JET SUBSTRUCTURE MEASUREMENTS IN PROTON-PROTON AND PROTON-GOLD COLLISIONS

by

ISAAC ARTHUR MOONEY

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2022

MAJOR: PHYSICS

Approved By:

Advisor Date

DEDICATION

To my mom,

who never discouraged or doubted me,

and didn't blink when her eight-year-old brought home The Universe in a Nutshell.

L.Y.L.Y.M.I.W.

ACKNOWLEDGMENTS

A PhD is basically a half-decade long apprenticeship. With that in mind, I first need to acknowledge my advisor, Professor Joern Putschke, whose behavior I have begun to consciously and unconsciously emulate as I mature from the 20-year-old he first met in 2014. His attention to detail, his wariness of overreach when making claims, and his preference for concise, unadorned language are all part of a healthy scientific mentality that I hope I can say I've improved on with his guidance (setting aside this flowery acknowledgement). However, perhaps at least as important has been his modeling of good mentorship including his flexibility, his even temper, and his concern for the personal aspects of his students' lives. I hope that the initial conditions of my career trajectory are strong, so that I can continue to model this good scientific and interpersonal behavior wherever I go. Additionally, I would like to thank the members of my dissertation committee: Professors Abhijit Majumder, Sergei Voloshin, and Christine Aidala who have, all-together, taught me six courses and much more physics besides. I especially thank Christine, my undergraduate advisor, for teaching me to be a considerate collaborator, and remaining a friend and mentor to me as I stayed in the area for this PhD.

That said, this dissertation would certainly not have been possible without a whole host of others who are too numerous to individually name. Obviously, as it relies on experimental data from the STAR Collaboration, I am indebted to the scientists, engineers, and technicians who designed, constructed, and maintain the Relativistic Heavy Ion Collider and the STAR detector and the physicists who collected data during long shifts in the control room; the computer scientists who ensure data production and accessibility, as well as the WSU High Performance Computing team; the Brookhaven National Laboratory administrative staff; and the funding agencies, domestic and otherwise, which support all of these operations and my own research.

The analyses presented here would have been poorer without competent colleagues both in the Jet-like Correlations Physics Working Group and in the larger collaboration, as well as in the Relativistic Heavy-Ion Group at WSU. This is especially true of the post-doctoral research fellows in Joern's group, Dr. Kolja Kauder and Dr. Raghav Kunnawalkam Elayavalli. If they had only tutored me in the relevant physics concepts and analysis techniques and handed me codebases in need of only slight adaptation for the problems at hand, that would have been enough. But as it happens, both of them taught me also to be brave: to ask questions without fear of embarrassment, to make friends in the field, and to apply for positions and tasks that might seem difficult. They are and will continue to be older brothers to me.

I also acknowledge some more informal mentors. I especially thank my Gandalf the Grey, Dr. Nick Elsey, for being a constant support, even after he left the field. I have already asked more questions than anyone should be expected to answer, but he has never intimated that he is annoyed at being asked. And Dr. Bryan Ramson, who pushes me to imagine what could be, and who has been simultaneously a mentor and best friend to me for eight years.

Family and friends who may not have directly contributed to this work, but who kept me sane for these six years, also deserve recognition. I sincerely include in the latter category my graduate student colleagues Veronica Verkest, and Grant McNamara, and our newest post-doctoral fellows Dr. Matt Kelsey and Dr. David Stewart, and thank them for productive physics discussions and for editing talks and papers. I would also like to thank the friends from around the world that I made during shifts at STAR and at conferences, as well as Adam, Amy, Caroline, Chris, Colleen, Emily, Jack, Karl, Pranav, Savannah, Shannon, Tayla, Tyler 1, Tyler 2, Vikesh, and Zach for their continued friendship and help in taking my mind off of physics. I was sometimes distant from my family because of my work, but they were always understanding, and were (and have been since my birth) vital to my success as a person and a researcher. Especially I thank my grandparents, Jim, Sue, and William, and my parents, Jo and Tim. I also thank my partner's family, Antonio, Barb, Ben, Joel, Paul, Simone, and Susan, for treating me like their own.

Finally, I thank my partner, Freida Blostein. For the past six years, while also completing her graduate studies in epidemiology, Freida has been my best friend, the other member of our own two-person book club, my roommate, my co-dog-parent, and my love. She has taught me to see the beauty in endeavors outside the physical sciences, and made me laugh and cry. She has shown me how to plan, to consider others, and to recognize biases in observational research and in myself. She has listened to me practice talks and given feedback, read through papers and abstracts, helped discuss quandaries in my research, and taken care of our dog while I take shifts at STAR. She encourages me to travel, to make room for change, to see art and do things. She is the two-body problem I am glad to have.

∼ *Isaac Mooney*

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

“O God, I could be bounded in a nutſhell, and
count my ſelfe a King of infinite ſpace; were it not that
I haue bad dreames.”

– Hamlet, *The Tragedie of Hamlet, Prince of Denmarke*

Our understanding of the physical world has advanced greatly in the past few centuries through reductionism and unification; by postulating that complicated phenomena are actually the result of simple component actions and interactions, it is possible both to make predictions about similar scenarios as well as generalizations that encompass seemingly disparate systems. As Richard Feynman said [1], “For example, the phenomena of sound could be completely understood as the motion of atoms in the air. So sound was no longer considered something in addition to motion.” This reductionism is possible in each case through the creation of a scientific model which, quoting John von Neumann’s formulation [2], is “a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena.”

The Standard Model of particle physics (Fig. 1.1) is the framework by which we currently understand the nature and interactions of elementary particles. The statement “all models are wrong but some are useful” [4] applies to this model as well – there are still questions of internal consistency and external validity [5–8], but nonetheless the Standard Model describes all of the fundamental forces that are relevant to current experimental resolution at the femtoscale, as well as the properties of all known elementary particles to a high precision. It is also predictive, with experiments later confirming, for example, the existence of a number of particles included in the model, including most recently the Higgs boson, fifty years after it was proposed [9].

1.1 Quantum chromodynamics (QCD)

The mathematical construct of the Standard Model is quantum field theory, which models both matter and fundamental forces as quantizations of fields. There are three quan-

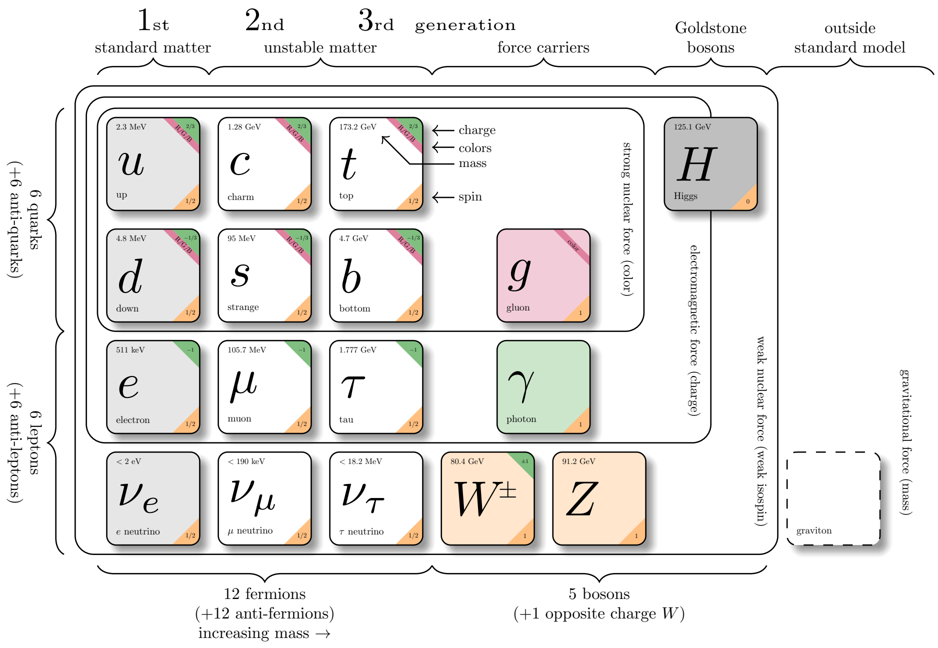


Figure 1.1: The Standard Model of particle physics. [3]

tized interactions: electromagnetic, weak, and strong, with four corresponding gauge bosons, which act as force carriers. This dissertation focuses on the strong interaction, mediated by the gluon, which binds the six quarks and their antiquarks into composite particles called hadrons. Among these hadrons are the protons (uud) and neutrons (udd) that make up the atomic nuclei in everyday elements.

The quantum field theory which describes the strong interaction is called quantum chromodynamics (QCD). This interaction is due to the color charge carried by the gluon and the quarks, which is analogous to the well-known electric charge of quantum electrodynamics (QED). Because the gluon interacts with itself, the theory is non-abelian, which complicates calculations. The gauge theory of QCD, which describes exact local symmetries of quarks and gluons (collectively, partons) under color interchange, admits three colors for quarks and eight color combinations for gluons. A ninth, color-singlet, free gluon is theoretically possible but experimentally ruled out. As for quarks, only the color-singlet state is possible, disallowing observation of individual quarks. Instead, in a QCD phenomenon called confinement,

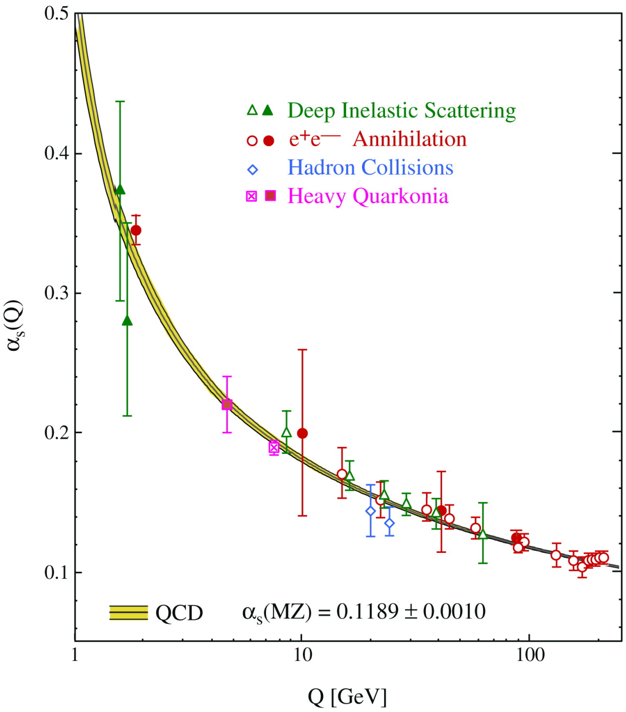


Figure 1.2: The running of the strong coupling, α_s, with momentum transfer Q, comparing QCD prediction (solid lines) to data (markers). [10]

the quarks can only be observed in hadronic bound states.

Another salient feature of QCD (and of all non-abelian gauge theories) which manifests at small color charge separations is called asymptotic freedom [11,12], resulting from the running of the strong coupling with the scale (Fig. 1.2). This alludes to the fact that the strength of interaction between charges approaches zero as energy goes to infinity or equivalently as separation goes to zero. As a consequence, calculations in the high-energy regime using perturbation theory are highly accurate (Fig. 1.2). On the other hand, perturbative methods are not applicable at low momentum transfer, where $\alpha_s \sim 1$. In this regime, the most successful approach has been lattice QCD [13] which regularizes the theory in the ultraviolet and infrared limits using a spacetime lattice with finite site spacing and volume, respectively, which allows the full path integrals to be evaluated numerically. Lattice QCD is quite computationally intensive, but has made many accurate predictions (see e.g. hadron spectra results in Fig. 15.9 in Ref. [14]).

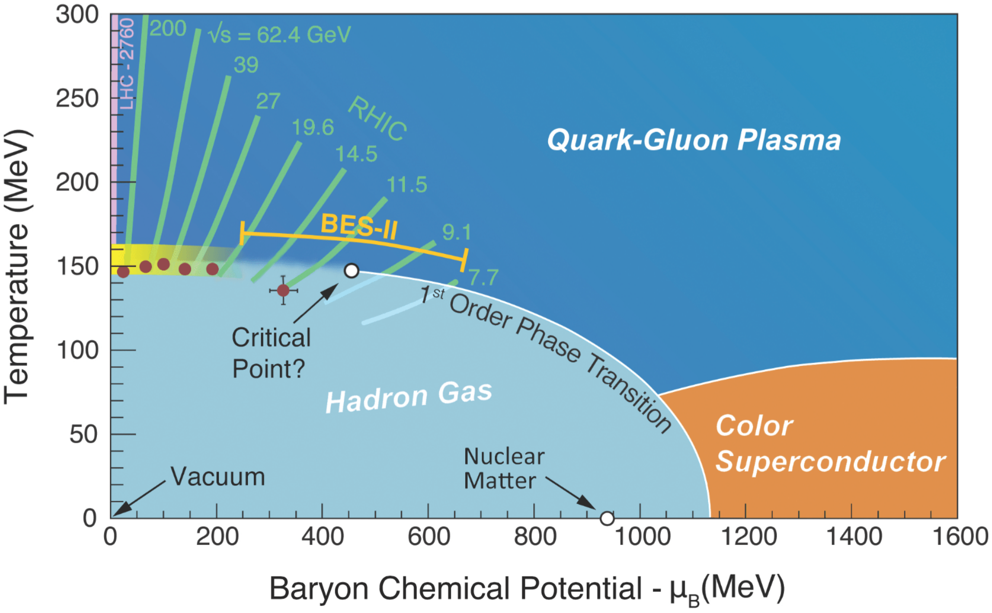


Figure 1.3: A sketch of the phase diagram of QCD matter as a function of temperature and baryon chemical potential (the abundance of quarks relative to antiquarks). [15]

1.2 Emergent properties of QCD

Despite QCD's color confinement, it is possible to form quasi-free quark matter states briefly if the system is heated above the binding energy of the quarks. At low density this temperature, called the Hagedorn temperature, is on the order of one trillion Kelvin (or roughly equivalent in energy to 150 million electron-volts, or MeV[1]) [16]. In order to provide this energy, modern particle colliders accelerate heavy ions such as lead or gold to greater than 99% the speed of light and collide them with energy per nucleon pair in the zero-momentum frame on the order of between 10^{11} and 10^{13} eV. The result in some of these collisions is the transition of the system from normal hadronic matter to a hot, strongly interacting ideal fluid with quark and gluon degrees of freedom called the quark-gluon plasma (QGP) (Fig. 1.3).

As this new state of matter exists for only roughly $10^{-24} - 10^{-23}\,\mathrm{s}$ $(1 - 10\,\mathrm{fm/c})$ [17] before its expansion and cooling causes a smooth crossover back to a hadronic state, it may

---

[1] For a definition of an electron-volt, and other collider physics units and variables used throughout this chapter, consult Appendix A.

initially seem impossible to study it. However, the collective dynamics of the plasma are imprinted on the collision products, for example in the characteristic azimuthal distributions of measured particles, due to the momentum anisotropy which is carried through the fluid from the initial spatial anisotropy of the colliding nuclei [18–20]. In rare cases, there are also partonic scatterings with large momentum transfer, Q^2, (called "hard" scatterings) which occur before the QGP hydrodynamizes at roughly $1\,\mathrm{fm/c}$ [21]. Since the scattered partons are colored, the colored thermal medium is opaque to them and they will therefore interact with the QGP while traversing a significant portion of it, acting like a tomographic probe of its microscopic structure.

1.3 Validation of jets as a tomographic probe of strongly interacting media

The high energy partons produced in a hard-scattering event are also highly virtual, or off-shell. This is the extent to which the relativistic energy-momentum relation ($E^2 = \mathbf{p}^2 + m^2$, corresponding to a hyperboloid or "shell" in energy-momentum space) is violated. Observable states must have zero virtuality, so the parent partons will subsequently evolve to lower virtuality mostly by radiating soft and collinear gluons, in what is called a "parton shower". Finally, the daughters will all hadronize before being detected as a spray of collinear particles called a jet.

Before jets can be used as a tool to study the microscopic structure of the QGP, there are a few considerations. First, they must be calculable. Given that the jets begin and end in hadrons, which are inherently non-perturbative (NP) objects, it is required that there is not much interplay between the long-distance and short-distance physics. With the large partonic momentum transfers, Q^2, possible in the collisions mentioned above, there indeed is an approximate separation of scales, called collinear factorization. Intuitively, the collision occurs over a small time scale in which a parton appears frozen at a given longitudinal momentum fraction ("Bjorken-x"), $x \approx Q^2/2q \cdot p$, of the parent hadron with momentum p, with small transverse momentum contributions. For an excellent review of factorization theorems with both physical motivation and proof, see Ref. [22]. Consider as a simple

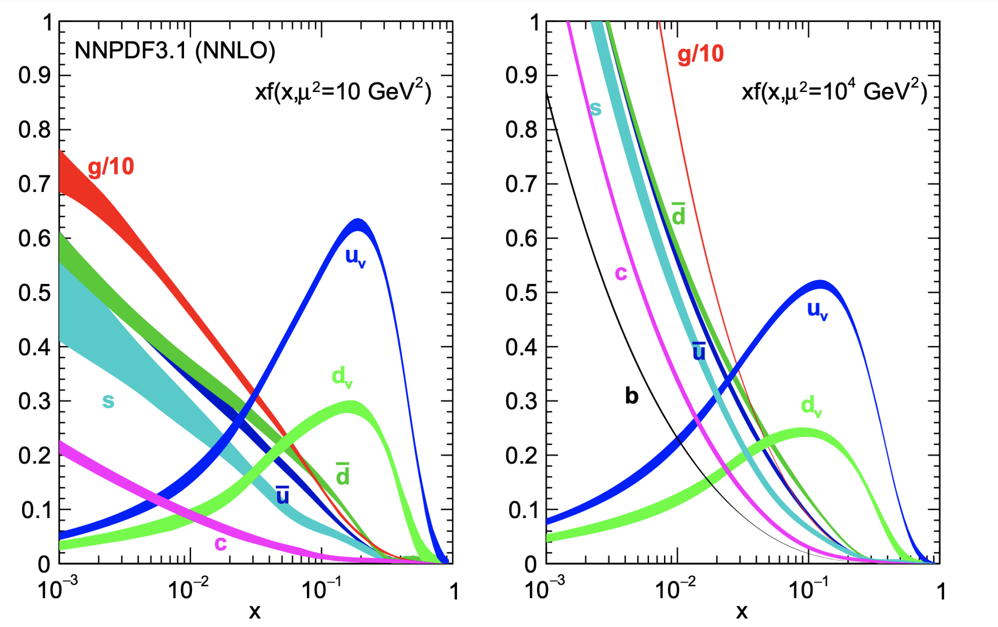


Figure 1.4: Example parton distribution functions at next-to-next-to-leading-order from NNPDF3.1 for two example scales, μ. Note that the gluon PDF has been scaled down by a factor of 10. [24]

example the factorized cross section of a hard proton-proton (pp) scattering:

$$\mathrm{d}\sigma_{\mathrm{pp}\to X} = \sum_{i,j} \int \mathrm{d}x\, \mathrm{d}x' f_i(x,\mu_f) f_j(x',\mu_f) \otimes \mathrm{d}\hat{\sigma}_{ij\to X}(x,x',\mu_f,\mu_r,\alpha_s(\mu_r)). \tag{1.1}$$

This cross section has now been expressed as a convolution of a process-dependent hard partonic cross section, $\hat{\sigma}$, and non-perturbative parton distribution functions (PDFs) (Fig. 1.4), $f_{i,j}$, which act as probabilities of finding partons with such momentum fractions. The scales μ_f and μ_r are the factorization and renormalization scales, respectively. The former separates the short-distance and long-distance physics and is somewhat arbitrary. In calculations, it is typically varied by some amount as an estimate of the uncertainty of the result [23].

The PDFs mentioned above are universal, and are typically parametrized and determined from global fits to collider data, using any process that contains a nucleon in the initial state (see Ref. [14], Table 18.1 for examples). Although $f_i(x,Q_0)$ for some momentum scale Q_0 is not calculable in perturbative QCD, its logarithmic evolution with Q can be determined in perturbation theory using the coupled DGLAP evolution equations [25–27]:

$$Q^2 \frac{\partial}{\partial Q^2} \begin{pmatrix} f_{q_i}(x, Q^2) \\ f_g(x, Q^2) \end{pmatrix} \\ = \frac{\alpha_s(Q^2)}{2\pi} \sum_j \int_x^1 \frac{\mathrm{d}\xi}{\xi} \begin{pmatrix} P_{q_i q_j}(x/\xi, \alpha_s(Q^2)) & P_{q_i g}(x/\xi, \alpha_s(Q^2)) \\ P_{g q_j}(x/\xi, \alpha_s(Q^2)) & P_{gg}(x/\xi, \alpha_s(Q^2)) \end{pmatrix} \begin{pmatrix} f_{q_j}(\xi, Q^2) \\ f_g(\xi, Q^2) \end{pmatrix}, \quad (1.2)$$

where the evolution of α_s with the scale is given by the QCD β function and the P_{ij} are the Altarelli-Parisi (AP) splitting kernels which can be thought of roughly as a probability to find a parton i splitting from a parent j by resolving the split at scale Q^2. Their leading-order (LO) values are given in most QCD textbooks, e.g. Ref. [28]. At low x values such that $\ln Q^2 \ll \ln(1/x)$, the latter logarithms which are ignored in DGLAP become large. This necessitates the application of the BFKL equations [29–32]. At even smaller x, the density of gluons becomes so large that the recombination $gg \to g$ dominates and BFKL also breaks down, necessitating nonlinear evolution equations. When recombination balances splitting, the proton is considered saturated, and forms a so-called color-glass condensate. Although the gluon occupancy is higher in proton-nucleus (pA) collisions than in proton-proton (pp) collisions [33], the proton-gold (pAu) measurements presented in this thesis in Ch. 6 result from initial parton dynamics with large x and Q^2, outside the kinematic regime of this phenomenon. See Ref. [34] for an excellent in-depth look at this low-x physics, which is beyond the scope of this thesis.

These non-perturbative PDFs appear at two points in a jet calculation. First, they enter in the factorization of the cross-section, shown below as an example for pp → jet + X and a jet substructure observable called angularity, τ_a [36] (which, for $a = 0$, is closely related to the jet mass as $M^2 = \tau_0 p_\mathrm{T}^2 + \mathcal{O}(\tau_0^2)$):

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\eta\,\mathrm{d}p_\mathrm{T}\,\mathrm{d}\tau_a} = \sum_{abc} f_a(x_a, \mu) \otimes f_b(x_b, \mu) \otimes H_{ab}^c(x_a, x_b, \eta, p_\mathrm{T}/z, \mu) \otimes \mathcal{G}_c(z, p_\mathrm{T}, R, \tau_a, \mu). \quad (1.3)$$

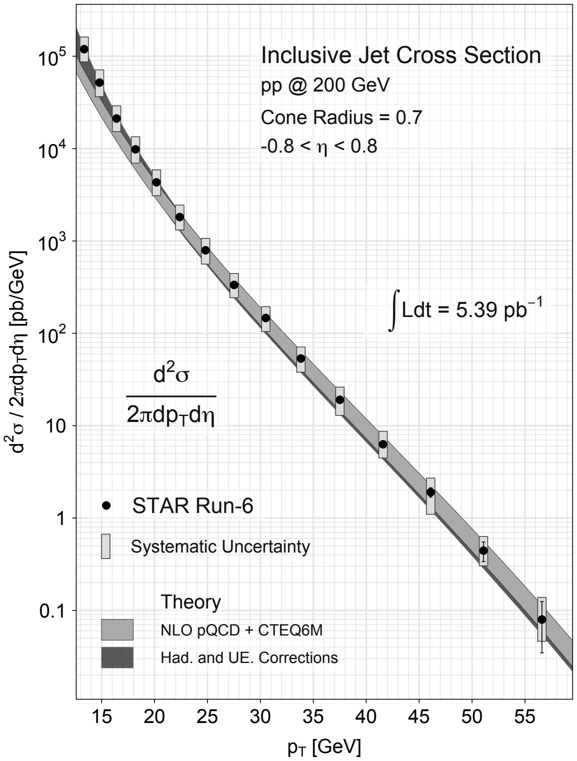


Figure 1.5: A comparison of the differential inclusive jet cross section from 200 GeV pp collisions in STAR Run-6 data to a next-to-leading-order (NLO) pQCD calculation with CTEQ6M PDFs and hadronization and underlying event corrections. [35]

Here, f is as before, and H is a hard function for production of a parton c similar to $\hat{\sigma}$ in eq. 1.1, with the new addition being a semi-inclusive jet function, $\mathcal{G}$, which takes into account contributions from radiation in the observed jet. Lastly, the final state of the jet must account for hadronization, which is described by analogous universal non-perturbative functions called fragmentation functions (FFs), $D_{k\to h}(z, \mu_F)$. There is a similar physical interpretation as the probability of finding a particle (in this case a hadron, h, in the final state) carrying some momentum fraction, z, in this case of the parent parton k's momentum. Together, PDFs and FFs make it possible to apply a solely perturbative parton-level jet calculation to collider data with high accuracy (Fig. 1.5) – that is, jets in vacuum are calibrated as an internal probe of the QGP. Any deviation from the pp baseline in heavy-ion collisions is then due to modification of the probe by cold or hot nuclear matter effects, which can be used to infer information about the structure of the nuclear matter. “Hot” nuclear matter (HNM) effects are those which are caused by the high-temperature quark-gluon plasma, while “cold” nuclear matter effects (CNM) are those caused by the large nucleus or nuclei not present in pp collisions. Relevant CNM effects for this thesis include modification to the proton PDFs due to the presence of the nucleus, and radiative energy loss as jets traverse the nucleus. They will be discussed in more detail in Ch. 2. In this thesis, we present measurements of jet substructure in pp and pAu collisions. The former allow us to study QCD radiation in vacuum; the latter allow us to determine the degree to which the cold nuclear environment plays a role in jet evolution. Both systems are also useful as baselines for heavy-ion collisions.

1.4 Overview

Before presenting the jet measurements performed for this thesis, I will discuss the current techniques for definition and experimental measurement of jets in Ch. 2, then in Ch. 3 give an overview of the RHIC accelerator and STAR detector from which data are obtained, and subsequently in Ch. 4 discuss the quality assurance selections made on the data. Finally, the measurements of jet substructure in pp and pAu collisions will be discussed in Ch. 5 and Ch. 6, respectively. I conclude and discuss potential future measurements in Ch. 7.

