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

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

2nd workshop on advancing the understanding of non-perturbative QCD using energy flow

11/7/2023

Supported in part by U.S. DEPARTMENT OF ENERGY | Office of Science

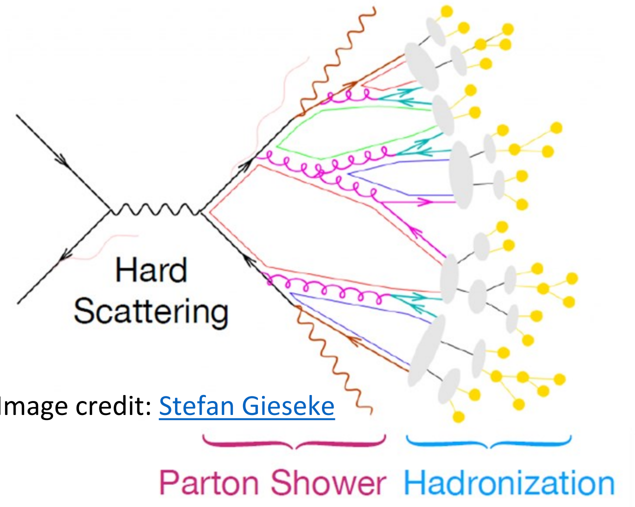


Image credit: Stefan Gieseke

Study the non-perturbative contribution in jets

- Soft component of the parton shower
  → CollinearDrop jet measurement

- Hadronization
  → Charge correlator ratio measurement

CollinearDrop grooming

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

Chien and Stewart JHEP 06 (2020) 64.

- 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 $M = |\Sigma_{i \in \mathrm{jet}}\ p_i| = \sqrt{E^2 - |\vec{p}|^2}$

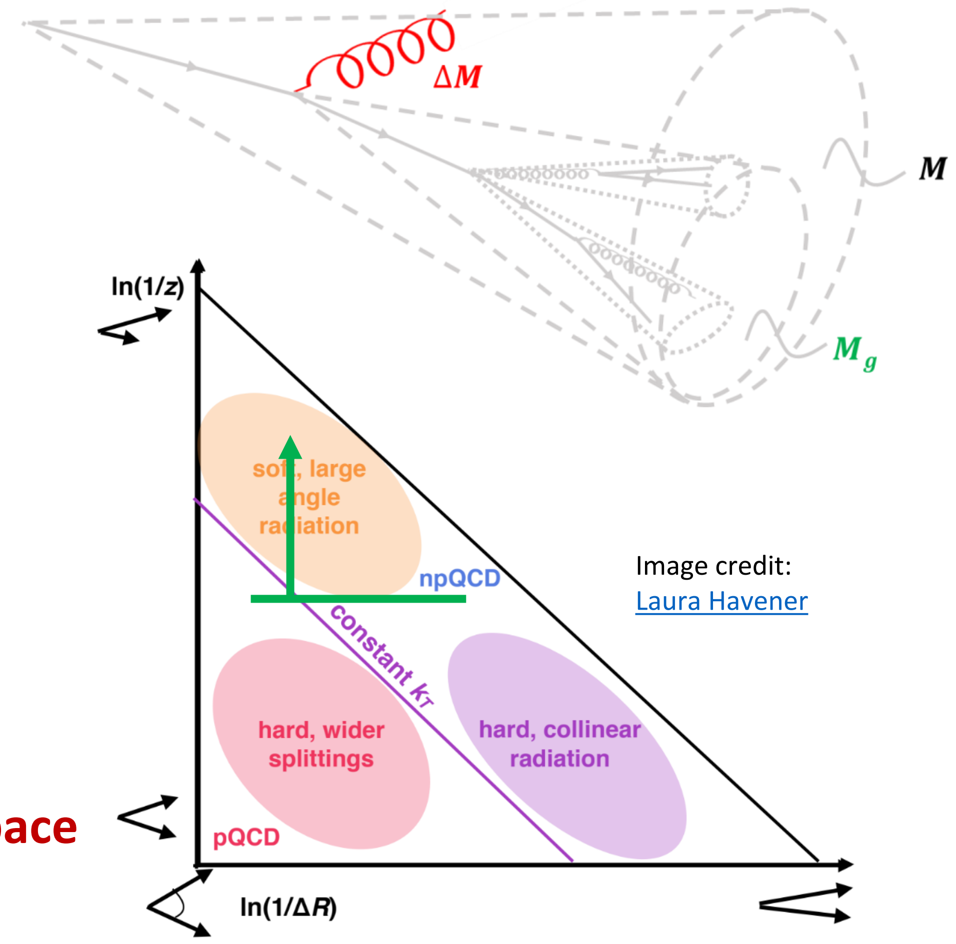


Image credit: Laura Havener

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

CollinearDrop vs SoftDrop correlation

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

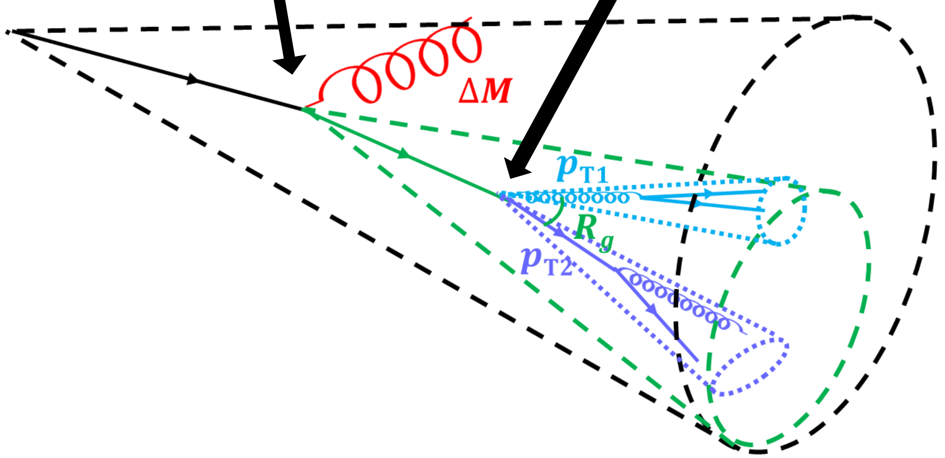
- How does the amount of soft radiation correlate with the angular and momentum scale of a hard splitting? → how an **early** emission affects a **later** splitting

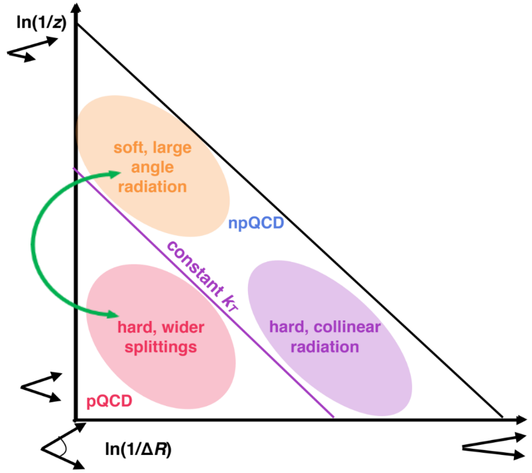
Captured by dM/M:
~ amount of **initial** wide and soft emission

Assuming angular ordering of parton shower

Captured by Zg and Rg:
~ imbalance and angular scale of a **later** splitting

$$\Delta M/M = \frac{M - M_g}{M}$$





Note: Hadronization effects smear the distribution but don't affect the correlations (see backup)

CollinearDrop vs SoftDrop correlation @ STAR: Analysis

- Reconstruct anti-k_T full 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
  - $Q^\kappa = \frac{1}{(p_{\mathrm{T}\text{jet}})^\kappa}\sum_{i\in\text{jet}} q_i \cdot (p_{\mathrm{T}i})^\kappa$
  - $M = |\Sigma_{i\in\text{jet}}\, p_i| = \sqrt{E^2 - |\vec{p}|^2}$
  - R_g: groomed jet radius
  - z_g: shared momentum fraction

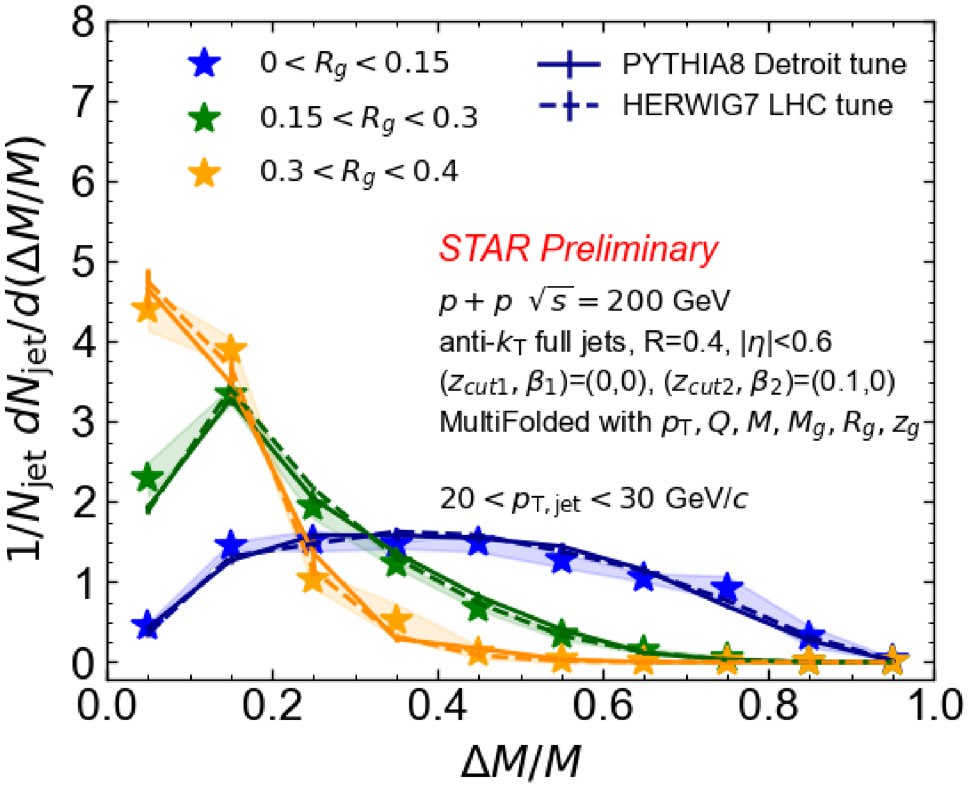
$$z_\mathrm{g} = \frac{\min(p_{\mathrm{T},1}, p_{\mathrm{T},2})}{p_{\mathrm{T},1} + p_{\mathrm{T},2}} > z_\mathrm{cut}(R_\mathrm{g}/R_\mathrm{jet})^\beta$$

