

Measurement of collinear drop jet mass and its correlation with groomed jet substructure observables in pp collisions

Youqi Song (youqi.song@yale.edu)

for the STAR Collaboration

DIS 2023, East Lansing, Michigan

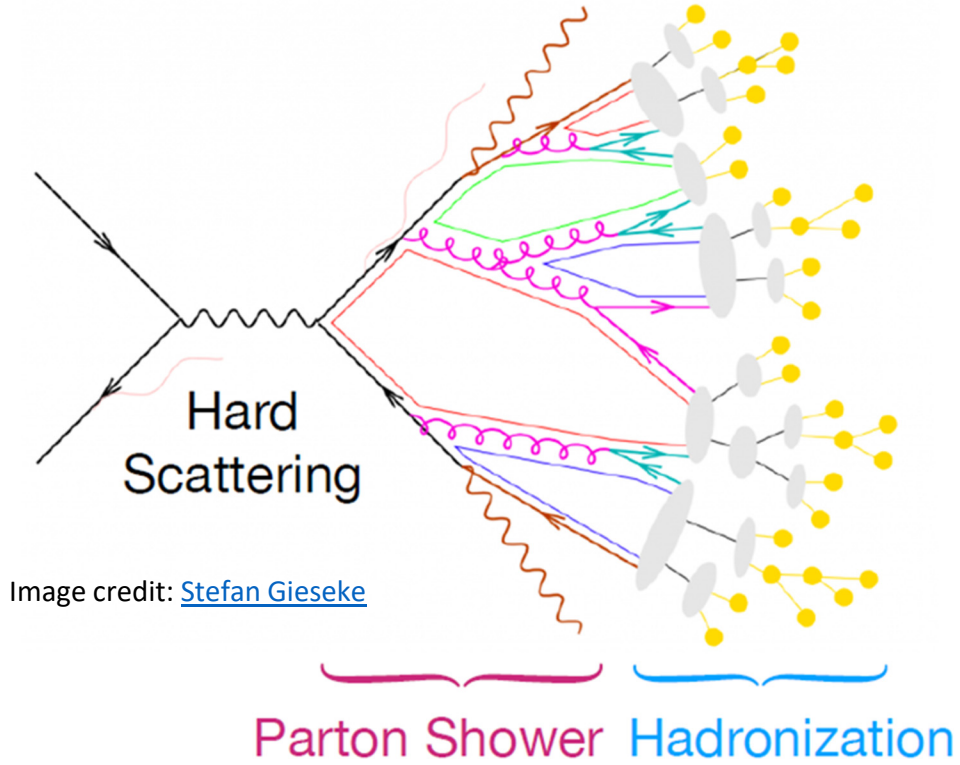
3/28/2023



WG4: QCD with Heavy Flavours
and Hadronic Final States

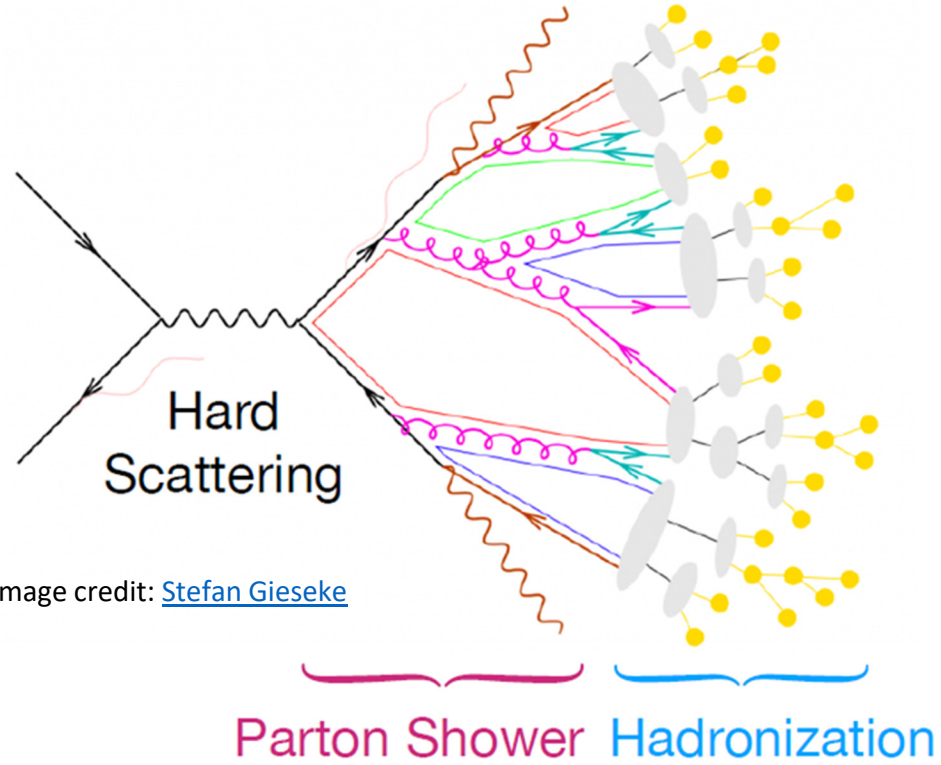


Motivation



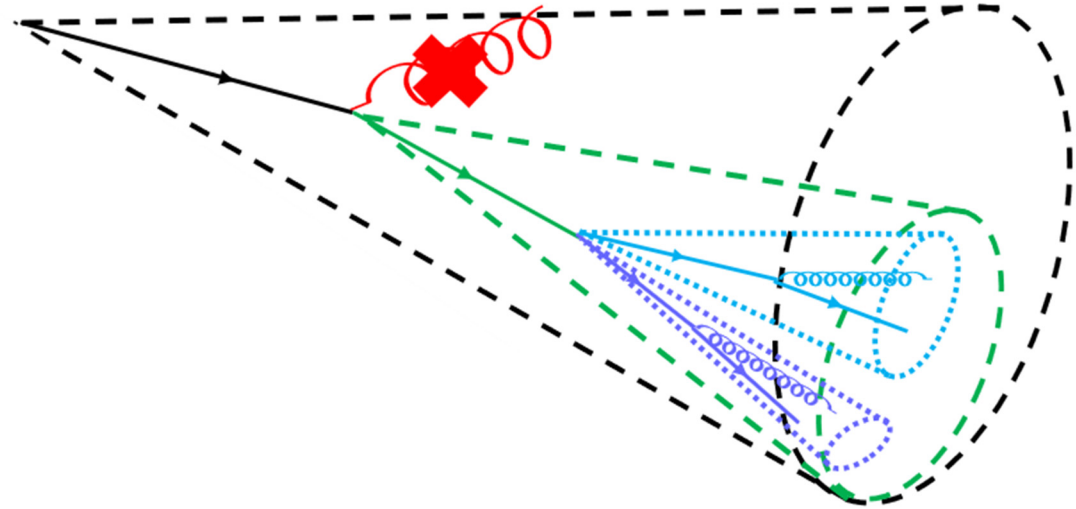
Jet is a multi-scale object

Motivation

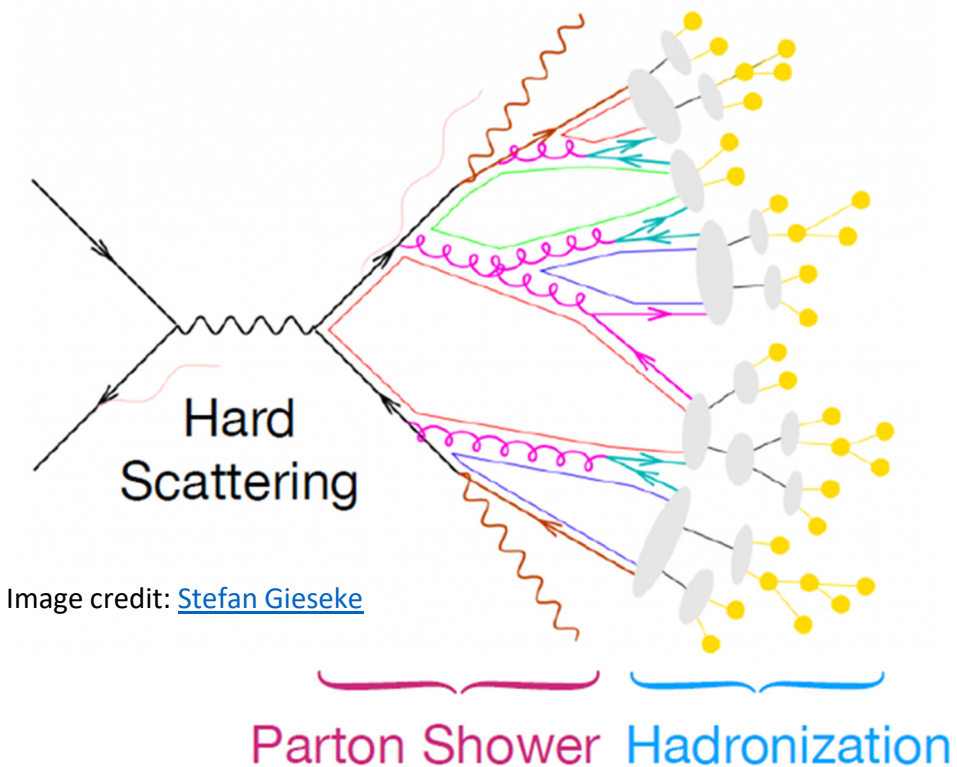


Jet is a multi-scale object

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

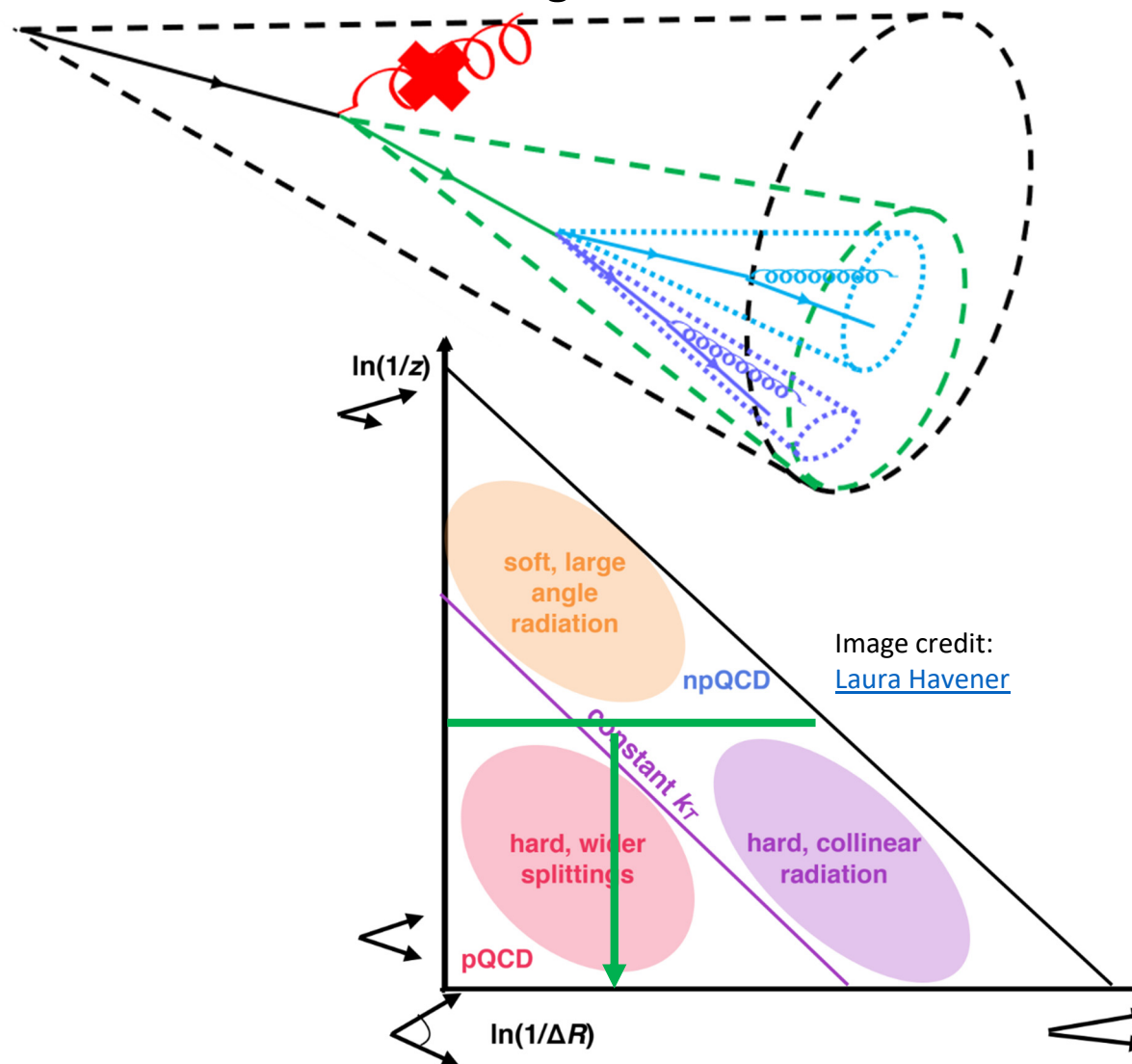


Motivation

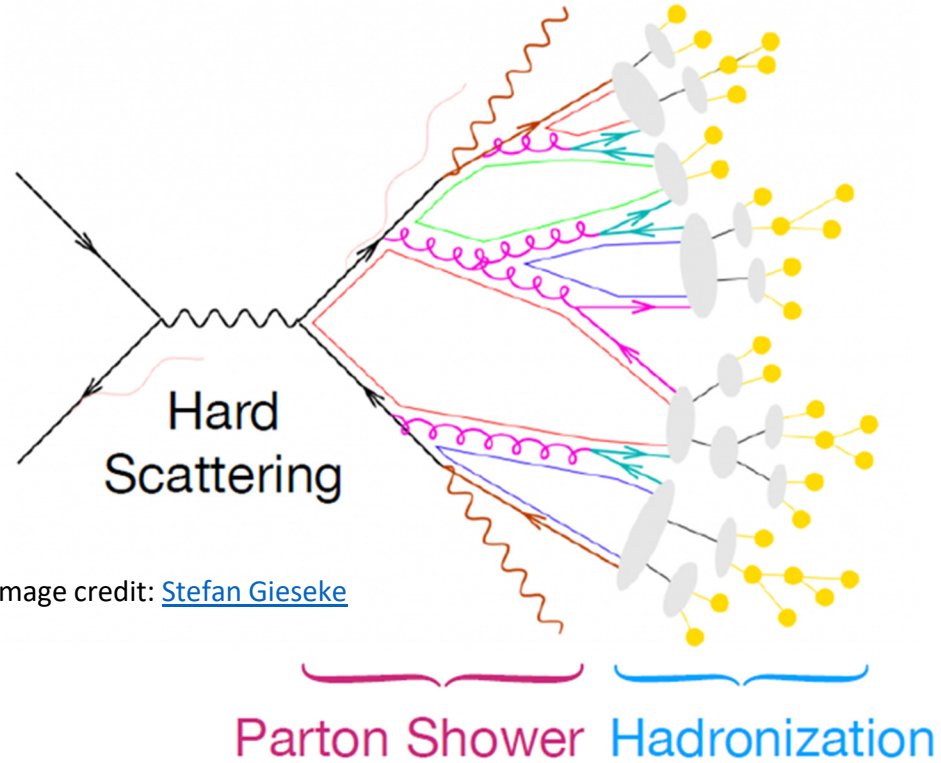


Jet is a multi-scale object

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



Motivation

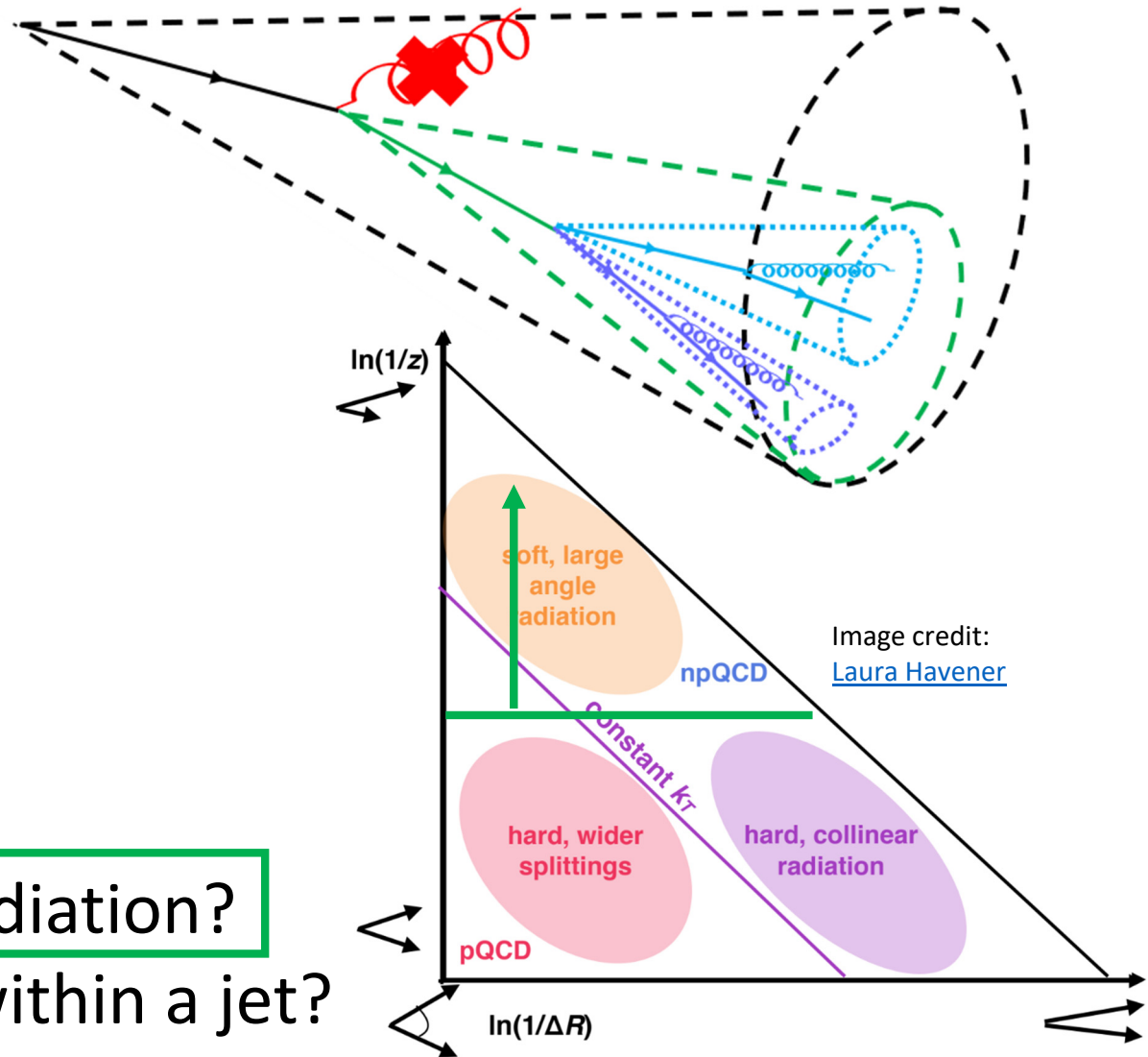


Jet is a multi-scale object

What about the wide-angle soft radiation?