CHAPTER 2 JETS IN VACUUM AND COLD NUCLEAR MATTER

2.1 Jet clustering algorithms

As mentioned in Sec. 1.2, the final state of partons from a hard scattering is a collection of hadrons, as well as electromagnetically interacting particles such as electrons and photons, which result from a parton shower (Fig. 2.1). The QCD matrix elements are modified by each splitting according to the AP splitting kernels, mentioned in Sec. 1.3, in a small-angle approximation. The same DGLAP evolution equations can be applied to this final state splitting, now with the interpretation of the distribution functions, $f_i(x,Q)$, as the probability density of daughter partons. The radiation pattern in the parton shower, and subsequently the jet, is fundamentally determined by these AP splitting kernels. For instance, it has been shown that measurements of the jet substructure observable z_{g}, which represents the momentum sharing fraction of daughters in the hardest split (see Sec. 2.3), approximate the AP splitting kernel [38]. Similarly, the invariant mass of the jet is largely determined in the collinear limit by the cross section for a hard quark producing the hardest gluon with pair invariant mass m^2 [39],

$$\frac{\mathrm{d}\sigma}{\mathrm{d}m^2\,\mathrm{d}z} = \Delta(Q,m^2)\frac{1}{m^2}\frac{\alpha_s}{2\pi}P_{qq}(z). \tag{2.1}$$

Δ here is called a Sudakov factor, which represents the effect of a leading-logarithmic resummation that eliminates a divergence at small invariant mass. The Sudakov factor, which is itself dependent on the (unregulated) AP splitting kernels, is also used to obtain the branching probabilities in semi-classical Monte Carlo parton shower simulations (Sec. 5.1.1).

To access information about the initiating parton and the shower in vacuum or as it traverses nuclear matter, it is helpful to define a jet, which is a collection of the final-state hadrons with some spatial and momentum correlation with respect to the initiating parton. Jet definitions include both the rule for collecting nearby particles into a single jet, as well as the convention for assigning momentum to the jet (see note 2). This jet definition should be applicable to theoretical calculations and simulation as well as experimental data to allow

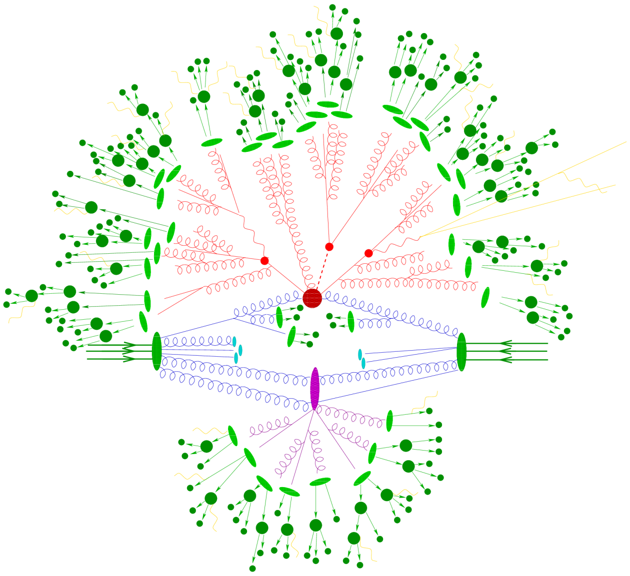


Figure 2.1: A schematic view of a hard collision and the resulting parton shower, hadronization, and hadronic decays. [37]

for straightforward comparison between them. The Snowmass Accord [40] in 1990 set a list of additional important considerations for jet definitions, including relative insensitivity to hadronization, and simplicity of implementation in experiment and theory, where in the latter they should be well-defined and produce finite cross sections at any order.

Algorithms in use today at the LHC and RHIC are required to be both infrared and collinear (IRC) safe (Fig. 2.2). The former refers to stability under addition of arbitrarily soft particles, which should not cause a merging of two jets that would otherwise be distinct; the latter refers to stability under arbitrarily close splittings which should not remove a potential jet from consideration due to, for example, a momentum threshold. From a theoretical perspective, this is important due to the following consideration: for a jet, at fixed order in α_s, the KLN theorem [42, 43] states that infrared divergences from real emissions should cancel the divergences from virtual corrections. If, for example, a collinear unsafe algorithm clusters two arbitrarily close particles into two different jets, this delicate cancellation of singularities would be broken, and the result of this calculation would be unphysical. From an experimental perspective, an insensitivity to arbitrarily soft or collinear radiation

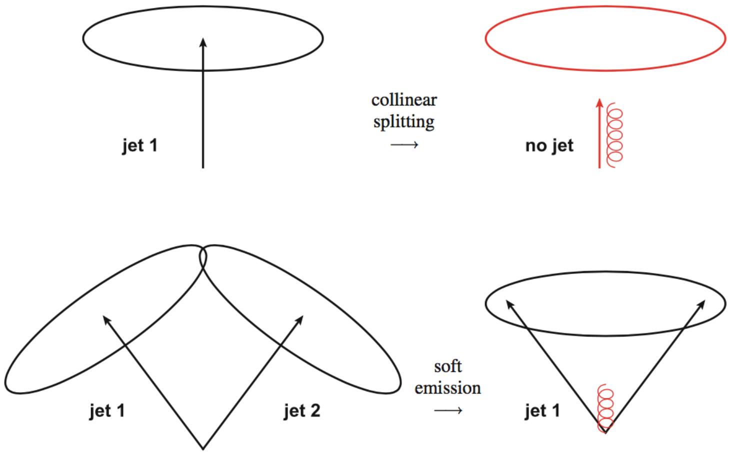


Figure 2.2: Upper: a collinear unsafe algorithm clustering arbitrarily close constituents. Lower: an infrared unsafe algorithm clustering arbitrarily soft radiation. [41]

is helpful so that detector performance concerns which do not strongly affect the total jet energy, such as track splitting and merging or tower boundary effects, do not change the experimental jet cross sections. Additional experimental considerations for jet algorithms, especially for high luminosity or heavy-ion collisions, include the clustering speed and the sensitivity to uncorrelated or correlated background such as pile-up and jet-induced medium recoil, respectively.

Today, so-called cone and sequential recombination algorithms are the two main classes in use. However, cone algorithms are typically not IRC safe, with the exception of SISCone [44] (or Seedless Infrared-Safe Cone), which is not commonly used because it is much slower than the standard sequential recombination algorithms. A sequential recombination algorithm works by choosing pairs of particles according to some minimized distance metric, combining them into a single object (called a pseudojet), and repeating until some criterion has been met (see below), at which point the algorithm stops and all remaining pseudojets are deemed jets. The current standard algorithms are the k_T family (k_T [45, 46], Cambridge/Aachen [47, 48], and anti-k_T [49]). These algorithms define distance metrics

$$d_{ij} = \min(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p})\frac{\Delta y^2 + \Delta\phi^2}{R^2},\ d_{iB} = p_{\mathrm{T},i}^{2p}, \tag{2.2}$$

where d_{ij} is a distance between particles i and j in k_{T} space[1] and d_{iB} is a distance between particle i and the beamline; p is a weight set by the algorithm (1, 0, and -1 for k_{T}, Cambridge/Aachen, and anti-k_{T}, respectively); and R is a parameter which roughly corresponds to the radius of the reconstructed jet, although not perfectly. The algorithm proceeds by finding the minimum d which may be between two particles or between a particle and the beamline. If it is the former, the particles are combined, with their four-momenta added[2], and the search for a minimum d is performed again. If it is the latter, the particle is deemed a jet and removed from consideration. When only jets remain, the process stops. The IRC safety of these algorithms is immediately demonstrable. For example, if either an arbitrarily soft or collinear particle is added to an event clustered with the k_{T} algorithm, it will have minimal d, so it will be clustered immediately and will not impact the final jets.

Because of the p-weighting, each of these three algorithms has different characteristics. As the anti-k_{T} algorithm looks for particles with smallest d, it will look to the hardest particles (with smallest p_{T}^{-2}) first. For example, given a hard particle, a, and two soft particles b and c with relatively small d_{bc}, with $d_{ab} < R$ and $d_{ac} > R$, the algorithm would first group a and b, leaving c to be clustered as part of a different jet. The result is that jets containing hard particles have a regular conical extent. The k_{T} algorithm, on the other hand, will cluster the soft particles first, so in the example above, b and c would likely be combined. Anti-k_{T}'s regular jet boundaries are desirable both from an experimental perspective, where regularity is important to ensure that the jet is entirely contained within the fiducial volume of the detector, and from a theoretical perspective, where an irregular boundary would increase the complexity of the components of the calculation (called non-global logarithms, or NGLs) relating to radiation entering the jet from outside its algorithmically determined boundary [51]. Anti-k_{T} has also been shown to be more resilient to uncorrelated background such as pile-up and underlying event (Fig. 2.3). For these reasons, anti-k_{T} is the standard algorithm

---

[1] For a definition of p_{T}, and other collider physics units and variables used throughout this chapter, consult Appendix A.

[2] This is the so-called E-scheme, which is the default in FastJet, but other choices for the pseudojet momentum are possible, e.g. the winner-take-all (WTA) scheme [50].

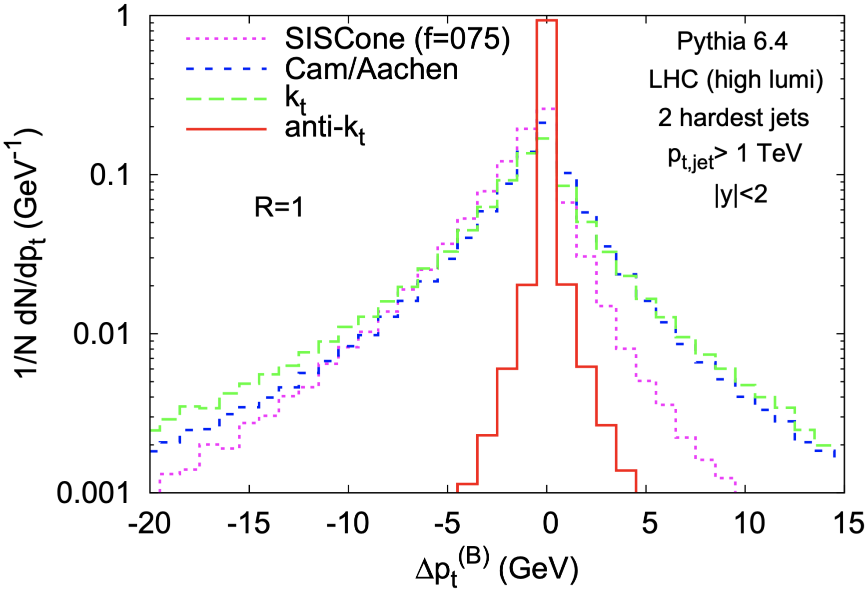


Figure 2.3: The change in jet p_{T}, in a simulation at LHC energy, when pile-up is added to the event and clustering is re-run. [49]

for jet-finding at modern colliders, although the other sequential recombination algorithms have found use; k_T is typically used to cluster soft particles for background estimation, while Cambridge/Aachen is used in the SoftDrop grooming algorithm mentioned in Sec. 2.3 because of its construction of an angular-ordered tree which approximately corresponds to the angular ordering of the parton shower.

2.2 Background contamination

Unfortunately, no jet clustering algorithms offer a way to distinguish the signal (particles in the jet directly descended from the original hard scattering) from the noise. Noise in this case can be either particles in the jet that were produced in additional soft interactions and from color reconnection of beam remnants, or particles from distinct events (pile-up). The intra-event background is referred to as the underlying event (UE) and in heavy-ion collisions it can be quite substantial in terms of the number of UE particles in a given unit area entering the jet. However, in pp collisions at RHIC, we expect on average roughly an additional 0.3 charged particles above 0.2 GeV/c (Fig. 2.4) in a high-p_{T} jet of radius 0.4 which typically has roughly 3 tracks. This is one motivation for keeping the jet radius

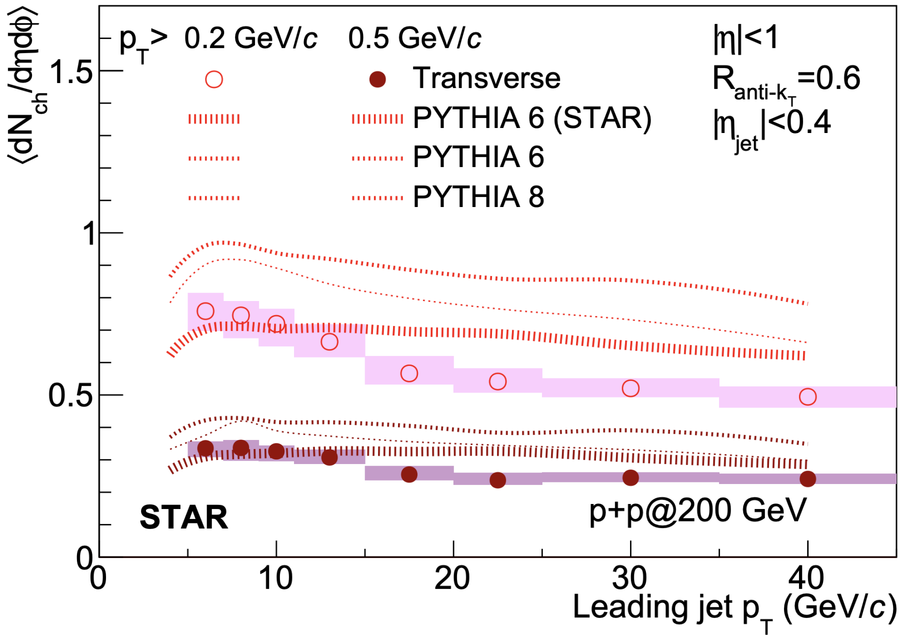


Figure 2.4: A STAR measurement of the underlying event associated with jets from 200 GeV pp collisions, in terms of average number of charged particles per unit rapidity and azimuth. [52]

small in heavy-ion collisions, as the anti-k_T jet area scales quadratically with the radius so larger-radius jets include significantly more background. Because UE is relatively small in pp collisions, additional techniques to subtract it do not need to be introduced in the analysis, and additionally the jet radius can be increased to study the effect of the larger phase space for radiation on the measured observables.

In heavy-ion collisions, the background p_T-density is defined as $\rho = \text{median}(p_{T,i}/A_i)$ for k_T clusters (pseudojets) i of area A which usually exclude the two hardest clusters from the median. ρ is typically pedestal-subtracted on a jet-by-jet basis, with the corrected jet p_T given by $p_T^{\text{sub}} = p_T^{\text{raw}} - \rho A$, where A is the area of the jet [53]. However, in small systems (e.g. pAu) collisions, the background is not much larger than in pp collisions and additionally has region-to-region fluctuations, σ, comparable to ρ, due to the sparse nature of the events. This is a complicating factor, because a simple pedestal subtraction may in this case misrepresent the jet by a significant amount depending on whether the jet is sitting on a large background fluctuation upwards or downwards, or not. If the background is not subtracted on a jet-by-jet basis, the jet may have a larger than physical p_{T}, but this slight shift and broadening due to the relatively small ρ can be incorporated into a detector response matrix (Sec. 5.3) which is used in a procedure for correcting the data on a statistical basis. Here, it will likely be dominated by detector effects such as tracking efficiency and momentum resolution [54]. Indeed, it was found in Ref. [55] that a jet-by-jet subtraction for the median mass density, ρ_{m}, led to a worsening of the jet mass resolution (Sec. 5.2) when compared to an ensemble correction given by a Bayesian unfolding procedure (Sec. 5.3). The analysis presented in Ch. 6 will use a similar approach, with no attempt at explicit jet-by-jet pedestal subtraction.

2.3 Jet substructure

Two jets with the same p_{T} may have quite different parton production and shower histories. For this reason, observables sensitive to the structure of jets have become popular and have now been used in hundreds of measurements, especially as a signal for heavy objects [56], the decays of which are significantly Lorentz-boosted, such that they can be measured within a single jet. The substructure of these jets therefore gives information about the initial heavy object. One substructure observable with broad applicability is the jet mass, defined as the magnitude of the four-momentum sum of jet constituents, $M = |\sum_{i\in J} p_i| = c^{-2}\sqrt{E^2 - \mathbf{p}^2c^2}$. As mentioned in Sec. 2.1 the jet mass is sensitive to the fragmentation pattern of partonic radiation which is in turn sensitive to the initiating parton's virtuality. However, jet substructure measurements may also be affected by non-perturbative physics such as UE and hadronization, making it difficult to compare directly to analytic calculations. Therefore, tools such as SoftDrop [57] have been developed to "groom" away unwanted constituents of a jet. As mentioned in Sec. 2.1, the algorithm begins by reclustering a jet with Cambridge/Aachen to obtain an angular-ordered tree which mimics the physical parton shower evolution. Then, beginning with the widest-angle pair (below, particles 1 and 2) in the tree and working inward along the higher p_{T} prong, it removes the softer of the two (see Fig. 2.5)

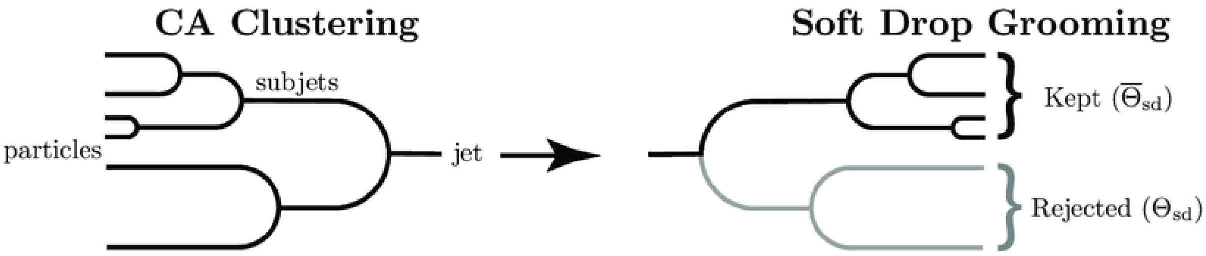


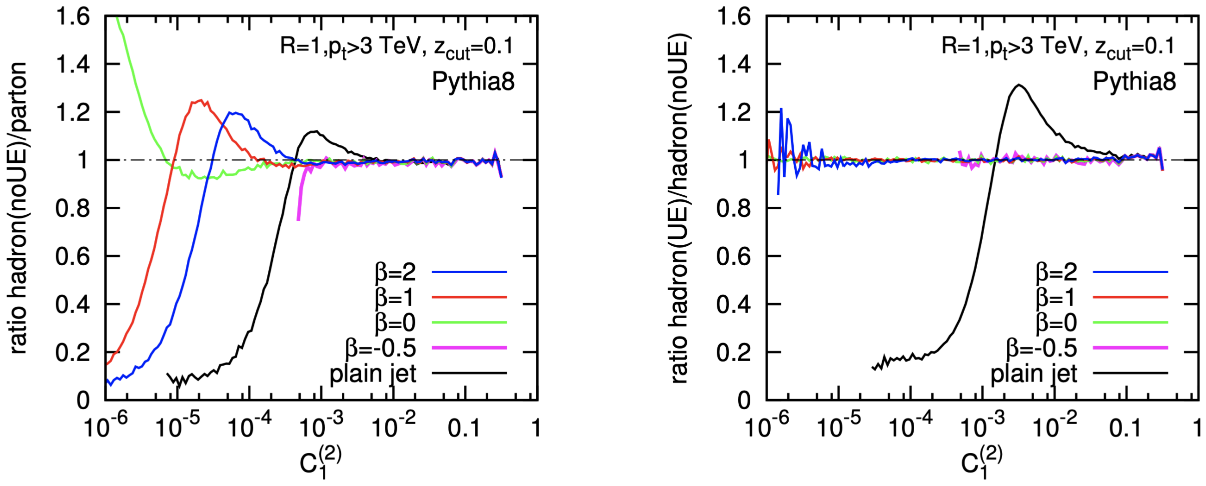
Figure 2.5: A schematic depiction of the SoftDrop algorithm applied to a jet which has been re-clustered with the Cambridge/Aachen algorithm. [58]

if it fails the criterion:

$$\frac{\min(p_{\mathrm{T},1}, p_{\mathrm{T},2})}{p_{\mathrm{T},1} + p_{\mathrm{T},2}} > z_{\mathrm{cut}} \left(\frac{\Delta_{12}}{R}\right)^{\beta}, \tag{2.3}$$

where z_{cut} is the momentum sharing fraction of the prong, Δ_{12} is the angular separation, R is the jet radius, and β is a parameter which determines the strength of grooming. The procedure is repeated on the next pair until there are no pairs remaining or a pair passes the above criterion. In the former case, in SoftDrop's grooming mode this jet is considered groomed, despite being unchanged, while in tagging mode this jet is dropped from the list of groomed jets. For the analyses in this thesis, tagging mode was chosen as it is IRC safe for the value of β selected, while this is not true of grooming mode.

The canonical SoftDrop grooming parameters, which we use for all SoftDrop measurements in Chs. 5 and 6, are $z_{\mathrm{cut}} = 0.1$ and $\beta = 0$. The theoretical motivation for $z_{\mathrm{cut}} = 0.1$ is that it is close to the value which minimizes higher-order effects for quark-initiated jets (Ref. [59], eq. 8.1), which are dominant at RHIC in the kinematic regime of the measurements presented in this thesis (see Ref. [60] and Sec. 5.3). The $\beta = 0$ value is chosen because at this value, NGLs do not enter the calculation of the mass of the groomed jet (Ref. [57], Sec. 3.4). Fig. 2.6 demonstrates that SoftDrop with these nominal parameters is indeed capable of reducing the contribution from non-perturbative physics, and we will show both ungroomed and SoftDrop-groomed jet observables in Ch. 5 and Ch. 6. In addition, by comparing ungroomed to groomed observables (e.g. the jet mass, M, to the mass of groomed



(a) (b)

Figure 2.6: The reduction in influence of (a) hadronization and (b) underlying event on a jet substructure observable, $C_1^{(2)}$, after SoftDrop grooming, in a simulation at LHC energies. [57]. See also eq. 1.2 within this reference for a definition of $C_1^{(2)}$, which is a 2-point correlator related to the jet mass.

jets, M_{g}) as a function of jet p_{T} and R, we can experimentally determine the phase space for which non-perturbative effects are significant.

2.4 Cold nuclear matter effects on jets

Although the underlying event is fairly small in pA collisions, there could still be a significant difference between jets measured in these small systems compared to jets in vacuum, due either to cold or hot nuclear matter effects. Despite indications that the latter may be present even in high-multiplicity pp collisions [61, 62], there is so far no sign [63–70] of the typical path-length dependent energy loss due to gluon bremsstrahlung that characterizes jets which have traversed the QGP in heavy-ion collisions [71]. However, modification (both suppression and enhancement) to mid-rapidity jet yields in small systems (proton- or deuteron-nucleus) collisions as a function of event activity measured at large rapidity ($3 \lesssim |\eta| \lesssim 5$) has been observed [72, 73] (see Ch. 6 for more details). Various potential explanations for this unexpected result have been developed, e.g. energy conservation [74] or a modified proton wave function [75]. The question of whether jet substructure could resolve

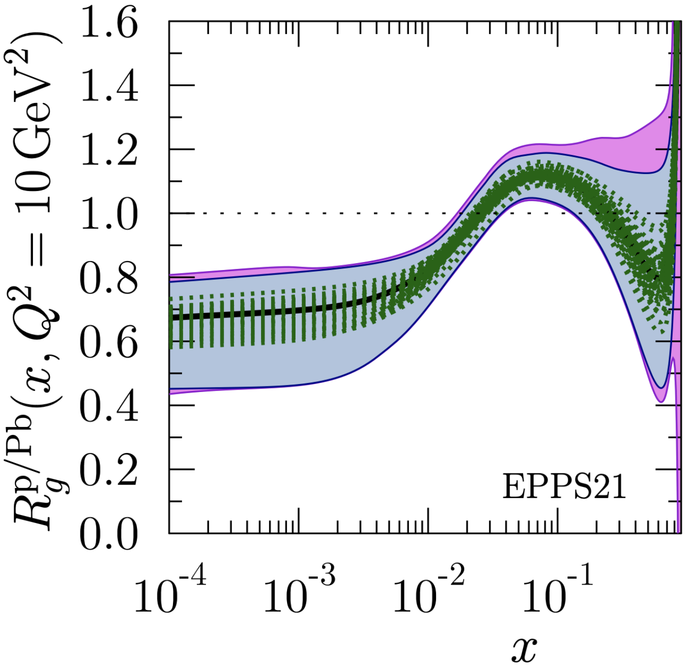


Figure 2.7: The average nuclear modification factor for gluons in the protons within lead nuclei as a function of the fraction of the proton momentum carried by the parton, at the scale $Q^2 = 10\,\mathrm{GeV}^2$, from the EPPS21 nPDF set. [76]

these or other nuclear matter effects was a motivation for the pAu jet substructure analysis presented in Ch. 6.

Other, better understood, CNM effects due to the presence of the large nucleus which could potentially influence jets include initial-state and final-state radiative energy loss [77] and modification of PDFs [76]. These modified PDFs (nPDFs), $f_i^{\mathrm{p/A}}$ for parton flavor i, are taken to be the free PDFs, f_i^{p}, multiplied by some nuclear modification factor, R_i^{A}, so that $f_i^{\mathrm{p/A}}(x, Q^2) = R_i^{\mathrm{A}} f_i^{\mathrm{p}}(x, Q^2)$. Similarly to the procedure mentioned in Sec. 1.3, the nPDFs (or equivalently, the nuclear modification factors) are parametrized, and global fits to data are performed, with evolution from the parametrization scale Q_0 to some scale Q obtained from the (N^n)LO DGLAP equations. As a function of Bjorken-x, or the fraction of the proton momentum carried by the parton (Sec. 1.3), the R_i^A have multiple features, two examples of which are shadowing and antishadowing. This refers to the suppression ($R_i^A < 1$) at low-x and enhancement ($R_i^A > 1$) in the mid-x region. These effects can be seen in Fig. 2.7

for $x \lesssim 0.02$ and $0.02 \lesssim x \lesssim 0.2$, respectively. However, at RHIC, the kinematic reach of jets extends to roughly $x = 0.03$ at lowest[3], so only anti-shadowing would be expected to influence the jets. Here, we would only expect a modification to jet yields since the hard matrix element and subsequent parton shower would be unchanged, while partonic energy loss in the nucleus would contribute to a potential modification to the substructure of jets similar to the heavy-ion case [80].

---

[3] This can be estimated by $x = (M_{\text{dijet}}/\sqrt{s})e^{\eta_{\text{dijet}} - \eta_{\text{shift}}}$ [78], where $\eta_{\text{dijet}} = (\eta_1 + \eta_2)/2$ for $R = 0.4$ jets is bounded by -0.6 and 0.6 at STAR; $\eta_{\text{shift}} \approx 0.03$ is the shift of the center-of-mass rapidity from the observation frame (Ch. 3, note 1); and M_{dijet} ranges from roughly 10 to 80 GeV at RHIC [79]. Together, this gives $0.03 \lesssim x \lesssim 0.75$.. Note, however, that due to the high p_{T} of jets in the analyses presented in this thesis, the minimum accessible Bjorken-x value is closer to $x \approx 0.1$.