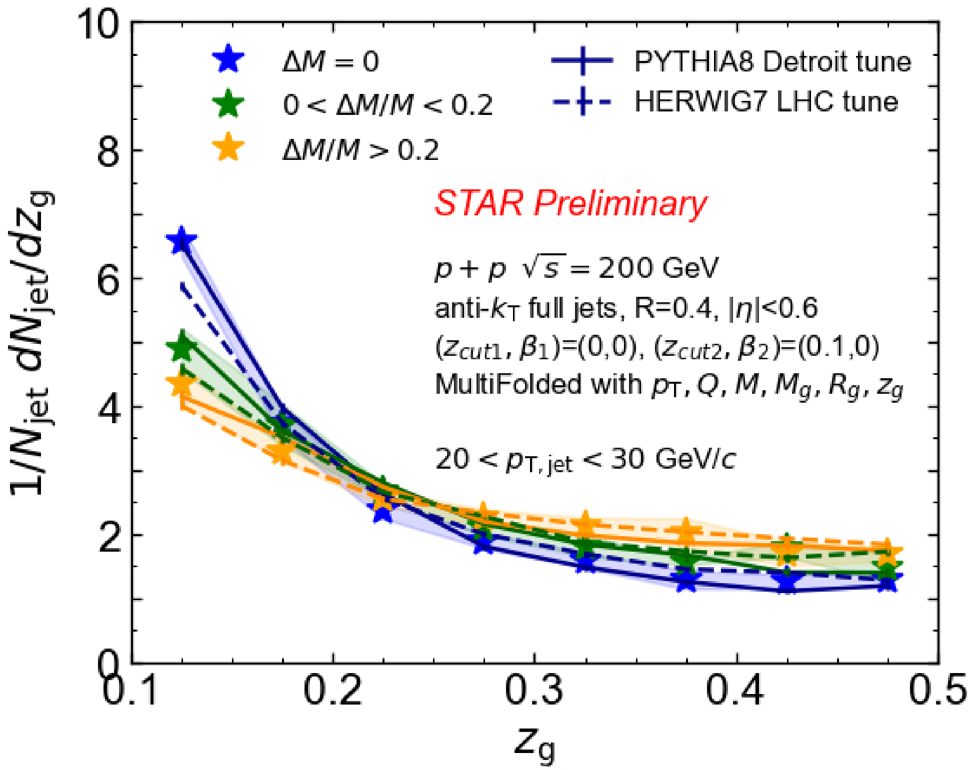
  - M_g: groomed jet mass

All 6 observables are simultaneously unfolded in an unbinned way!

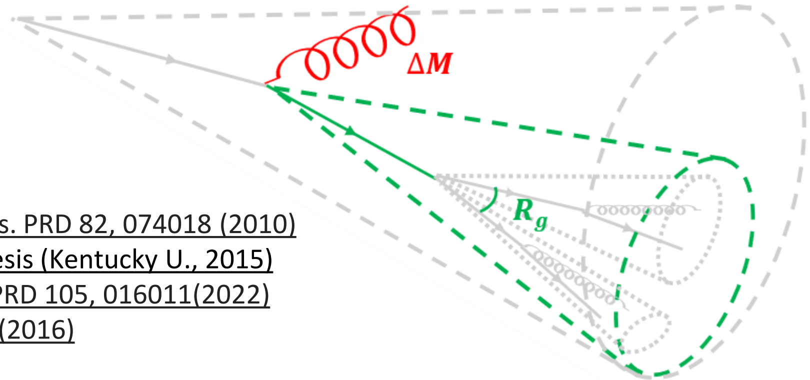
- Uncertainties due to prior choice accounted for through 6D reweighting based on PYTHIA8 or HERWIG (see backup)

CollinearDrop vs SoftDrop correlation @ STAR: 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 phase space and momentum imbalance of **later** splittings
- MC models describe the trend of 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: Bellm, et al. EPJC 76, 196 (2016)

- See arXiv:2307.07718 for more details and more results!

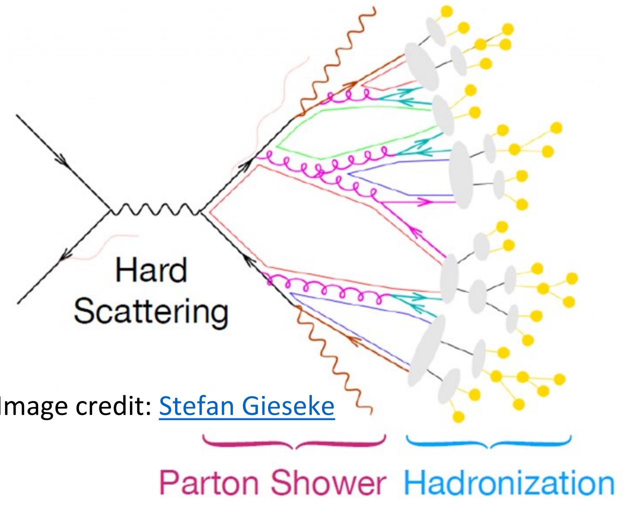


Image credit: Stefan Gieseke

Study the non-perturbative contribution in jets

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

Charge correlator ratio

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.

- Measure in pp collisions at RHIC & LHC and in ep(A) at the future EIC!

Chien et al. PRD 105 051502 (2022)

Talk by Charles earlier in this session

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

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

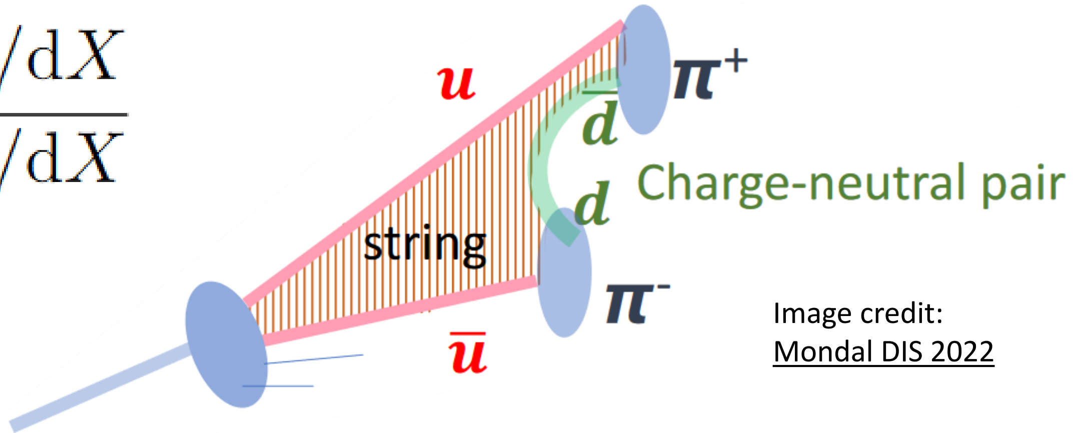


Image credit:
Mondal DIS 2022

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 makes it less susceptible to the background.

- Measure in heavy-ion collisions!

Charge correlator ratio @ STAR: Analysis

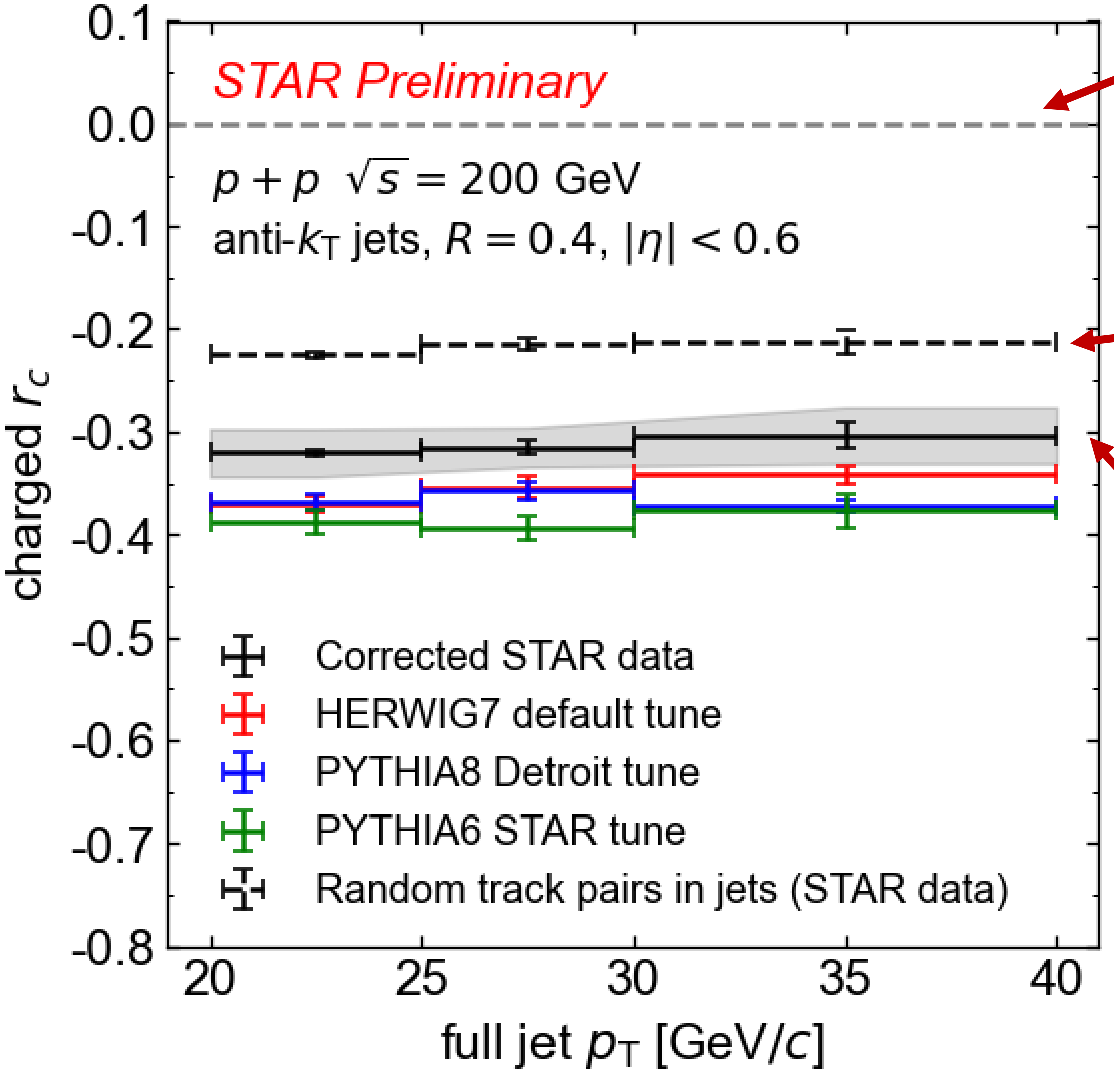
- Reconstruct anti-k_T **full jets** with R=0.4 from 200 GeV pp collisions
- At detector level, the decay of neutral hadrons affects the ordering of particles in jet, so we consider a **charged $\boldsymbol{r_c}$** measurement. See backup slides for an example.

$$r_c(X) = \frac{\mathrm{d}\sigma_{h_1h_2}/\mathrm{d}X - \mathrm{d}\sigma_{h_1\bar{h}_2}/\mathrm{d}X}{\mathrm{d}\sigma_{h_1h_2}/\mathrm{d}X + \mathrm{d}\sigma_{h_1\bar{h}_2}/\mathrm{d}X}$$

h_1h_2: same charge **tracks**,
$h_1\overline{h_2}$: opposite charge **tracks**

- In addition, we performed a mistagged subtraction to account for incorrectly identifying tracks that are not leading/subleading, followed by a bin-by-bin reweighting procedure to account for the jet energy scale.

