Systematic exploration of multi-scale jet substructure in $p+p$ collisions at $\sqrt{s} = 200$ GeV by the STAR experiment

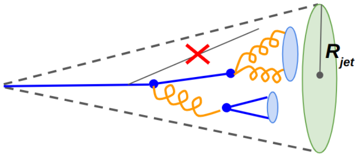
Monika Robotková for the STAR Collaboration

Nuclear Physics Institute, Czech Academy of Sciences
Czech Technical University in Prague

Hard Probes 2023, Aschaffenburg, Germany
26 - 31 March, 2023

Jet substructure and SoftDrop

- Study of jet substructure can help understand partonic fragmentation and hadronization processes
- Our goal is to access parton showers through experimental observables
- Grooming technique called SoftDrop used to remove soft wide-angle radiation from the jet in order to mitigate non-perturbative effects
- Connects parton shower and angular tree



$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut}\theta^{\beta},$$

where $\theta = \frac{\Delta R_{12}}{R_{jet}}$

$p_{T,1}, p_{T,2}$ - transverse momenta of the subjets
z_{cut} - threshold (0.1)
β - angular exponent (0)
ΔR_{12} - distance of subjets in the rapidity-azimuth plane

- Iterative SoftDrop used to study first, second, and third splits

Substructure observables

Momentum and angular observables

z_g	**shared momentum fraction**	$z_g \equiv \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1}+p_{T,2}}$
R_g	**groomed radius**	first ΔR_{12} that satisfies SoftDrop condition
k_T	**splitting scale**	$k_T = z_g p_{T,jet} \sin R_g$

Mass observables

M	**jet mass**	$M = \lvert \sum_{i \in jet} p_i \rvert = \sqrt{E^2 - \lvert\vec{p}\rvert^2}$
M_g	**groomed jet mass**	jet mass after grooming
μ	**groomed mass fraction**	$\mu \equiv \frac{\max(m_{j,1}, m_{j,2})}{M_g}$

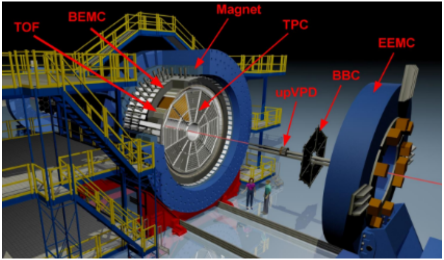
STAR experiment

TPC - *Time Projection Chamber*

- Detection of charged particles for jet reconstruction
- Transverse momenta of tracks: $0.2 < p_{\mathrm{T}} < 30$ GeV$/c$

BEMC - *Barrel Electromagnetic Calorimeter*

- Detection of neutral particles for jet reconstruction
- Granularity $(\Delta\eta \times \Delta\phi) = (0.05 \times 0.05)$
- Tower requirements: $0.2 < E_{\mathrm{T}} < 30$ GeV



Full azimuthal angle, $|\eta| < 1$

Dataset:
p+p, $\sqrt{s} = 200$ GeV, 2012
Algorithms:
anti-k_{T}, Cambridge/Aachen
Jets:
Full jets, $20 < p_{\mathrm{T,jet}} < 50$ Gev$/c$

Detector effects correction

- Measurement is affected by finite efficiency and resolution of the instrumentation
- Our goal is to deconvolve detector effects and obtain true distribution from measured one

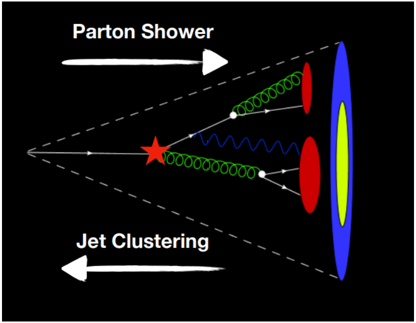
(2+1)D unfolding (D'Agostini. arXiv:1010.0632(2010))

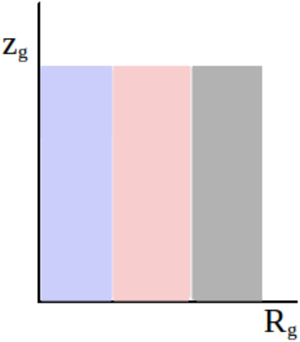
- 2D unfolding via Iterative Bayesian unfolding
- Correction on ensemble level for the 3rd dimension

MultiFold (Andreassen et al. PRL 124, 182001 (2020))