CHAPTER 3 EXPERIMENTAL SETUP

The data used in the analyses presented in this thesis consist of recorded particle collisions, or "events", in the Solenoidal Tracker at RHIC (STAR) detector, which is located on the circular Relativistic Heavy Ion Collider (Fig. 3.1) operated by Brookhaven National Laboratory on Long Island, in New York. Before collisions can be recorded, the technically difficult act of accelerating nuclei (including protons) of various species to as much as 99.999% the speed of light must be accomplished. Although RHIC [82] is quite flexible, having to date accelerated nuclei of ${}^{1}\text{H}^{1}$, ${}^{2}\text{H}^{1}$, ${}^{3}\text{He}^{2}$, ${}^{16}\text{O}^{8}$, ${}^{27}\text{Al}^{13}$, ${}^{63}\text{Cu}^{29}$, ${}^{96}\text{Zr}^{40}$, ${}^{96}\text{Ru}^{44}$, ${}^{197}\text{Au}^{79}$, and ${}^{238}\text{U}^{92}$ to a broad range of energies between 3.85 and 254.9 GeV/nucleon, this thesis will focus on protons and gold nuclei accelerated to approximately 100 GeV/nucleon.

3.1 The RHIC complex

This acceleration process begins with ion sources which feed charged ions to either the 200 MeV Linear Accelerator [83] for protons or the adjacent Electron Beam Ion Source (EBIS) [84] for heavier species. As they are transferred through successively larger rings (first the Booster Synchrotron, then the Alternating Gradient Synchrotron, and finally into each of the two kilometer-diameter RHIC storage rings), the nuclei are stripped of successively more electrons as they pass through sheets of foil [85], and nuclei and protons alike are both accelerated to successively higher kinetic energy via radio-frequency (RF) electromagnetic cavities which apply accelerating voltages to these charges. Because of the varying strength of the field, the beam is divided into bunches which experience roughly the maximum force aligned with their direction of motion. At full capacity, there are 111 such bunches, each a few millimeters in transverse size by a few meters in length, separated by roughly 30 meters, circling in each direction (clockwise and counterclockwise) in the RHIC storage ring, each with on the order of 10^{11} particles in the case of protons and 10^{9} in the case of gold. This factor of 100 difference is due to the requirement for stable bunches of similar charge per bunch.

Modern synchrotrons, including RHIC, maintain the closed orbit of particles via magnetic

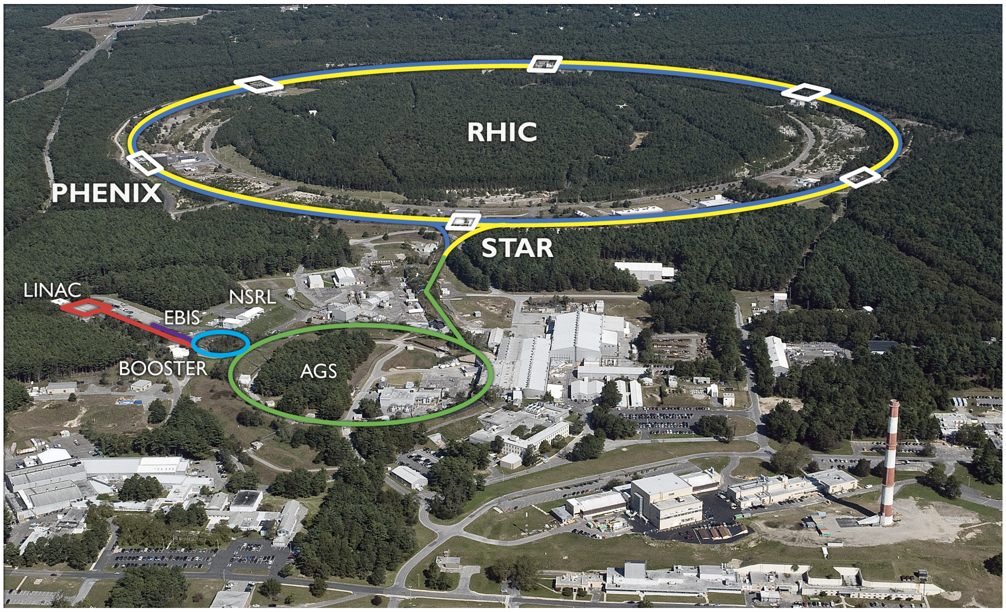


Figure 3.1: The RHIC complex. [81]

fields along the arcs of a rounded polygon, with the straight sections being reserved mostly for acceleration (as mentioned above) and detectors (see Sec. 3.2). As the kinetic energy of the particles increases, the magnetic fields must also increase in sync to maintain the same radius of curvature. This is accomplished at RHIC with superconducting magnets, cooled below 4.6 K by supercritical helium. The energy of protons or ions is limited by the maximum field of these magnets, which is roughly 3.5 T, and the species' charge-to-mass ratio, Z/A (accounting for the difference in maximum CM energy between pp ($Z/A = 1$, $\sqrt{s} = 510\,\text{GeV}$) and AuAu ($Z/A = 0.4$, $\sqrt{s_{\text{NN}}} = 200\,\text{GeV}$) collisions)[1]. There are two rings of these superconducting magnets (in contrast to particle-antiparticle colliders which only require one, but have less flexibility in species collided), with 396 dipole magnets to control bending of the beam, 492 quadrupole magnets to focus it, and 852 smaller magnets [82].

The rounded hexagon of RHIC has six possible interaction regions (IRs) along the straight

---

[1] Although this beam "rigidity" accounts for the maximum energy of each beam, it is possible to collide different ion species with differing rigidities at roughly the same energy per nucleon, using clever accelerator techniques [86]. Because of this, the pAu data taken in 2015 (Sec. 4) was able to use an Au beam of 97.74 GeV and p beam of 103.88 GeV, which resulted in a rapidity shift of the center of mass (Appendix A) of only approximately 0.03, which allows for more of the products of the collision to be measured at mid-rapidity than if there were a shift on the order of that in LHC pPb collisions of 0.465.

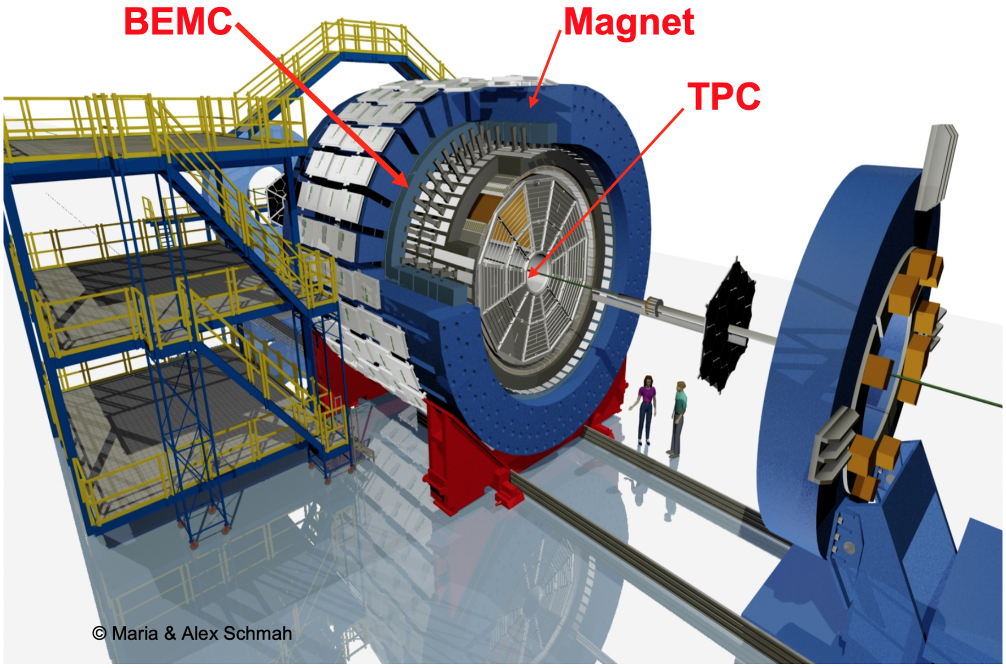


Figure 3.2: A schematic depiction of the STAR detector, with three important subsystems highlighted. Note there is an artificial cutaway from the barrel to show the components underneath the magnet and the endcap, and the large-rapidity detectors are pulled away to show the TPC. [90]

segments where the beams cross, in addition to its six arc segments. Of the IRs, one is occupied by the RF cavities, and another is occupied by a beam polarization measurement called HJET [87] used during polarized proton runs. The remaining four were previously all occupied by data-taking detectors, although now the only remaining active experiment is STAR. A new experiment called sPHENIX will replace the previous Pioneering High Energy Nuclear Interaction eXperiment (PHENIX) detector [88] and start data-taking in the first quarter of 2023. Planning has also advanced for a future electron-ion collider (EIC), which will be located at the current RHIC site, with CD-1 status recently being granted by the Department of Energy [89] marking the beginning of the preliminary design phase. Construction is planned to begin in 2024 with first data in the early 2030s.

3.2 The STAR detector

STAR [91] (shown in Fig. 3.2) is a cylindrical detector with full coverage in azimuth and a wide rapidity[2] coverage of roughly $|\eta| < 5$. As the bunches delivered by the accelerator cross near the center of the barrel and collisions occur, the products of the collision are measured by various subsystems. Those subsystems which are relevant for this analysis will be presented in detail in the following subsections. Subsystems are often upgraded or replaced, so presented below is the state of the detector for data-taking in 2012 and 2015 during which time the data analyzed in this thesis were taken. Where relevant differences occur between these two years, they will be highlighted and discussed.

3.2.1 The Time Projection Chamber

Charged particles in an external magnetic field will curve, due to the Lorentz force, which gives a simple relation between the momentum transverse to the magnetic field and the measured radius of curvature of the helix, for a given magnetic field, B, and with typical charge magnitude $q = 1$, $p_{\mathrm{T}}[\mathrm{GeV/c}] \approx \frac{0.3\mathrm{GeV/c}}{\mathrm{eTm}}(qBr\,[\mathrm{eTm}])$. The field is provided by an 1,100 ton, nominally 0.5 T, water-cooled, room-temperature solenoidal magnet. The device for measuring the radius of curvature is the Time Projection Chamber (TPC) [92], a schematic of which is shown in Fig. 3.3, nested inside the magnet. The basic architecture is comprised of two concentric cylinders, the inner and outer field cage rings, which are divided into 182 equipotentials to maintain the uniformity of the electric field in the volume; a central membrane acting as a cathode at 28 kV; and an endcap anode at 0 V on each side separated from the central membrane by 210 cm.

The enclosed volume contains an ionizable gas from which electrons are freed by the charged particles from the collision and subsequently drift along an electric field (uniform, along the z, or beampipe direction – see Appendix A) to have their positions read. The gas mixture used is P10 (90% argon, 10% methane) which was chosen for its high drift velocity (5.45 cm/μs) for a low applied electric field (nominally 135 V/cm for STAR). The gas is held

---

[2] For a definition of rapidity, and other collider physics units and variables used throughout this chapter, consult Appendix A.

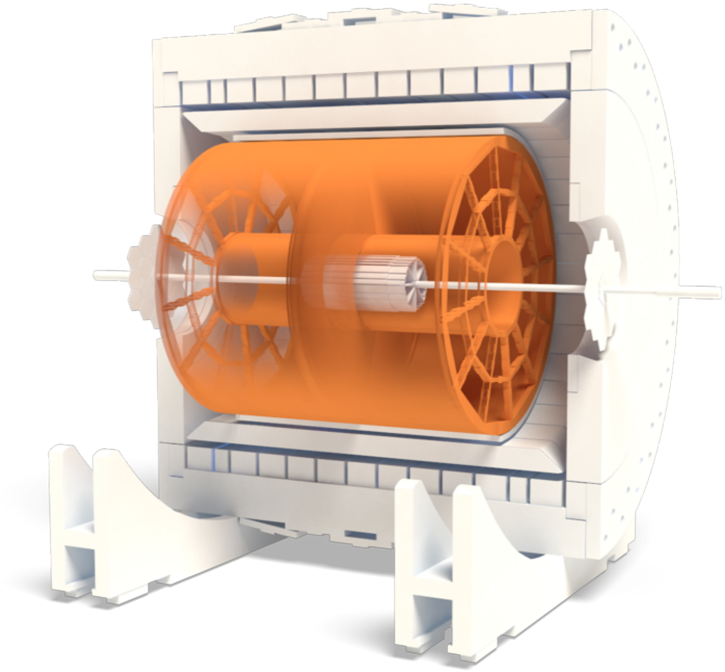


Figure 3.3: A schematic depiction of the Time Projection Chamber, in orange, situated within the STAR detector. [93]

slightly above atmospheric pressure to prevent inclusion of impurities such as oxygen and water which increase electron absorption. Because atmospheric pressure fluctuates, and due to small impurities, the drift velocity in the gas can change, so calibrations are done using lasers multiple times per day. However, because the electric field is chosen to correspond to the peak of the drift velocity curve, small variations in conditions do not have a large effect. P10 has a characteristic transverse ($x - y$) and longitudinal (z) diffusion for a given magnetic field. For the nominal $B = 0.5\,\mathrm{T}$, the transverse diffusion over one half of the TPC is 3.3 mm while the longitudinal diffusion over the same length is 5.2 mm. For the drift velocity mentioned above, this corresponds to a spread in time of 230 ns [92].

Electrons are read out at the endcaps by multi-wire proportional chambers with pad planes, divided into twelve wedges called sectors, with 3 mm spacing due to the support structure between sectors. The readout is divided into four sections, from low to high z: a gated grid, shield grid, anode grid, and finally a pad plane (Fig. 3.4). The anode wires provide

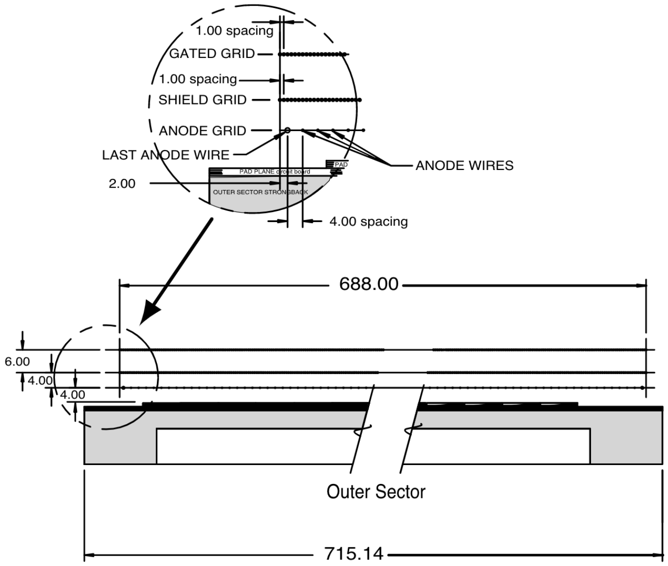


Figure 3.4: A diagram of the multi-wire proportional chamber in an outer subsector of the TPC, with listed dimensions in millimeters. [92]

amplification of the incoming primary ionization electrons by a factor of a few thousand to be measured at the pad planes. The shield grid confines the field from the anode to the avalanche area. The gated grid prevents the ions from the avalanche from returning to the TPC volume and distorting the field there. Each sector (Fig. 3.5) has thousands of these pads, with roughly 100 to a row, and 45 total rows, extending radially from the inner field cage. For a few reasons, the geometry of the inner 13 rows is different from the outer subsector of 32 rows. The inner subsector has smaller pads which are designed to improve the two-hit resolution (a concern due to the high density of tracks in the inner region, especially in heavy-ion collisions, which can lead to overlapping secondary avalanches). However, for technical reasons, it is not possible to pack rows more tightly to achieve the spatially contiguous readout that exists in the outer subsector. The main benefit of this feature of the outer subsector is higher statistics for track ionization energy loss ($\mathrm{d}E/\,\mathrm{d}x$) measurements for particle identification.

After hits have been recorded in the readout pads, a three-dimensional map is created, using the time of the collision and the calibrated drift velocity to deduce the z position of

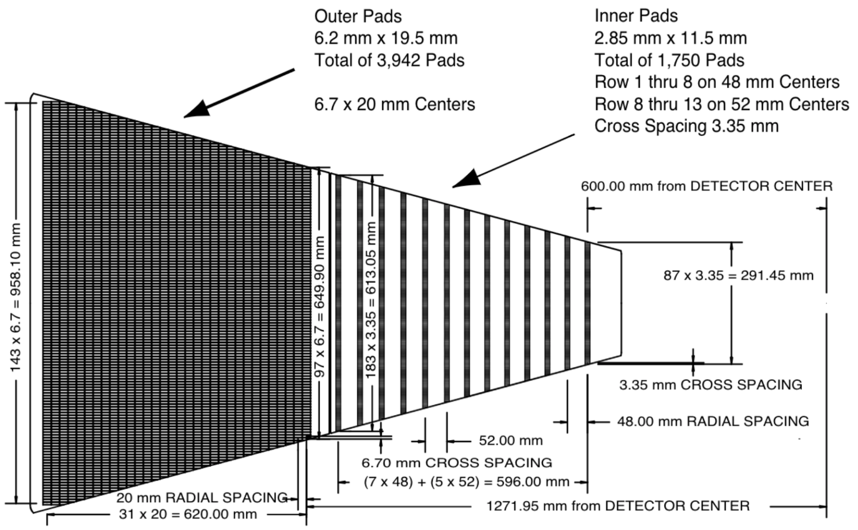


Figure 3.5: A diagram of one of the twelve sectors of the TPC. [92]

each hit. Hit positions are additionally corrected for known distortions in the electric and magnetic fields, which is especially important for later tracking of high-momentum particles since the error on the transverse momentum due to $x - y$ resolution is proportional to p_{T}. Then, working backward from the outermost pad row, an algorithm correlates hits into tracks. These tracks are fit with a helix, which later allows for extraction of the momentum and charge of the particle. The primary vertex position is estimated by the average position of the beginning of these “global tracks”. Tracks with a distance of closest approach (DCA) less than 3 cm are refit with the requirement that they pass through this vertex to obtain so-called “primary tracks”, which are the tracks used for the analyses in this thesis.

Not all charged particles are found by the tracking procedure, due to the physical limitations of the detector (e.g. the support structure separating the TPC endcap sectors, which is not readout-enabled) and to procedural requirements. For instance, tracking efficiency reduces quite sharply for low-momentum tracks as they loop with a small radius in the magnetic field, and therefore do not cover the minimum number of pad rows (10) to be deemed a track rather than a potential broken fragment of a larger track. For those tracks which are found, momentum has an intrinsic precision, due mostly at low-p_{T} to stochastic multiple

Coulomb scattering in the gas which causes deviation from the naïve helical shape, and at high-p_{T} to $x - y$ hit position resolution, as mentioned above. The resolution near its optimum at mid-p_{T} of roughly 1 GeV/c is a few percent. Poor resolution at large track p_{T} is the reason for an analysis requirement of tracks with $p_{\mathrm{T}} < 30\,\mathrm{GeV/c}$ (Sec. 4.2).

3.2.2 The Barrel Electromagnetic Calorimeter

Electromagnetic calorimetry is a useful tool for measuring the energy of particles which interact with the material in the detector, some of which are neutral and therefore are not measured by the TPC. Ideally, calorimeters are segmented finely enough that each segment, or "tower", typically encompasses one or very few electromagnetic showers. However, to span the $60\,\mathrm{m}^2$ of barrel area with towers of width near the Molière radius, which contains 90% of the shower width (1.6 cm for Pb), would be impractical. Instead the STAR Barrel Electromagnetic Calorimeter (BEMC) [94] is made of 4800 towers with transverse size of roughly $10\,\mathrm{cm} \times 10\,\mathrm{cm}$ near $\eta = 0$, and increasing as $|\eta|$ increases due to the fact that the towers project back toward $z = 0$. For this reason, the towers contain a Shower Maximum Detector (SMD), which is a gas-filled (90% Ar, 10% CO_2) wire proportional counter, located between 5 and 7 radiation lengths X_0 (depending on the position in pseudorapidity), which is near the longitudinal position of the maximum energy deposition for showers from high-energy electromagnetically-interacting particles. The SMD provides an improved spatial resolution to the BEMC on the order of the Molière radius for low occupancy, while using roughly one tenth the number of channels necessary if the towers were the size of the Molière radius. Additionally, it is capable of particle discrimination since electromagnetic and hadronic showers have much different transverse and longitudinal profiles.

Each tower is composed of 20 lead plates and 21 scintillating tiles, alternating between absorber – which induces a radiation cascade – and scintillator – which measures the shower and converts it to a light signal – with a gap after five such pairs of these for the SMD. With the exception of the first two scintillators which are 6 mm thick, each layer is 5 mm, which is roughly one radiation length in lead ($X_0 \approx 5.6\,\mathrm{mm}$). Therefore, the entire tower is about

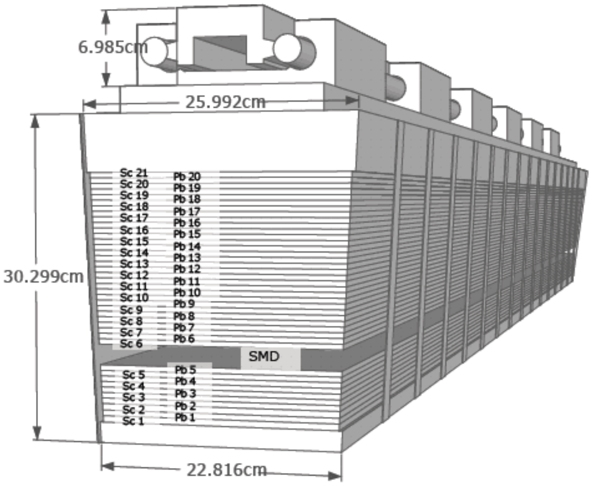


Figure 3.6: A diagram of one of the 120 modules in the STAR BEMC, with the location of the SMD labeled. [93]

$20X_0$ deep, which is enough to contain most electromagnetic showers entirely, although 95% longitudinal containment for electrons and photons exceeds $20X_0$ for $E \gtrsim 1\,\mathrm{GeV}$ [95]. Forty of these towers (20 in η by 2 in ϕ each with angular size 0.05×0.05) comprise one of the 120 modules ($\Delta\eta \times \Delta\phi = 1 \times 0.1$) (Fig. 3.6). The BEMC provides complementarity to the TPC given that its energy resolution (nominally $\sigma_E/E \approx 1.5\% + 15\%/\sqrt{E}$) improves as energy increases, while the TPC momentum resolution degrades as momentum increases ($\sigma_{p_\mathrm{T}}/p_\mathrm{T} \propto p_\mathrm{T}$).

3.2.3 Large-rapidity subsystems

Although the measurements in this thesis directly make use of only the particles measured at mid-rapidity by the TPC and the BEMC, the detectors at large-rapidity are important for triggering and centrality estimation. One of these is the Vertex Position Detector (VPD) [97] which consists of two identical scintillating detectors along the beampipe at ±5.7 m from the center of the detector. Coincident PMT signals from the plastic scintillators are used as a measure of the start and location of a collision, with a time (position) resolution of approximately 80 ps (2.5 cm) in pp collisions. For uses of the VPD in the analysis, see Secs. 4.1 and 4.5.

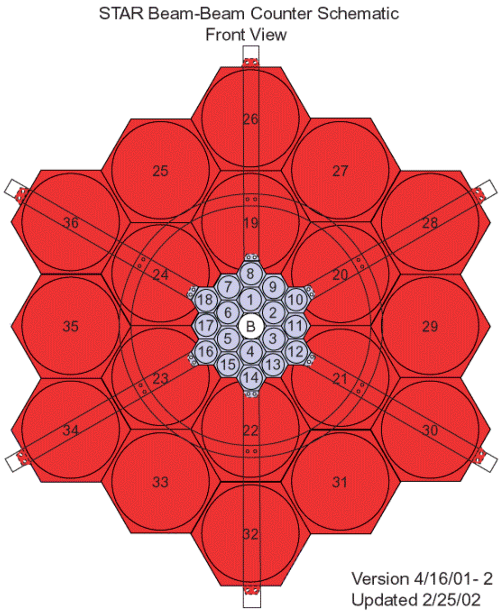


Figure 3.7: A diagram of one of the BBCs. [96]

Another detector which is used is the Beam-Beam Counter (BBC) [98] (Fig. 3.7), a large-rapidity hexagonal scintillating detector ($2.2 < |\eta| < 5.0$) at $z = \pm 3.7\,\mathrm{m}$ which was originally intended as a polarimeter for polarized pp collisions, but can also be used as a minimum-bias trigger in pp collisions to replace the Zero Degree Calorimeter (ZDC) since the latter measures spectator neutrons which are not present in such collisions. Additionally, in the pAu analysis presented in Ch. 6, the Au-going BBC (specifically the inner ring, covering $-5.0 < \eta < -3.4$, denoted iBBCE) is used as a measure of the activity in an event which avoids a trivial correlation with the measured jets since it is completely separated from them in η.

CHAPTER 4 DATA SELECTION

The analysis in Ch. 5 uses data from pp collisions at $\sqrt{s} = 200\,\mathrm{GeV}$ taken in 2012 ("Run 12"), with total delivered luminosity of $74\,\mathrm{pb}^{-1}$ (see Ref. [99] for definition and units). The analysis in Ch. 6 uses data from pAu collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ taken in 2015 ("Run 15"), with total delivered luminosity of $1.27\,\mathrm{pb}^{-1}$. By design, the choices made for both the pp and pAu analyses were very similar. When not otherwise noted in this chapter, the procedure being discussed will refer to both analyses.

4.1 Event selection

Events were required to have passed an online trigger called "Jet Patch 2" (JP2), requiring that the EMC records an ADC value above a threshold in at least one of the 18 partially overlapping 1×1 (in $\eta \times \phi$) areas of 400 towers each, called jet patches. In 2012, this threshold was 36 (roughly $E_{\mathrm{T}} > 7.3\,\mathrm{GeV}$ [100]). In 2015, the thresholds were 500, 42, and 40, for east ($\eta < 0$), middle, and west jet patches, respectively. It is therefore practically impossible for a jet to trigger an east jet patch in this dataset (see Sec. 6.1 for more details). However, for $R = 0.4$ jets, the overlapping middle jet patches extend to the edge of the jet acceptance, so for these and larger radius jets there is no effect of this larger threshold. Since high-energy triggers are rare, the JP2 was not prescaled, meaning that every event which passed the JP2 condition was recorded, with no concern for bandwidth. This resulted in a recorded luminosity of $22.9\,\mathrm{pb}^{-1}$ in 2012 and $0.42\,\mathrm{pb}^{-1}$ in 2015. However, in the 2015 dataset, one crate had a long-lasting calibration problem (Fig. 4.1) so the events in which this crate was miscalibrated were removed from consideration so that jet topological selection was not biased. These runs (16142059 – 16149001) account for roughly 1/5 of the 35 days of data acquisition. Additionally, in the 2012 (2015) dataset, 208 (205) runs which were outliers in some quantities (Ref. [101], Sec. 4.1), including track quality (e.g. DCA) and kinematics (e.g. $\langle p_{\mathrm{T,tracks}} \rangle$), for instance, were removed.

