Probing the parton shower with multi-differential jet substructure measurements in pp collisions at STAR

Youqi Song (youqi.song@yale.edu)

for the STAR Collaboration

BOOST 2023, Berkeley, CA

8/2/2023

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

East Bay BOOST 23
31 Jul - 4 Aug Lawrence Berkeley National Lab

Jets and SoftDrop grooming

Jet is a multi-scale object

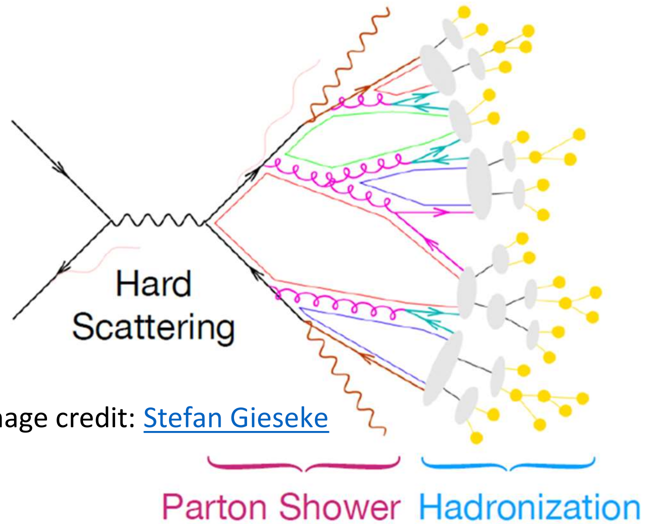


Image credit: Stefan Gieseke

To enhance perturbative contributions, SoftDrop grooming is often used to remove soft wide-angle radiation

Larkoski, et al. JHEP 05 (2014) 146.
Dasgupta et al. JHEP 09 (2012) 29.

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

The Lund Plane

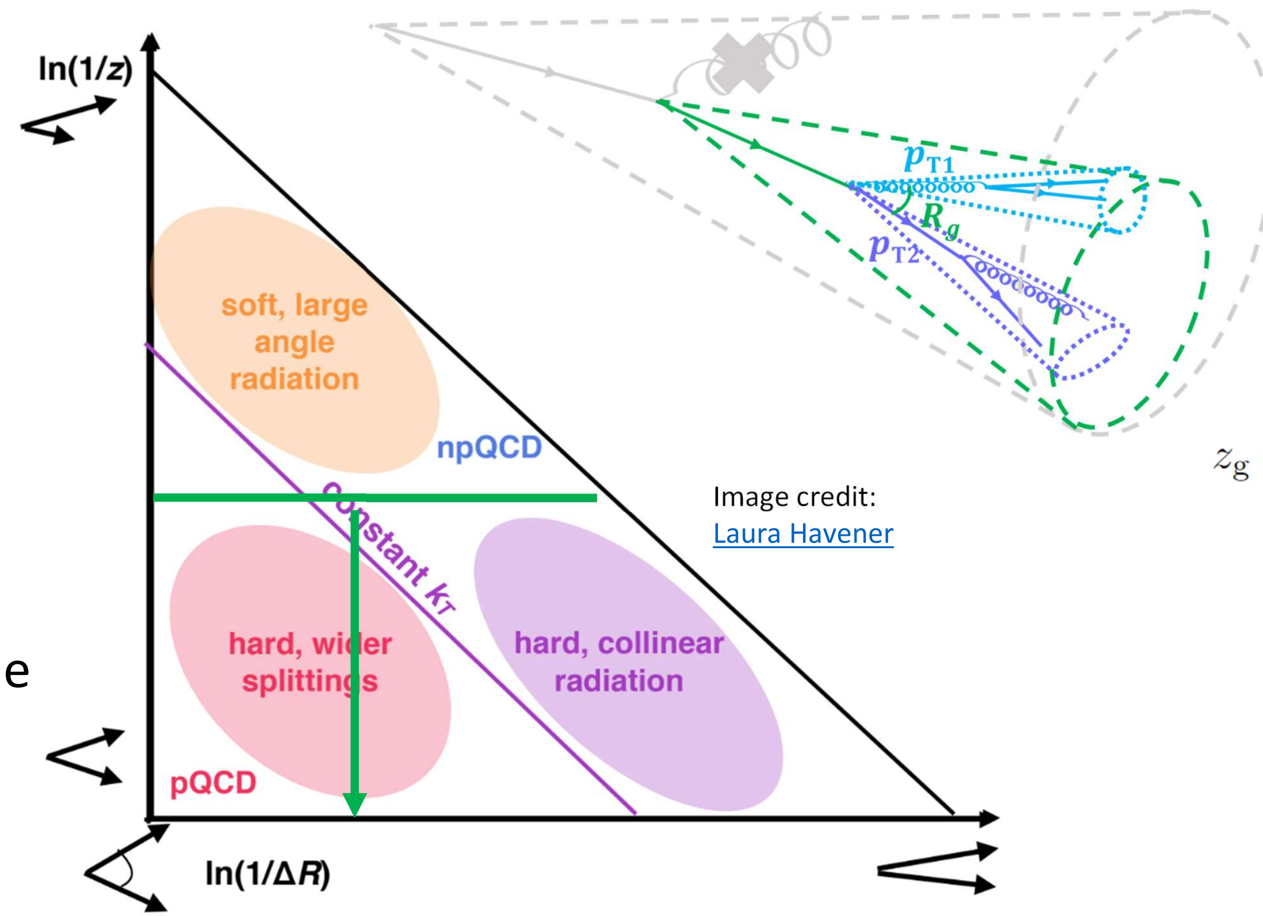
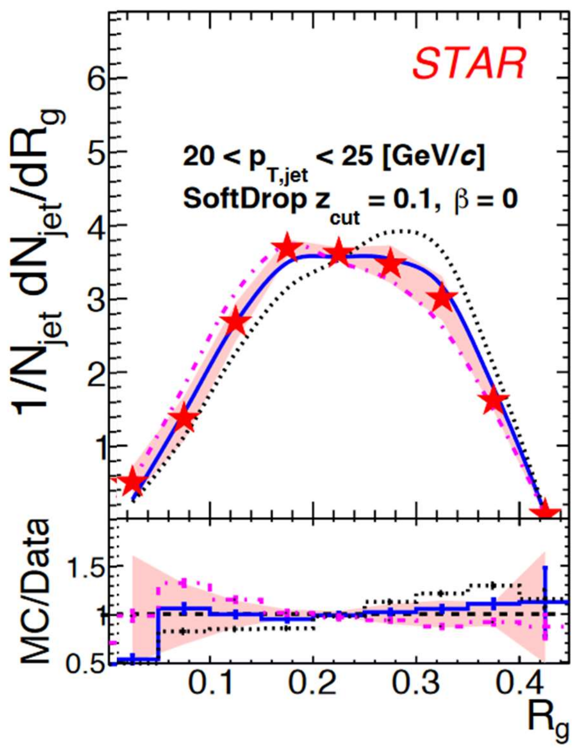


Image credit: Laura Havener

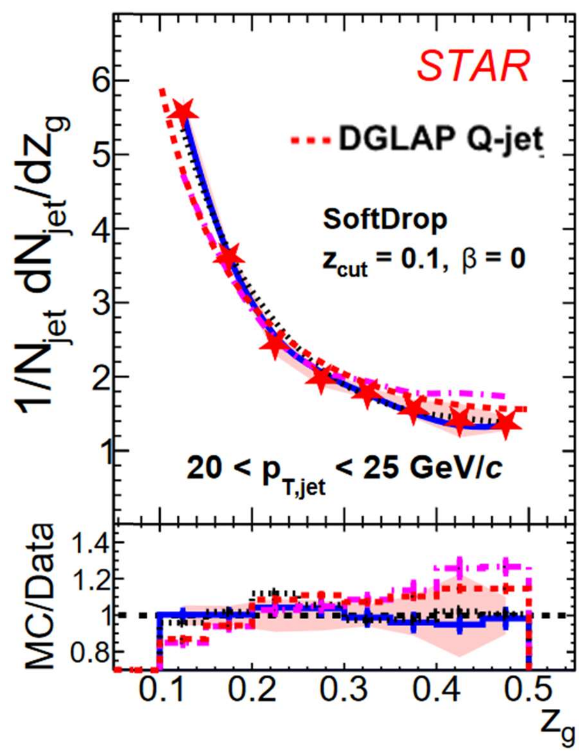
Jets and SoftDrop grooming

- STAR has measured SoftDrop jet substructures

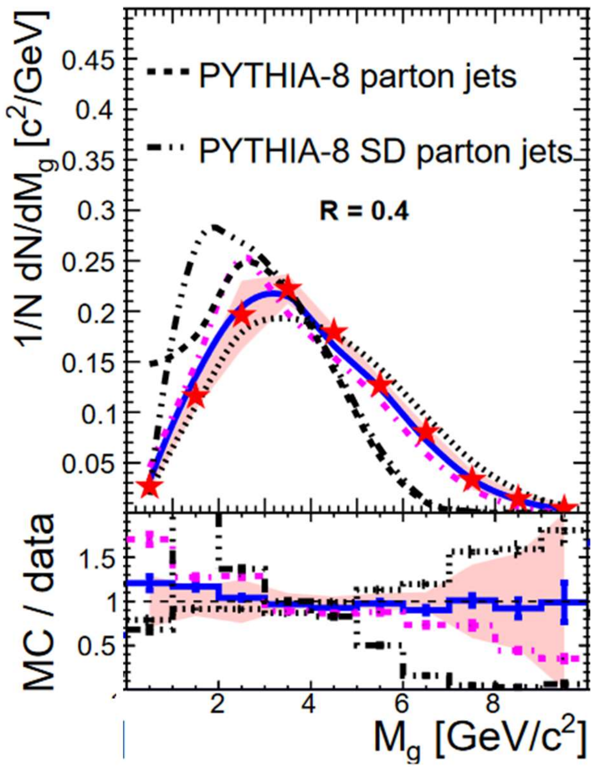
R_g



z_g



M_g



STAR. PLB 811 (2020) 135846, STAR. PRD 104, 052007(2021)

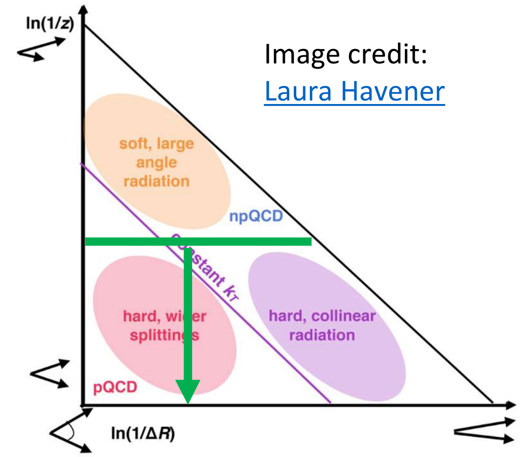
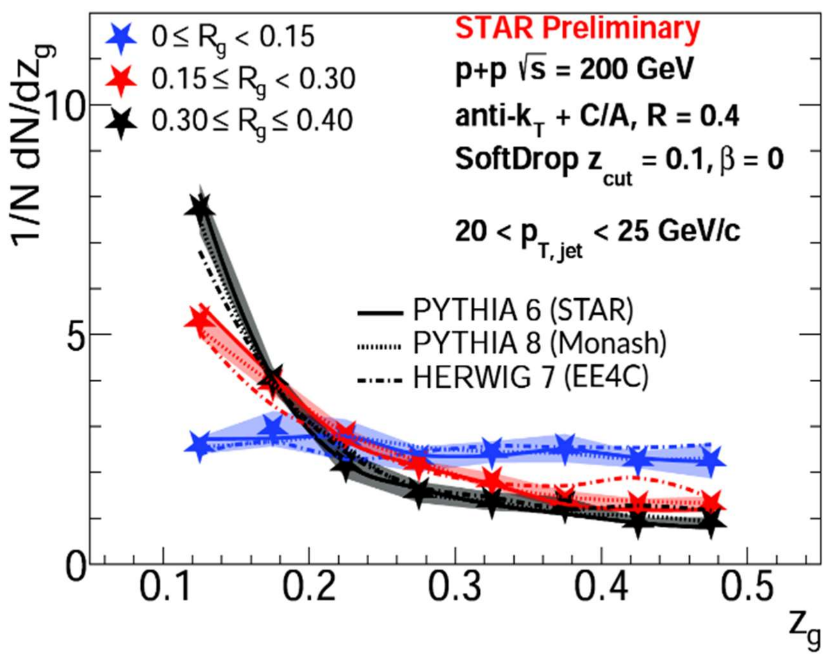


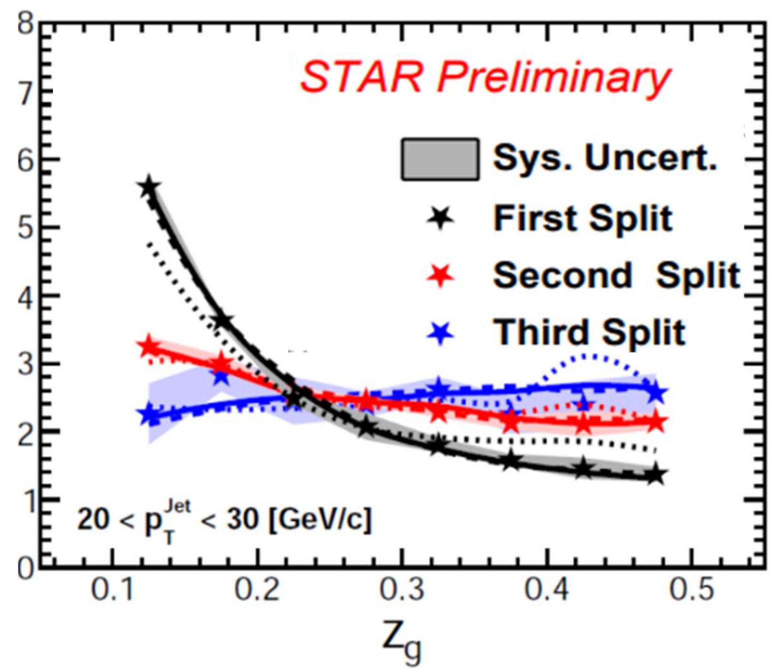
Image credit: Laura Havener

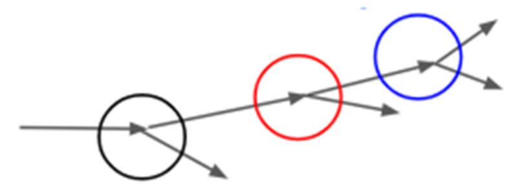
Jets and SoftDrop grooming

- STAR has measured SoftDrop jet substructures



... and in a multi-dimensional fashion!