What is the soft-hard correlation within a jet?

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



Motivation



1. **soft** radiation within the jet \rightarrow collinear drop jet observable

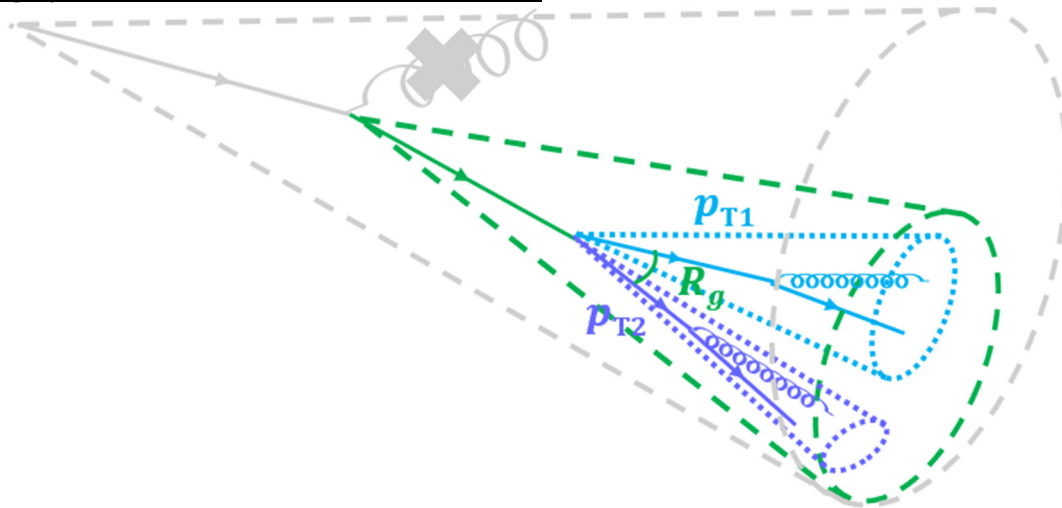
Motivation

1. **soft** radiation within the jet → collinear drop jet observable

SoftDrop: removes wide-angle soft radiation

Larkoski, et al. JHEP 2014, 146 (2014).

Dasgupta et al. JHEP 2013, 29 (2013).



Require subjet momentum fraction to pass

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

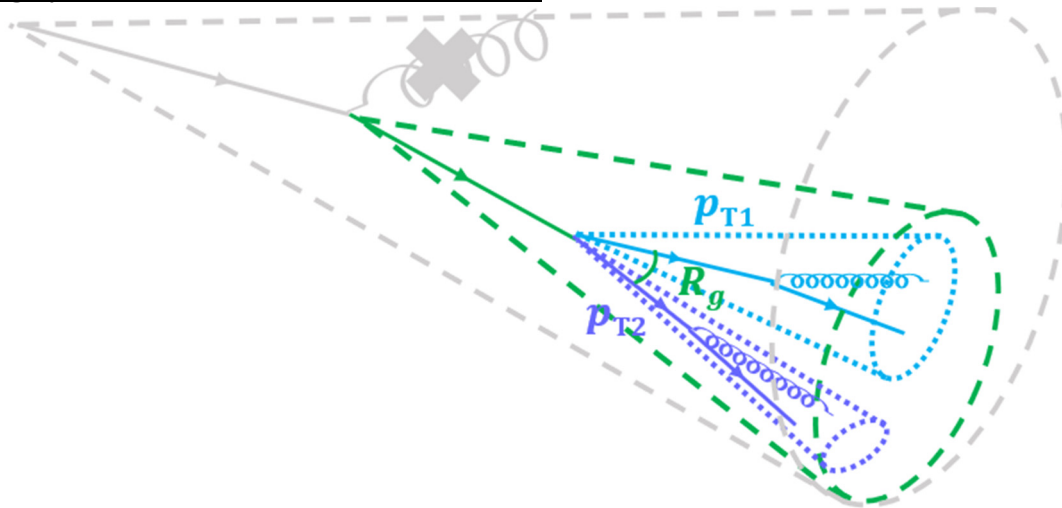
Motivation

1. **soft** radiation within the jet \rightarrow collinear drop jet observable

SoftDrop: removes wide-angle soft radiation

Larkoski, et al. JHEP 2014, 146 (2014).

Dasgupta et al. JHEP 2013, 29 (2013).



Require subjet momentum fraction to pass

$$z_g = \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{\text{cut}} (R_g/R_{\text{jet}})^\beta \quad \begin{matrix} z_{\text{cut}} = 0.1 \\ \beta = 0 \end{matrix}$$

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

Motivation

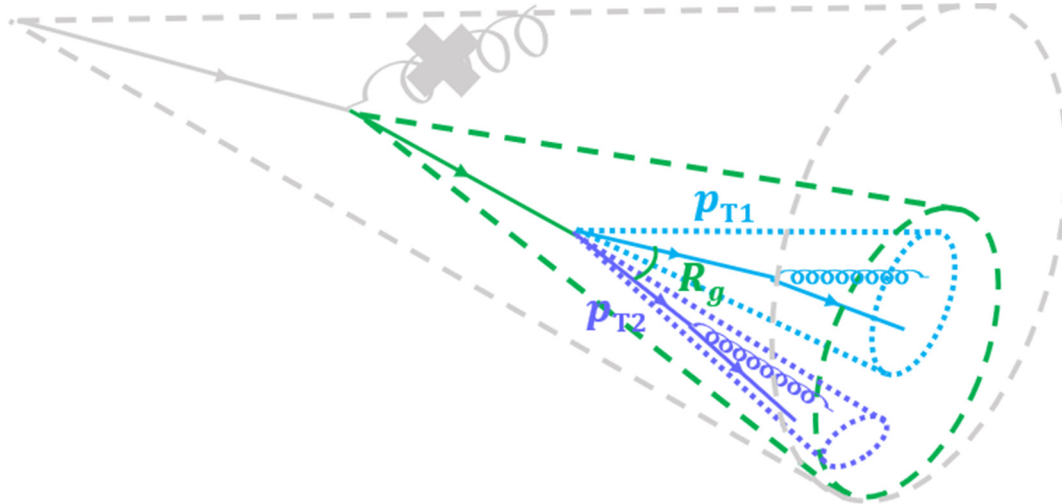
1. **soft** radiation within the jet → collinear drop jet observable



SoftDrop: removes wide-angle soft radiation

Larkoski, et al. JHEP 2014, 146 (2014).

Dasgupta et al. JHEP 2013, 29 (2013).

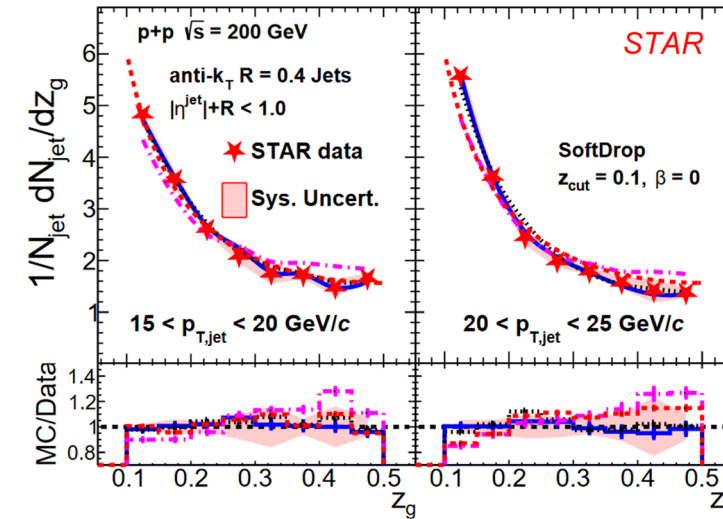


Require subjet momentum fraction to pass

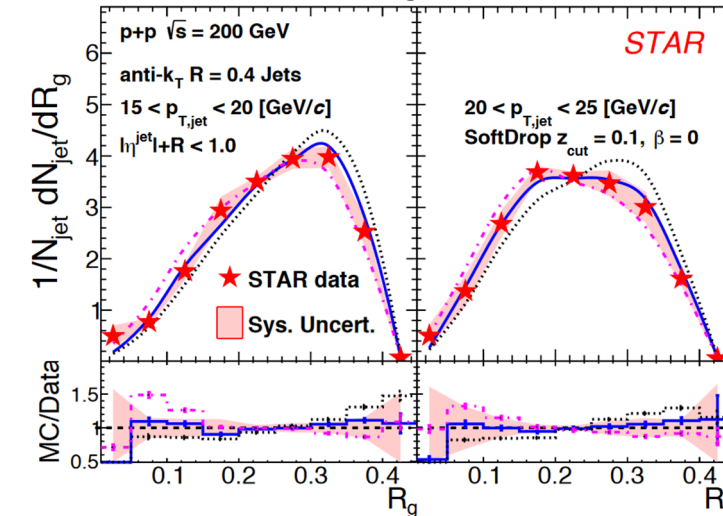
$$z_g = \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{\text{cut}} (R_g/R_{\text{jet}})^\beta \quad z_{\text{cut}} = 0.1 \quad \beta = 0$$

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

z_g



R_g



STAR. PLB 811 (2020) 135846

Motivation



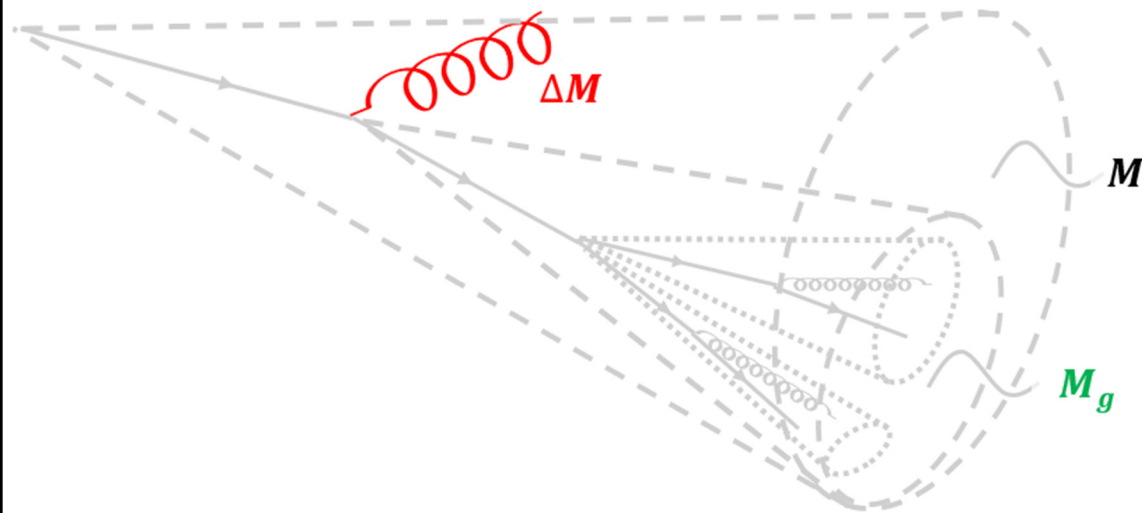
1. **soft** radiation within the jet \rightarrow collinear drop jet observable

Motivation

1. **soft** radiation within the jet \rightarrow collinear drop jet observable

Collinear Drop: probes the soft component

Chien and Stewart JHEP 2020, 64 (2020).

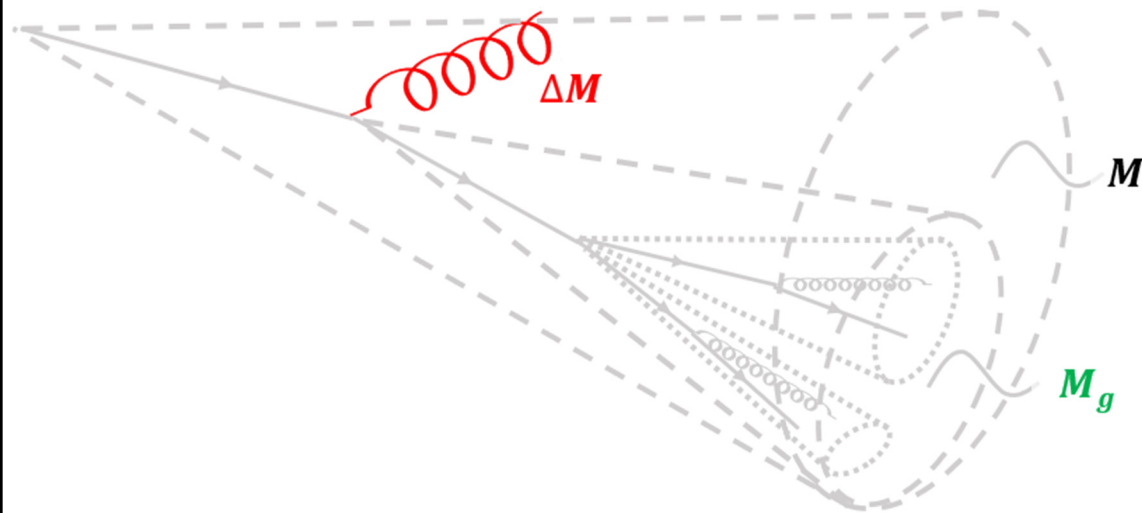


Motivation

1. **soft** radiation within the jet \rightarrow collinear drop jet observable

Collinear Drop: probes the soft component

Chien and Stewart JHEP 2020, 64 (2020).

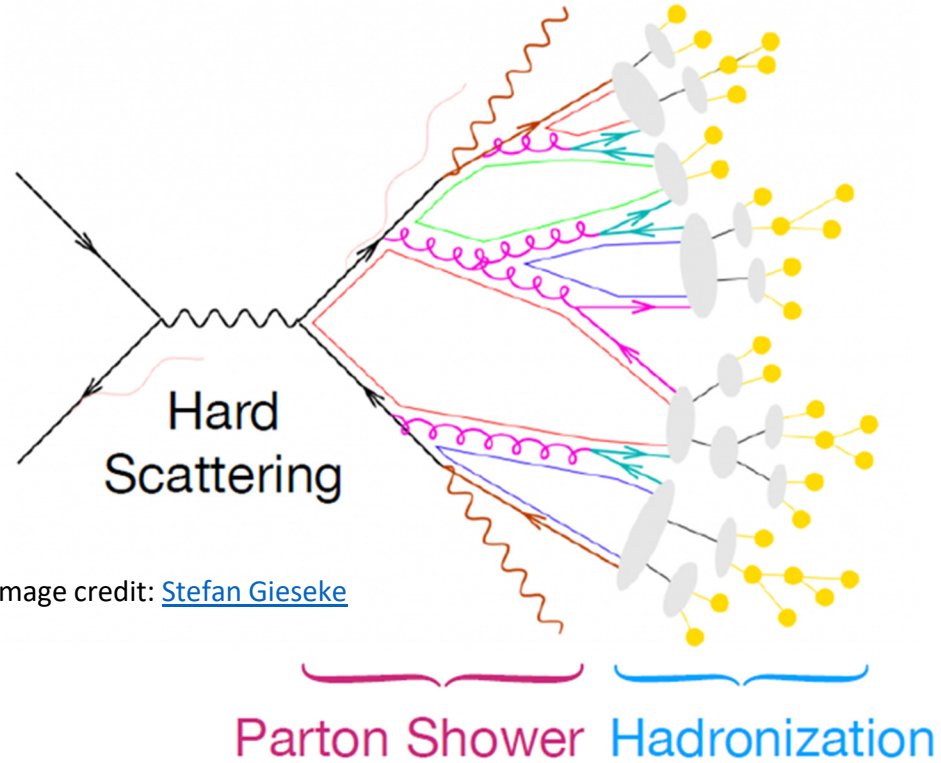


- General case: difference of an observable with two different SoftDrop selections $(\mathbf{z}_{cut\ 1}, \boldsymbol{\beta}_1)$ and $(\mathbf{z}_{cut\ 2}, \boldsymbol{\beta}_2)$
- For this analysis, $(\mathbf{z}_{cut\ 1}, \boldsymbol{\beta}_1) = (\mathbf{0}, \mathbf{0})$ and $(\mathbf{z}_{cut\ 2}, \boldsymbol{\beta}_2) = (\mathbf{0.1}, \mathbf{0})$: difference in the original and SoftDrop groomed observable