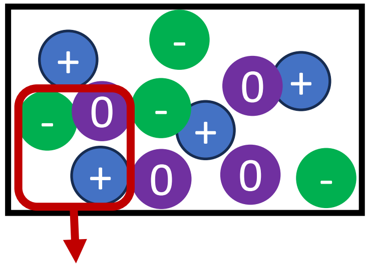
Charge correlator ratio @ STAR: Result



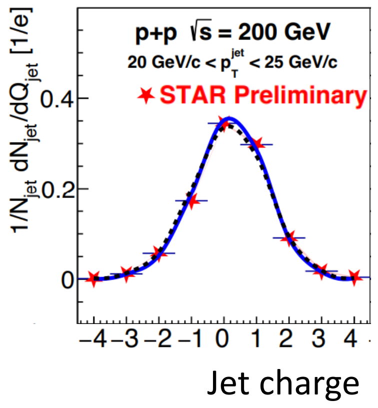
Random pairs in an uncorrelated (infinite) bath with no net charge: $\boldsymbol{r_c = 0}$

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: $\boldsymbol{r_c \approx -0.2}$
- leading and subleading tracks in jet. additional correlation from fragmentation: $\boldsymbol{r_c \approx -0.3}$

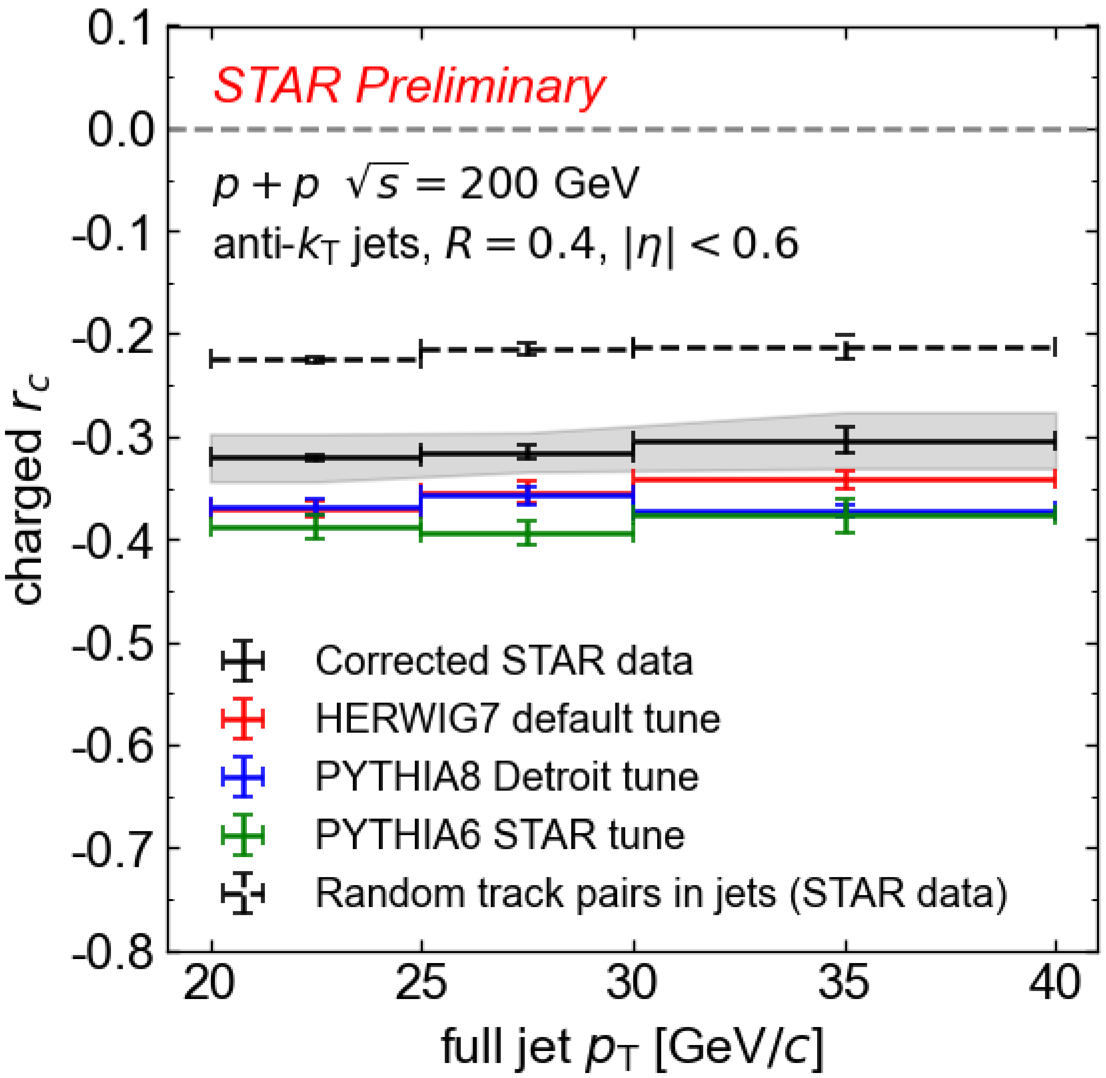


clustered into a jet



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: Bellm, et al. EPJC 76, 196 (2016)

Charge correlator ratio @ STAR: Result



Weak dependence on jet $\boldsymbol{p}_{\mathbf{T}}$

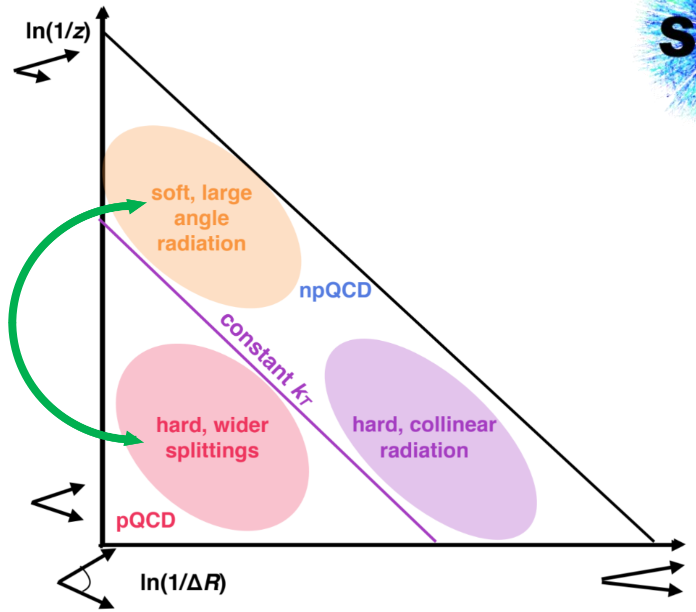
Models based on Lund string fragmentation and cluster hadronization both underpredict $\boldsymbol{r}_{\boldsymbol{c}}$ in data.

→ In the future, study $\boldsymbol{r}_{\boldsymbol{c}}$ as functions of observables sensitive to pQCD→npQCD transition! ($\boldsymbol{k}_{\mathbf{T}}, \boldsymbol{t}_{\boldsymbol{f}}, \boldsymbol{z}$...)

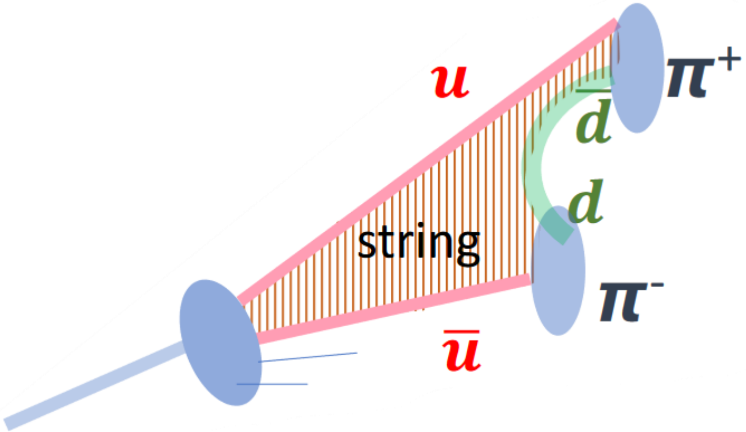
PYTHIA6 Perugia + STAR tune: Skands. PRD 82, 074018 (2010)
J. K. Adkins, PhD thesis (Kentucky U., 2015)
PYTHIA8 Detroit tune: Aguilar et al. PRD 105, 016011(2022)
HERWIG7: Bellm, et al. EPJC 76, 196 (2016)

Conclusions

- Study the **soft**-**hard** correlation within jets with CollinearDrop vs SoftDrop jet observables
  - The **NP 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 to RHIC data!



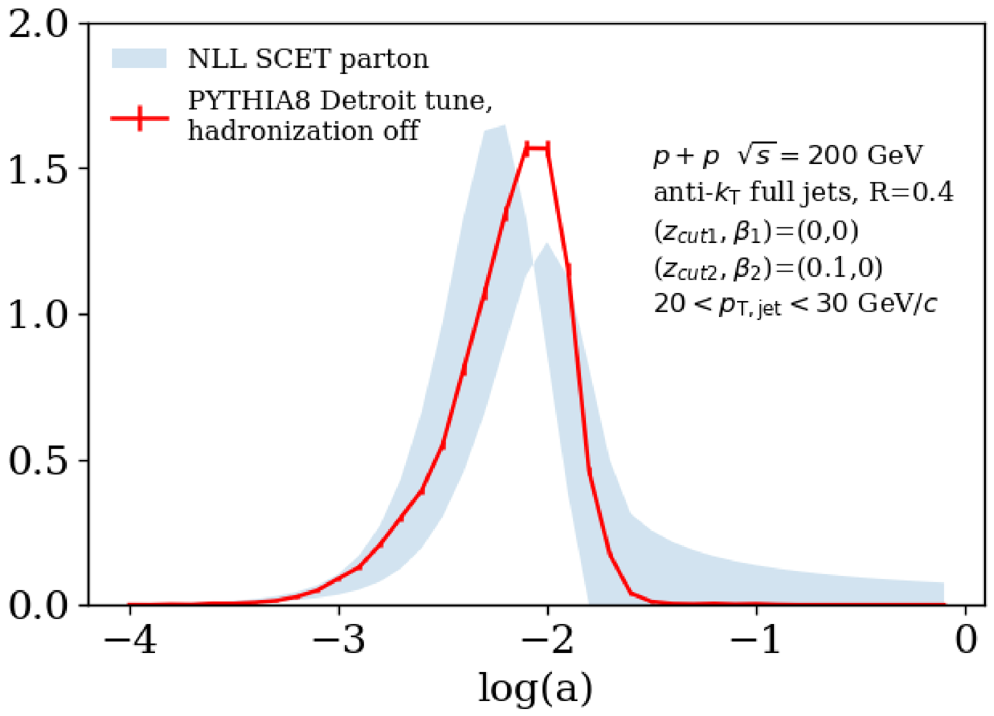
- Study **hadronization** with the charge correlator ratio
  - Data show a **weaker correlation** between leading and subleading particles in jet than models
  - How does r_c depend on other jet substructure observables? Can potentially provide more distinguishing power for models and help us better understand the npQCD evolution!