→ study dynamics of the parton shower!

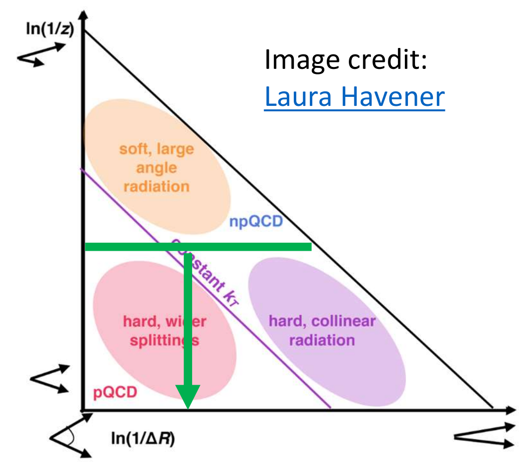


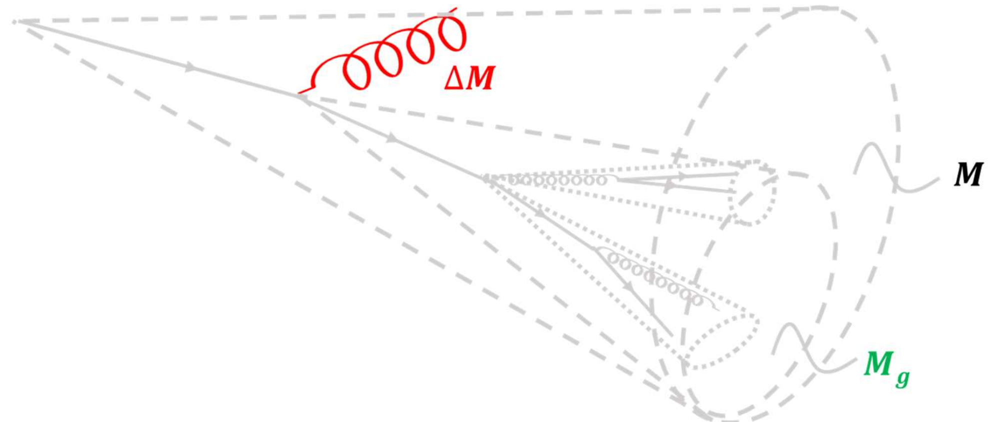
Image credit: Laura Havener

CollinearDrop grooming: probes the soft component

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



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

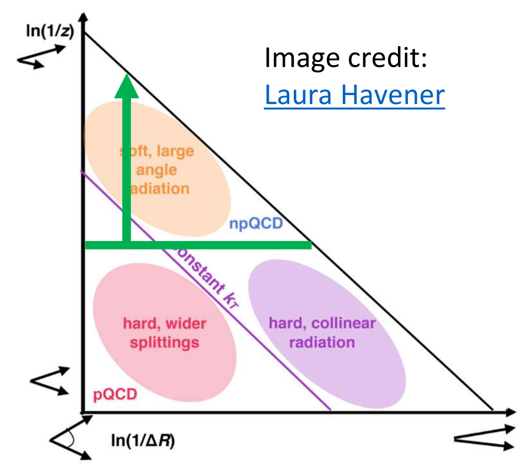


Image credit: Laura Havener

CollinearDrop vs SoftDrop correlation:

probes the soft-hard correlation

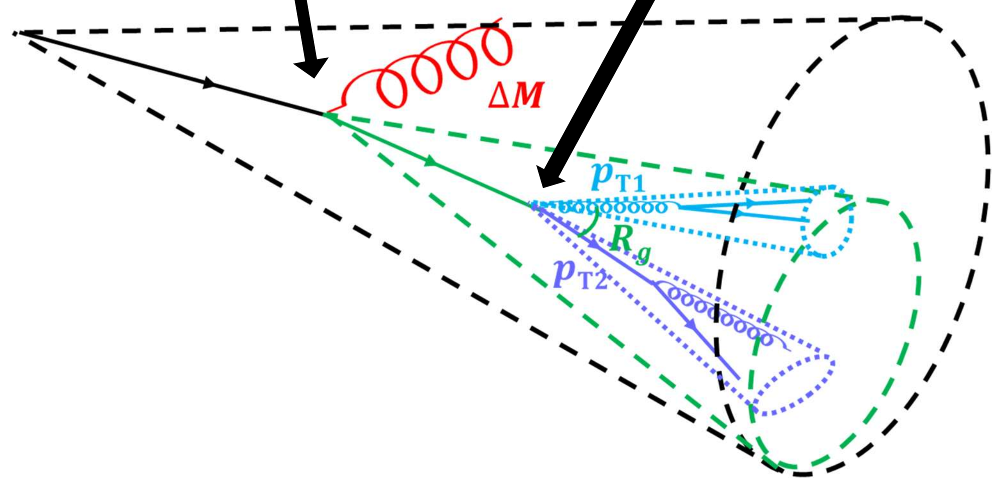
- How does the amount of soft radiation correlate with the angular and momentum scale of a hard splitting? → how an **early** emission affects a **later** splitting

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

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



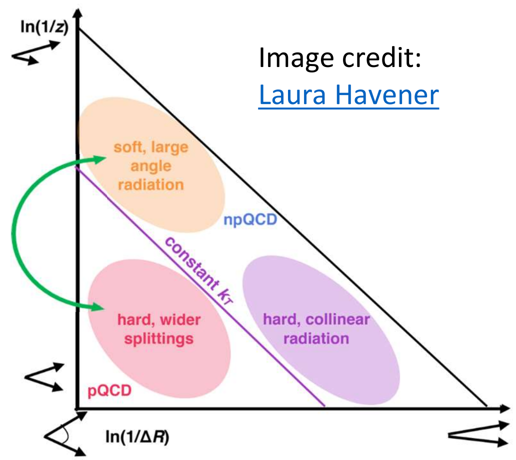


Image credit: Laura Havener

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

Jet reconstruction at STAR

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

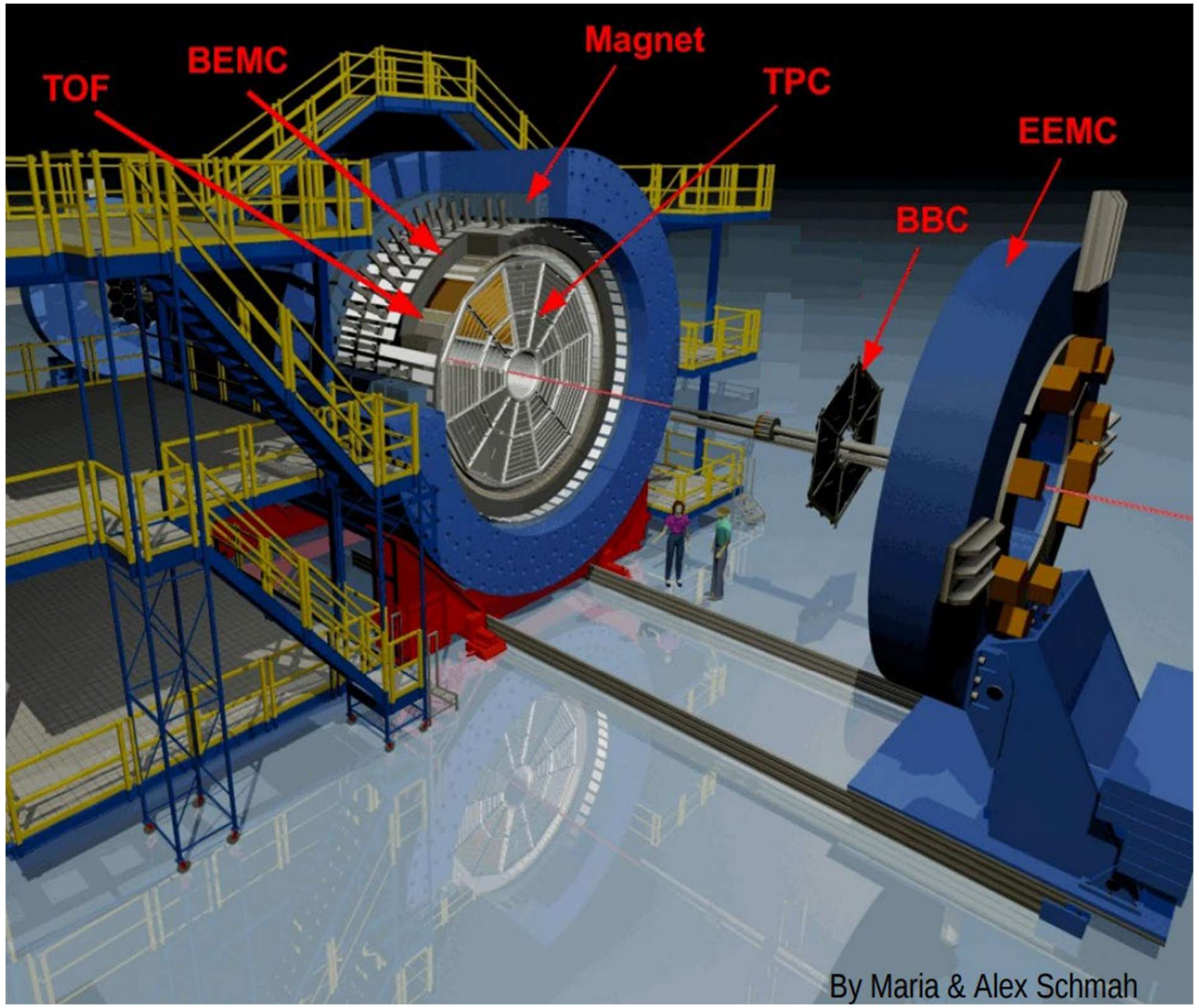
- **TPC** (Time Projection Chamber)
  - For **charged** particle track reconstruction
  - $|\eta| < 1$, full azimuthal coverage

- **BEMC** (Barrel ElectroMagnetic Calorimeter)
  - For **neutral** energy measurement and triggering
  - $|\eta| < 1$, full azimuthal coverage

➢ Reconstruct anti-k_T **full jets**
  - Jet resolution parameter **R=0.4**
  - $|\eta_{jet}| < 0.6$

Additional selections

- Tracks (Towers): $0.2 < p_{\mathrm{T}}(E_{\mathrm{T}}) < 30$ GeV/c
- Jets
  - $p_{\mathrm{T}} > 15$ GeV/c, $M > 1$ GeV/c^2
  - Passes SoftDrop with $z_{cut} = 0.1$ and $\beta = 0$



Unfolding method

- Jet measurements need to be corrected for detector effects for comparison with theory/model
- 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
  - $Q^{\kappa} = \frac{1}{(p_{\mathrm{T}\,\mathrm{jet}})^{\kappa}} \sum_{i \in \mathrm{jet}} q_i \cdot (p_{\mathrm{T}i})^{\kappa}$
  - $M = |\Sigma_{i \in \mathrm{jet}}\, p_i| = \sqrt{E^2 - |\vec{p}|^2}$
  - 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

