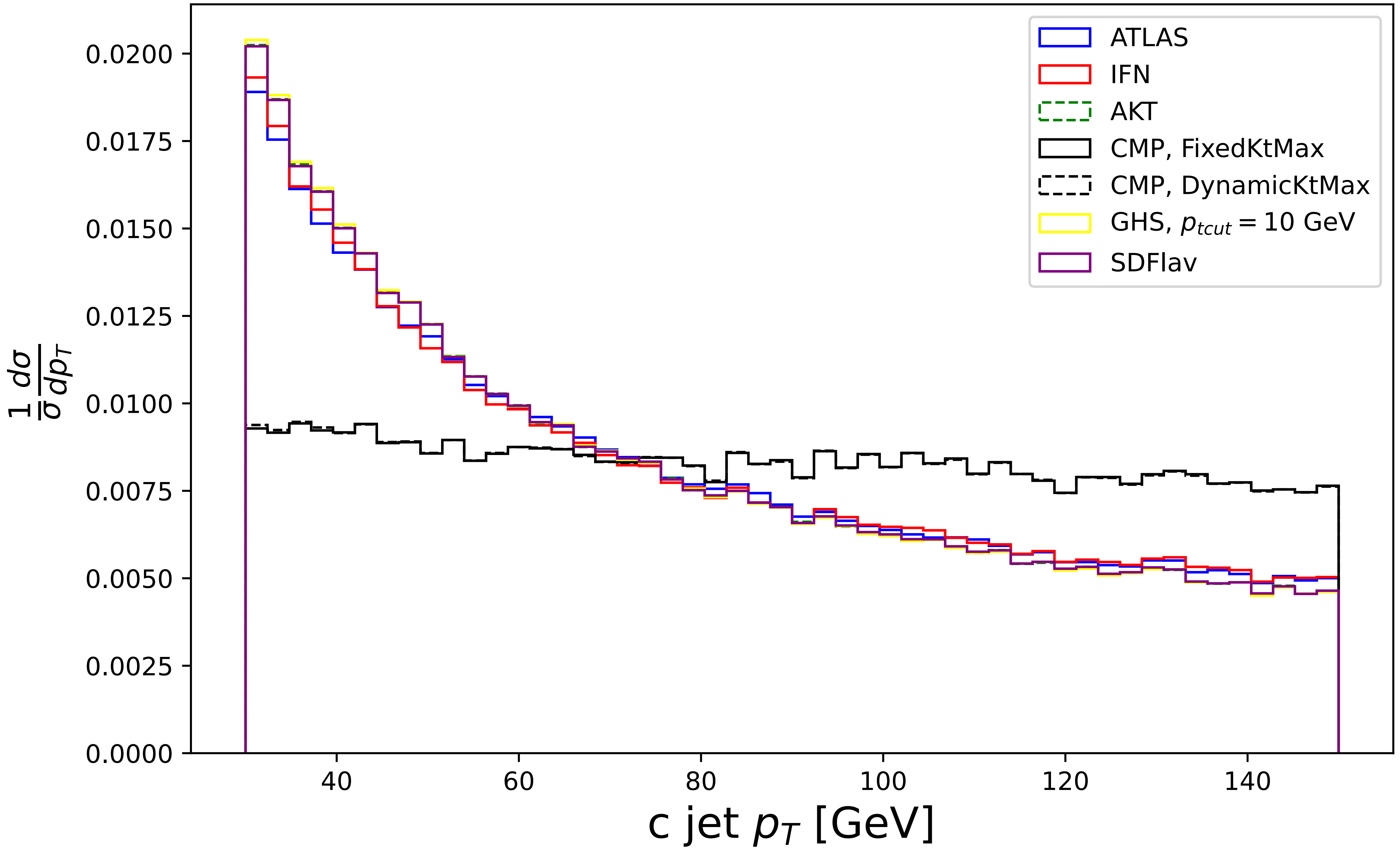
The dataset

Decided to take the Z' sample ($m_{Z'} = 4$ TeV) and look at 40 GeV $< p_T < 150$ GeV c jets i.e. "QCD junk" *not* from the hard event

It provides a good representation of c jets where the algorithms play a non-trivial role