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

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

Motivation

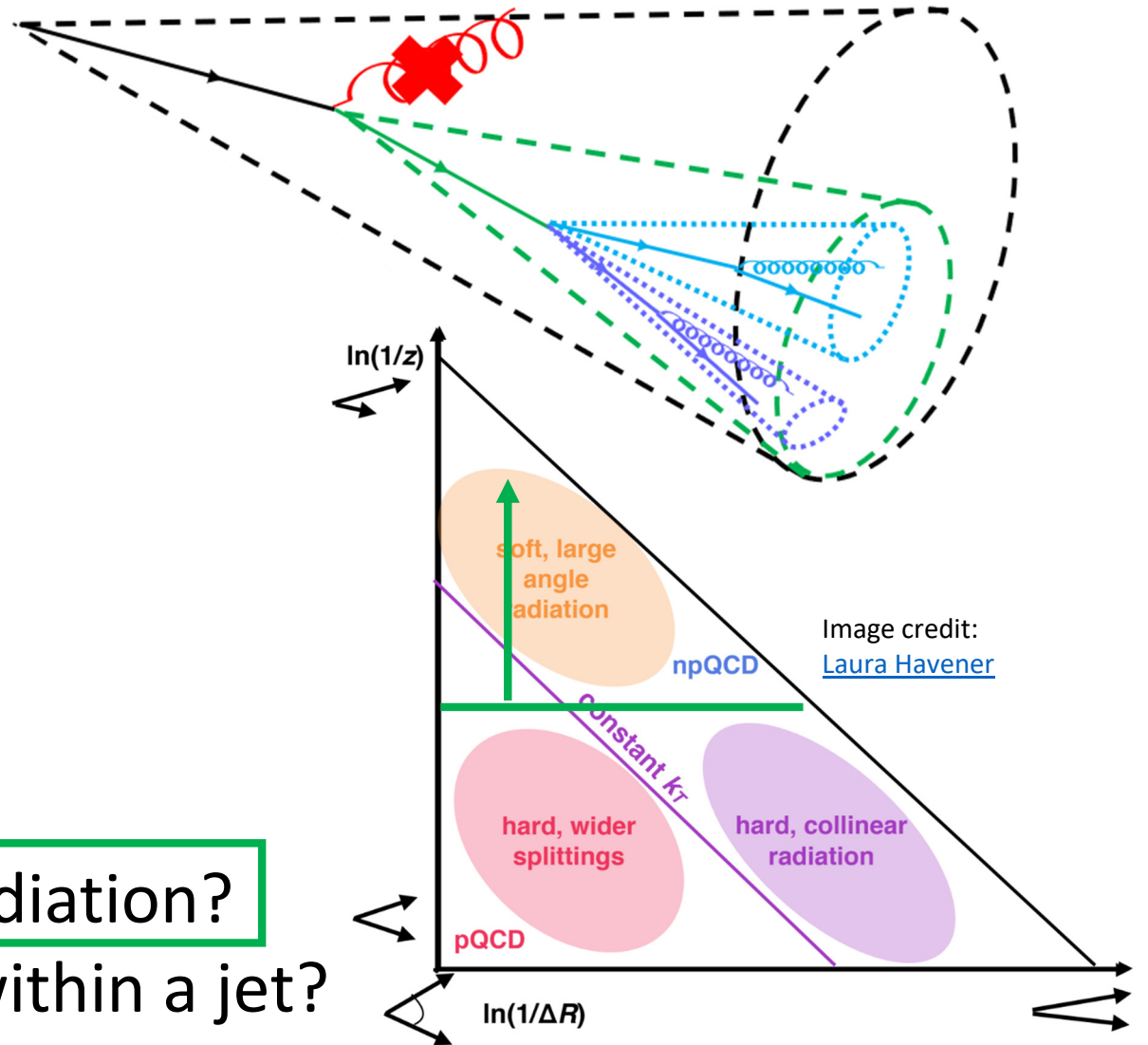


Jet is a multi-scale object

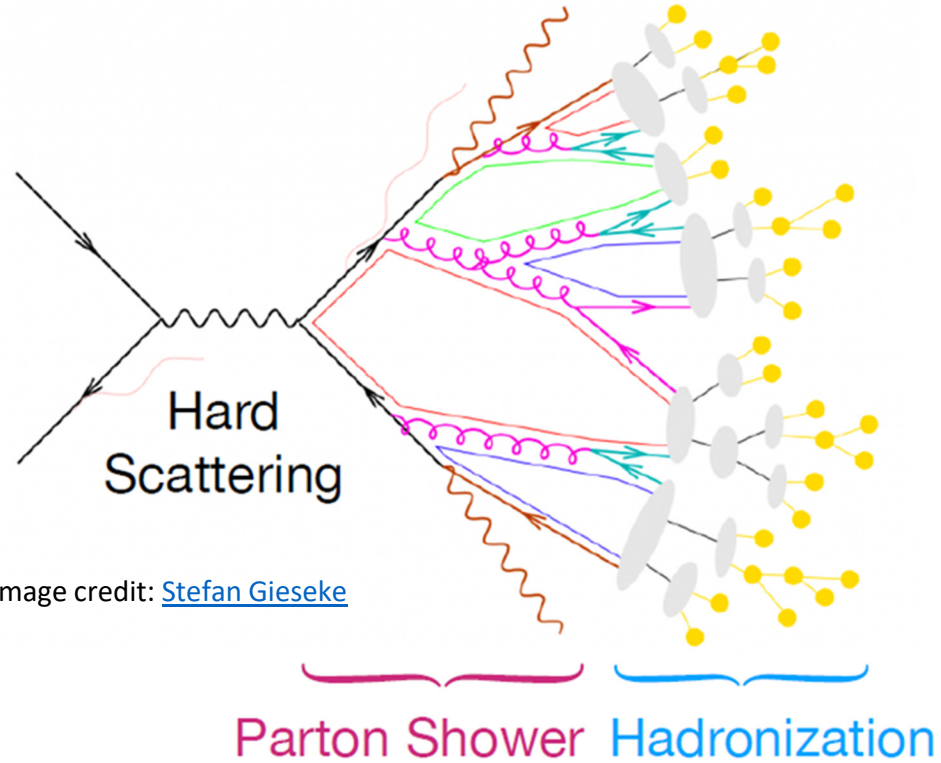
What about the wide-angle soft radiation?

What is the soft-hard correlation within a jet?

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



Motivation

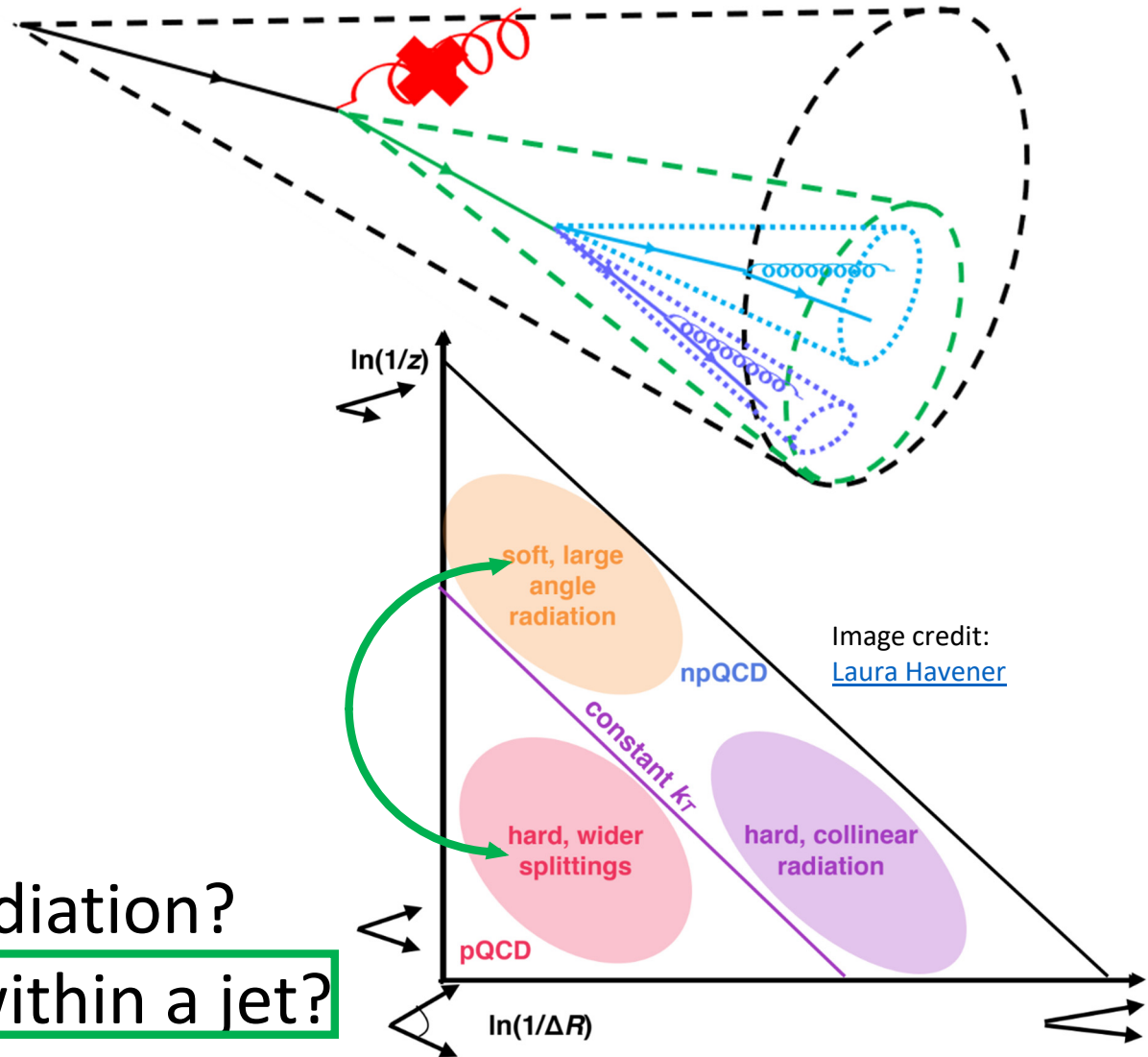


Jet is a multi-scale object

What about the wide-angle soft radiation?

What is the soft-hard correlation within a jet?

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





Motivation

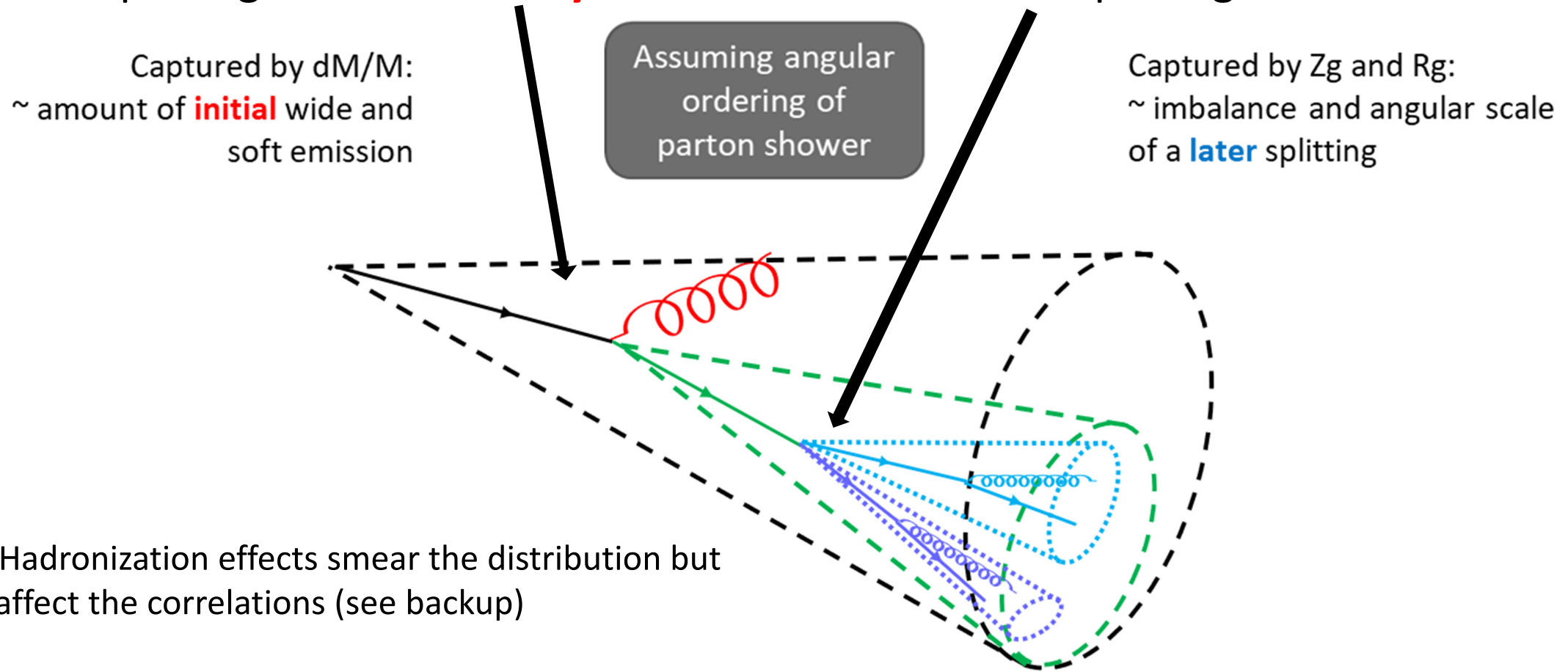
2. **soft-hard correlation** → collinear drop jet vs groomed jet observables

- 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

Motivation

2. **soft-hard correlation** → collinear drop jet vs groomed jet observables

- 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





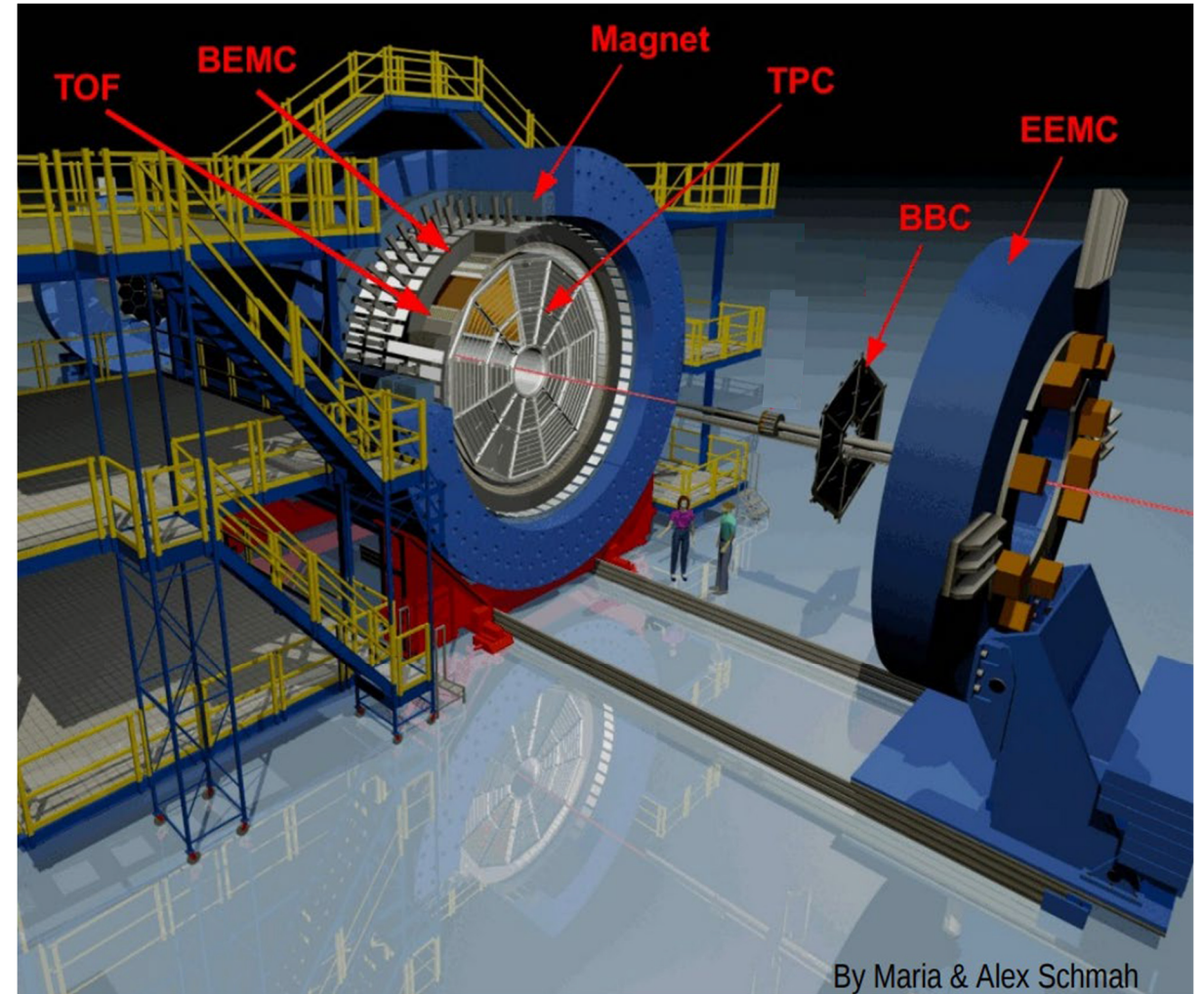
Jet reconstruction at STAR

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**
 - $|\eta_{\text{jet}}| < 0.6$

Additional selections

- Tracks (Towers): $0.2 < p_T(E_T) < 30 \text{ GeV}/c$
- Jets
 - $p_T > 15 \text{ GeV}/c$, $M > 1 \text{ GeV}/c^2$
 - Passes SoftDrop with $z_{\text{cut}} = 0.1$ 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 multiple observables → **Correlation** information is retained!
- First application of MultiFold on RHIC data!



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 multiple observables → **Correlation** information is retained!
- First application of MultiFold on RHIC data!

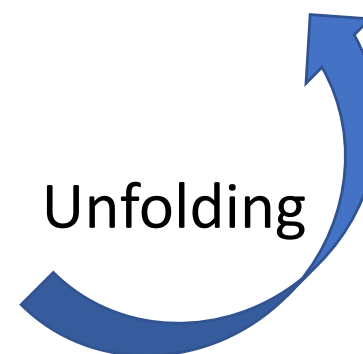
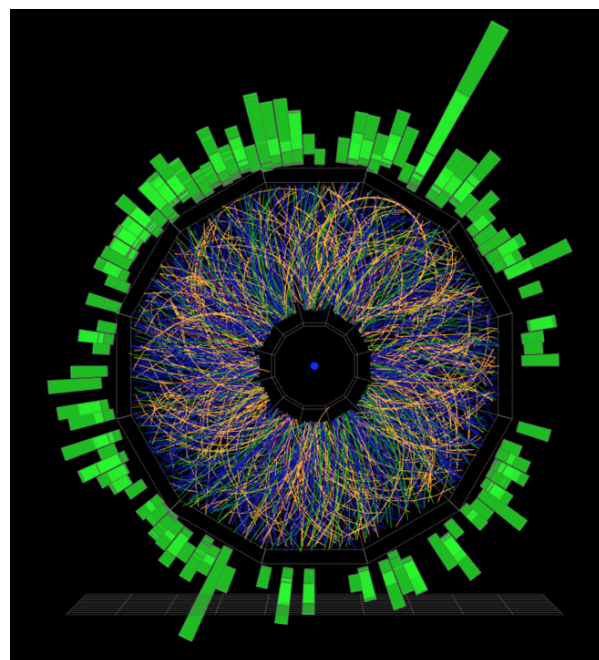
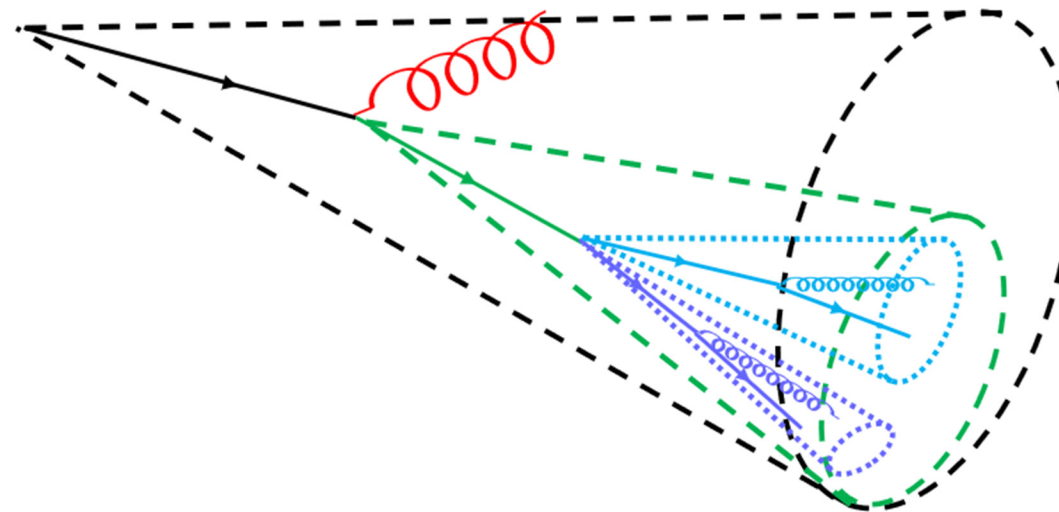
- Jet observables
 - p_T : transverse momentum
 - $Q^\kappa = \frac{1}{(p_{T\text{jet}})^\kappa} \sum_{i \in \text{jet}} q_i \cdot (p_{Ti})^\kappa$ → Choose $K=2$
 - $M = |\sum_{i \in \text{jet}} p_i| = \sqrt{E^2 - |\vec{p}|^2}$
 - 4-momentum of the constituent i
 - R_g : groomed jet radius
 - z_g : shared momentum fraction
 - 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)



