



Probing the Parton Shower and Hadronization with Novel Jet Substructure Measurements at STAR

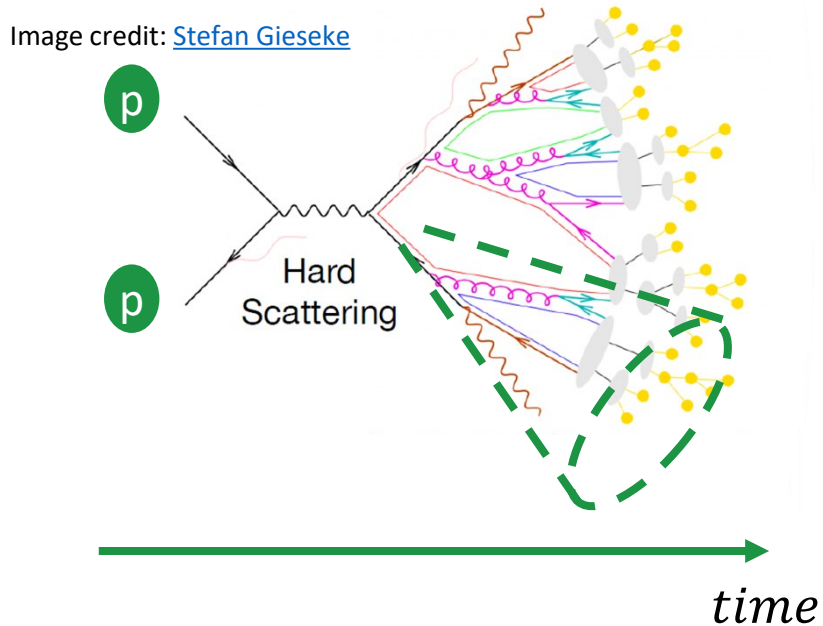
Youqi Song (youqi.song@yale.edu)
for the STAR Collaboration

The 39th Winter Workshop on Nuclear Dynamics,
Jackson, WY
02/13/2024





Jets



- Jets are collimated sprays of final-state hadrons (rare!).
- Jets can serve as **proxies for the hard-scattered parton** (with a proper choice of clustering algorithm, e.g., anti- k_T).
- Jet substructure measurements can be used to **probe parton shower and hadronization**.
- Note: This presentation will focus on measurements in pp.



Jet substructures

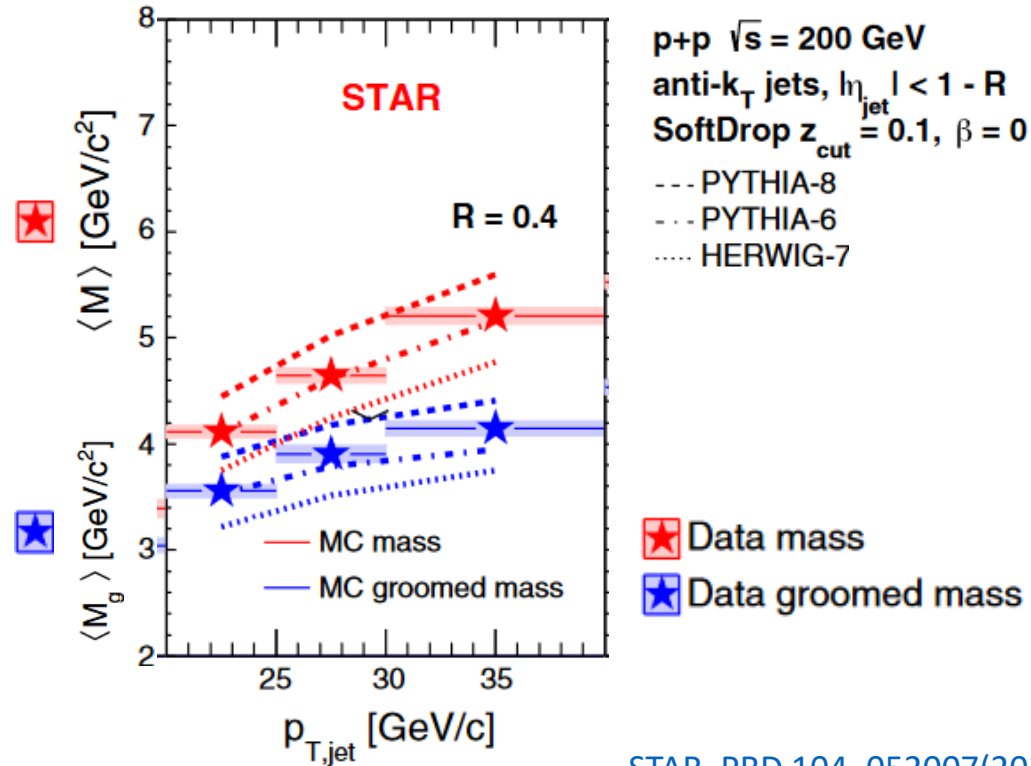
- Use subscript “ g ” to denote observables obtained through **SoftDrop grooming** which removes soft and wide-angle radiation from the jet

Example:

Ungroomed jet mass M

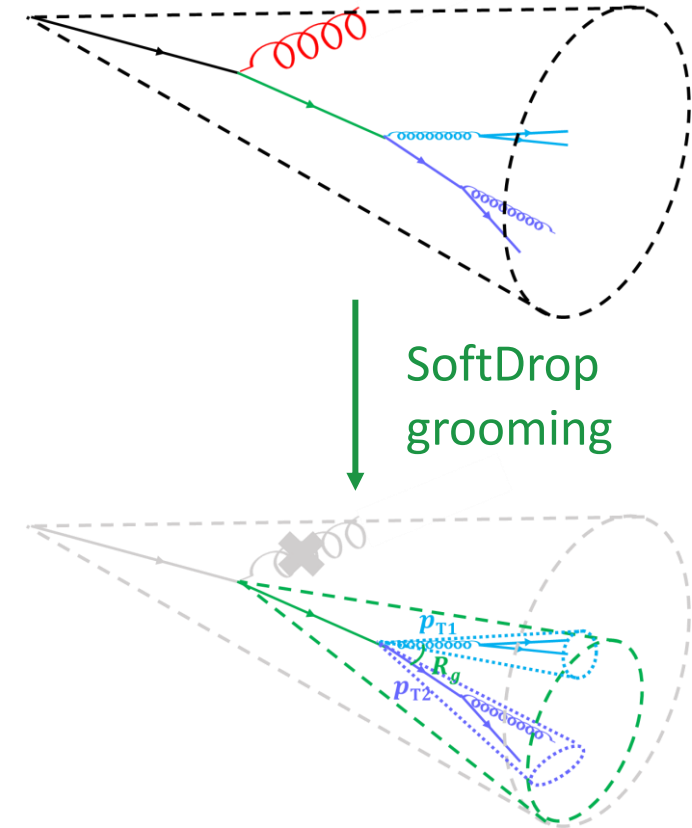
vs

SoftDrop groomed jet mass M_g

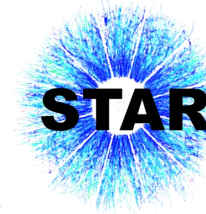


[STAR. PRD 104, 052007\(2021\)](#)

Effect:



$$M = \left| \sum_{i \in \text{jet}} p_i \right| = \sqrt{E^2 - |\vec{p}|^2}$$

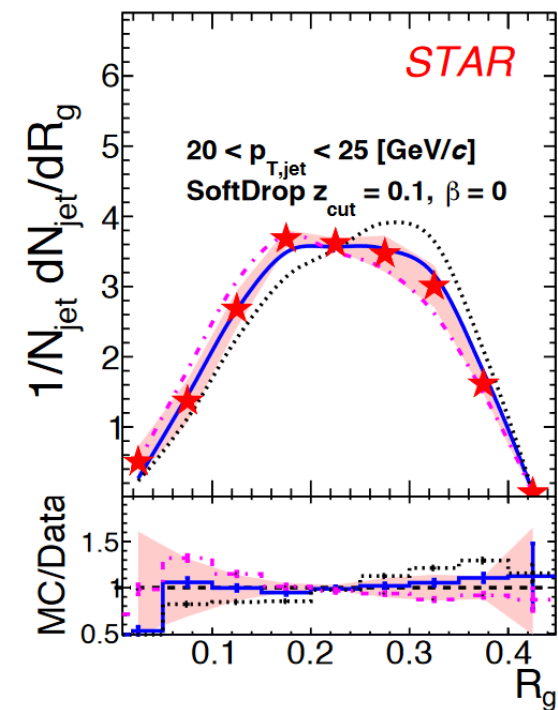


Jet substructures

- Use subscript “ g ” to denote observables obtained through **SoftDrop grooming** which removes soft and wide-angle radiation from the jet

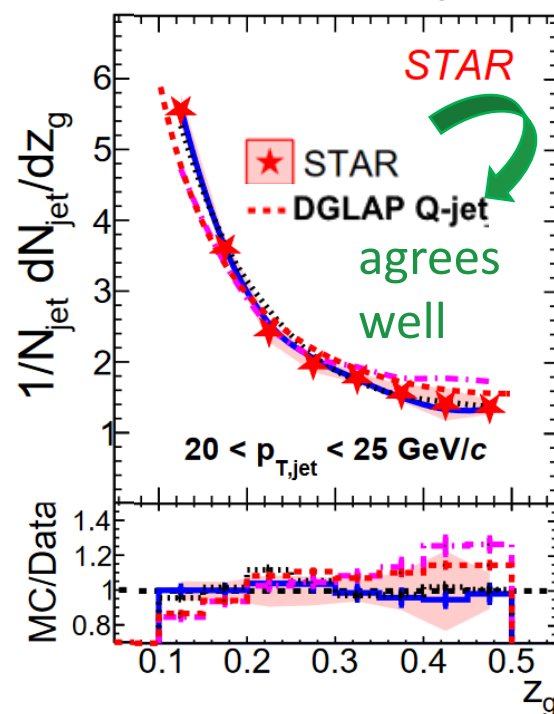
Example:

Groomed jet radius R_g



WWND, 02/13/2024

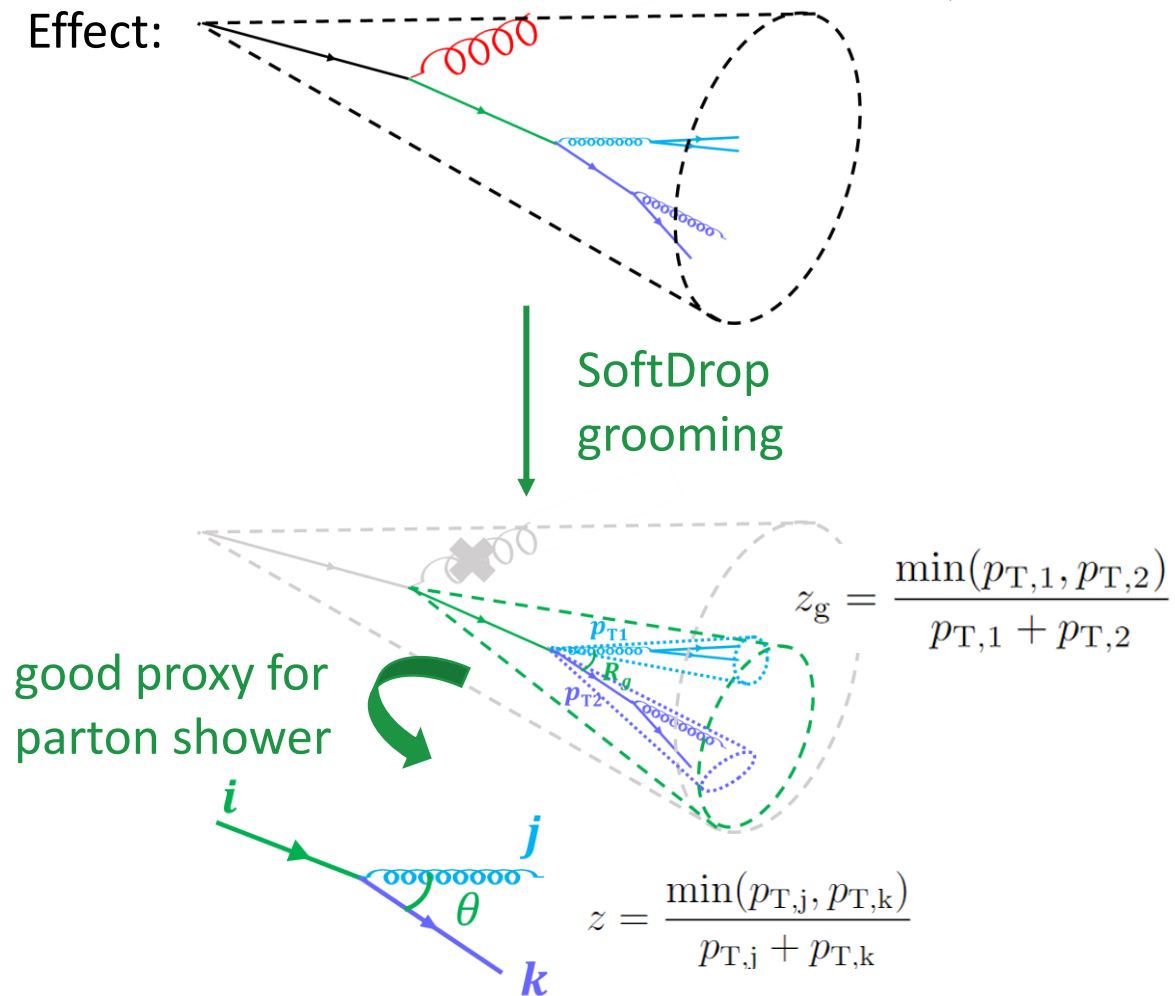
Momentum imbalance z_g



[STAR. PLB 811 \(2020\) 135846](#)

Youqi Song

Effect:



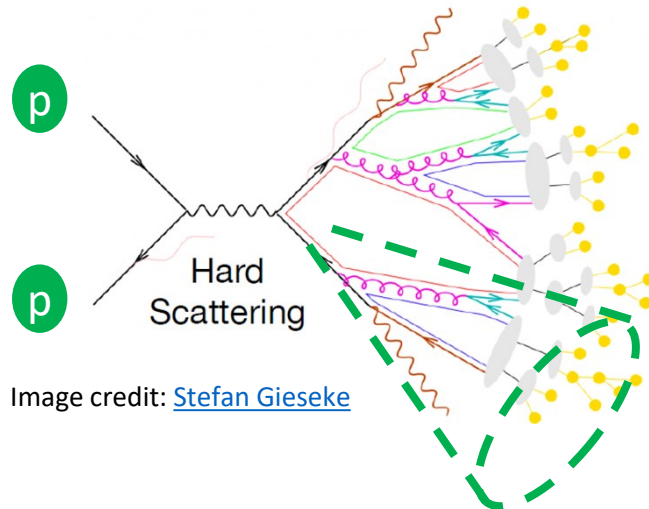


Image credit: [Stefan Gieseke](#)

Study the non-perturbative contribution in jets

- Soft component of the parton shower
→ CollinearDrop jet measurement
- Hadronization
→ Charge correlator ratio measurement



Motivation: CollinearDrop

Aims to probe the **soft and wide-angle radiation** within a jet

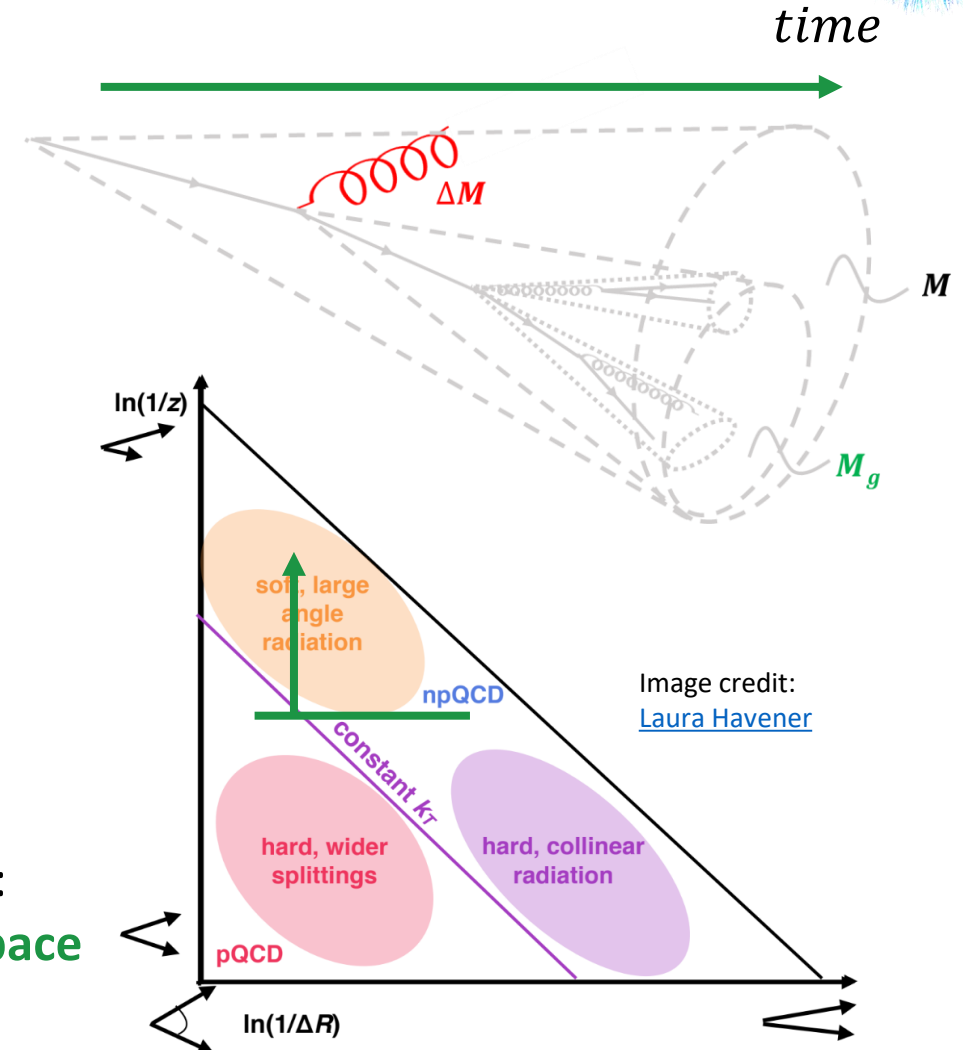
- General case: difference of an observable with two different SoftDrop selections $(z_{cut\ 1}, \beta_1)$ and $(z_{cut\ 2}, \beta_2)$
- For this analysis, $(z_{cut\ 1}, \beta_1) = (0,0)$ and $(z_{cut\ 2}, \beta_2) = (0.1,0)$: difference in the original and SoftDrop groomed observable

- Observables: e.g.,
$$\Delta M/M = \frac{M - M_g}{M}$$

where
$$M = \left| \sum_{i \in \text{jet}} p_i \right| = \sqrt{E^2 - |\vec{p}|^2}$$

soft and wide-angle radiation:
interesting **region of phase space**
that deserves more study!

Chien and Stewart JHEP 06 (2020) 64.