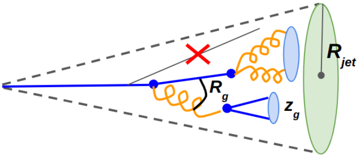
- Machine learning method
- New tool at RHIC
- All observables are simultaneously unfolded in an unbinned way

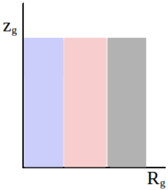
Correlation between substructure observables at the first split



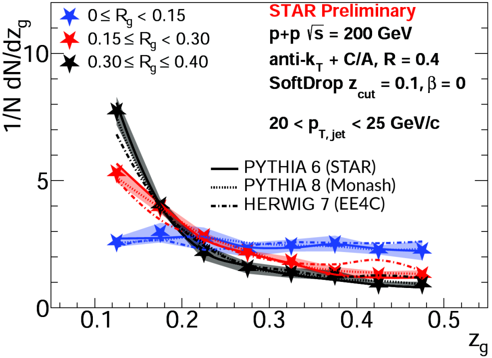


z_g vs. R_g at the first split



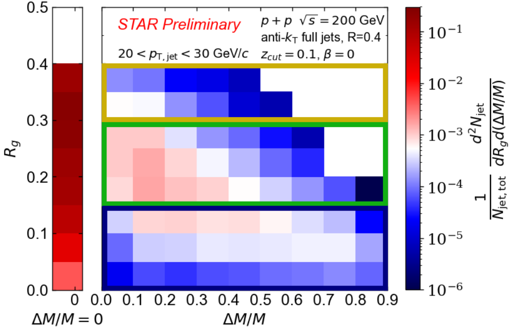


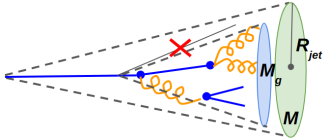
$$z_g \equiv \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1}+p_{T,2}}$$



- When we move from collinear hard splitting to softer wide angle splitting, z_g distribution becomes **steeper** and more **perturbative**
- MC models describe the trend of the data

R_g vs. $\Delta M/M$ at the first split



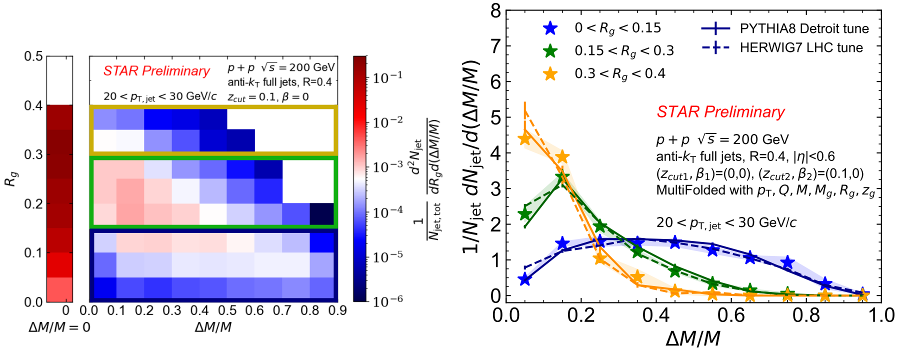


$$\Delta M = M - M_g \text{ [GeV]}$$

Collinear Drop

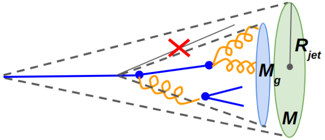
- Probes the soft component of the jet
- Difference of an observable with two different SoftDrop settings of parameters $(z_{\text{cut},1}, \beta_1)$ and $(z_{\text{cut},2}, \beta_2)$
- Our case: $(z_{\text{cut},1}, \beta_1) = (0, 0)$, $(z_{\text{cut},2}, \beta_2) = (0.1, 0)$

R_g vs. $\Delta M/M$ at the first split



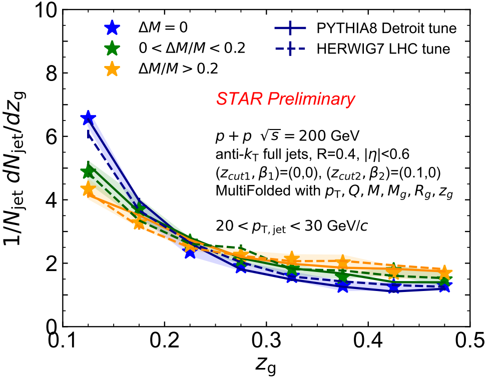
- The $\Delta M/M$ distribution is **anti-correlated** with R_g, which is consistent with angular ordering of the parton shower
- Large groomed jet radius → little/no soft wide angle radiation (small $\Delta M/M$) in the shower
- MC models describe the trend of the data