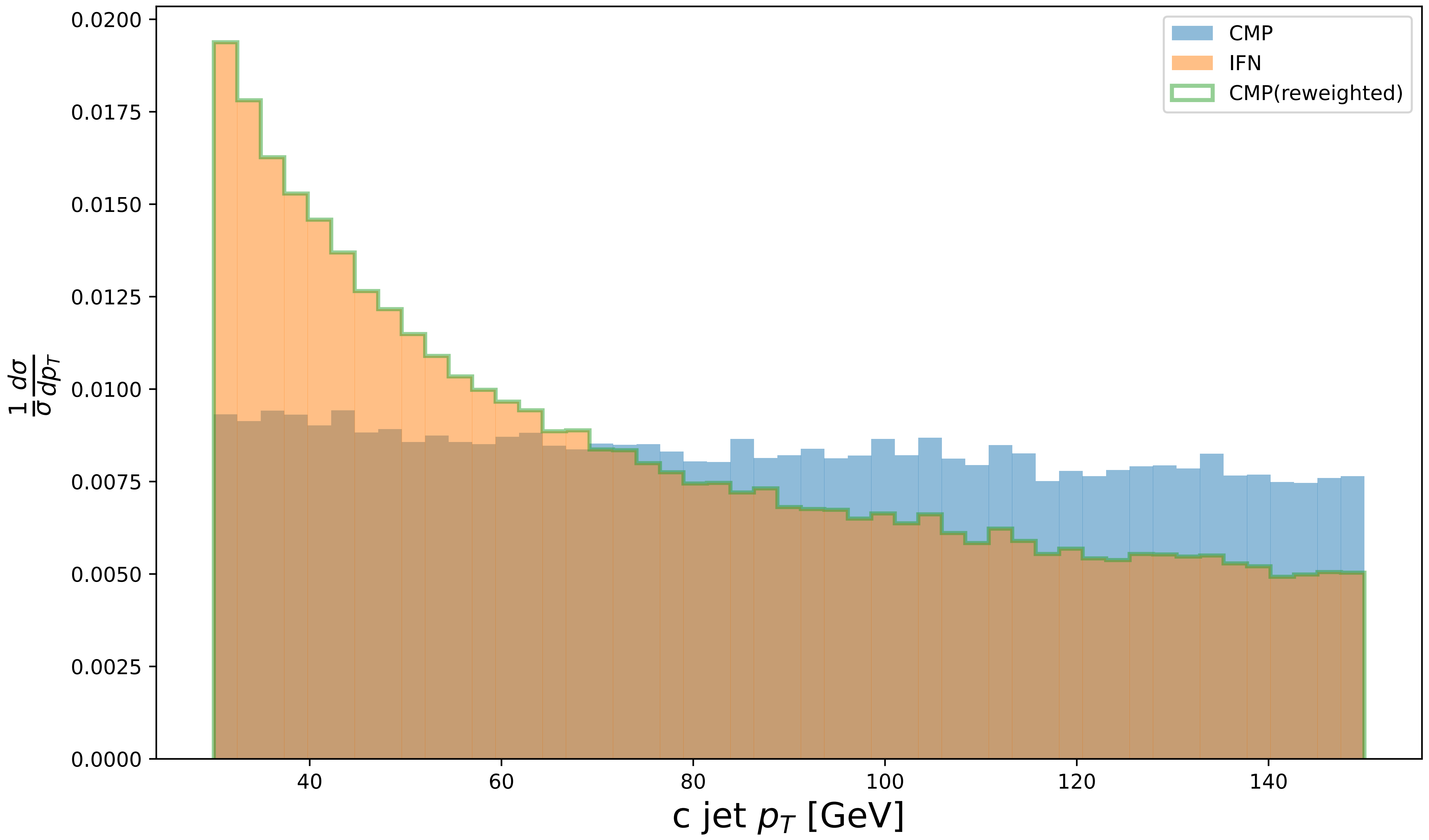
Its a particular "stress test" for CMP because of a presence of a very large $k_{T\mathrm{max}}$ scale — are there similar stress tests for other algos?

