Systematic Effects in Jet Tagging with the ATLAS Detector

ML4Jets

DESY, November 9th 2023

Kevin Greif, on behalf of the ATLAS collaboration

Jet Tagging in 2019…

arxiv:1902.09914

SciPost Physics Submission

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

"Indeed, we will see that we can consider jet classification based on deep learning at the pure performance level an essentially solved problem.
For a systematic experimental application of these tools **our focus will be on a new set of questions related to training data, benchmarking, calibration, systematics, etc.**"

Jet Tagging in 2019…

arxiv:1902.09914

SciPost Physics Submission

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

This Talk!

"Indeed, we will see that we can consider jet classification based on deep learning at the pure performance level an essentially solved problem.*

For a systematic experimental application of these tools **our focus will be on a new set of questions related to training data, benchmarking, calibration, systematics, etc.**"

…and Jet Tagging in 2023

arxiv:2205.05550

Search for Higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Colla

arxiv:2204.12413

Search for a massive scalar resonance decaying to a light scalar and a Higgs boson in the four b quarks final state with boosted

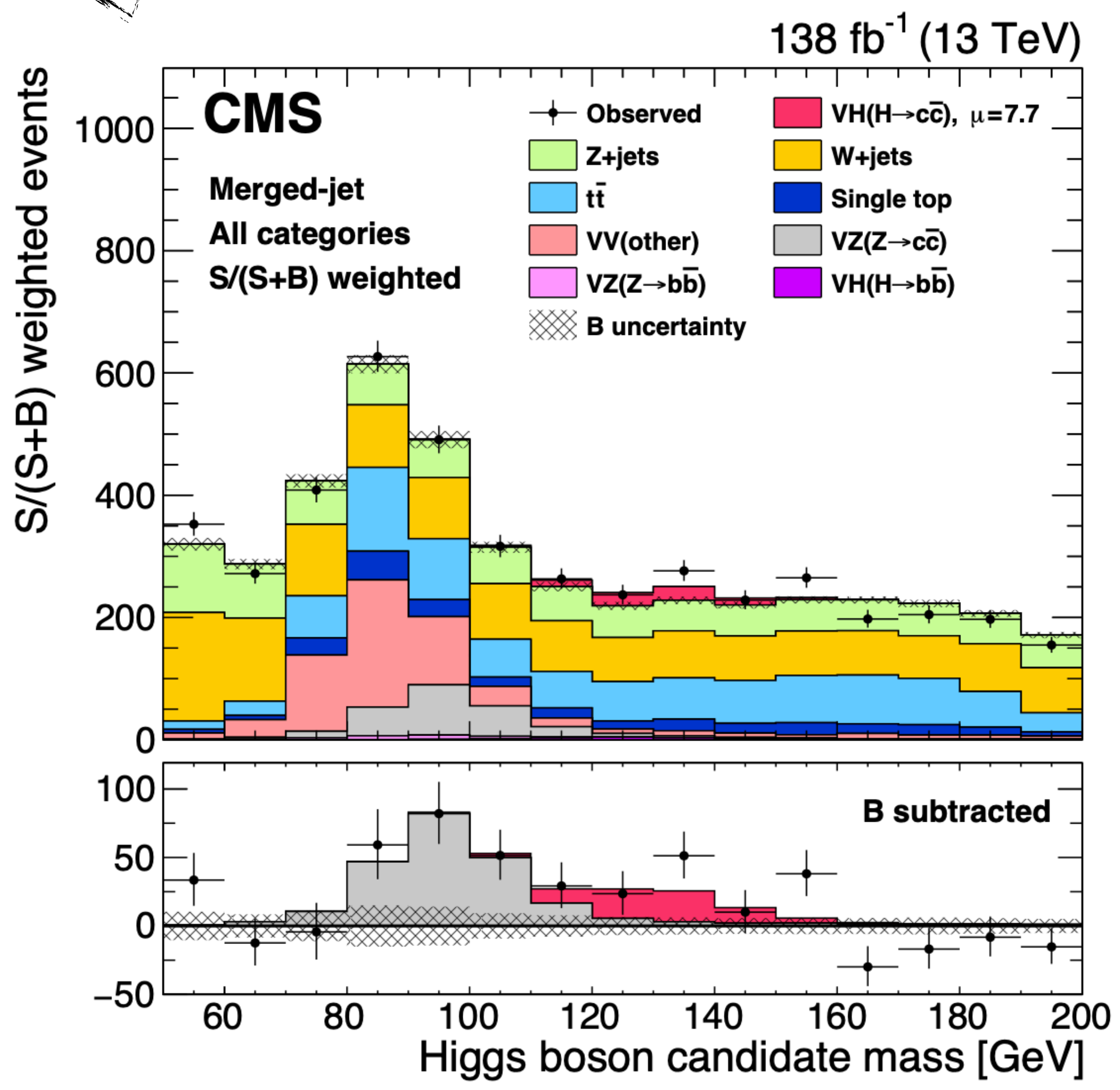
The CMS Colla

arxiv:2205.02817

Differential $t\bar{t}$ cross-section measurements using boosted top quarks in the all-hadronic final state with 139 fb^{-1} of ATLAS data

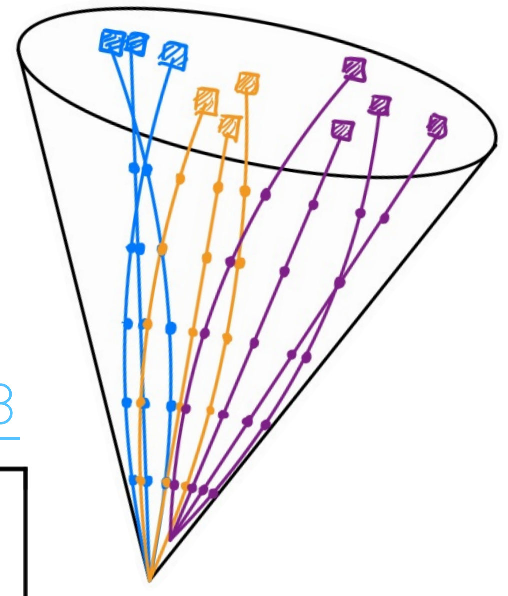
The ATLAS Collaboration

+ many others

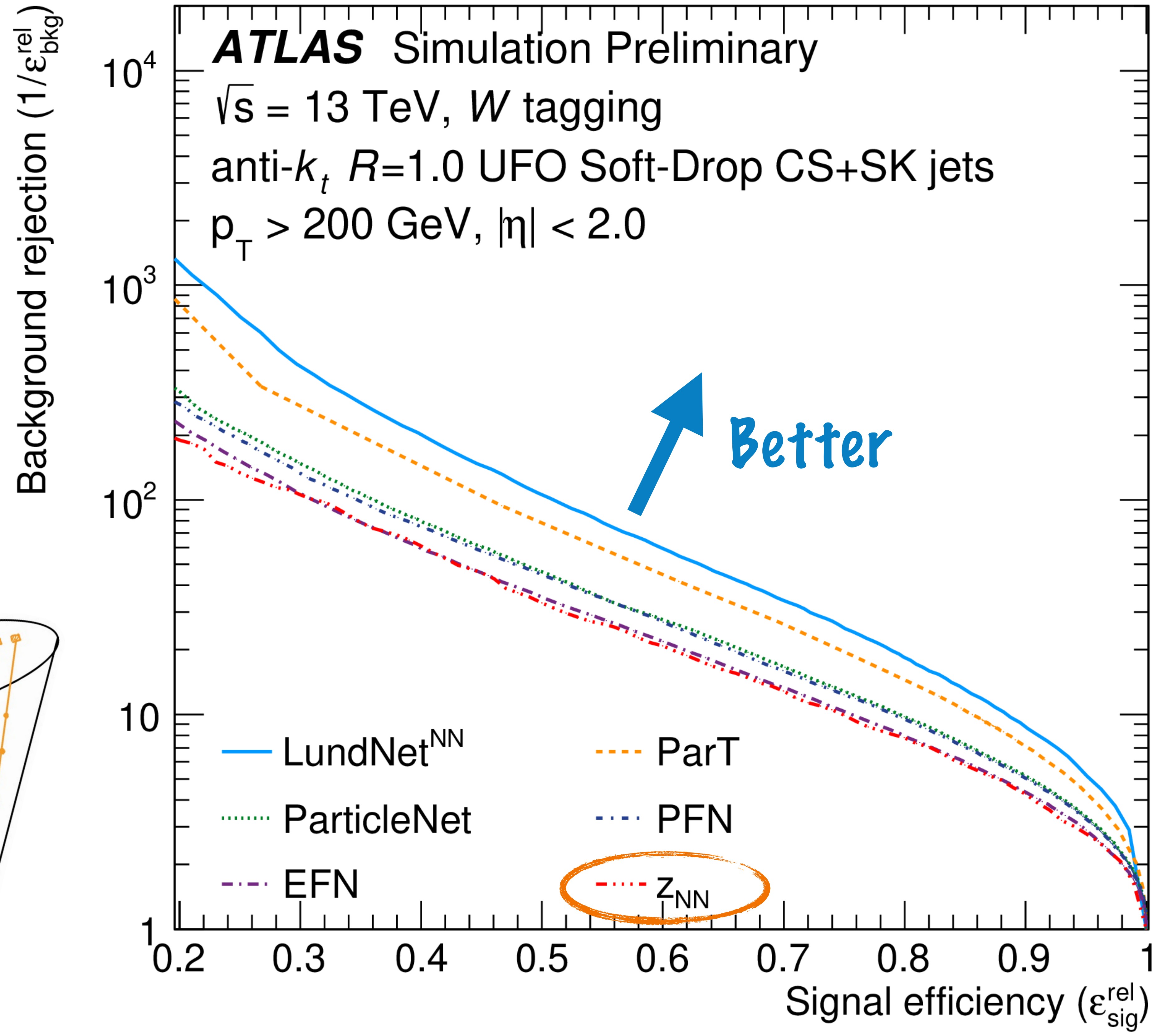


- Point cloud based jet tagging is enabling previously un-imaginable measurements
- The networks are getting ever bigger and more powerful