z_g vs. $\Delta M/M$ at the first split



$$\Delta M = M - M_g \text{ [GeV]}$$

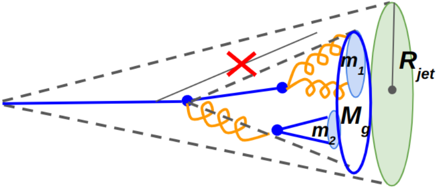
- The more mass that is groomed away relative to the ungroomed mass, the **flatter** and more **non-perturbative** the z_g distribution is

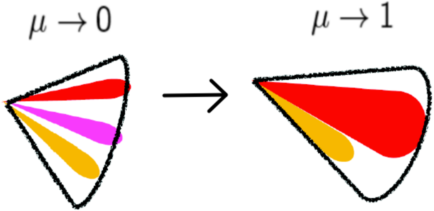


- The first splitting that passes SoftDrop can be non-perturbative → application of the $\Delta M = 0$ selection can filter out the jets with large non-perturbative contribution

μ vs. R_g at the first split

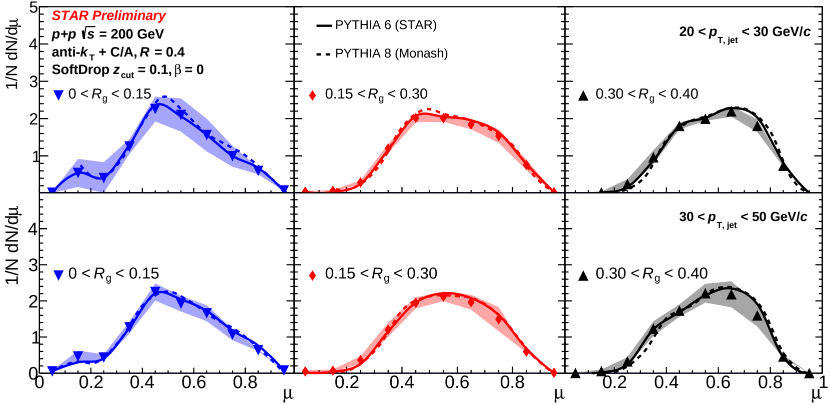
$$\mu \equiv \frac{\max(m_{j,1}, m_{j,2})}{M_g}$$





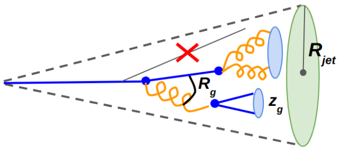
μ allows us to study mass sharing of the hard splitting

μ vs. R_g at the first split for two different $p_{T,jet}$ bins



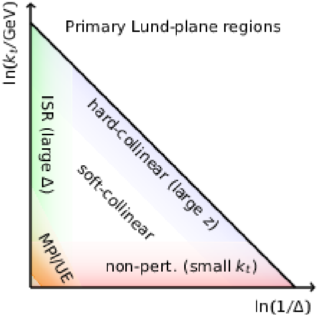
- Dependence on R_g much **weaker** than $\Delta M/M$, largely independent of $p_{T,jet}$, MC models agree with data
- μ shifts to smaller values at smaller angles, indicating a faster reduction of virtuality in the jet shower