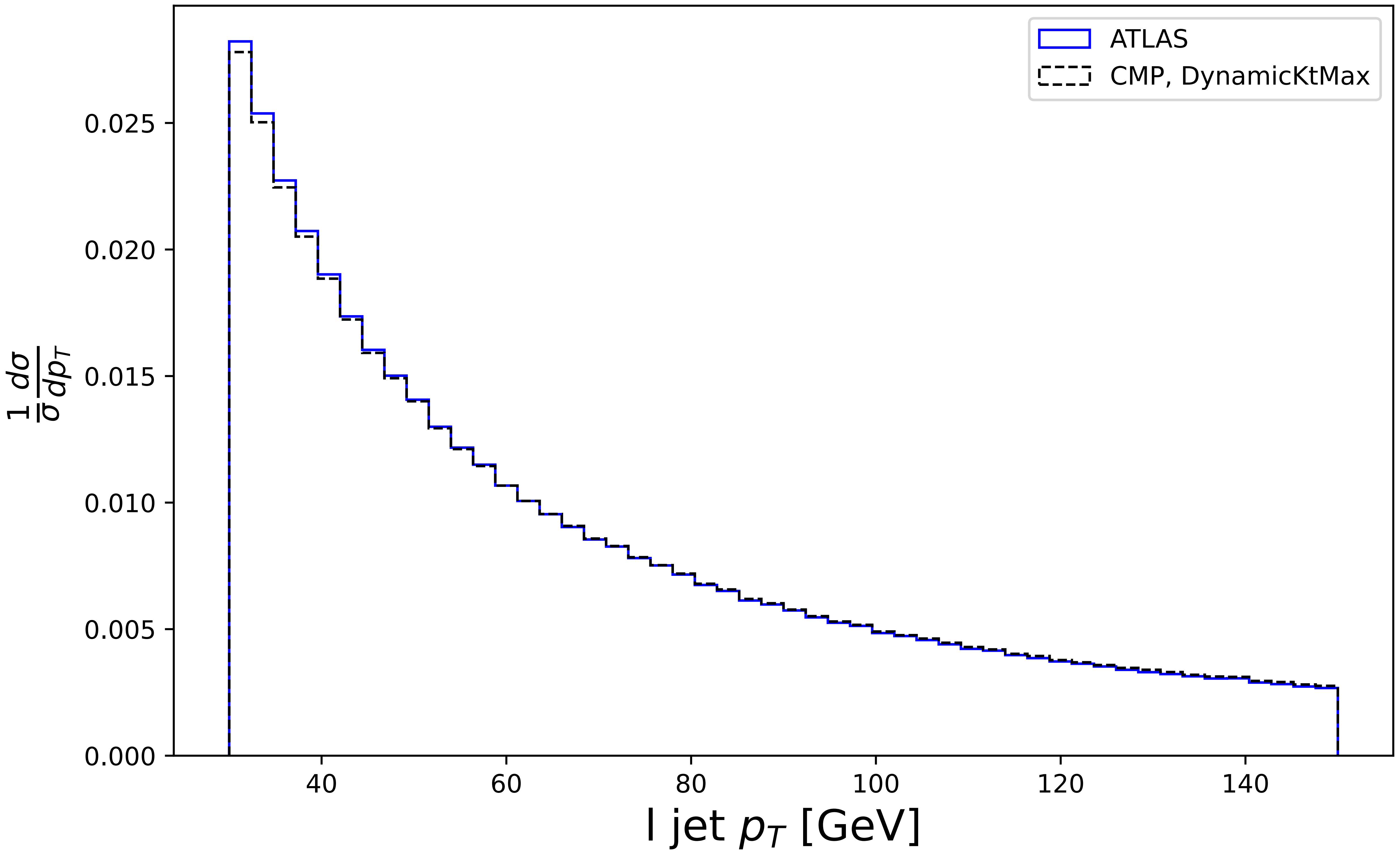




The dataset

Due to the significantly different shape of CMP c jets in pt, and the fact that there are only ~5e4 of them, applied reweighting to match IFN c jet pt and multiplied weights $\cdot 2$ to train classifier vs 1e5 light CMP jets (for all other algos 1e5 jets used in both categories)