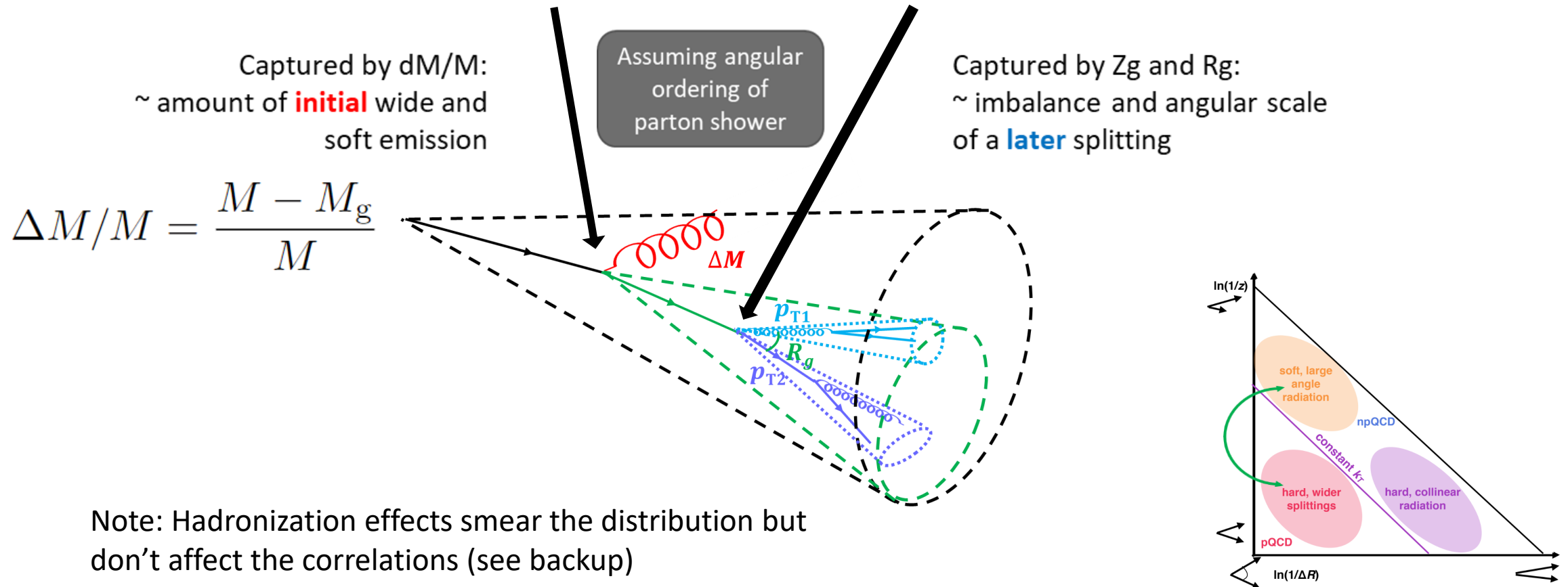




Motivation

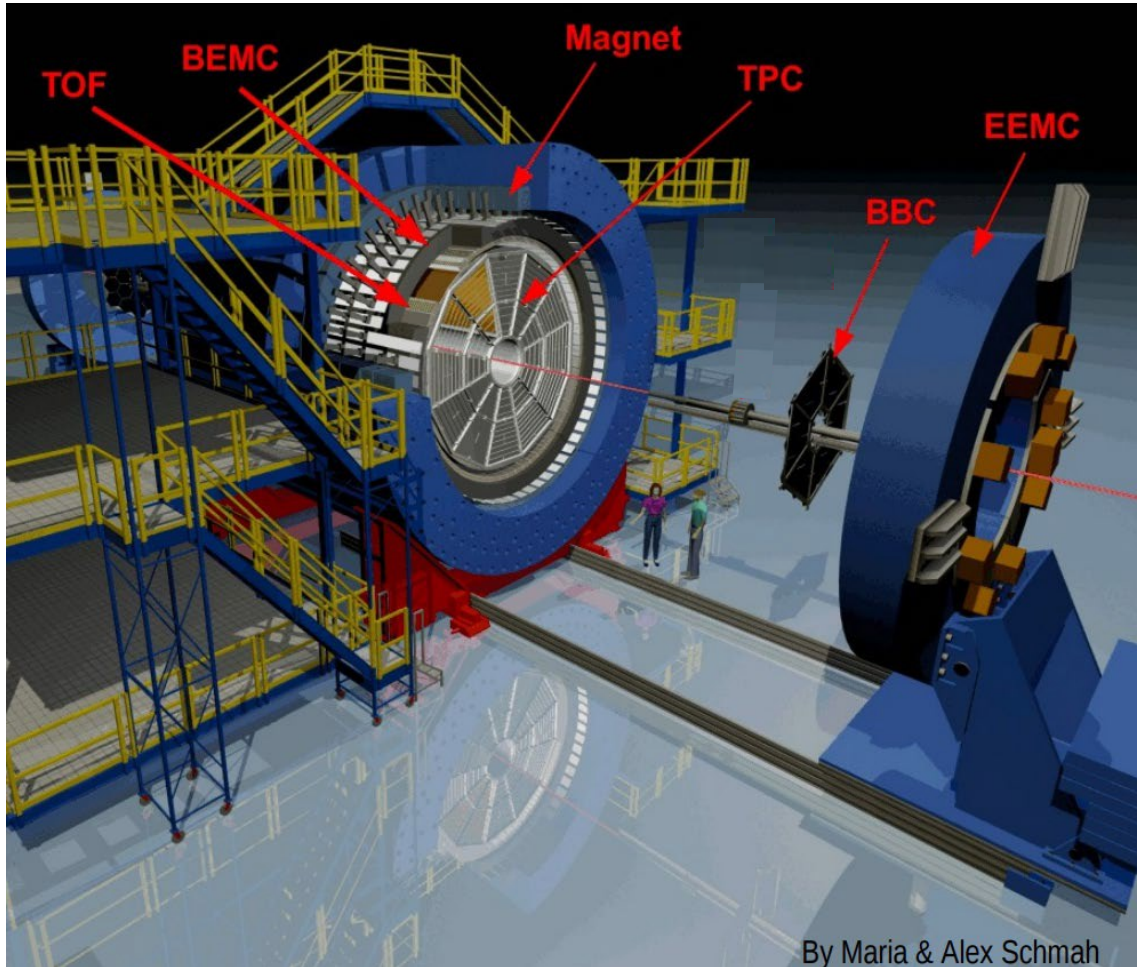
Aims to probe the **soft-hard correlation** within a jet

- How does the amount of soft radiation correlate with the angular and momentum scale of a hard splitting? → how an **early** emission affects a **later** splitting





The STAR detector



Important subdetectors for **200 GeV pp** collisions data-taking during 2012 RHIC run

- **TPC** (Time Projection Chamber)
 - For **charged** particle track reconstruction
 - $|\eta| < 1$, full azimuthal coverage
- **BEMC** (Barrel ElectroMagnetic Calorimeter)
 - For **neutral** energy measurement and triggering
 - $|\eta| < 1$, full azimuthal coverage



Analysis method

- Reconstruct anti- k_T full (charged tracks + neutral energy towers) jets with $R = 0.4$ from 200 GeV pp collisions
- Unfolding methods:
 - Iterative Bayesian unfolding ([D'Agostini. arXiv:1010.0632\(2010\)](#))
 - **MultiFold** ([Andreassen et al. PRL 124, 182001 \(2020\)](#))
 - Machine learning driven
 - Unbinned
 - **Simultaneously** unfolds many observables
→ **Correlation** information is retained!
- **First** application of MultiFold on RHIC data!

Jet observables

- p_T : transverse momentum
- Q^κ : jet charge $Q^\kappa = \frac{1}{(p_{T\text{jet}})^\kappa} \sum_{i \in \text{jet}} q_i \cdot (p_{Ti})^\kappa$
- M : jet mass $M = |\sum_{i \in \text{jet}} p_i| = \sqrt{E^2 - |\vec{p}|^2}$
- M_g : groomed jet mass
- R_g : groomed jet radius
- z_g : shared momentum fraction

$$z_g = \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{\text{cut}} (R_g / R_{\text{jet}})^\beta$$

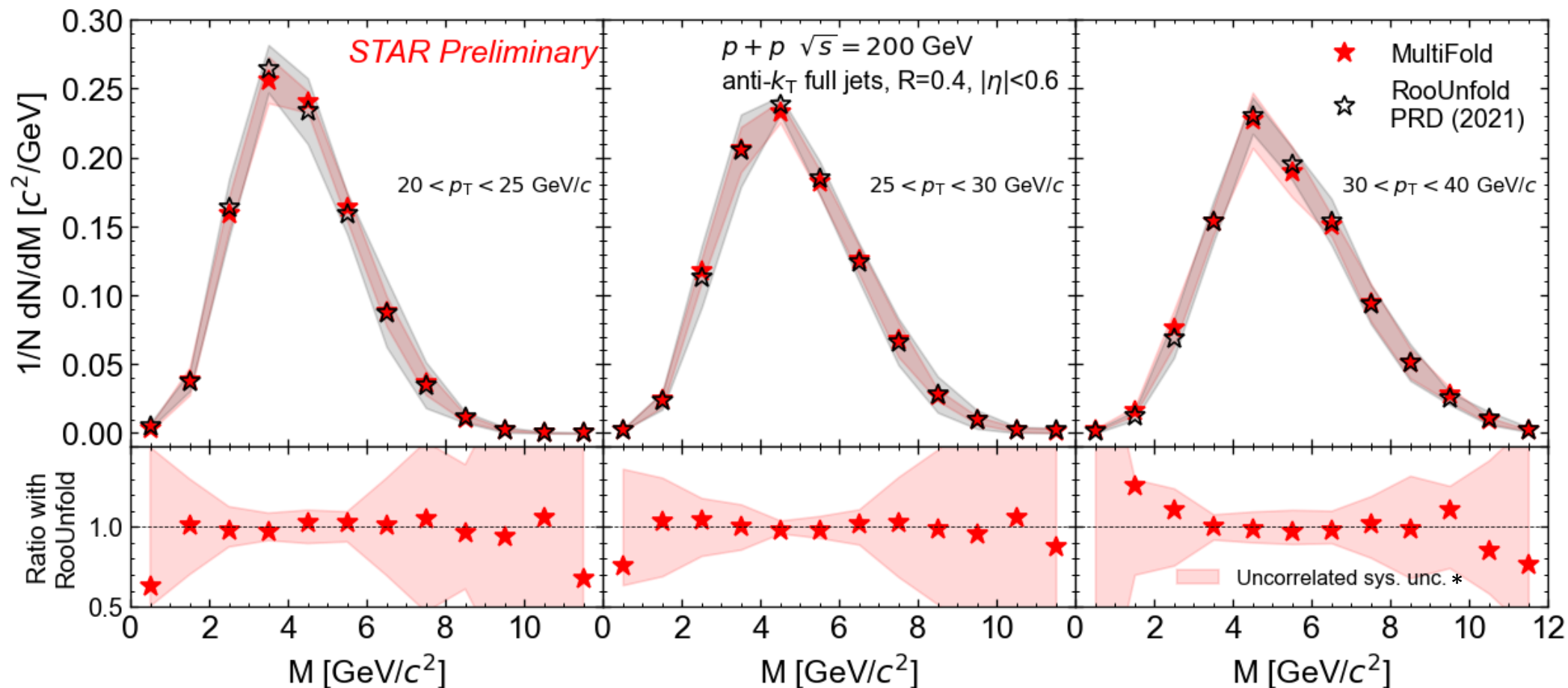
All 6 observables are simultaneously unfolded in an unbinned way!

- Uncertainties due to prior choice accounted for through 6D reweighting based on PYTHIA8 or HERWIG (see backup)

Analysis method: Validation

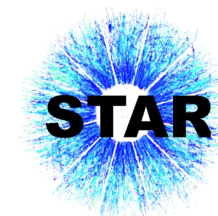
$$M = \left| \sum_{i \in \text{jet}} p_i \right| = \sqrt{E^2 - |\vec{p}|^2}$$

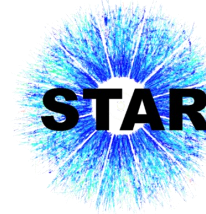
MultiFolded result agrees with **RooUnfolded** result ([STAR Collaboration. PRD 104, 052007\(2021\)](#)) [HEPData](#)



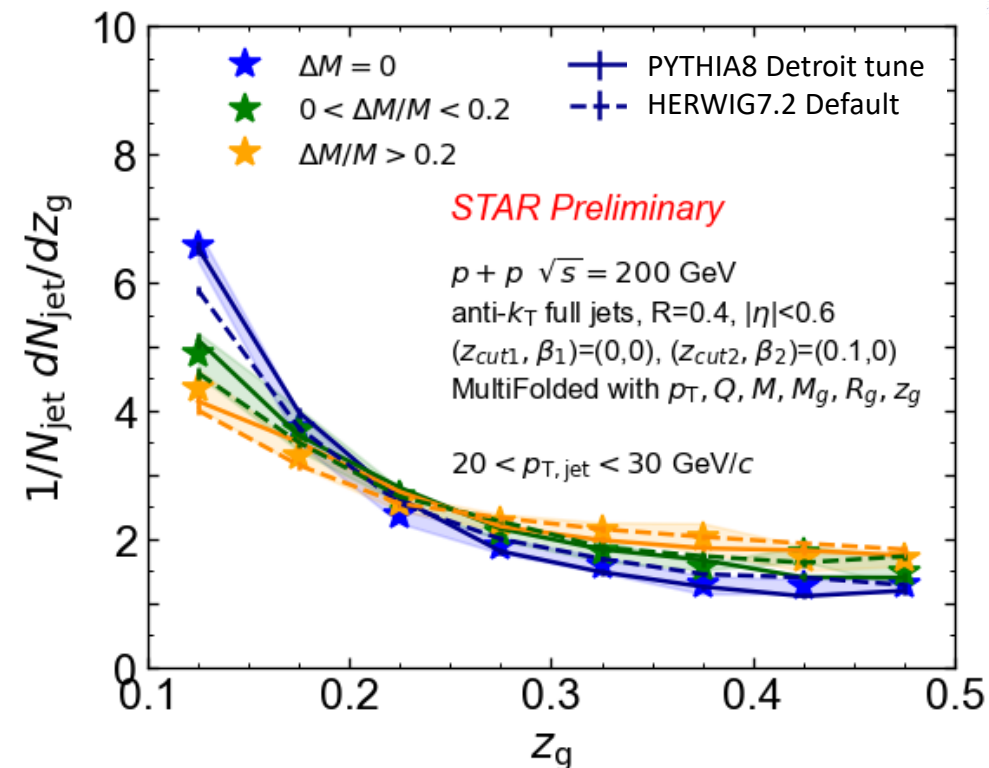
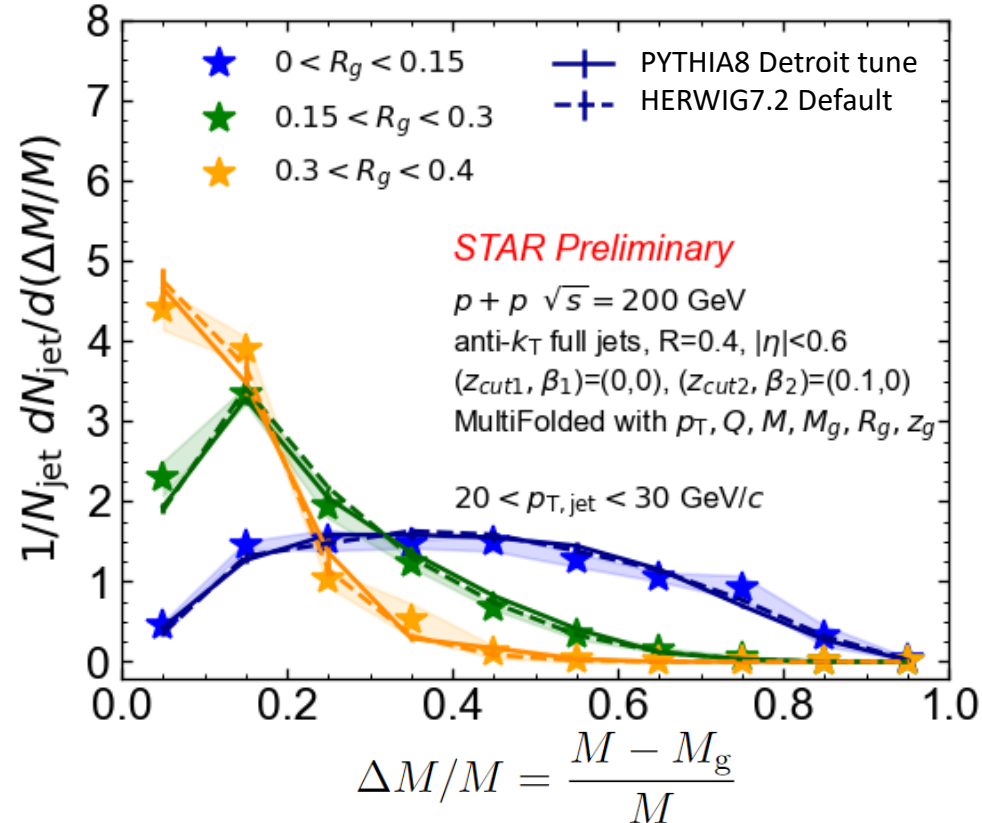
... but **MultiFold** also gives us high-dimensional correlation between observables!

* 2D reweighting used for prior variation, to be consistent with RooUnfolded measurement

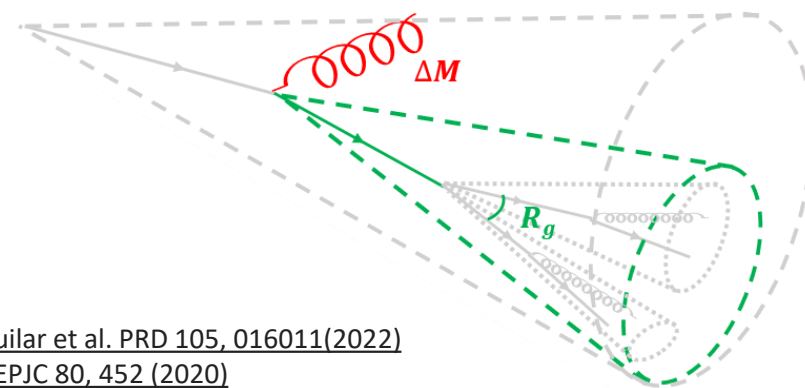




Results



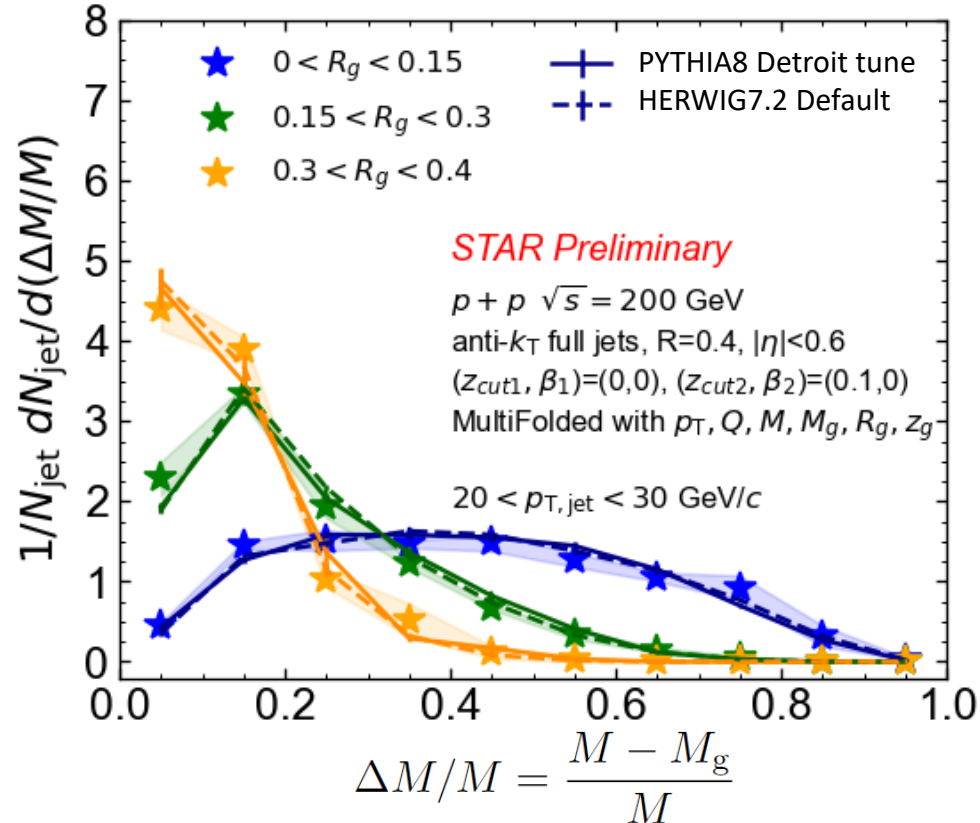
- $\langle \Delta M/M \rangle$ anti-correlated with $\langle R_g \rangle \rightarrow$ consistent with angular ordered parton shower
- **Early** soft wide-angle radiation constrains the angular and momentum phase space of **later** splittings
- MC models describe the trend of data
- See [arXiv:2307.07718](https://arxiv.org/abs/2307.07718) for more details and more results!