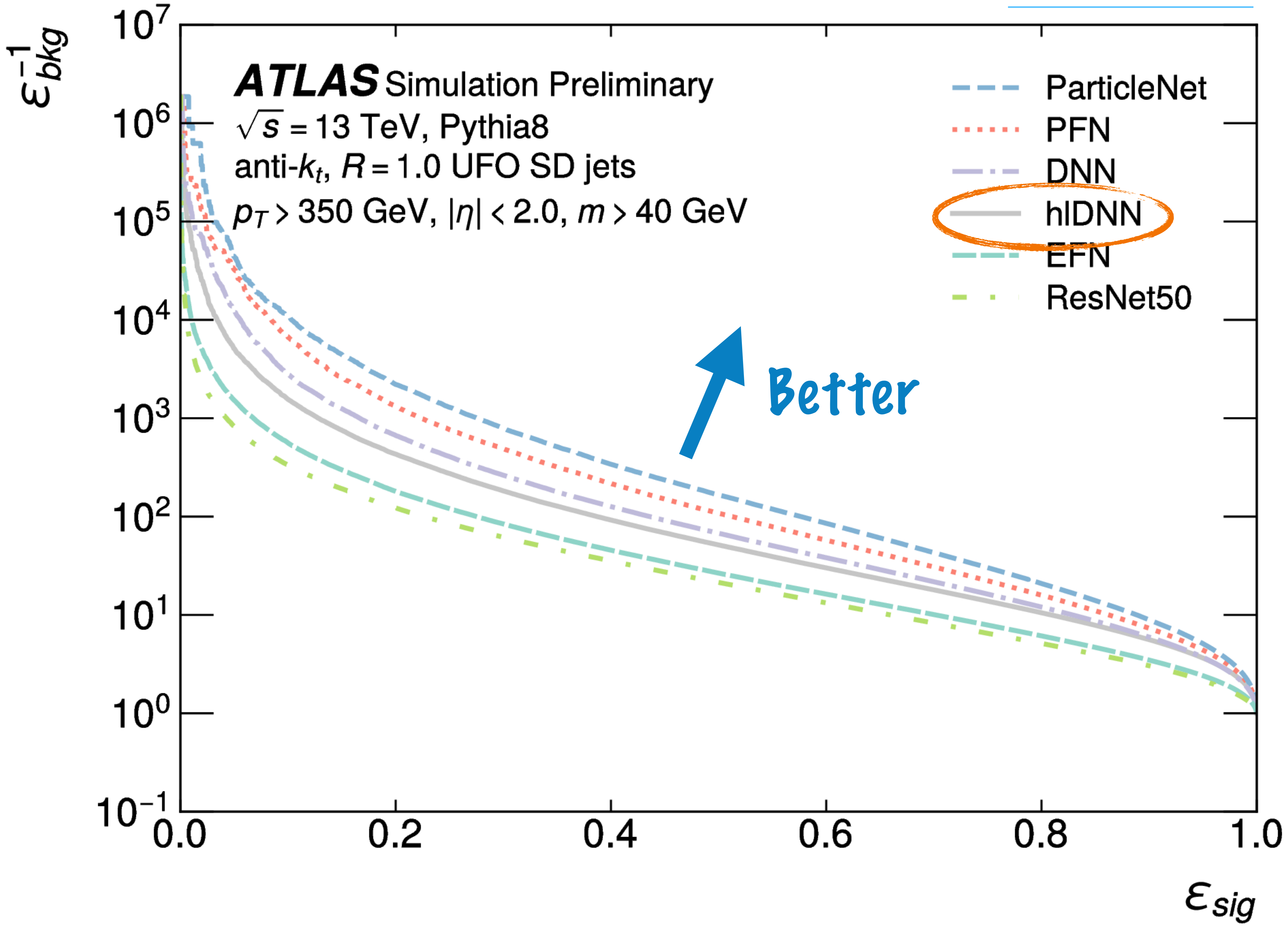
Point Cloud Taggers in ATLAS

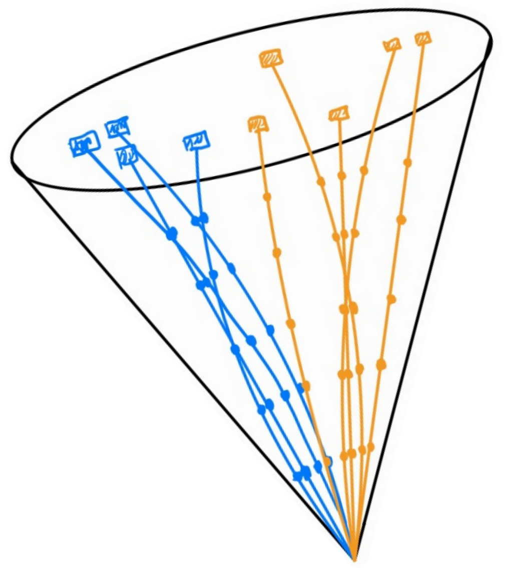


cds:2864131 cds:2866592



cds:2825328

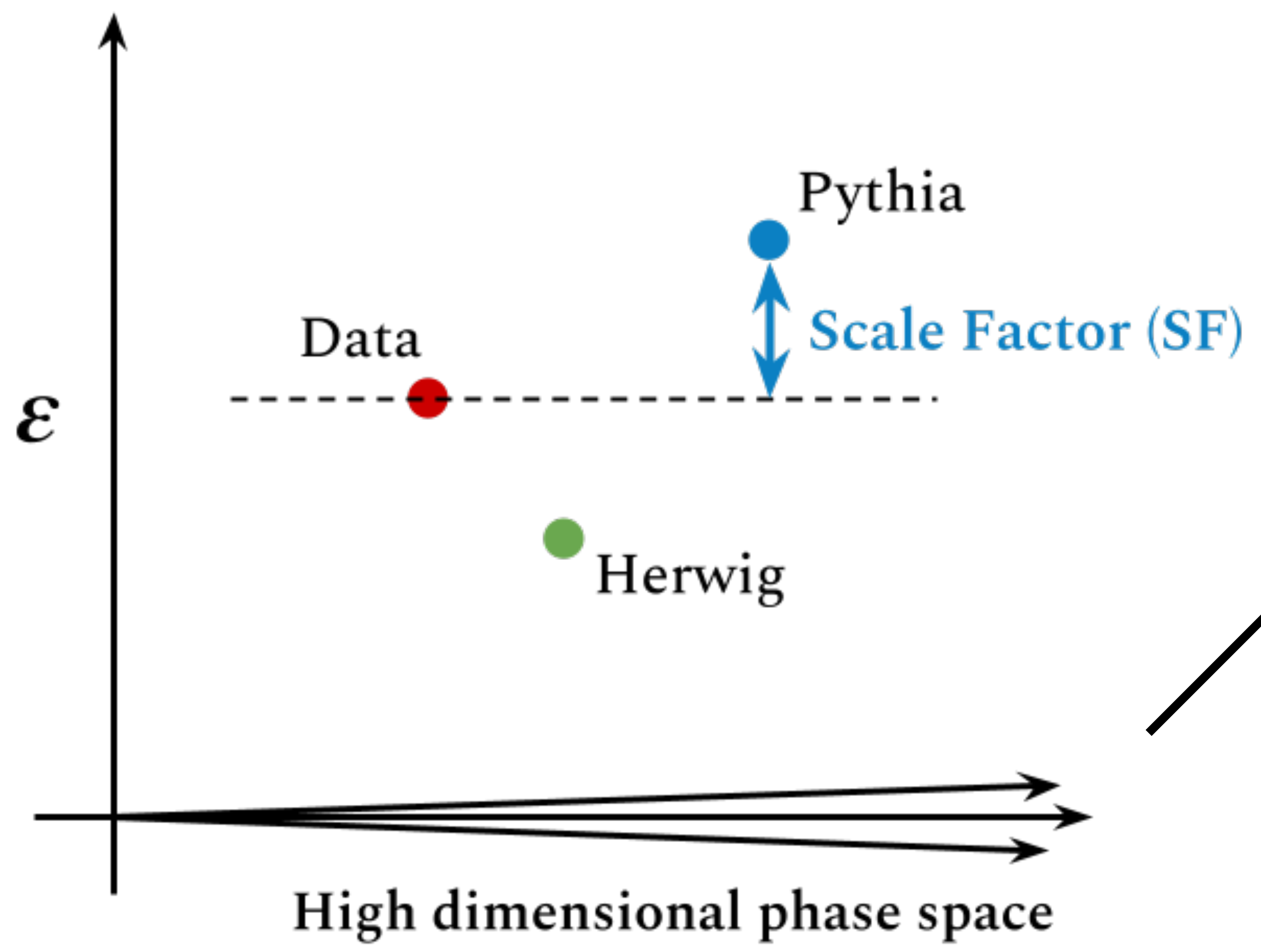




- Large performance gains for point-cloud taggers over high-level quantity baselines
- Also see Samuel's talk on q/g tagging from yesterday!

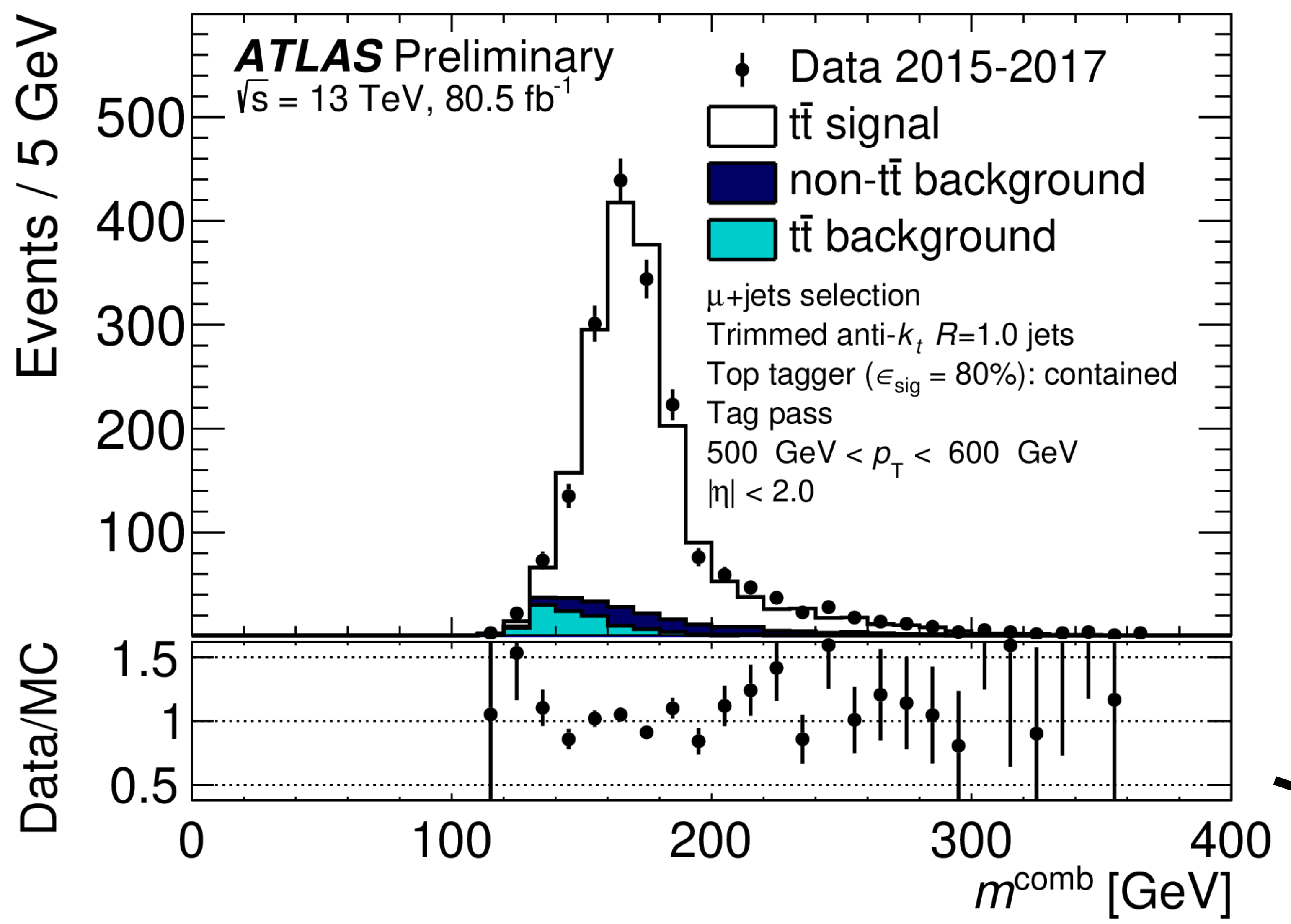
A Brief Aside on Scale Factors

- Both ATLAS and CMS train taggers on MC, but need to know efficiency in data
- Measure **scale factor** to correct MC efficiency to data efficiency