Events were then required to pass certain quality criteria for consideration. Since tracks must hit a minimum number of pad rows, the reconstructed vertex is required to be relatively

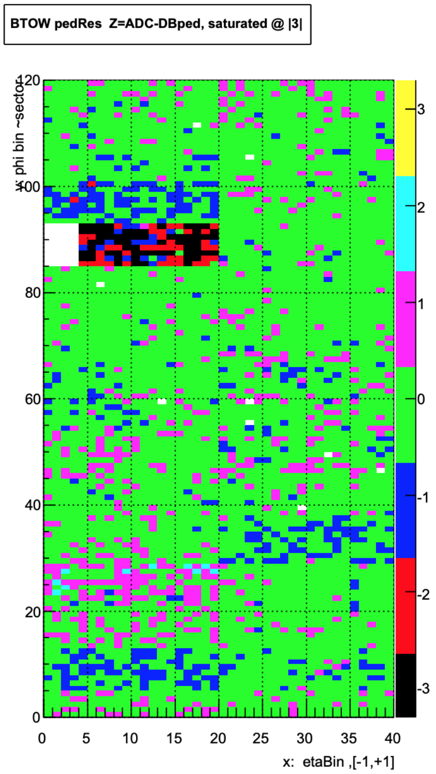
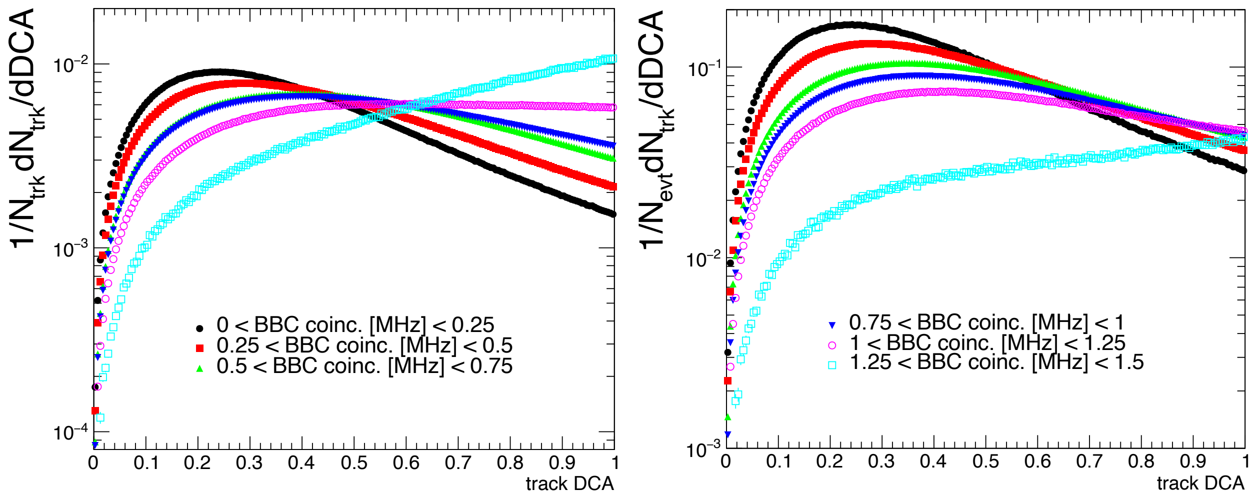


Figure 4.1: BEMC online monitoring plot for pedestal-subtracted ADC as a function of η and ϕ tower position, for example run 16147065, showing the bad calibration of crate 4.

close to $z = 0$ ($|v_z| < 30\,\mathrm{cm}$), so that collision products on either side of the vertex are reconstructed with similar efficiency. Additionally, in the 2015 data, during quality analysis it was noticed that the track distance of closest approach (DCA) to the vertex was highly dependent on luminosity, so a strict selection on the difference between the prompt VPD signal and the slower TPC signal ($|v_z^{\mathrm{VPD}} - v_z^{\mathrm{TPC}}| < 3\,\mathrm{cm}$) was imposed. This reduces events with out-of-time pileup, thus decreasing luminosity dependence of tracking (Fig. 4.2). It was also required in the 2015 data that events have an ADC sum in the inner BBC ring on the east side (iBBCE) less than 6.4×10^4, as this is about where the signal saturates in the BBC. However, this selection removes very few events in this analysis. Finally, in both datasets, events containing tracks determined to have $p_{\mathrm{T}} > 30\,\mathrm{GeV/c}$ were not considered due to poor momentum resolution of high-momentum tracks (Sec. 3.2.1). To maintain consistency between tracks and towers, events containing towers with $E_{\mathrm{T}} > 30\,\mathrm{GeV}$ were also removed from consideration.

For the pAu analysis, the event activity (EA), determined in the iBBCE, was initially



(a)

(b)

Figure 4.2: Luminosity dependence of track–vertex DCA in JP2-triggered 2015 data without (a) and with (b) a VPD–TPC Δv_z within $\pm 3\,\text{cm}$.

[102] required to be above the 50th percentile (“high”) or below (“low”) for a comparison of jet substructure between these two classes of activity, due to a limitation in statistics in the simulation that prevented narrower ranges. When more simulation was produced in 2022, it became possible to select narrower (10%) ranges. In this analysis, no attempt was made to connect the event activity measured in the BBC to the geometry of the initial collision, since there is a much looser correlation of impact parameter, b, to the measured multiplicity than there is in AuAu collisions (see Ch. 5 of Ref. [101] for a detailed explanation). Therefore, we report EA rather than centrality ranges.

4.2 Track selection

The aim of track selection is to reduce background and improve precision and accuracy of reconstruction, namely in the momentum resolution. To reduce background, tracks are required to be primary (as mentioned in Sec. 3.2.1), with the distance in three dimensions of the corresponding global track to the vertex required to be less than 1 cm. To improve reconstruction accuracy, the tracks are required to contain 20 or more fit points (including the vertex), and use at least half of the possible fit points along the track’s path

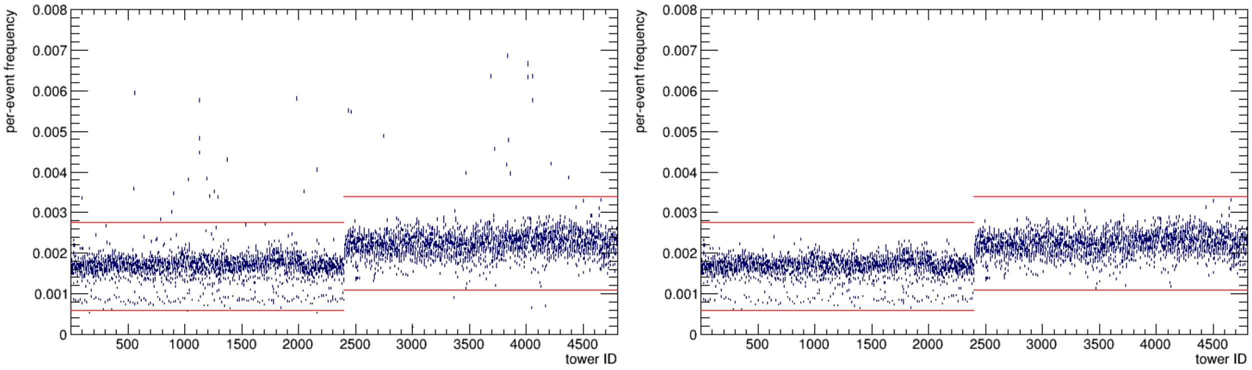
($N_{\mathrm{hits,fit}}/N_{\mathrm{hits,possible}} \geq 0.52$[1]) since it would otherwise potentially be falsely found to be two separate tracks each with half of the fit points, but still potentially more than the minimum. Tracks below 200 MeV/c loop in the TPC without hitting enough pad rows, and tracks above 30 GeV/c have poor momentum resolution, as mentioned in the previous section, so we require $0.2 < p_{\mathrm{T}} < 30$ GeV/c. Finally, to ensure the same acceptance between the TPC and the BEMC, we require $|\eta| < 1$.

4.3 Tower selection

Although track selection criteria can be chosen in advance, and applied to all tracks uniformly to obtain a set of good tracks, the tower selection must be done on a run-by-run and tower-by-tower basis, since each of the 4800 towers can have either a hardware or calibration error at any given time, for any duration. As mentioned in Sec. 4.1, the calibration problem of a particular crate led to the exclusion of the runs for which it existed. That leaves hardware errors, which manifest as either “hot/cold” or “dead” towers. The former term refers to towers which fire, on average, too frequently/infrequently or with too high/low an ADC value; the latter term refers to towers which do not fire, e.g. because of physical damage. Status tables are provided which indicate known problems with certain towers at certain times, although some towers may not be caught by the procedure which generates the status tables, and some may be falsely identified as having a bad status. For this reason, it is typical to remove towers in addition to those with bad status, especially given the large effect a single hot tower can have on a jet spectrum. The additional malfunctioning towers were removed from the analyses in this thesis by placing a 3σ selection on the per-tower average firing frequency (Fig. 4.3), E_{T}, and $E_{\mathrm{T}} > 2$ GeV, from the minimum-bias data. In the pAu data, for example, the procedure removes 394 bad towers, while 28 more that were not considered bad by the procedure are removed for having bad status at any time (neglecting those from bad runs that are already omitted) during the run period, as determined by the status tables.

---

[1] Before the iTPC [103] was installed, this threshold corresponded to $((1/2)N^{\mathrm{max}}_{\mathrm{hits,fit}} + 1)/N^{\mathrm{max}}_{\mathrm{hits,fit}}$, where $N^{\mathrm{max}}_{\mathrm{hits,fit}} = 45 + 1$, corresponding to the 45 pad rows and 1 vertex point.



(a) (b)

Figure 4.3: Average firing (greater than 0.2 GeV) frequency of towers in the 2015 minimum-bias data, before (a) and after (b) removal of bad towers. The red lines denote $\mu \pm 3\sigma$. Note that towers in the east half of the barrel (towerID > 2400) and the west half of the barrel are handled separately due to the slight systematic difference between the two.

The remaining good towers have similar kinematic requirements as do tracks in the TPC, namely $0.2 < E_{\mathrm{T}} < 30\,\mathrm{GeV}$ and $|\eta| < 1$. Additionally, since charged particles deposit energy in the electromagnetic calorimeter as well, there must be a procedure for subtracting the energy due to this track, or else a 10 GeV charged particle can appear in the detector as a 10 GeV track and a few-GeV tower (depending on the amount of energy it deposits in the BEMC). Previously, the standard subtraction was taken to be the energy of a minimum ionizing particle (MIP) for each track which extrapolates through (matches to) a tower. These minimum ionizing particles are charged hadrons (e.g. $\pi^{\pm}$, $K^{\pm}$, and $p/\bar{p}$), which often exit the BEMC having only deposited some small fraction of their energy. The MIP peak (most likely deposited energy) for charged hadrons in the STAR detector with $p_T \approx 1\,\mathrm{GeV/c}$ is located at about 20 to 30 MeV [104] (Fig. 4.4). However, when the average minimum ionizing energy is subtracted from towers, there is still residual contamination due to particles which deposited more of their energy, the variation of which affects the jet energy resolution (JER). So instead of subtracting MIPs from track-matched towers, we subtract the full $p_{\mathrm{T}}c$ of the track from the tower, unless the tower energy would go negative, in which case the energy of the tower is set to zero [105]. Although this is certainly an overestimate[2] of

---

[2] Overestimation is a concern because a neutral particle could strike the same tower, so subtracting more

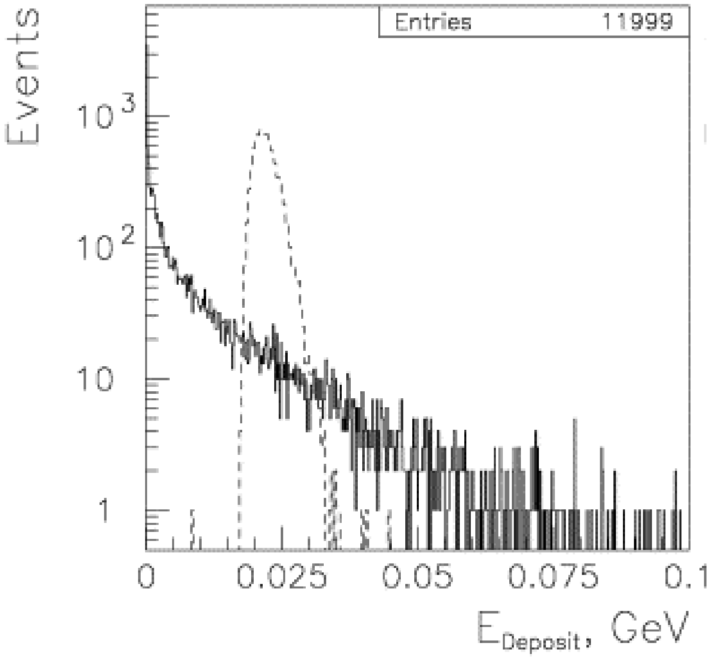


Figure 4.4: The MIP peak (dashed) for 1 GeV/c charged hadrons overlaid on the distribution of deposited energy from neutral particles (solid). [104]

the energy deposited in towers – since the BEMC depth is only approximately 80% of one hadronic interaction length, λ – the residual fluctuation is minimal, so the JER is improved by roughly 25%. An additional benefit is that the neutral energy fraction (NEF) of jets is closer to the theoretically expected value. The uncertainty of this approach is estimated by varying the subtraction from the nominal $1.0 p_{\mathrm{T}} c$ to $0.5 p_{\mathrm{T}} c$ (see Sec. 5.4).

4.4 Jet selection

After we obtain a list of tracks and towers for an event which pass selection criteria, they are clustered with the anti-$\mathrm{k_T}$ algorithm (Sec. 2.1) to obtain a list of jets. Jets considered for the analyses are those which are entirely within the fiducial acceptance of the detector ($|\eta_{\mathrm{jet}}| < 1 - R$), which are not significantly biased by the JP2 trigger, and which are likely not to be due to background. To reduce the bias on the jet momentum spectrum from the trigger, we consider only jets with $p_{\mathrm{T}} > 15$ GeV/c, as this is relatively far removed from the trigger turn-on at about 7 GeV/c. To reduce the likelihood of a jet being from beam

---

than the energy that the charged hadron deposited would result in at least some subtraction of the neutral particle's energy, which is unnecessary since it was not tracked by the TPC.

background or cosmic rays, neither of which would produce primary tracks which extrapolate to a vertex, we exclude jets with more than 90% of their energy from towers. For the jet mass analysis, we additionally exclude jets with $M < 1\,\mathrm{GeV/c^2}$. This selection is discussed in Sec. 5.2. Groomed jets have no selection on mass, other than the requirement on the corresponding ungroomed jet. We do however require for M_g and R_g measurements that groomed jets have a $z_g > 0.1$, which is equivalent to using the SoftDrop algorithm in tagging mode (and is IRC safe for $\beta = 0$; see Secs. 2.1 and 2.3 for more detail about IRC safety and SoftDrop grooming parameters). For the z_g observable, we include jets which did not pass the SoftDrop criterion in the $z_g = 0$ bin. See Table 4.1 for the number of jets after all of the selections mentioned in this chapter. Given the reduced statistics for the pAu analysis when split roughly by a factor of ten for a given iBBCE decile, in Ch. 6 we will show comparisons to the published pp analysis in the same p_T selections shown in the table, but for broader (30%) EA ranges (see Sec. 4.1 for a definition, and cf. Fig. 6.4, where 30% corresponds to three of the deciles marked in that figure with vertical lines). And for comparisons between activity selections in pAu, where we want to separate the activity as much as possible to have a chance to observe an effect if it exists, we will use narrow activity ranges (e.g. 0–10%), but broaden the p_T range.

Table 4.1: Number of $R = 0.4$ jets considered in the analyses for each transverse momentum range, and for a given (0–10%) activity range for the pAu data.

	$N_\mathrm{jets} \times 10^4$		
p_T range [GeV/c] / Collision species	$20--25$	$25--30$	$30--40$
pp	12.6	2.8	0.8
pAu (0–10%)	0.5	0.09	0.02

4.5 Embedding

Data are corrected for detector effects by a Bayesian unfolding procedure that will be described in Sec. 5.3, which allows for comparison between these measurements and theoretical calculations, models, and data from other experiments. This correction approach relies on an accurate simulation of both the physical processes involved and the interactions of the collision products with the STAR detector. For the former, we nominally use the PYTHIA-6 [106] Monte Carlo event generator (discussed in more detail in the next chapter), which has been tuned and subsequently conclusively validated for inclusive jets from pp collisions at STAR (see Appendix C for more detail). These are our so-called "particle-level" events. For the detector interactions, we run the PYTHIA particles through a GEANT-3 detector simulation [107] (producing what we call "detector-level" events). However, even in pp collisions, there will be some activity due to e.g. beam background and cosmic events that the simulation will not capture. For this reason, the detector-level simulation is embedded into real data from the run period being studied. In the case of the pp analysis, these were zerobias-triggered pp collisions[3]. The agreement between this detector-level simulation and the data from pp collisions is excellent, and will be shown in the next chapter. In the pAu analysis, the embedding data were VPDMB30-triggered (Sec. 3.2.3) pAu events. Although the VPDMB30 trigger is a minimum-bias (MB) trigger, this additional activity is found to increase the reference multiplicity (ref. mult.) at mid-rapidity beyond that of the data, which also results in a non-negligible difference in the jet mass. To correct for this, an additional weighting (described in more detail in Ch. 6) is done to force the ref. mult. distribution in the embedded simulation events to match that in the data events. After this weighting, the detector-level embedding agrees with the data, which allows it to be used for Bayesian unfolding.

---

[3] A zerobias trigger relies only on the RHIC clock, i.e. on the coincidence of the two beams, and is heavily prescaled.

CHAPTER 5 JET MASS IN pp COLLISIONS

In this chapter, I present the first inclusive measurement of jet mass in pp collisions at RHIC, which was published in 2021 (Ref. [108]). One motivation for the analysis was to obtain a baseline for a similar measurement in heavy-ion collisions [55] (Sec. 1.3) since the jet mass is related to the initiating parton's virtuality, and can be modified through interactions with the QCD medium (typically decreasing more quickly) [109]. Although many measurements and calculations have been performed at LHC kinematics for pp collisions (see Refs. 10–16 and 28–35 of Ref. [108]), RHIC presents a unique opportunity due to its lower $p_{\mathrm{T,jet}}$ range, which increases the relative effect of hadronization and underlying event (UE) [110] giving an indication of the size of the hadronization correction necessary for perturbative calculations in this kinematic regime, and putting UE tunes of Monte Carlo event generators to the test (Sec. 5.1.1). For these reasons, independently of a heavy-ion measurement, this analysis is necessary as a test of both perturbative and non-perturbative QCD, and state of the art models. Additionally, we are able to isolate the perturbative aspect of the jet utilizing SoftDrop grooming techniques (Sec. 2.3) which reduces the contribution of NP effects such as UE and hadronization, and allows for a more direct comparison to analytic calculations [111]. Finally, a comparison between the ungroomed and groomed jet mass allows us to determine the degree of importance of NP effects across the experimentally examined phase space.

The chapter is organized as follows. First, we discuss the simulation (Sec. 5.1) that we assume as a prior for our physics expectation in this kinematic regime. We start with this because we use it in detector simulations (Sec. 4.5), and because it determines some of the selections made on the jets in data. Using one of these simulations, we discuss the modeled jet mass scale and resolution in the STAR detector (Sec. 5.2), followed by the Bayesian unfolding which corrects the data to particle-level by accounting for detector effects (Sec. 5.3). Next, we determine the systematic uncertainties on the data (Sec. 5.4) and present the fully corrected results (Sec. 5.6), and finally discuss an interesting application (Sec. 5.7) of the data to the improvement of one of the simulations mentioned in the next section, demonstrating the iterative nature of scientific understanding.

5.1 Simulation

5.1.1 Models

Because of the narrow window of applicability of perturbation theory in QCD, we must supplement these calculations with phenomenological models of the physics that occurs at lower energy (larger distance) scales. These are packaged together into general purpose Monte Carlo event generators like PYTHIA [112] and HERWIG [113]. Each model uses different mechanisms for carrying out e.g. the parton shower and hadronization (for more details, beyond the scope of this thesis, consult the comprehensive review in Ch. 43 of Ref. [14]). HERWIG's parton shower, for instance, is angular-ordered, while PYTHIA-6 (8) uses a virtuality (p_T) ordering. PYTHIA's hadronization is done using the Lund string model which keeps track of color connections between the partons as they shower and their potential increases linearly with distance. String breaks occur when this potential is greater than the energy to form a $\mathrm{q\bar{q}}$ pair, and continue until the invariant mass of all pairs is roughly at the mass of hadrons at which point these string-connected partons are algorithmically assigned physical masses and become hadrons. HERWIG, on the other hand, forces gluons to split into $\mathrm{q\bar{q}}$ pairs at some scale, then forms color singlet clusters from the final quarks. If a cluster has too high an invariant mass, it fissions into lighter clusters (Ref. [114], Sec. 7.2), which then decay into hadrons.

Even within a single one of these models, there are many parameters that have no *a priori* correct value, especially given the multiple physical scales that must be described correctly at once. A single collection of choices for all parameters in the model is called a tune, and typically is obtained by comparing the output of the model with measurements of some large set of observables (Sec. 5.7). Here, a measurement of jet mass can be especially helpful in tuning parton shower and hadronization parameters. Unfortunately, most tuning exercises within PYTHIA and HERWIG are performed solely using LHC and Tevatron data.

In theory, parameters can be extrapolated to lower energies, but in practice, there is room for improvement of the extrapolations by including data from lower energies. The default tune for HERWIG-7.0 is in fact explicitly (in the authors' own words) "not suitable for lower energy runs (below $\sqrt{s} = 300\,\mathrm{GeV}$)" [115]. With that in mind, we compare to PYTHIA-8 and HERWIG-7 with LHC tunes not because we expect excellent agreement, but rather to motivate the community to develop tunes that describe RHIC data as well. Exactly such a task was undertaken for PYTHIA-8 (see Sec. 5.7), using (among other measurements) the corrected jet mass and groomed jet mass data presented in this chapter. For more detail on the model settings and tunes used in this analysis, please consult Appendix C.

5.1.2 Particle decays

Because the GEANT simulation (Sec. 4.5) handles decays of particles which are long-lived in the reference frame of the detector (since the parents may reach the detector material before decaying), the particle-level PYTHIA-6 which is passed to it, and which is used as the prior in the Bayesian unfolding, does not include the (mostly weak) decays of the following particles (and their anti-particles): π^0, π^+, η, K^+, K^0_s, K^0_L, Λ^0, Σ^0, Σ^-, Σ^+, Ξ^-, Ξ^0, Ω^-[1]. This has an impact on the jet mass – as can be evaluated by simulation (Fig. 5.1) – partly due to the fact that decay products of parents originally inside (outside) the jet may fall outside (inside) of the jet cone, reducing (increasing) the mass in this case, and partly due to the particles' rest mass assignment (Sec. 5.1.3). However, this effect is not a problem as long as we are consistent in comparing only to undecayed Monte Carlo or calculations. This is due to the fact that the data is corrected with an undecayed PYTHIA-6 prior which at detector-level also produces good agreement between simulation and data when using this rest mass assignment (Fig. 5.11).

5.1.3 Rest mass assignment

Since the jet mass is the magnitude of the sum of four momenta of the particles in the jet, the rest mass of these particles contributes to the overall jet mass. However, due to

---

[1] For the $\pi^\pm$, $K^\pm$, and K^0_L, this is the program default, since their proper lifetimes are larger than the threshold of $1\,\mathrm{m/c}$.

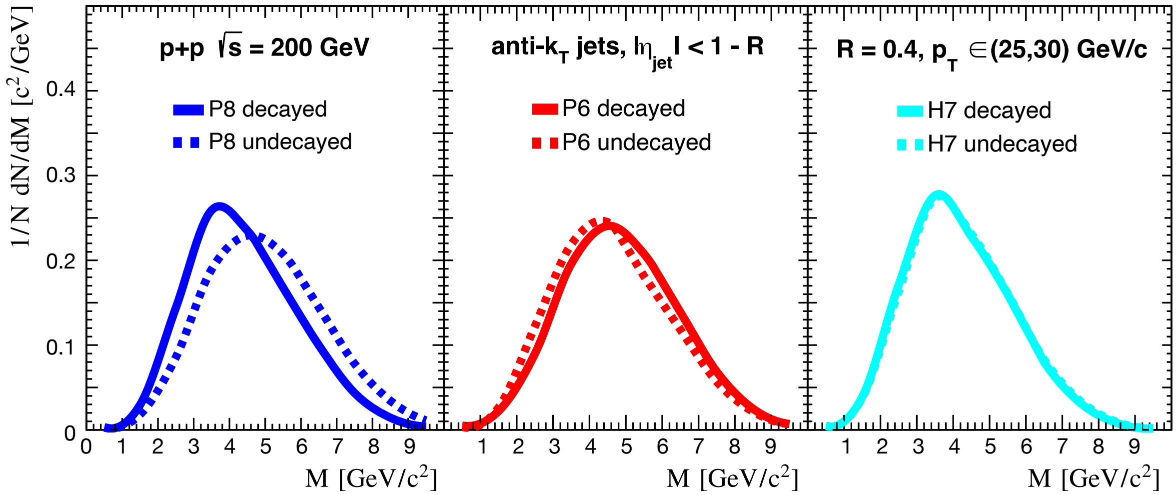


Figure 5.1: The effect of decays of long-lived particles on the jet mass for three different Monte Carlo event generators, at a given jet p_{T} and R. Left: PYTHIA-8. Middle: PYTHIA-6. Right: HERWIG-7.

the difficulty of PID for high-momentum particles, where e.g. the $\mathrm{d}E/\,\mathrm{d}x$ distributions for various hadrons converge (see Ref. [92], Fig. 11), the analysis does not attempt any particle identification, so we must make some choice for assignment of this rest mass in data. We could either assume all rest masses are negligible and set them to zero, or we could assume that towers contribute zero rest mass while tracks are all charged pions (which are the most common charged hadrons [116]) and take $m_{\pi^\pm}$. The latter is more realistic (and has been used in previous analyses, e.g. Ref. [55]) and therefore is our choice for both pp and pAu analyses. The rest mass of particles in simulation, just before they interact with the detector, is known exactly and is given by the PDG [14] value, which gives us a third choice of rest mass assignment which we use for only the particle-level simulation. See Fig. 5.2 for a study of the effect of each choice on the jet mass at particle-level as a function of jet p_{T}. It is evident that there is a large difference between the jet mass given by the simulation "truth" in the red markers and the other possible mass assignments (massless in blue, and $m_{\mathrm{ch}} = m_{\pi^\pm}$ in green). This implies that, especially given that weak decays are turned off, there is a significant quantity of heavier particles in these jets being misassigned $m = 0$ or $m = m_{\pi^\pm}$. Since the tower assignment is $m = 0$, "heavier" in this case can also include

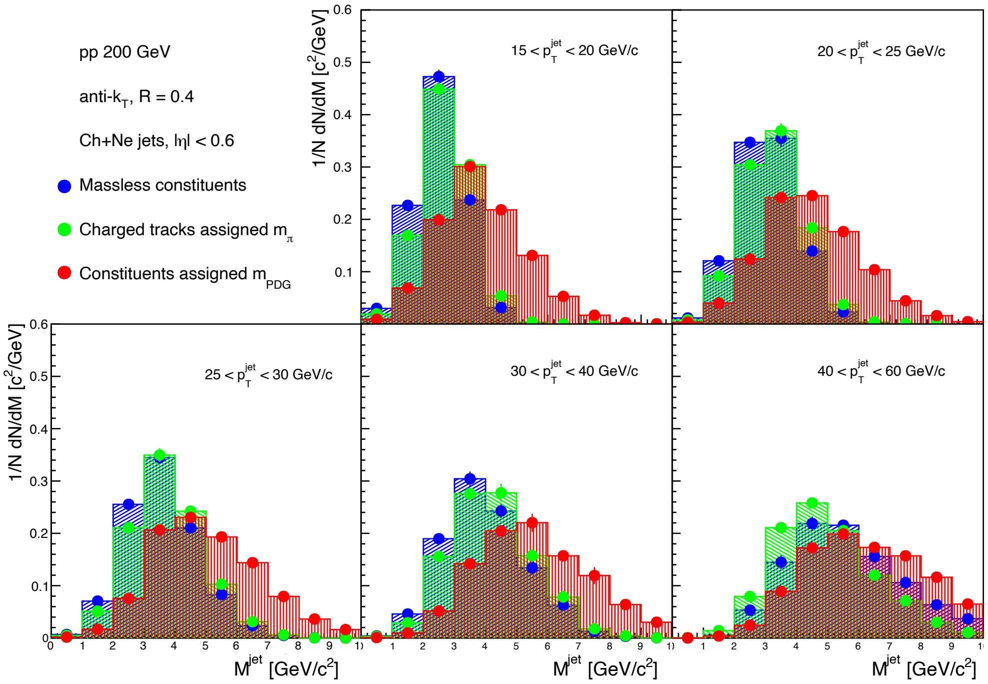


Figure 5.2: The effect of particle rest mass assignment on jet mass using a particle-level PYTHIA-8 simulation with final-state hadrons having no weak decays.

relatively light, and abundant, neutral hadrons such as π_0s. This misassignment leads to some discrepancy between particle-level simulation and data, but it is folded into the jet mass scale, appearing as if the jets are "losing" more mass in the detector than they otherwise would (since the actual rest mass is often higher than 0 or $m_{\pi^\pm}$, e.g. for kaons or protons). The Bayesian unfolding procedure (Sec. 5.3) then corrects for this difference as well, and gives results comparable to particle-level simulation with the correct mass assignment. It should be noted also that it is possible to turn off hadronization in the Monte Carlo models to show the effect this has on the jet mass, for which study we use PYTHIA-8. In this case, for technical reasons [117], the rest mass of the final state partons is not the PDG value. For example, a charm quark weighs 1.5 GeV/c^2 rather than the 1.27 GeV/c^2 given by PDG [14]. In any case, the difference for light quarks, which dominate in the kinematic regime of this analysis, is negligible.