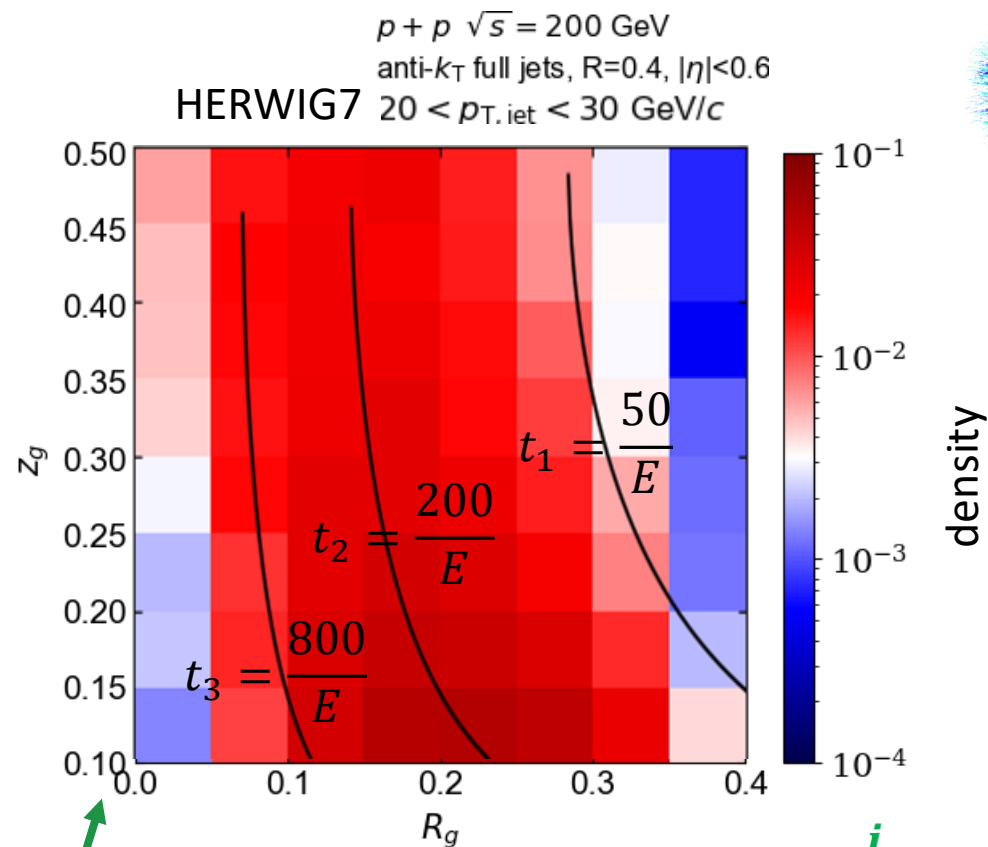
PYTHIA8 Detroit tune: [Aguilar et al. PRD 105, 016011\(2022\)](#)

HERWIG7.2: [Bellm, et al. EPJC 80, 452 \(2020\)](#)

Discussion



- $\langle \Delta M/M \rangle$ anti-correlated with $\langle R_g \rangle \rightarrow$ consistent with angular ordered parton shower
- **Early** soft wide-angle radiation constrains the angular and momentum phase space of **later** splittings
- MC models describe the trend of data
- See [arXiv:2307.07718](https://arxiv.org/abs/2307.07718) for more details and more results!



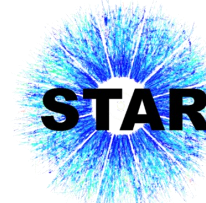
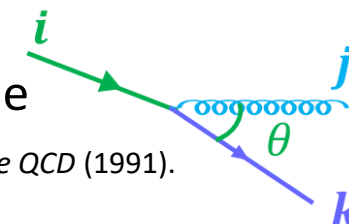
- Lines of constant formation time

Y. L. Dokshitzer, et al. *Basics of Perturbative QCD* (1991).

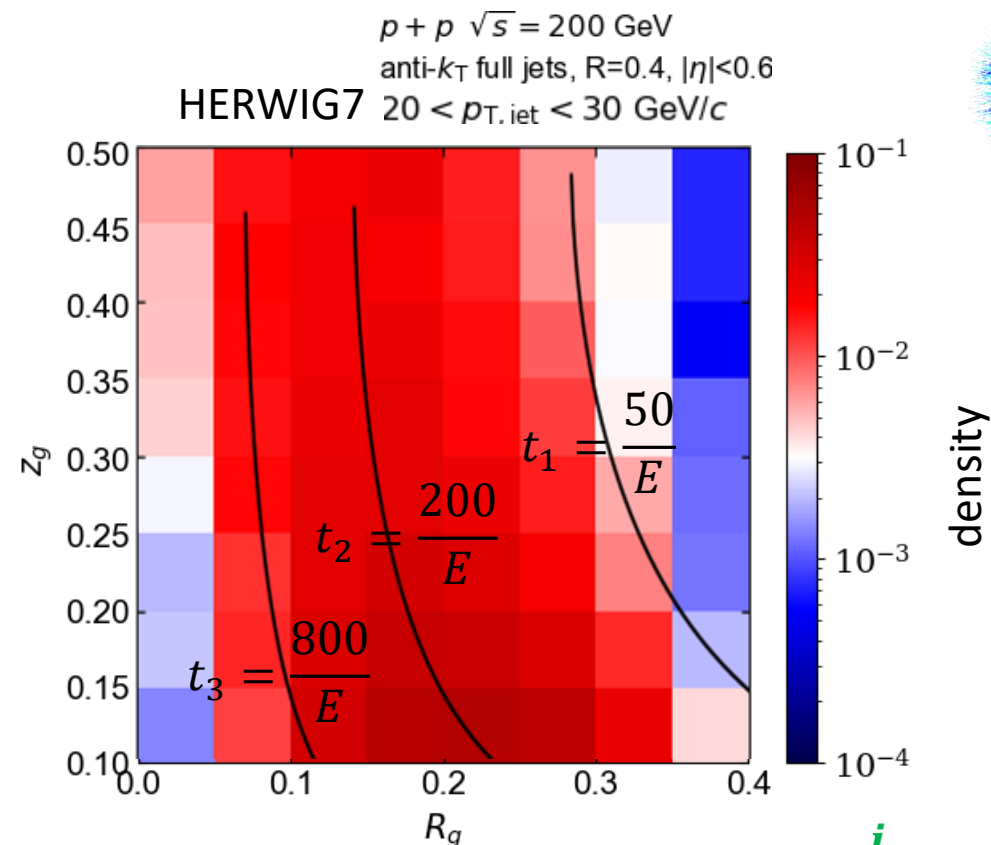
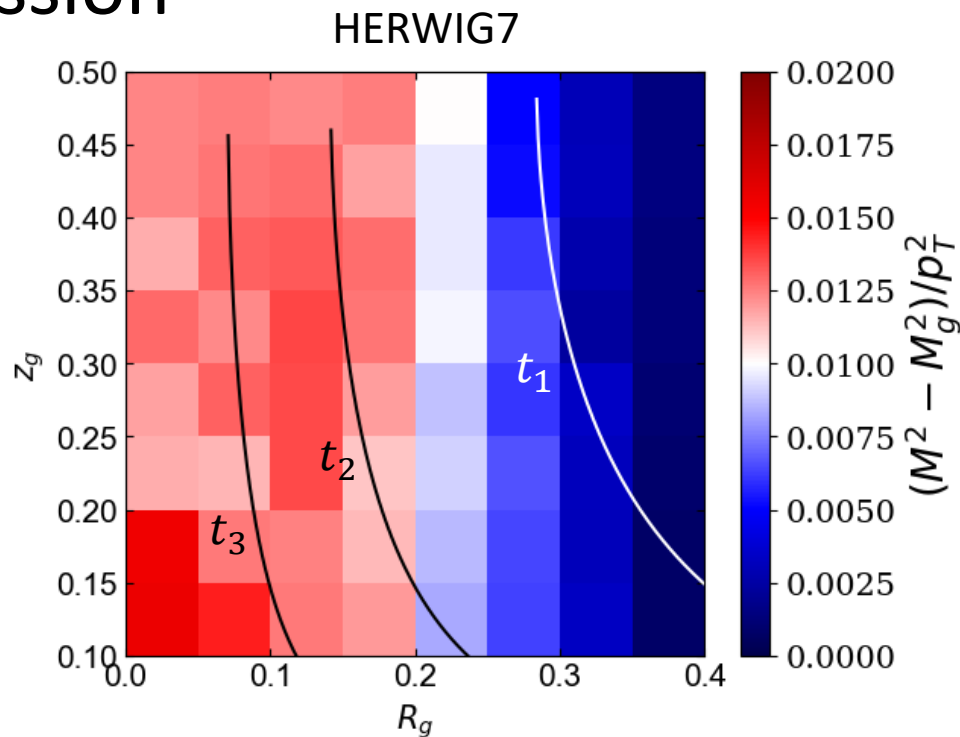
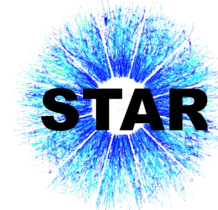
Using $t_F \sim \frac{1}{2Ez(1-z)(1-\cos(\theta))}$
 solve for $z(\theta, z < 0.5)$:

$$z = \frac{1}{2} \left[1 - \sqrt{1 - \frac{2}{tE(1-\cos(\theta))}} \right]$$

$$z = \frac{\min(p_{T,j}, p_{T,k})}{p_{T,j} + p_{T,k}}$$



Discussion



- CollinearDrop groomed mass strongly correlated with R_g and weakly with z_g
- To shed a lot of mass at **early** stage, the hard splitting needs to happen **late**
- PYTHIA largely agrees with HERWIG (see backup).

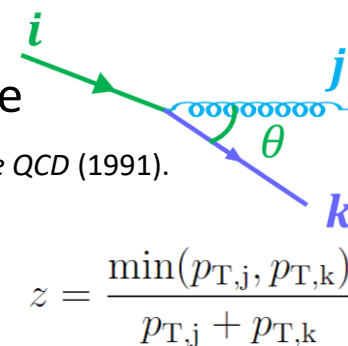
Stay tuned for data!

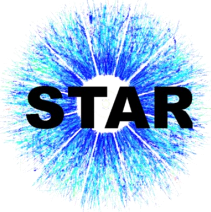
- Lines of constant formation time

Y. L. Dokshitzer, et al. *Basics of Perturbative QCD* (1991).

Using $t_F \sim \frac{1}{2Ez(1-z)(1-\cos(\theta))}$
 solve for $z(\theta, z < 0.5)$:

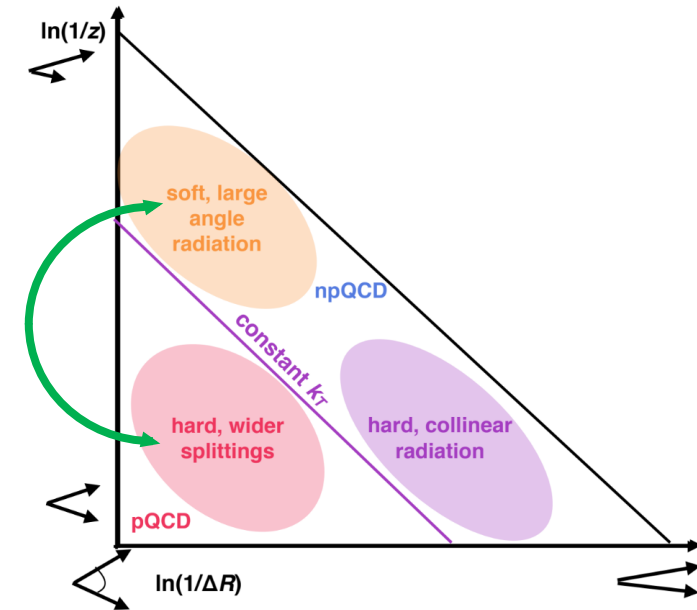
$$z = \frac{1}{2} \left[1 - \sqrt{1 - \frac{2}{tE(1-\cos(\theta))}} \right]$$





Conclusion (Part 1)

- Study the **soft-hard** correlation within jets with CollinearDrop vs SoftDrop jet observables
 - The **early-stage radiation** is correlated with the **later-stage splittings**
 - MultiFold allows for access of multi-dimensional correlations on a jet-by-jet basis. First application on RHIC data!
 - For example, we can select on a specific region of phase space by cutting on **3** different observables, and then study the **4th**



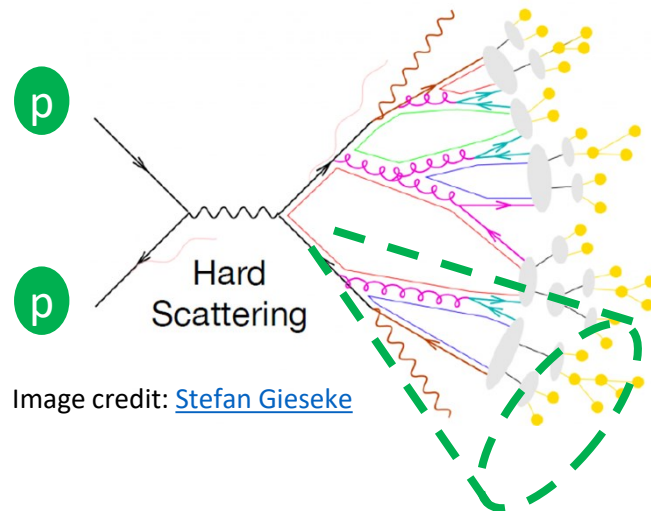


Image credit: [Stefan Gieseke](#)

Study the non-perturbative contribution in jets

- Soft component of the parton shower
→ CollinearDrop jet measurement
- Hadronization
→ Charge correlator ratio measurement



Motivation

The **charge correlator ratio** r_c can probe for evidence of **string-like fragmentation**, by distinguishing the charge signs of **leading and subleading** particles within jets.

[Chien et al. PRD 105 051502 \(2022\)](#)

$$r_c(X) = \frac{d\sigma_{h_1 h_2}/dX - d\sigma_{h_1 \bar{h}_2}/dX}{d\sigma_{h_1 h_2}/dX + d\sigma_{h_1 \bar{h}_2}/dX}$$

$h_1 h_2$: same charge hadrons,
 $h_1 \bar{h}_2$: opposite charge hadrons

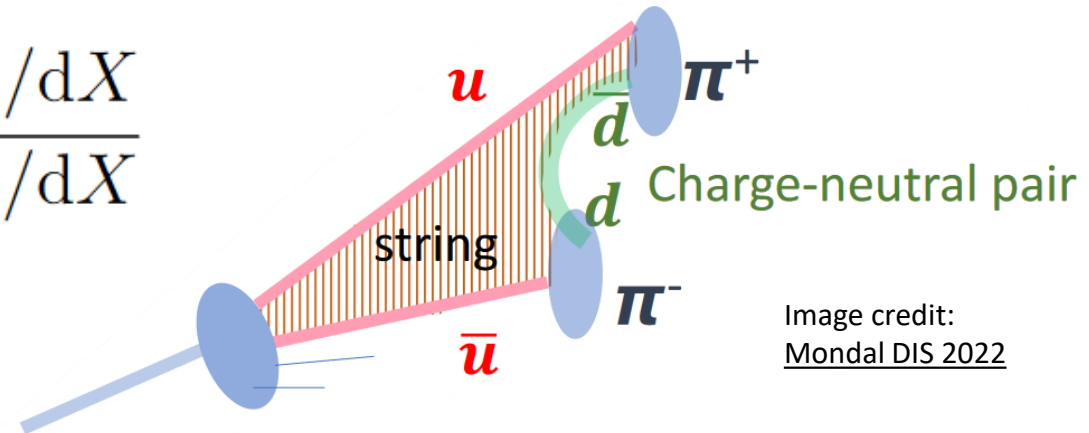


Image credit:
[Mondal DIS 2022](#)

This configuration above favors $r_c \rightarrow -1$.

This measurement can also establish a baseline for studying **medium modification of hadronization in the QGP!** The choice of leading dihadrons might make it less susceptible to the background.

→ In the future, interesting to pursue this in heavy ions!



Analysis method

- Reconstruct anti- k_T **full jets** with $R = 0.4$ from 200 GeV pp collisions
- At detector level, the decay of neutral hadrons affects the ordering of particles in jet, so we consider a **charged** r_c measurement. See backup slides for an example.

$$r_c(X) = \frac{d\sigma_{h_1 h_2}/dX - d\sigma_{h_1 \bar{h}_2}/dX}{d\sigma_{h_1 h_2}/dX + d\sigma_{h_1 \bar{h}_2}/dX}$$

$h_1 h_2$: same charge **tracks**,

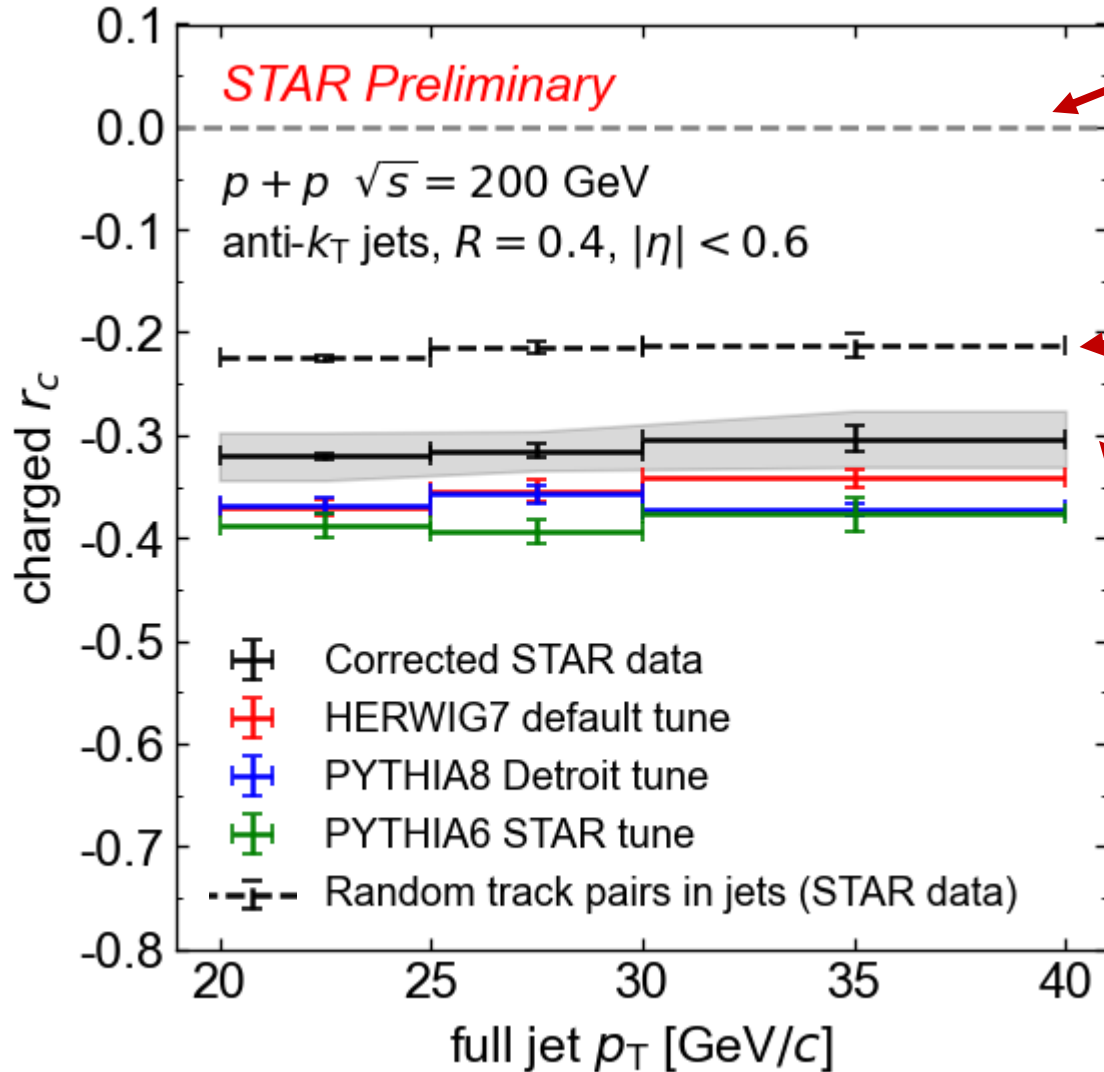
$h_1 \bar{h}_2$: opposite charge **tracks**

→ Slightly changed the definition as proposed by theorists, but comparison with MC models is still meaningful!