Backup

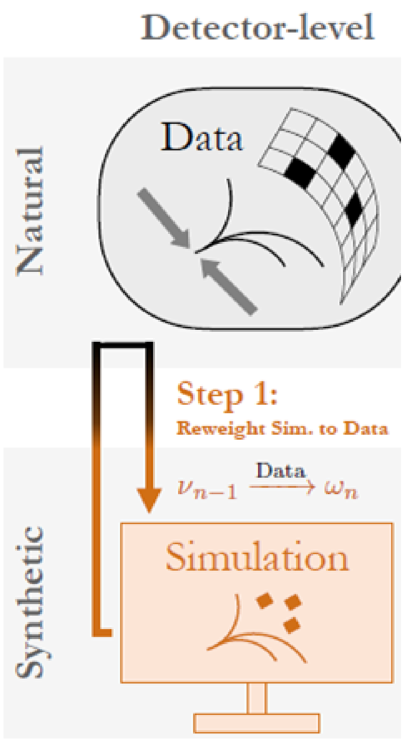
CollinearDrop groomed jet mass

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

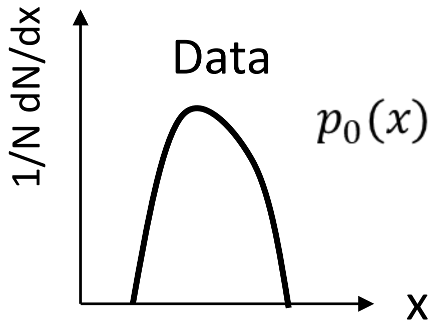


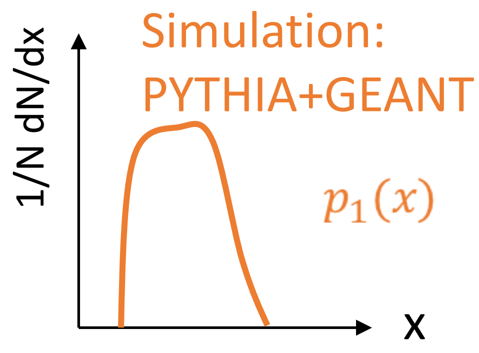
$$a = (M^2 - M_g^2)/p_T^2$$

Method: machine learning



Detector-level





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

Method: machine learning

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

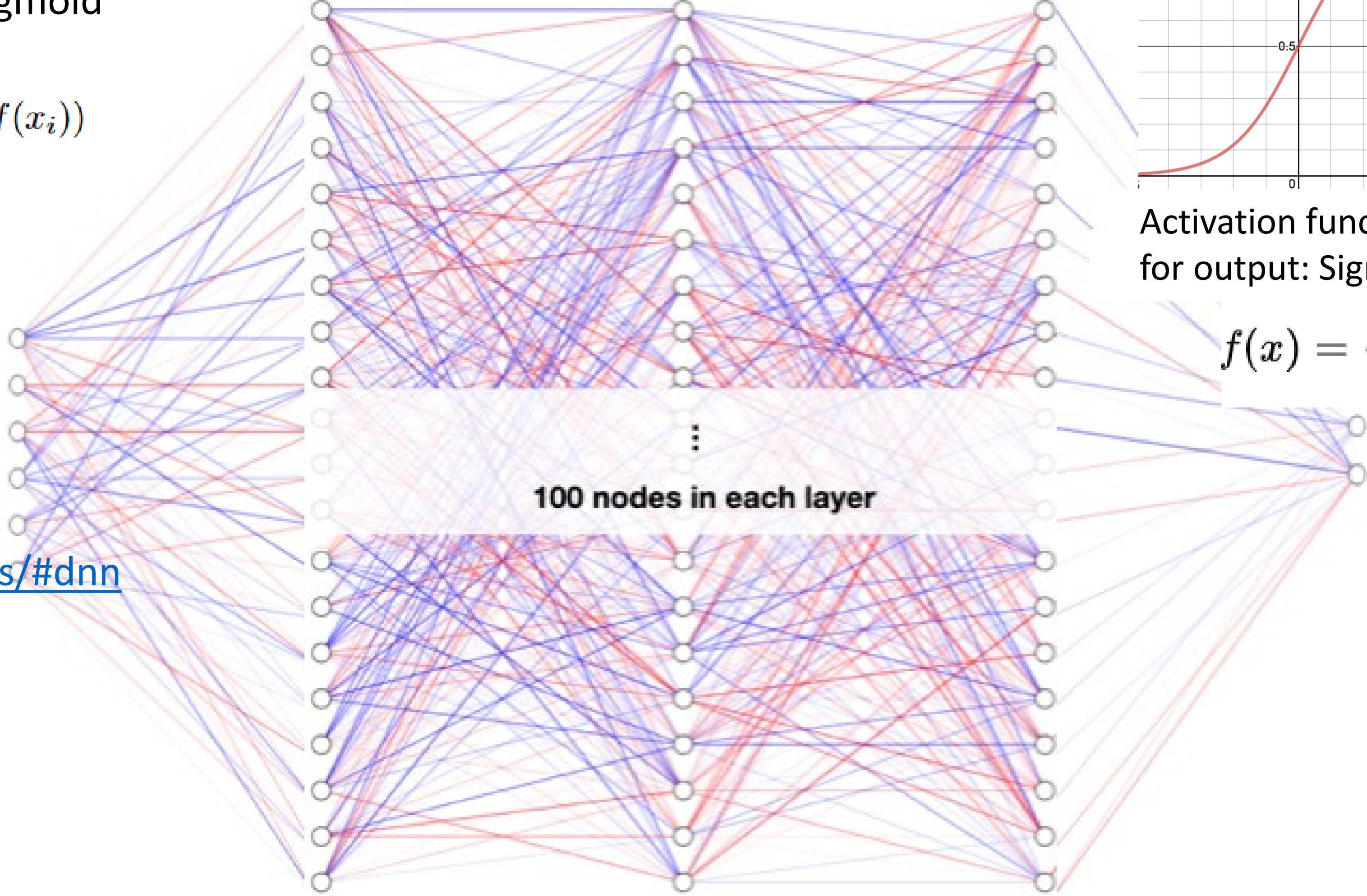
$$\text{loss}(f(x)) = -\sum_{i \in 0} \log f(x_i) - \sum_{i \in 1} \log(1 - f(x_i))$$

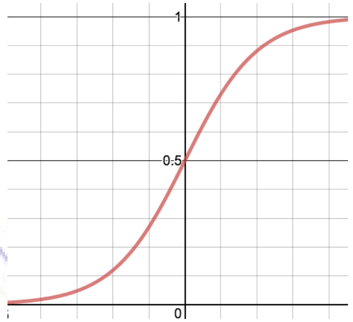
- Optimization algorithm: Adam
  https://arxiv.org/pdf/1412.6980.pdf
- Nodes per dense layer: [100,100,100]
- Output dimension: 2
- Input dimension: 6