$\log(k_T)$ vs. R_g at the first split



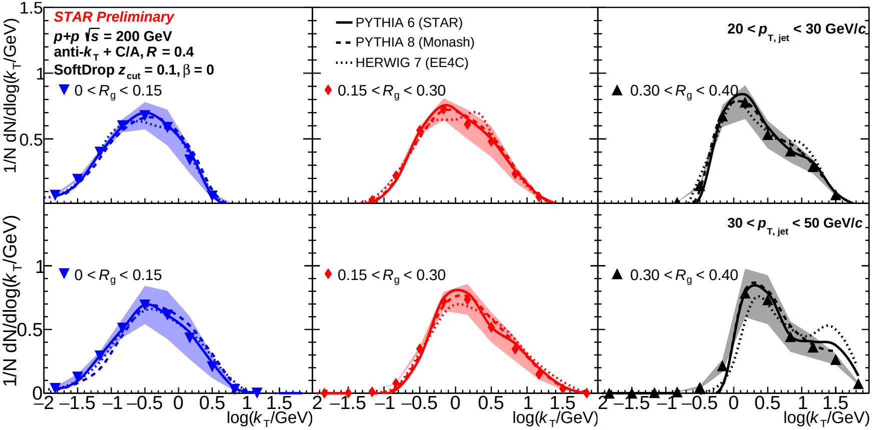
$$k_T = z_g p_{T,\text{jet}} \sin R_g$$

Dreyer, Salam, Soyez, JHEP 12 (2018) 064



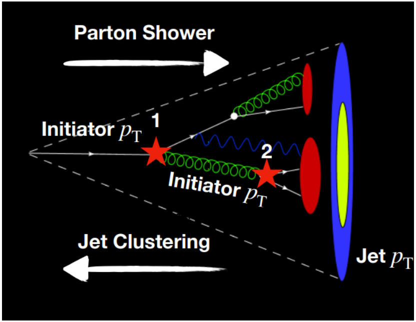
Cutting on R_g moves us to different k_T → we are probing different parts of the Lund Plane

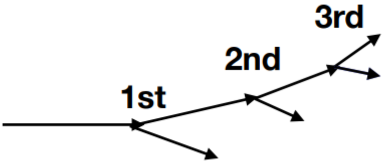
$\log(k_T)$ vs. R_g at the first split for two different $p_{T,jet}$ bins



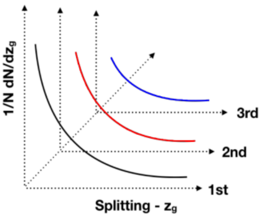
- $\log(k_T)$ has a **strong** dependence on R_g and **weak** dependence on $p_{T,jet}$, MC models describe the trend of the data
- 0 value corresponds to 1 GeV → we move from **non-perturbative** to **perturbative** region

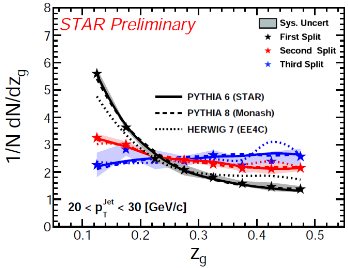
Evolution of the splitting observables as we travel along the jet shower

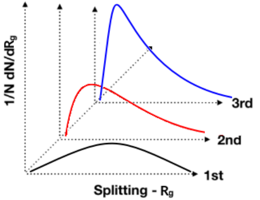


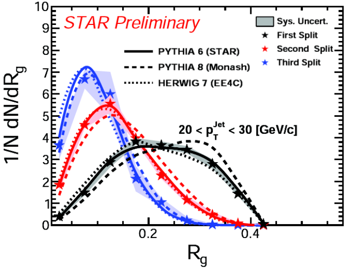


z_g and R_g distributions at 1st, 2nd, and 3rd splits









- Going from 1st $\rightarrow$ 3rd split
  - z_g distribution becomes **flatter**
  - R_g distribution becomes **narrower**
- **Collinear emissions are enhanced when going from 1st to 3rd split**

Summary

Correlation at the first split

- New methods for the unfolding were applied (MultiFold, (2+1)D unfolding)
- z_g, $\Delta M/M$, $\log(k_T)$ have a **weak** dependence on $p_{T,jet}$ and a **strong** dependence on R_g
- Study of different Lund Plane regions allows us to observe the correlations between jet substructure observables

Splits along the shower

- Observed significantly harder/symmetric splitting at the third/narrow split compared to the first and second splits

Selecting on the split number along the jet clustering tree results in similar change in z_g distributions as selecting on R_g or $\Delta M/M$ at the first split

Jet substructure measurements at RHIC energies allow to disentangle perturbative (early, wide splits) and mostly non-perturbative dynamics (late, narrow splits) within jet showers, and test validity of MC models

Thank you for your attention!