Result

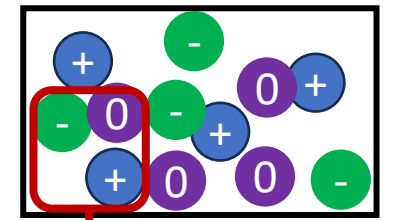
$$r_c(X) = \frac{d\sigma_{h_1 h_2}/dX - d\sigma_{h_1 \bar{h}_2}/dX}{d\sigma_{h_1 h_2}/dX + d\sigma_{h_1 \bar{h}_2}/dX}$$



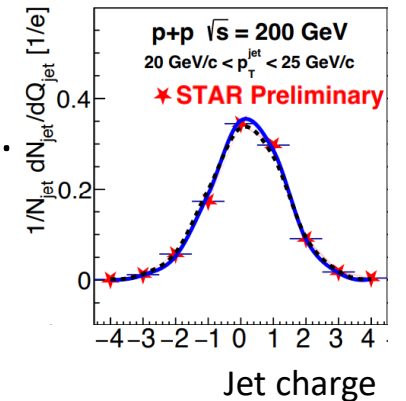
Random pairs in an uncorrelated (infinite) bath with no net charge: $r_c = 0$

Data show a preference of opposite sign pairs over same sign pairs, in:

- pair of random tracks within a jet; influenced by jet charge ~ 0 on average: $r_c \approx -0.2$
- leading and subleading tracks in jet. additional correlation from fragmentation: $r_c \approx -0.3$



clustered into a jet

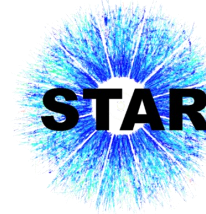


PYTHIA6 Perugia + STAR tune: Skands, PRD 82, 074018 (2010)

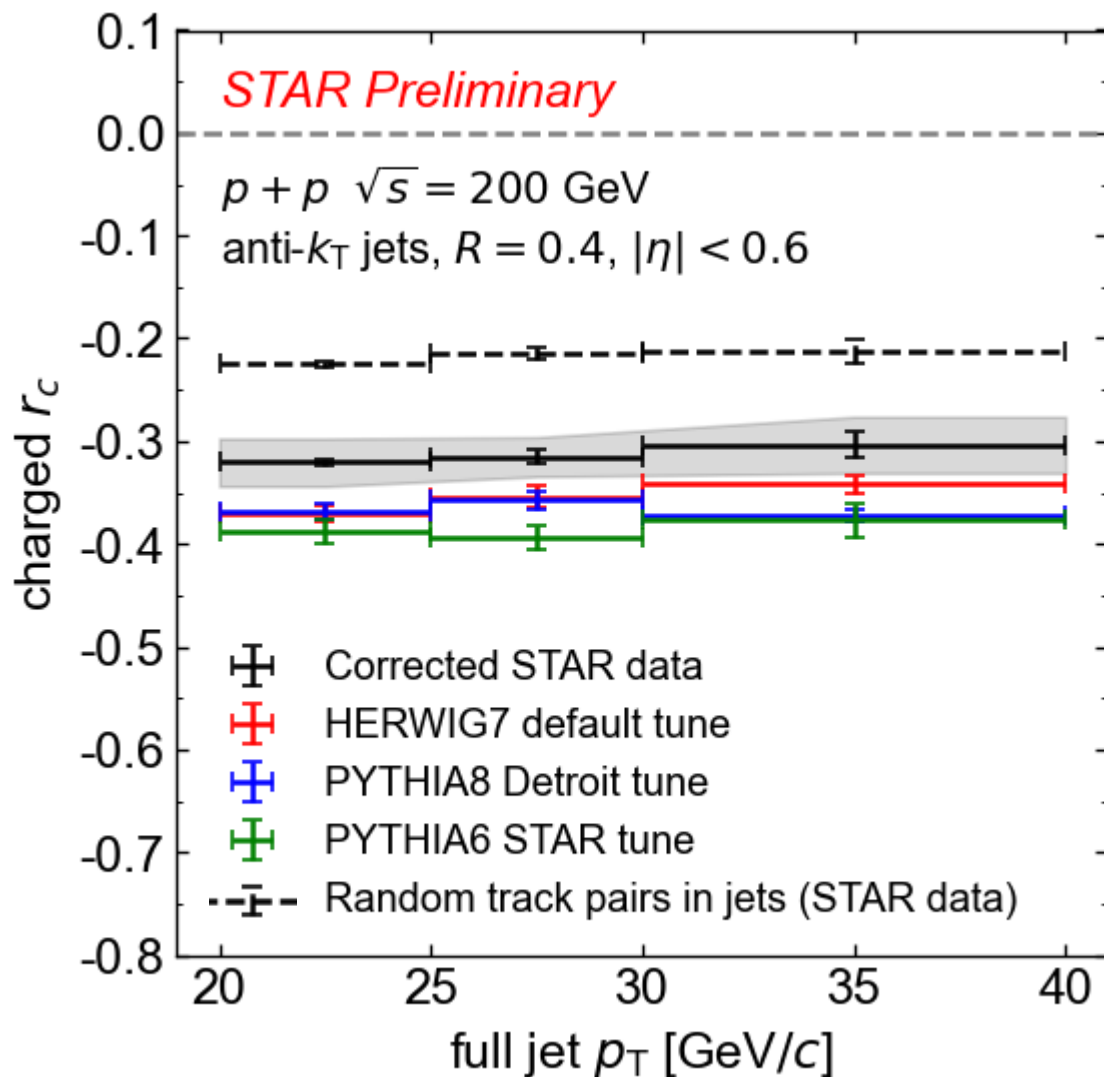
J. K. Adkins, PhD thesis (Kentucky U., 2015)

PYTHIA8 Detroit tune: Aguilar et al. PRD 105, 016011(2022)

HERWIG7.2: Bellm, et al. EPJC 80, 452 (2020)



Result



Weak dependence on jet p_T in 20-40 GeV

Models based on Lund string fragmentation and cluster hadronization both underpredict r_c in data.

PYTHIA6 Perugia + STAR tune: [Skands. PRD 82, 074018 \(2010\)](#)

[J. K. Adkins, PhD thesis \(Kentucky U., 2015\)](#)

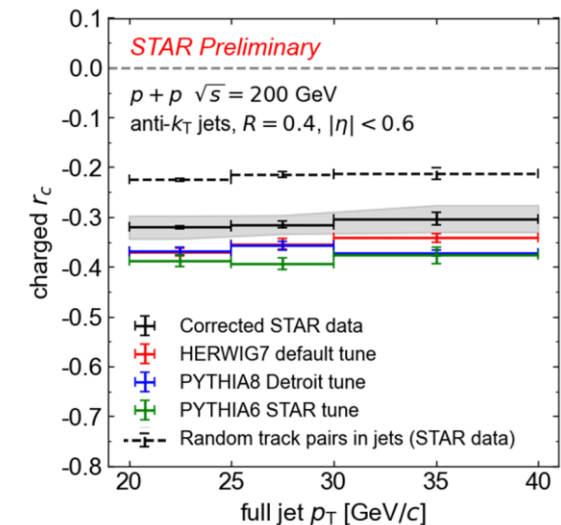
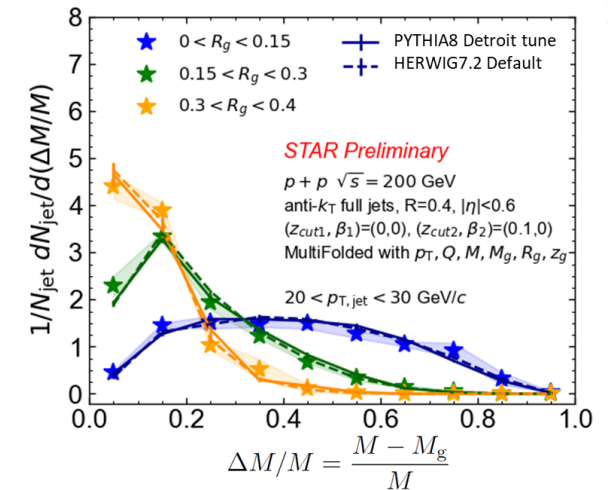
PYTHIA8 Detroit tune: [Aguilar et al. PRD 105, 016011\(2022\)](#)

HERWIG7.2: [Bellm, et al. EPJC 80, 452 \(2020\)](#)



Conclusions

- Study the **soft-hard** correlation within jets with CollinearDrop vs SoftDrop jet observables
 - The **early-stage radiation** is correlated with the **later-stage splittings**
 - MultiFold allows for access of multi-dimensional correlations on a jet-by-jet basis. First application on RHIC data!
 - For example, we can select on a specific region of phase space by cutting on **3** different observables, and then study the **4th**
- Study **hadronization** with the charge correlator ratio
 - Data show a **weaker correlation** between leading and subleading particles in jet than models
 - In the future, study r_c as functions of observables sensitive to pQCD \rightarrow npQCD transition! ($\mathbf{k}_T, \mathbf{t}_f, \mathbf{z} \dots$)

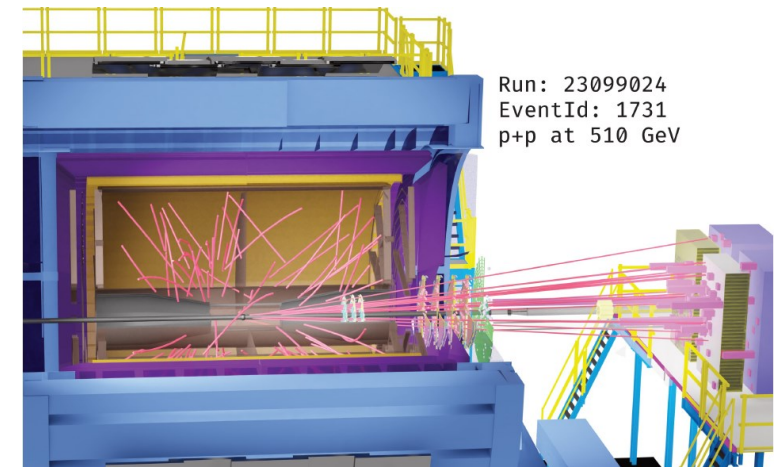
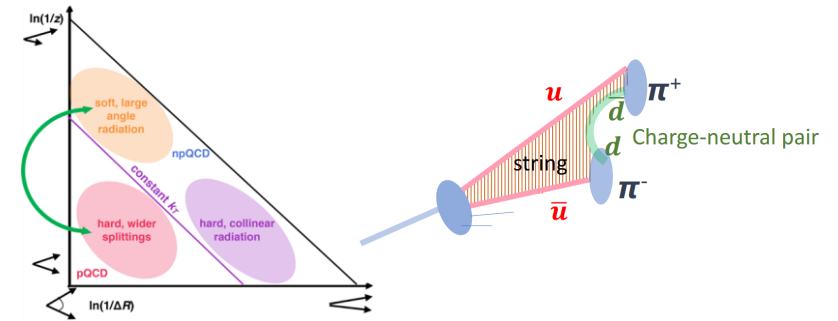




Jet substructure measurements at RHIC → Unique sensitivity to **non-perturbative QCD** effects. See [talk](#) by Isaac Mooney for **more** measurements from STAR!

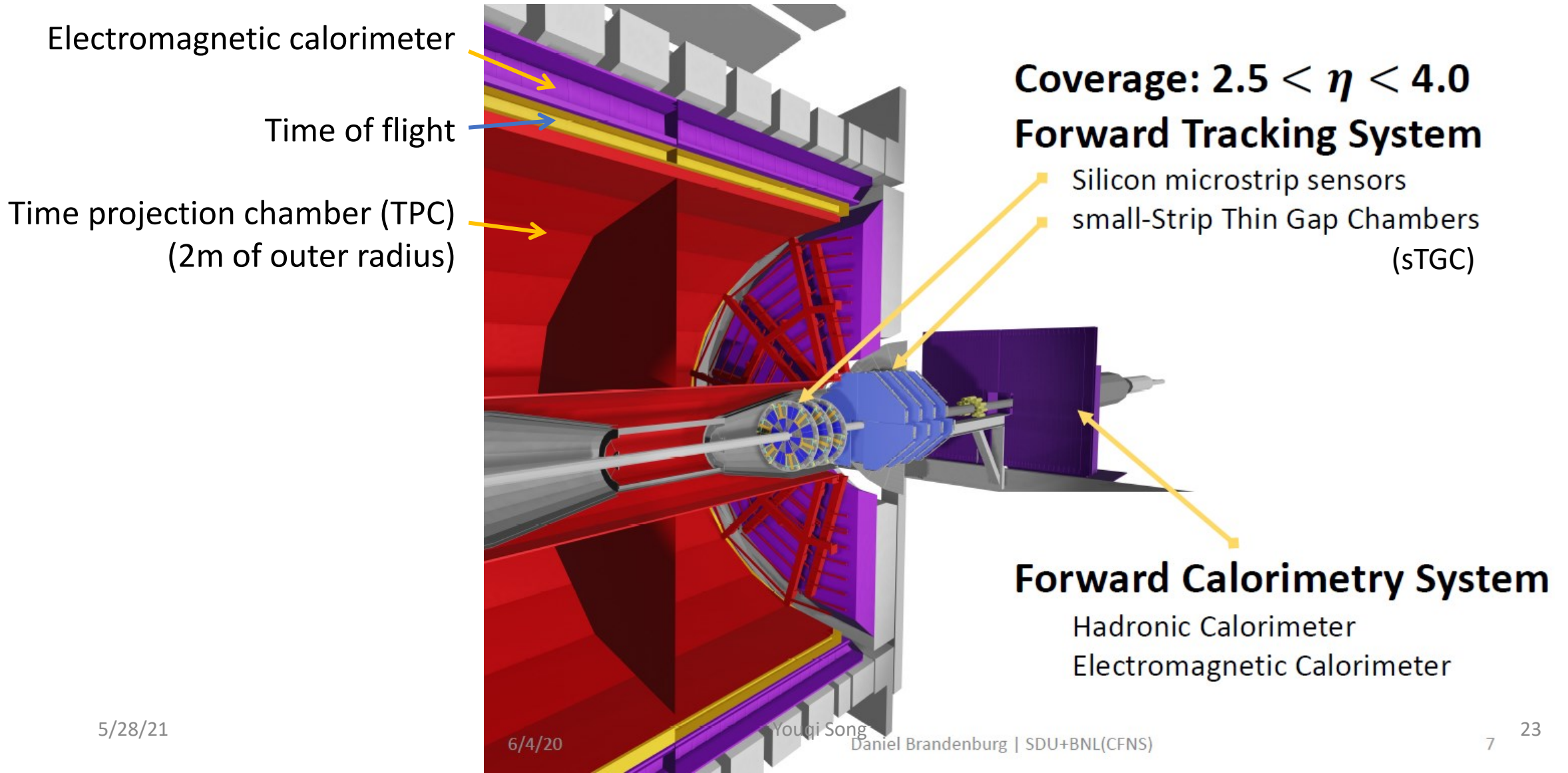
Outlook

STAR in the upcoming run → Unique capability to measure jets in **both midrapidity and forward region** at RHIC energies with the forward upgrade, since 2022. Stay tuned for more data!



Backup

STAR forward upgrade: forward detector



Jet substructures

- Use “ z_g ” to denote observables obtained through **SoftDrop grooming** which removes soft and wide-angle radiation from the jet

Procedure:

- Undo jet clustering
- From the last step of the clustering, check if subjet momentum fraction passes:

$$z_g = \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{\text{cut}} (R_g / R_{\text{jet}})^\beta \quad z_{\text{cut}} = 0.1$$

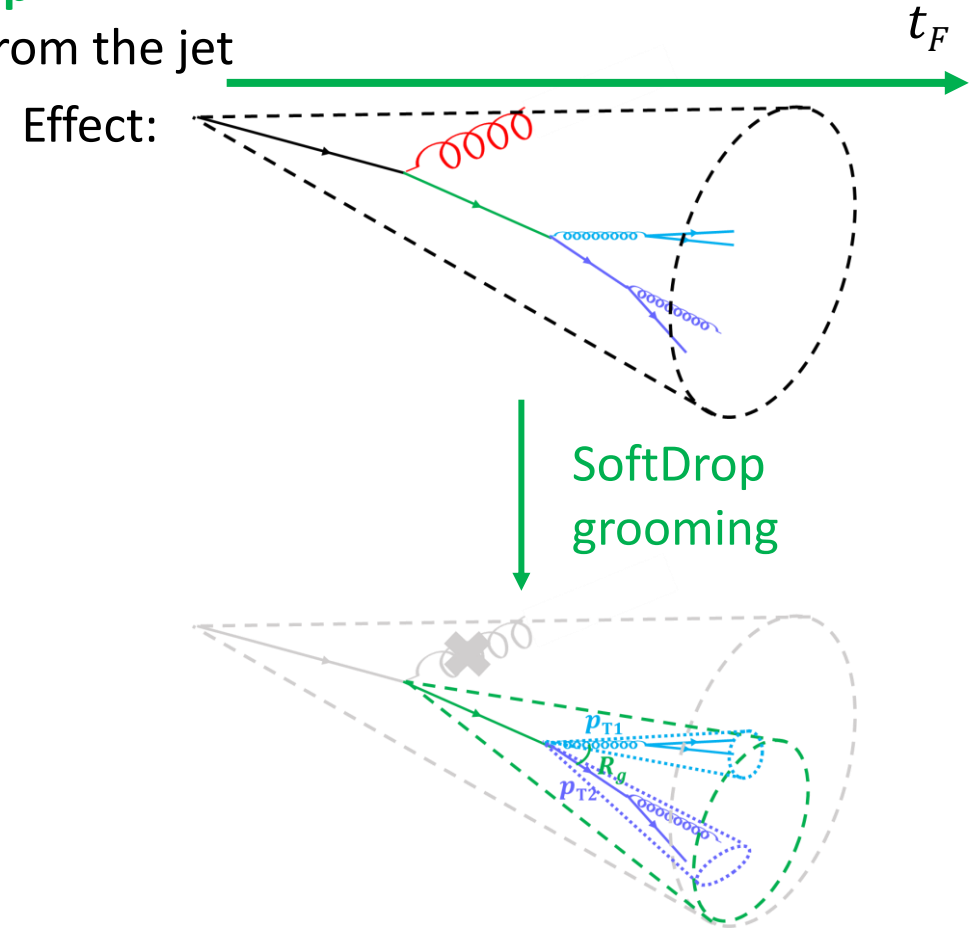
$$\beta = 0$$

$$\rightarrow \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > 0.1$$

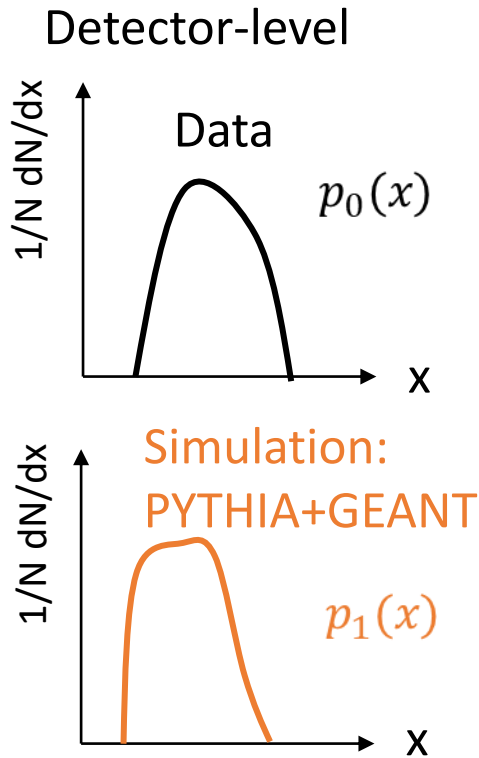
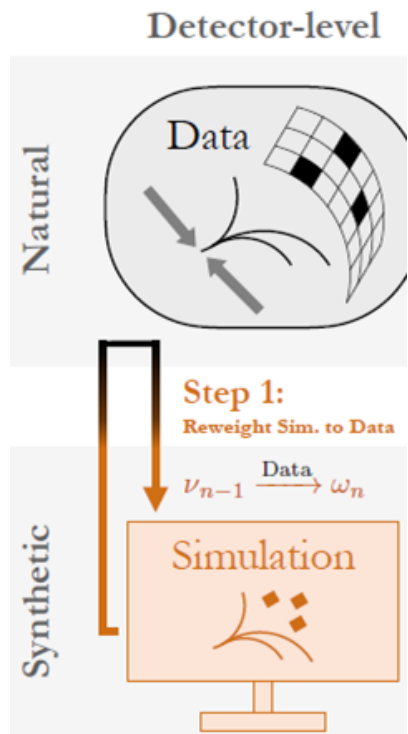
- If so, grooming is done; if not, remove the softer subjet and follow through the second to last step

[Larkoski, et al. JHEP 2014, 146 \(2014\).](#)

[Dasgupta et al. JHEP 2013, 29 \(2013\).](#)



Method: machine learning



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 \rightarrow Reweighting histograms
 \rightarrow Classification \rightarrow Neural network

Where does the machine learning part come in?