Project to one dimension sensitive to efficiency

cds:2724149



Fit normalizations (N) of MC distributions to data

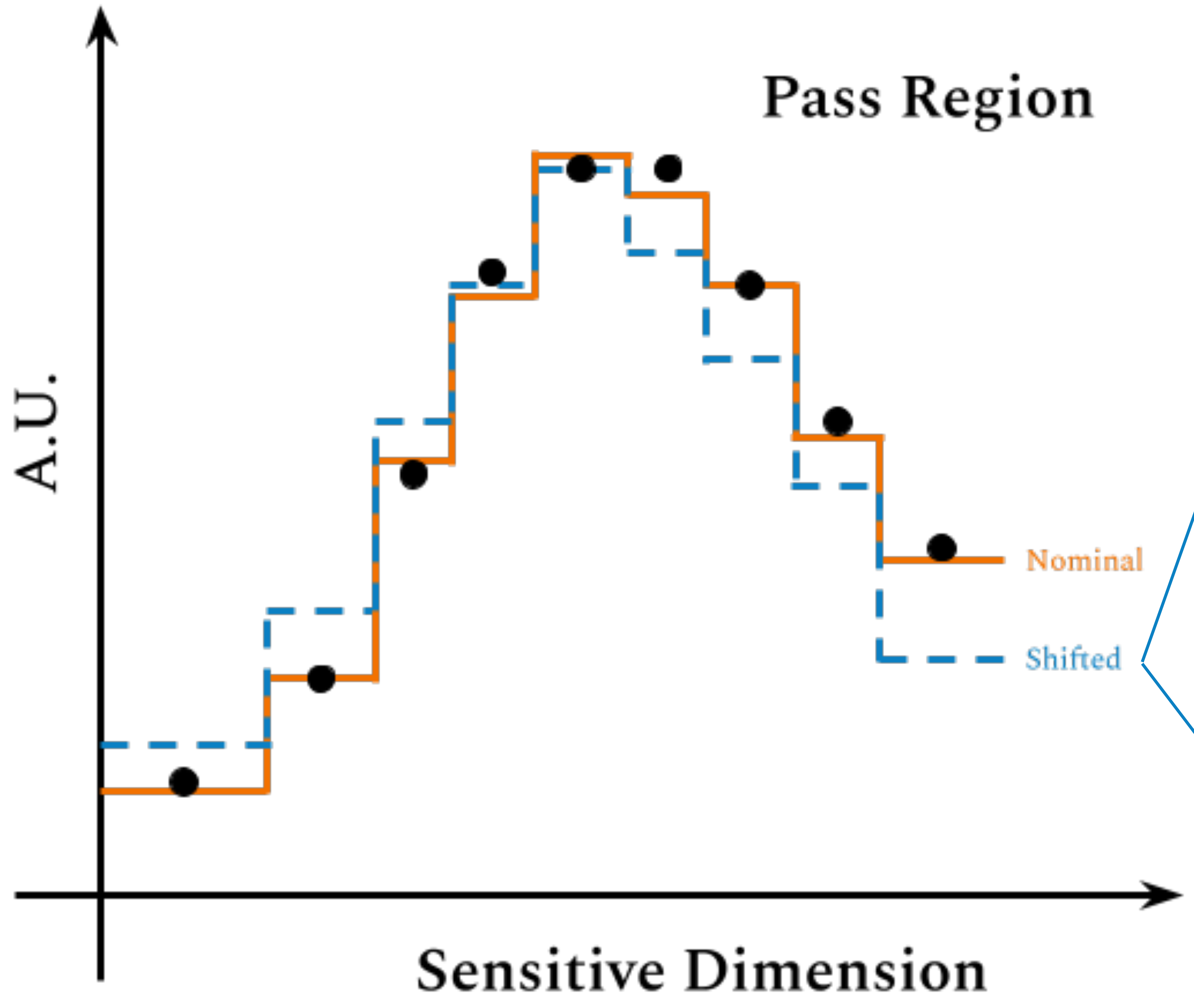
$$\epsilon_{\text{data}}(p_T) = \frac{N^{\text{tagged}}_{\text{fitted signal}}(p_T)}{N^{\text{tagged}}_{\text{fitted signal}}(p_T) + N^{\text{not tagged}}_{\text{fitted signal}}(p_T)}$$

$$\text{SF}(p_T) = \frac{\epsilon_{\text{data}}(p_T)}{\epsilon_{\text{MC}}(p_T)}$$

Scale Factor Uncertainties

$$\mathrm{SF}(p_\mathrm{T}) = \frac{\epsilon_\mathrm{data}(p_\mathrm{T})}{\epsilon_\mathrm{MC}(p_\mathrm{T})}$$

Like any measurement SFs have **uncertainties**:



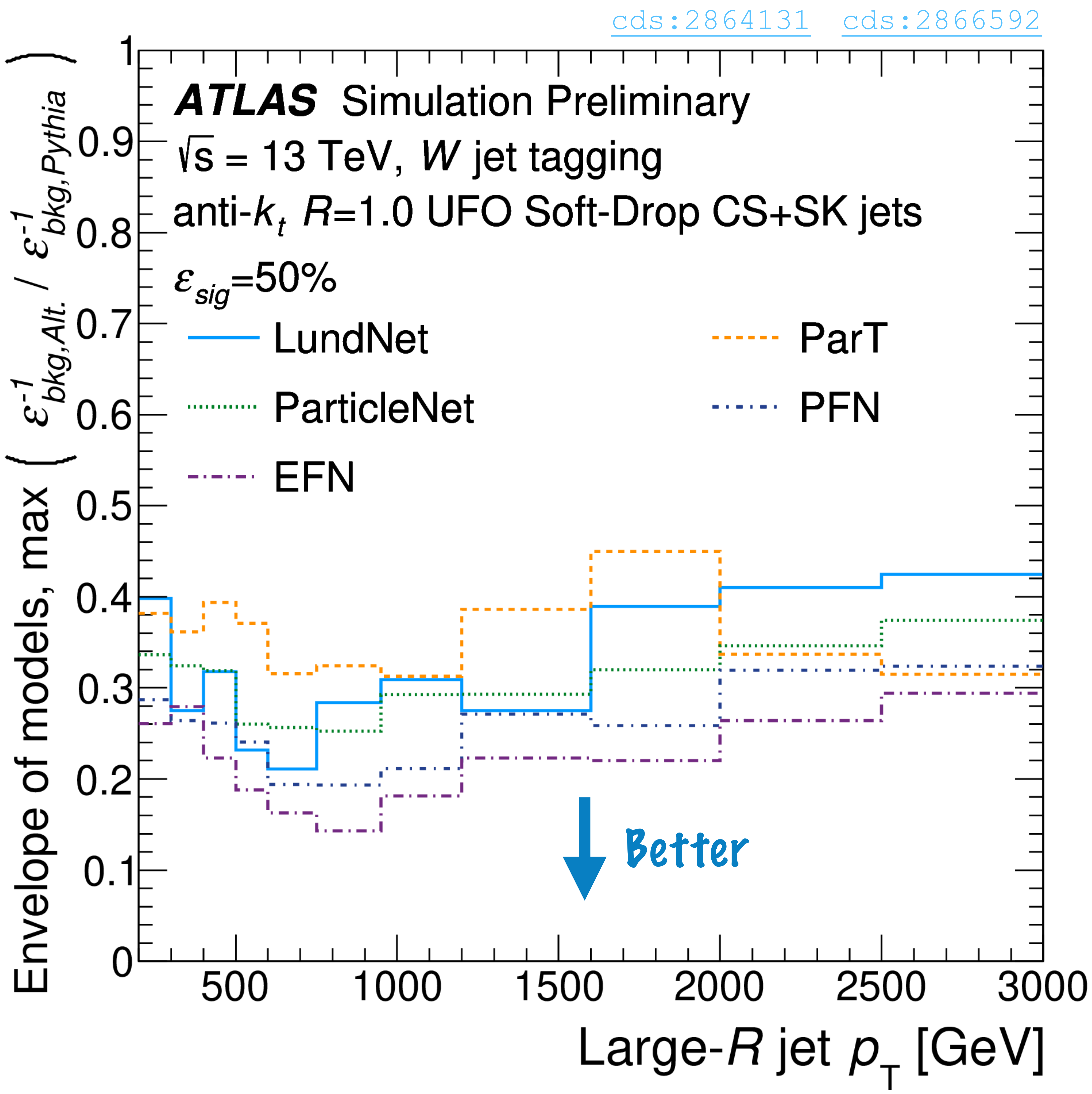
Theoretical

- Parton shower + hadronization modeling
- Renormalization scale
- Cross sections
- PDFs

Experimental

- Jet mass scale / resolution or similar
- Statistical
- Other SFs (e.g. b-tagging)
- Luminosity

W Tagger Modeling Dependence



- The most powerful taggers (LundNet, ParT) show variations in performance of up to 40%
- Likely to produce larger SF uncertainties