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

Activation function for dense layers: Rectified linear unit

$$f(x) = x^+ = \max(0, x)$$





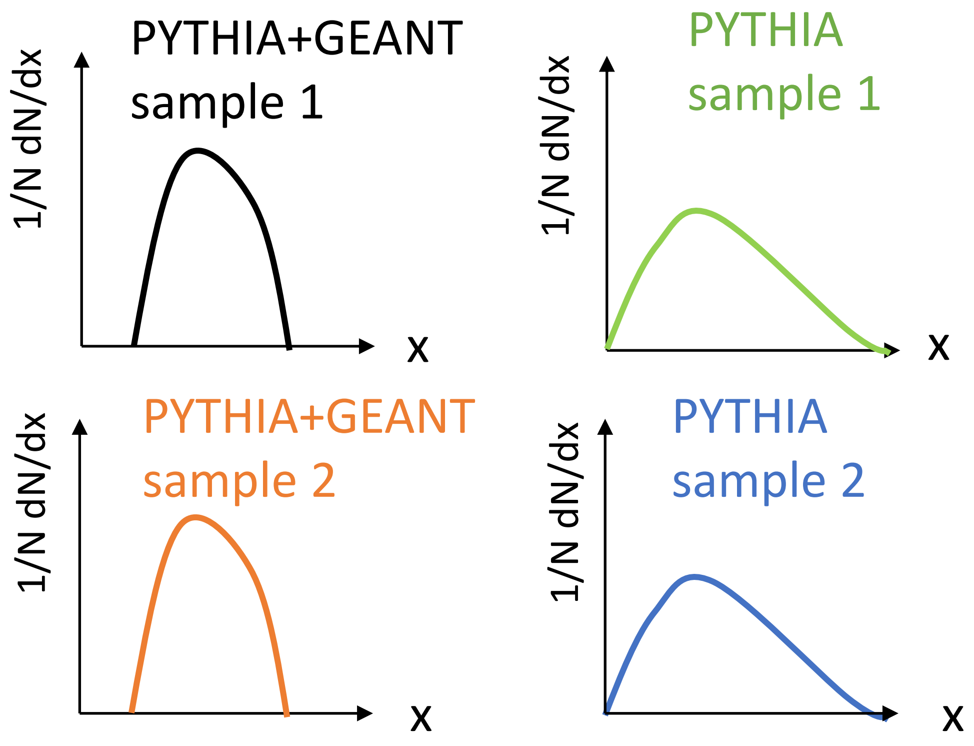
Activation function for output: Sigmoid

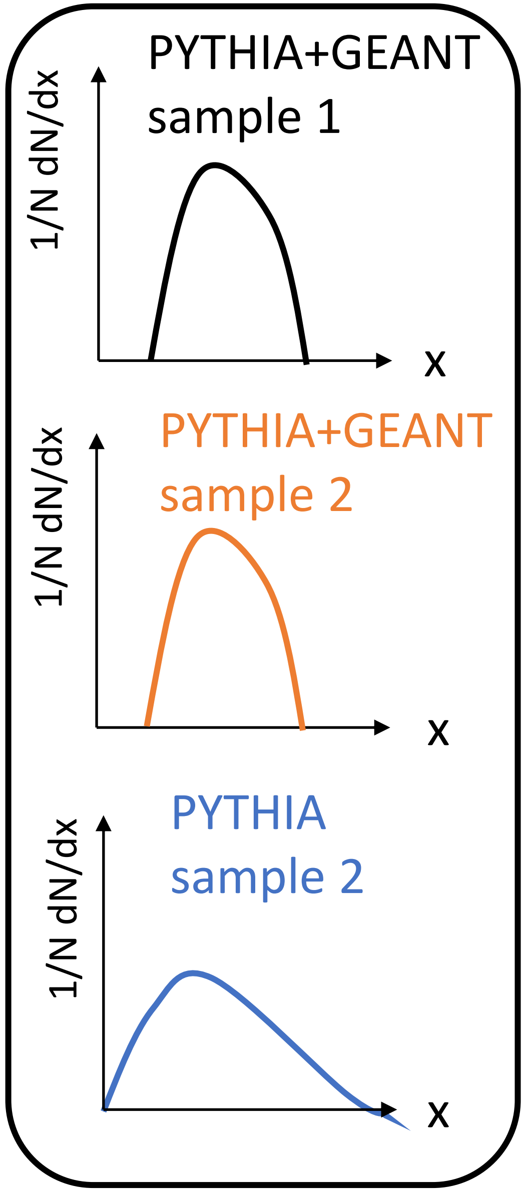
$$f(x) = \frac{1}{1 + e^{-x}}$$

Closure test for unfolding

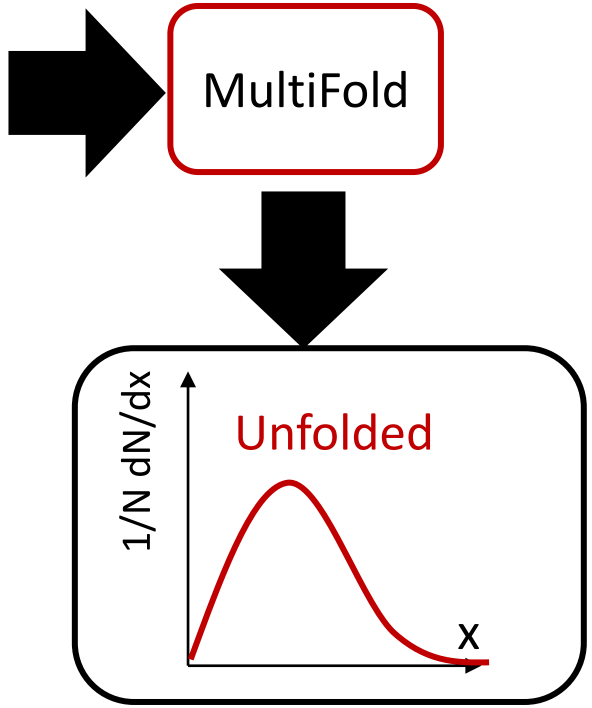
- Step 1: Separate matched jets from PYTHIA and PYTHIA+GEANT into 2 samples

Detector-level | Particle-level





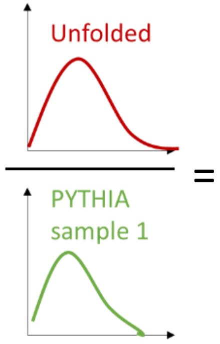
- Step 2: Unfold

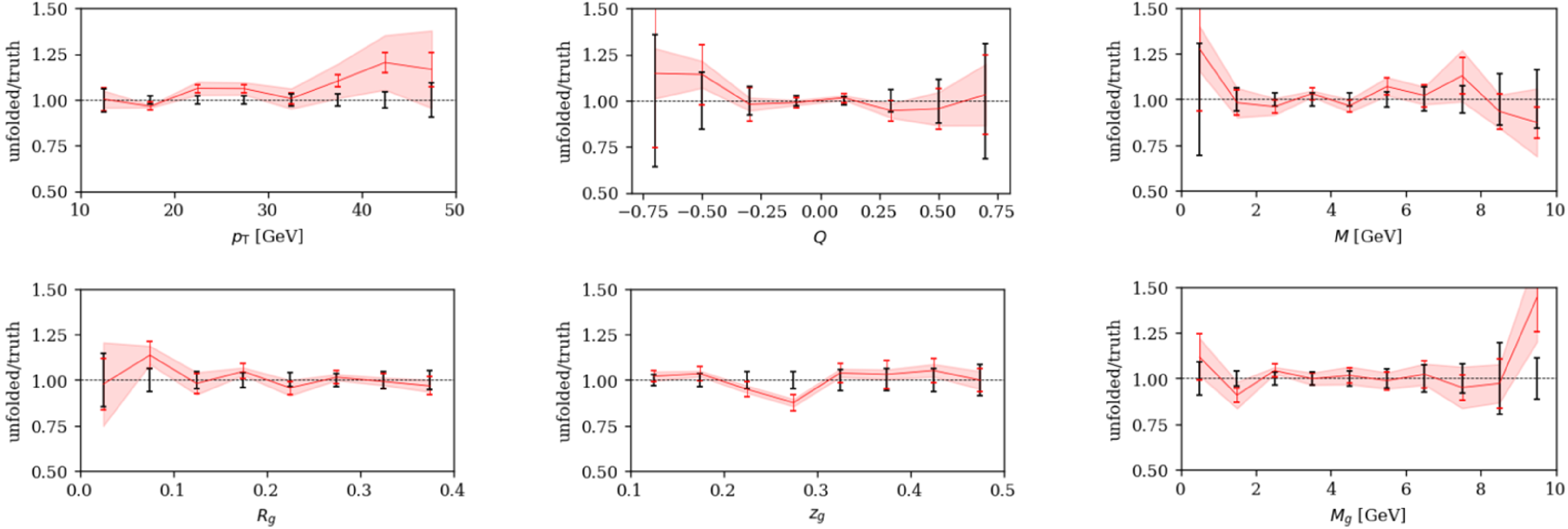


Closure test for unfolding

- Decent **closure** for all substructure observables

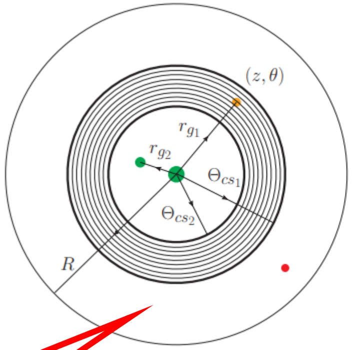
Unfolding unc. on data (not including misses)
Stat. unc. on sample 1
Stat. unc. on sample 2