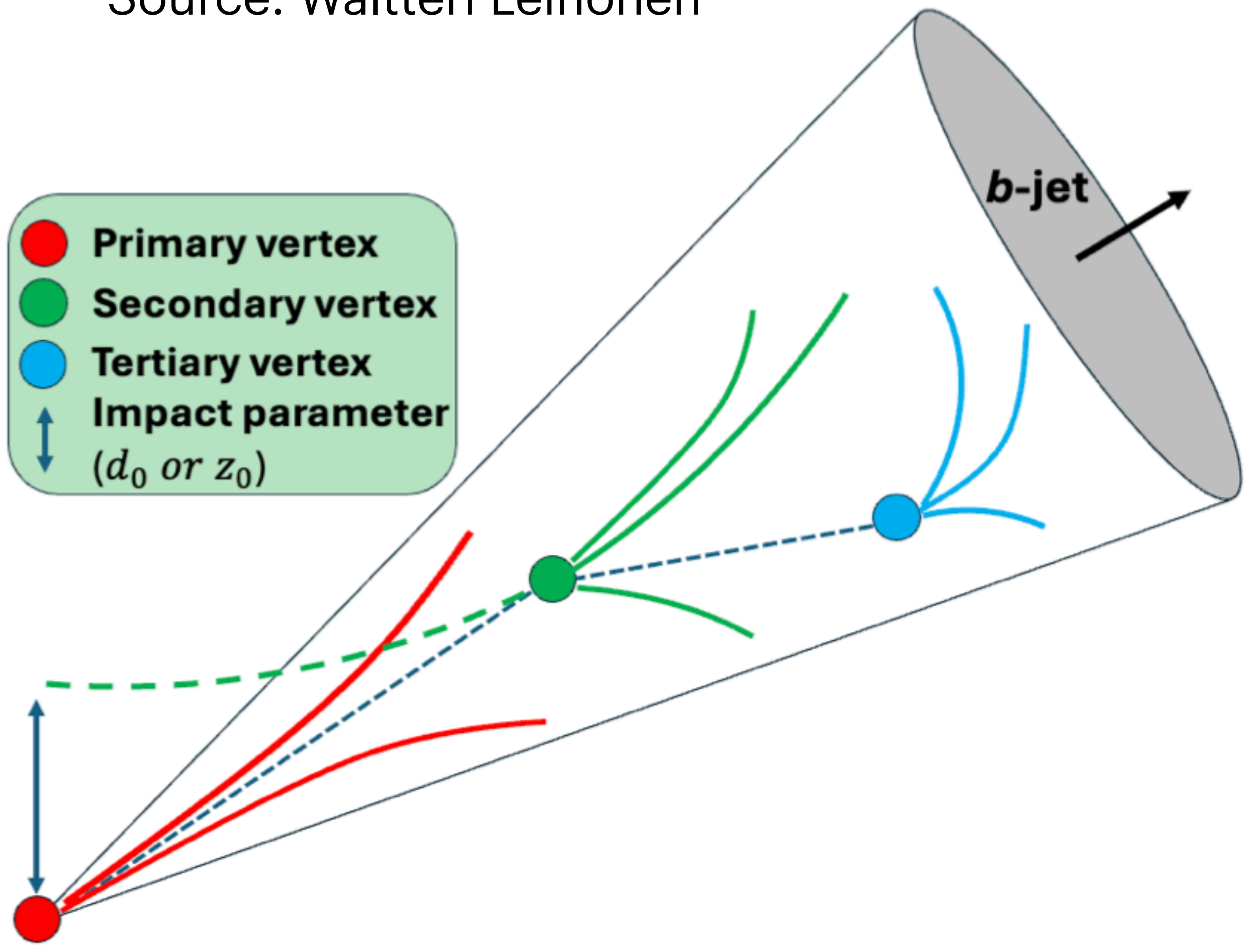
light jet sample had a nice p_T shape so it is left untouched!

Crucial point of jet flavour tagging philosophy: For a multipurpose jet flavour tagger in ATLAS/CMS we want to tag only based on what is **inside** the jet (or relatively close)

Otherwise the tagger has to be retrained, or at least calibrated for every analysis separately

This is impossible if the labels strongly depend on other properties of the event e.g. $k_{T\mathrm{max}}$ in CMP

Source: Waltteri Leinonen



It would be very interesting to quantify this "locality" of the algorithms, which experimentalists would want to be as large as possible. A study comparing how things change in the same p_T window but with a smaller Z' mass is underway.