Method: machine learning

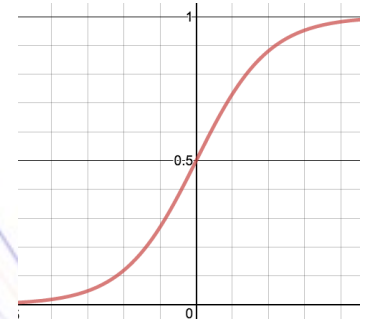
- Architecture: Dense neural network
- Activation function for dense layers: Rectified linear unit
- Activation function for output layer: Sigmoid
- Loss function: Binary cross entropy

$$\text{loss}(f(x)) = - \sum_{i \in 0} \log f(x_i) - \sum_{i \in 1} \log(1 - f(x_i))$$

- Optimization algorithm: Adam
<https://arxiv.org/pdf/1412.6980.pdf>
- Nodes per dense layer: [100,100,100]
- Output dimension: 2
- Input dimension: 6

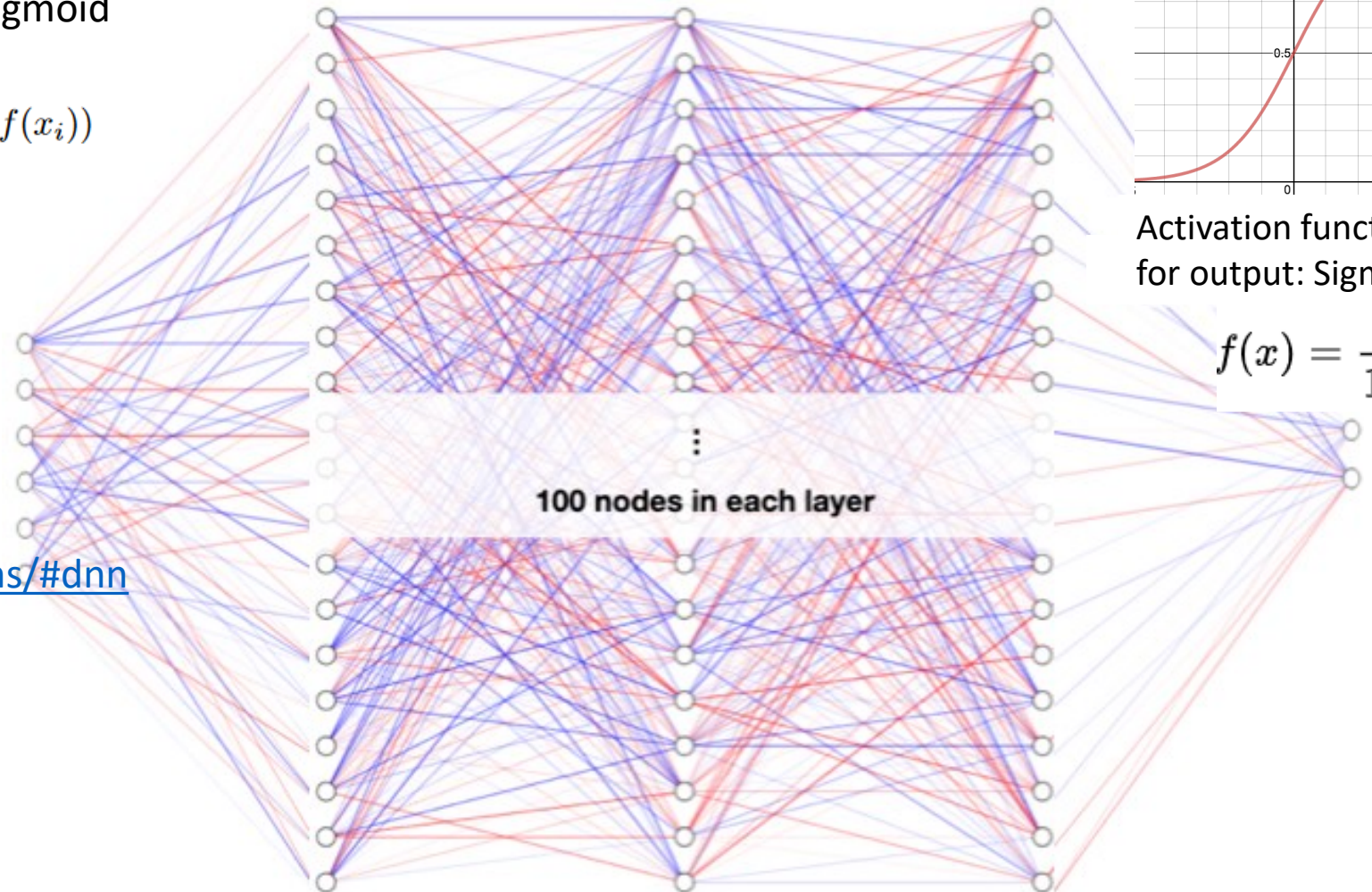
- All hyperparameters are default:
<https://energyflow.network/docs/archs/#dnn>

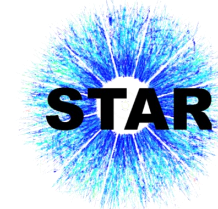
Activation function for dense layers: Rectified linear unit
 $f(x) = x^+ = \max(0, x)$



Activation function for output: Sigmoid

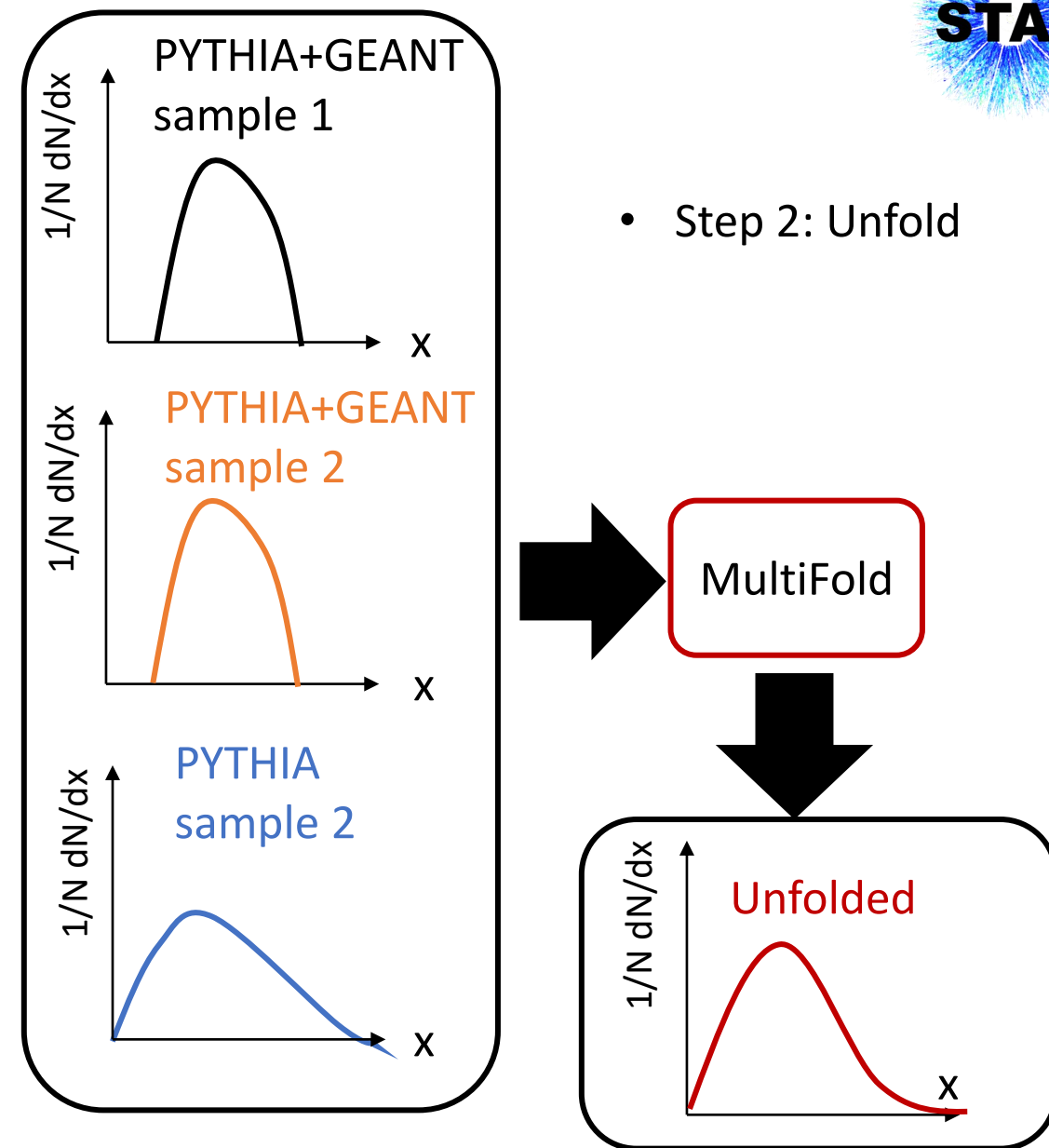
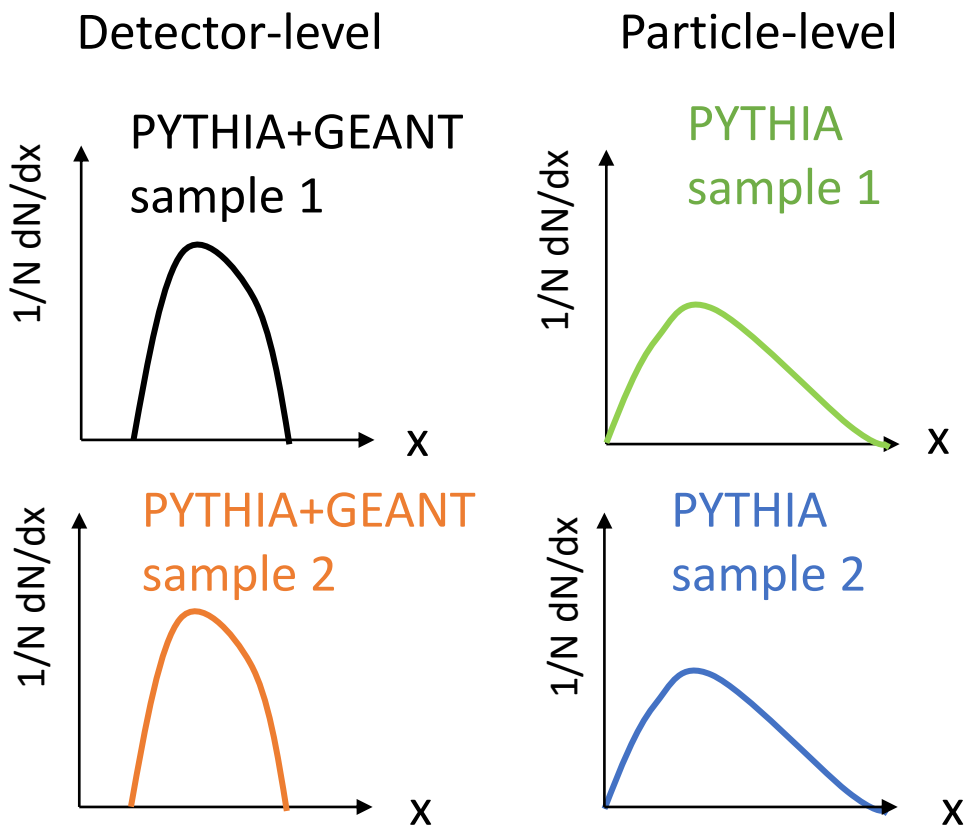
$$f(x) = \frac{1}{1 + e^{-x}}$$



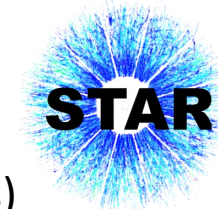


Closure test for unfolding

- Step 1: Separate matched jets from PYTHIA and PYTHIA+GEANT into 2 samples



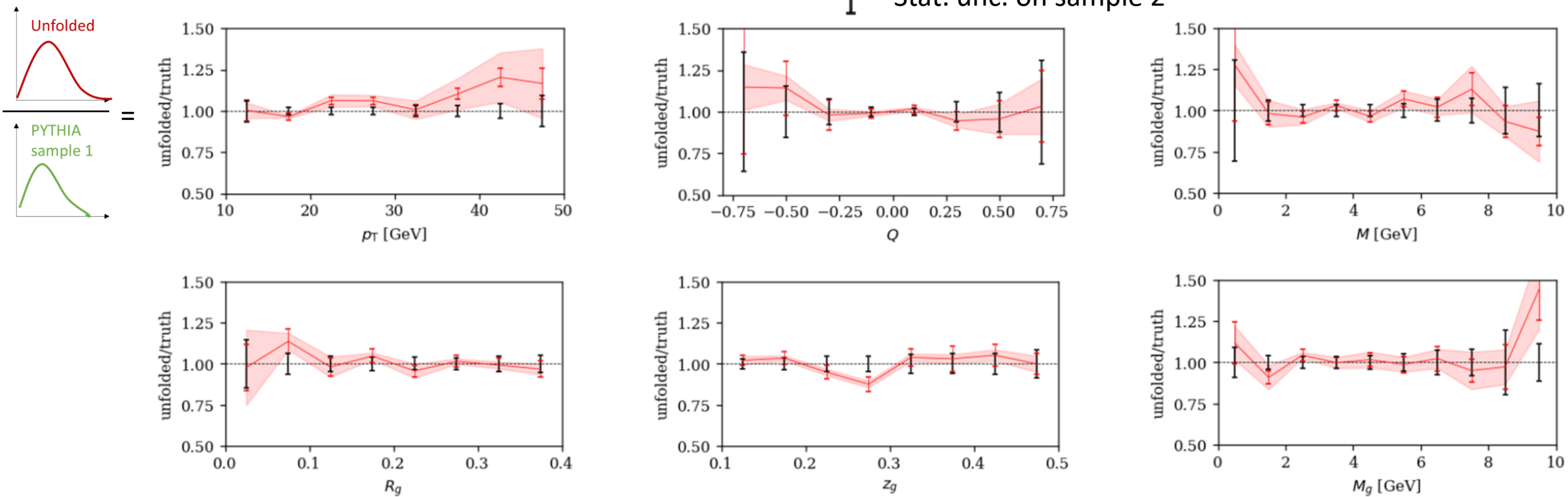
- Step 2: Unfold



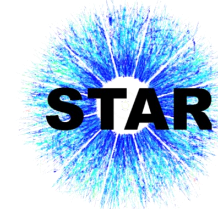
Closure test for unfolding

- Decent **closure** for all substructure observables

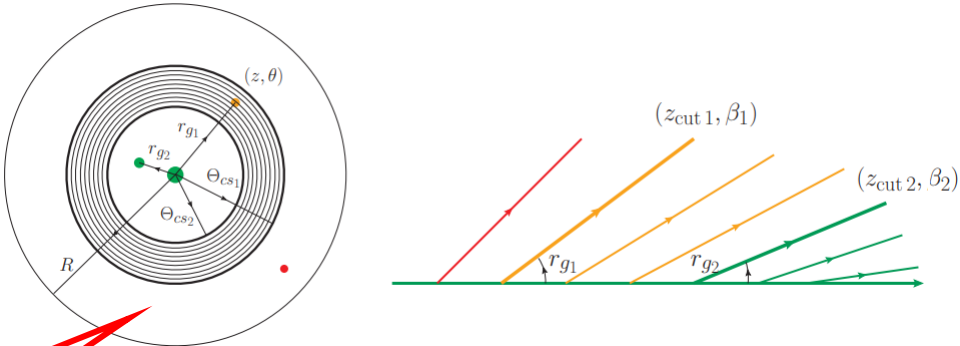
Unfolding unc. on data (not including misses)
Stat. unc. on sample 1
Stat. unc. on sample 2



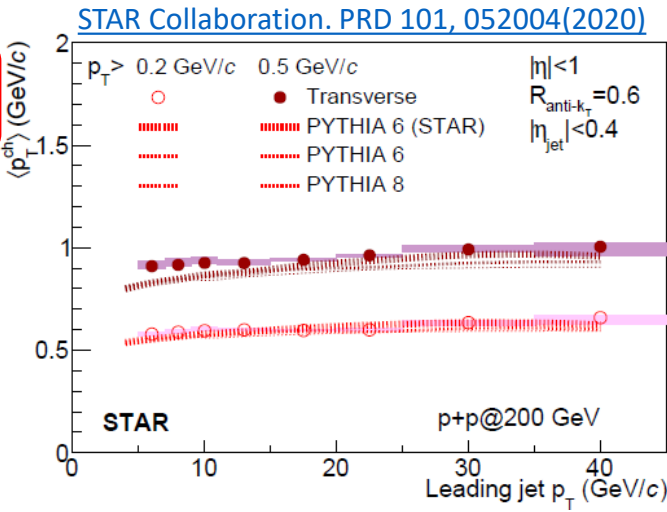
* 2D reweighting used for prior variation



Underlying event

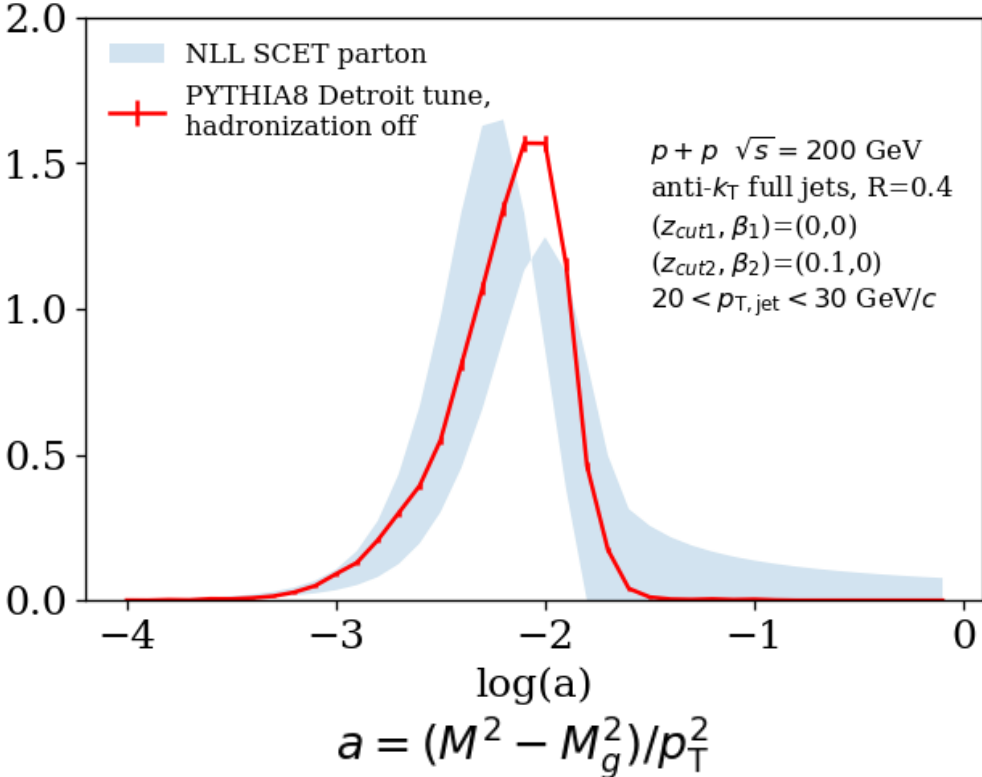


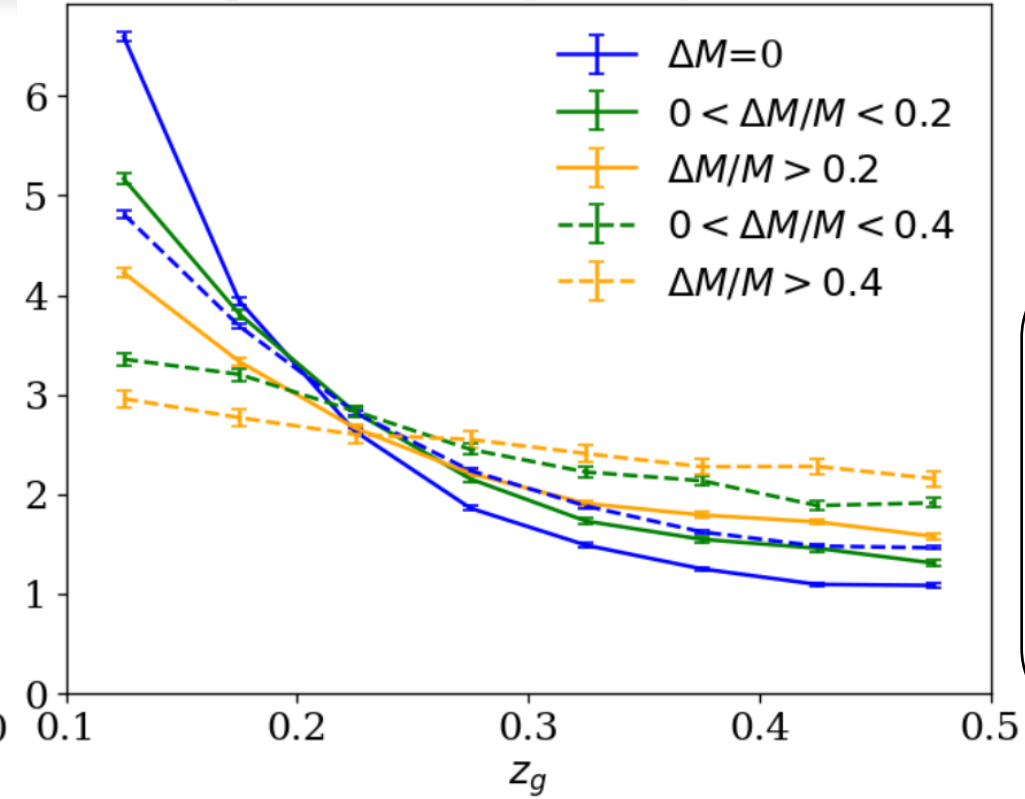
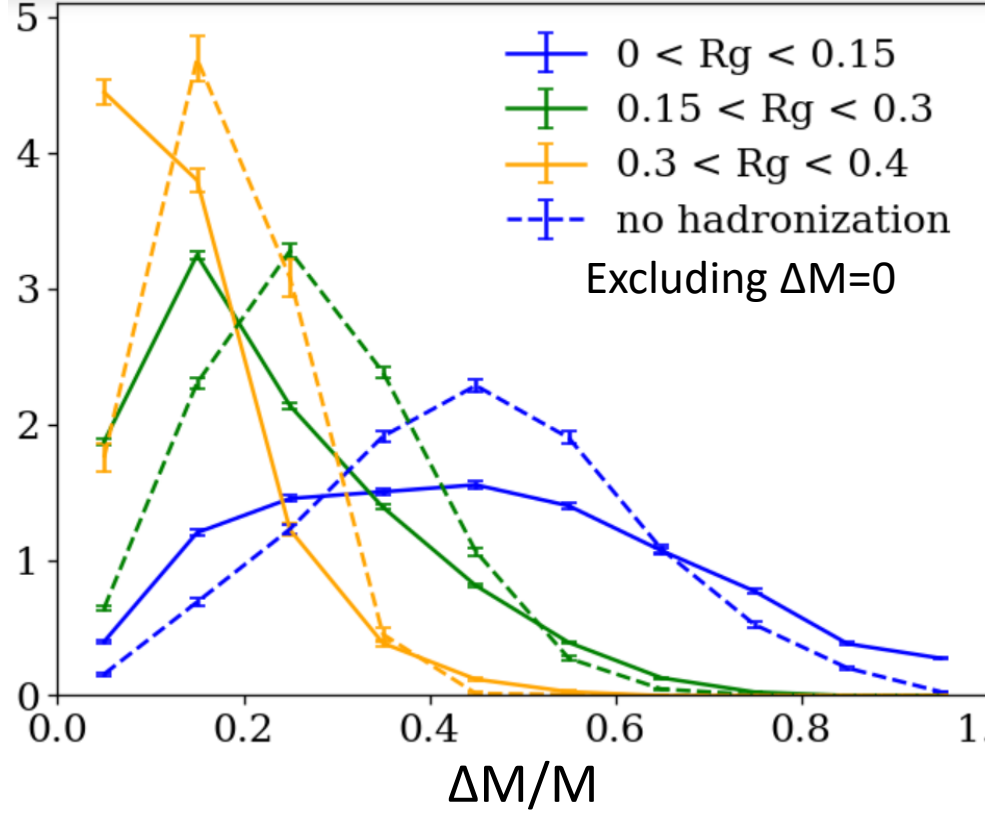
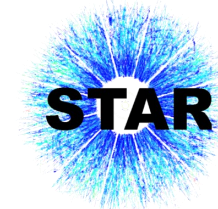
Remove underlying effects and pileup, not necessary for 200 GeV collisions