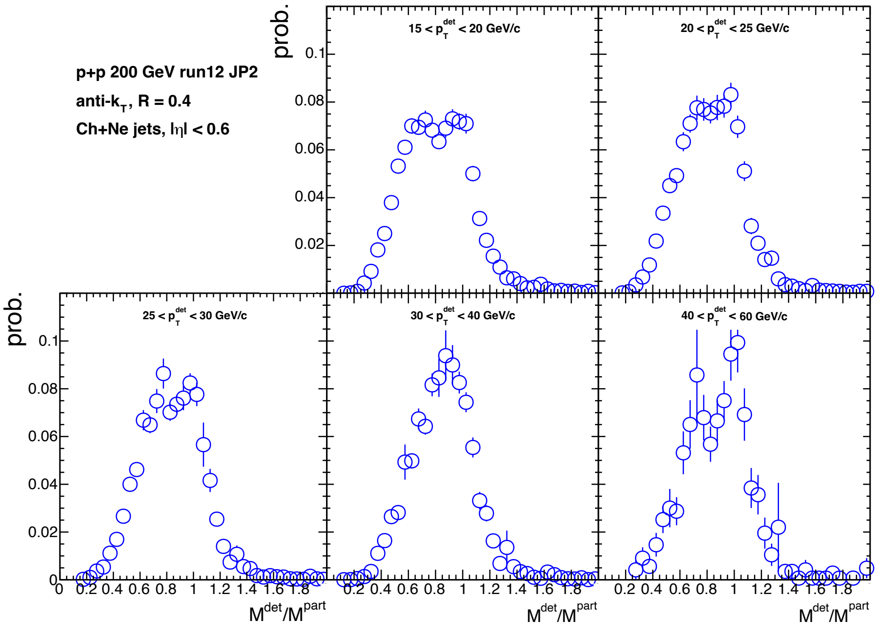


Figure 5.3: The jet mass scale and resolution for a variety of $p_T^{\text{det.jet}}$ selections, with no selection on the particle-level $p_{\text{T,jet}}$.

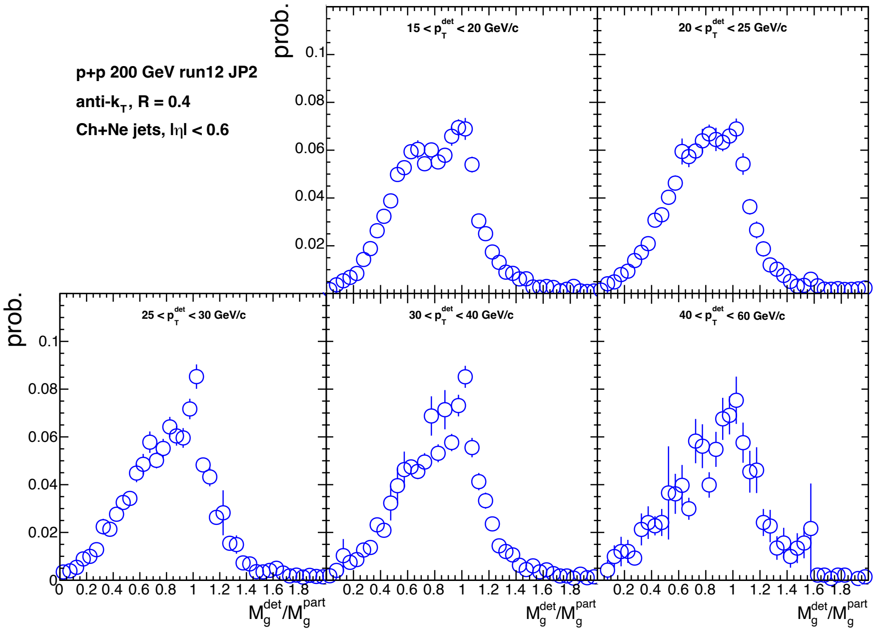
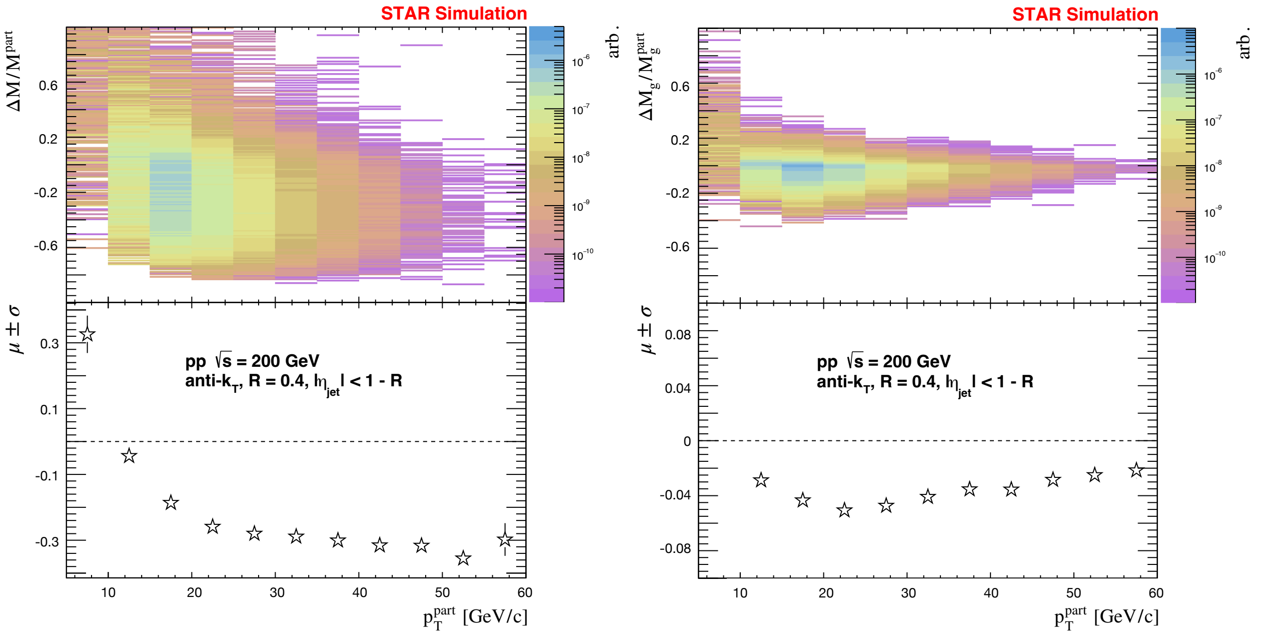


Figure 5.4: The same as Fig. 5.3 but for the groomed jet mass, M_{g}.

5.2 Jet mass scale and resolution

After applying the jet selections mentioned in Sec. 4.4, we can estimate the detector performance using PYTHIA-6 and PYTHIA-6+GEANT events embedded into real zero-bias data from 200 GeV pp collisions in 2012 (cf. Sec. 4.5). For each event, we match particle-level jets to detector-level jets geometrically, requiring the jet axes to be separated by not more than the jet radius, and taking the highest-p_{T} match when more than one jet satisfies this requirement. Fig. 5.3 (5.4) shows the ratio of (groomed) jet mass at detector-level to jet mass at particle-level for various p_{T} selections. See Appendix D for the figures for $R = 0.2$ and $R = 0.6$. There is a bimodality to the distribution for $15 < p_{\mathrm{T}} < 20\,\mathrm{GeV/c}$ that is not evident in the higher-p_{T} selections where the shape is more Gaussian. Similar behavior is seen for the groomed jet mass, although with more populated tails. From these distributions, we can extract the jet mass scale (JMS) and jet mass resolution (JMR), or the mean and width of these distributions, respectively. These are shown in Fig. 5.5. We see that in addition to the fairly large shift of the $\mathrm{JM_{(g)}S}$ from zero, there is a jet p_{T} dependence, which follows the p_{T} dependence of the jet momentum scale caused by tracks lost due to tracking efficiency, a greater percentage of which are lost as p_{T} (and by extension the average number of low-momentum constituent tracks) increases. This p_{T}-dependence of the JMS, along with the p_{T} dependence of the mass itself – see, e.g. Fig. 5.22, motivates the use of a two-dimensional response matrix for the Bayesian unfolding procedure (Sec. 5.3).

The detector performance so far has been shown as a function of jet p_{T}. However, for some subset of jet masses, we may expect a much different performance. Very-low-mass jets, for example, should also have fewer constituents, with higher $p_{\mathrm{T,cons}}$ on average. Due to momentum-dependent single tracking efficiency and momentum resolution, these jets have potentially been affected by the detector in different ways than the overall jet population. It was seen that for these jets, the detector performance worsens significantly (Fig. 5.6), both in terms of JMR (the width of the distribution) and bimodality. The lower peak, causing the bimodality, is partly due to the misassignment of rest mass at detector-level (Sec. 5.1.3).



(a) (b)

Figure 5.5: Upper panels: fractional change of the jet mass (a) and groomed mass (b) in the detector, as a function of particle-level jet p_{T}. Lower panels: The mean for each p_{T} (star marker) with RMS shown as a vertical error bar (typically smaller than the size of the marker). Note: Δ is defined as detector-level minus particle-level.

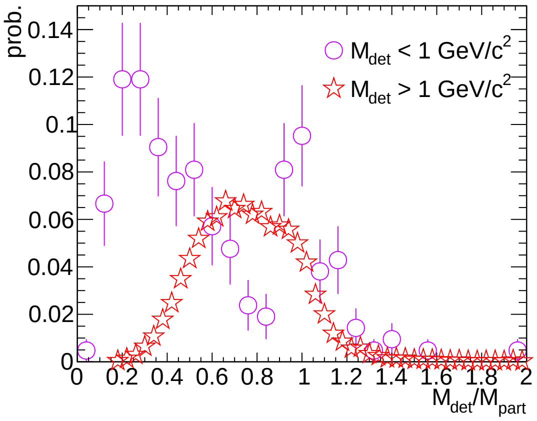
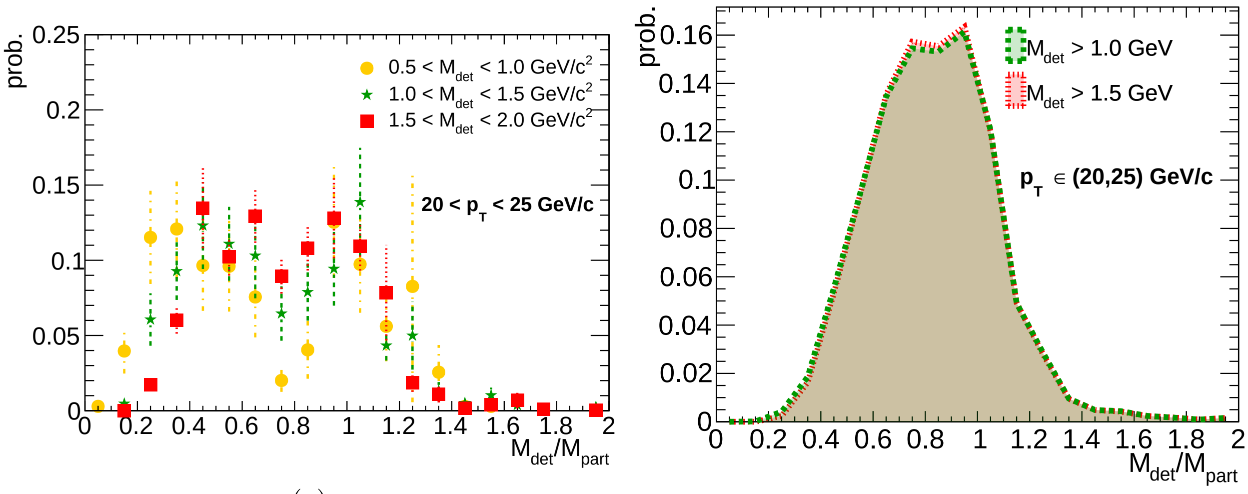


Figure 5.6: The jet mass scale and resolution from PYTHIA-6 ("part", or particle-level) and PYTHIA-6+GEANT ("det", or detector-level), shown for an example p_{T} selection, for detector-level jets failing (magenta circles) and passing (red stars) the jet mass requirement.

As an exaggerated example, take a jet with $M_{\mathrm{part}} = 3.3\,\mathrm{GeV/c^2}$ with three constituents, all protons, which are then each assigned m_{π^+}. The mass of the measured jet even excluding any detector effects which tend to decrease the mass, would be about $0.5\,\mathrm{GeV/c^2}$, which would give $M_{\mathrm{det}}/M_{\mathrm{part}} \approx 0.15$. The same rest mass misassignment on a jet which is measured with, say, $M_{\mathrm{det}} = 4\,\mathrm{GeV/c^2}$ implies an original particle-level jet with $M_{\mathrm{part}} = 6.8\,\mathrm{GeV/c^2}$, which would give $M_{\mathrm{det}}/M_{\mathrm{part}} \approx 0.6$. This explains the approach to a normal distribution as detector-level jet mass increases. See Appendix B for a more detailed discussion of the low-jet-mass detector performance.

The resulting selection that we make for this analysis, and that of Ch. 6, of $M_{\mathrm{det}} > 1\,\mathrm{GeV/c^2}$ is slightly arbitrary (as can be seen in Fig. 5.7a, which demonstrates the gradual improvement of the jet mass resolution). However, raising the threshold to, say, $1.5\,\mathrm{GeV/c^2}$ does not dramatically improve the *overall* jet mass resolution or scale (Fig. 5.7b), and does reduce statistics by roughly 3%. On the other hand, reducing the threshold to, say, $0.5\,\mathrm{GeV/c^2}$ increases the number of jets by less than 1%, so $M = 1\,\mathrm{GeV/c^2}$ was chosen as the threshold. It was also determined that a selection on the groomed jet mass would be detrimental, because of the irregular phase space selection that would be caused by having a requirement



(a)

(b)

Figure 5.7: (a): The jet mass scale and resolution from PYTHIA-6 (“part”, or particle-level) and PYTHIA-6+GEANT (“det”, or detector-level), for three different populations of low mass detector-level jets. (b): The negligible effect on the overall jet mass scale and resolution of changing the jet selection threshold to $M > 1.5\,\mathrm{GeV/c^2}$. Both panels shown for the same example p_T selection.

on both the ungroomed jet and its corresponding groomed jet. Also, given that the groomed mass is bounded from above by the ungroomed mass, this bin of the histograms is closer to the diagonal in the response matrix, which increases its importance for the Bayesian unfolding procedure.

5.3 Bayesian unfolding

The unfolding procedure we use [118] is an iterative Bayesian inference method implemented in the RooUnfold package [119], which returns the most probable physical distributions (“causes”, C) for some observable(s) which, when modified by the detector, resulted in the measured data (“effects”, E), using Bayes’ Theorem for each possible cause (bin) i of the total number n_c, and effect (bin) j:

$$P(C_i|E_j) = \frac{P(E_j|C_i)P(C_i)}{\sum_{l=1}^{n_\mathrm{c}} P(E_j|C_l)P(C_l)}. \tag{5.1}$$

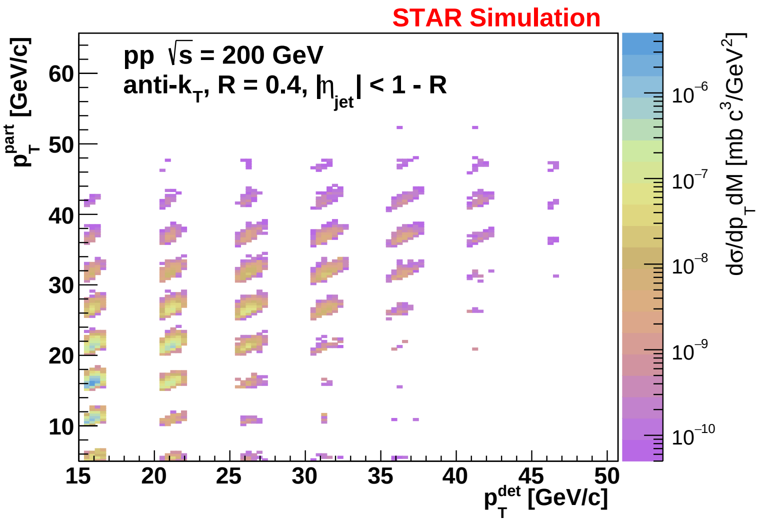


Figure 5.8: The four-dimensional response matrix used to correct the raw jet mass data to particle-level in the Bayesian unfolding procedure. Each 5 GeV/c bin in detector- and particle-level p_{T} contains a two-dimensional histogram of the mass at detector-level (x-axis) vs. particle-level (y-axis) for that range in p_{T}.

In our case, the initial iteration's prior assumption, $P(C)$, is taken as the particle-level distribution given by the simulation (Sec. 4.5), while the likelihood function, $P(E|C)$, is given by the so-called response matrix discussed below. After marginalizing over the E_j in the posterior using the data, the procedure is iterated $n-1$ further times, with the posterior of each step as the new prior in the next. Compared to a bin-by-bin correction in which an efficiency is applied to each bin of a measured distribution to account for e.g. detector effects, the Bayesian approach has many advantages. For instance, the particle- and detector-level distributions may have completely different domains (which is useful for both the jet mass and p_{T} in the unfolding) and be binned differently; and bin migration is automatically accounted for.

The iterative approach allows for less reliance on the initial prior given by a particular Monte Carlo event generator (in this case, PYTHIA-6), and for more influence of the off-diagonal elements. However, at some point, for $n \gg 1$, the unfolding becomes numerically

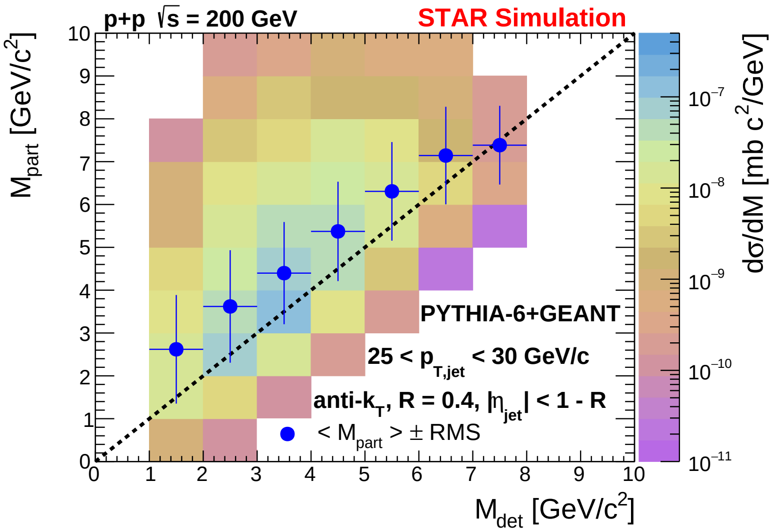
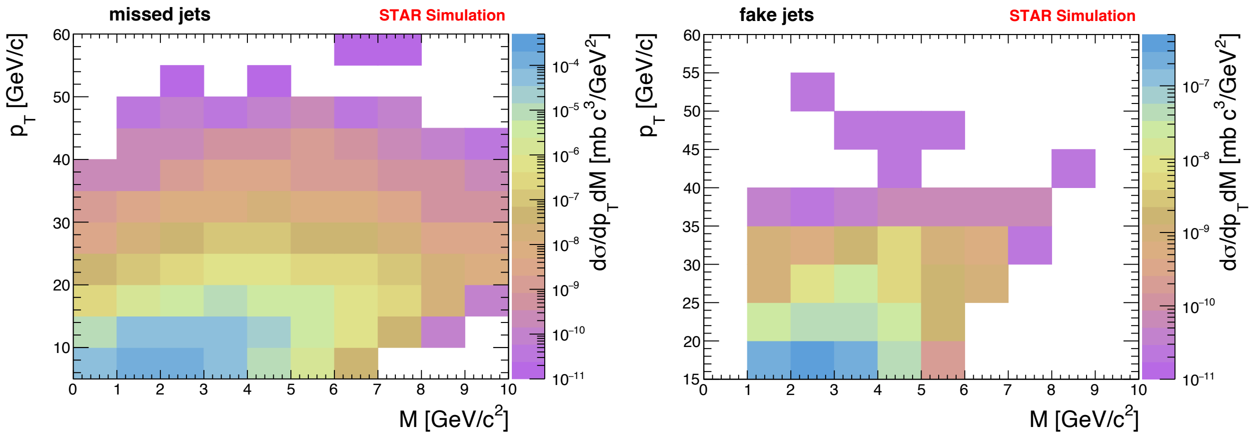


Figure 5.9: For visualization purposes, it is better to look at a single one of the p_{T} bins of Fig. 5.8. We present that here, for $25 < p_{\mathrm{T}}^{\mathrm{part,det}} < 30\,\mathrm{GeV/c}$ as an example. The blue circles represent the mean particle-level jet mass corresponding to a given detector-level mass; the vertical bars are the RMS (the horizontal bars span the width of the bin).

unstable due to statistical fluctuations. The nominal regularization parameter for both analyses presented in this thesis is $n = 4$ because this gives a good Monte Carlo closure (more detail later in this section). However, this number was also varied as a systematic uncertainty (Sec. 5.4), because it is somewhat arbitrary. Statistical fluctuations are also mitigated by requiring greater than 20 entries per bin in the Monte Carlo closure two-dimensional spectra (and the data), then removing the same low-statistics bins from the response matrices used in the closure and the unfolding as well. The response matrix (Figs. 5.8, 5.9) is constructed from the matched particle-level and detector-level jets from simulation, as mentioned in Sec. 5.2. Any jets which fall outside of the detector's fiducial acceptance or fail our selection criteria at detector-level enter the response as "missed" jets (Fig. 5.10a) which act as an efficiency scaling of the particle-level bins. Jets which are found at detector-level but have no match at particle-level enter as "fake" jets (Fig. 5.10b) which are treated as an additional background.



(a)

(b)

Figure 5.10: (a): The missed jet distribution which is used as a bin-by-bin efficiency to scale the response in Fig. 5.8. (b): The fake jet distribution which is treated as an additive background in the Bayesian unfolding procedure.

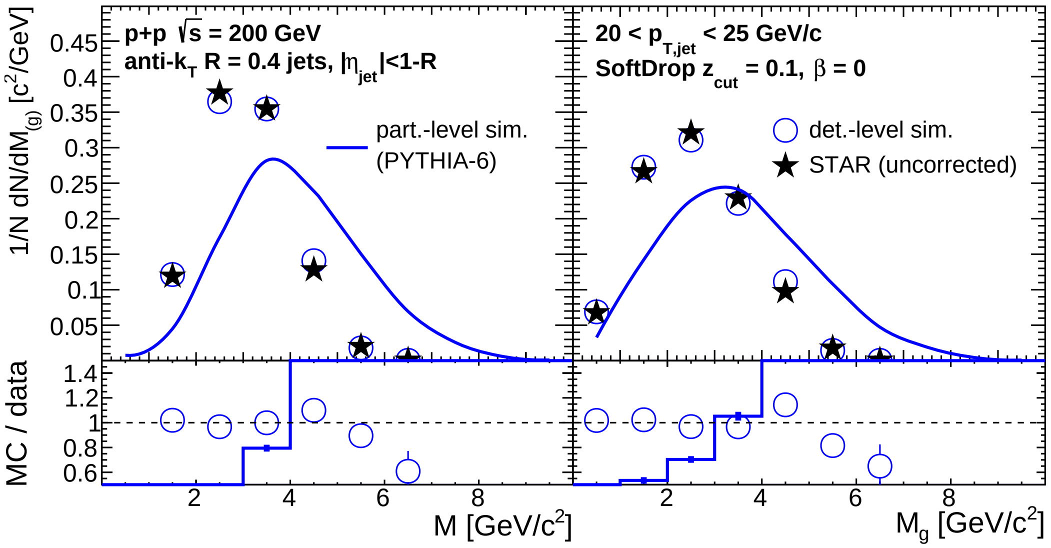


Figure 5.11: Jet mass (left) and groomed mass (right) for particle-level simulation (blue line), detector-level simulation (blue circles), and uncorrected data (black stars), for the selection $20 < p_{\mathrm{T}} < 25\,\mathrm{GeV/c}$. The lower panels show a ratio between Monte Carlo simulation and data. Statistical uncertainties are mostly smaller than the size of the markers.