LundNet

samples = Lund declustering tree graphs

nodes = Lund coordinates $(\ln k_t, \ln \Delta, \ln z, \ln m, \psi)$

classifier = LundNet5 (GNN) (Frédéric A. Dreyer, Huilin Qu, 2012.08526)

$$\Delta^2 = \Delta y^2 + \Delta\phi^2$$

$$k_T = \min(p_{Ti}, p_{Tj})\Delta$$

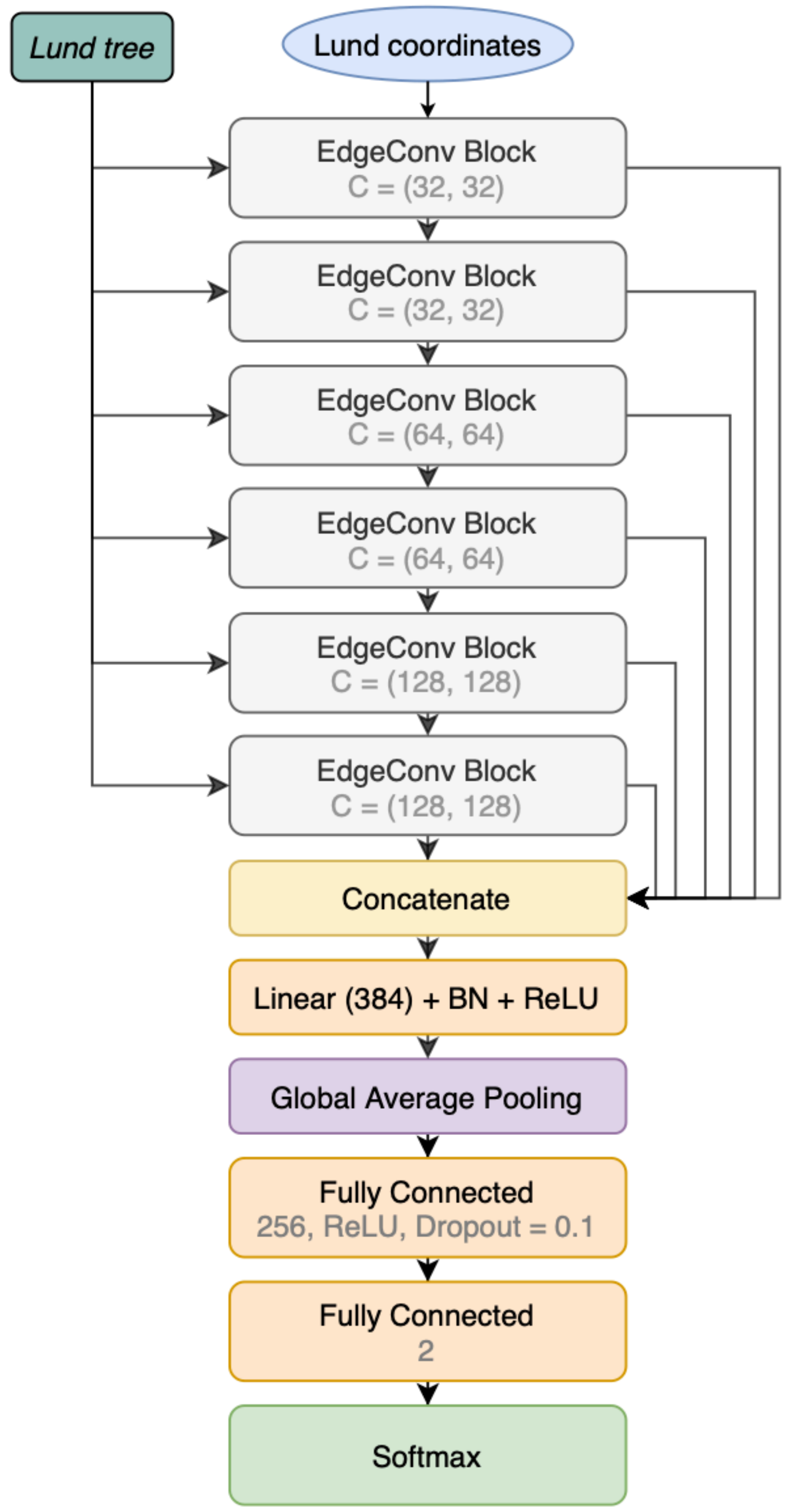
$$z = \frac{\min(p_{Ti}, p_{Tj})}{p_{Ti} + p_{Tj}}$$