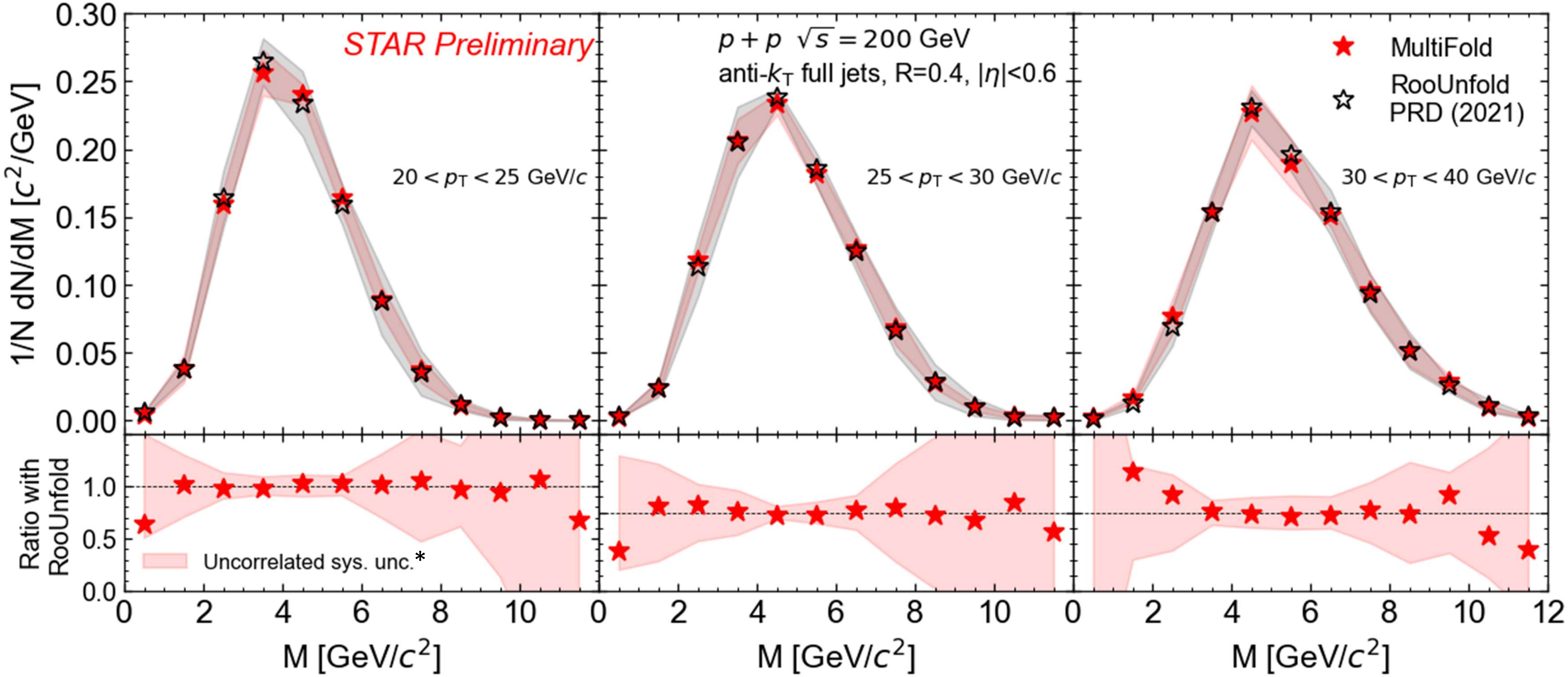
All 6 observables are simultaneously unfolded in an unbinned way!

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

Does MultiFold work on our data?

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

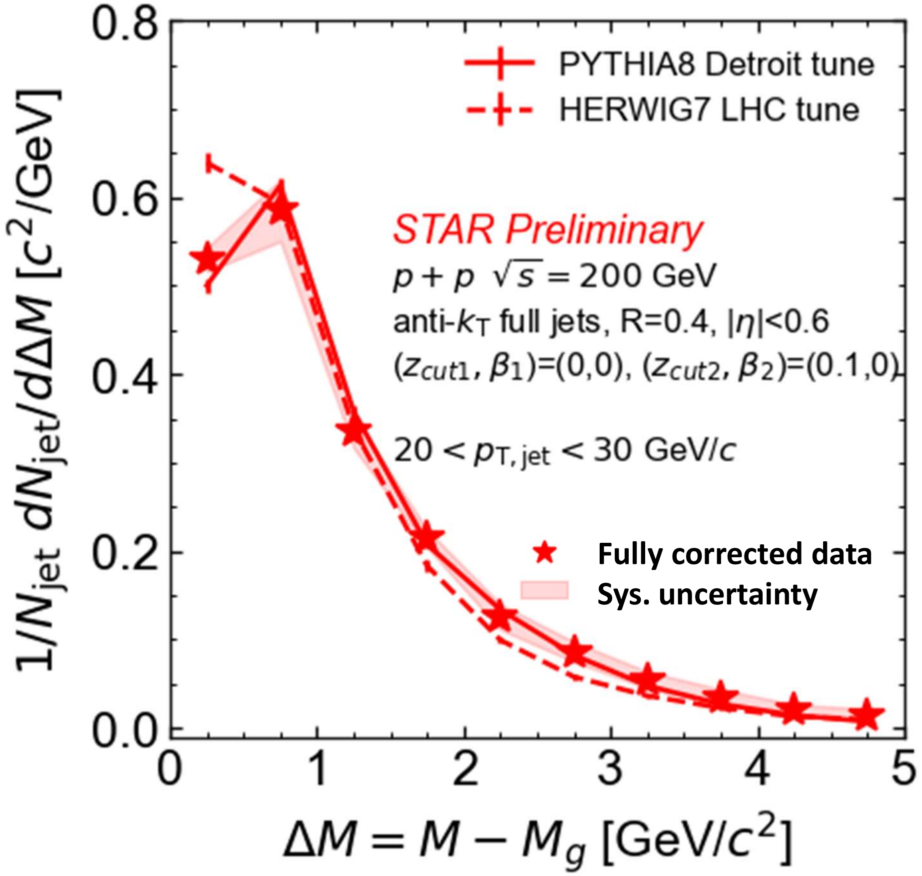
MultiFolded result agrees with **RooUnfolded** result (STAR Collaboration. PRD 104, 052007(2021)) HEPData

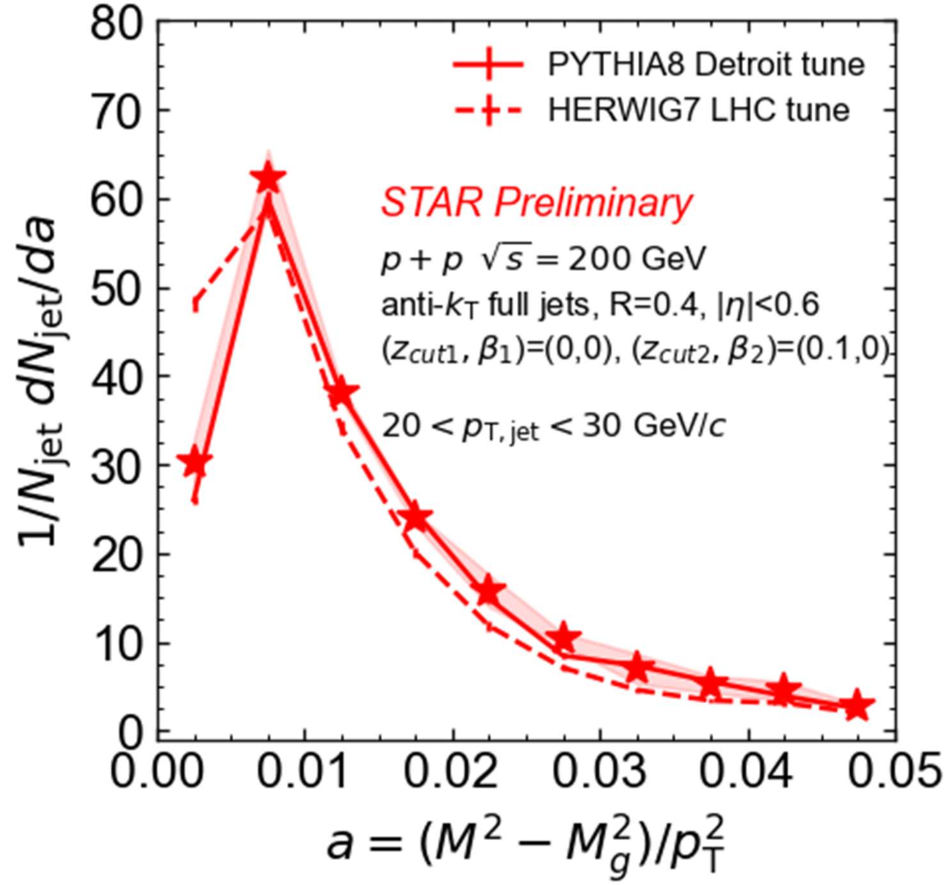


… but **MultiFold** also gives us high-dimensional correlation between observables!

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

CollinearDrop groomed jet mass



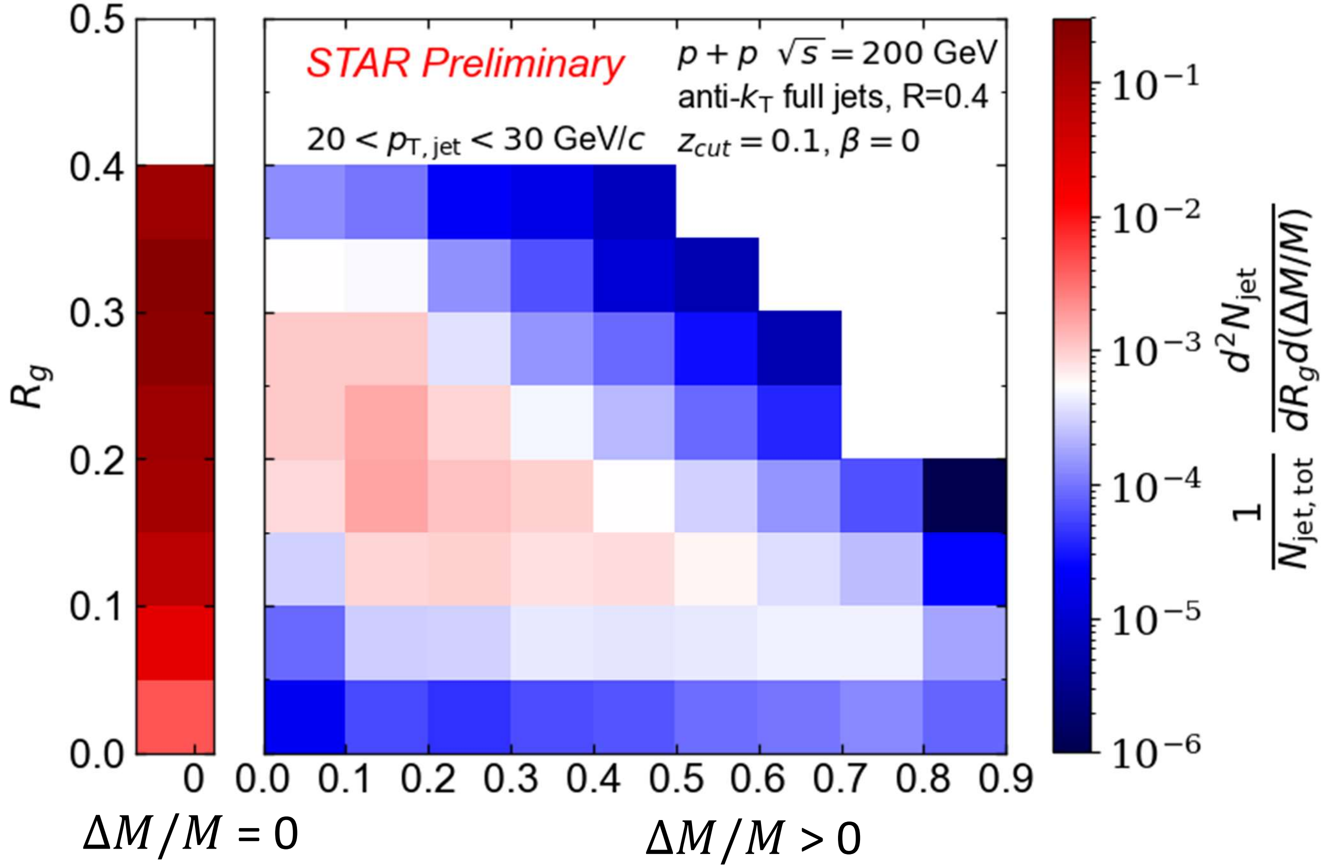


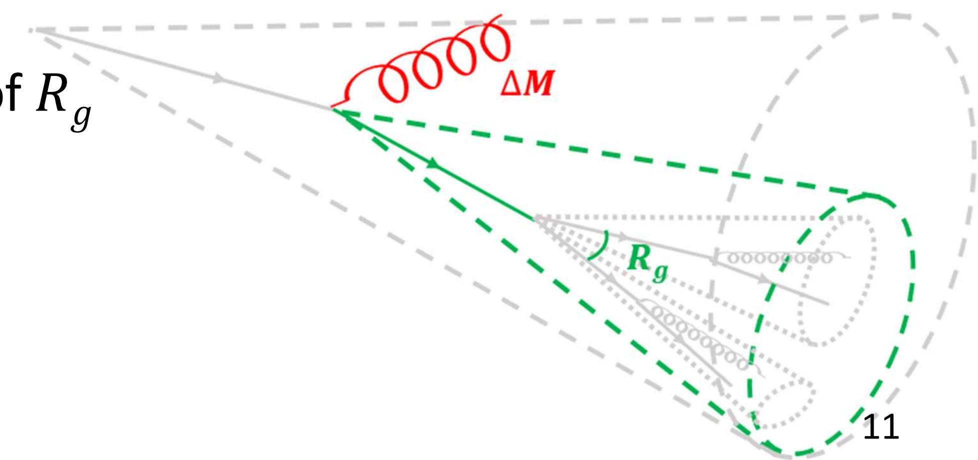
PYTHIA6 Perugia + STAR tune: Skands, PRD 82, 074018 (2010)
J. K. Adkins, PhD thesis (Kentucky U., 2015)
PYTHIA8 Detroit tune: Aguilar et al. PRD 105, 016011(2022)
HERWIG7: Bellm, et al. PRC 76, 1-8 (2016)