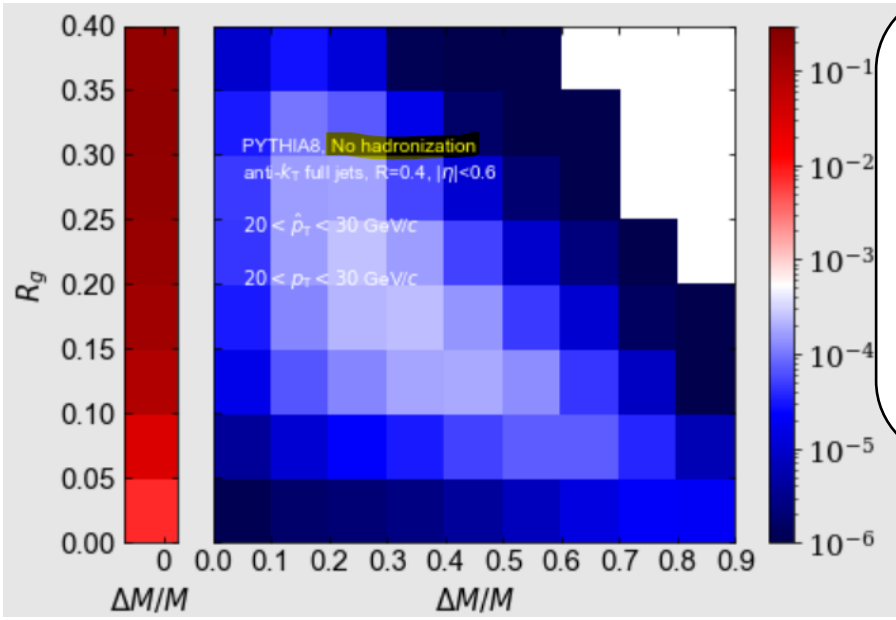
CollinearDrop groomed jet mass

- Theoretical calculation (next-to-leading log precision, using SCET calculational framework, and not including hadronization) agrees with PYTHIA8



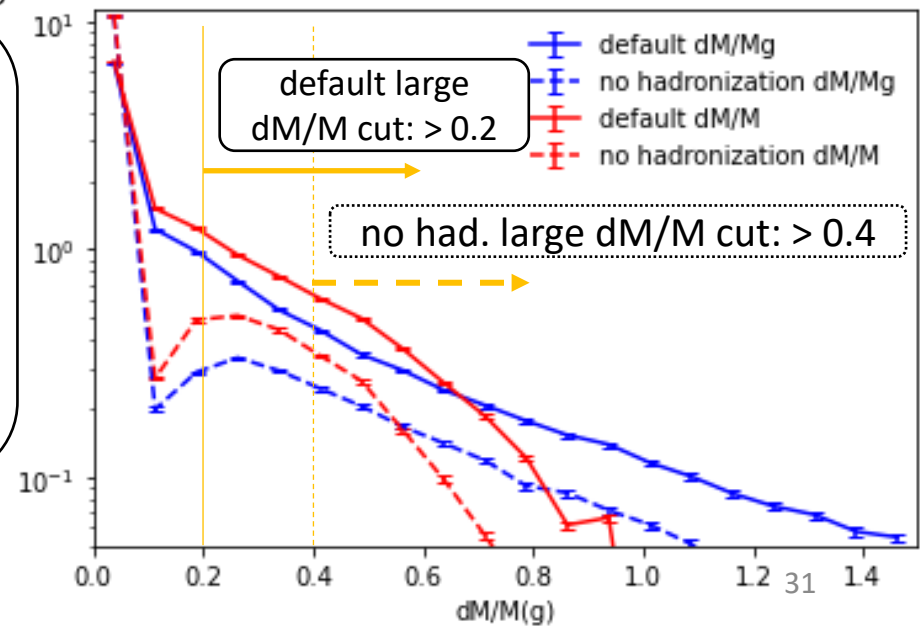


PYTHIA8 Detroit tune
 $20 < p_{T,jet} < 30$ GeV

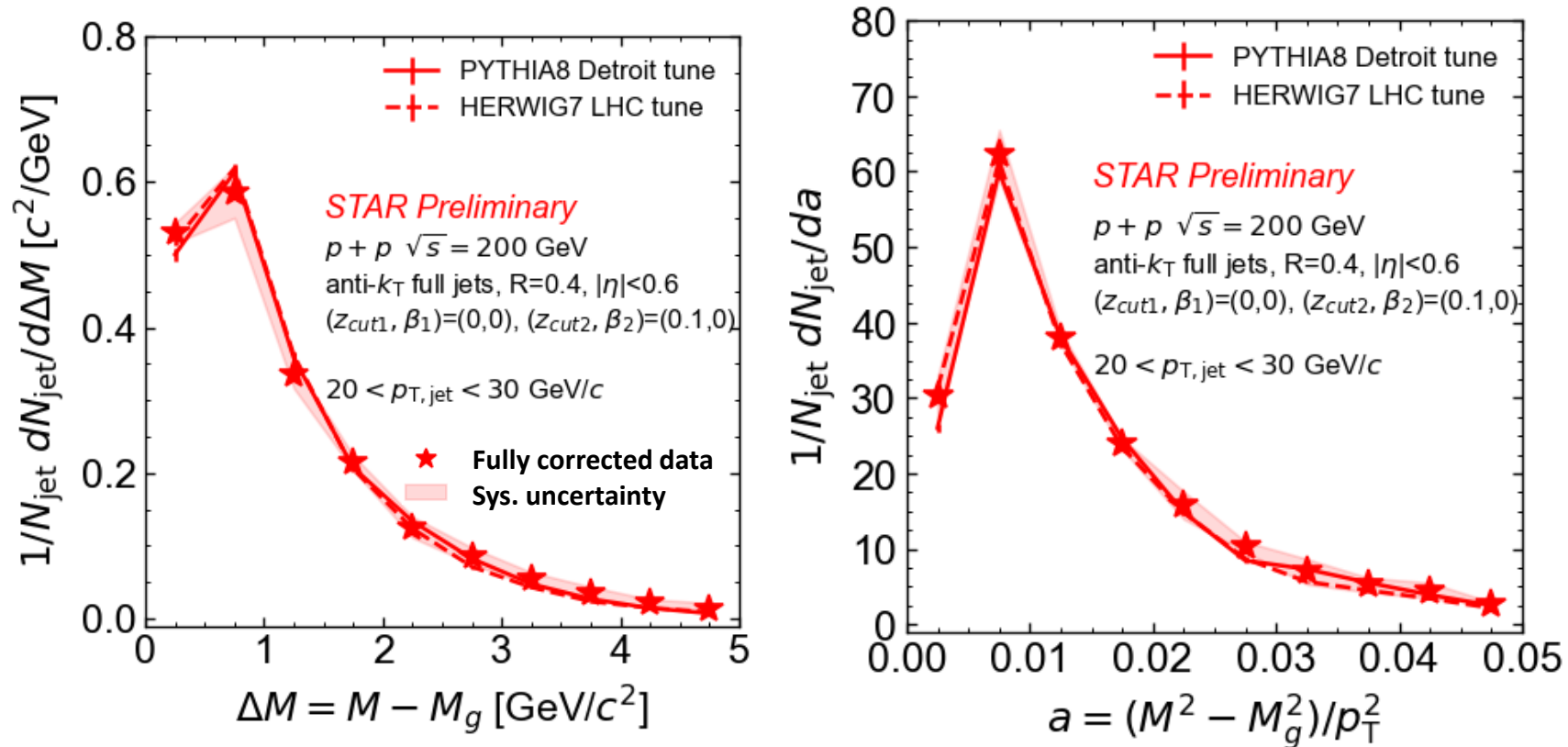


Hadronization effects study:
Hadronization smears/shifts the distributions, but the correlation with and without hadronization is the same.

Youqi Song



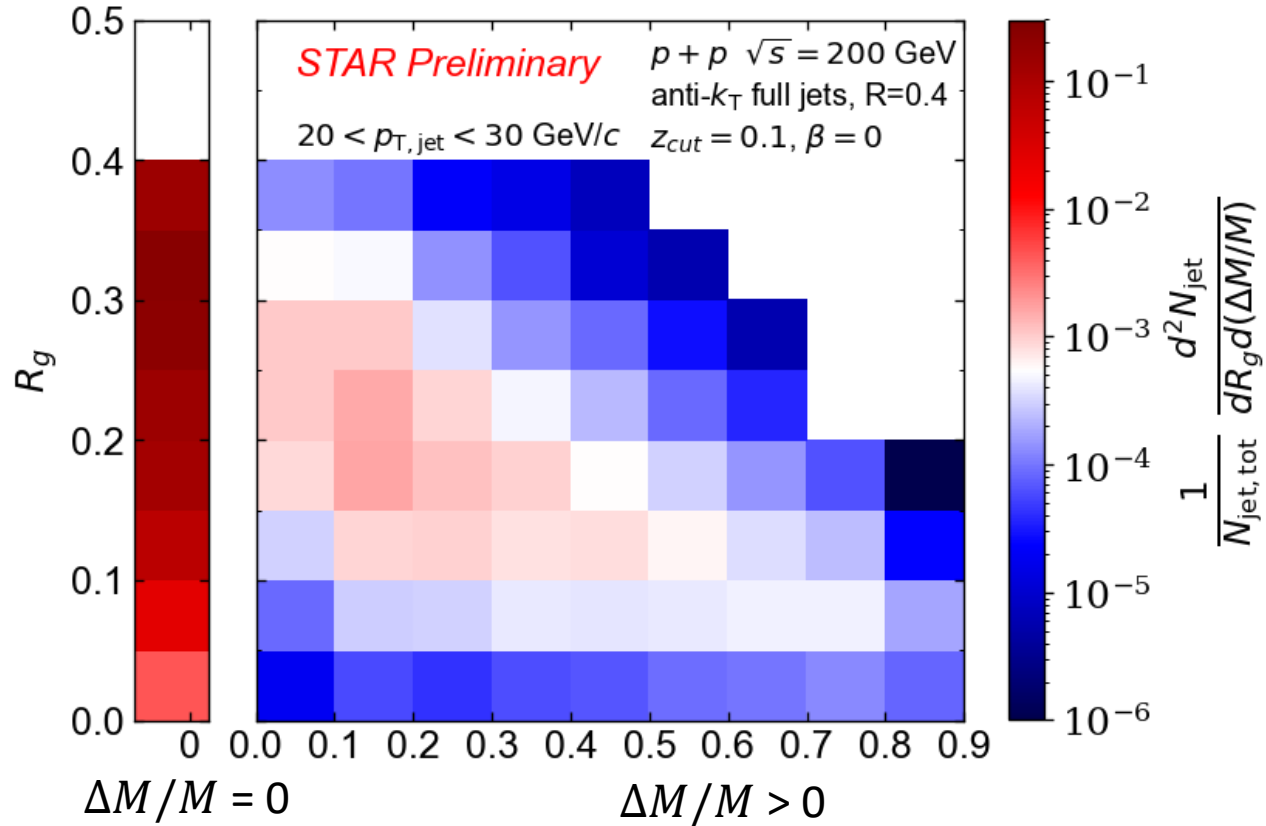
CollinearDrop groomed jet mass



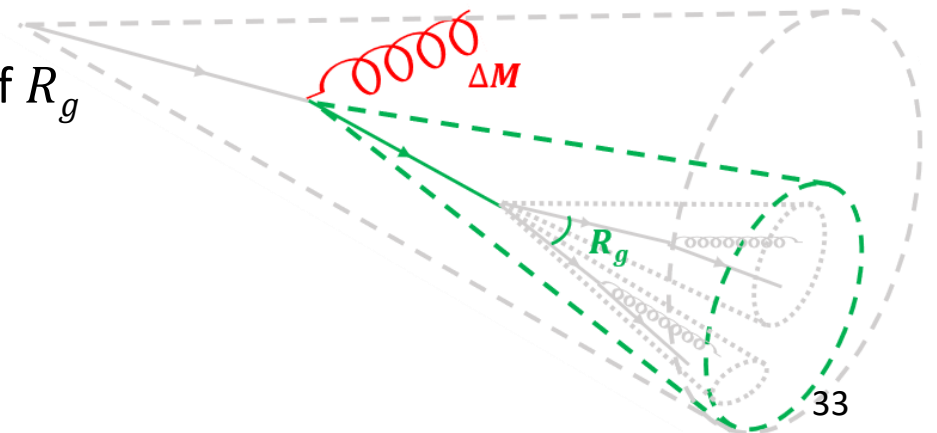
Measurement excludes jets with $\Delta M = 0$

- First CollinearDrop groomed jet measurement, sensitive to soft radiation within jets.
- MC predictions qualitatively consistent with data.
- MultiFold allows us to correlate (combinations of) unfolded quantities.

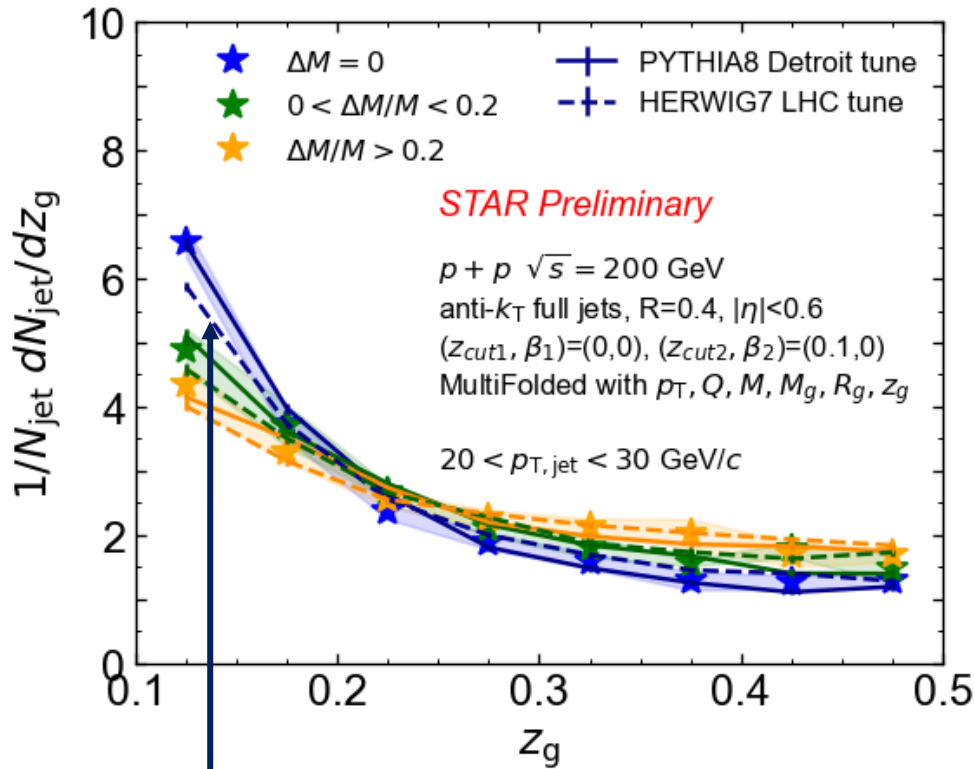
Soft radiation vs hard splitting angle



- The mean of $\Delta M/M$ distribution is anti-correlated with mean of R_g
 → consistent with angular ordered parton showers



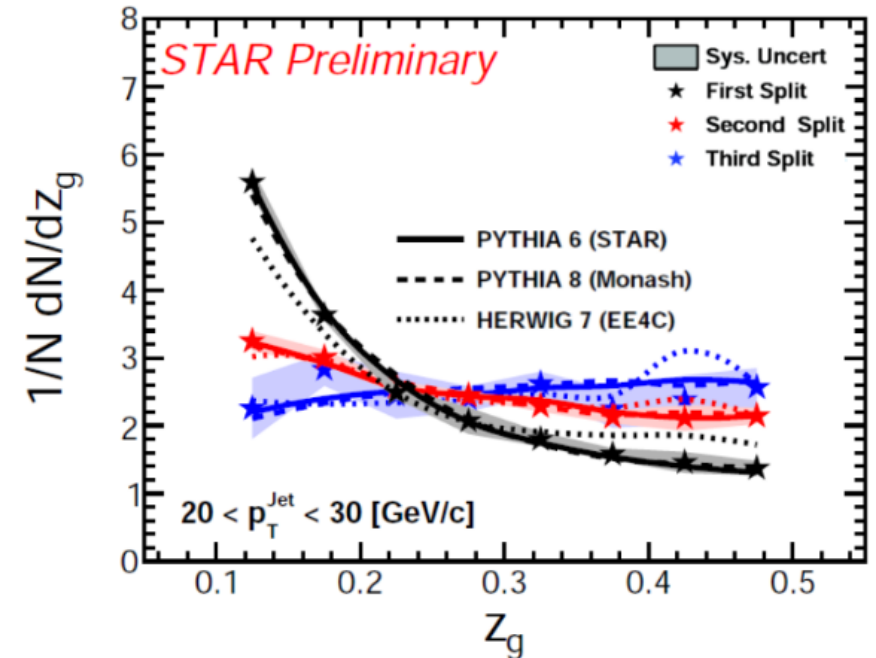
Soft radiation vs hard splitting momentum imbalance