Beyond Modeling Dependence

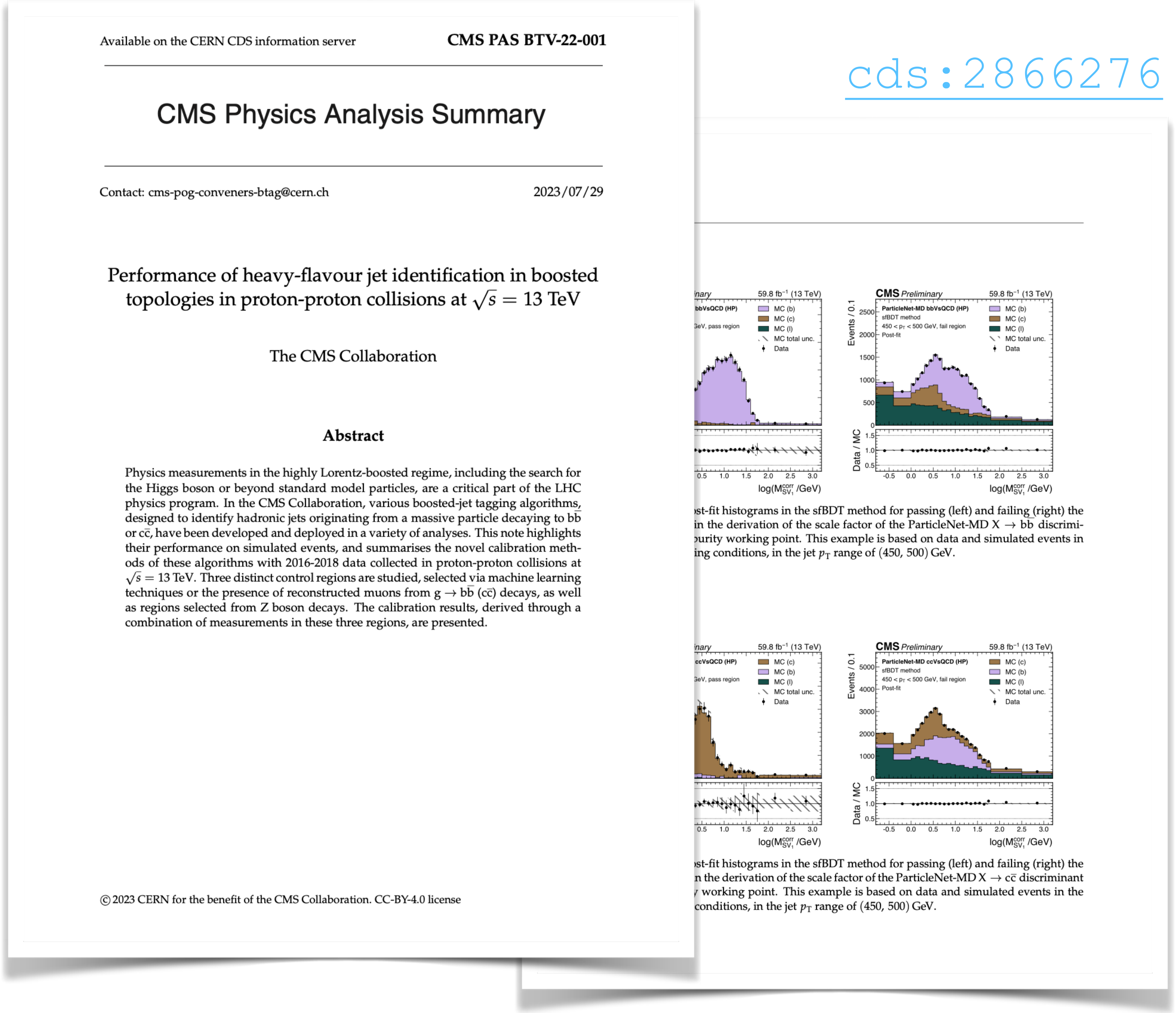
Modeling uncertainties were dominant for simple high-level quantity based taggers.

What about constituent based taggers?

cds:2724149

Systematic Group	W tagger p_T bins [GeV]			
	[200,250]	[250,300]	[300,350]	[350,600]
Statistical	0.01	0.02	0.03	0.04
Theory	< 0.01	< 0.01	< 0.01	< 0.01
$t\bar{t}$ modeling	0.21	0.20	0.15	0.12
Large-R jet	0.01	0.01	< 0.01	< 0.01
Other experimental	< 0.01	< 0.01	< 0.01	< 0.01
b-tagging	< 0.01	< 0.01	< 0.01	< 0.01
Total Uncertainty	0.21	0.20	0.15	0.12

cds:2866276



Available on the CERN CDS information server — CMS PAS BTV-22-001

CMS Physics Analysis Summary

Contact: cms-pog-conveners-btag@cern.ch — 2023/07/29

Performance of heavy-flavour jet identification in boosted topologies in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

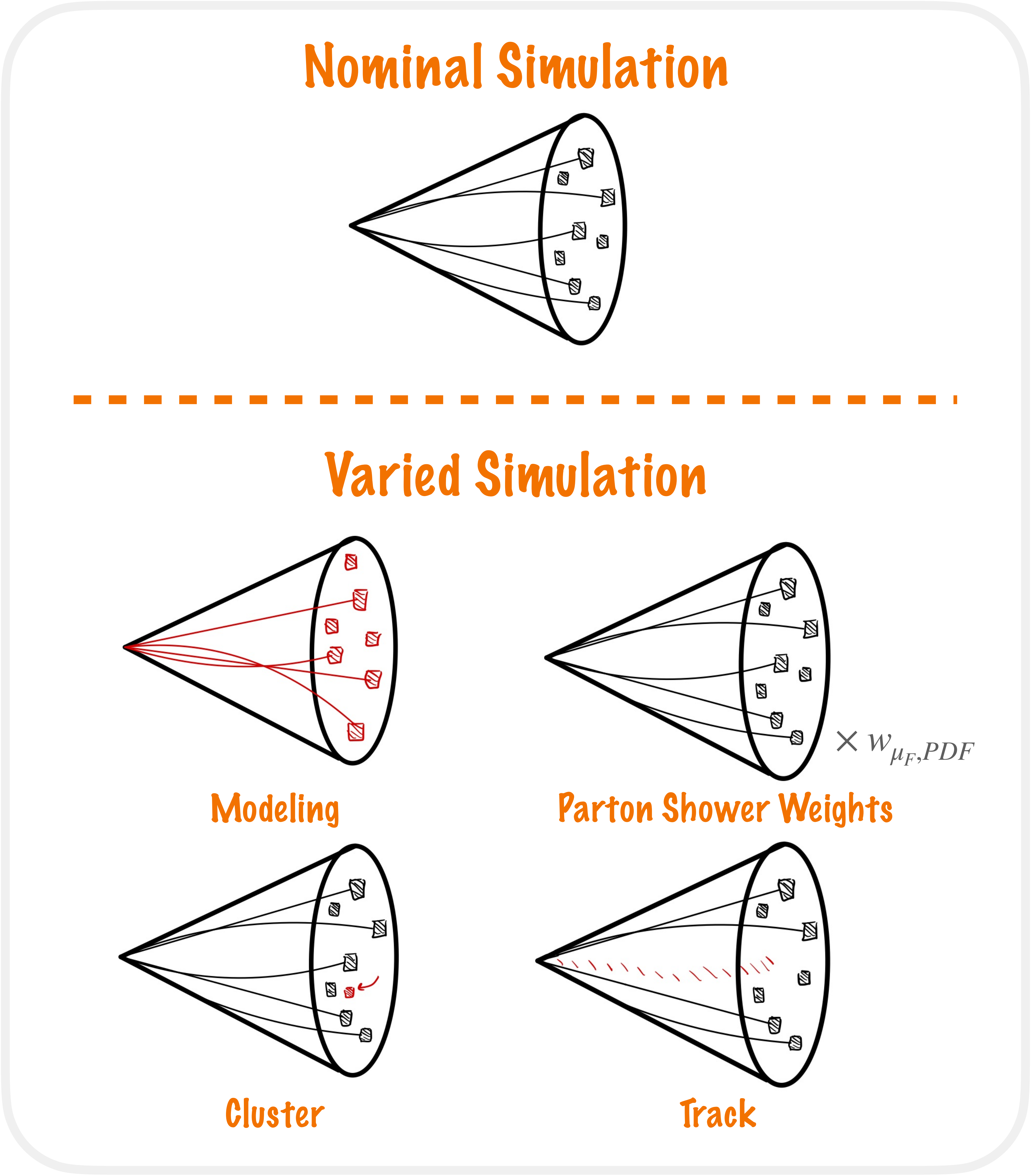
Physics measurements in the highly Lorentz-boosted regime, including the search for the Higgs boson or beyond standard model particles, are a critical part of the LHC physics program. In the CMS Collaboration, various boosted-jet tagging algorithms, designed to identify hadronic jets originating from a massive particle decaying to $b\bar{b}$ or $c\bar{c}$, have been developed and deployed in a variety of analyses. This note highlights their performance on simulated events, and summarises the novel calibration methods of these algorithms with 2016-2018 data collected in proton-proton collisions at $\sqrt{s} = 13$ TeV. Three distinct control regions are studied, selected via machine learning techniques or the presence of reconstructed muons from $g \to b\bar{b}$ ($c\bar{c}$) decays, as well as regions selected from Z boson decays. The calibration results, derived through a combination of measurements in these three regions, are presented.