$$m = (p_i + p_j)^2$$

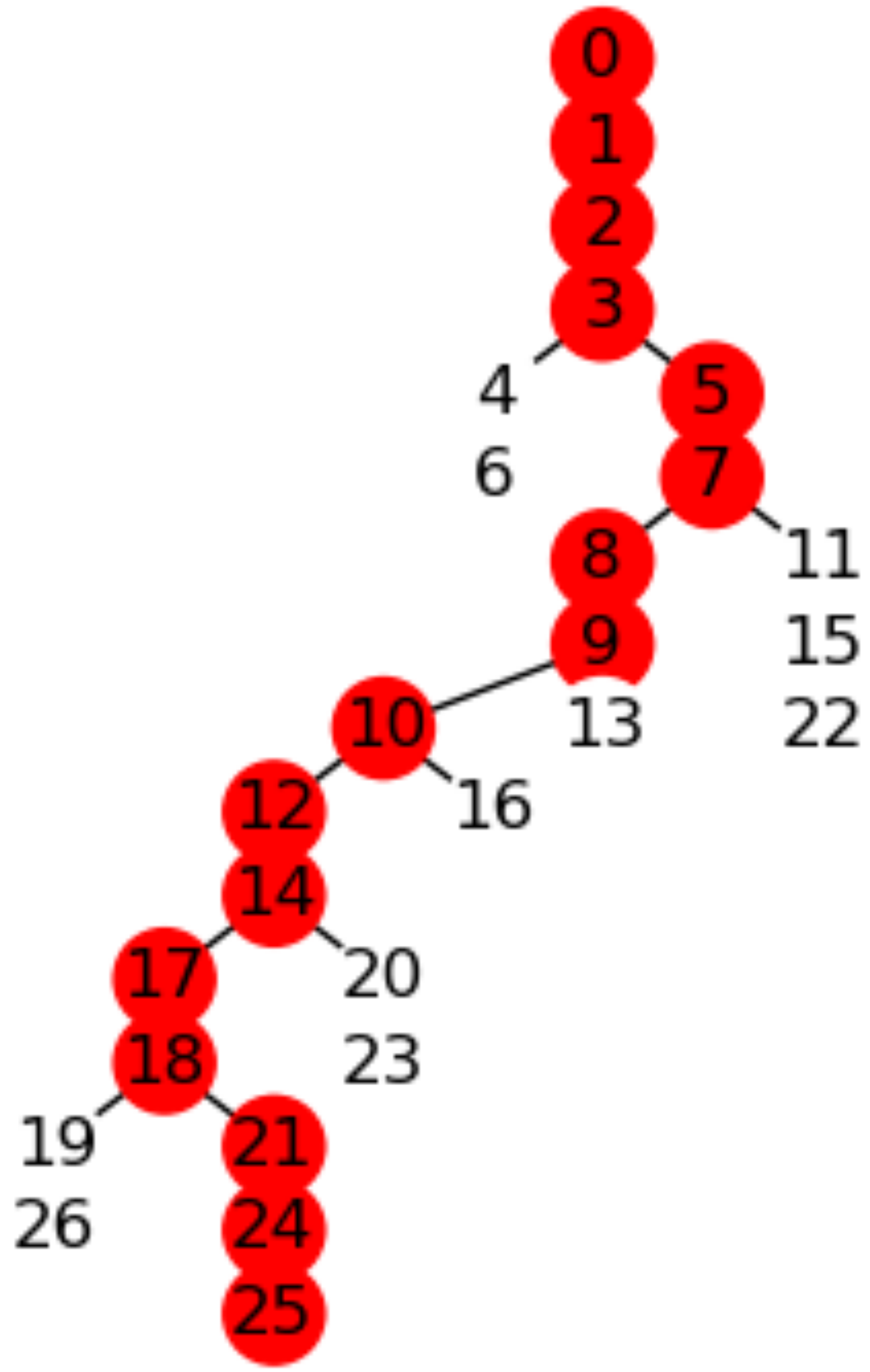
$$\psi = \arctan\left(\frac{\Delta y}{\Delta\phi}\right)$$

→ here, we only use jet substructure to try and classify the labels

interesting to see (in my opinion), but ultimately one would use more information than this