Steeply falling \sim DGLAP $1/z$: pQCD

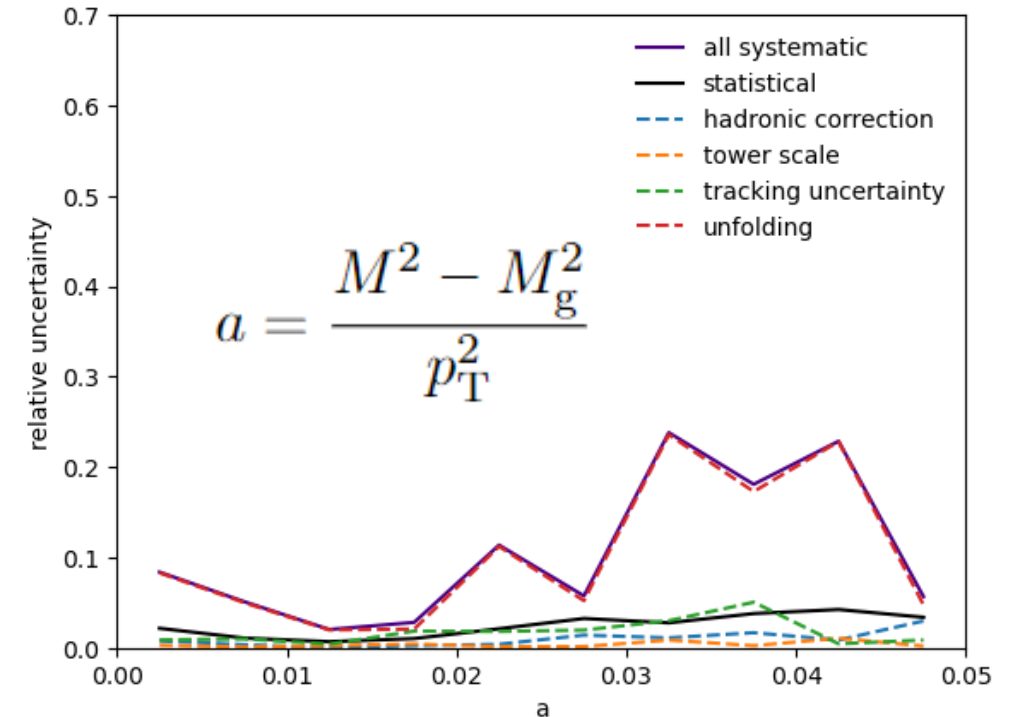
→ The first splitting that passes SoftDrop can still be non-perturbative, but if we apply the $\Delta M = 0$ selection, we can filter out some npQCD contribution due to the parton splitting

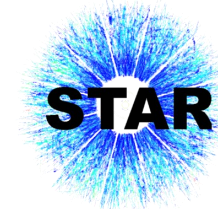
- The more mass that is groomed away relative to the original mass, the flatter the Z_g distribution is
 - Demonstrates that **early** soft wide angle radiation constrains the momentum imbalance of & the amount of npQCD contributions to **later** splittings
- MC models describe the trend of data



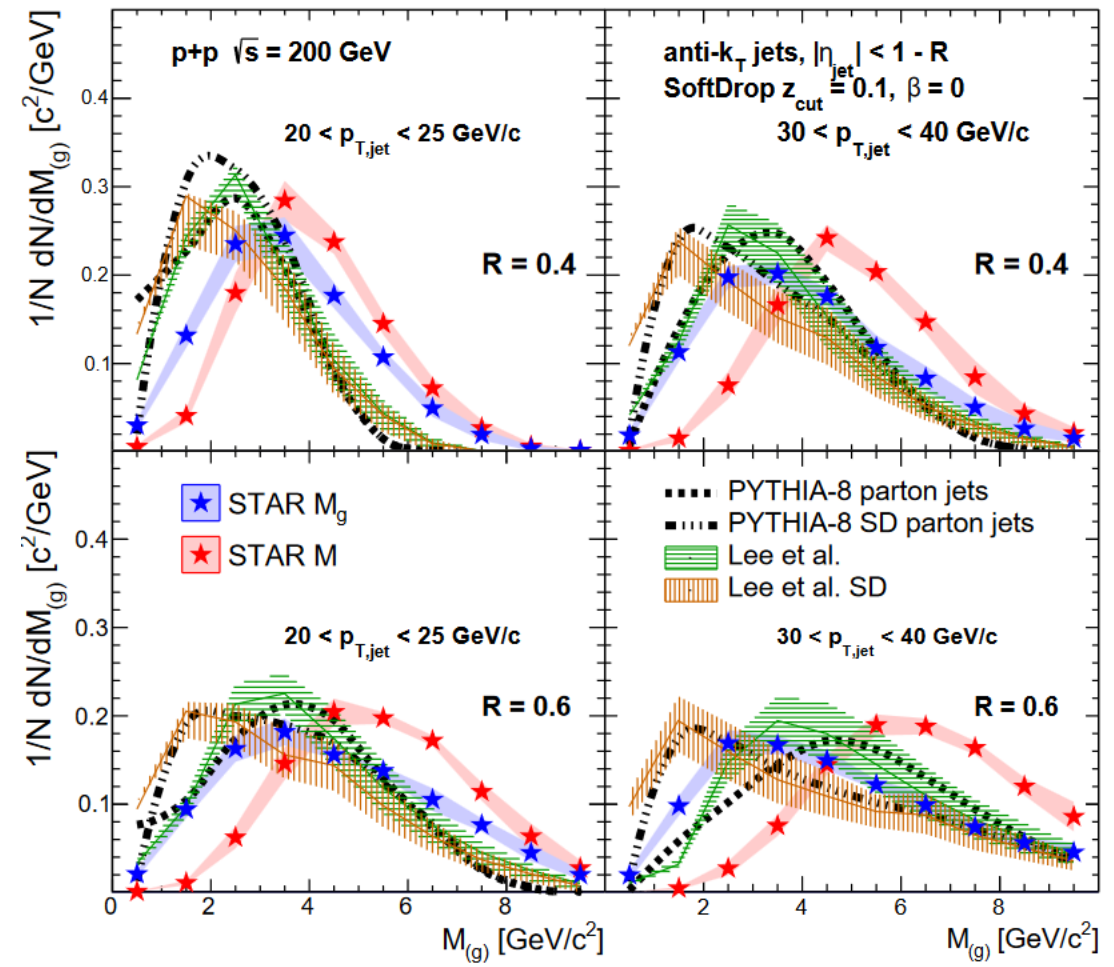
Systematic uncertainties

- Detector systematics
 - Hadronic correction 100% \rightarrow 50%
 - Tower scale +3.8%
 - Tracking uncertainty -4%
- Unfolding systematics
 - Unfolding seed
 - Iteration number variation
 - Prior shape variation to HERWIG7 and PYTHIA8
 - Nominal: prior = (generation, simulation)
= (PYTHIA6, PYTHIA6 + GEANT3 + embedding)
 - Varied to: prior \rightarrow reweight \otimes nominal prior ,
with $\text{reweight}(p_T, Q, M, M_g, R_g, z_g) = \frac{\text{Herwig truth}(p_T, Q, M, M_g, R_g, z_g)}{\text{Pythia6 truth}(p_T, Q, M, M_g, R_g, z_g)}$

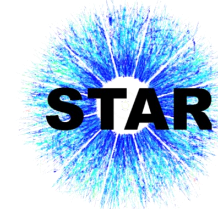




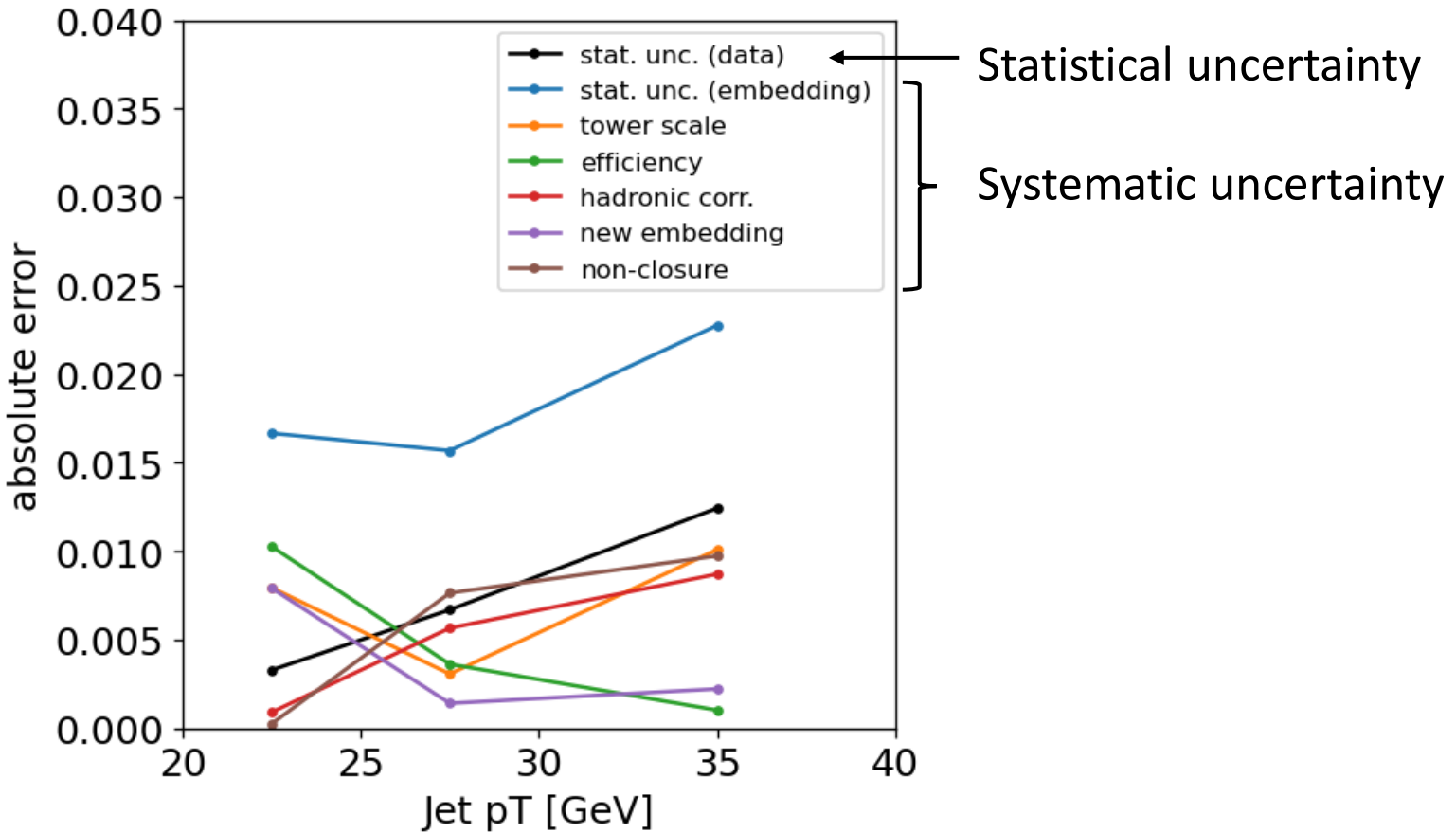
Jet mass: Comparison with models and calculations

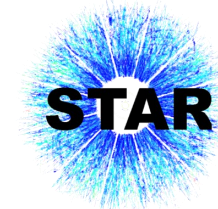


[STAR Collaboration. PRD 104, 052007\(2021\)](#)



Systematic uncertainty





Analysis method

Problem: pi0s (and other neutral hadrons) decay at the detector-level

Example:

jet at truth level

pi+ 6 GeV
pi0 4 GeV
pi- 3 GeV
pi- 1 GeV

jet at detector level

track(+) 6 GeV
track(-) 3 GeV
neutral 2 GeV
neutral 2 GeV
track(-) 1 GeV

leading/subleading is neutral
→ don't consider this jet

leading/subleading are both charged
→ include this jet for analysis
→ mistagged jet (shouldn't include this jet for analysis, but cannot identify it from data)

- How should we account for the neutral background?

207 found in Sec.2.1 of this [STAR analysis note](#). In addition to leptons, protons
208 and anti-protons, several other particles are also deemed as stable particles
209 at the particle level. Their list is available below.

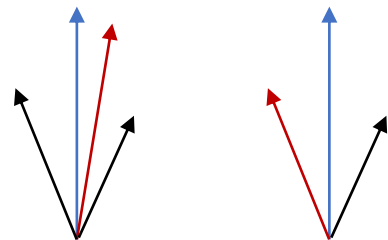
Solution: Switch to a “charged jet” measurement

$$\pi^0, \pi^\pm, \eta, K^+, K_S^0, K_L^0, \Sigma^\pm, \bar{\Sigma}^\pm, \Lambda, \bar{\Lambda}, \Xi^-, \bar{\Xi}^+, \Omega^-, \bar{\Omega}^+ \quad (3)$$

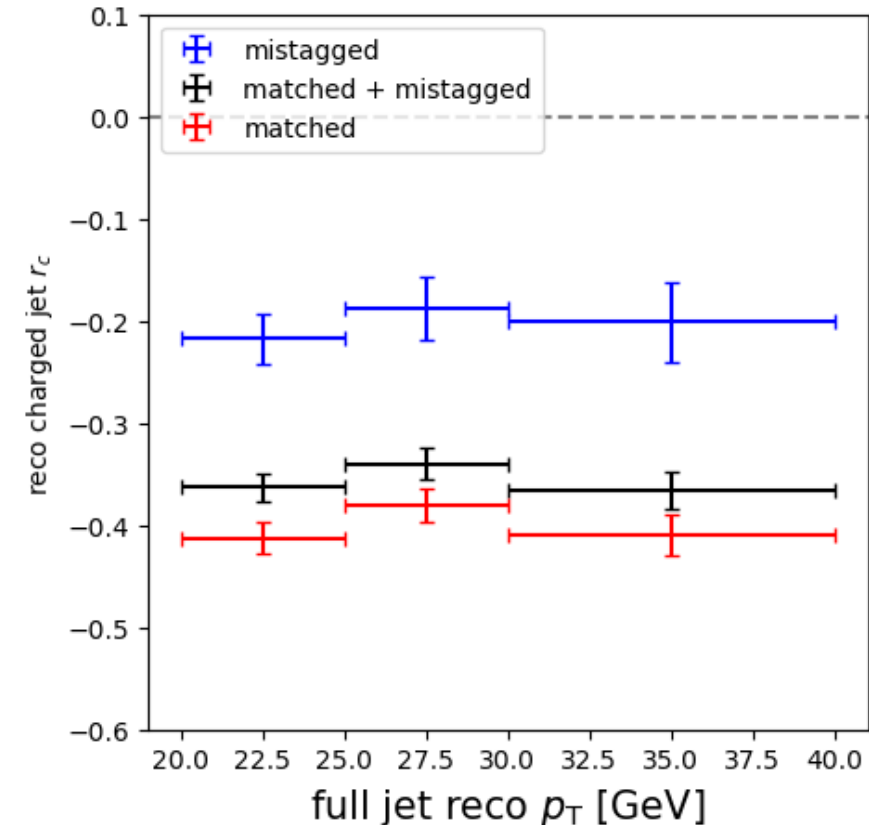
Analysis method

- Correction is needed!
 - mistagged subtraction to account for incorrectly identifying tracks that are not leading/subleading
 - bin-by-bin reweighting procedure to account for the jet energy scale
- Example: Why is mistagged subtraction necessary?

PYTHIA → GEANT



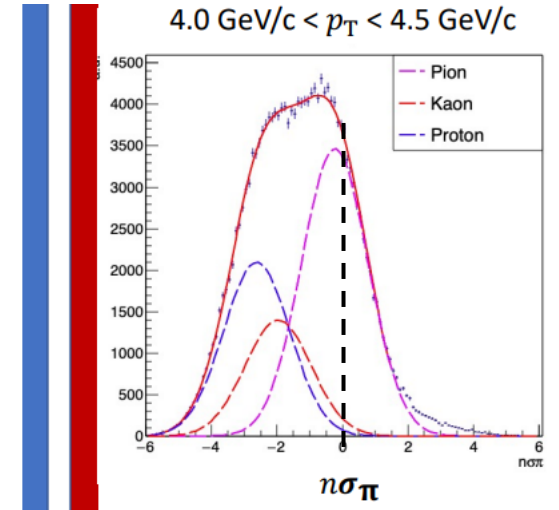
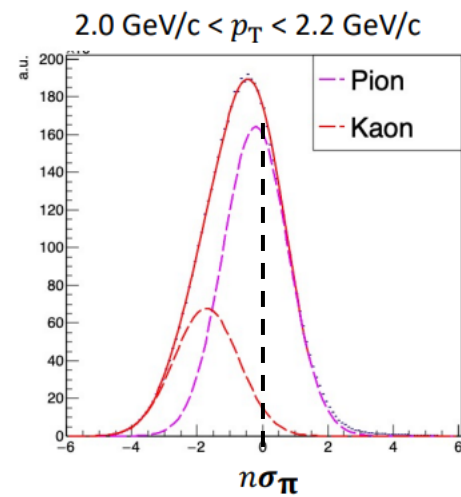
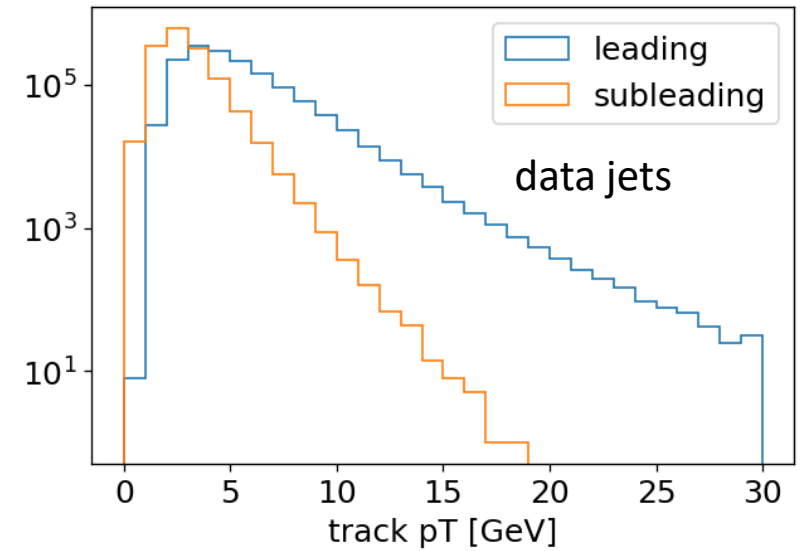
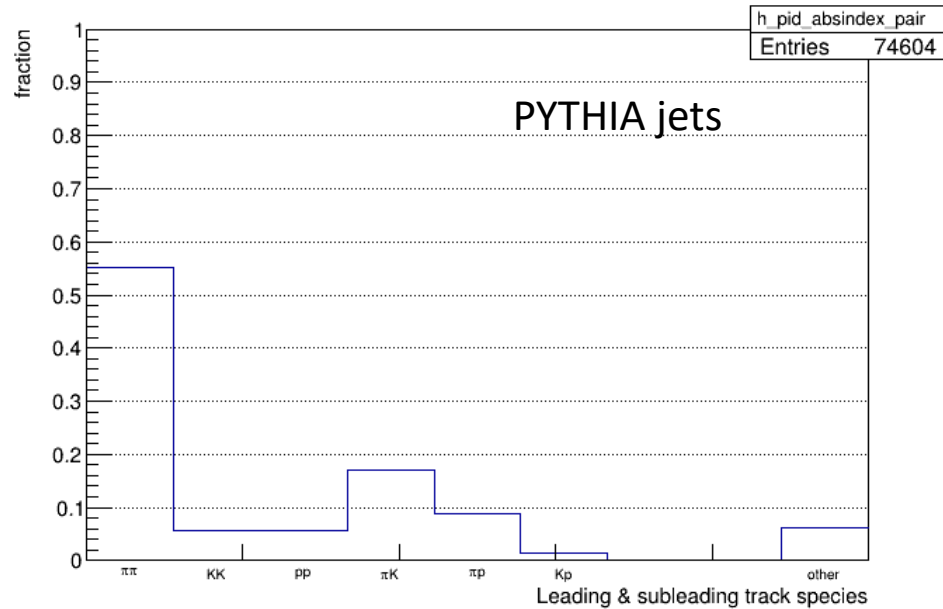
Mistagged example



Study done with embedding

Possibility of doing PID?