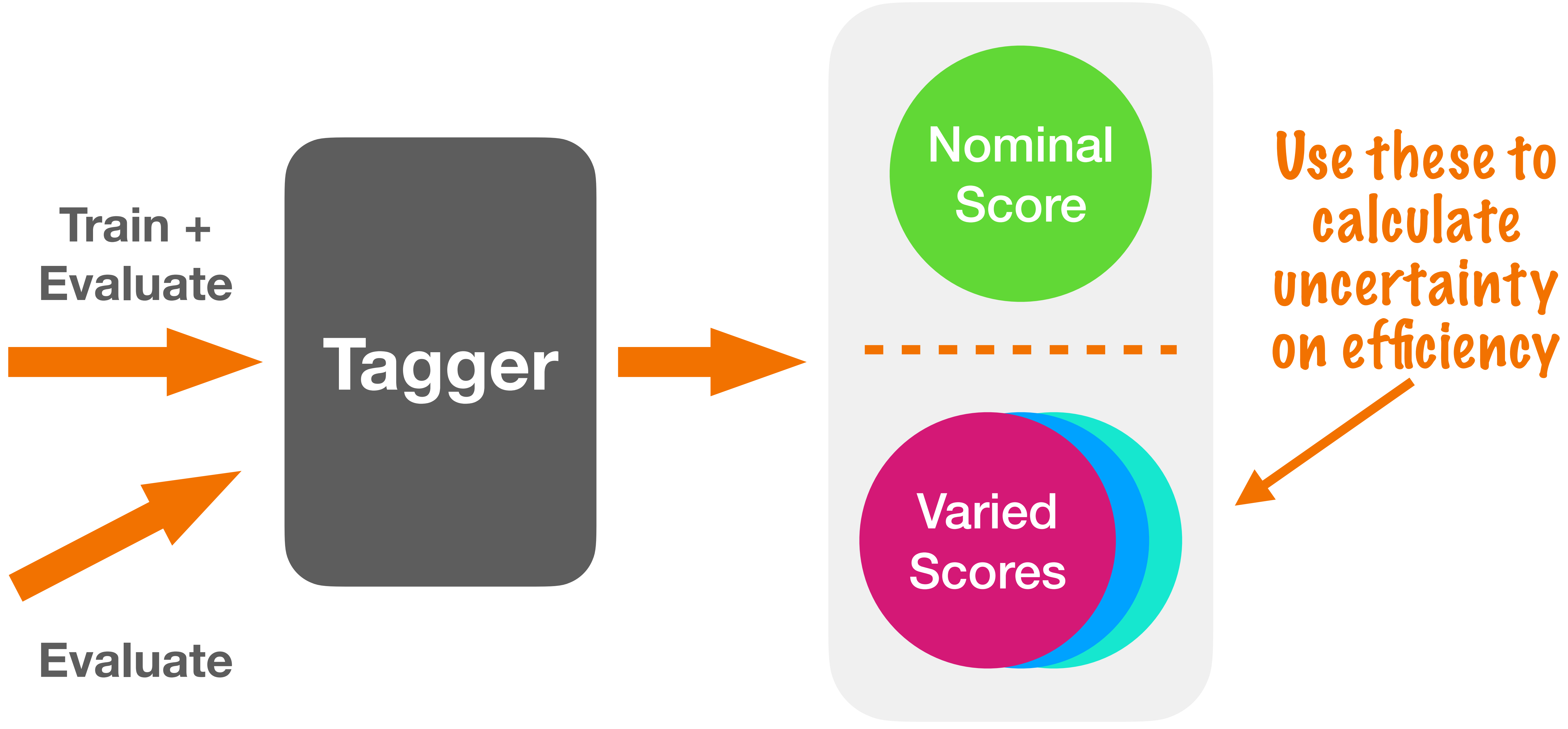
© 2023 CERN for the benefit of the CMS Collaboration. CC-BY-4.0 license

Measuring scale factors is difficult, and only possible within collaborations.

Can we find something approximate everyone can use?

Bottom-up Uncertainties



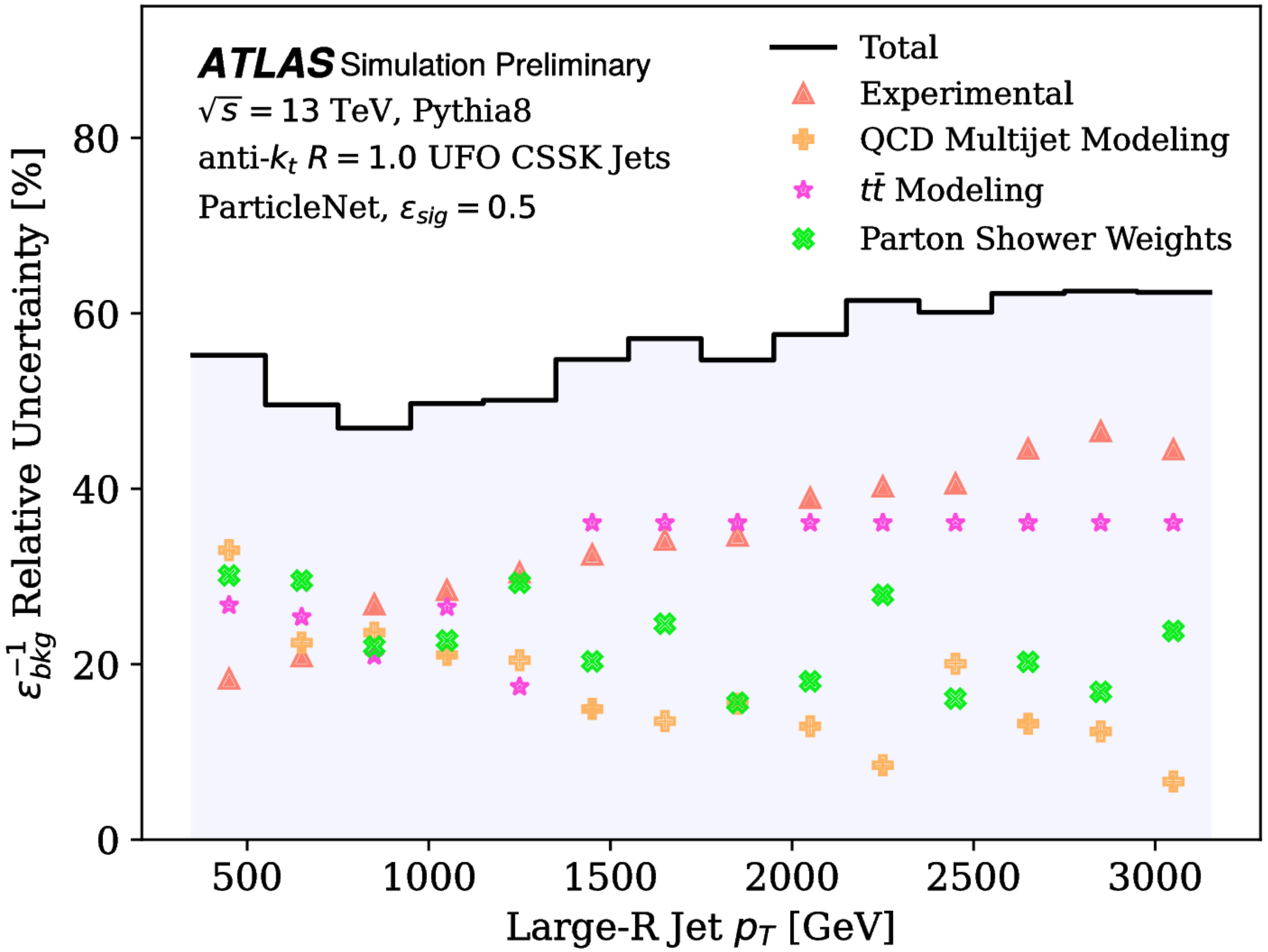


Benefits

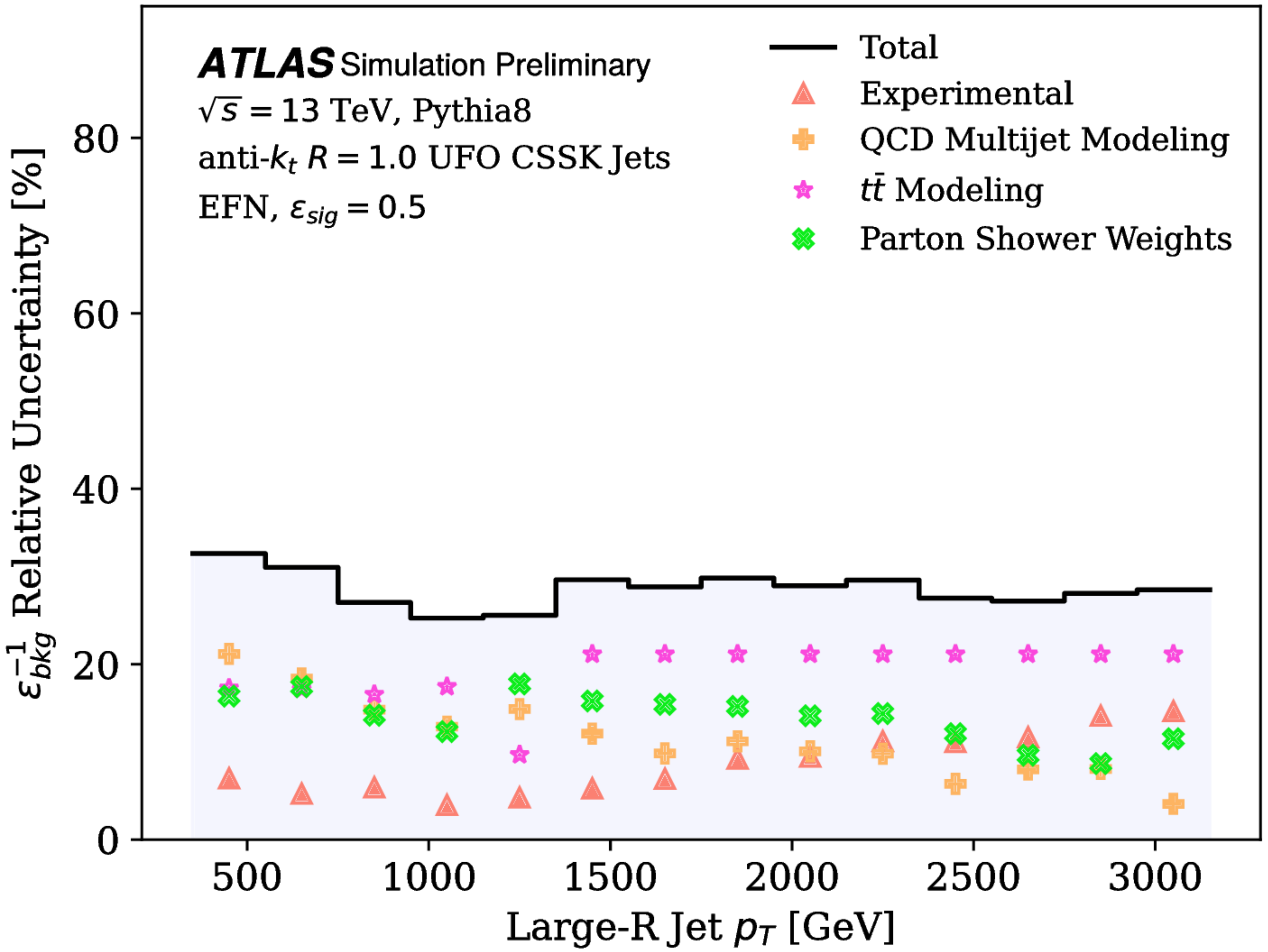
- No data required
- Once varied simulation is generated, can be used for arbitrary tagger
- Can define uncertainties on tagger efficiency with no signal enriched region in data

Top Tagger Uncertainties

Note these **are not** scale factor uncertainties. Expected to be conservative, but relative sensitivity of taggers is important.