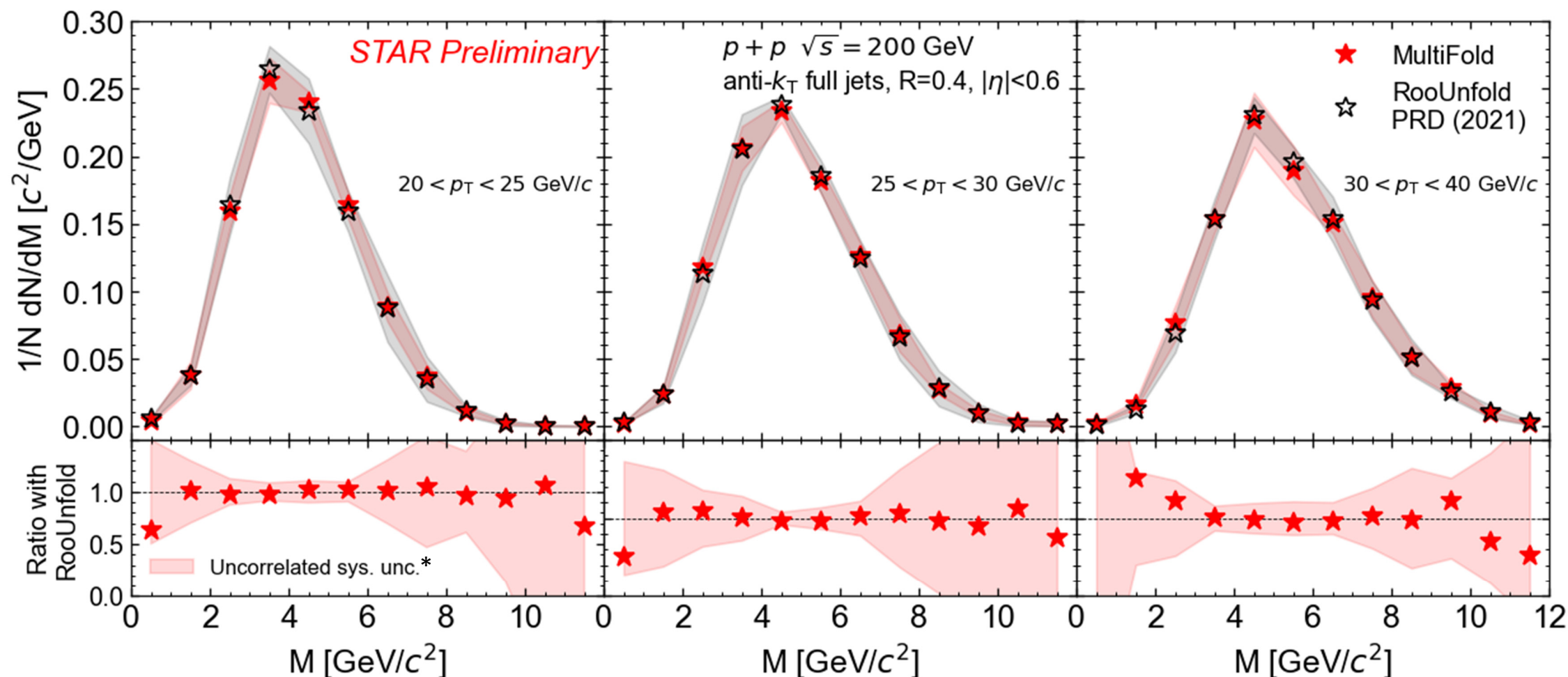
Does our method
work?



Fully corrected jet M

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

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



... but **MultiFold** also gives us correlation between observables!

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



Looking at wide-angle soft radiation

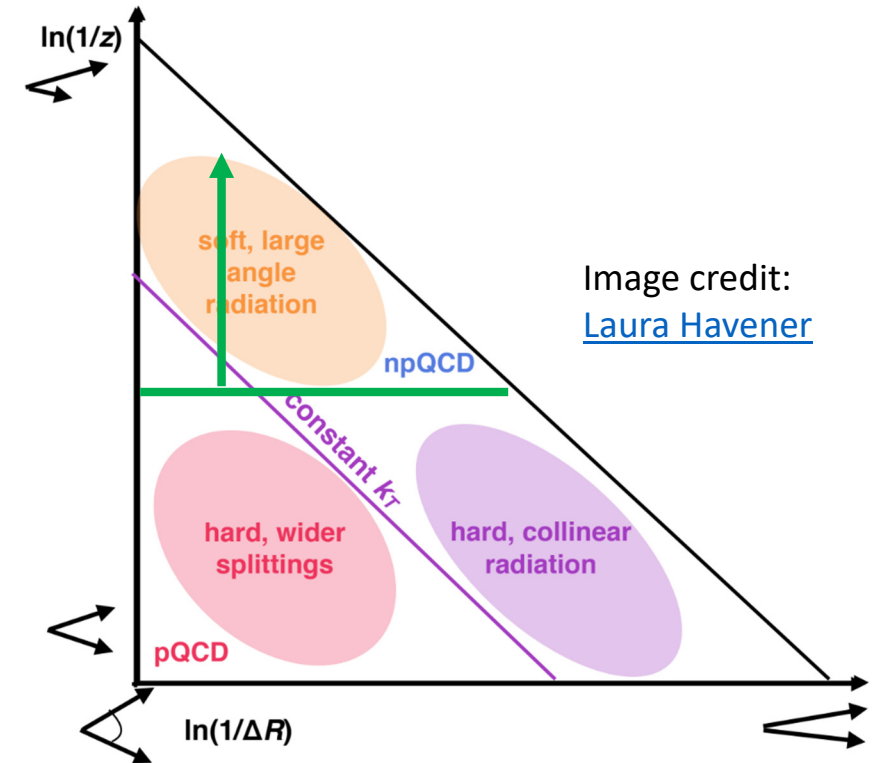
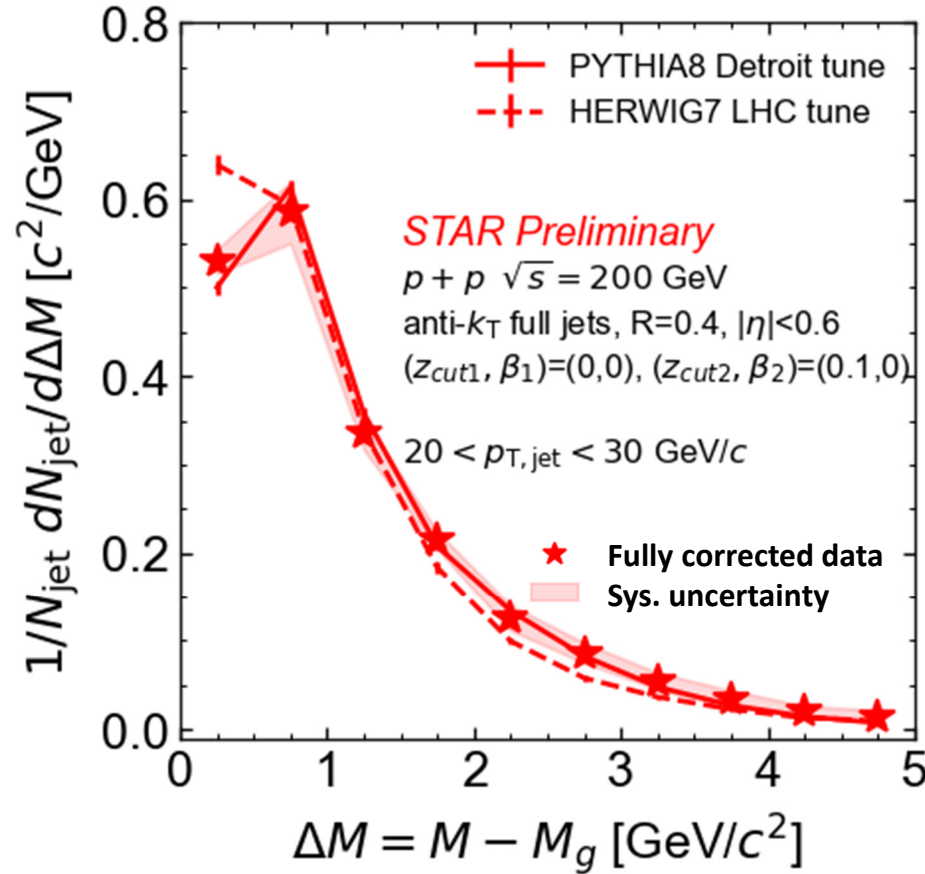


Image credit:
[Laura Havener](#)

Collinear drop groomed jet mass

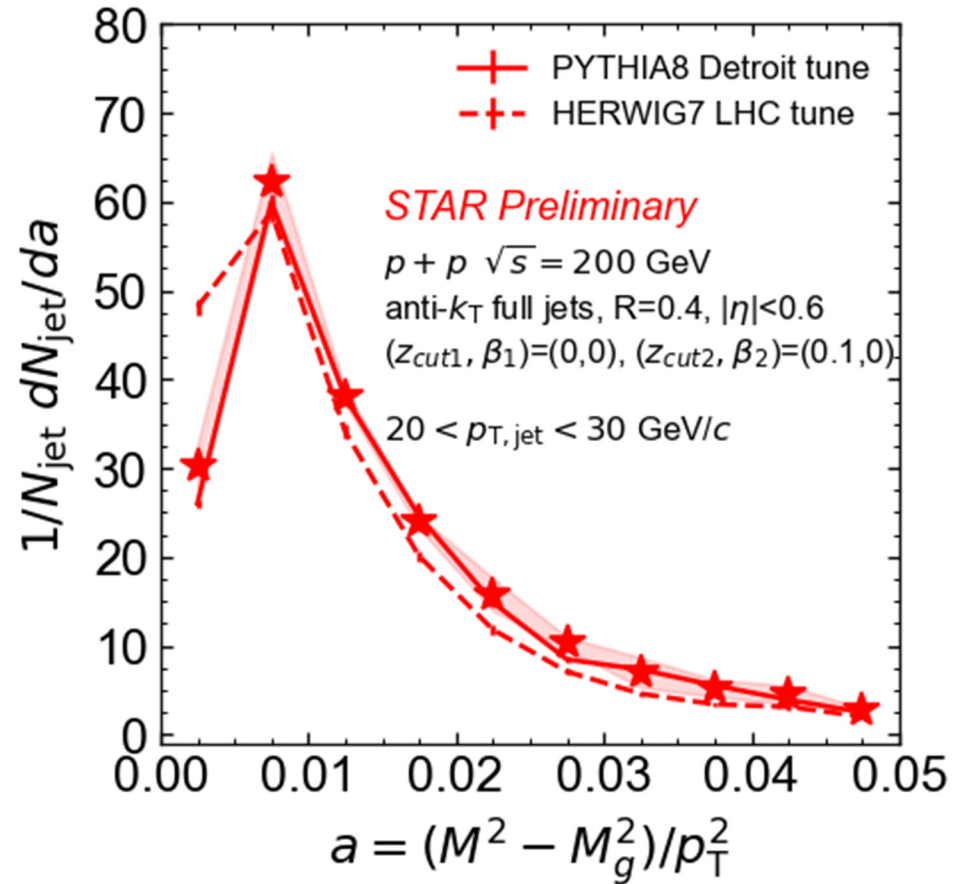
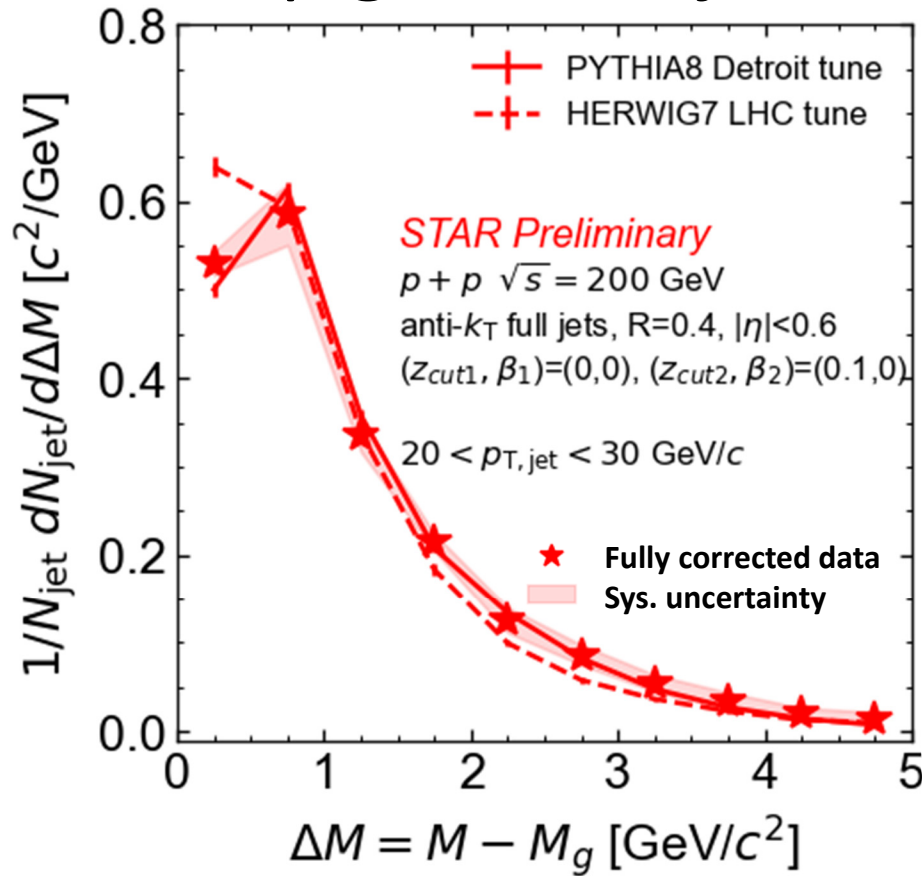


PYTHIA6 Perugia 2012 STAR tune: Skands, PRD 82, 074018 (2010)
 PYTHIA8 Detroit tune: Aguilar et al. PRD 105, 016011(2022)
 HERWIG7: Bellm, et al. PRC 76, 1-8 (2016)

Measurement excludes jets with $\Delta M = 0$ (45.5% of jets in this jet p_T range)

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

Collinear drop groomed jet mass

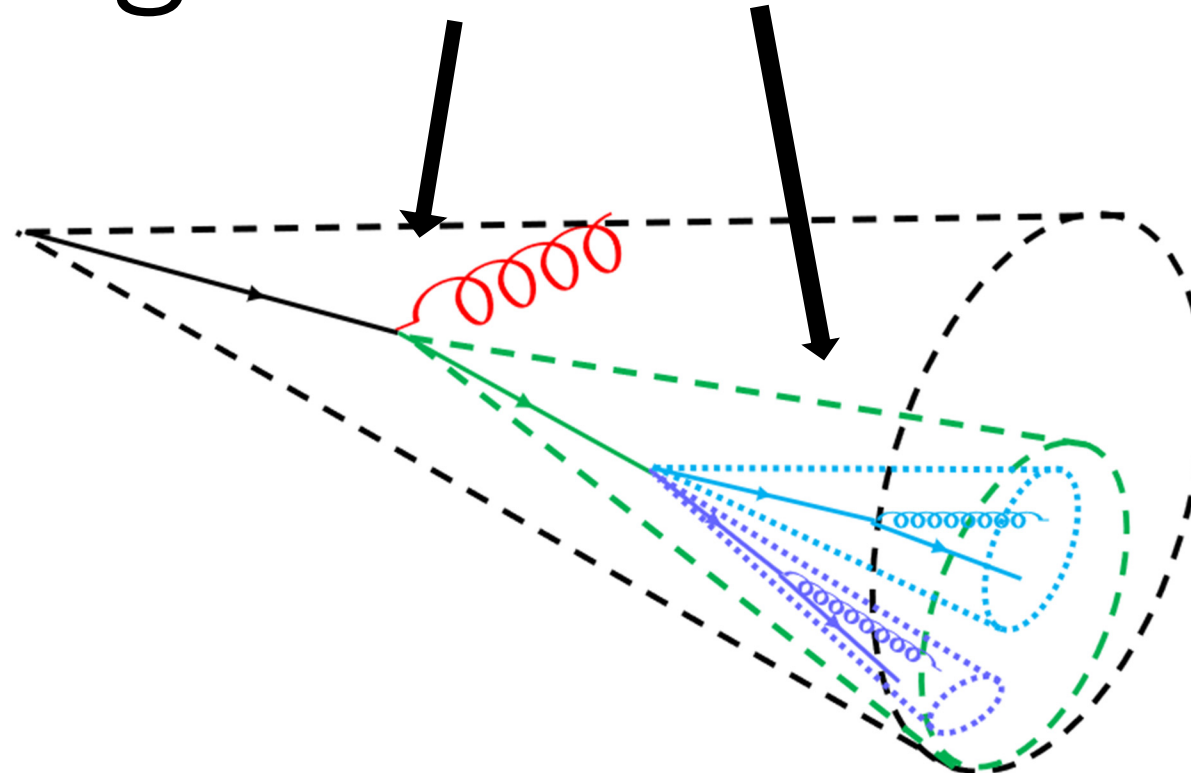


PYTHIA6 Perugia 2012 STAR tune: Skands, PRD 82, 074018 (2010)
 PYTHIA8 Detroit tune: Aguilar et al. PRD 105, 016011(2022)
 HERWIG7: Bellm, et al. PRC 76, 1-8 (2016)