Back up

Jet clustering algorithms

- Jets are defined using algorithms

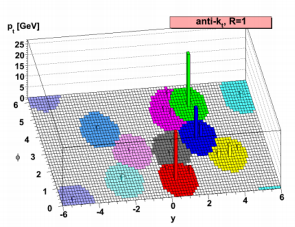
Anti-k_T algorithm

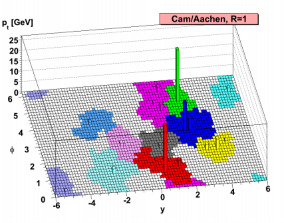
- $d_{ij} = \frac{\min(1/p_{Ti}^2, 1/p_{Tj}^2)\Delta R_{ij}^2}{R}$, $d_{iB} = 1/p_{Tj}^2$
- Clustering starts from the particles with the highest transverse momentum

Cambridge/Aachen (C/A) algorithm

- $d_{ij} = \Delta R_{ij}^2 / R^2$, $d_{iB} = 1$
- Particles are clustered exclusively based on angular separation, ideal to be used to resolve jet sub-structure

d_{iB} - distance of the particle i from the beam
p_T - transverse momentum
ΔR_{ij} - distance between the particle i and j
R - jet resolution parameter



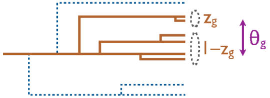


Cacciari, Salam, Soyez, JHEP 0804:063 (2008)

SoftDrop

- Grooming technique used to remove soft wide-angle radiation from the jet
- Connects parton shower and angular tree

1. Jets are first found using the anti-k_T algorithm
2. Recluster jet constituents using the C/A algorithm
3. Jet j is broken into two sub-jets j_1 and j_2 by undoing the last stage of C/A clustering
4. Jet j is final SoftDrop jet, if sub-jets pass the condition on the right, otherwise the process is repeated



Larkoski, Marzani, Thaler, Tripathee, Xue, Phys. Rev. Lett. 119, 132003 (2017)

- **Shared momentum fraction** z_g

$$z_g = \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut}\theta^{\beta},$$

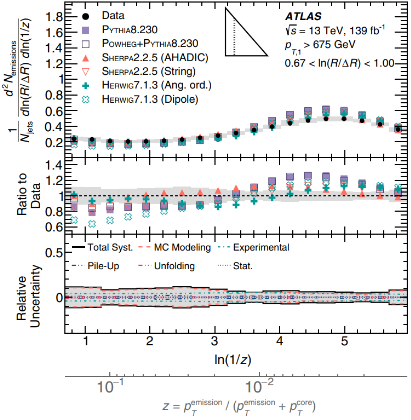
where $\theta = \frac{\Delta R_{12}}{R}$

- **Groomed radius** R_g - first ΔR_{12} that satisfies SoftDrop condition

$p_{T,1}, p_{T,2}$ - transverse momenta of the subjets
z_{cut} - threshold (0.1)
β - angular exponent (0)
ΔR_{12} - distance of subjets in the rapidity-azimuth plane

Lund Plane measurement

- Previous ATLAS measurement uses Lund jet plane
- Significant differences in varying hadronization models at high $p_{\mathrm{T,jet}}$ at the LHC $\rightarrow$ we want to study this at lower $p_{\mathrm{T,jet}}$, where non-perturbative effects are expected to be larger
- While Lund jet plane integrates over all splits, we focus on the first split



ATLAS, Phys. Rev. Lett. **124**, 222002 (2020)

Data analysis

- $p + p$ **collisions at** $\sqrt{s} = 200$ **GeV, 2012**
- $\sim$11 million events analyzed

Event and track selection

- Transverse momenta of tracks: $0.2 < p_{\mathrm{T}} < 30$ GeV/c
- Tower requirements: $0.2 < E_{\mathrm{T}} < 30$ GeV

Jet reconstruction

- Jets reconstructed with anti-k_{T} algorithm, reclustered with the C/A algorithm
- Transverse momenta of jets: $15 < p_{\mathrm{T,jet}} < 40$ GeV/c
- Resolution parameters: $R = 0.4$, $R = 0.6$
- SoftDrop parameters: $z_{\mathrm{cut}} = 0.1$, $\beta = 0$

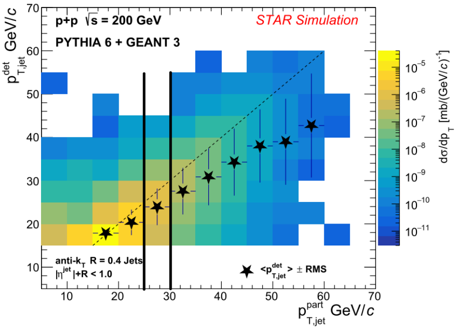
$$\frac{\min(p_{\mathrm{T},1}, p_{\mathrm{T},2})}{p_{\mathrm{T},1} + p_{\mathrm{T},2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R}\right)^{\beta}$$