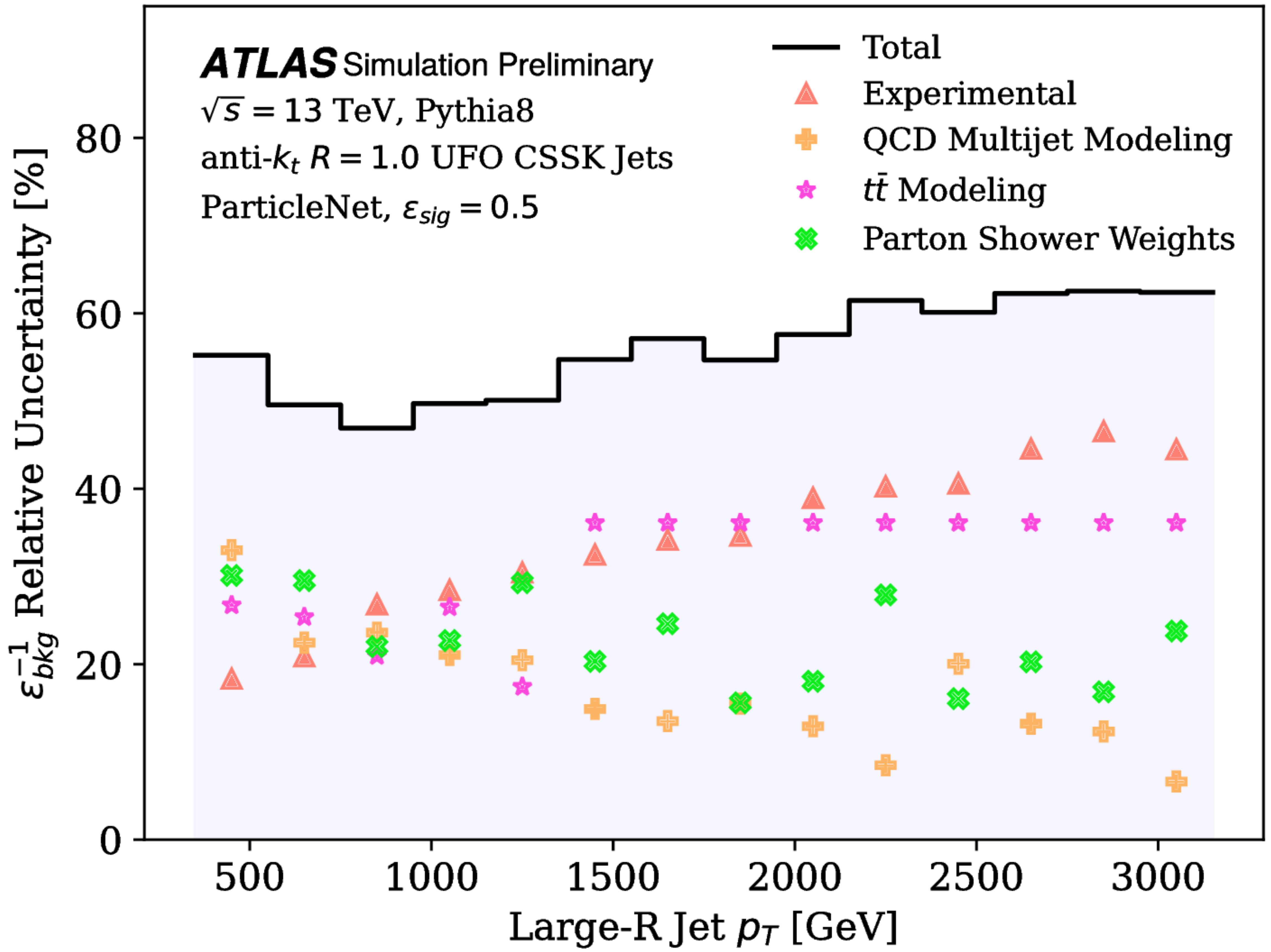
Large and Powerful GNN



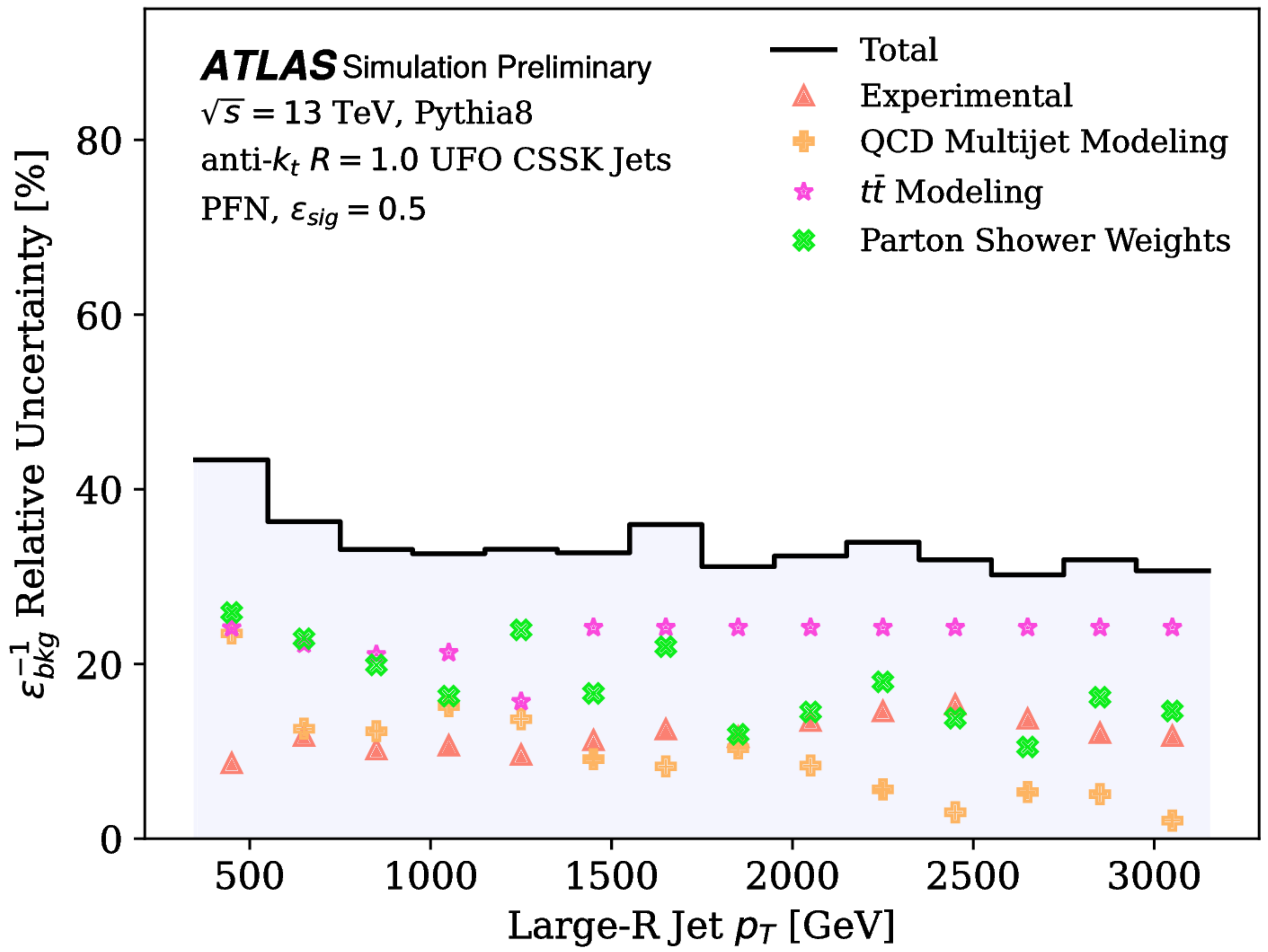
Theory Motivated IRC Safe Tagger

Top Tagger Uncertainties

Note these **are not** scale factor uncertainties. Expected to be conservative, but relative sensitivity of taggers is important.