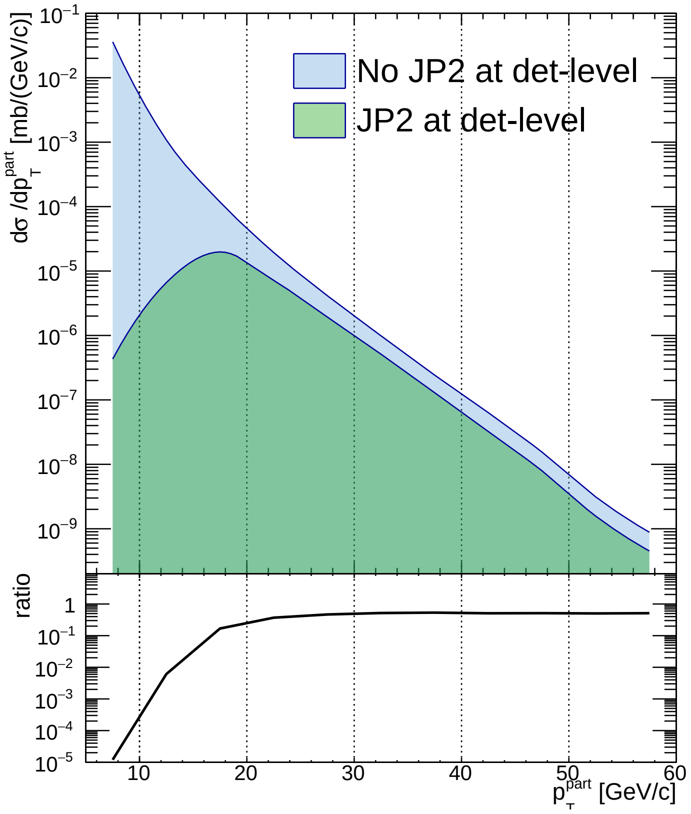


Figure 5.12: The effect on the particle-level p_T spectrum (our prior) due to the JP2 trigger turn-on's effect on the matching to detector-level jets, modeled in PYTHIA-6+GEANT simulation. The bias (in the shape) is reduced to a reasonable amount by $p_\mathrm{T} \approx 15\,\mathrm{GeV/c}$.

For the unfolding procedure, it is also important that the detector-level simulation has sufficient agreement with the data. Fortunately, the detector-level simulation we have from STAR-tuned-PYTHIA-6+GEANT matches the jet substructure in data quite well (Fig. 5.11). We also note that the jet mass at particle-level is much larger, as expected, due mostly to track loss in the detector, which reduces the four-momentum sum.

Jet kinematic selections were discussed briefly in Sec. 4.4. One such choice which deserves more explanation relates to the jet p_T selection. For the Bayesian unfolding, jets at detector-level in simulation, and in the data, are required to have $p_\mathrm{T} > 15\,\mathrm{GeV/c}$ due to the effect the JP2 trigger turn-on has on the particle-level jet spectrum below this value (Fig. 5.12). Above this threshold, the spectrum is relatively unbiased in shape (overall yield is not important for this per-jet analysis). However, the proportion of neutral momentum in the jet is still biased to higher values by the trigger until 20 GeV (Fig. 5.13), which leads to both a biased jet fragmentation, and a potentially increased gluon jet content in this bin and by extension

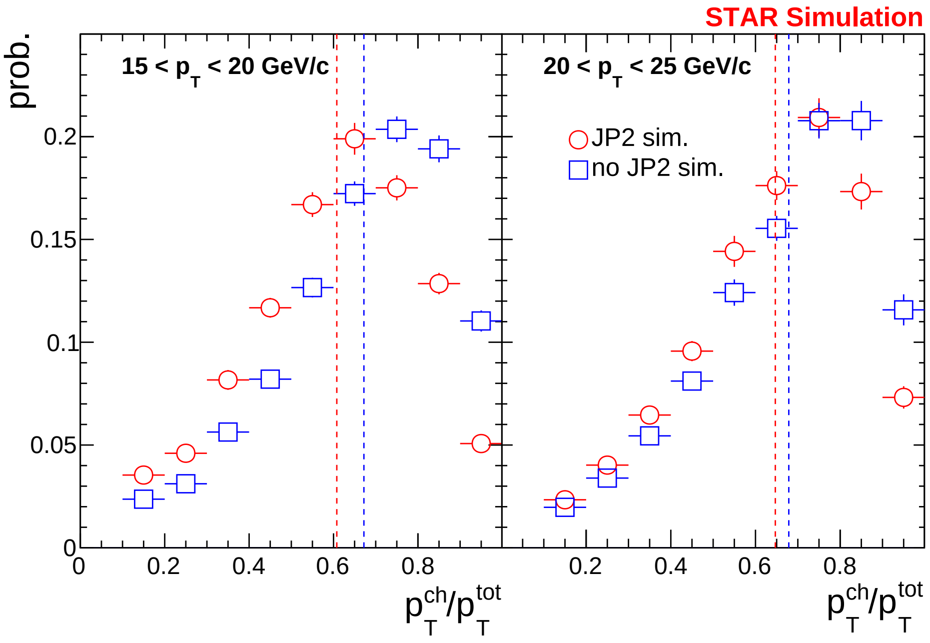
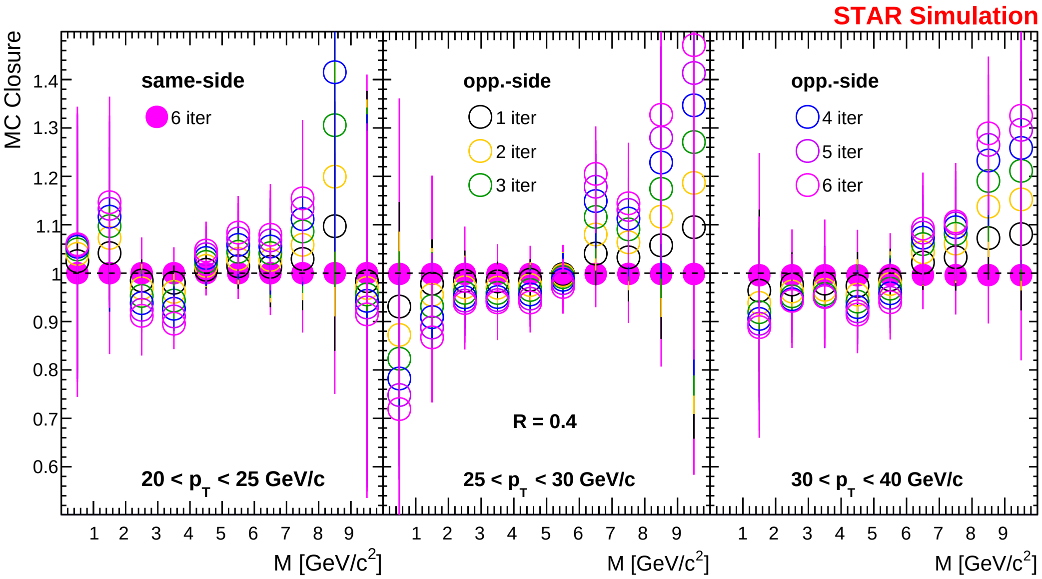


Figure 5.13: The effect on the jet's charged momentum fraction due to the JP2 trigger's requirement of some amount of neutral energy, modeled in PYTHIA-6+GEANT simulation. The bias is minimal by $p_{\mathrm{T}} = 20\,\mathrm{GeV/c}$ (right panel). Vertical lines denote the means of the distributions.

an increased jet mass (since gluon jets have larger mass on average than quark jets due to their increased color factor, $C_A/C_F = 9/4$ – see, e.g., Sec. 6.4 of Ref. [28], and Ref. [120]). Typically, in the absence of detector biases, quark jets are dominant for jet $x_{\mathrm{T}} \gtrsim 0.2$ (the kinematic regime of this thesis; see Ref. [60] for details). Because of the potential bias, although jets are allowed to enter the response matrix if $p_{\mathrm{T}} > 15\,\mathrm{GeV/c}$, the corrected result is only shown (Sec. 5.6) for $p_{\mathrm{T}} > 20\,\mathrm{GeV/c}$.

Before we move on to actually unfold the data, we should test on pseudodata that the procedure works. For this, we have "same-side" and "opposite-side" Monte Carlo closure tests. For each, we split the detector-level jets in simulation randomly into two populations, call them A and B. For the same-side MC closure, we simply test that the unfolding mechanically works, with a tautology: if one measures a set of effects from a known set of causes, then unfolding should simply return the distribution of the causes. If unfolding detector-level population A with response A does not return the particle-level population A,



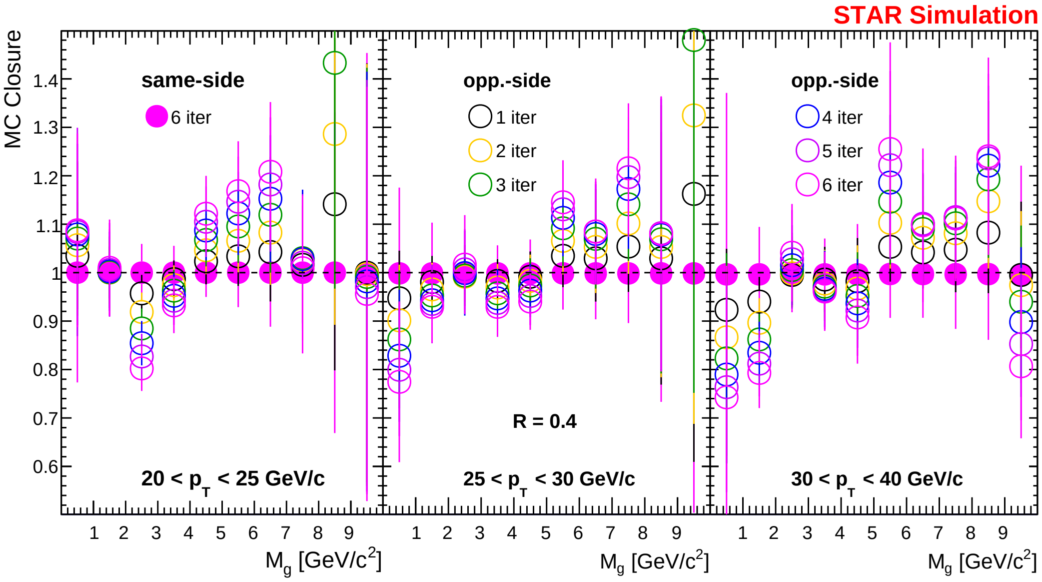


Figure 5.14: MC closure test for the two-dimensional jet mass unfolding procedure, for three selections in jet p_{T} (the same as will be shown for the results). The full circles represent the "same-side" closure (see text for definition), while the open circles represent the "opposite-side" closure. For the latter, we show the dependence on the regularization parameter of the Bayesian unfolding. Upper: jet mass. Lower: groomed jet mass.

within numerical precision, then something is wrong. For the opposite-side MC closure, we test that the unfolding works on pseudodata that is very similar to, but not exactly the same as, the detector-level in the response. This is done by unfolding e.g. population B with the response constructed from population A. If this does not return corrected pseudodata within roughly 10% of the particle-level (at least near the highly populated bins in the distributions), then the model is failing its "testing", to borrow machine learning terminology. The MC closure test for the (groomed) jet mass unfolding is shown in Fig. 5.14. Almost all points are consistent with unity, and the few that are not are either within 10% or far on the tails of the jet mass distribution, where statistical fluctuations have a larger effect. We also note that for 4 iterations (blue markers), which is the nominal for the RooUnfold package, the MC closure does not exhibit quite as poor disagreement on the tails of the distribution as it does for further iterations. Therefore we select 4 iterations as the nominal for the analysis.

5.4 Systematic uncertainties

The approach to evaluating systematic uncertainties for this analysis, and that of Ch. 6, is to vary relevant aspects of the particle-level or detector-level simulation and rerun the entire analysis chain, including unfolding, using that varied parameter, distribution, or effect. Then, dividing the nominal result by the varied one gives a percentage uncertainty for each bin in the distribution. The uncertainties can conceptually be broken into two categories: uncertainties on the detector performance (Sec. 5.4.1) and uncertainties on the Bayesian unfolding procedure (Sec. 5.4.2). In this section we show the uncertainties for $R = 0.4$ jets; the uncertainties for $R = 0.2$ and 0.6 can be found in Appendix D.

5.4.1 Detector uncertainties

There are three sources of uncertainty due to the STAR detector that were considered for this analysis. One is from tracking efficiency, which has a 4% absolute uncertainty [60]. This was simulated by randomly vetoing an additional 4% of all tracks. Another uncertainty is due to the BEMC tower gain which has a calibration uncertainty of 3.8% [100]. This was simulated by augmenting all tower energies by that amount. Finally, we have an uncertainty

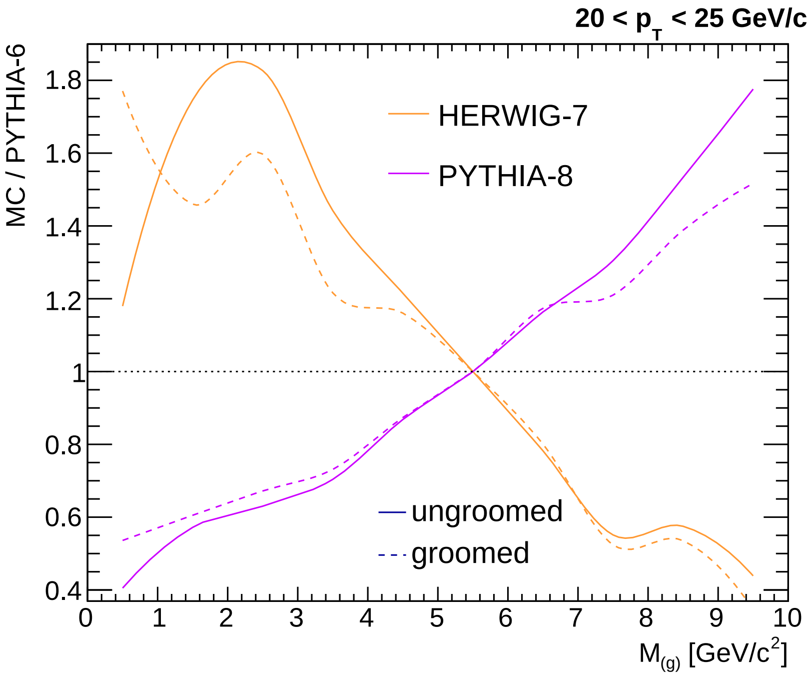


Figure 5.15: The ratios between two models and the nominal prior for the (groomed) jet mass distribution, used as a shape adjustment for evaluation of the systematic uncertainty related to the choice of prior. One set of these distributions exists for each p_{T} range.

due to the fact that charged hadrons typically only deposit some fraction of their energy in the BEMC. We nominally assume 100% of a charged track's p_{T} will be absorbed within the tower that it extrapolates to, although we know that in reality this is not the case (Sec. 4.3). So for this systematic uncertainty, we vary the 100% correction down to 50%. The value of 50% was chosen because it is known that charged hadrons deposit about 30% of their energy on average in the STAR BEMC [60], but this is not the case for electrons, which deposit a significantly larger percentage of their energy. Additionally, the JP2 trigger should have a bias toward towers in which the hadrons deposited more of their energy than average, in order to satisfy the relatively large energy requirement. So 50% was chosen as a reasonable estimate. Each of these uncertainties is, in most bins, subleading to the unfolding uncertainties.

5.4.2 Unfolding uncertainties

The unfolding procedure introduces multiple uncertainties. As mentioned in Sec. 5.3, the number of iterations of the procedure is nominally taken as 4 (the program default). However, we vary this down to 2, below which the Bayesian approach is no longer iterative, and up to 6, beyond which the MC closure test begins to fail (Fig. 5.14). Another set of uncertainties due to unfolding relates to the prior spectrum shape. We use PYTHIA-6 and PYTHIA-6+GEANT for our particle-level prior and detector-level likelihood – to use the Bayesian terminology – respectively (Sec. 4.5). However, to reduce our dependence on these model assumptions, we "smear" the particle-level and detector-level p_{T} spectra by Gaussian functions dependent on the p_{T} resolution, as well as the particle-level (groomed) jet mass spectrum, using the ratio of the nominal PYTHIA-6 to two other distinct Monte Carlos, HERWIG-7 and PYTHIA-8 (Sec. 5.1.1). Specifically, the shift to the particle- or detector-level p_{T} for a jet i is $p_{\mathrm{T},i}^{\mathrm{shifted}} = p_{\mathrm{T},i}^{\mathrm{nom}} - r$, where r is a random number sampled from $f(x) = \exp\left(-\frac{1}{2}\frac{x^2}{(p_{\mathrm{T},i}^{\mathrm{nom}})^2\mathrm{JMR}^2(p_{\mathrm{T},i})}\right)$. The jet mass spectrum ratio between models that reweights the mass distributions is shown in Fig. 5.15. After reweighting, we unfold in only one dimension (mass or groomed mass) for each p_{T} bin, to control for jet p_{T} dependence, since we already smear the p_{T} spectrum as a separate uncertainty.

Fig. 5.16 shows each of the uncertainties mentioned above, for each selection in p_{T} for which we report results. One immediately notes the characteristic shape as a function of mass, with a valley near roughly $4\,\mathrm{GeV}/c^2$, and cliffs on either side for low and high mass. This is due to the shapes of the corrected jet mass distributions, which peak at around this value. Because most uncertainties manifest themselves as a shift of the corrected jet mass to lower or higher values, at the peak this effect is reduced, while a small shift left or right on the steeper tails will cause a much larger difference from the nominal value. It is also clear that variation of the particle-level spectra dominates other sources of uncertainty in most jet mass bins. Lastly, we see that the uncertainty is typically slightly smaller for the groomed jet mass, especially at low-M_{g} due to the shift in the peak, and due to the

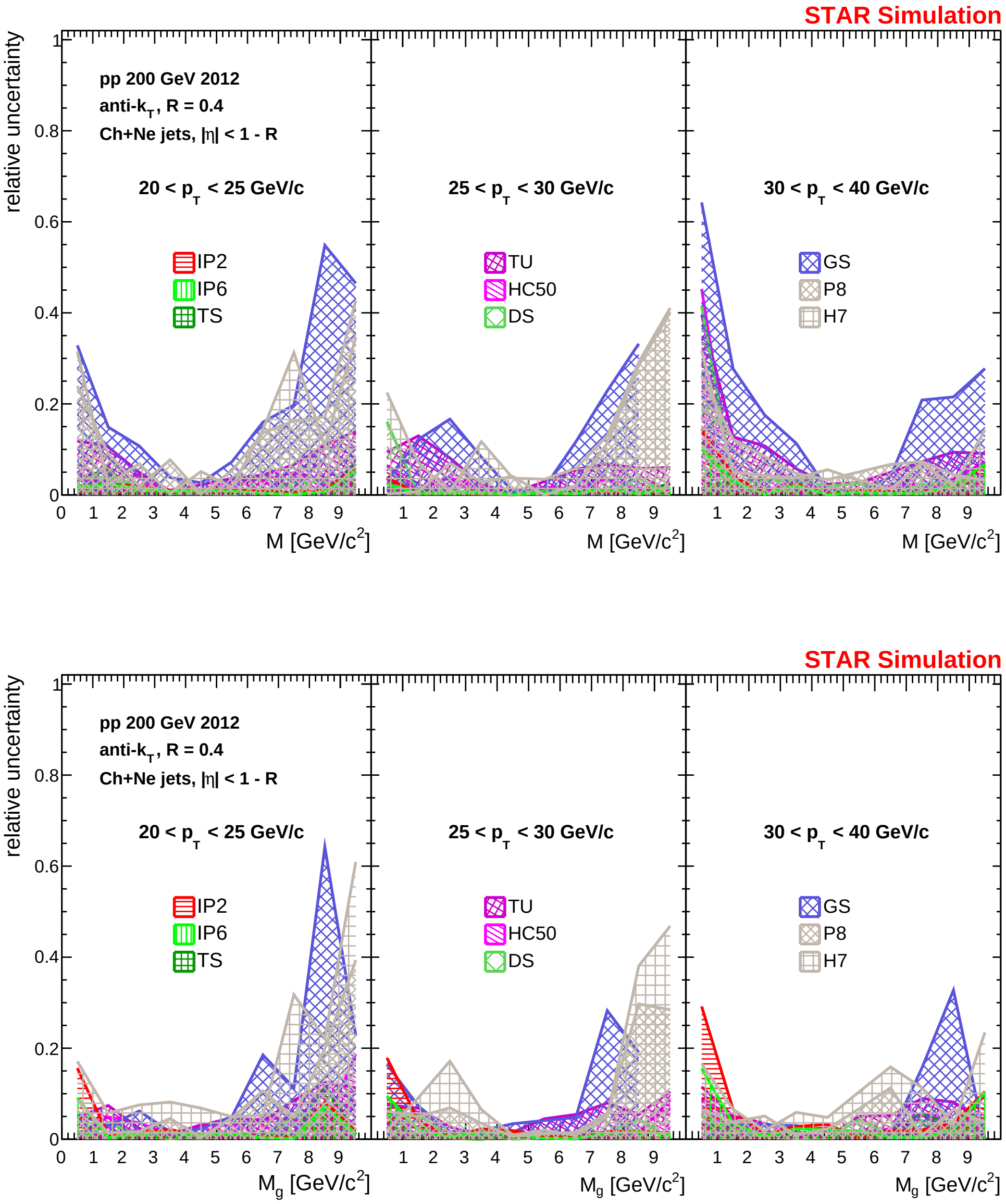


Figure 5.16: The relative uncertainties ($0.1 = 10\%$), by source. Legend abbreviations: IP2(6) refers to variation of the regularization parameter from the nominal 4 to 2(6). “TS” stands for tower scale variation. “TU” stands for tracking efficiency uncertainty. “HC50” stands for a 50% hadronic correction. “D(G)S” refers to smearing of the detector (particle)-level p_T spectrum. “P8” and “H7” refer to variation of the particle-level M or M_g spectrum using PYTHIA-8 and HERWIG-7, respectively. Upper: uncertainty on M. Lower: uncertainty on M_g. Note: due to a plotting bug internal to the program used to make these plots (ROOT6), the “GS” uncertainty for $25 < p_T < 30\,\mathrm{GeV/c}$ is only shown until the penultimate bin.

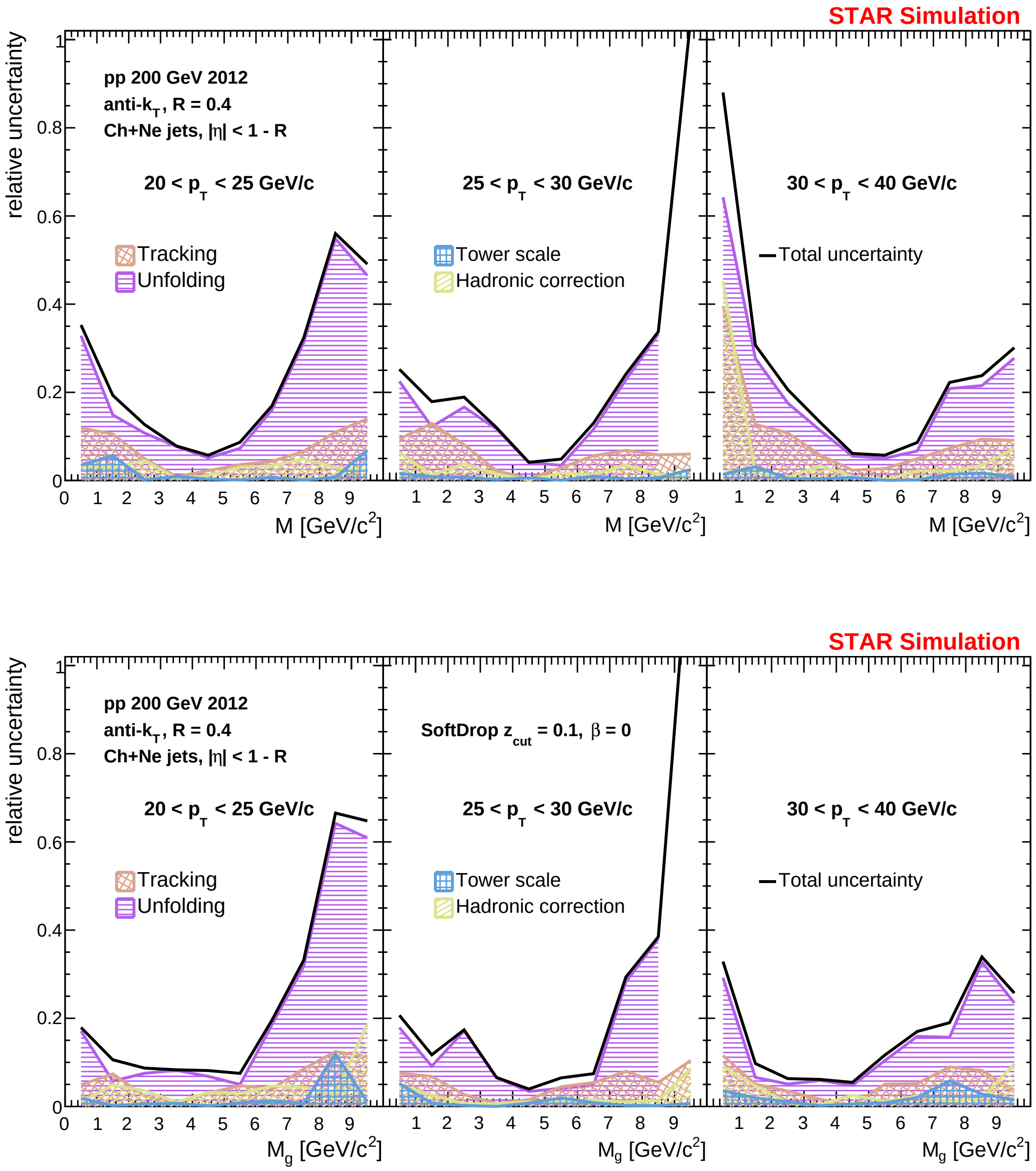


Figure 5.17: This figure is the same as Fig. 5.16, except with a maximum envelope of all unfolding uncertainties taken, and with a quadrature sum of the four uncertainties listed, shown as a black line. Upper: uncertainty on M. Lower: uncertainty on M_g. Note that due to a plotting bug internal to the program used to make these plots (ROOT6), the "Unfolding" uncertainty for $25 < p_\mathrm{T} < 30\,\mathrm{GeV/c}$ is only shown until the penultimate bin. However, it would closely track the black line since it is the largest contribution to the quadrature sum there.

removal via the grooming procedure of low-p_{T} tracks with poorer tracking efficiency. We then take the maximum envelope of all unfolding uncertainties, shown in Fig. 5.17, along with a total systematic uncertainty, which is given by the quadrature sum of this and the detector uncertainties, since we assume they are independent. In most bins, unfolding is the dominant source of uncertainty. Table 5.1 lists the uncertainties due to the detector performance and the overall uncertainty due to the unfolding procedure, as well as the total uncertainty due to all sources, for selected bins of (groomed) jet mass for a more quantitative comparison.