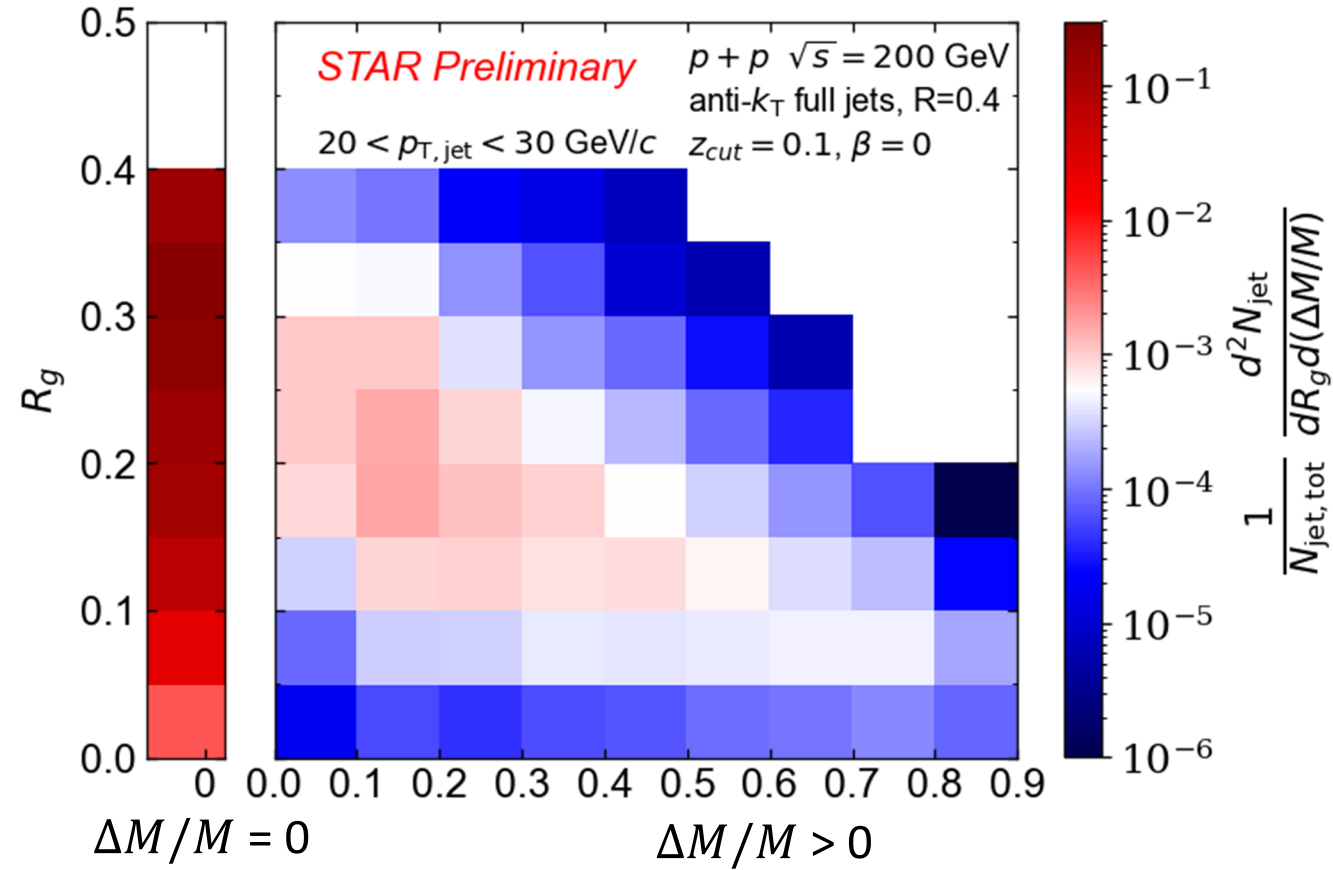
Measurement excludes jets with $\Delta M = 0$ (45.5% of jets in this jet p_T range)

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

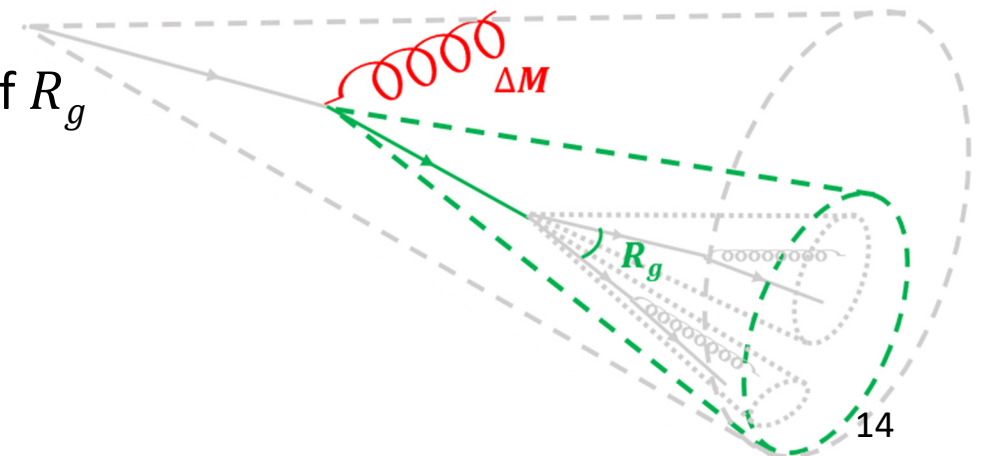
Probing soft-hard correlation



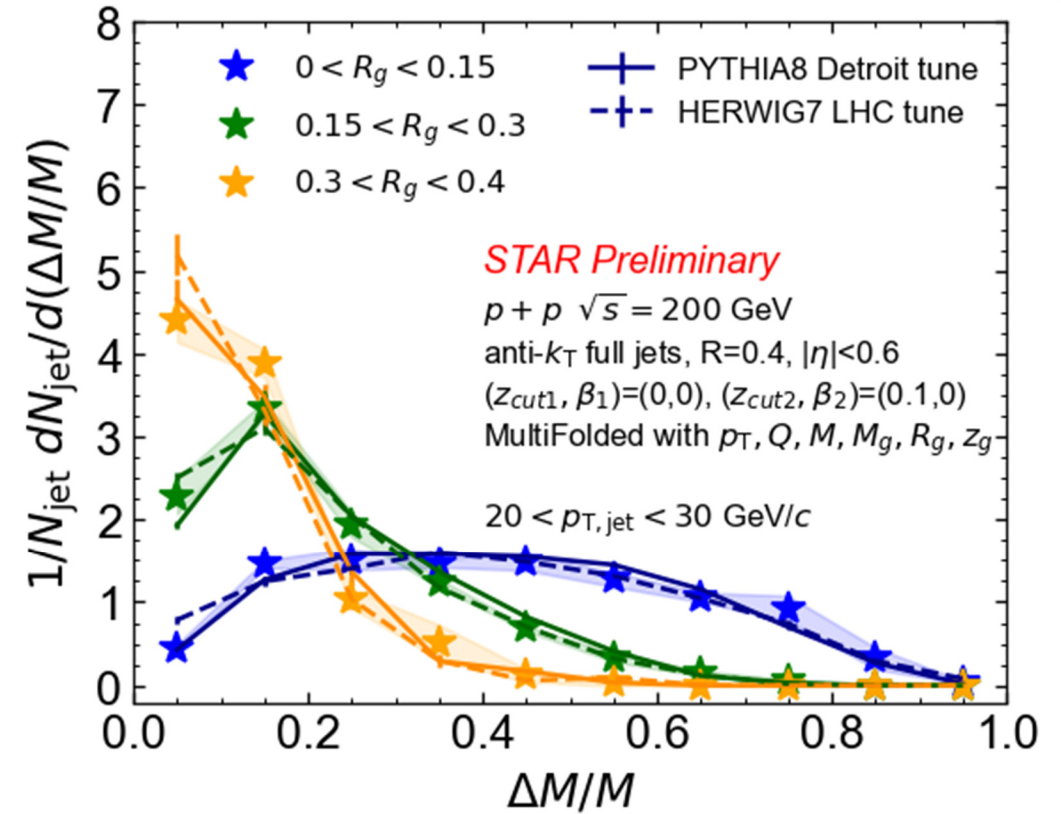
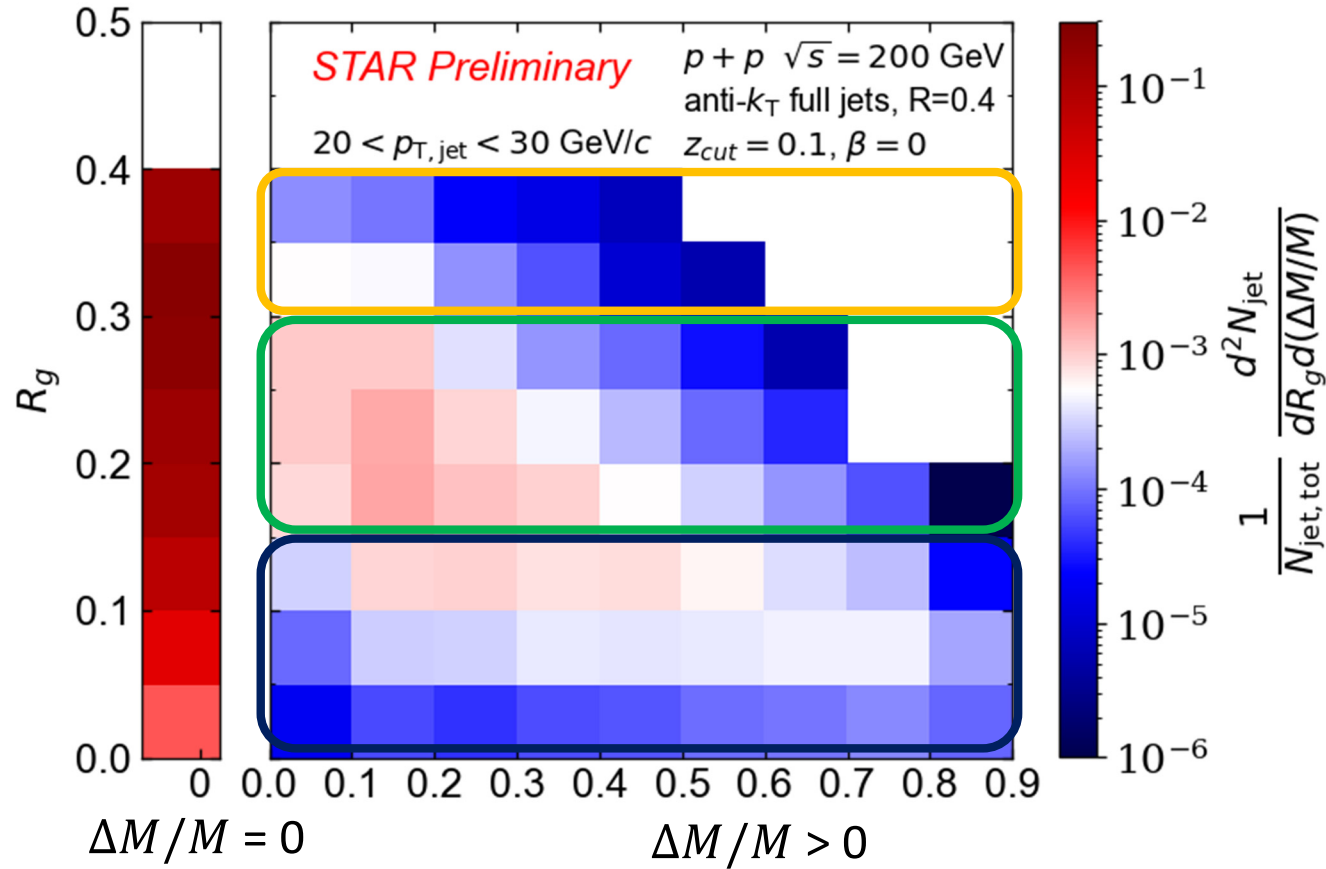
Soft radiation vs hard splitting angle



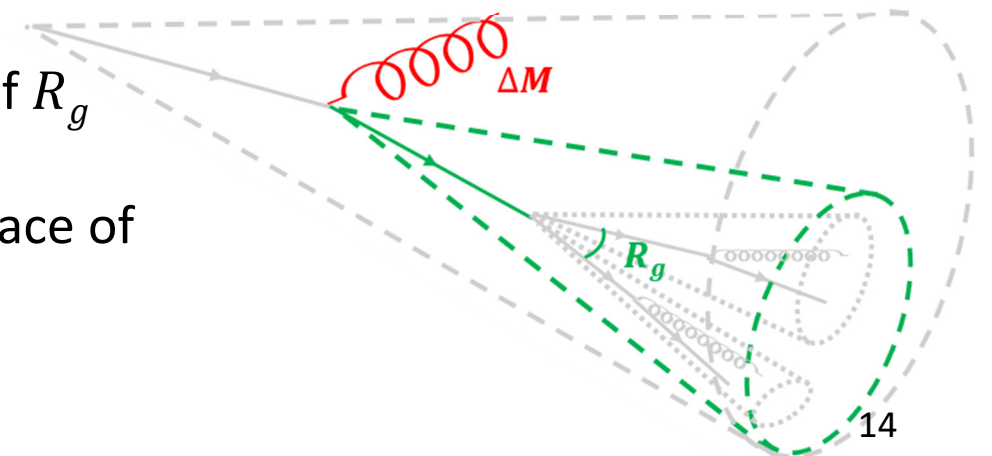
- The mean of $\Delta M/M$ distribution is anti-correlated with mean of R_g
→ consistent with angular ordered parton showers



Soft radiation vs hard splitting angle



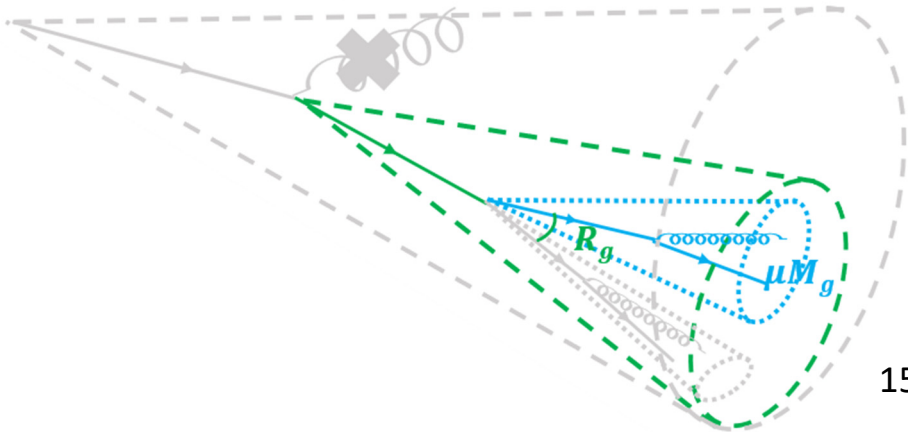
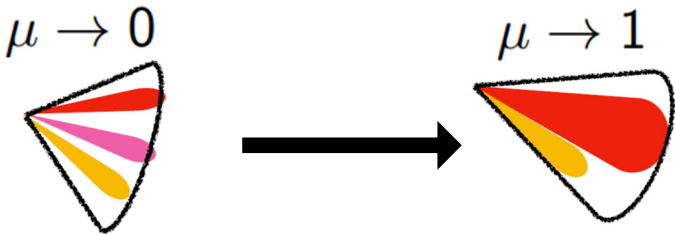
- The mean of $\Delta M/M$ distribution is anti-correlated with mean of R_g
 → consistent with angular ordered parton showers
- **Early** soft wide-angle radiation constrains the angular phase space of **later** splittings
- MC models describe the trend of data



Hard radiation vs hard splitting angle

μ	groomed mass fraction	$\mu \equiv \frac{\max(m_{j,1},m_{j,2})}{m_g}$
-------	------------------------------	--

Dasgupta M. et al, JHEP (2013) 2013:29

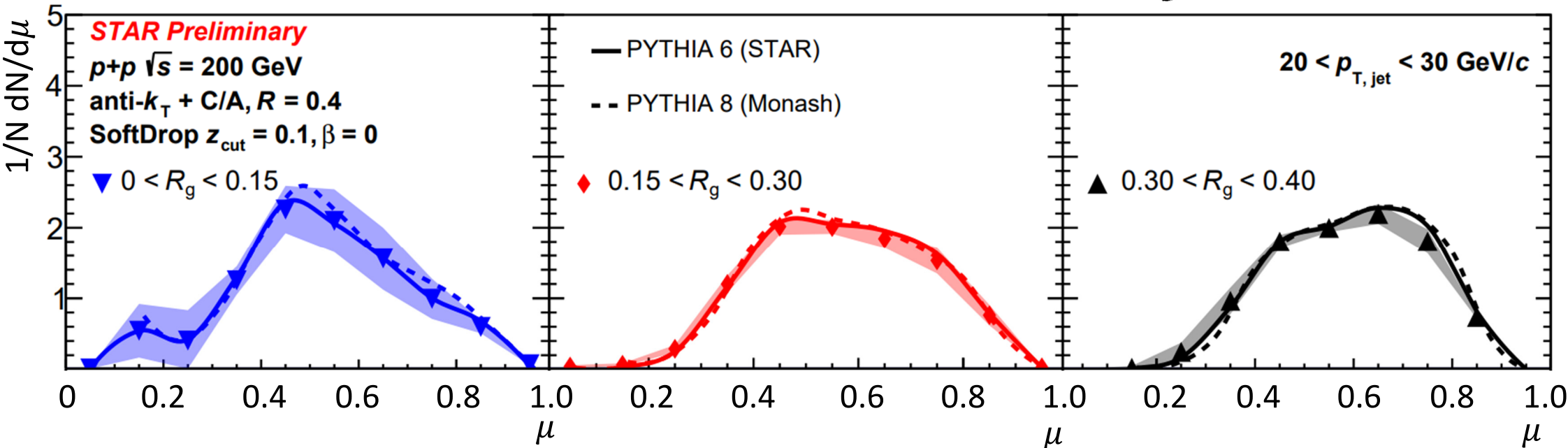
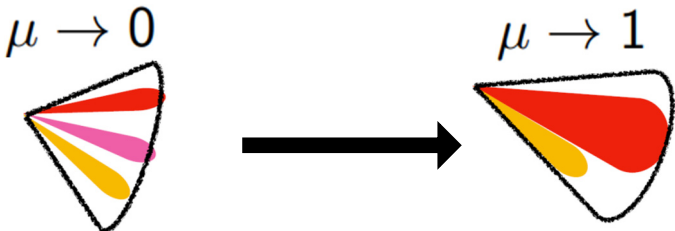