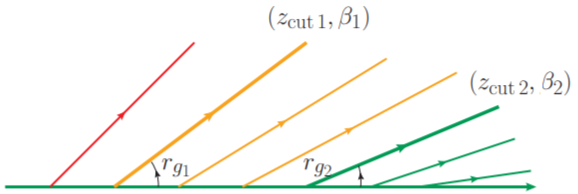




* 2D reweighting used for prior variation

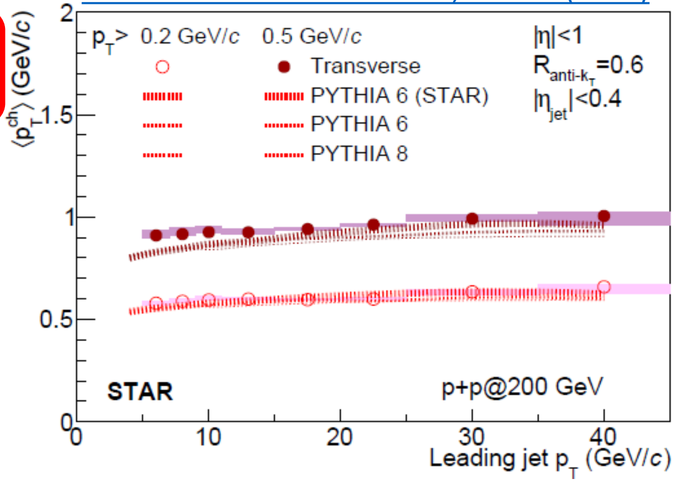
Underlying event

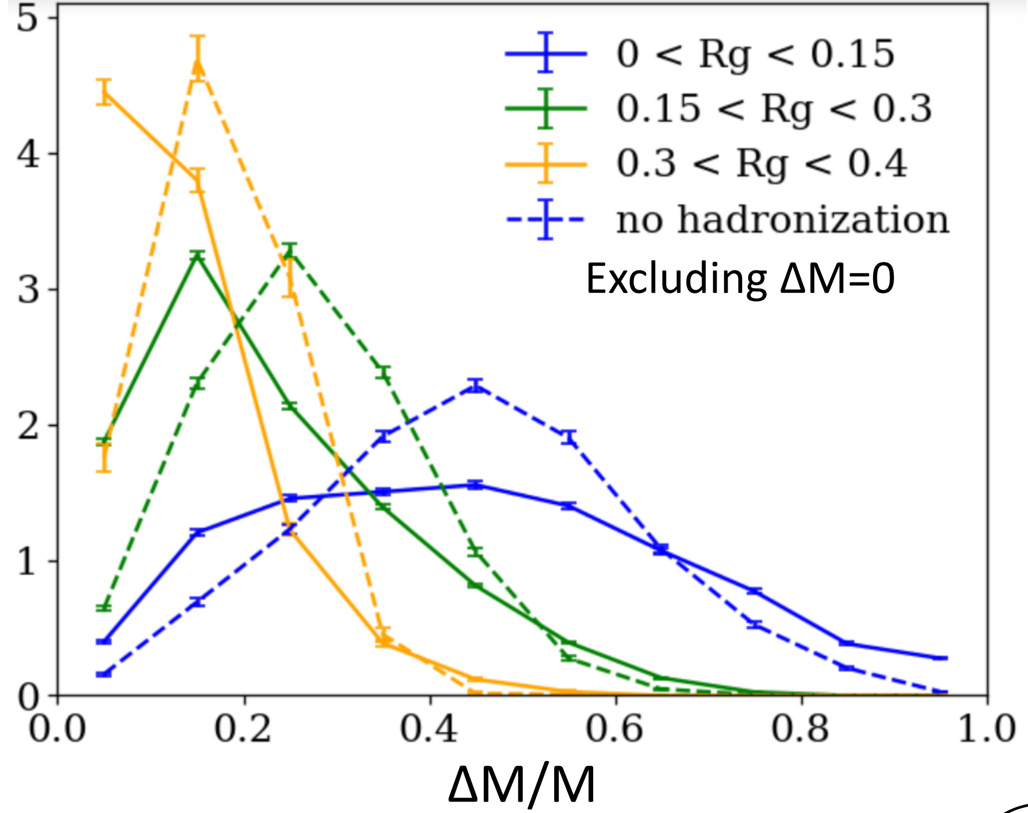


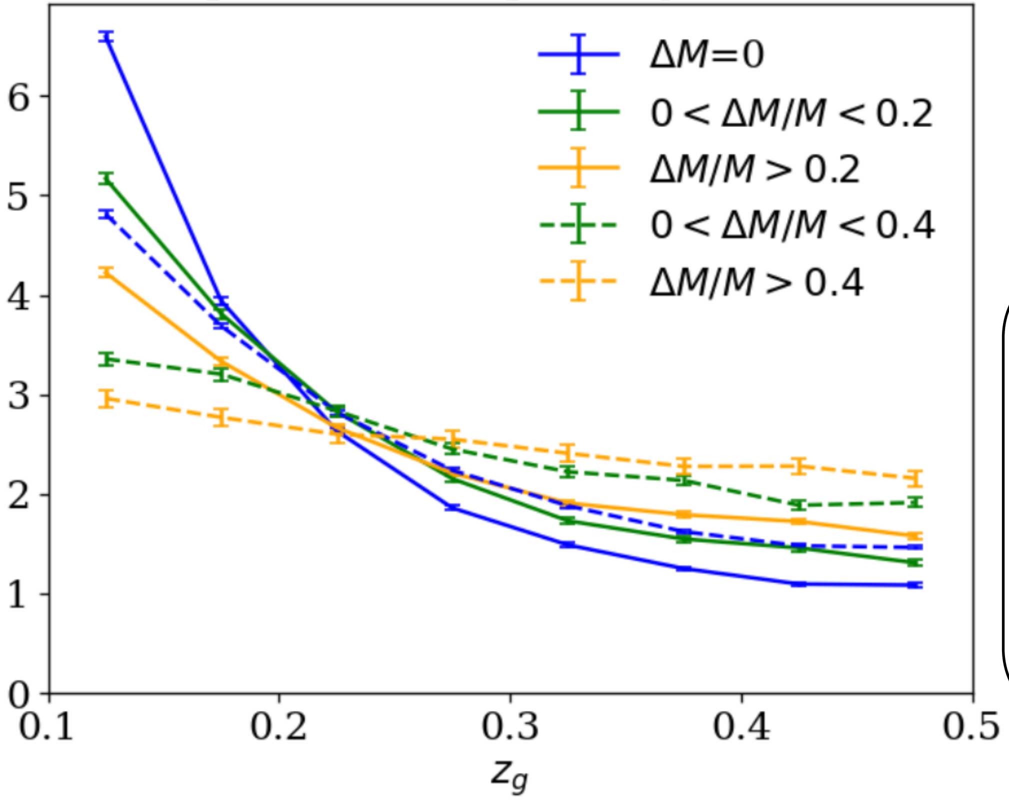


Remove underlying effects and pileup, not necessary for 200 GeV collisions

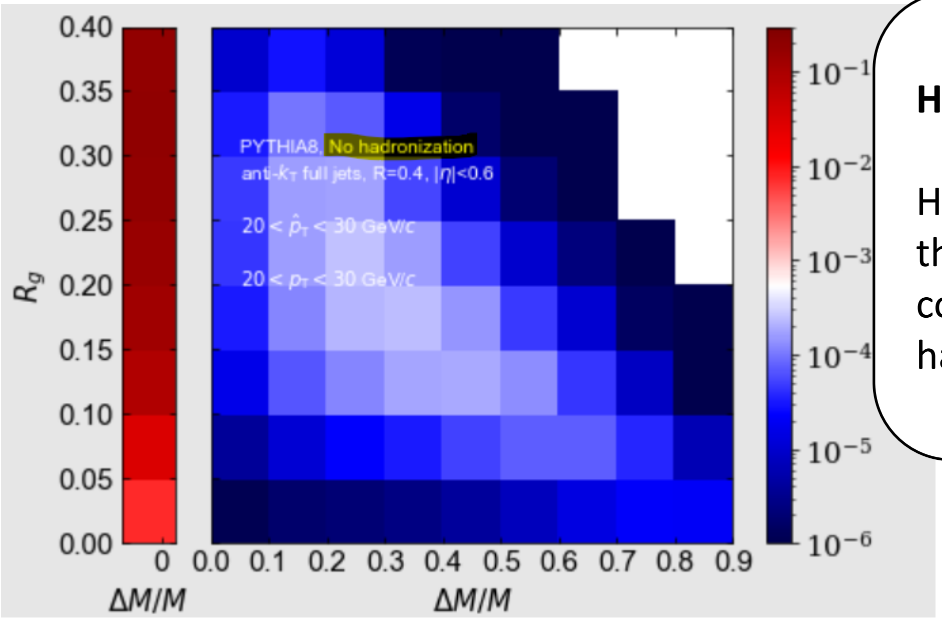
[STAR Collaboration. PRD 101, 052004(2020)]





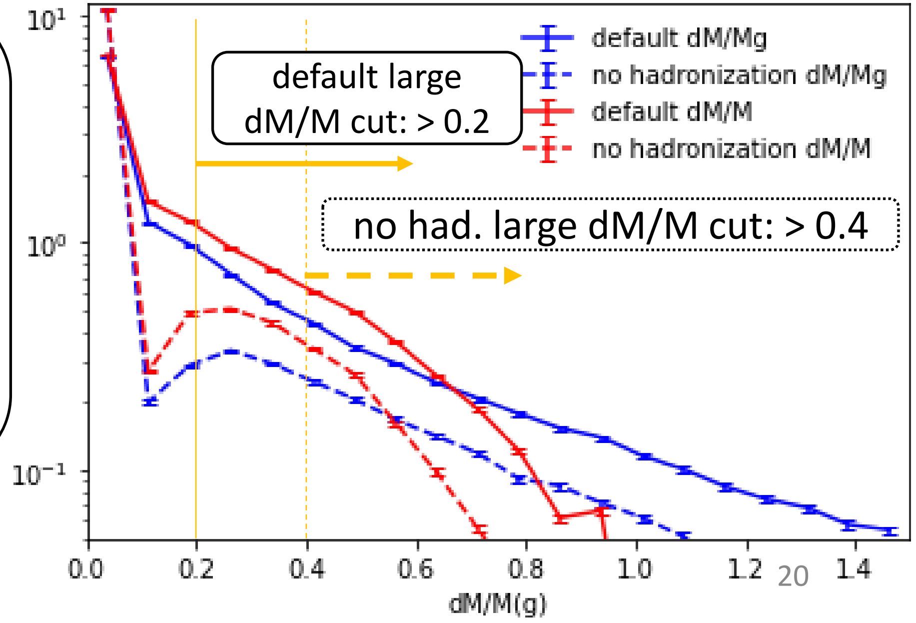


PYTHIA8 Detroit tune
20 < pT,jet < 30 GeV



Hadronization effects study:

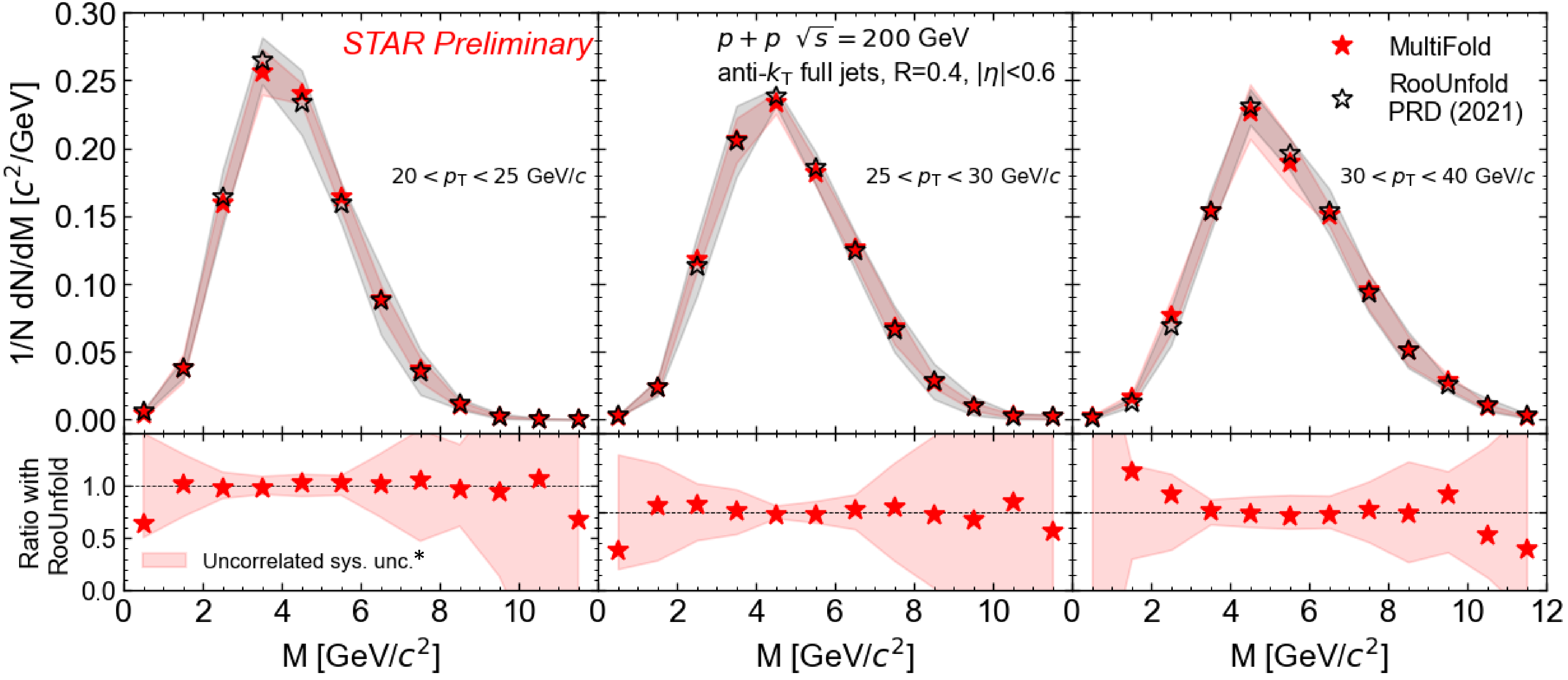
Hadronization smears/shifts the distributions, but the correlation with and without hadronization is the same.