Table 5.1: Systematic uncertainties for an example jet population, with $25 < p_{\mathrm{T,jet}} < 30\,\mathrm{GeV}/c$, and three example selections of (groomed) jet mass corresponding to the low-mass tail, the peak region, and the high-mass tail, from top to bottom. The "Total Systematics" column is obtained by adding the four preceding columns in quadrature. Upper: jet mass. Lower: groomed jet mass. [108]

Source / Range in M	Hadronic Correction	Tower Gain	Tracking Efficiency	Unfolding Procedure	Total Systematics
$(1,2)\,\mathrm{GeV}/c^2$	1.3%	0.9%	13.0%	12.2%	17.9%
$(4,5)\,\mathrm{GeV}/c^2$	0.1%	0.6%	0.4%	4.1%	4.1%
$(7,8)\,\mathrm{GeV}/c^2$	3.6%	0.4%	6.9%	22.9%	24.1%

Source / Range in M_{g}	Hadronic Correction	Tower Gain	Tracking Efficiency	Unfolding Procedure	Total Systematics
$(1,2)\,\mathrm{GeV}/c^2$	2.6%	0.7%	6.8%	9.1%	11.7%
$(4,5)\,\mathrm{GeV}/c^2$	1.0%	0.8%	1.5%	3.4%	4.0%
$(7,8)\,\mathrm{GeV}/c^2$	1.1%	0.2%	8.0%	28.3%	29.4%

5.5 Statistical uncertainties

One last consideration is necessary before the fully corrected results can be shown. The RooUnfold package treats missed jets as true counts for the purposes of statistical uncertainty propagation, although they should in reality be treated as a scaling to the uncertainty, since they also scale the bin content by some efficiency. So instead of the uncertainty on a bin scaling as $\sqrt{\varepsilon^{-1}N_{\mathrm{meas}}}$, for kinematic efficiency $\varepsilon = N_{\mathrm{meas}}/N_{\mathrm{tot}}$ (Fig. 5.18), it should scale as $\varepsilon^{-1}\sqrt{N_{\mathrm{meas}}}$. To achieve this, we scale the overall statistical uncertainty of each mass bin

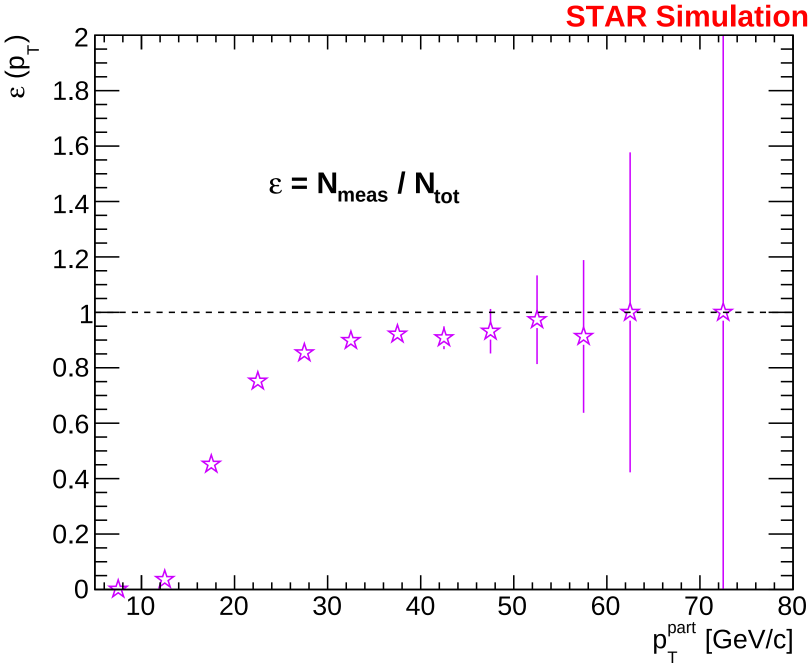


Figure 5.18: Kinematic efficiency, ε, caused by missing jets to the JP2 trigger simulation and jet selections, as modeled by PYTHIA-6 and PYTHIA-6+GEANT. This figure is similar to Fig. 5.12, but in that figure we examine only the effect of the trigger, while here a jet may be lost also due to selection criteria, which is more applicable for the purpose of scaling the uncertainty on the corrected results (see text).

for a given jet p_{T} by $\sqrt{\varepsilon^{-1}(p_{\mathrm{T}})} = \sqrt{1 + N_{\mathrm{miss}}(p_{\mathrm{T}})/N_{\mathrm{match}}(p_{\mathrm{T}})}$. The overall effect of this procedure is minimal for two reasons. First, the largest contribution from missed jets occurs at low p_{T} (as seen in Fig. 5.18), due to jets at particle-level whose detector-level counterpart failed the trigger requirement after detector inefficiencies were applied. Second, due to the steeply falling jet spectrum, the statistical uncertainties are already small for the lowest p_{T} selection (negligible compared to systematic uncertainty even after scaling – see Fig. 5.19). Nonetheless, we correct for this, using scaling factors of approximately 15%, 8%, and 6%, from lowest to highest of the p_{T} selections shown in Sec. 5.6.

5.6 Results

Fig. 5.20 shows the fully corrected jet mass results, published in 2021 [108], for the nominal jet resolution parameter $R = 0.4$ and three selections of $p_{\mathrm{T}}^{\mathrm{jet}}$: $20 < p_{\mathrm{T}} < 25\,\mathrm{GeV/c}$, $25 < p_{\mathrm{T}} < 30\,\mathrm{GeV/c}$, and $30 < p_{\mathrm{T}} < 40\,\mathrm{GeV/c}$. We first notice that increasing the jet p_{T} increases the jet mass distribution mean and width (by roughly 20% in either case, for mass and groomed mass), which is expected due to the increase in available phase space for radiation at these higher jet momenta. As mentioned in Sec. 5.3, quark jets dominate in these p_{T} selections, and the quark jet fraction increases as a function of p_{T} (see Fig. 1 of

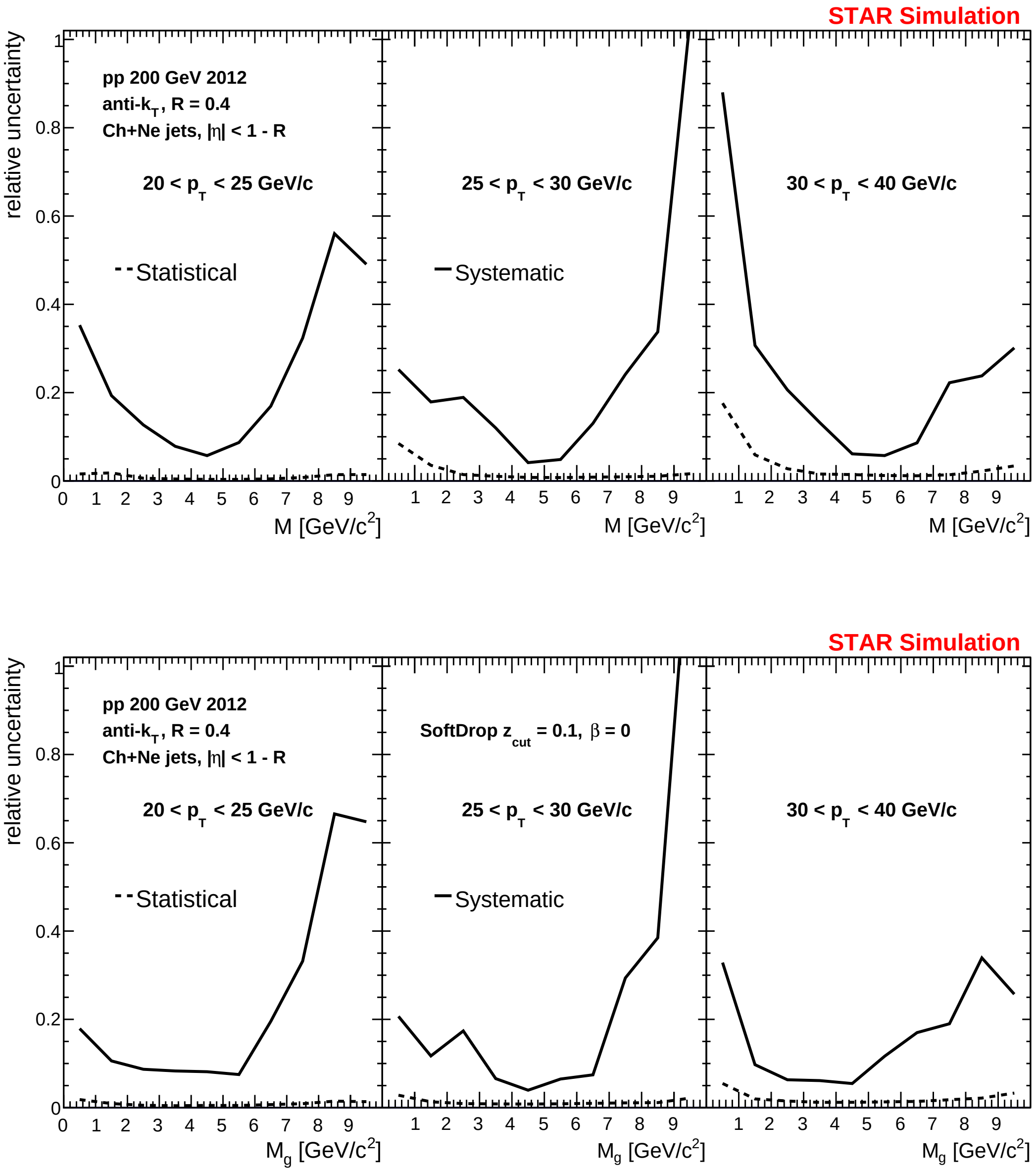


Figure 5.19: A bin-by-bin comparison of the total systematic uncertainty (the black lines from Fig. 5.17) with the statistical uncertainty after it has been scaled by the kinematic efficiency factor described in the text. Upper: uncertainty on M. Lower: uncertainty on M_{g}.

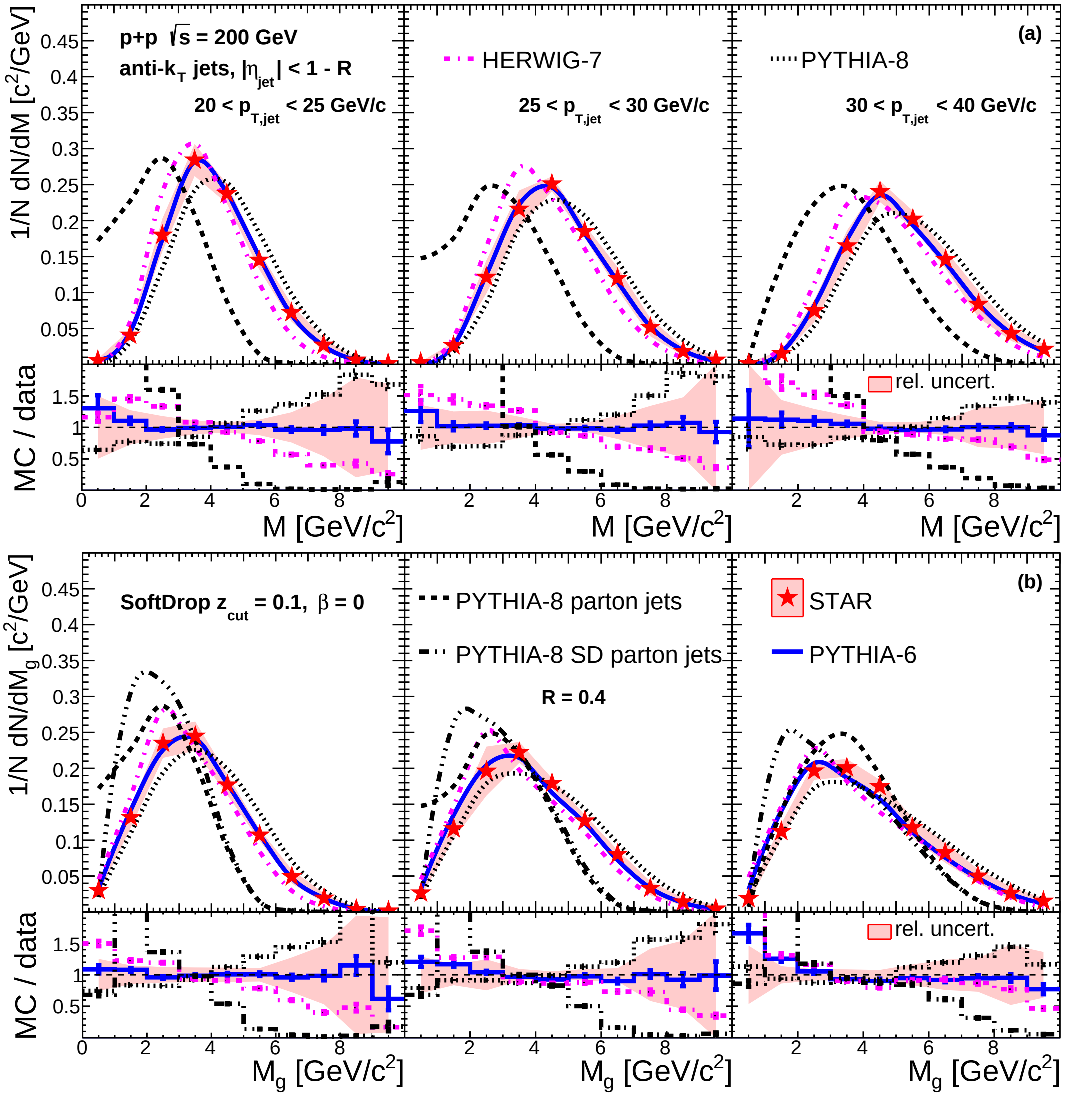


Figure 5.20: Fully corrected (a) jet mass and (b) groomed jet mass for three selections of jet p_{T}, and fixed jet radius $R = 0.4$, shown in red star markers. The data are compared to PYTHIA-6 (solid blue line), PYTHIA-8 (dotted black line), and HERWIG-7 (dot-dashed magenta line), at hadron-level. Included also are two PYTHIA-8 parton-level curves: one for ungroomed mass (dashed black line) and another for groomed mass (dash-dotted black line). Data are compared to models via ratio in the bottom panels, where total relative uncertainty is shown in the red band centered at unity. Note: statistical uncertainties are smaller than the size of the markers (cf. Fig. 5.19). [108]

Ref. [60]). This should mitigate the shift to higher mass somewhat, since quark jets have lower mass than gluon jets (Sec 5.3).

In all panels and bins except the low-groomed-mass region of the $30 < p_{\mathrm{T}} < 40\,\mathrm{GeV/c}$ panel, the PYTHIA-6 is consistent with the data. This is sensible since it was tuned to STAR data producing good agreement between detector-level and data (Fig. 5.11), but only a single UE parameter was changed for this tune, so the agreement in terms of jet substructure was not trivially guaranteed. The HERWIG-7 and PYTHIA-8 curves, on the other hand, tend to under- and over-predict the data, respectively, although in the highest-p_{T} selection they exhibit better agreement with the groomed jet mass. We also compare to PYTHIA-8 with hadronization turned off[2] (Appendix C), where we see a large discrepancy with the jet mass prediction due to the significant effect of hadronization, which is reduced when comparing the groomed jet mass to the ungroomed PYTHIA-8 parton-level curves, since grooming reduces the effects of NP radiation and hadronization on jet substructure observables. In addition, the agreement of the groomed data and ungroomed parton-level simulation improves as p_{T} increases, due to the increased reduction of NP effects by SoftDrop grooming at higher jet p_{T} (Fig. 5.22).

Fig. 5.21 shows the fully corrected jet mass measurements for $30 < p_{\mathrm{T}} < 40\,\mathrm{GeV/c}$ and three selections of jet radius, $R = 0.2$, 0.4, and 0.6. The distributions exhibit a similar, although more dramatic dependence: for increasing R, the mean and RMS of the (groomed) jet mass distribution increase significantly. Again, this is as expected, given that larger radius jets are more likely to be gluon jets, since smaller radii are more likely to exclude some of the radiation from the wider showers initiated by gluons.

The dependence of the mean (groomed) jet mass on p_{T} and R is presented in Fig. 5.22 for clarity. We see that the groomed jet mass data is in all cases smaller than the ungroomed jet mass, but that the discrepancy increases slowly as a function of p_{T}, and increases rapidly

---

[2] Note: it is possible for the groomed jet mass distribution to have a larger mean than the ungroomed distribution in the same p_{T} selection because we veto groomed jets with $z_{\mathrm{g}} < z_{\mathrm{cut}}$, and these jets often have small M_{g}.

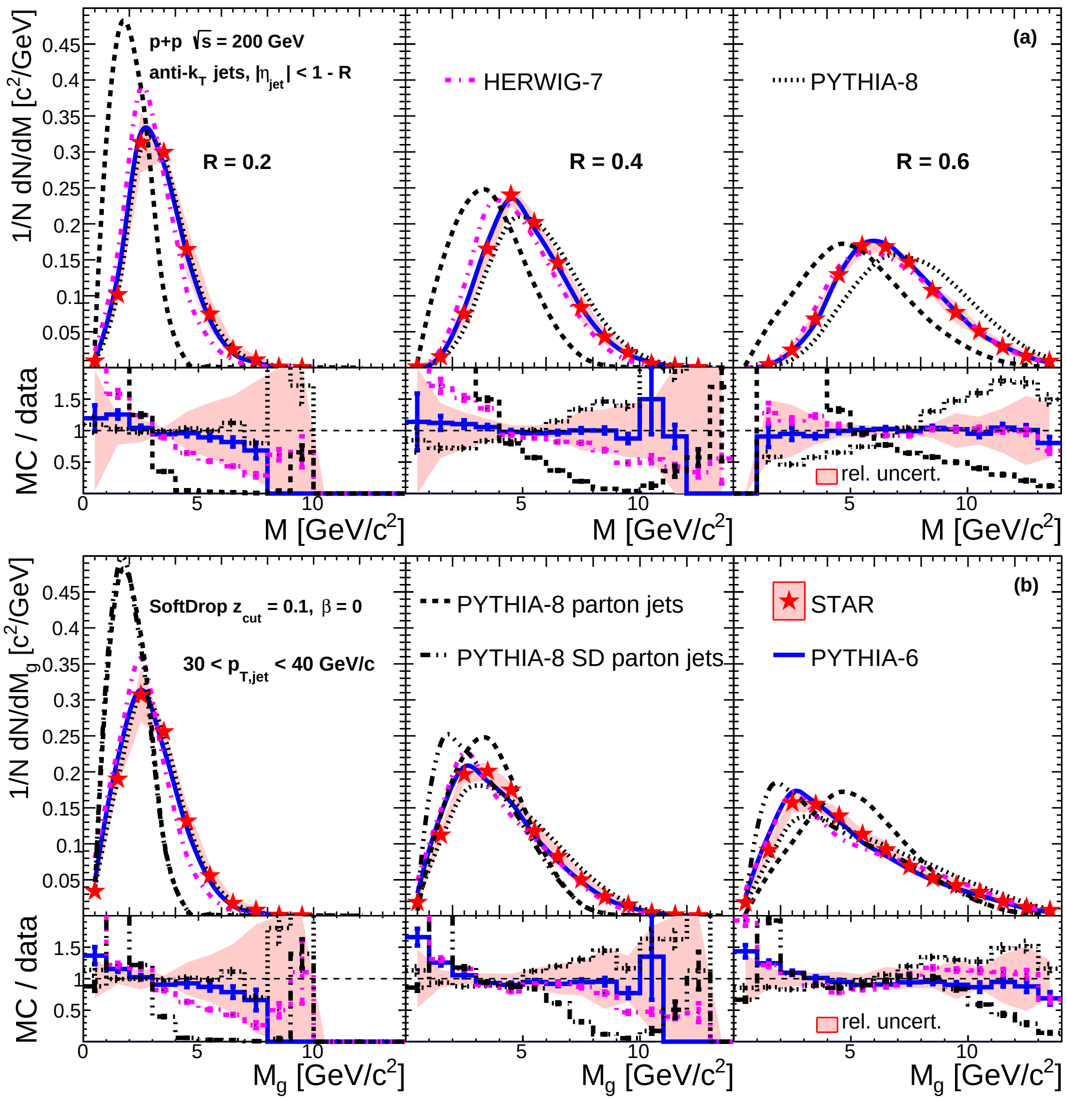


Figure 5.21: See caption for Fig. 5.20. The difference in this figure is that we fix jet p_{T} ($30 < p_{\mathrm{T}} < 40\,\mathrm{GeV/c}$) and vary jet radius: $R = 0.2$, 0.4, and 0.6, from left to right. [108]

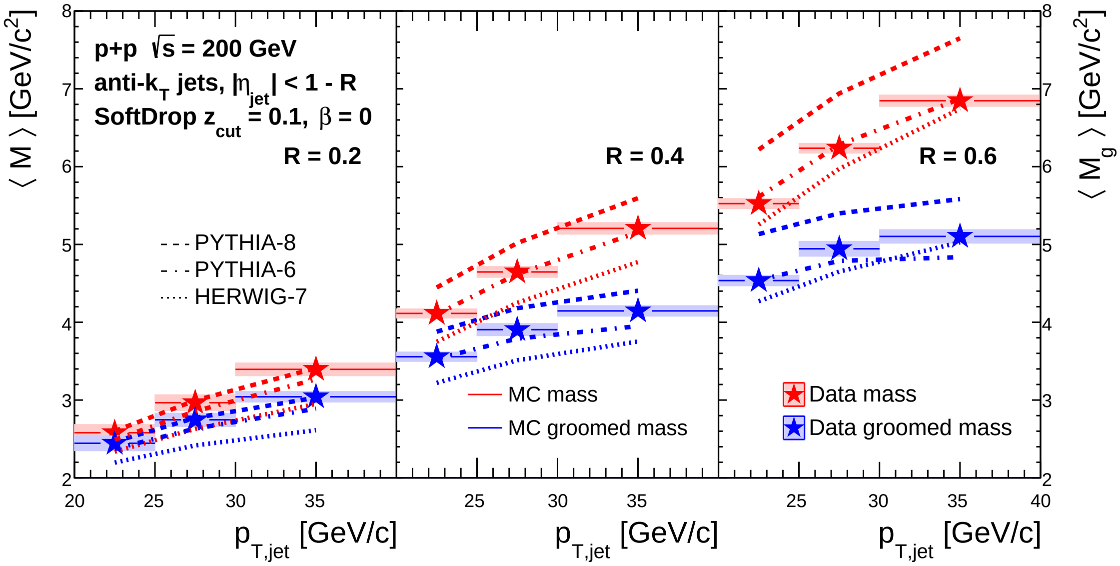


Figure 5.22: The means of the fully corrected (groomed) jet mass distributions in each selection of R (panels) and p_{T} (x-axis), shown as (blue) red filled stars, with an accompanying band corresponding to the systematic uncertainty on the mean. Statistical uncertainties are smaller than the size of the markers. Additionally, the same quantity for the MCs is shown for comparison: PYTHIA-6 (dot-dashed line), HERWIG-7 (dotted line), and PYTHIA-8 (dashed line). [108]

as a function of R. The former dependence of $\langle M \rangle - \langle M_{\mathrm{g}} \rangle$ is expected because a higher-p_{T} jet is more collimated, therefore including more of the wide-angle radiation in the cone to be reduced by the SoftDrop procedure. The latter dependence is also expected given that a jet of a larger radius collects more soft, wide-angle radiation, which is then reduced by SoftDrop, resulting in a larger difference between ungroomed and groomed jets for larger radii. And from another perspective, the larger relative importance of R compared to p_{T} is as expected, given the following approximation for the jet mass (modified from Ref. [121], eq. 1, using eq. 2.6 from Ref. [122]):

$$M^2 \sim p_{\mathrm{T,jet}} \sum_{i \in J} p_{\mathrm{T},i} \Delta R_{iJ}^2, \tag{5.2}$$

where the sum is over the constituents i in the jet, J, and the distance ΔR is between the constituent and the jet axis. So the angular scale of radiation in the jet increases the mass linearly, while the momentum scale increases the jet mass as a square root. Finally, we note that the agreement of the simulation's mean jet mass with that in data is the same as described for Figs. 5.20 and 5.21.

Lastly, we compare the result to a parton-level calculation at next-to-leading-logarithm (NLL) accuracy, by Lee et al. (Fig. 5.23). Since NP effects are largest for small R, p_{T}, and M, the calculation was only provided for $R = 0.4$ and 0.6, for $M > 1\,\mathrm{GeV/c^2}$. For this reason, we only show these two radius selections in the figure, and for the ungroomed jet mass calculation, we extrapolate linearly from the last point provided to the origin, in order to have a consistent normalization between the data and theory. Note that the normalization on the data is different than in Fig. 5.20 due to the different range on the x-axis. Uncertainty on the Lee et al. calculation is given by QCD scale variation by a factor of two around the nominal, while keeping the relation between collinear and soft, and hard and jet scales fixed (see Ref. [23]). We see from the figure that even by $R = 0.6$, the discrepancy between the calculation and data implies that NP effects are still sizable. However, the groomed jet mass in data and the ungroomed jet mass from theory are more comparable at larger radius and p_{T}, since grooming reduces the jet mass more in these regions due to the larger contribution from NP radiation. Finally, we note that the calculation is broadly consistent with PYTHIA-8 at parton-level (with the exception of the groomed jet mass in the lowest R and p_{T} selection), which suggests that the other NP effects which PYTHIA models such as UE do not contribute as much to the discrepancy between calculation and data as does the effect of hadronization. Therefore, these results should be used to obtain an R-dependent NP shape function to apply to theoretical quantities as a data-driven correction for hadronization (e.g. F_κ in Ref. [123]).