Hard radiation vs hard splitting angle

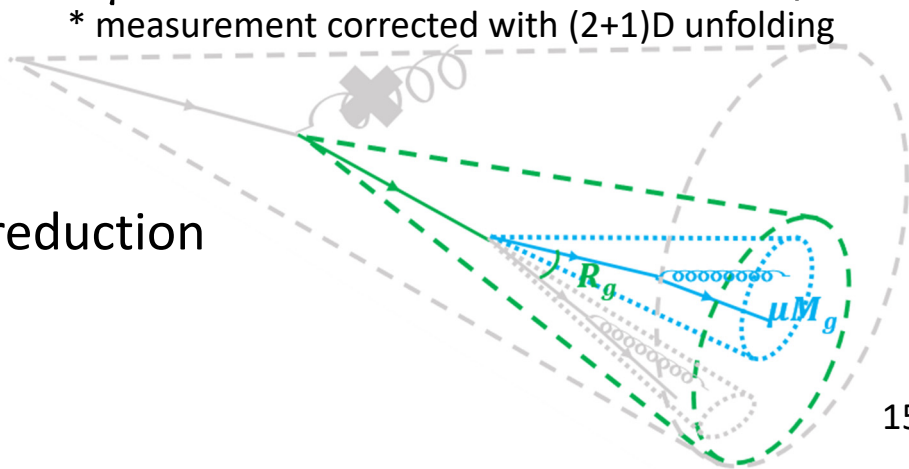


μ	groomed mass fraction	$\mu \equiv \frac{\max(m_{j,1}, m_{j,2})}{m_g}$
-------	------------------------------	---

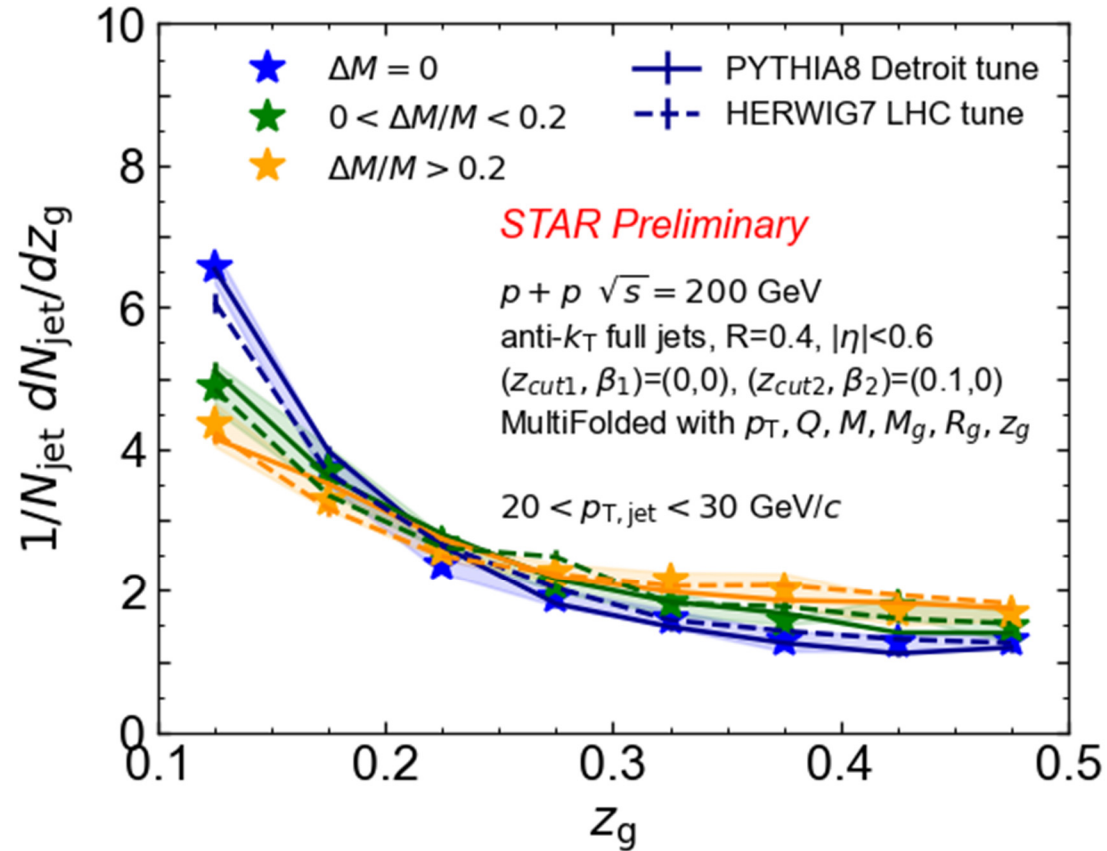
Dasgupta M. et al, JHEP (2013) 2013:29



- μ has a weaker dependence on R_g compared to $\Delta M/M$. MC models describe the trend of data.
- Shift of μ to smaller values at smaller R_g indicates a faster reduction of virtuality in the jet shower.

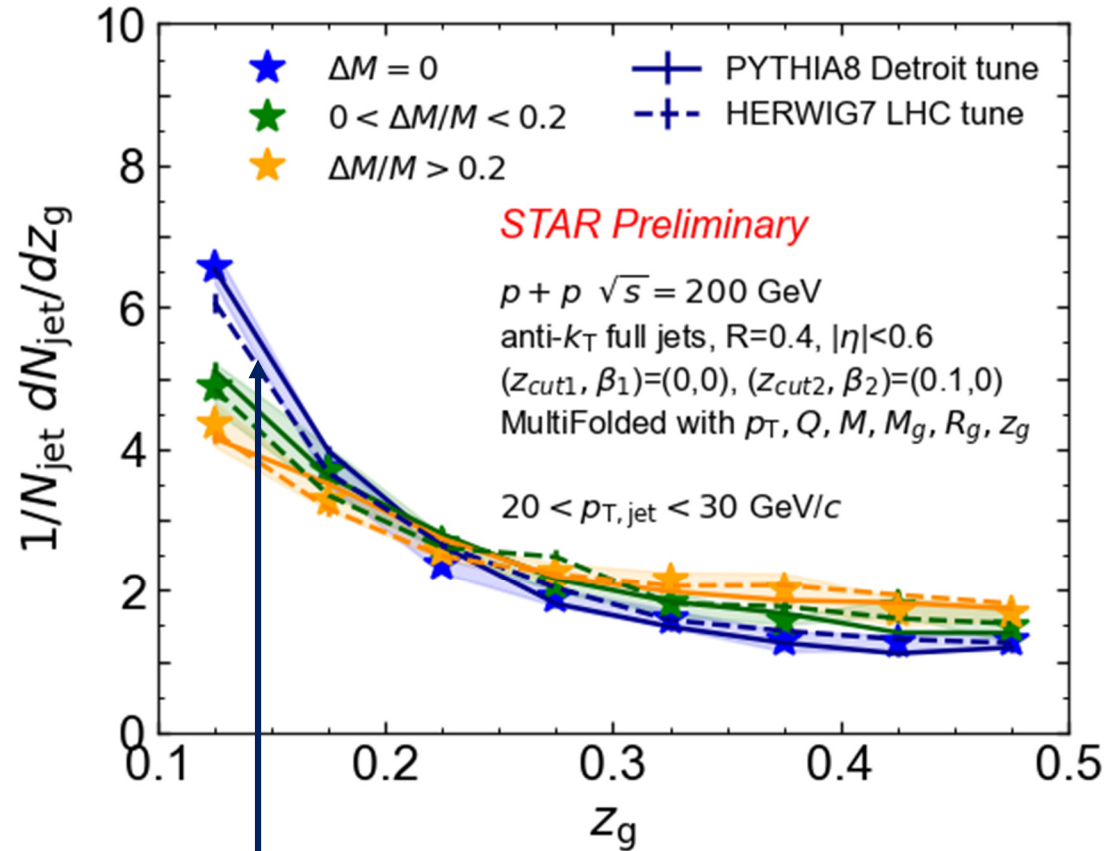


Soft radiation vs hard splitting momentum imbalance



- 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 & the amount of npQCD contributions to **later** splittings
- MC models describe the trend of data

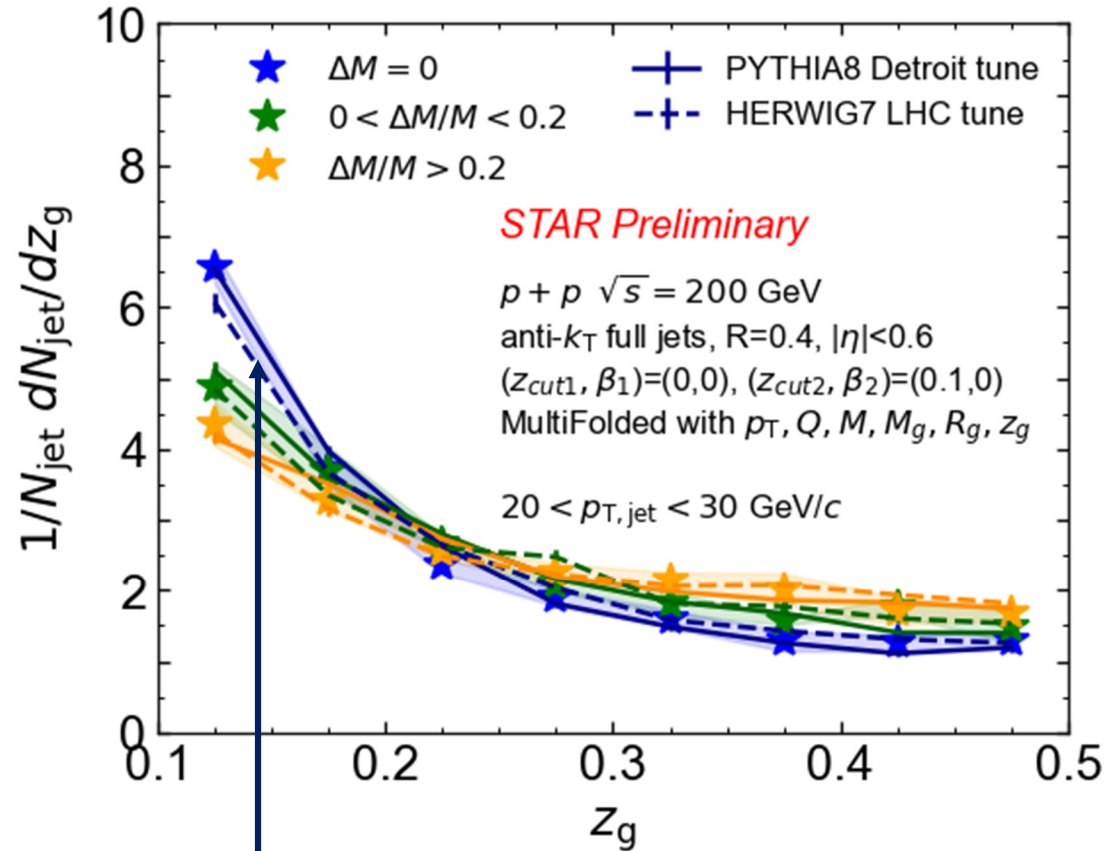
Soft radiation vs hard splitting momentum imbalance



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

- 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

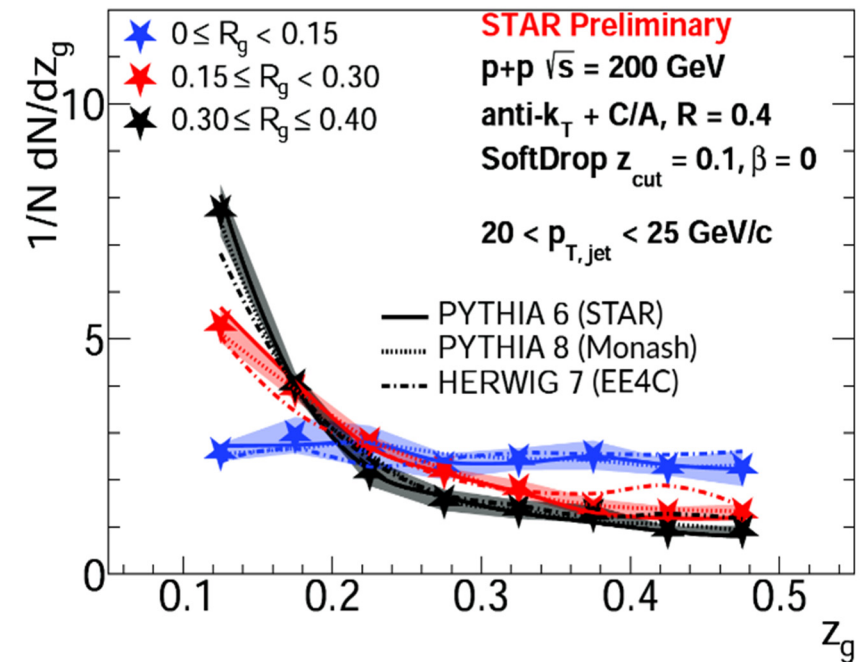
Soft radiation vs hard splitting momentum imbalance



Steeply falling $\sim \text{DGLAP } 1/z$: pQCD

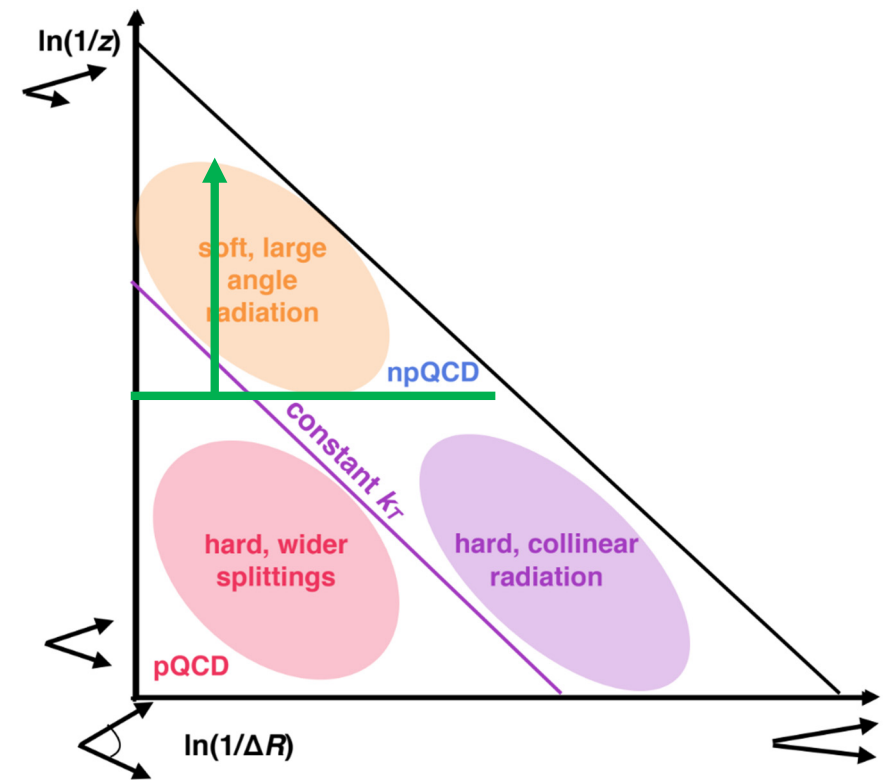
R_g and z_g are correlated, $\Delta M/M$ affects R_g
 \rightarrow correlation between $\Delta M/M$ and z_g

- 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



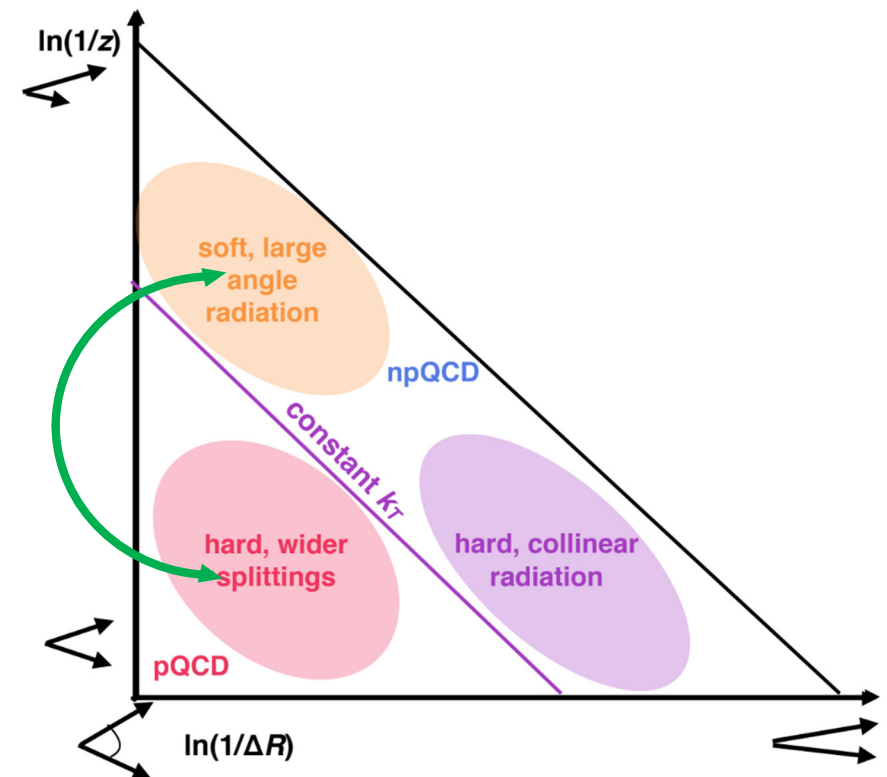
Conclusions

- Probing **soft** wide-angle radiation within jets
 - First fully corrected collinear drop jet measurement is presented



Conclusions

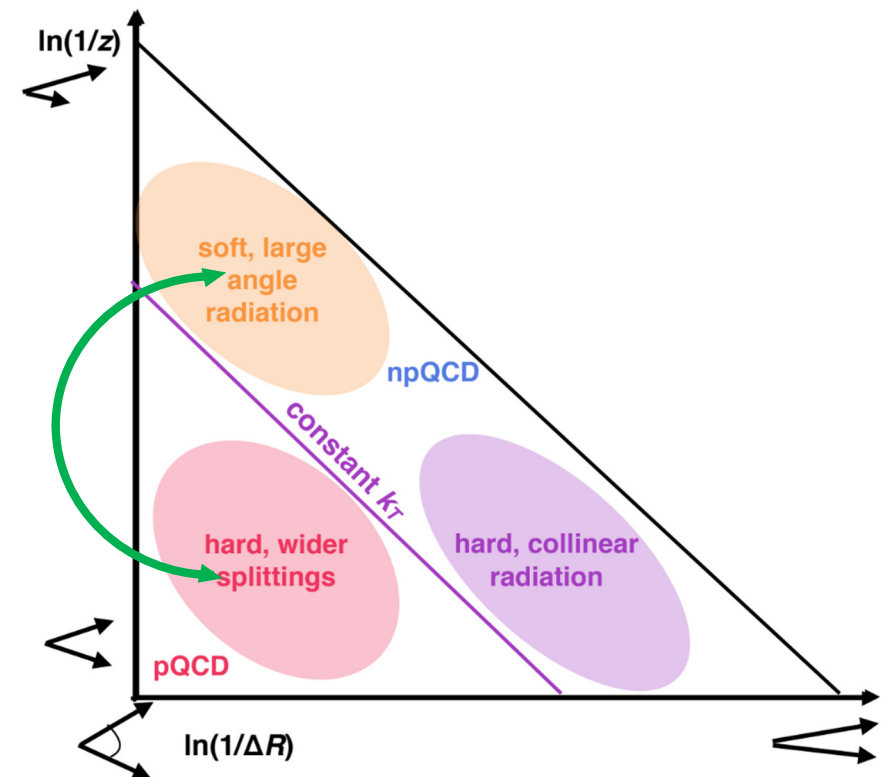
- Probing **soft** wide-angle radiation within jets
 - First fully corrected collinear drop jet measurement is presented
- Probing **soft-hard** correlation within jets
 - MultiFold allows for access of multi-dimensional correlations on a jet-by-jet basis. First application to RHIC data!
 - Jets with a large **perturbative** contribution (DGLAP 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