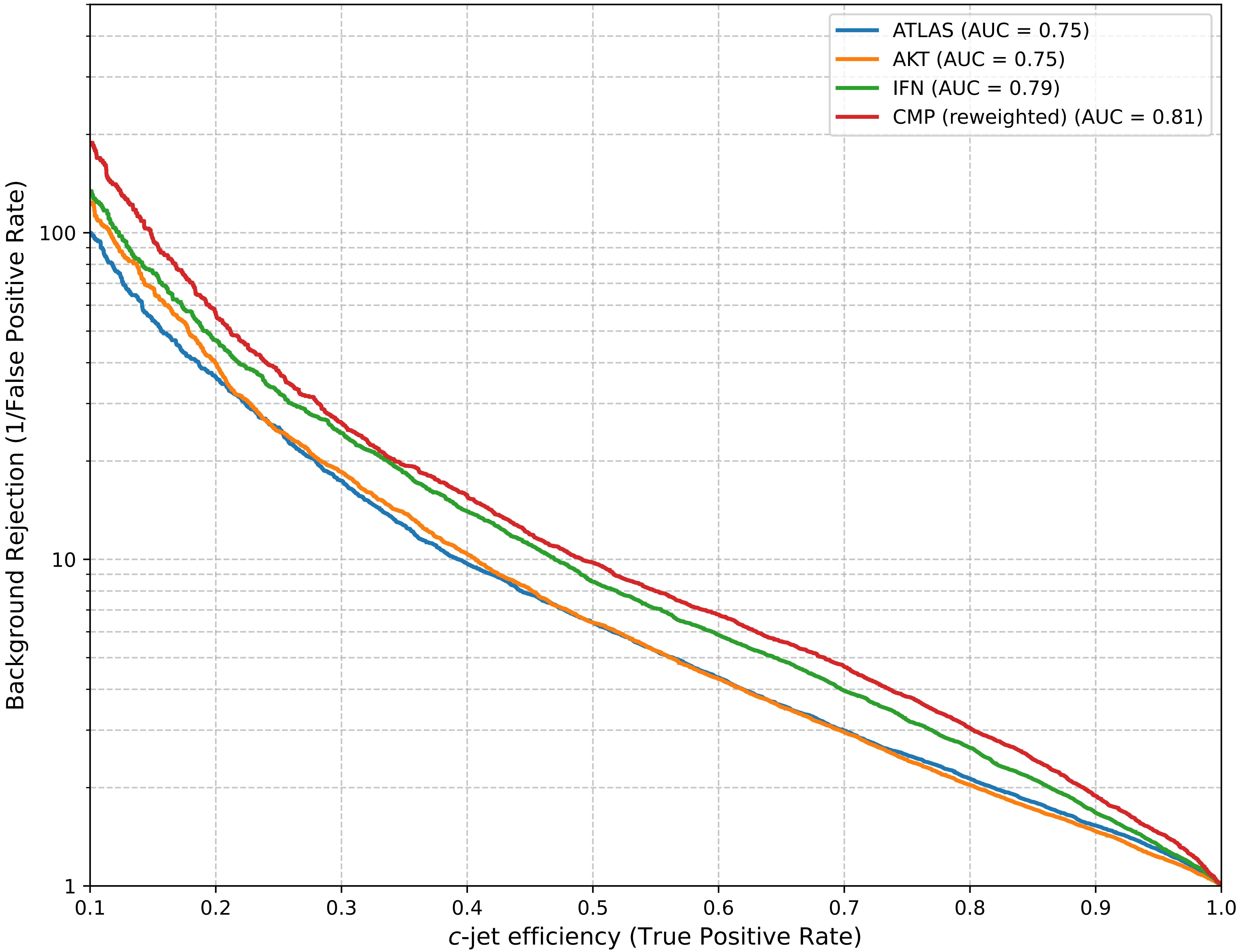


Truth Lund Tree



- primary emissions (but the network doesnt know that)

ROC curves



ROC curves are evaluated on the labels that the network was trained on in each case

Removing jets with a soft D hadron (neutralised by IFN) seems to improve performance!

The fact that the substructure is cleaner (more “learnable”) is some kind of selling point of moving to the algorithms

Summary

There is an additional unexplored requirement on jet flavour algorithms for a “collaboration-wide” flavour tagger: “locality”

A substructure based tagger appears to be **less** confused by IRC safe labels! More studies are underway

Samples are being generated to see how things change in the same p_T window but with a less massive Z' as some measure of “locality”

backup: IFN paper study of $t\bar{t}$

Only difference to the ATLAS study is that it is truth level and there is no p_{Tcut} in B hadrons for the any-flavour label; it is clear that just retraining on $t\bar{t}$ is pointless, as the interesting jets are not represented well enough