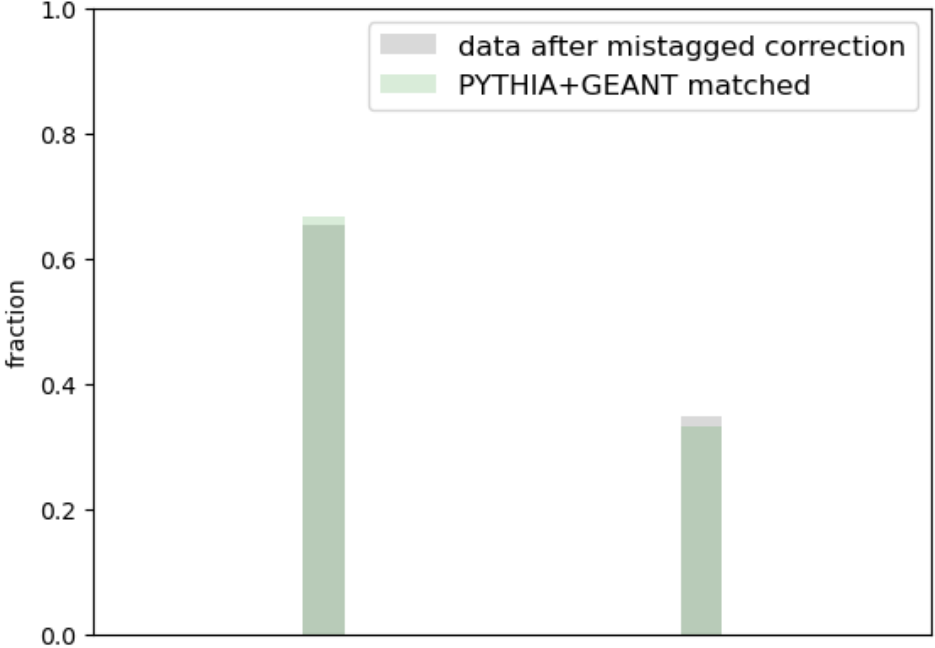
- It seems challenging to do event-by-event PID for high p_T π , K , p
- Pion selection
 - Cutting on $n(\sigma_{\pi}) > 0$ might suffer $\sim 50\%$ efficiency
 - The inclusive case already has $\sim 60\%$ purity



[Gabe's talk at HP 2023](#)

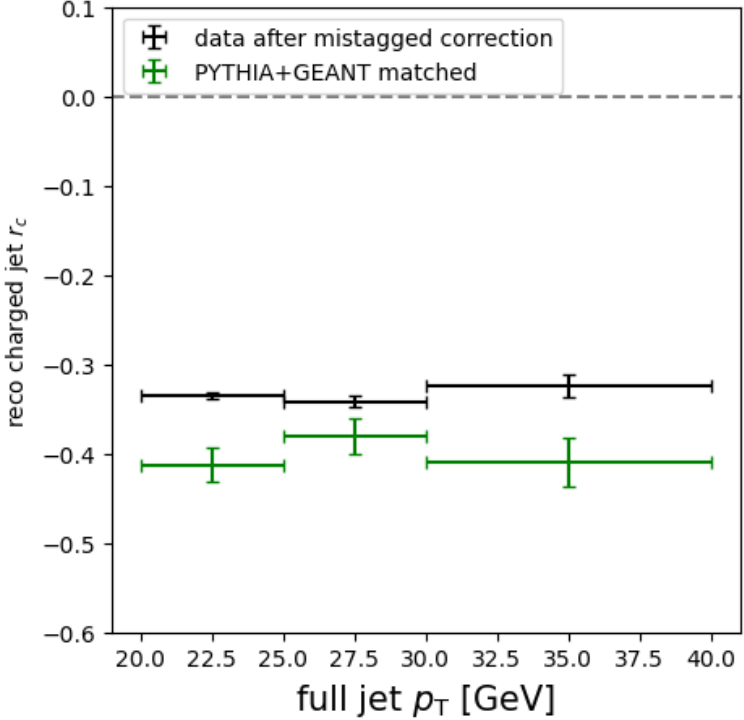
More details on the pT correction procedure

- For each detector-level pT bin, reweight the charge sign distribution in embedding, to match data after mistagged correction.



Leading dihadrons are opposite sign Leading dihadrons are same sign

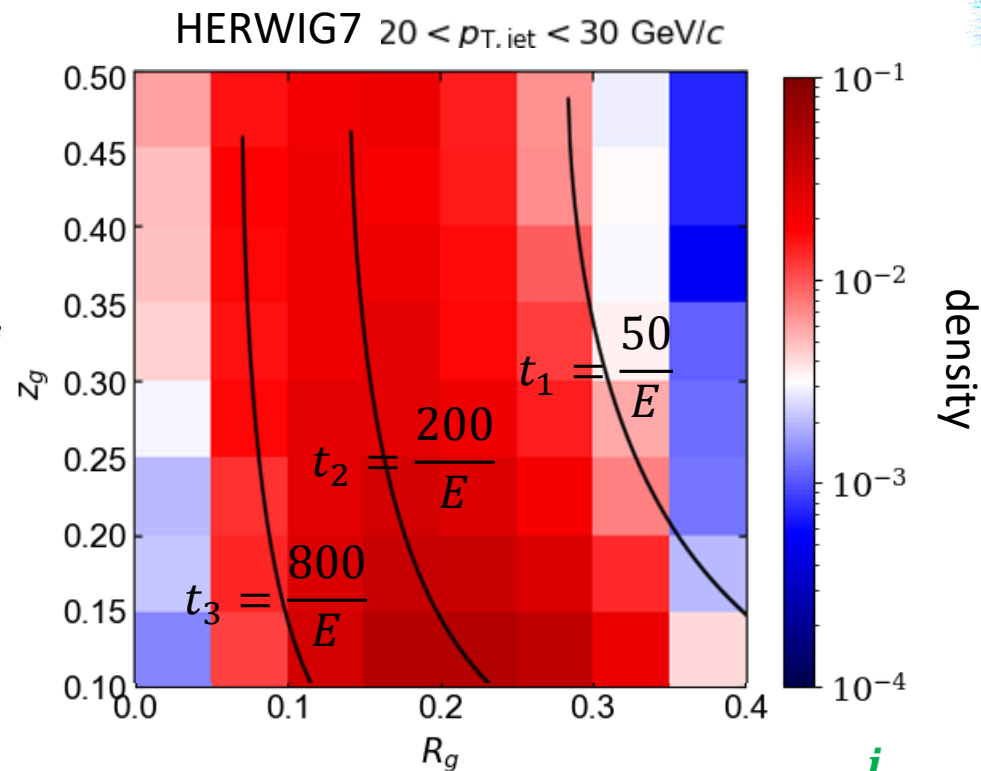
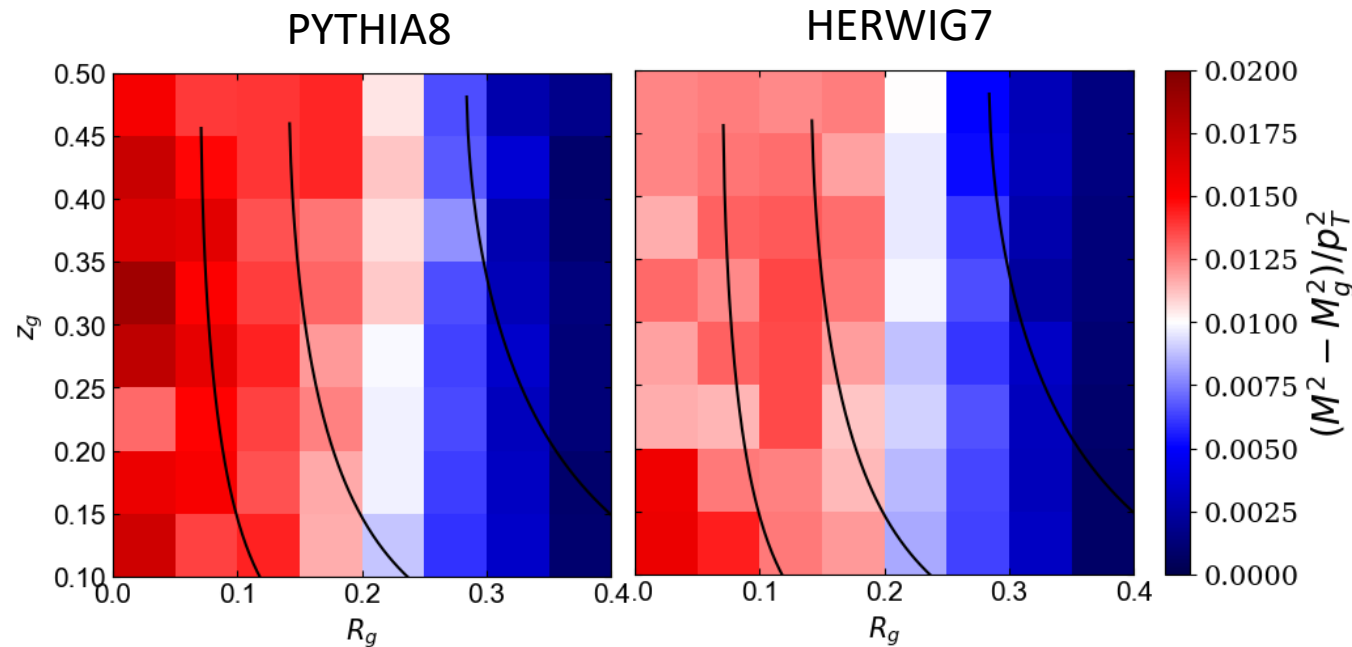
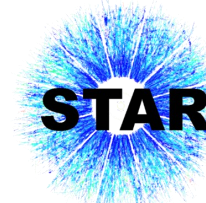
Example: $20 < p_{T, \text{reco}} < 25$ GeV
 Reweights = [0.94, 1.13]



- This means, if we weight the opposite pair jets down (0.94) and the same sign jets up (1.13) in PYTHIA+GEANT, then we can get the PYTHIA+GEANT r_c to match data (after mistagged subtraction).

- Since jets are matched between PYTHIA and PYTHIA+GEANT, the reweights automatically carry onto the PYTHIA jets too. This matching essentially serves the role of a response matrix, since it also contains the information such as the truth jet pT distribution given a reconstructed pT.

Discussion

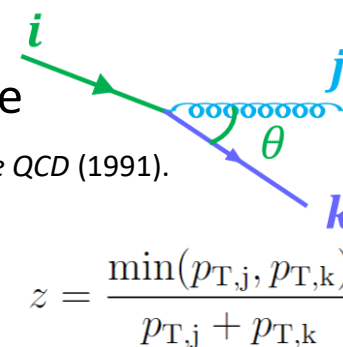


- Lines of constant formation time

Y. L. Dokshitzer, et al. *Basics of Perturbative QCD* (1991).

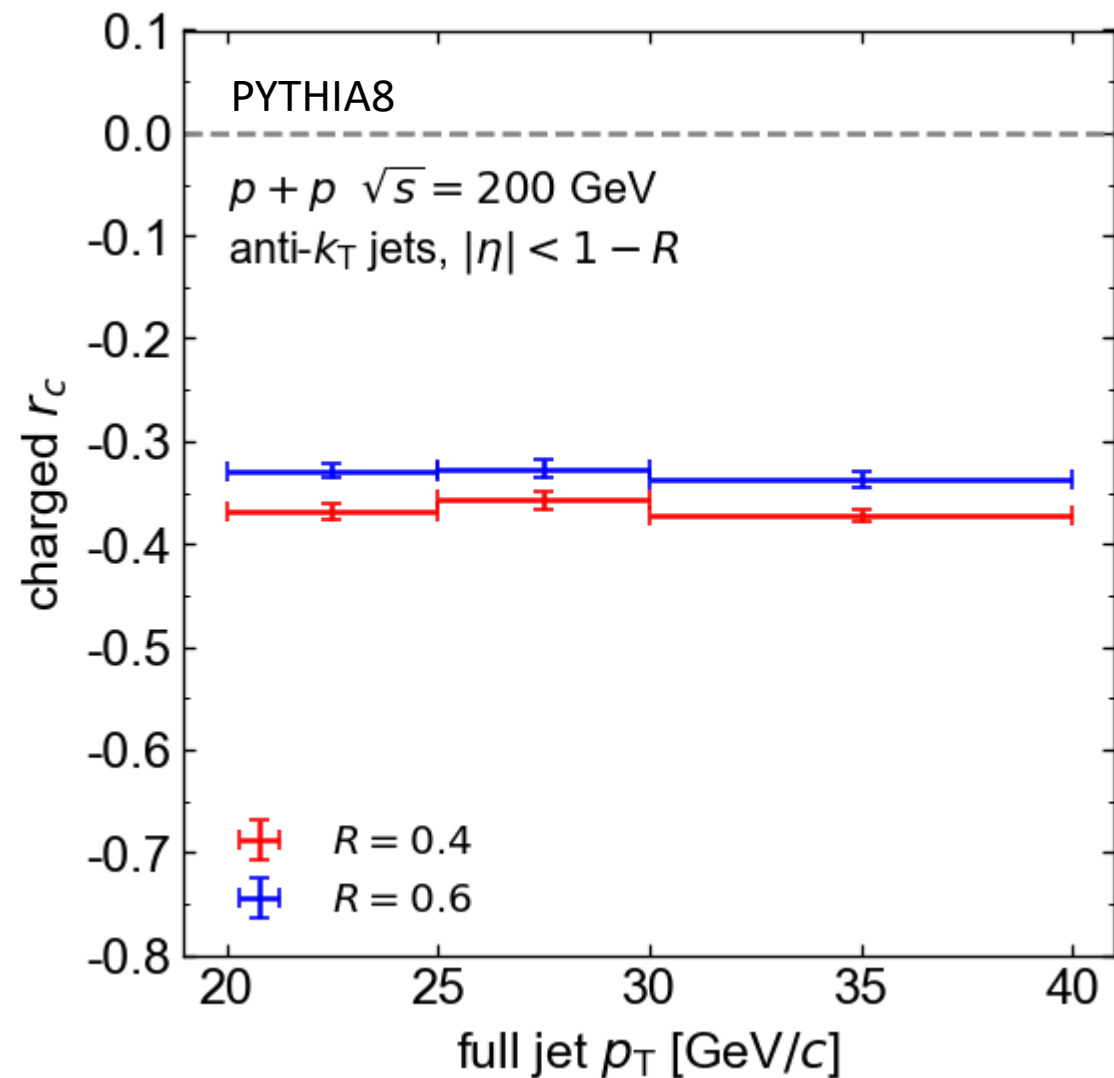
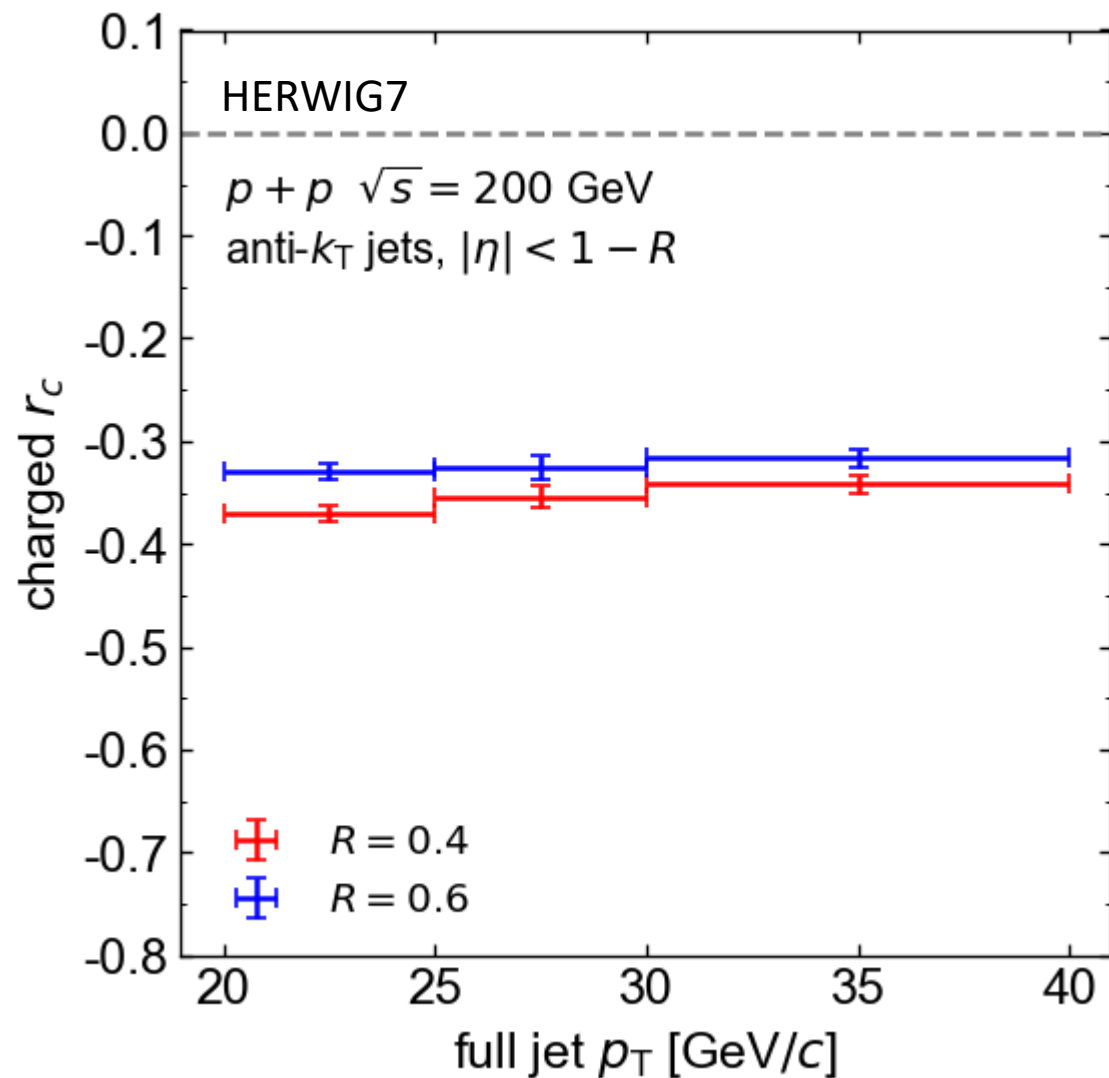
Using $t_F \sim \frac{1}{2Ez(1-z)(1-\cos(\theta))}$
 solve for $z(\theta, z < 0.5)$:

$$z = \frac{1}{2} \left[1 - \sqrt{1 - \frac{2}{tE(1-\cos(\theta))}} \right]$$



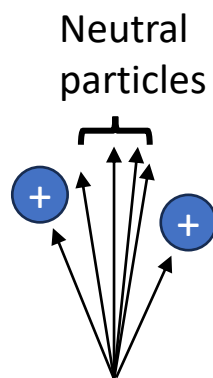
$$z = \frac{\min(p_{T,j}, p_{T,k})}{p_{T,j} + p_{T,k}}$$

Discussion



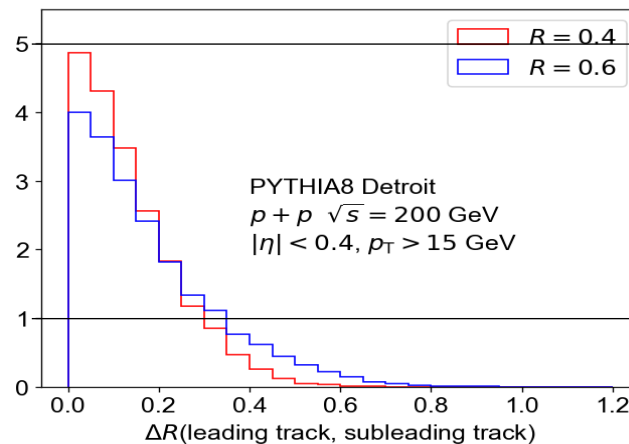
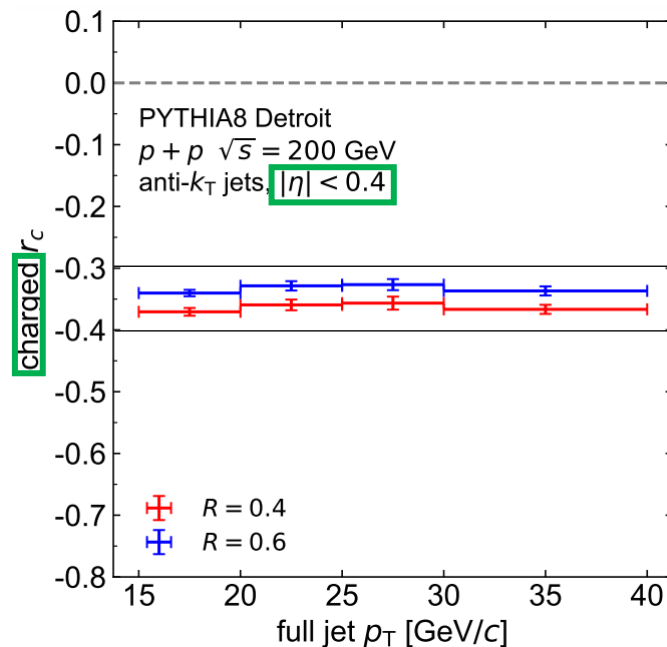
Discussion

- r_c is less negative with larger jet $R \rightarrow$ likely that “background” track pairs are included with larger R
- Potentially introduce a jet neutral energy fraction requirement to reduce this effect
 - Fragmentation bias? Or does this bring us closer to the “original definition”?



Example of a jet with a large R and large neutral core

Definition used in our measurement



Original definition as proposed by Chien et al. PRD 105 051502 (2022)