Measurement excludes jets with ΔM = 0 (45.5% of jets in this jet p_T range)

- First CollinearDrop groomed jet measurement, sensitive to soft radiation within jets
- MC predictions qualitatively consistent with data; some tension from HERWIG7 in small ΔM region
- MultiFold allows us to correlate (combinations of) unfolded quantities

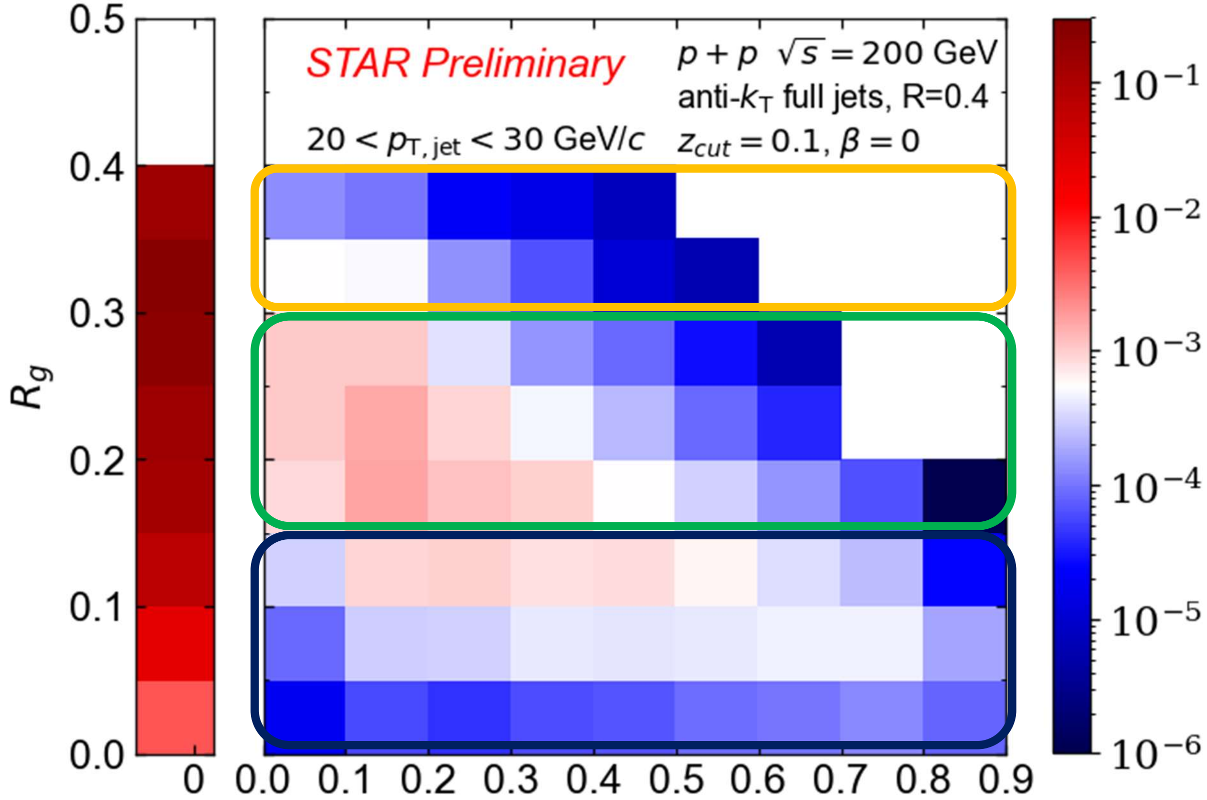
Soft radiation vs hard splitting angle



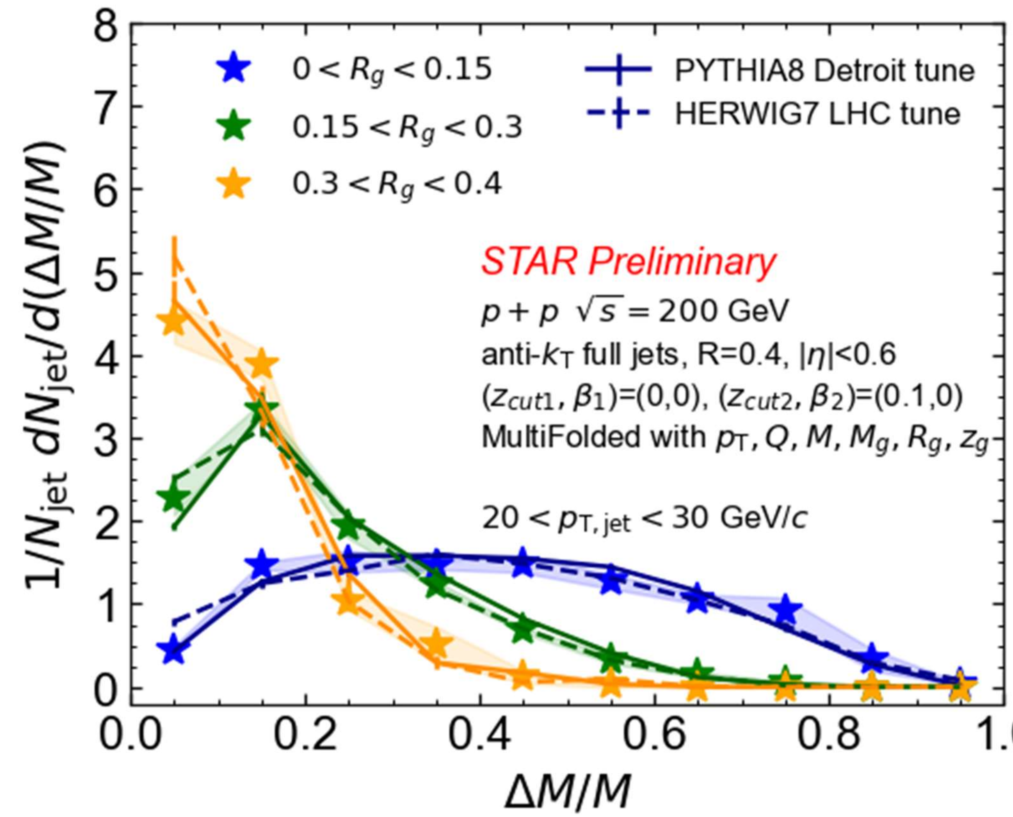


- The mean of $\Delta M/M$ distribution is anti-correlated with mean of R_g → consistent with angular ordered parton showers

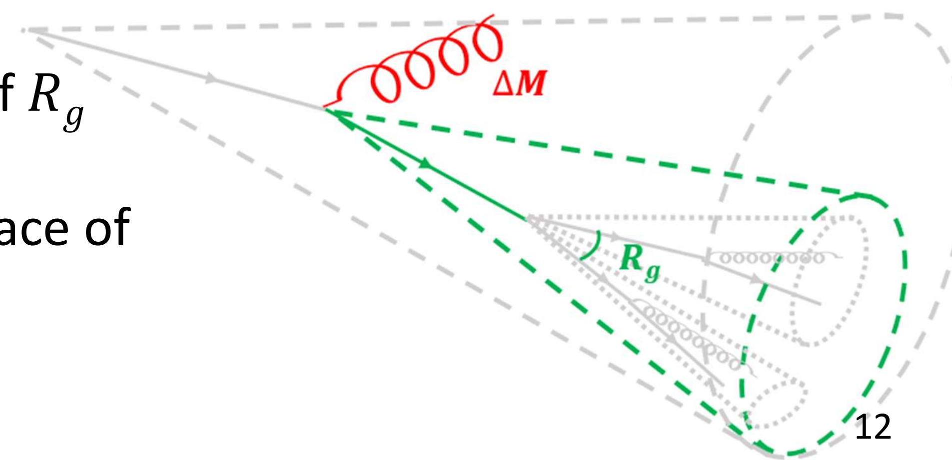
Soft radiation vs hard splitting angle



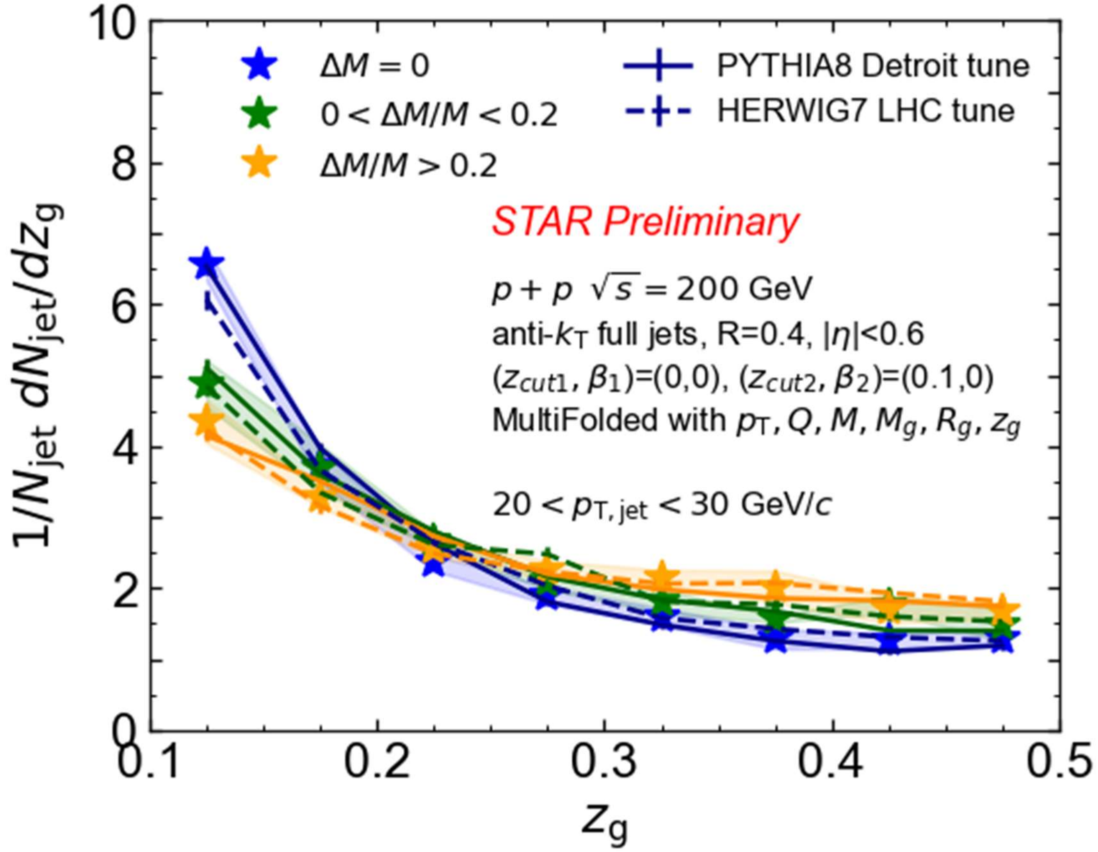
$\Delta M/M = 0$ $\Delta M/M > 0$



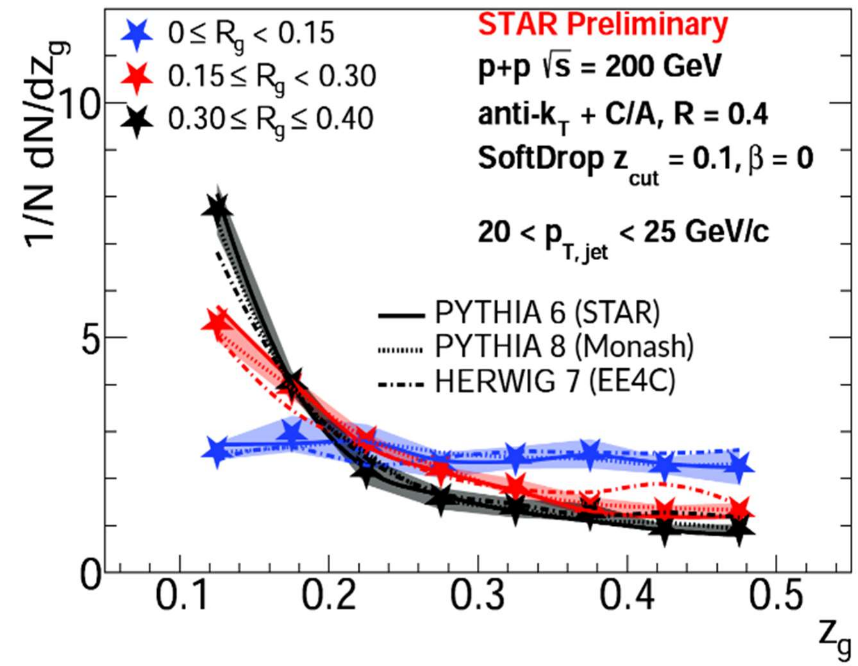
- The mean of $\Delta M/M$ distribution is <u>anti-correlated</u> with mean of R_g
  → consistent with angular ordered parton showers
- **Early** soft wide-angle radiation constrains the angular phase space of **later** splittings
- MC models describe the trend of data