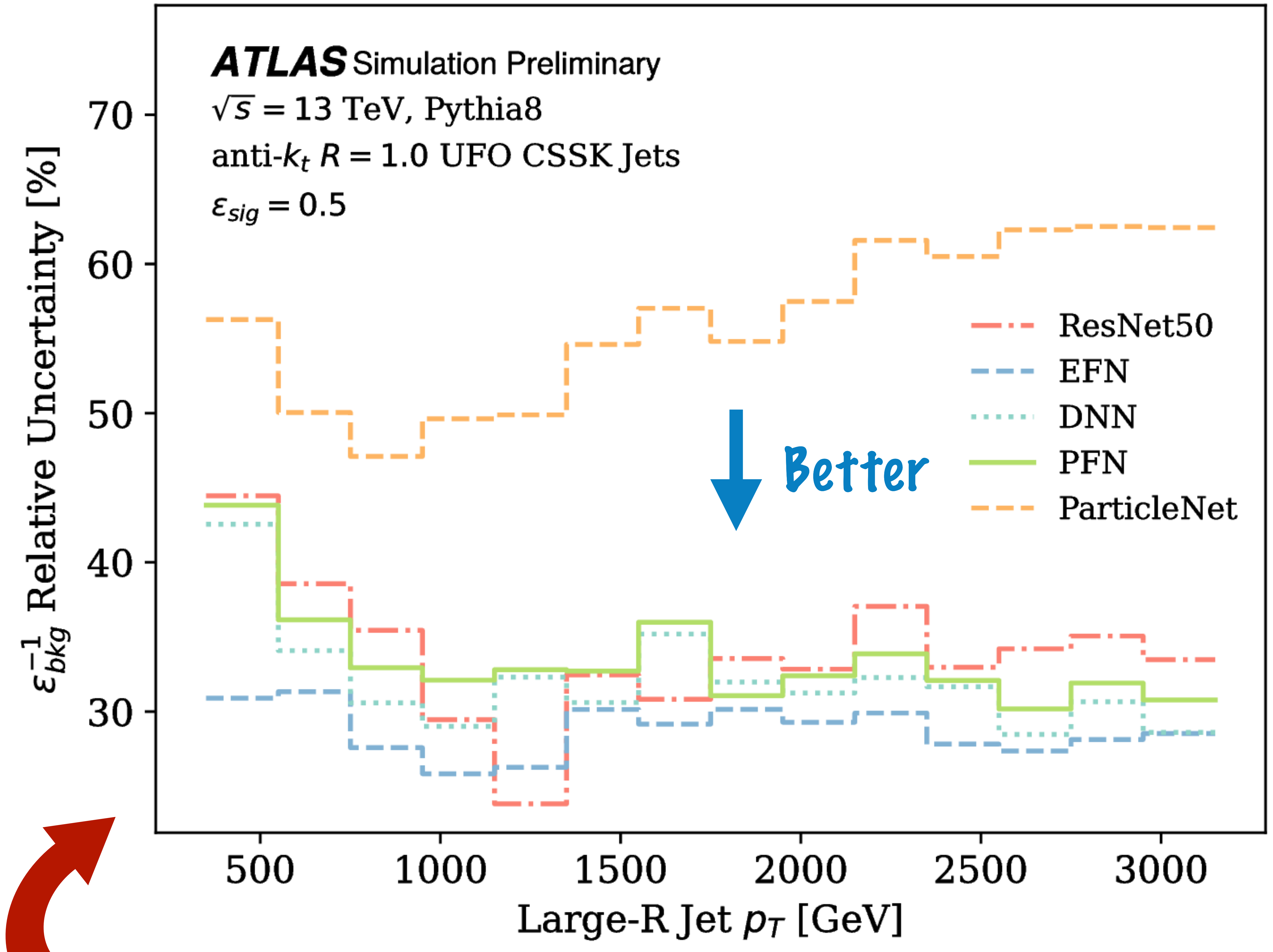
Large and Powerful GNN

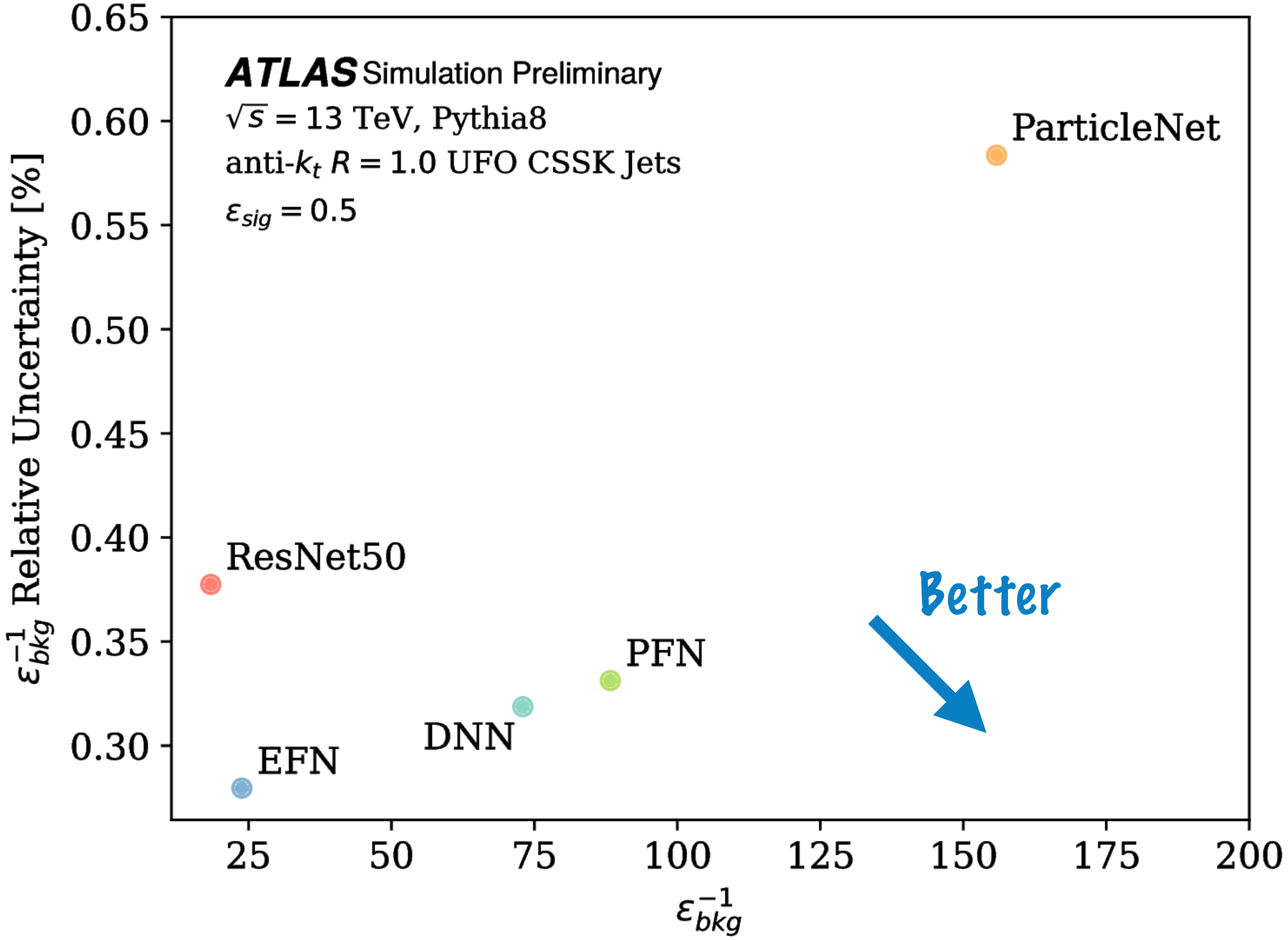


Deep Set (PFN)

- ParticleNet more sensitive to modeling
- Dramatically more sensitive to experimental variations

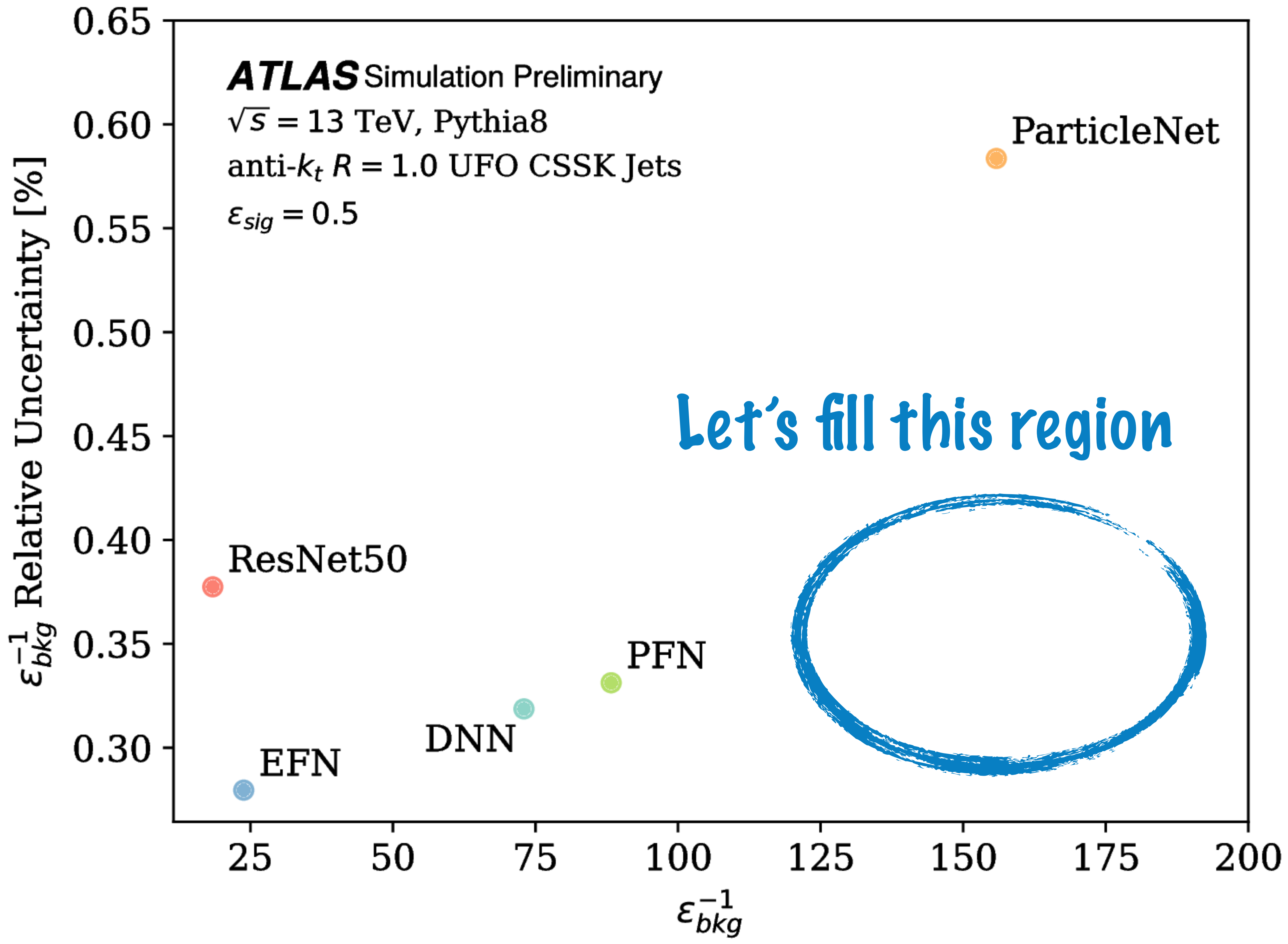
Uncertainty Comparison





Larger uncertainties here are expected to produce larger SF uncertainties

Conclusions



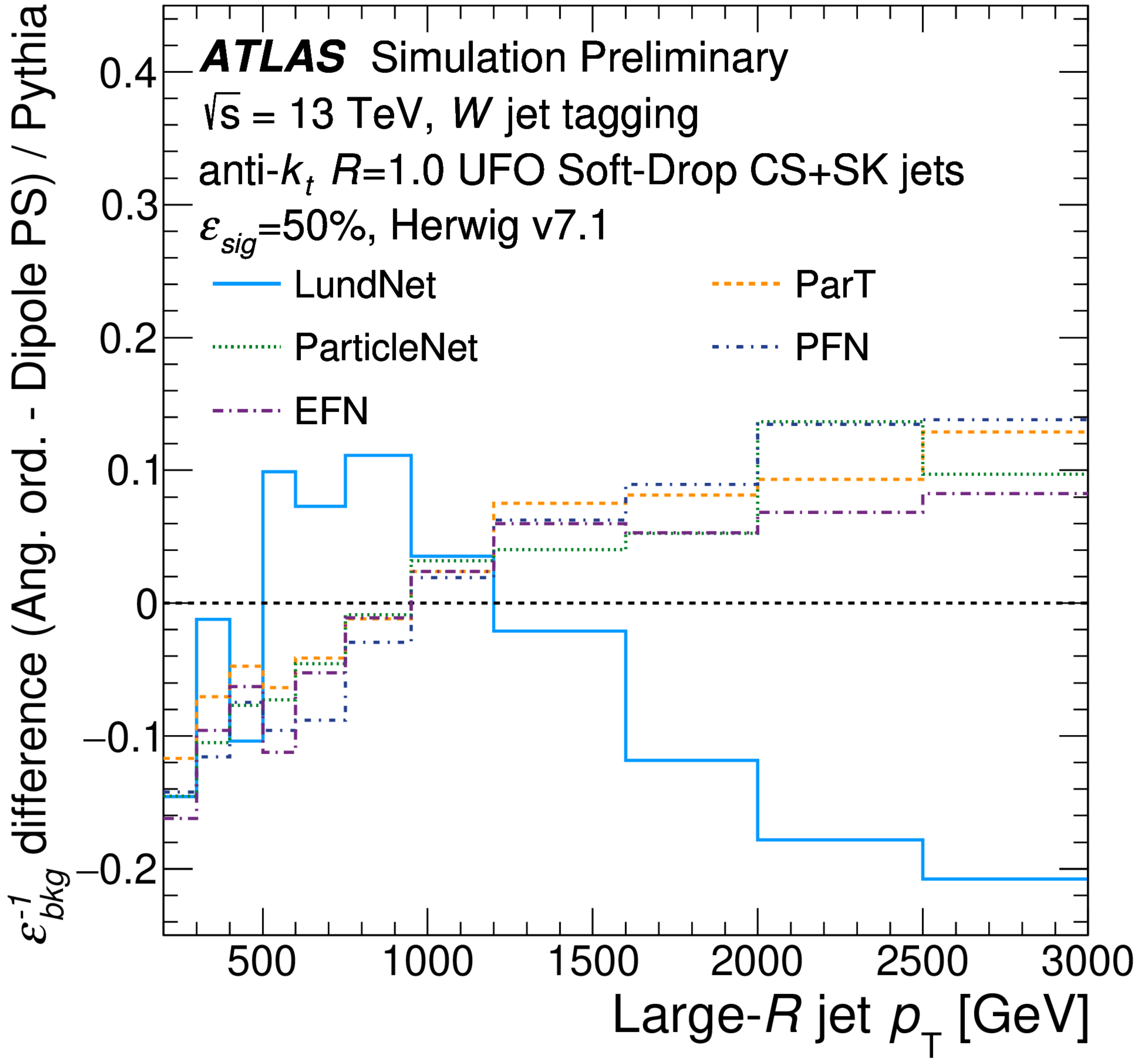
Working on public data!