Does MultiFold work on our data?

$$M = |\Sigma_{i\in\text{jet}}\, p_i| = \sqrt{E^2 - |\vec{p}|^2}$$

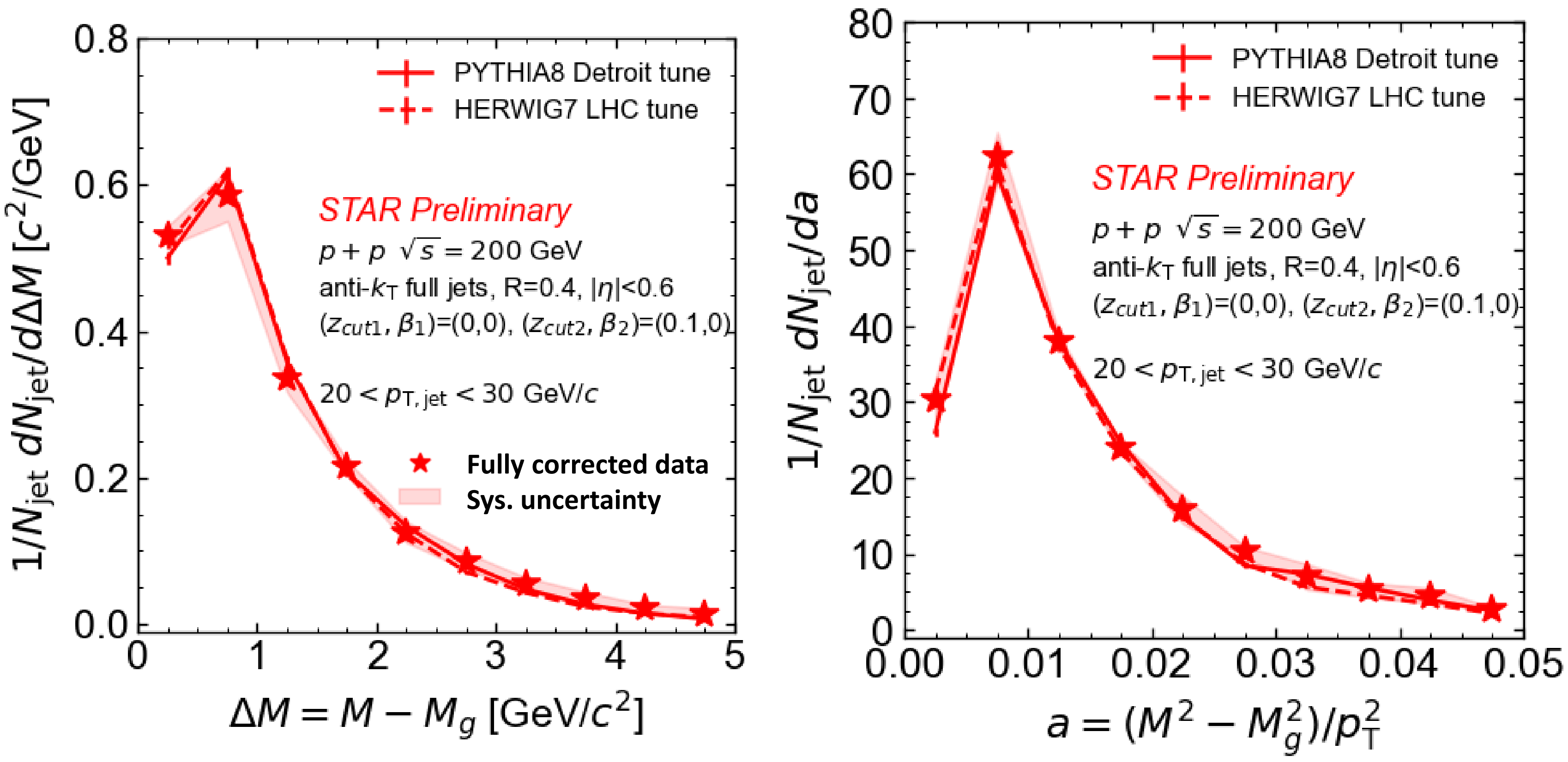
MultiFolded result agrees with **RooUnfolded** result ([STAR Collaboration. PRD 104, 052007(2021)]) [HEPData]



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

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

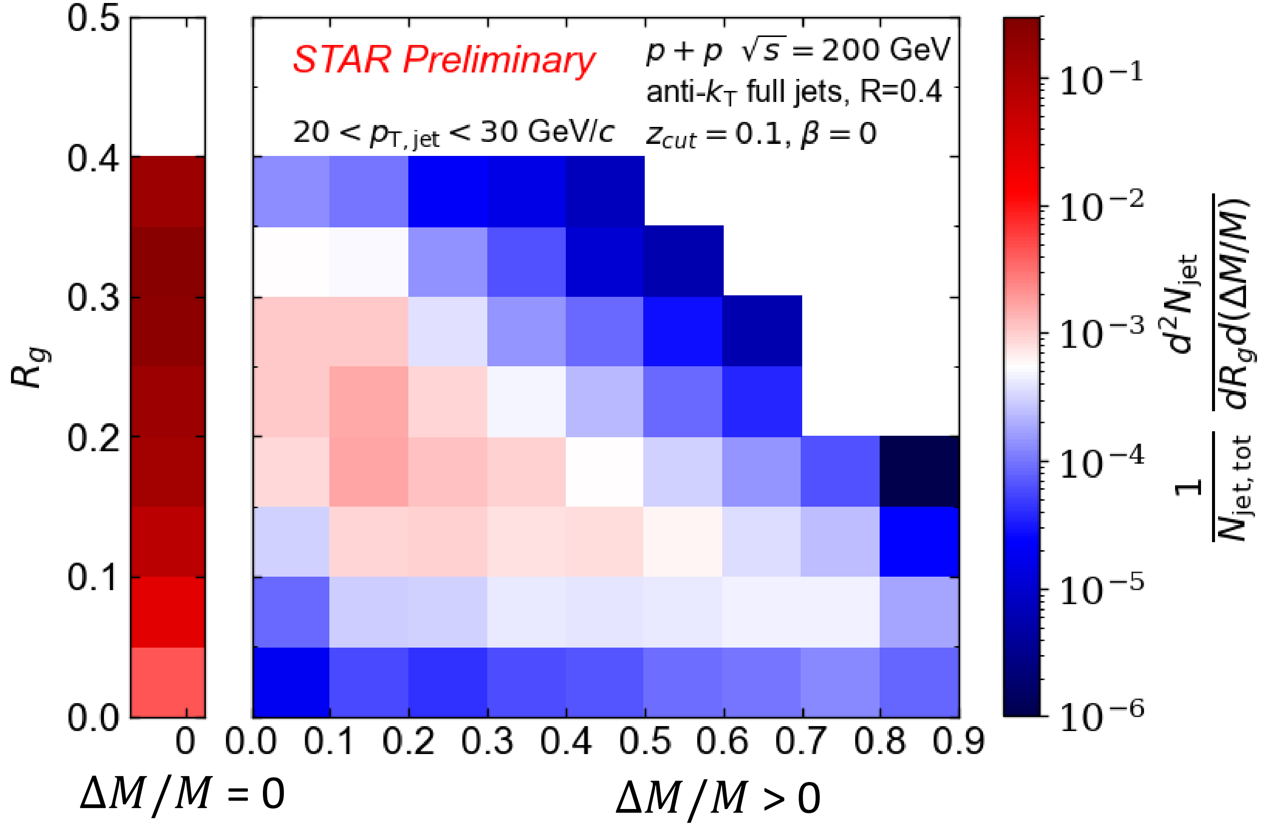
CollinearDrop groomed jet mass



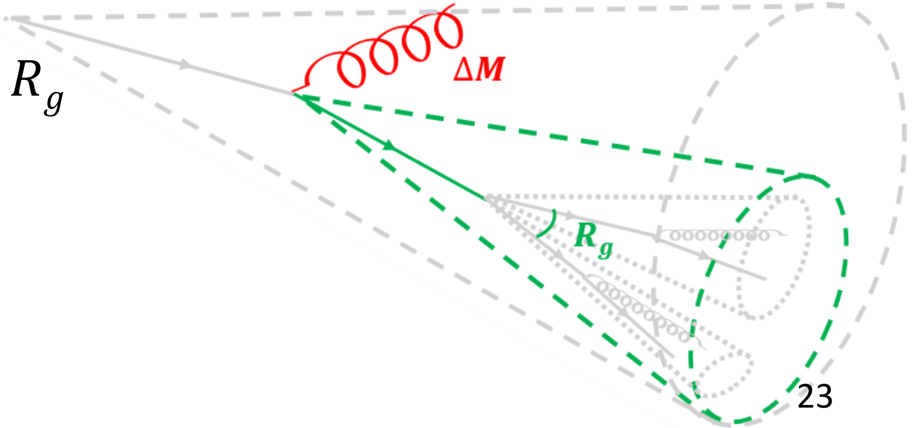
Measurement excludes jets with ΔM = 0

- 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.

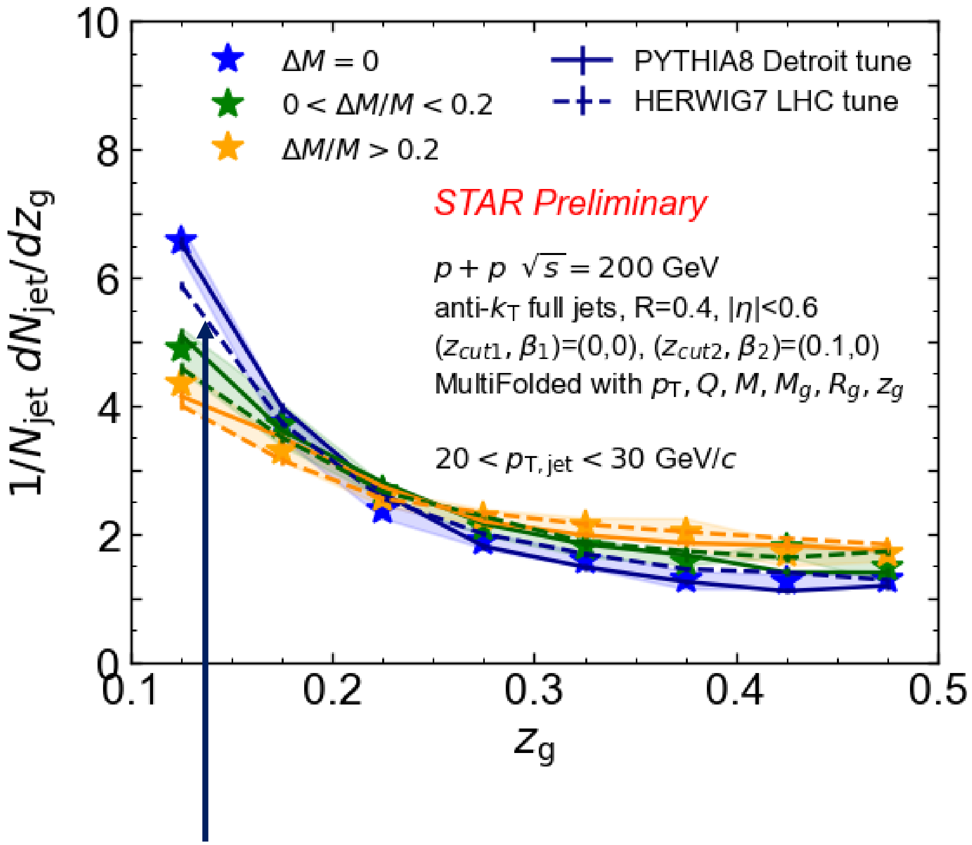
Soft radiation vs hard splitting angle



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



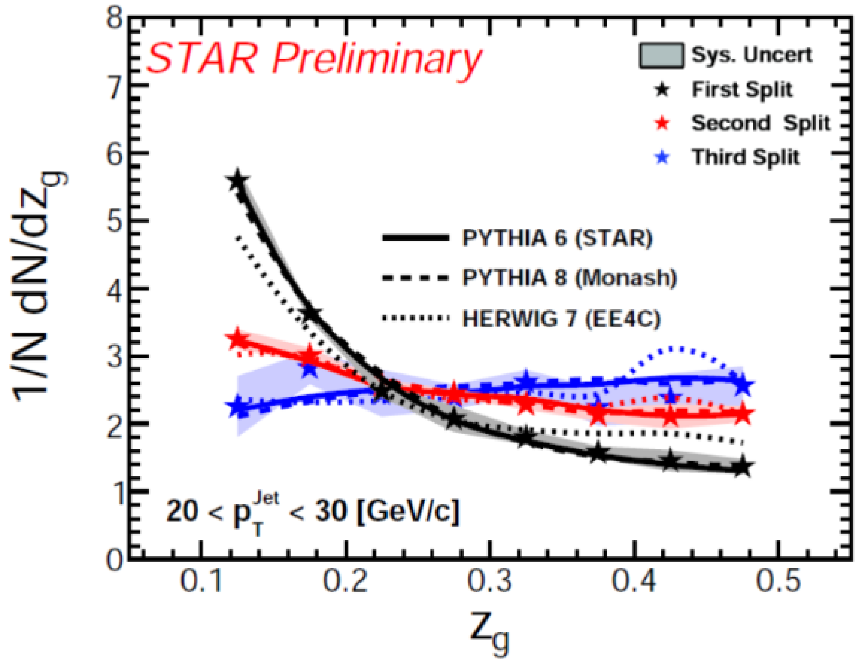
Soft radiation vs hard splitting momentum imbalance