Soft radiation vs **hard** splitting momentum imbalance



- 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 **later** splittings
- MC models describe the trend of data

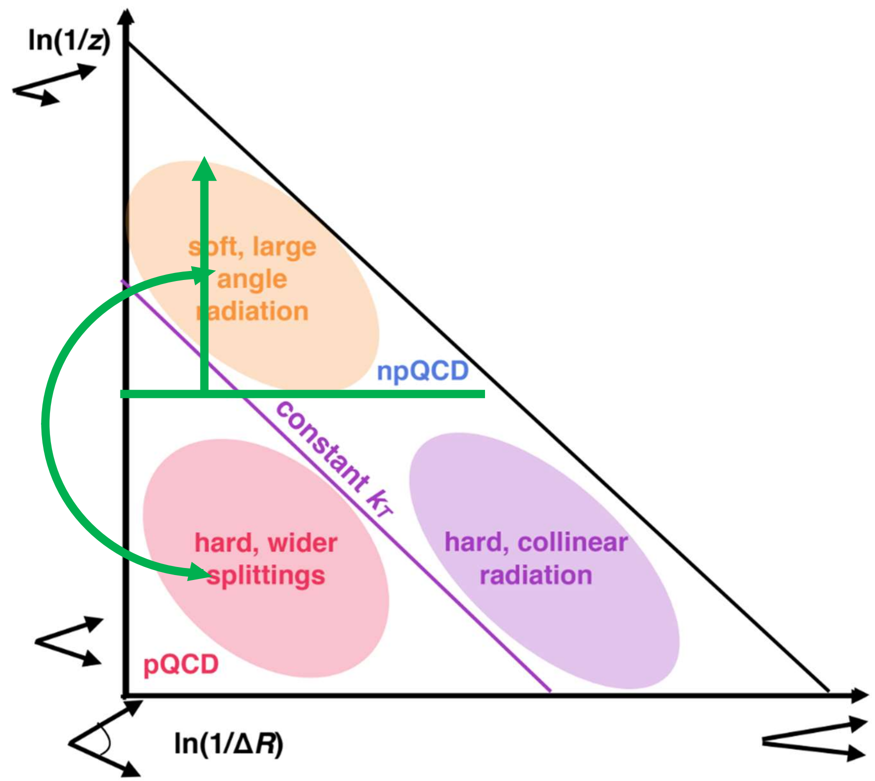


R_g and z_g are correlated, $\Delta M/M$ affects R_g
→ correlation between $\Delta M/M$ and z_g

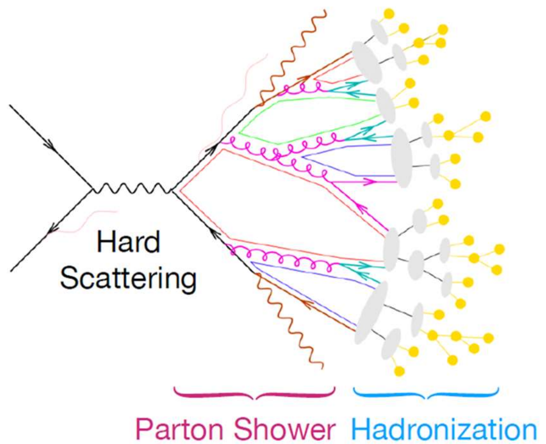
Summary

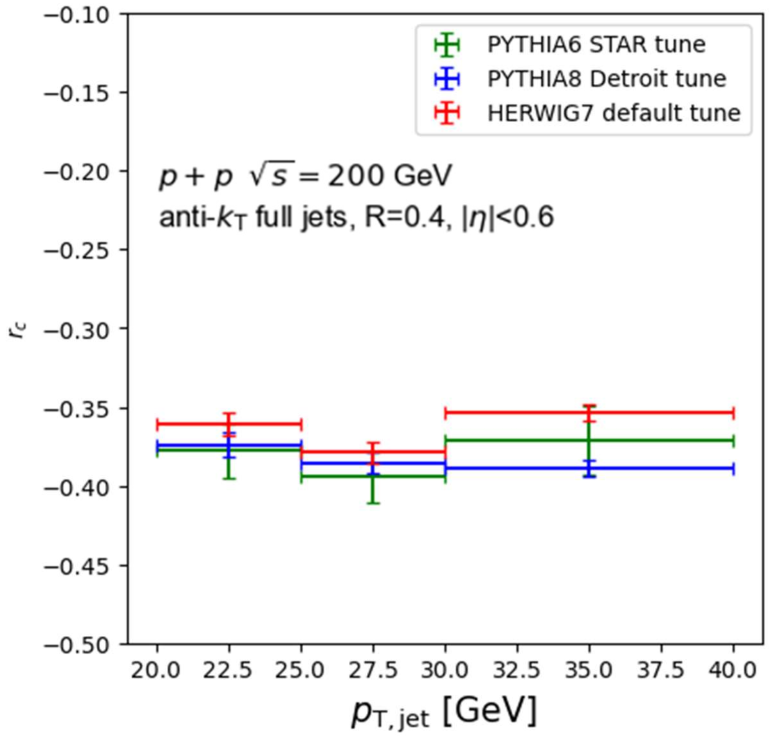
- Probing **soft** wide-angle radiation within jets
  - Fully corrected CollinearDrop jet measurement is presented
- Probing **soft**-**hard** correlation within jets
  - MultiFold allows for access of multi-dimensional correlations on a jet-by-jet basis. First application to RHIC data!
  - Jets with a more asymmetric splitting are more likely to have small early-stage radiation
  - Anti-correlation between the amount of early-stage radiation and the angular scale of a later-stage splitting is observed

Improve understanding of jet substructure and the correlations between different substructure observables!



What’s next?





- Study hadronization with jet substructure by measuring r_c

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

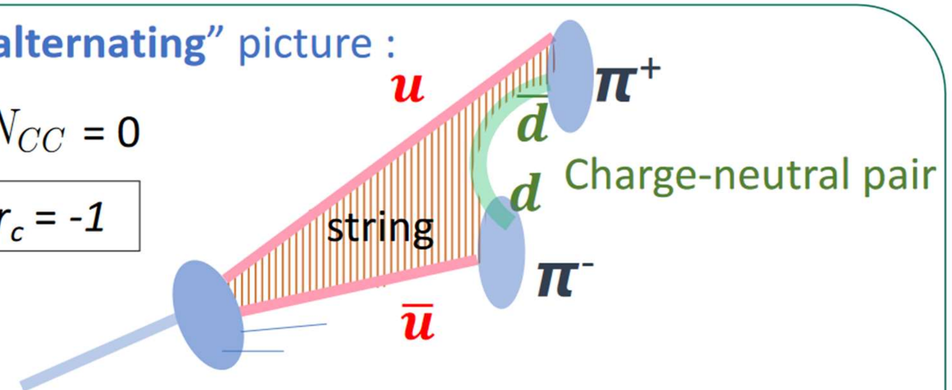
Chien et al. PRD 105 051502 (2022)

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

“**alternating**” picture :

N_{CC} = 0

r_c = -1



Partonic final state : u and $\bar{u}$

Combine charge-neutral pair : $\bar{d}$ and d

“**random**” picture :

no charge correlation

$N_{CC} = N_{C\bar{C}}$

r_c = 0

r_c is a measure of the fraction of “string-like hadronization”

Mondal DIS 2022

- To be validated with STAR data!

What's next?

- Correct for detector effects for r_c?

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?
- Is a "charged jet" measurement meaningful?

- Discussions are welcomed!

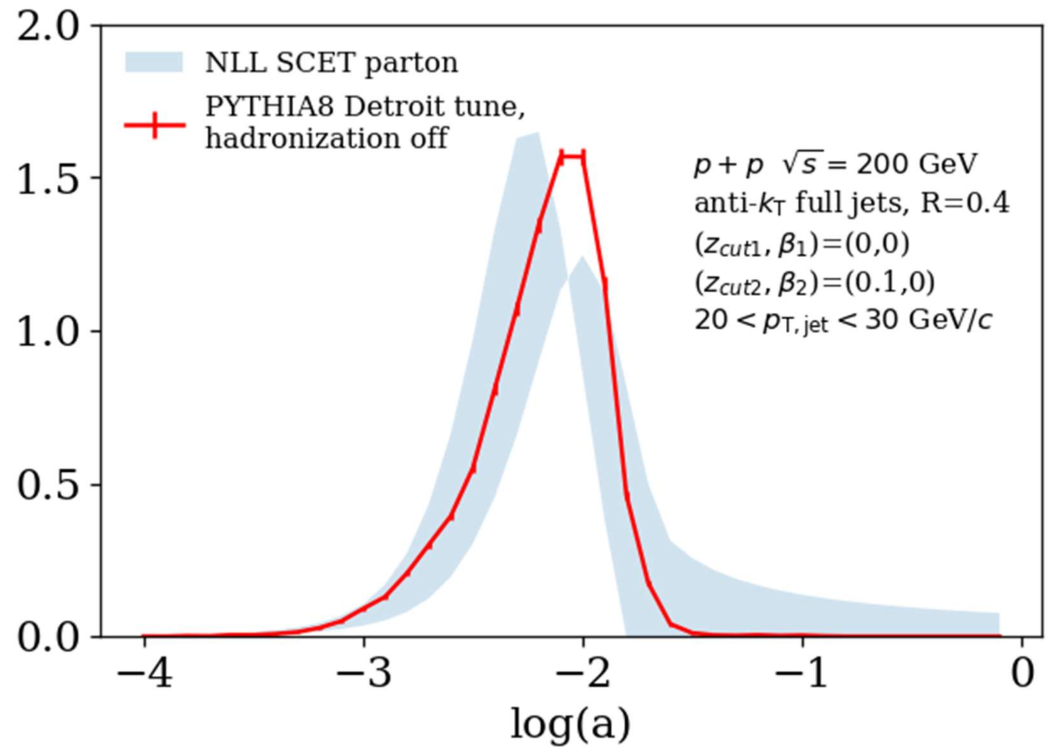
Jets at RHIC are great for the study of parton shower

... and have an enhanced sensitivity for non-perturbative effects like hadronization!