- Powerful ML based jet tagging is deployed and producing physics!
- However, the more powerful the tagger, the larger the uncertainties
    - Could be limiting for some analyses

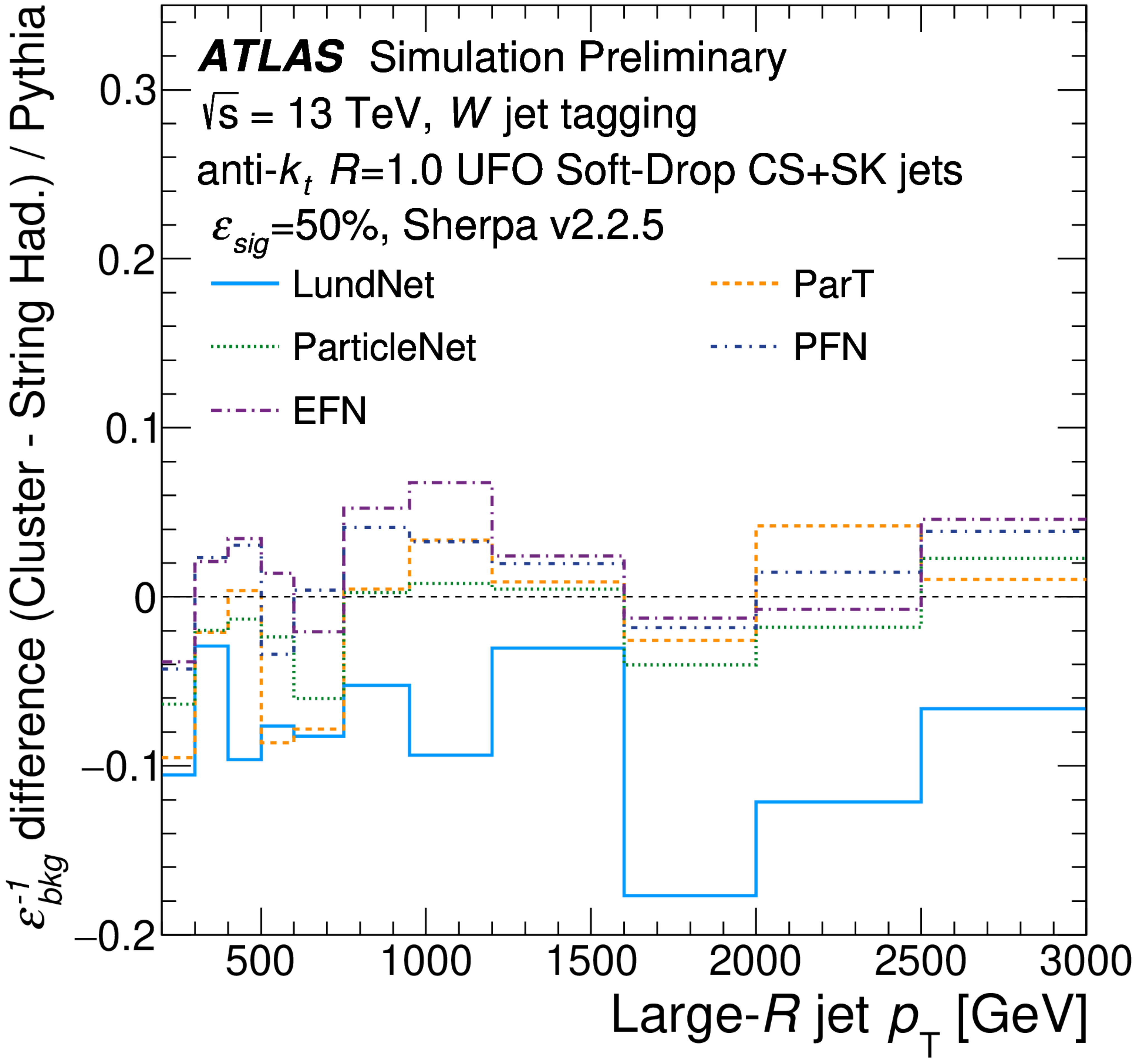
The new frontier is high performance and low uncertainties

Backup

W Tagger Modeling Dependence



Parton Shower



Hadronization

Top Tagging Systematic Variations

Modify nominal
Alternative samples
Pythia shower weights

Experimental	Theoretical

Experimental

- Calorimeter Clusters[1]
  - Energy Scale (Up / Down)
  - Energy Resolution
  - Position resolution
- Tracks
  - Fake rate
  - Efficiency
  - Sagitta bias

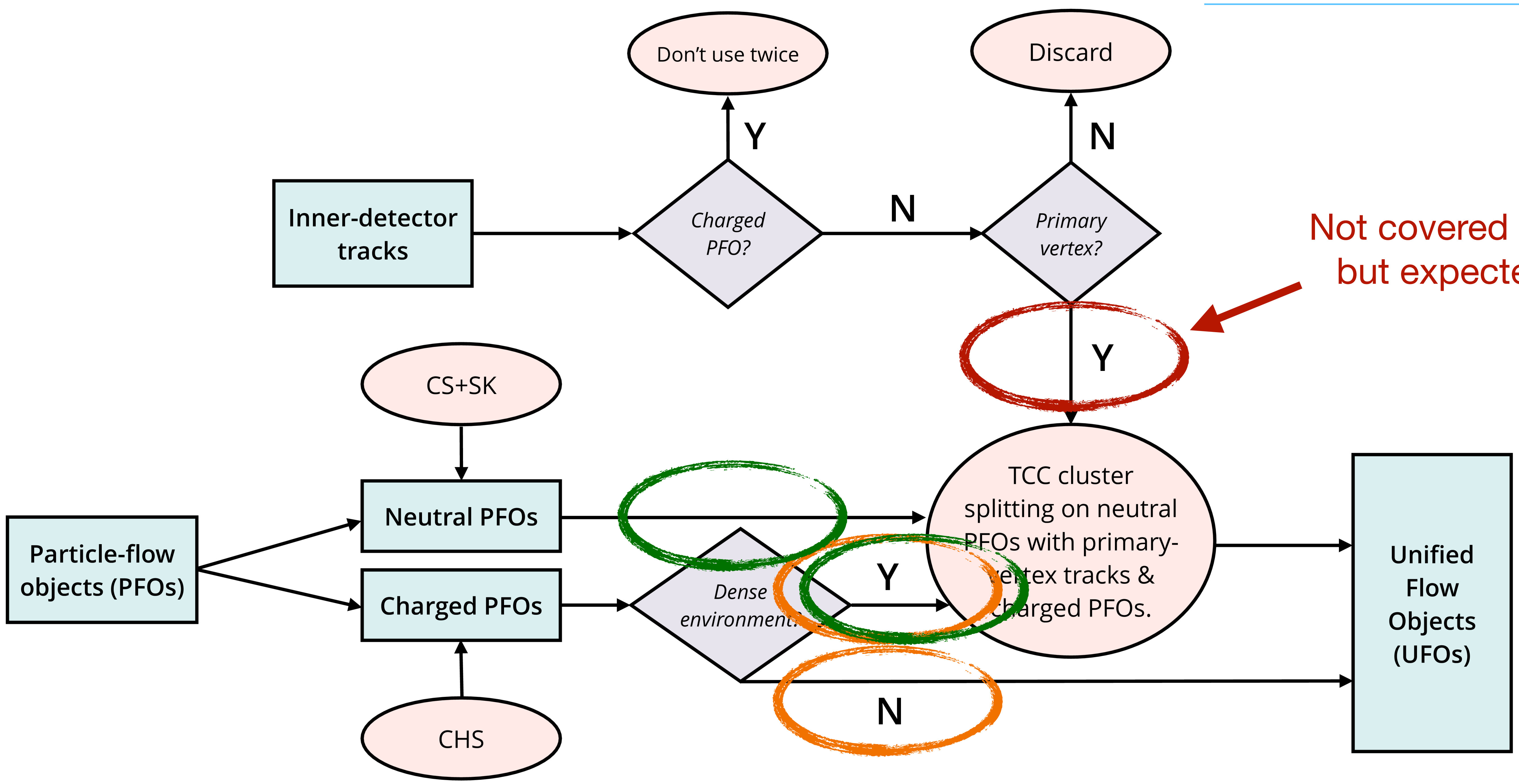
1 - arxiv: 1912.0983, arxiv:1903.02942, arxiv:2108.09043

Theoretical

- $t\bar{t}$ modeling
  - Compare Pythia to Herwig in SM $t\bar{t}$ samples
- QCD multijet modeling
  - Compare Herwig angular ordered to dipole parton shower
  - Compare Sherpa cluster to string based hadronization model
- Renormalization scale
  - Vary scale up/down by factors of 2
- PDFs
  - Vary PDFs up/down

Experimental Uncertainties

arxiv:2009.04986



Tracks

- Apply to charged and "merged" UFOs
- Track fake rate and efficiency
- Track bias

Calorimeter Clusters

- Apply to neutral and "merged" UFOs
- Cluster energy scale and resolution
- Cluster position resolution

Top Tagger Uncertainties

Note these **are not** scale factor uncertainties. Expected to be conservative, but relative sensitivity of taggers is important.