Steeply falling ~ DGLAP 1/z: pQCD

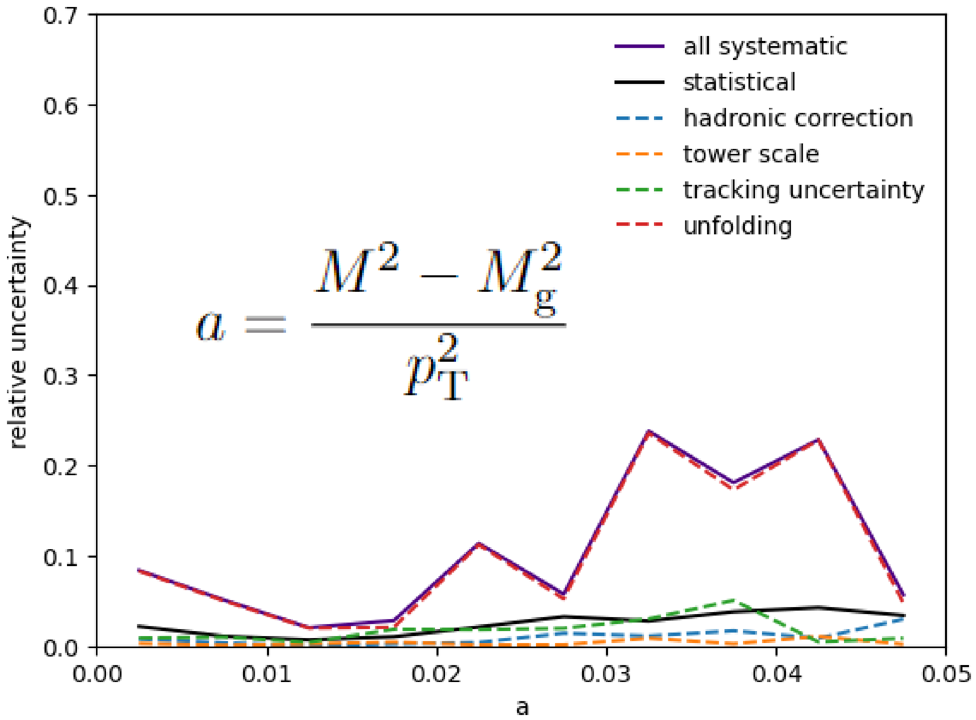
→ The first splitting that passes SoftDrop can still be non-perturbative, but if we apply the $\Delta M = 0$ selection, we can filter out some npQCD contribution due to the parton splitting

- The more mass that is groomed away relative to the original mass, the flatter the Zg 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

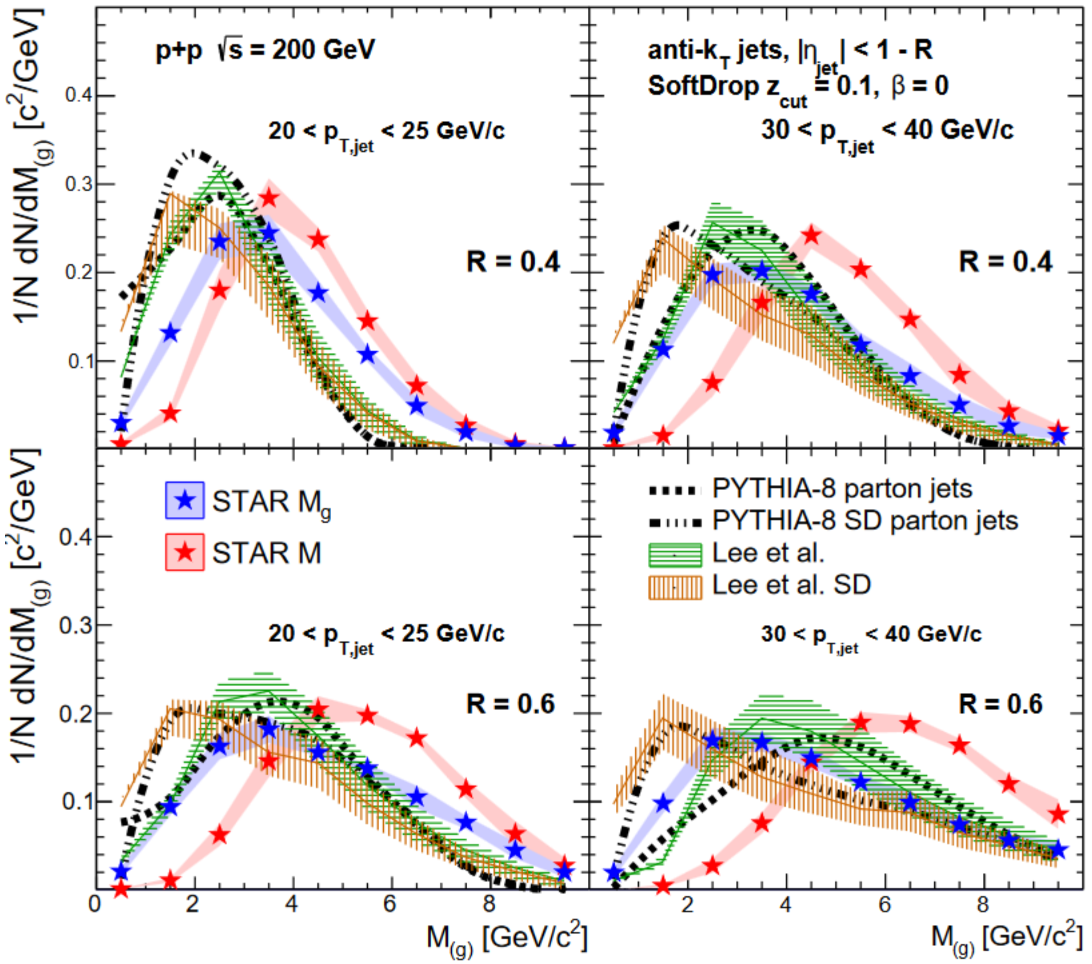


Systematic uncertainties

- Detector systematics
  - Hadronic correction 100% → 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: $\text{prior} \to \text{reweight} \bigotimes \text{nominal prior}$,
      with $$\text{reweight}(p_\mathrm{T}, Q, M, M_\mathrm{g}, R_\mathrm{g}, z_\mathrm{g}) = \frac{\text{Herwig truth}(p_\mathrm{T}, Q, M, M_\mathrm{g}, R_\mathrm{g}, z_\mathrm{g})}{\text{Pythia6 truth}(p_\mathrm{T}, Q, M, M_\mathrm{g}, R_\mathrm{g}, z_\mathrm{g})}$$



Jet mass: Comparison with models and calcaulations



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

Correction for detector effects

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

Example:

jet at truth level	jet at detector level
pi+ 6 GeV	track(+) 6 GeV
pi0 4 GeV	track(-) 3 GeV
pi- 3 GeV	neutral 2 GeV
pi- 1 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)

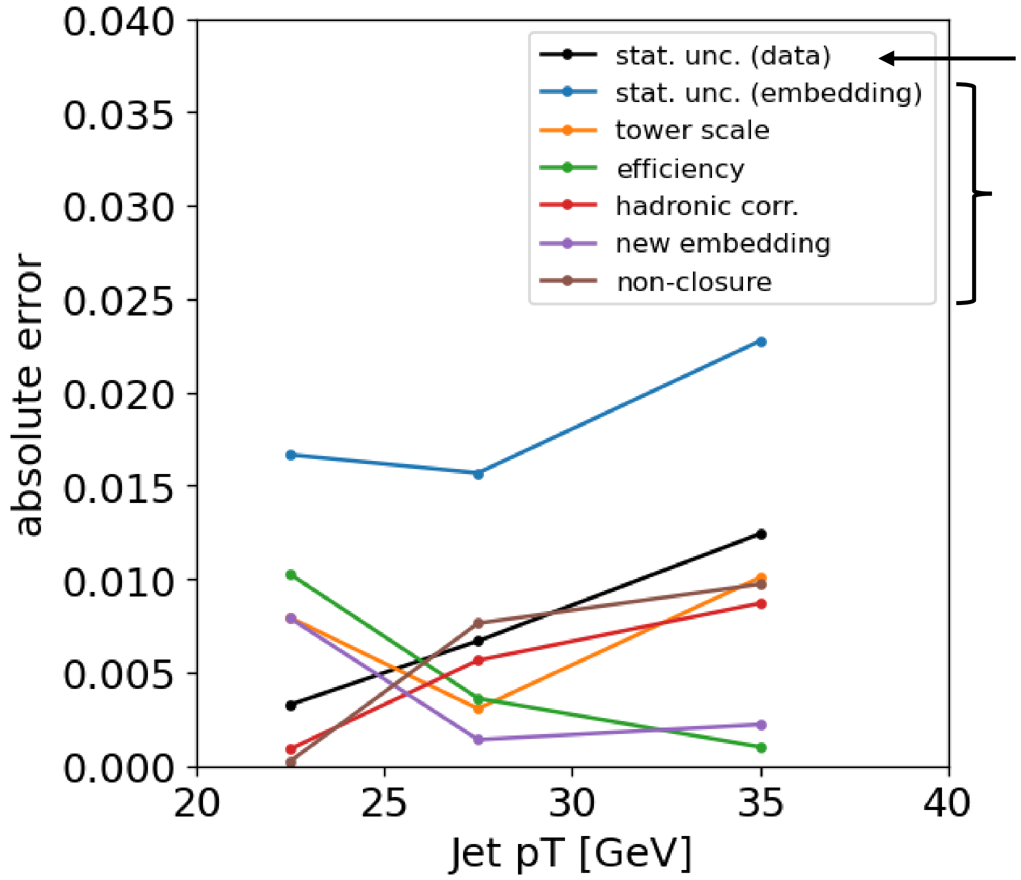
- How should we account for the neutral background?

Solution: Switch to a "charged jet" measurement

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

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

Systematic uncertainty



Statistical uncertainty

Systematic uncertainty