Backup

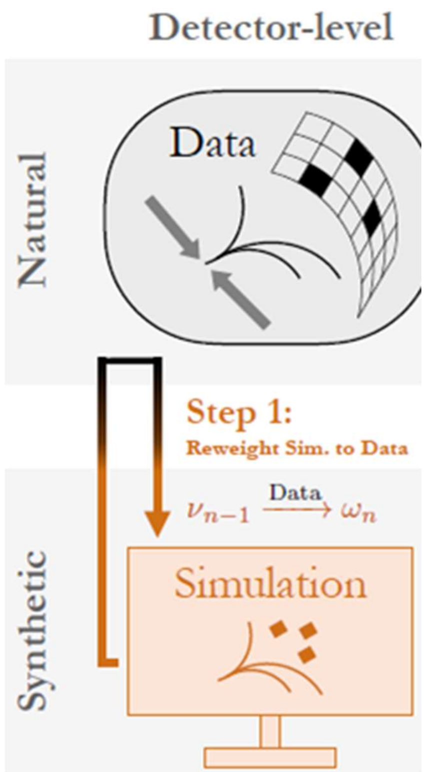
CollinearDrop groomed jet mass

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

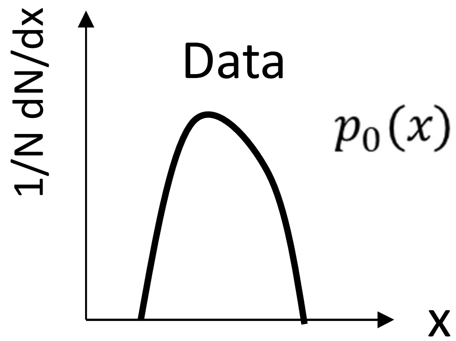


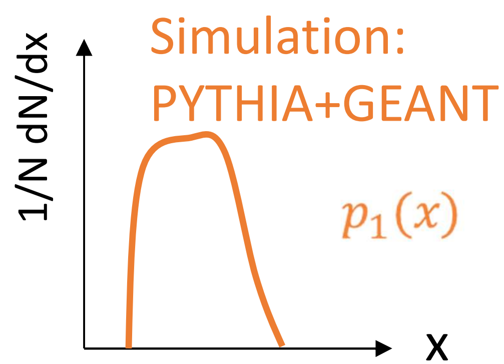
$$a = (M^2 - M_g^2)/p_T^2$$

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

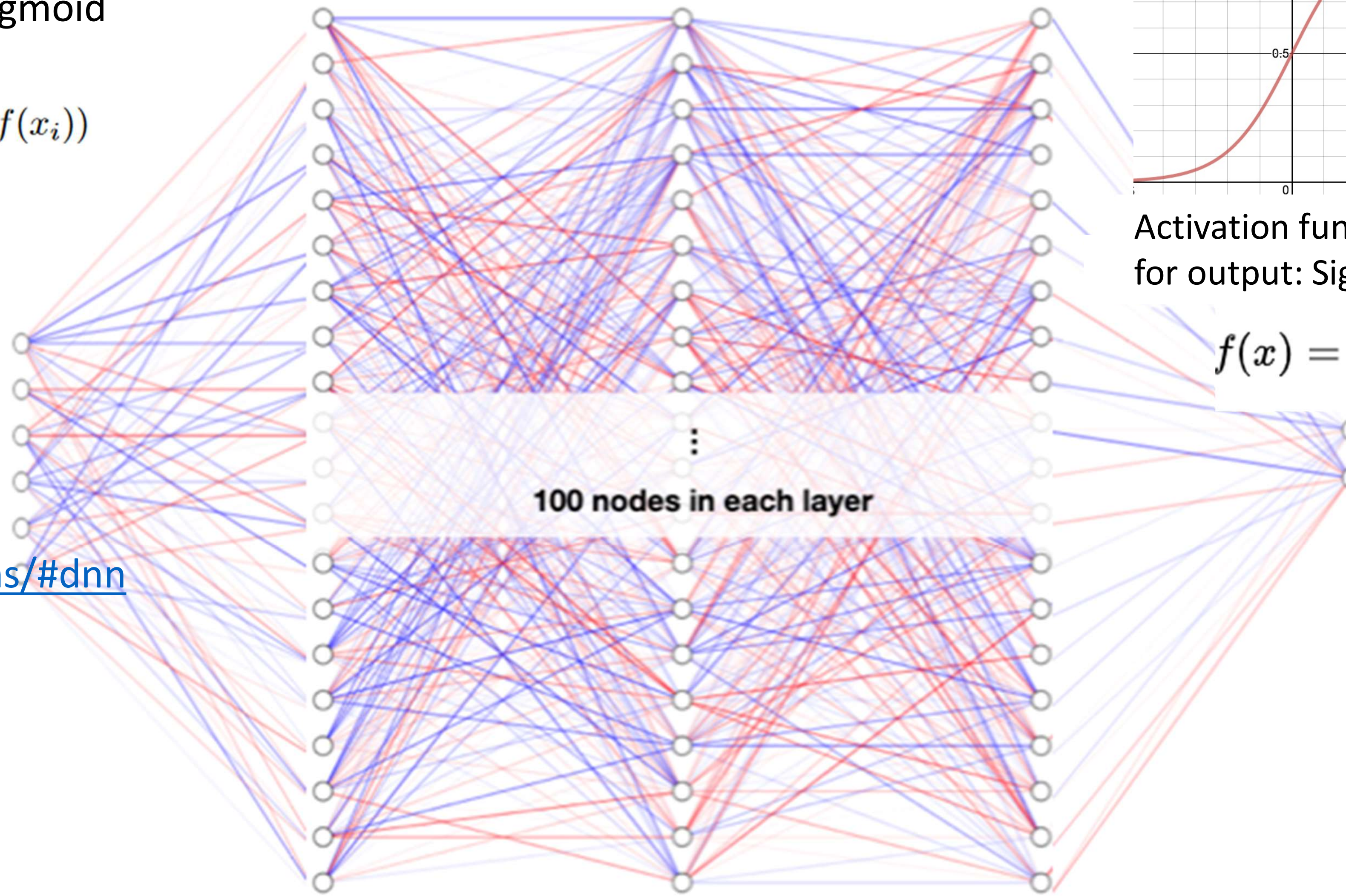
- Architechture: Dense neural network Activation function for dense layers: Rectified linear unit
- Activation function for output layer: Sigmoid
- Loss function: Binary cross entropy

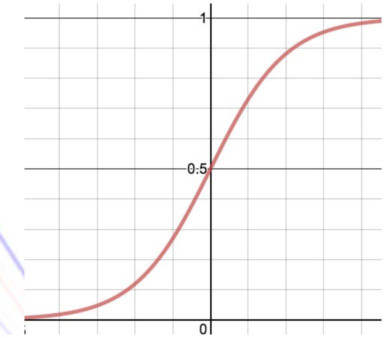
$$\mathrm{loss}(f(x)) = -\sum_{i\in 0}\log f(x_i) - \sum_{i\in 1}\log(1 - f(x_i))$$

- 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

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



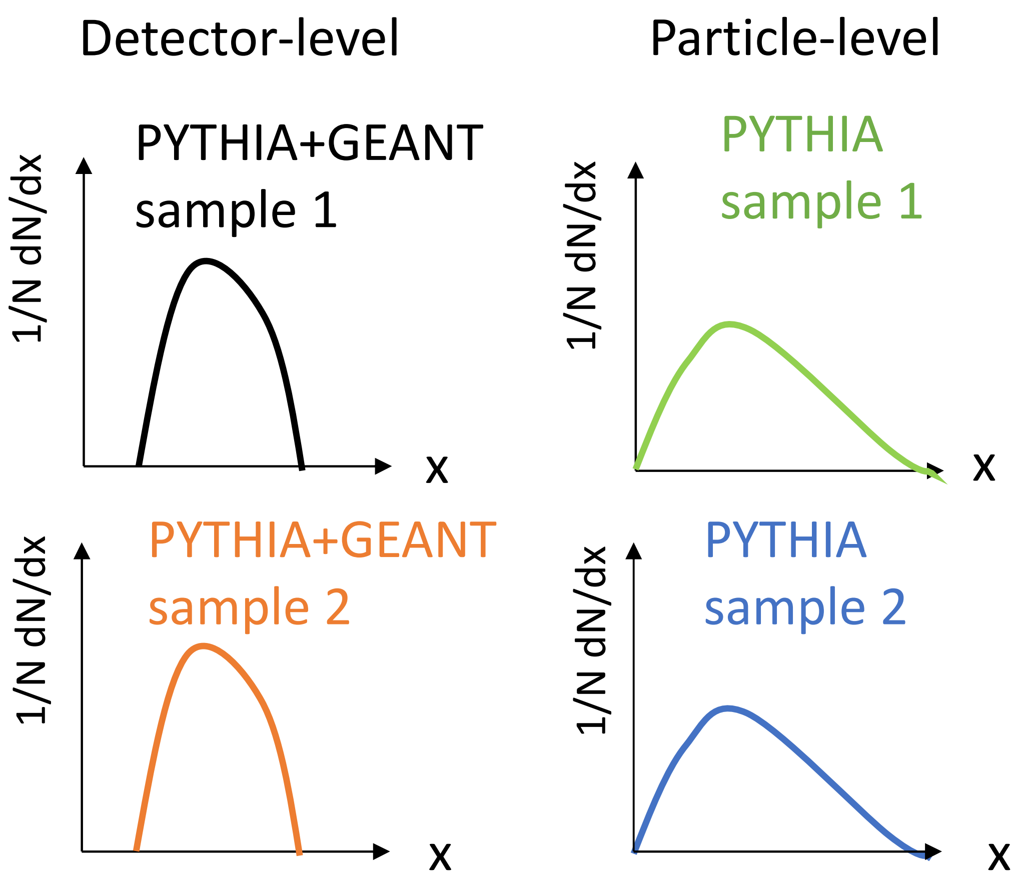


Activation function for output: Sigmoid

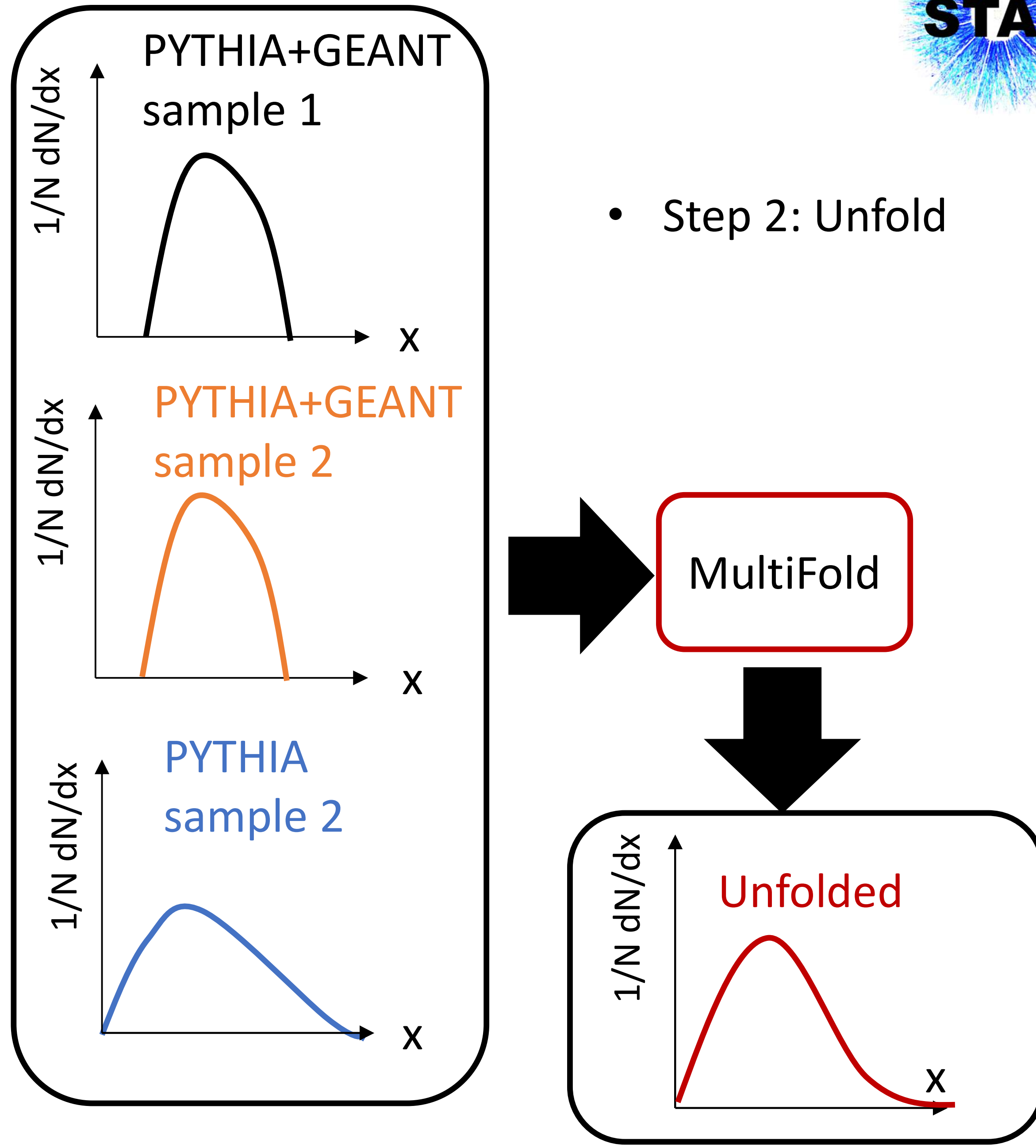
$$f(x) = \frac{1}{1 + e^{-x}}$$

Closure test for unfolding

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

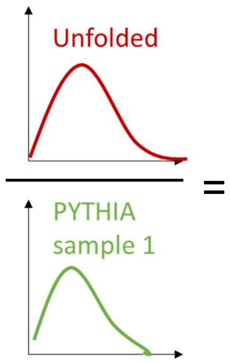


- Step 2: Unfold

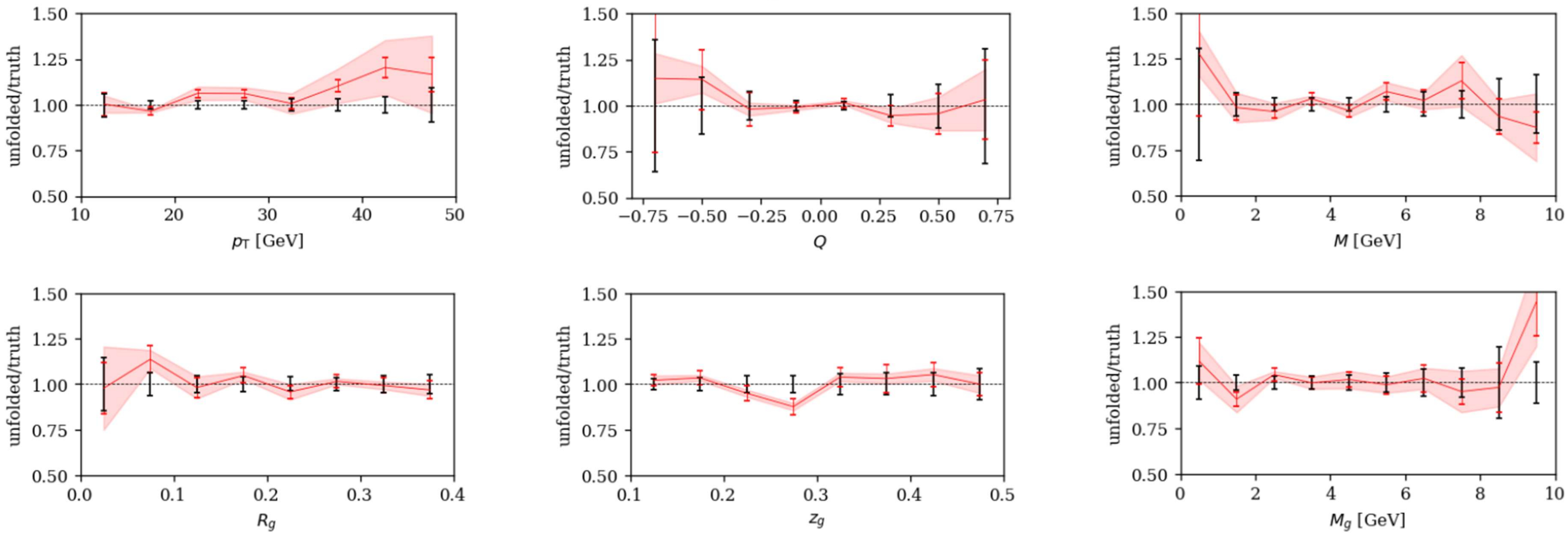


Closure test for unfolding

- Decent **closure** for all substructure observables

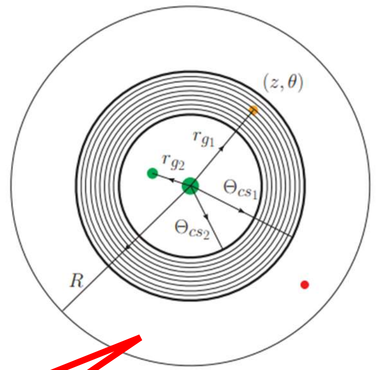


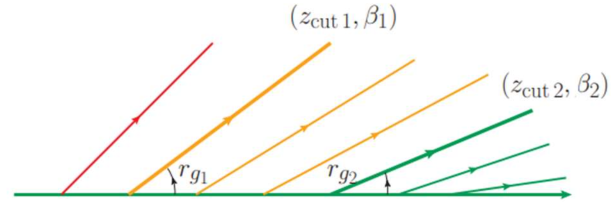
Unfolding unc. on data (not including misses)
Stat. unc. on sample 1
Stat. unc. on sample 2