2D Bayesian Unfolding

- 2D iterative Bayesian method implemented in the RooUnfold
- Procedure has following steps:
  1. The jets at the detector and particle level are reconstructed separately
  2. Jets are matched based on $\Delta R < 0.6$
  3. Jets without match - missed jet (particle level) and fake jets (detector level)
  4. Response between detector level and particle level for observables is constructed
- We use RooUnfold response which contains Matches and Fakes
  - Unfolding is done separately for p_{T}^{det} intervals 15-20, 20-25, 25-30, 30-40 GeV/c
- Then unfolded spectra are weighted with values from our projection and put together
- Together with trigger missed and unmatched weighted spectra we get our fully unfolded spectrum

Correction in 2+1D for z_g, R_g, and $p_{T,jet}$

- Results are in 3D → z_g vs. R_g is unfolded in 2D and correction for $p_{T,jet}$ in 1D is needed
  - For each particle-level $p_{T,jet}$ bin, we do projection of this bin into detector-level $p_{T,jet}$, and get the weights from detector-level $p_{T,jet}$ bins



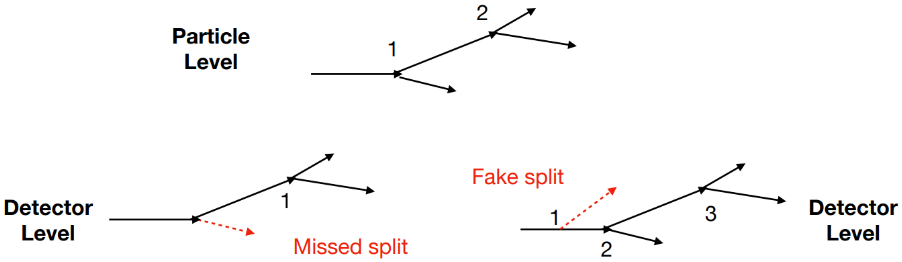
STAR, Phys. Lett. B 811 (2020) 135846

- We unfold z_g vs. R_g via iterative Bayesian unfolding in 2D using RooUnfold and unfolded spectra for each detector-level $p_{T,jet}$ bin are weighted and summed
- Additional corrections for trigger and jet finding efficiencies are applied

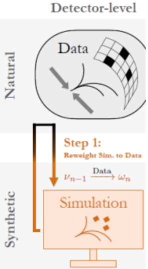
Details on systematic uncertainties available in back up

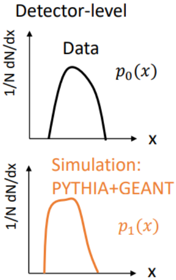
Correction in 2+1D for $p_{\mathrm{T,jet/initiator}}$, z_g, R_g

- Splits can be affected by detector efficiency and resolution
- Observables at a given split are smeared
- Splitting hierarchy is modified going from particle level to detector level



MultiFold





Where does the machine learning part come in?

E.g., Iteration 1, step 1:

Weights: $w(x) = p_0(x)/p_1(x)$ Ok for 1D

$$\approx f(x)/(1 - f(x))$$ (Andreassen and Nachman PRD 101, 091901 (2020))

where f(x) is a neural network and trained with the binary cross-entropy loss function

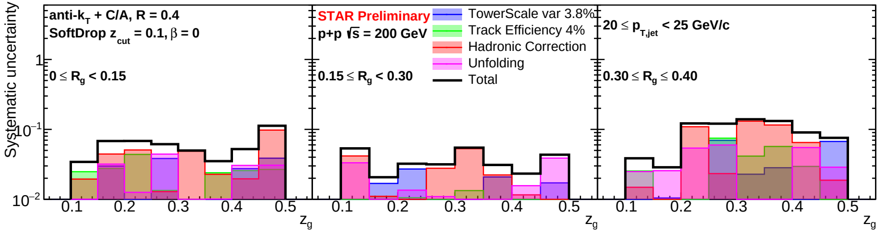
to distinguish jets coming from data vs from simulation

Unfolding → Reweighting histograms → Classification → Neural network

See backup slides for details of the neural networks.

Systematic uncertainties

- Systematic uncertainties estimated by varying the detector response
  - Hadronic correction - fraction of track momentum subtracted is varied
  - Tower scale variation - tower gain is varied by 3.8%
  - Tracking efficiency - efficiency is varied by 4%
  - Unfolding - iterative parameter is varied from 4 to 6
- Systematics due to prior shape variation will be included in the final publication



$0 \leq R_g < 0.15$ $\quad$ $0.15 \leq R_g < 0.30$ $\quad$ $0.30 \leq R_g \leq 0.40$