Conclusions

- Probing **soft** wide-angle radiation within jets
 - First fully corrected collinear drop 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 large **perturbative** contribution (DGLAP 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

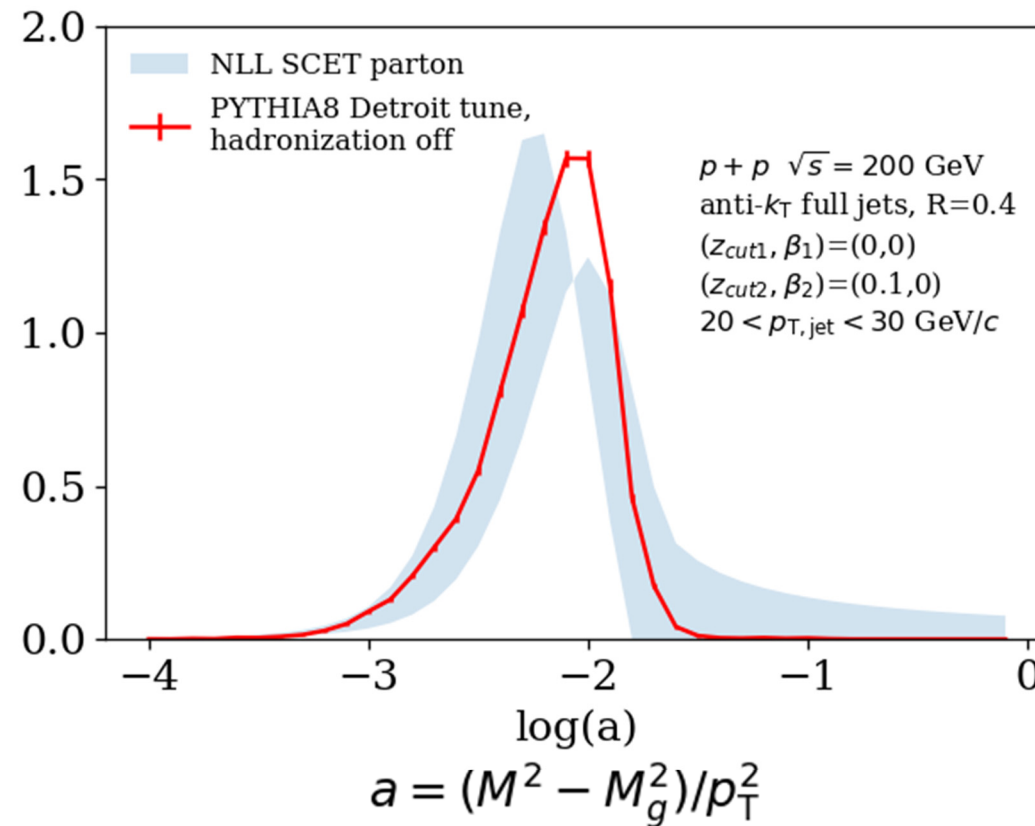
Improve understanding of jet substructure and the correlations between different substructure observables!



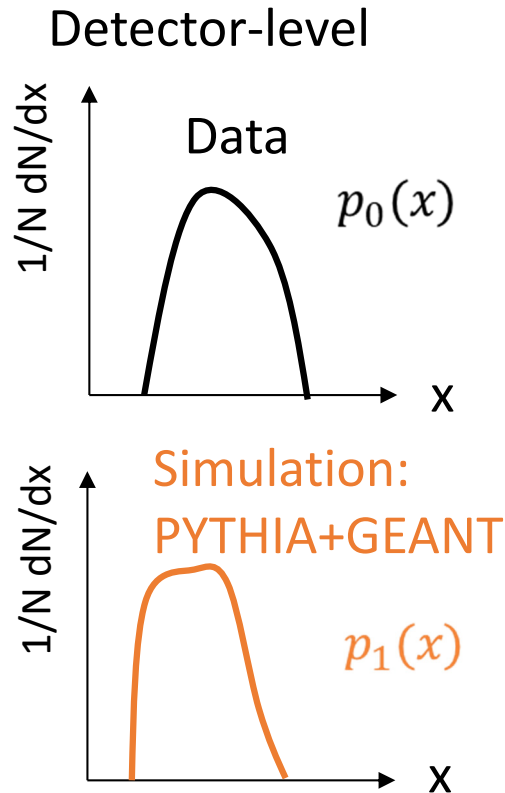
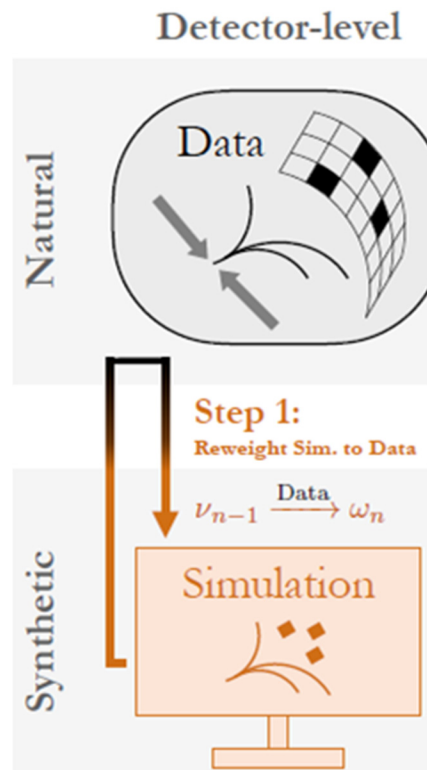
Backup

Collinear drop groomed jet mass

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



Method: machine learning



E.g., Iteration 1, step 1:

Weights: $w(x) = p_0(x)/p_1(x)$

Ok for 1D

$$\approx f(x)/(1 - f(x)) \quad \text{(\a Andreassen and Nachman PRD 101, 091901 (2020))}$$

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

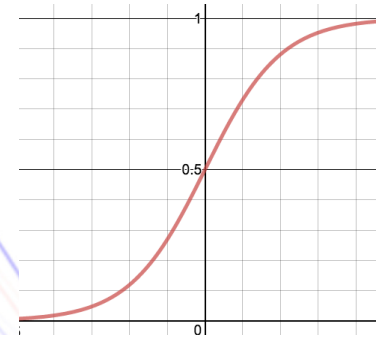
- Architecture: Dense neural network
- Activation function for dense layers: Rectified linear unit
- Activation function for output layer: Sigmoid
- Loss function: Binary cross entropy

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

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

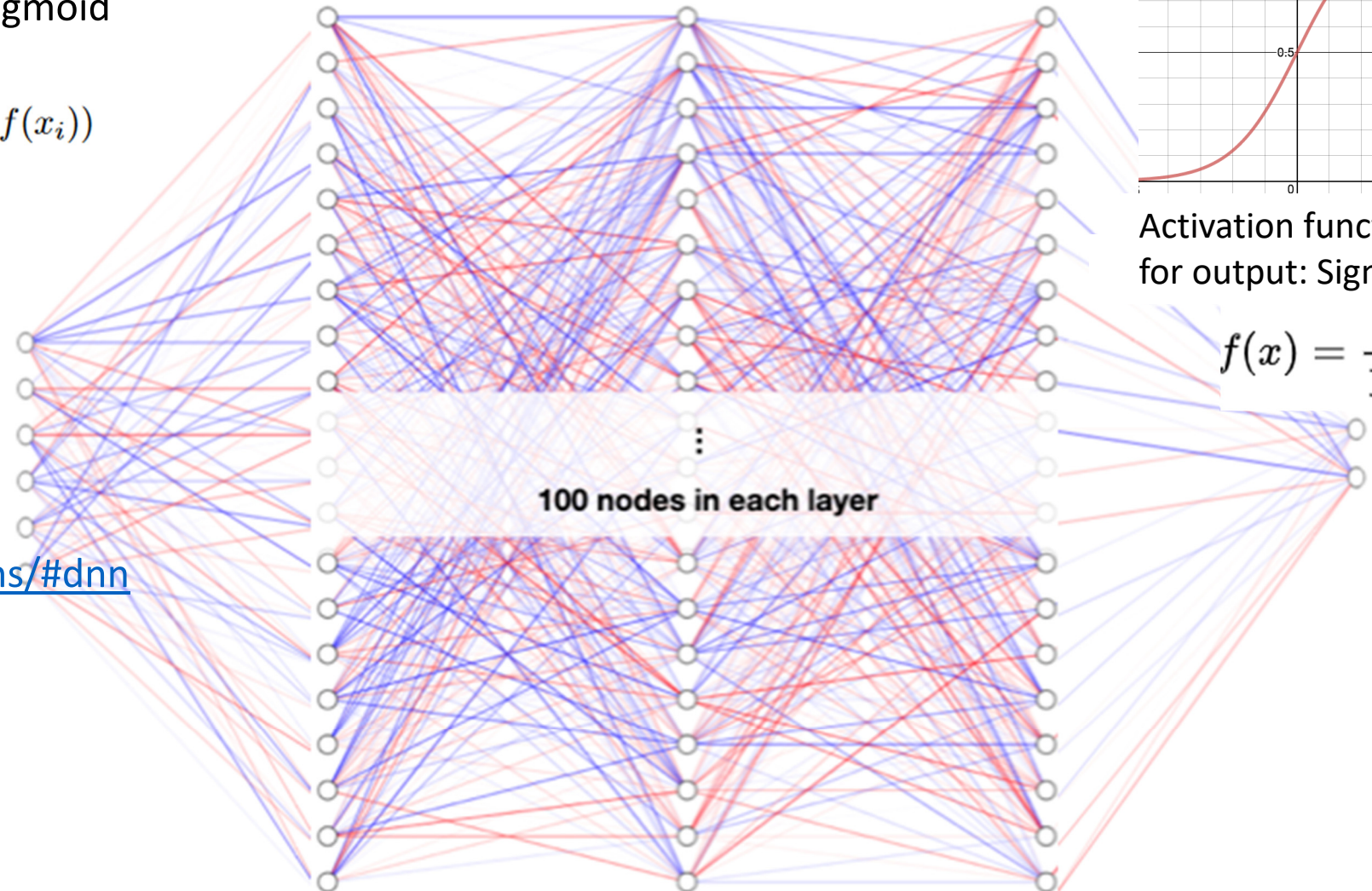
- All hyperparameters are default:
<https://energyflow.network/docs/archs/#dnn>

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



Activation function
for output: Sigmoid

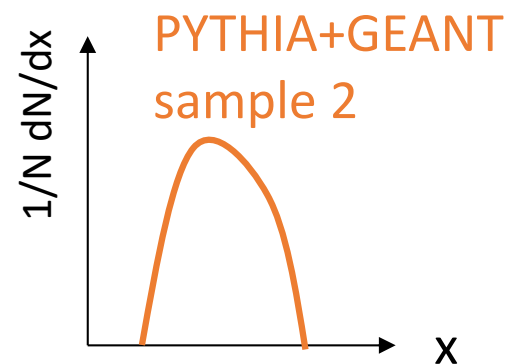
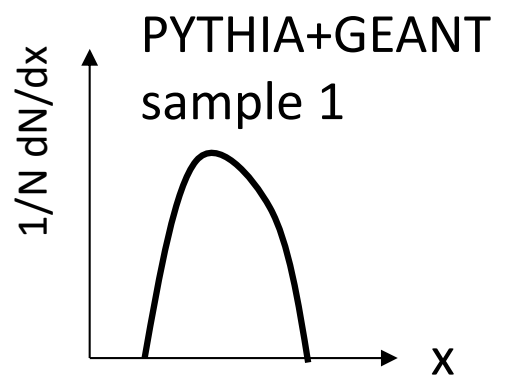
$$f(x) = \frac{1}{1 + e^{-x}}$$



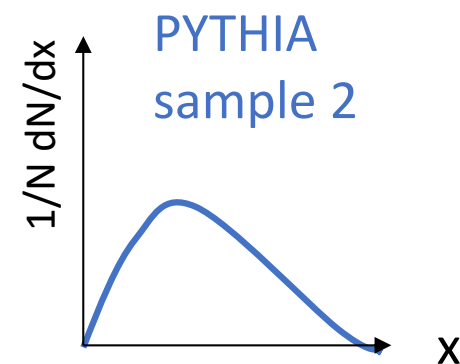
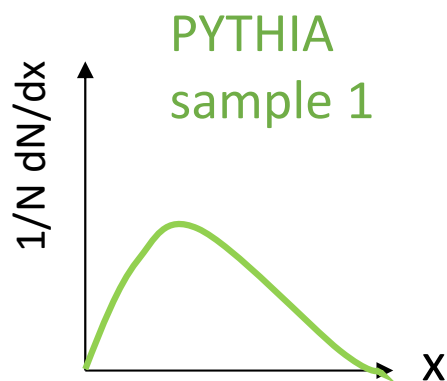
Closure test for unfolding

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

Detector-level

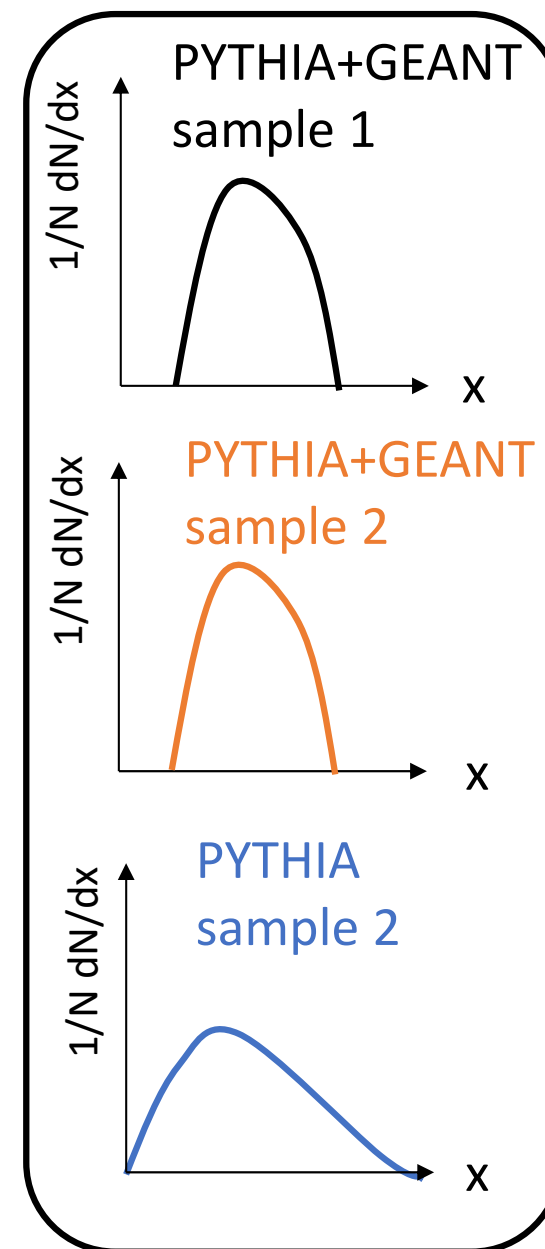


Particle-level

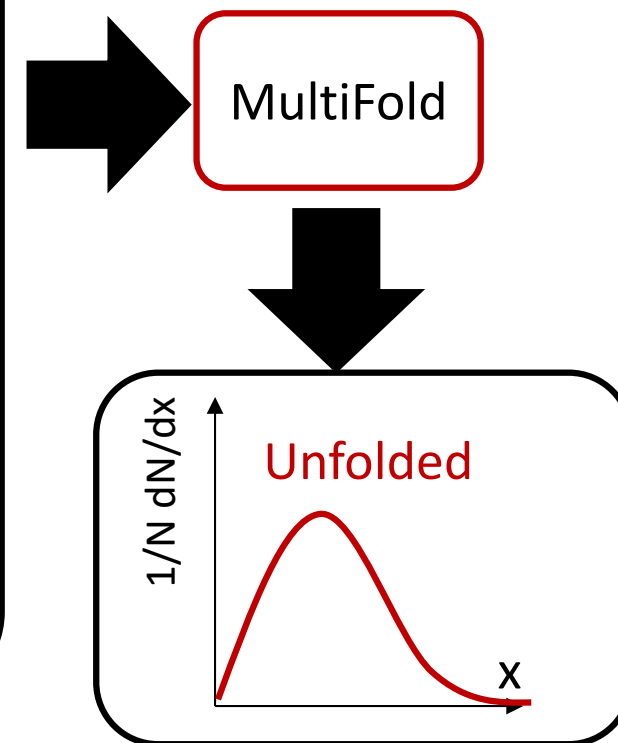


DIS, 3/28/2023

Youqi Song






- Step 2: Unfold

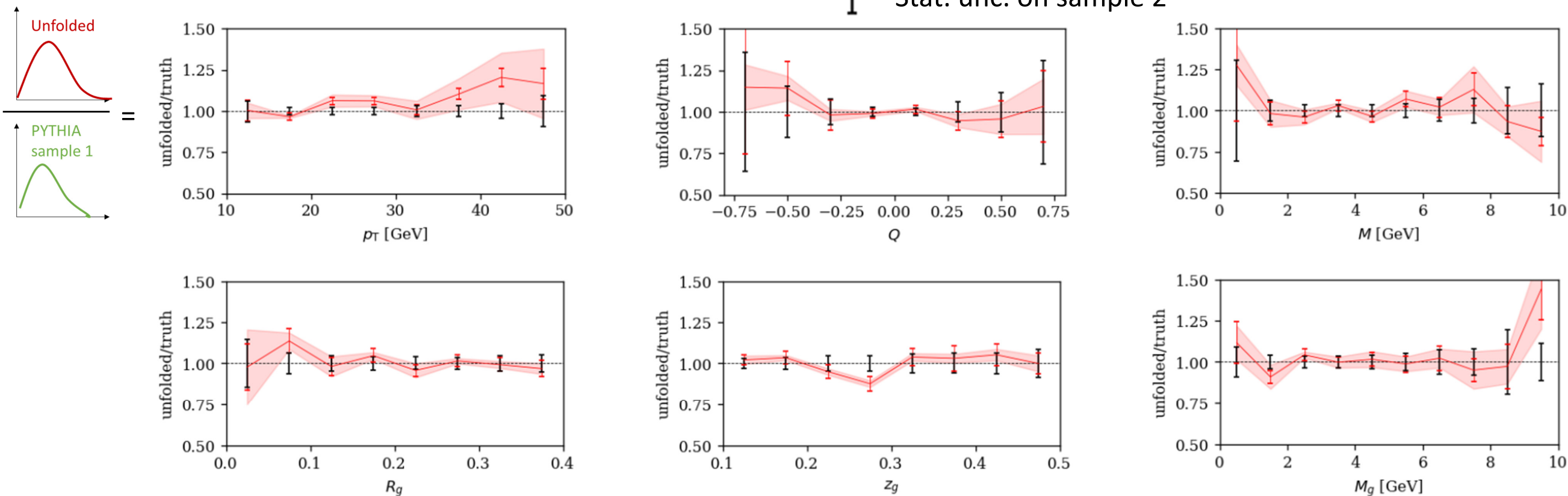


40

Closure test for unfolding

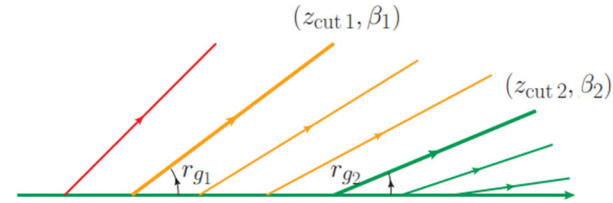
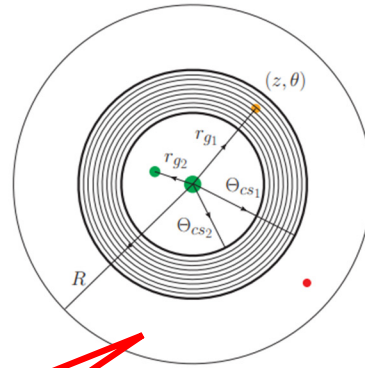
- Decent **closure** for all substructure observables