* 2D reweighting used for prior variation

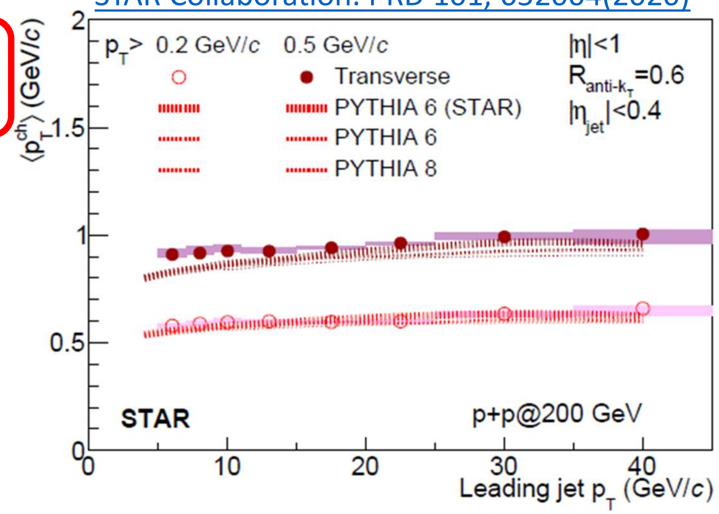
Underlying event

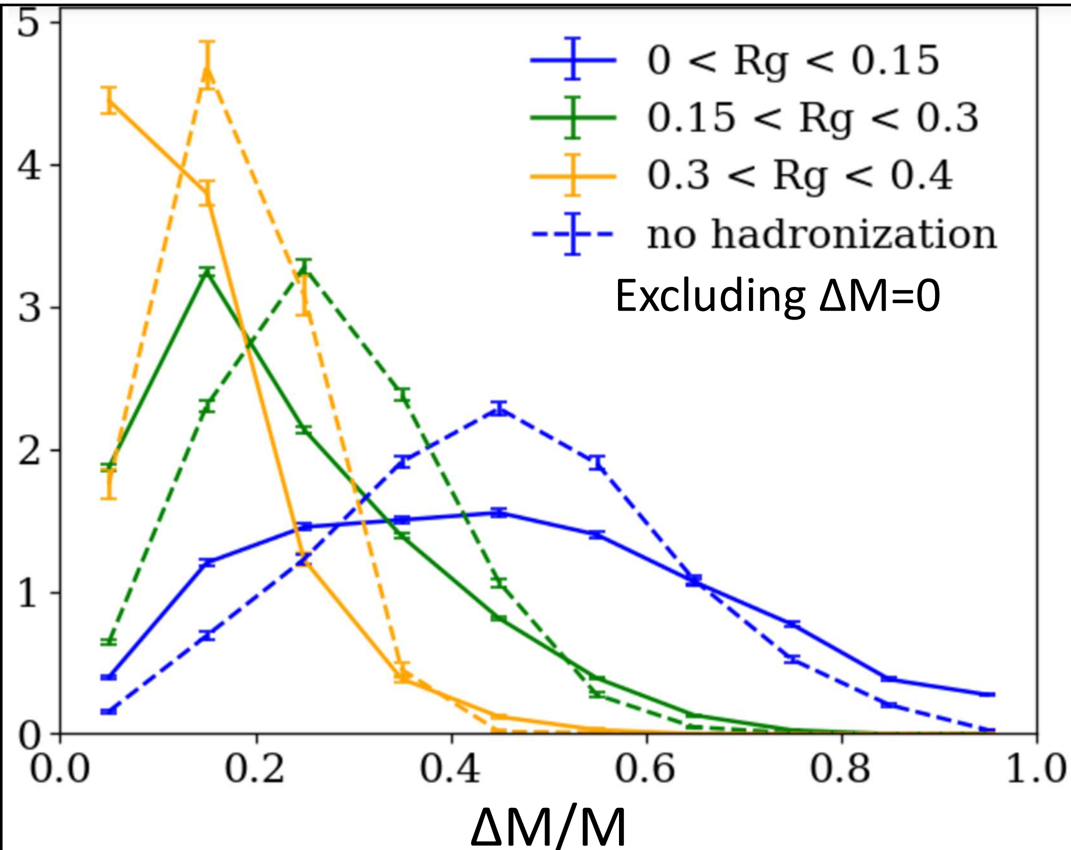


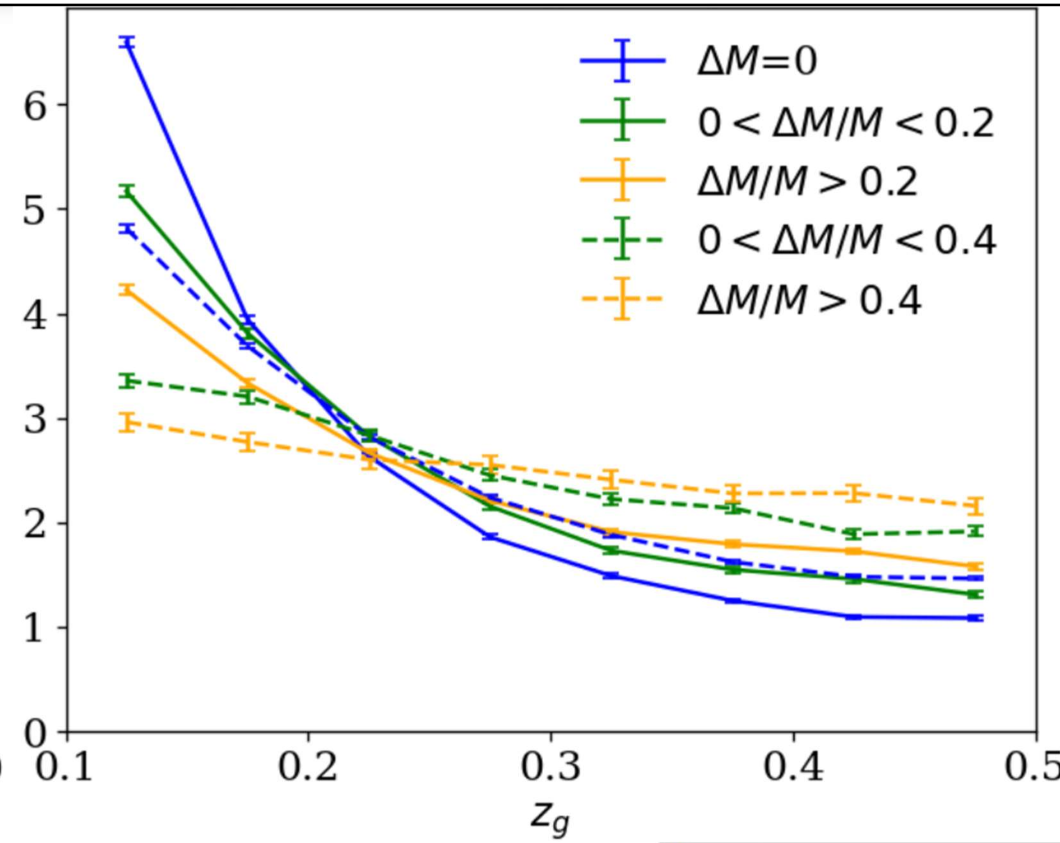


Remove underlying effects and pileup, not necessary for 200 GeV collisions

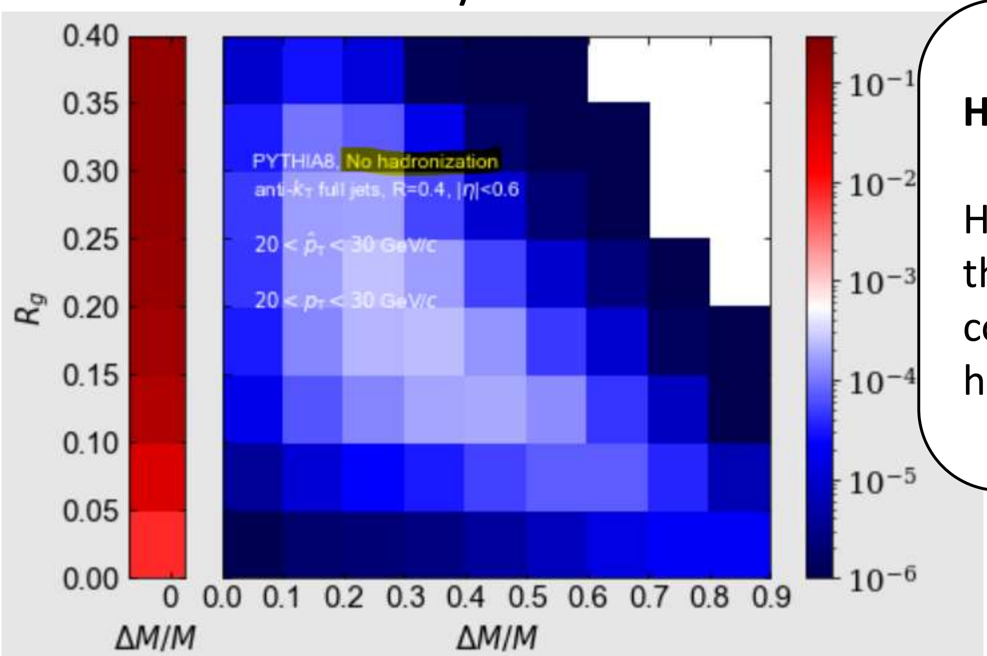
STAR Collaboration. PRD 101, 052004(2020)





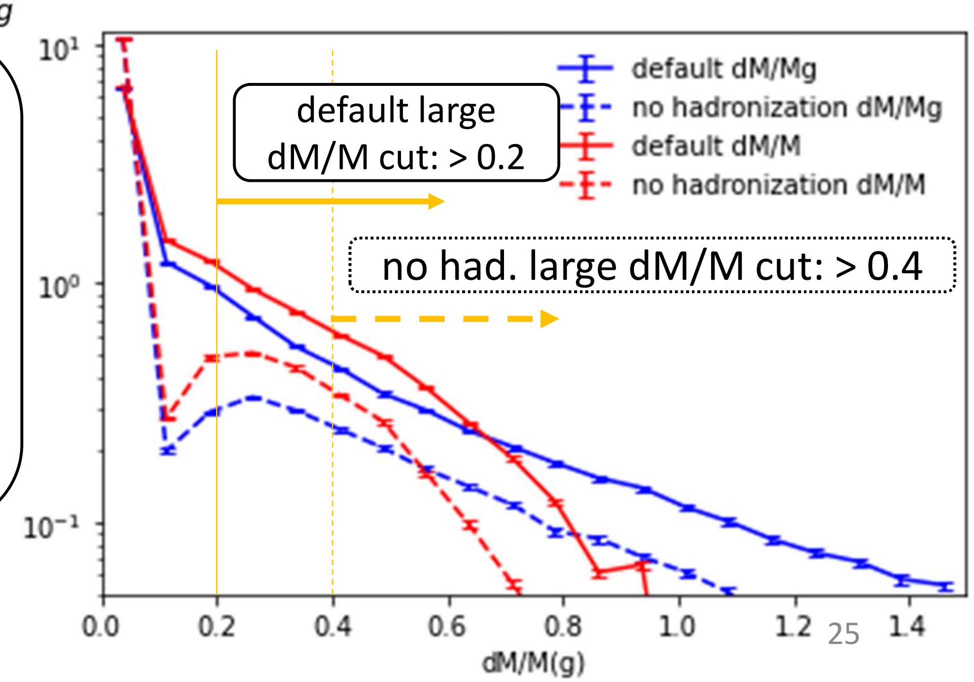


PYTHIA8 Detroit tune
20 < pT,jet < 30 GeV



Hadronization effects study:

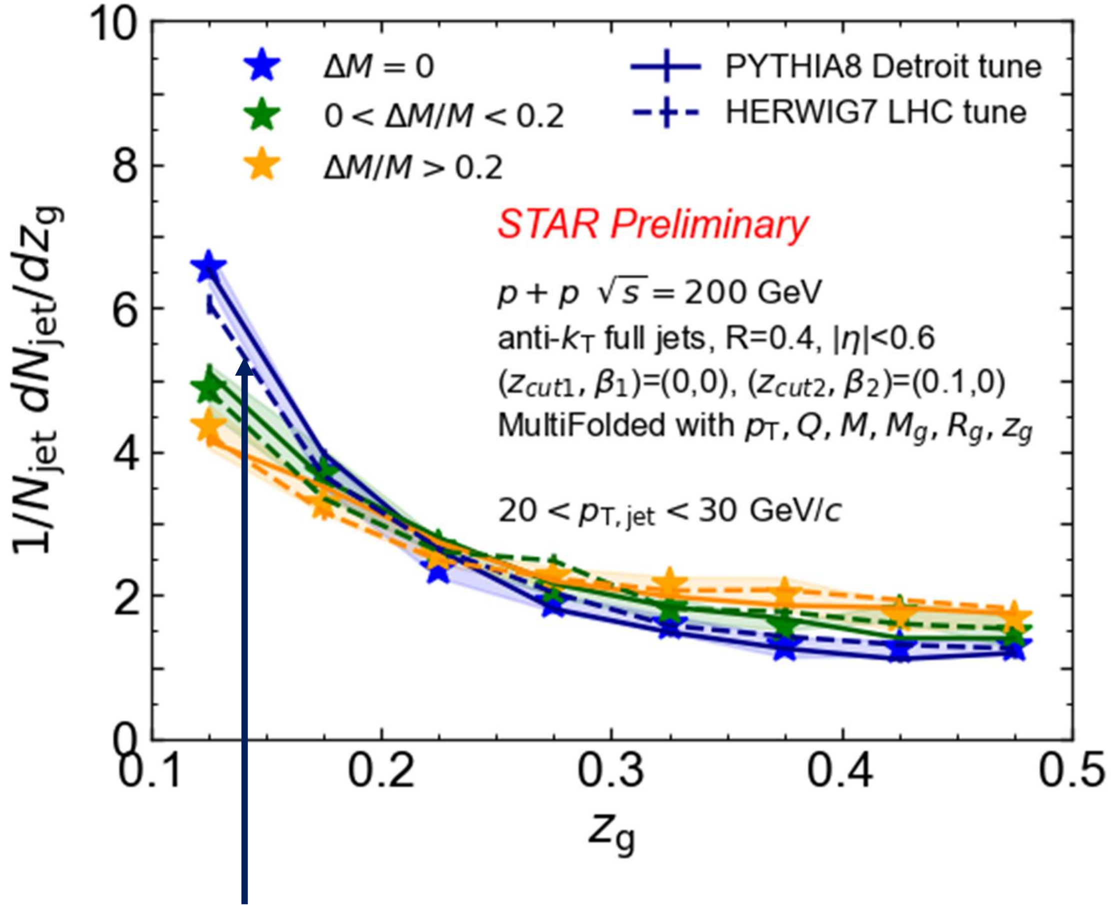
Hadronization smears/shifts the distributions, but the correlation with and without hadronization is the same.



Youqi Song

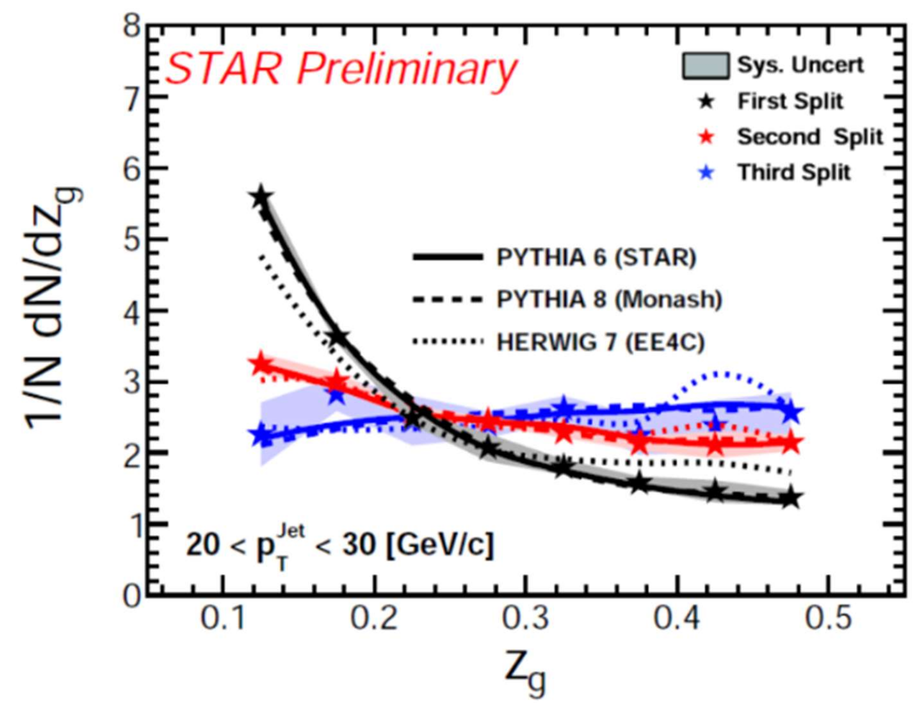
- Among truth jets with $20 < p_T < 30$ GeV, 44% of PYTHIA6 jets, 48% of PYTHIA8 jets, and 43% of HERWIG jets have dM=0.
- Among reco jets with $20 < p_T < 30$ GeV, 36% of PYTHIA6 embedding jets and 37% of data jets have dM=0. (Higher pT jets have a smaller fraction of dM=0).

Soft radiation vs **hard** splitting momentum imbalance



Steeply falling ~ DGLAP 1/z: pQCD

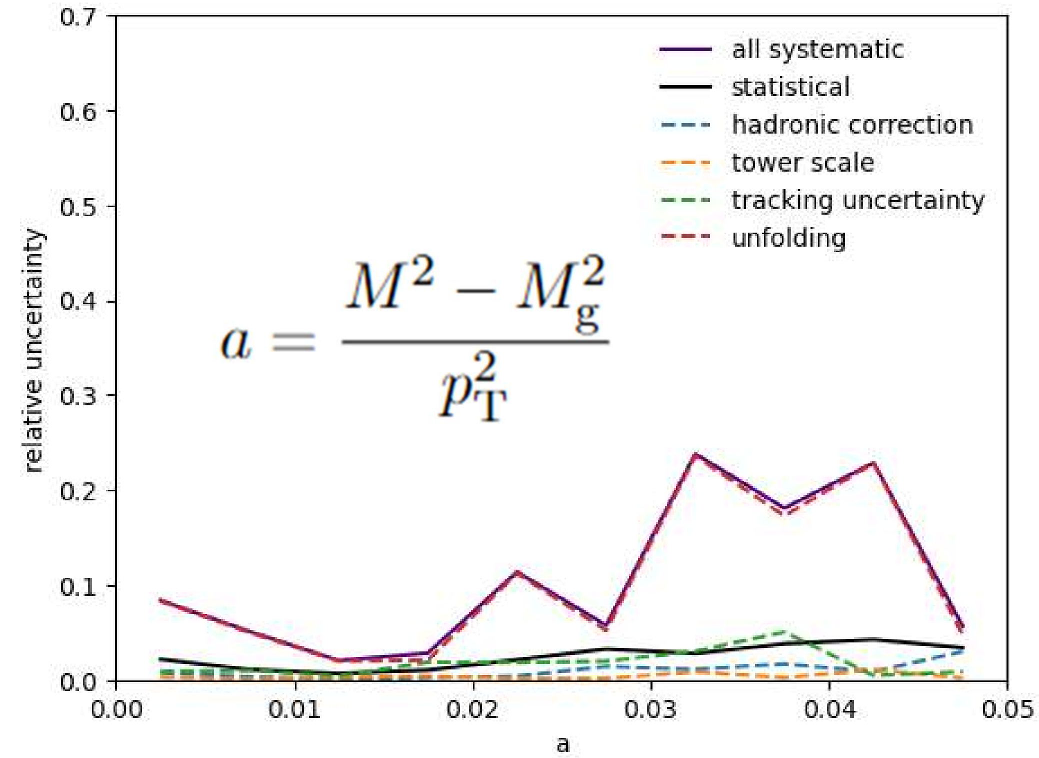
- The more mass that is groomed away relative to the original mass, the flatter the Zg distribution is
  - Demonstrates that **early** soft wide angle radiation constrains the momentum imbalance of & the amount of npQCD contributions to **later** splittings
- MC models describe the trend of data



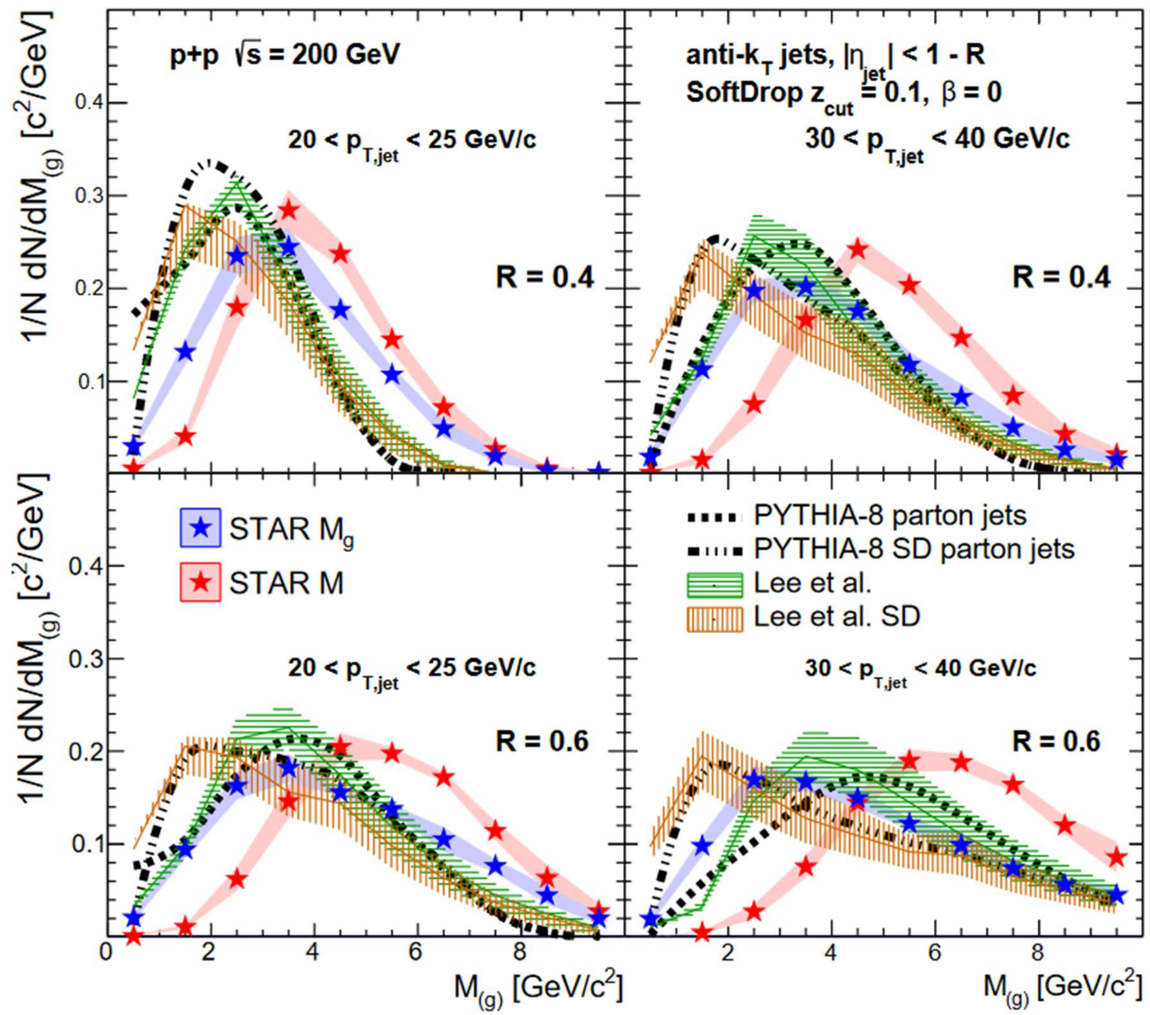
→ The first splitting that passes SoftDrop can still be non-perturbative, but if we apply the $\Delta M = 0$ selection, we can filter out some npQCD contribution due to the parton splitting

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: $\mathrm{prior} \to \mathrm{reweight} \bigotimes \mathrm{nominal\ prior}$,
      with $\mathrm{reweight}(p_\mathrm{T}, Q, M, M_\mathrm{g}, R_\mathrm{g}, z_\mathrm{g}) = \frac{\mathrm{Herwig\ truth}(p_\mathrm{T}, Q, M, M_\mathrm{g}, R_\mathrm{g}, z_\mathrm{g})}{\mathrm{Pythia6\ truth}(p_\mathrm{T}, Q, M, M_\mathrm{g}, R_\mathrm{g}, z_\mathrm{g})}$



Jet mass: Comparison with models and calcaulations



STAR Collaboration. PRD 104, 052007(2021)