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

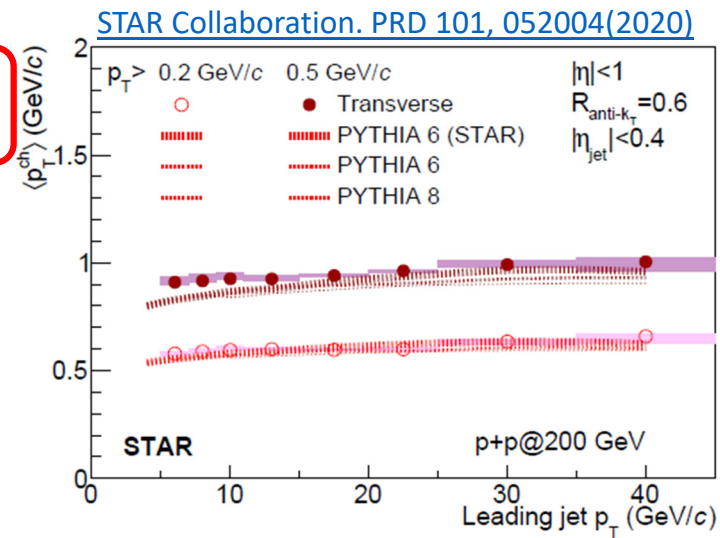


* 2D reweighting used for prior variation

Underlying event

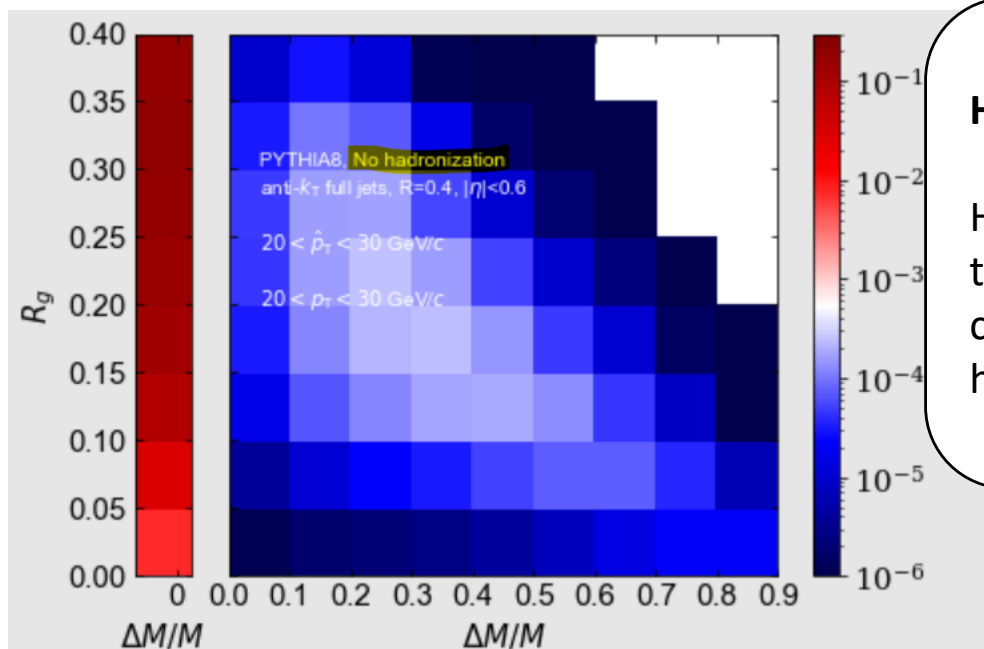
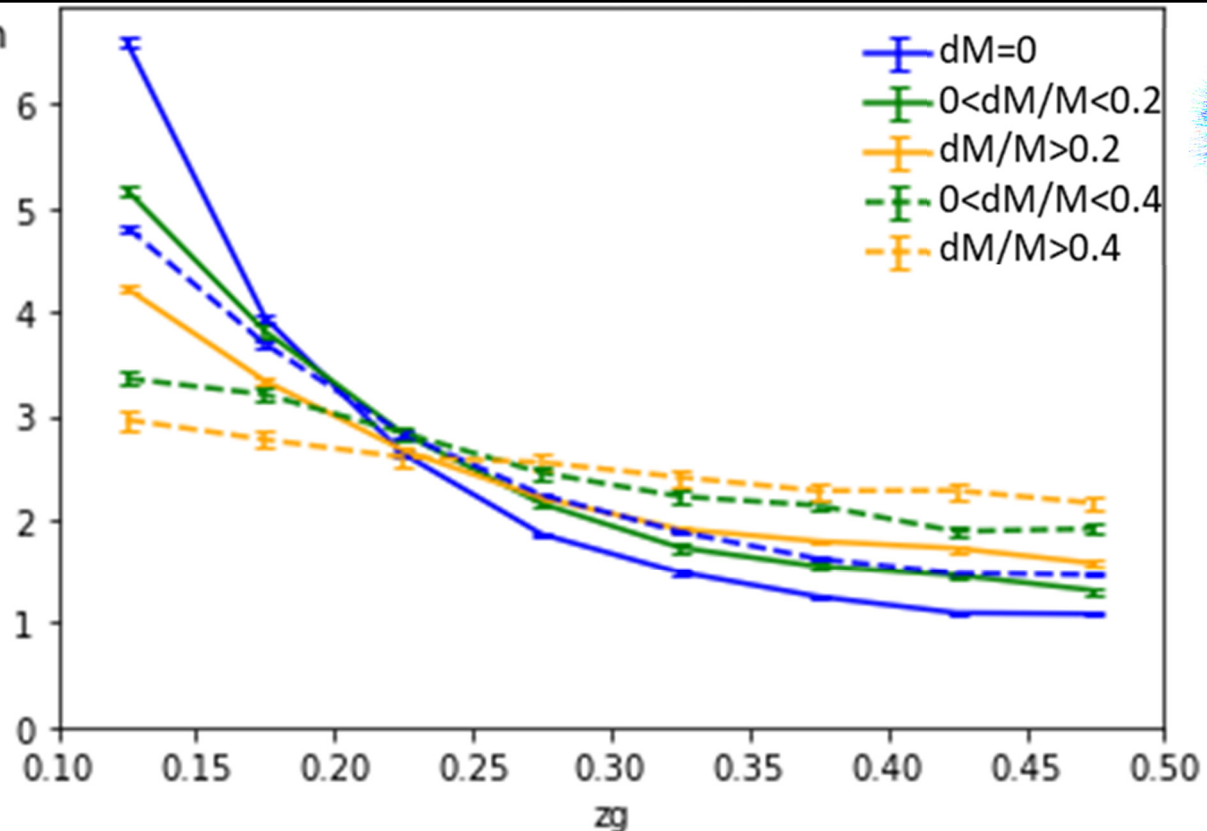
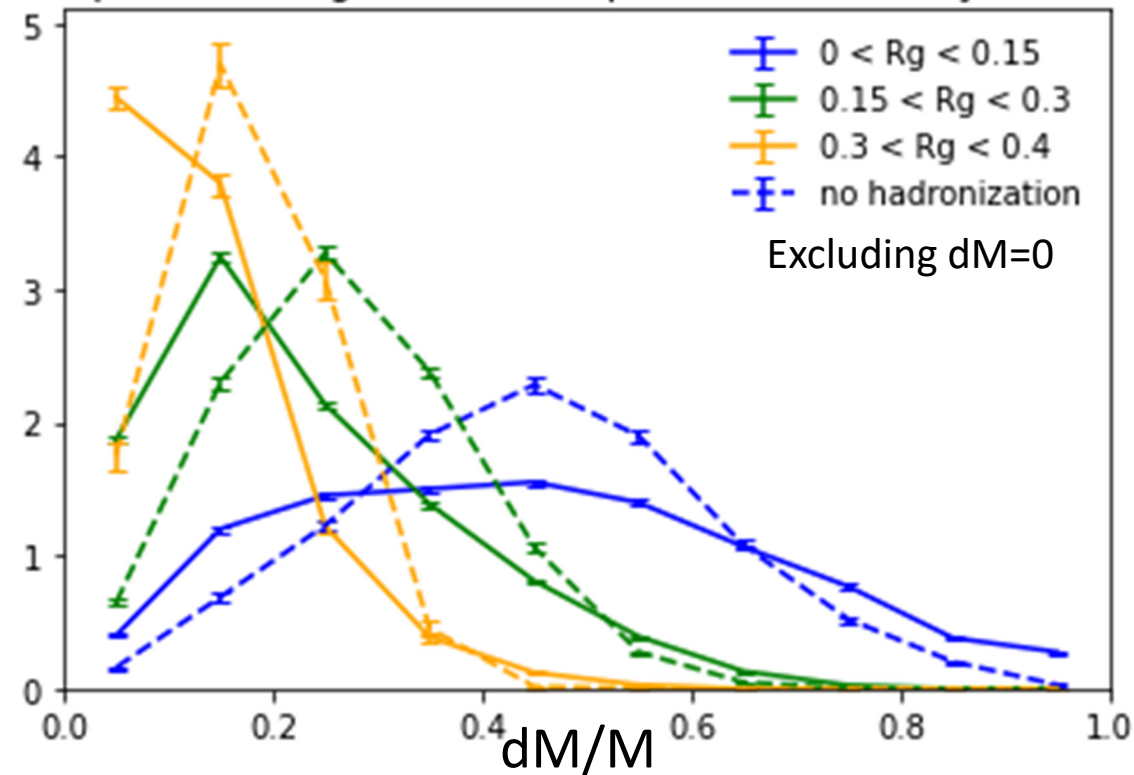


Remove underlying effects and pileup,
not necessary for 200 GeV collisions





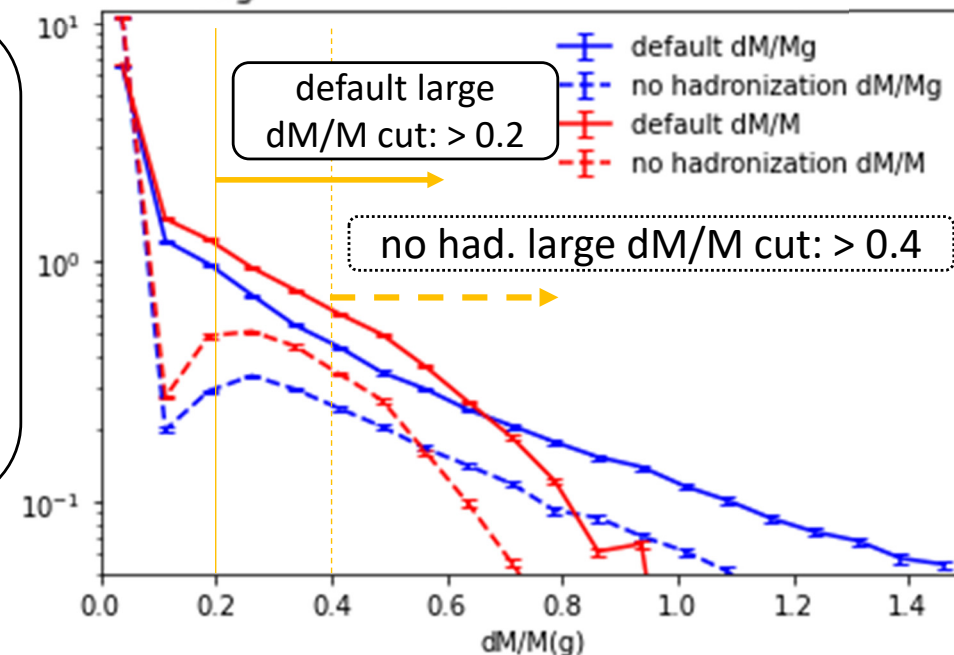
20 < pt < 30 and ng > 1 and 20 < pTHat < 30 GeV, Pythia truth



Hadronization effects study:

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

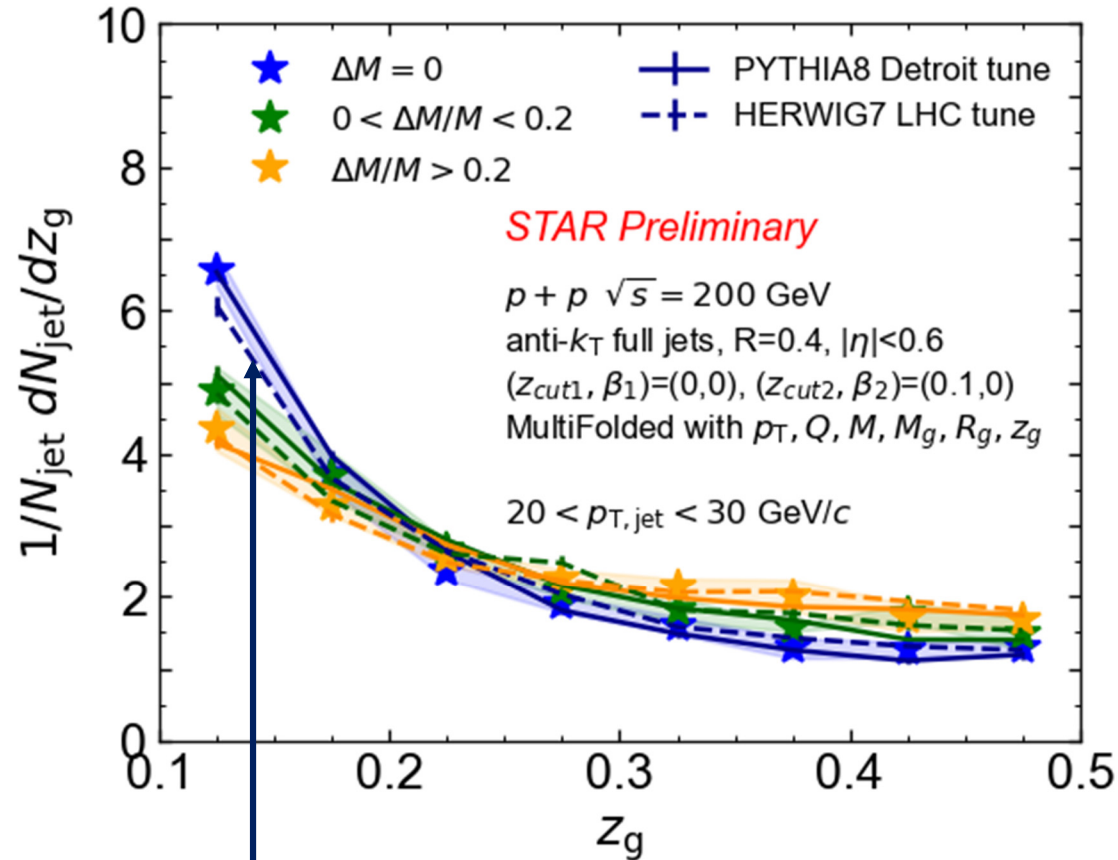
Youqi Song





- 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 p_T jets have a smaller fraction of $dM=0$).

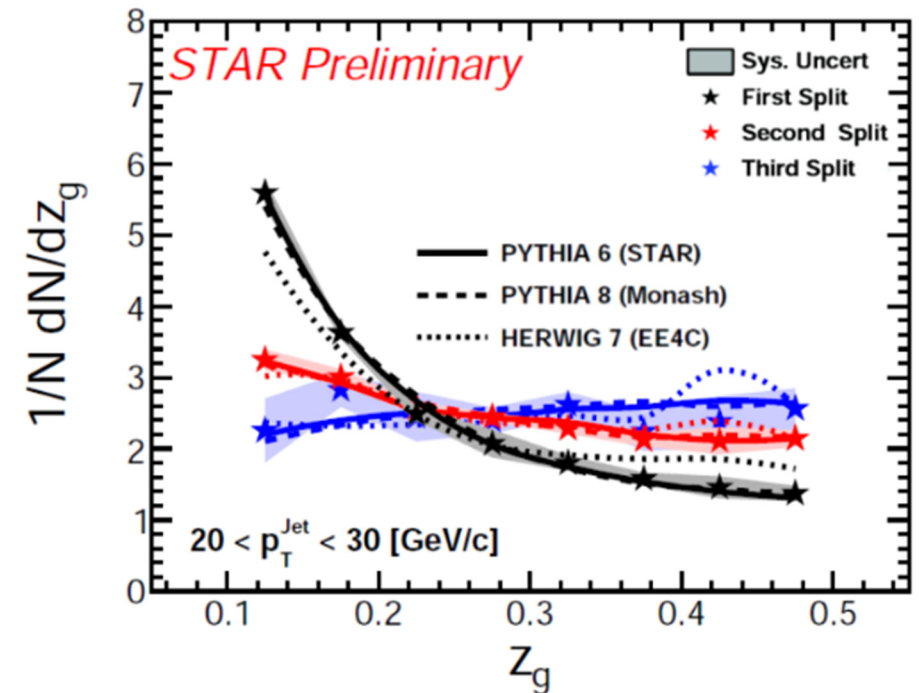
Soft radiation vs hard splitting momentum imbalance



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 $\text{reweight}(p_T, Q, M, M_g, R_g, z_g) = \frac{\text{Herwig truth}(p_T, Q, M, M_g, R_g, z_g)}{\text{Pythia6 truth}(p_T, Q, M, M_g, R_g, z_g)}$