5.7 PYTHIA-8 Detroit tune

As mentioned in Sec. 5.1.1, until now PYTHIA-8 has not had a tune appropriate for RHIC energies. A task force, of which I was a member, was formed within STAR to create such a tune, called the "Detroit tune", published in Ref. [124]. A wide variety of data were

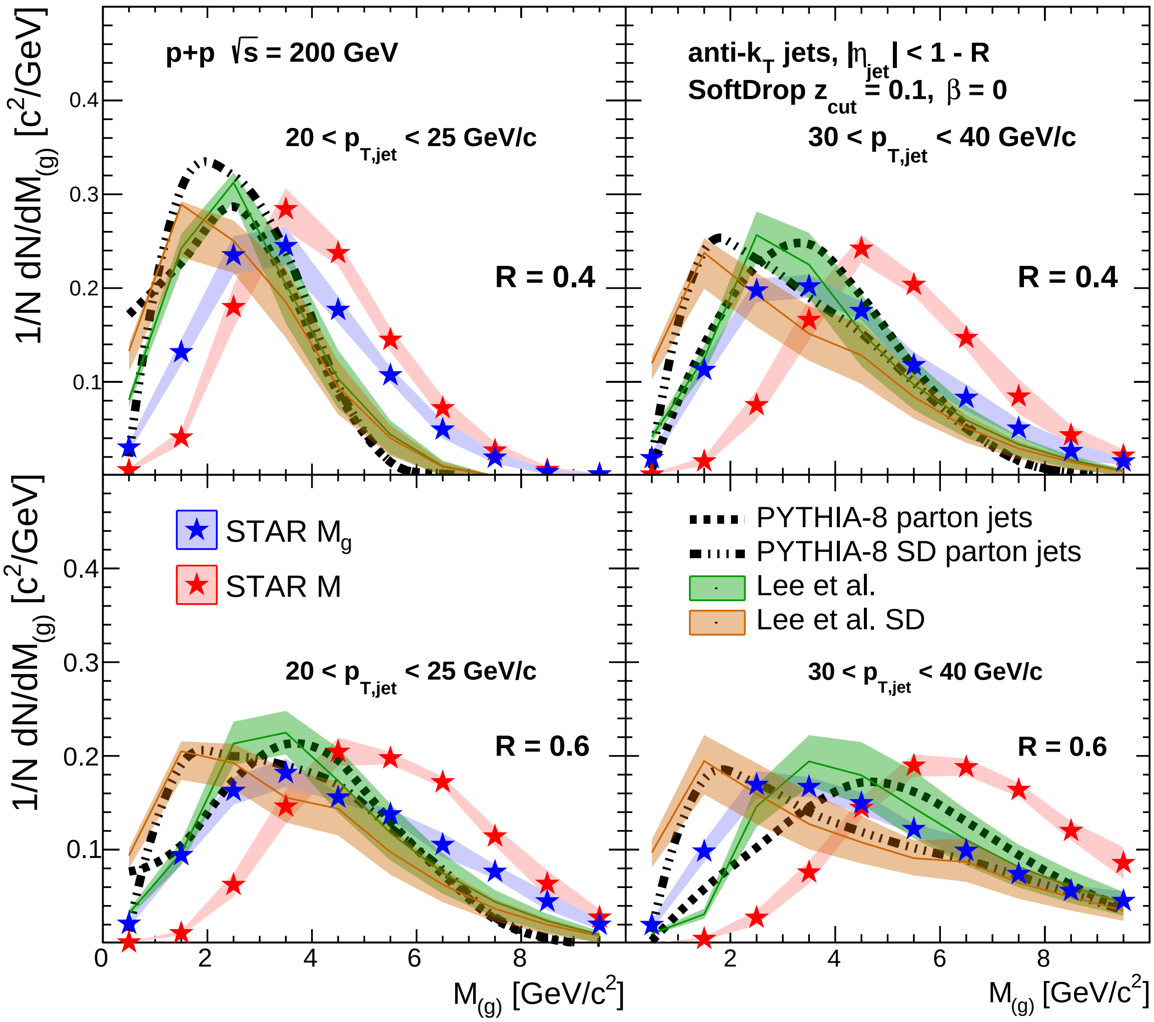
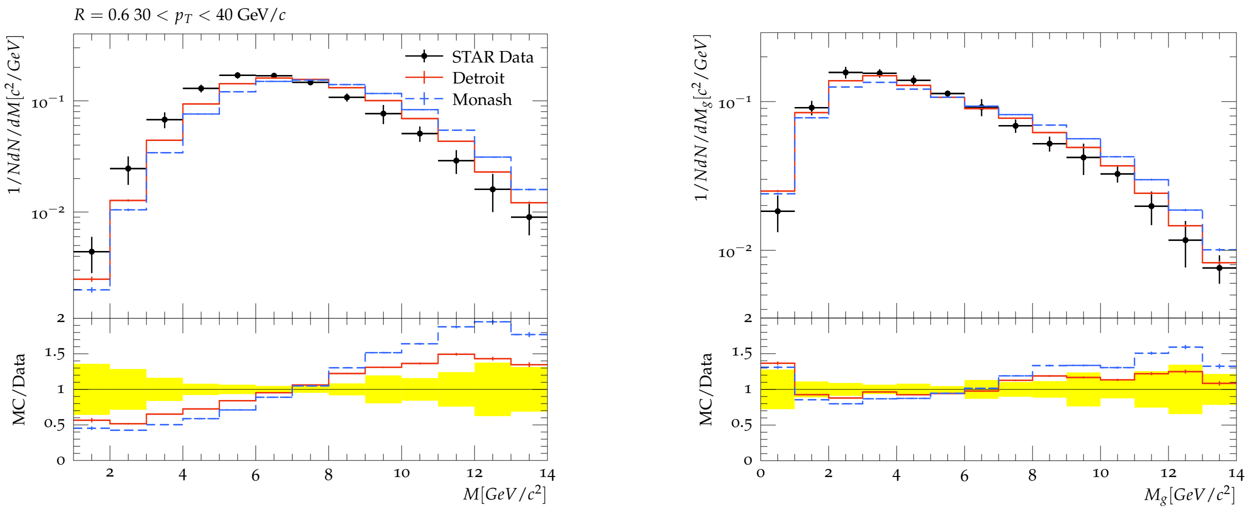


Figure 5.23: Comparison between fully corrected hadron-level STAR data (star markers with shaded bands for systematic uncertainty, and statistical uncertainties smaller than the markers) and a parton-level calculation (described in Sec. 5.6; shaded bands, corresponding to QCD scale variation uncertainty), as well as parton-level PYTHIA-8 (ungroomed: dashed line, groomed: dot-dashed line). Left to right, p_T increases. Top to bottom, R increases. Groomed jet mass is shown in blue (data) and orange (theory), while jet mass is in red (data) and green (theory). [108]



(a)

(b)

Figure 5.24: The positive effect on inclusive jet substructure – in this example, jet mass (a), and groomed jet mass (b) – of tuning underlying event parameters in PYTHIA-8 to RHIC (and CDF) data [124].

used from STAR and PHENIX at $\sqrt{s} = 200\,\text{GeV}$, and CDF up to $\sqrt{s} = 1.96\,\text{TeV}$, with observables such as pion cross sections, Drell-Yan dimuon yields, charged particle multiplicities, and jet substructure. Two of these jet substructure observables were the jet mass and groomed mass[3]. Although only underlying event (multi-parton interaction) parameters were tuned in the model, there was an improvement in the description of these jet substructure observables (Fig. 5.24). In the future, a similar procedure could equally well be applied to HERWIG-7. A complementary approach could also be taken in either PYTHIA or HERWIG, by including the jet mass, which is more difficult to model than, e.g., the z_g, along with other jet substructure observables in a tuning exercise focused more narrowly on the parton shower parameters.

---

[3] The data from the paper is freely accessible on HepData (https://doi.org/10.17182/hepdata.102953) for anyone interested in using it.

CHAPTER 6 JET SUBSTRUCTURE IN pAu COLLISIONS

In this chapter, I present an analysis, in preparation for publication, of the jet substructure in pAu collisions at RHIC. The physical case for these measurements was discussed in Sec. 2.4, but here we highlight the most intriguing prior work directly motivating this analysis. In 2015, PHENIX published a measurement of centrality-dependent jet yield modification in dAu collisions using a similar measurement of event activity to that in STAR (Sec. 3.2.3) – namely, the sum of the charge in a beam-beam counter located at $-3.9 < \eta < -3.1$, where the negative direction is Au-going, as it is in STAR. The result (Fig. 6.1) shows a strong dependence of jet yields on the centrality (similar to a measurement by ATLAS [72]), with a significant suppression in central collisions compared to pp, and a significant enhancement of yields in peripheral collisions. The enhancement is beyond the level predicted by a model including nPDF effects [125], while the suppression was shown to be consistent with an energy-loss calculation. Since then, a reanalysis has been done which will be published as an erratum to the original paper; although an enhancement at low-p_{T} remains [125], there is now no clear sign of suppression in central collisions. Nevertheless, the analysis presented in this chapter was initiated as an attempt to address whether these jet yield modifications in small systems collisions are indeed due to final-state effects, which may be apparent in the measured jet substructure.

As mentioned in Ch. 4, the pAu analysis is intentionally almost identical in structure to the pp analysis, so that results are directly comparable. For this reason, the analysis steps will not be discussed in detail in this chapter (see Appendix E), except for two differences: the first being a physical difference, the second being procedural. First, when selecting event activity classes using the signal in the inner ring of the east (Au-going) side of the BBC (Sec. 3.2.3), we will need to account for the increased activity at mid-rapidity as a function of this large-rapidity activity (Sec. 6.1) which is not related to the physics effects in question. Next, we need to account, via a reweighting procedure, for the additional activity in the simulation which is unphysical, due to the addition of a PYTHIA event to a real data

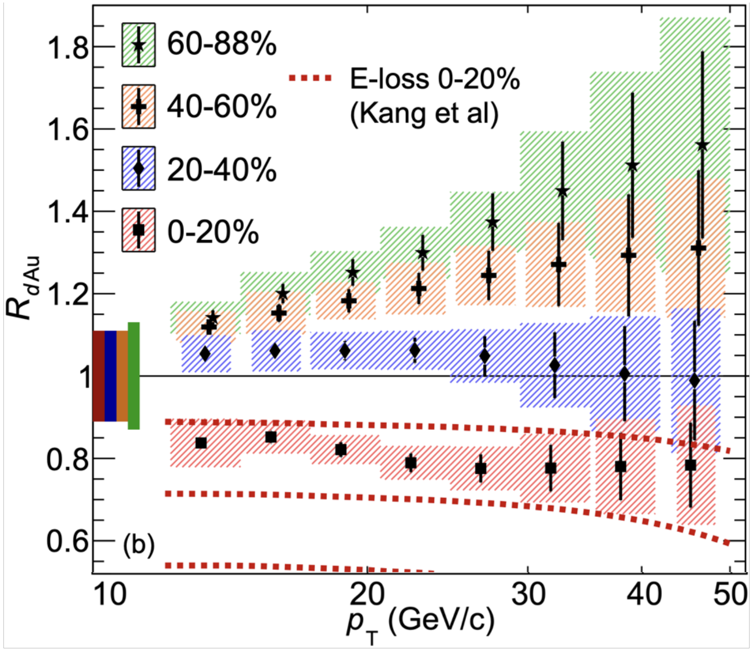


Figure 6.1: PHENIX measurement of the yield modification in dAu compared to pp collisions, for various centrality classes. [73]

event without simulating the additional response of the BBC, resulting in potential event misclassification (Sec. 6.2). We finish by presenting the results (Sec. 6.3) and discussing their implications (Sec. 6.4).

6.1 pAu data

During 2015 data acquisition, the Jet Patch 2 trigger requirement on the east side of the BEMC was much higher than in 2012, for cold-QCD studies. For this reason, we needed to verify that there was not a resulting geometric bias on jets in the raw data. For this test (Fig. 6.2), we can compare the jet pseudorapidity distribution in data and in our embedding sample, where in the latter we remove the simulated jet patch firing, simply requiring $p_T >$ 15 GeV/c. Overlaid on the figure is the physical position of each jet patch. In the range of the fiducial jet acceptance, the middle jet patch, which has a typical ADC threshold, entirely overlaps with the east jet patch, so we would not expect any effect, and indeed, this seems to be the case. However, this is something to consider in the future when $R = 0.2$ jets are analyzed, as their pseudorapidity distribution will extend to ±0.8.

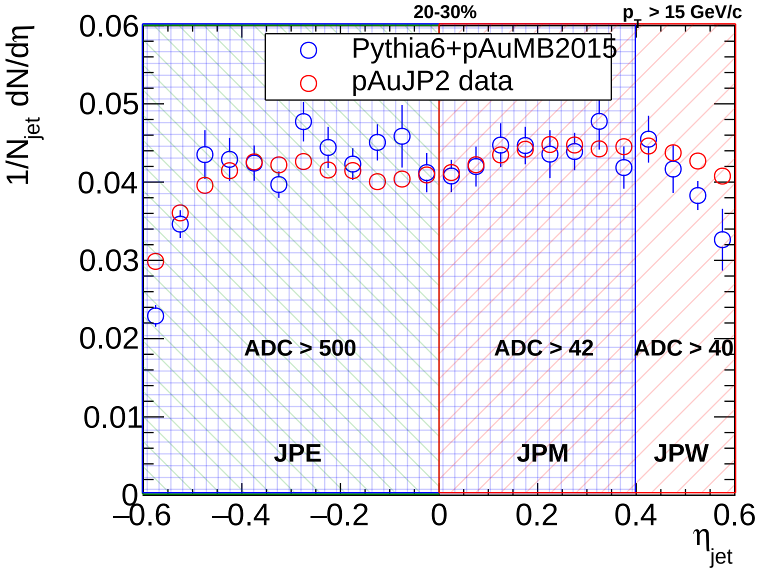


Figure 6.2: A test of the effect on the rapidity distribution of the high JP2 trigger threshold on the east side, comparing data (which has the requirement) and embedding (which does not have the requirement). An example selection of 20–30% EA is shown. The jet patches on the east, middle, and west, have −45°, boxed, and +45° fills, respectively.

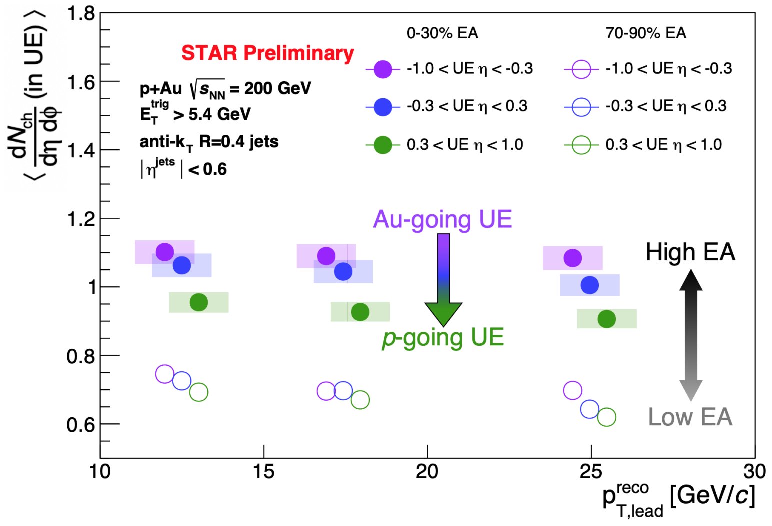


Figure 6.3: The dependence on large-rapidity EA of the average number of charged particles per unit rapidity and azimuth, located away from a hard trigger in φ. High- (low-)EA events are denoted with closed (open) circles. [126]

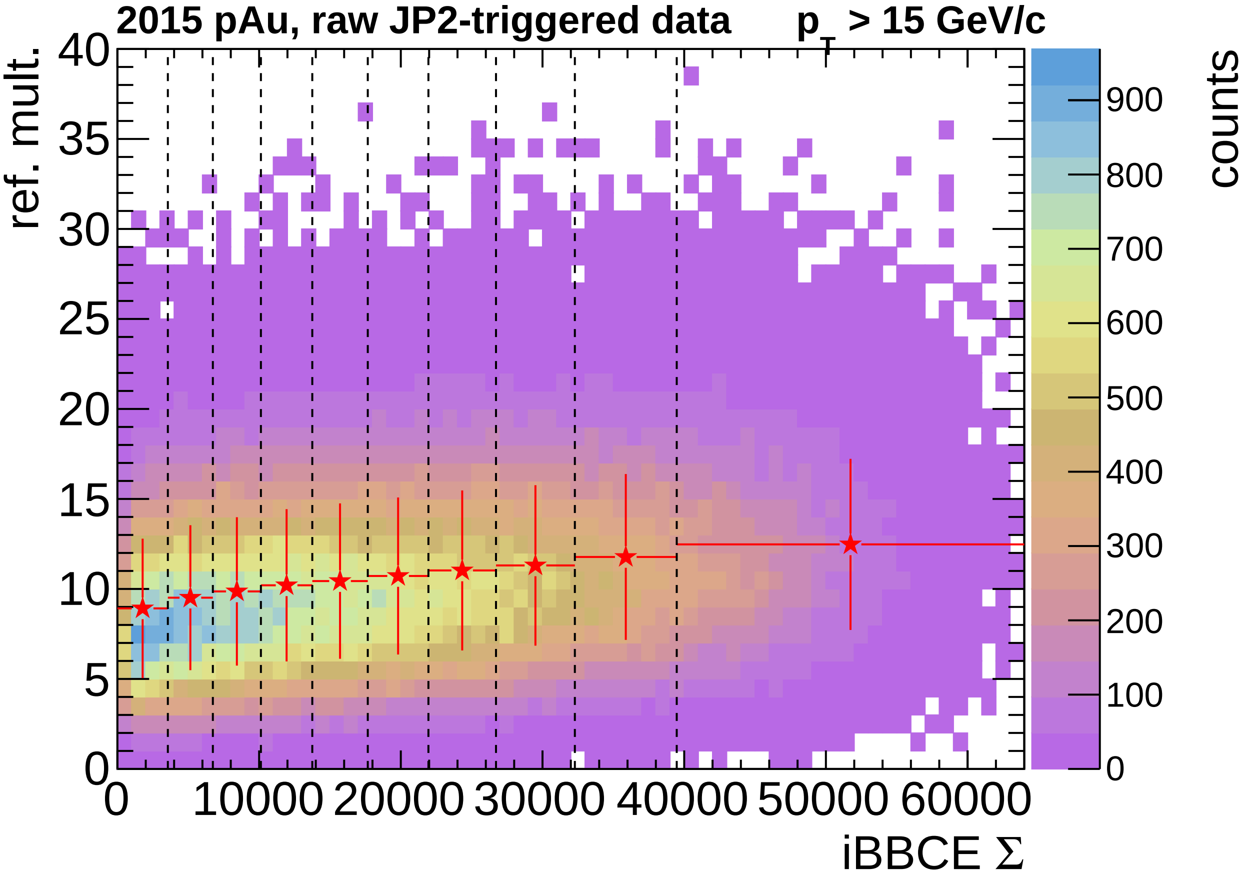


Figure 6.4: The correlation between measures of activity at large-rapidity (iBBCE ADC sum, x-axis), and mid-rapidity (charged multiplicity in $|\eta| < 0.5$, y-axis), for events which passed the JP2 requirement and have at least one jet passing selection criteria. The dashed black lines represent the deciles of the minimum-bias iBBCE distribution, while the filled stars represent the mean ref. mult. for a given decile, with horizontal bars denoting the width of that bin, and vertical bars denoting the standard deviation of the ref. mult. in that range.

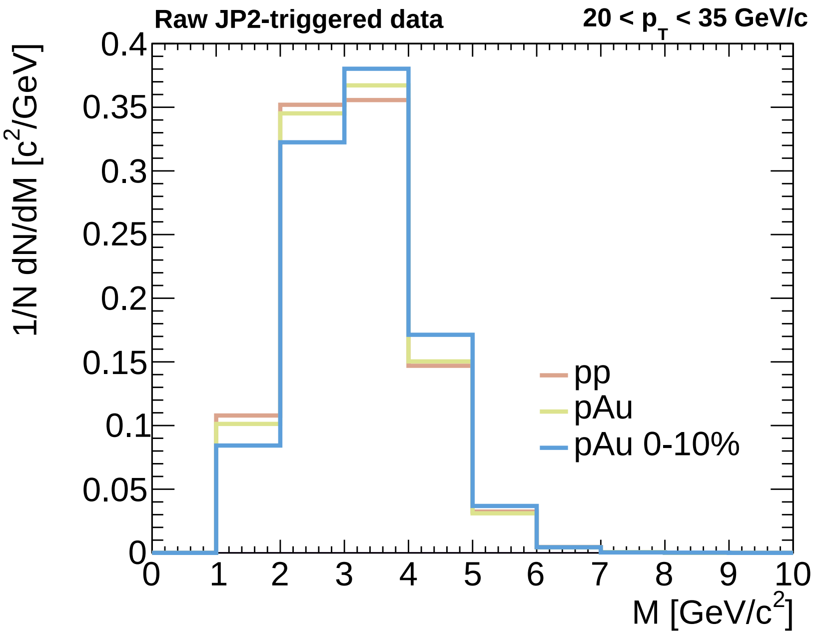


Figure 6.5: A comparison of raw JP2-triggered data between pp and pAu collisions in the red and yellow lines, respectively, as well as a high-EA selection of pAu events (blue line).

There is one more aspect in which the data from pAu and pp collisions differ – namely, the increased soft underlying event in pAu collisions (Fig. 6.3), especially for higher EA selections, since EA at large-rapidity (Sec. 4.1) and UE at mid-rapidity (Sec. 2.2) are positively correlated (Fig. 6.4). Although the difference in underlying event at mid-rapidity between EA classes due to this correlation is small – from low to high EA, on average there is an increase of one charged particle in every three units of $\eta \times \phi$, and a jet of $R = 0.4$ has $A \approx 0.5$ meaning an additional charged particle is added only every sixth jet on average – there will be some increase of the jet mass. That is what we observe (Fig. 6.5) in the raw data, with a more pronounced effect at the highest event activities. Although this effect could be minimized with a pedestal subtraction (and indeed, it was observed that FastJet's area-based subtraction method greatly reduces the activity-dependence in the raw data), it is more effective to leave the correction to the Bayesian unfolding procedure, as mentioned in Sec. 2.2.

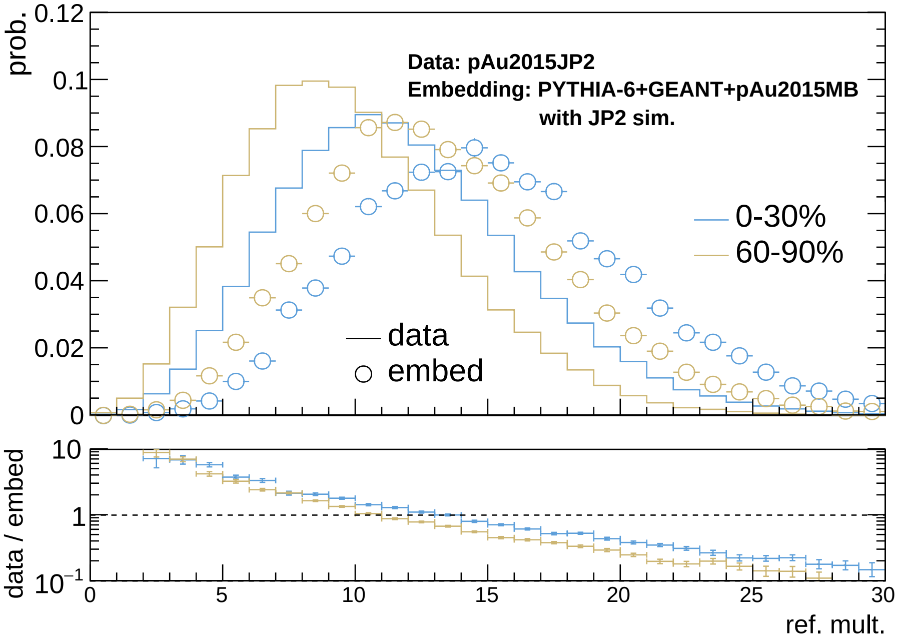


Figure 6.6: Upper: The reference multiplicity distributions in data (lines) and embedding (circles) for two ranges of event activity (0–30% in blue, 60–90% in tan). Lower: The ratios between embedding and data, which are used as event reweighting values in the response matrix.

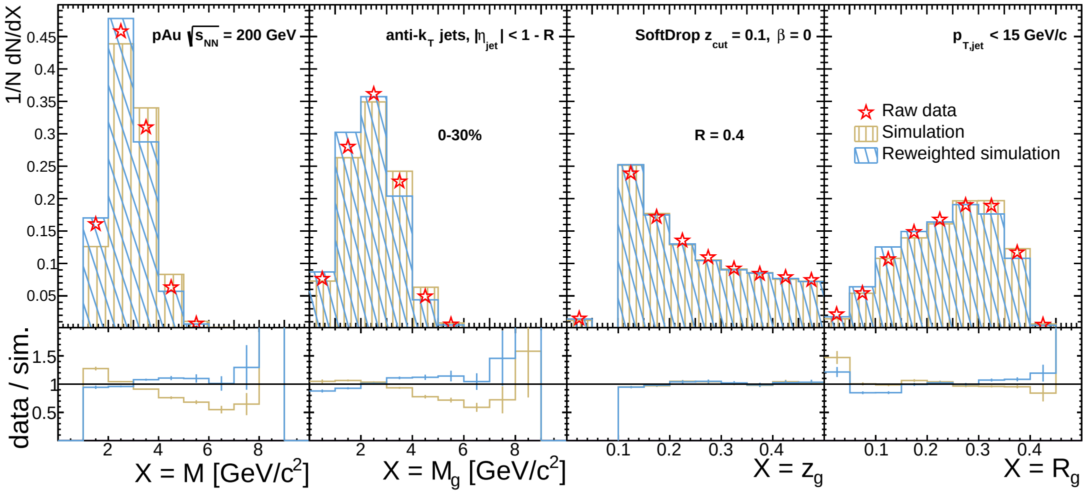


Figure 6.7: Comparison of raw data (open stars) to pAu embedding before (tan) and after (blue) reweighting of the simulation is done, for an example EA range of 0–30%.

6.2 Embedding reweighting

As mentioned in Sec. 4.5, the simulation used for the Bayesian unfolding of pAu data consists of PYTHIA-6 events at particle-level which are processed through a GEANT-3 simulation of the detector, then embedded into minimum-bias background from real pAu data taken in 2015. However, by adding the activity from PYTHIA onto that of a real event, the multiplicities in the simulation are higher (Fig. 6.6). But the event activity measure as determined by the BBC detector (Sec. 3.2.3) does not have a corresponding increase, since the BBC response is not modeled in the simulation, and thus the EA is only given by what exists in the MB pAu data. This leads to, for instance, a systematically larger jet mass in simulation (Fig. 6.7) for a fixed EA selection. In order to improve agreement between the response matrix and the data which we will unfold, we correct for this difference using a reweighting procedure. The reference multiplicity (ref. mult.) is calculated as the number of charged particles passing selection criteria in $|\eta| < 0.5$ in the event[1], and compared between data and embedding for a given EA selection (Fig. 6.6). This ratio is used as an event weight so that, e.g., an event in simulation with a ref. mult. of 5 will have roughly a factor of two more influence on the result, and hence the correlation between mid-rapidity and large-rapidity activity are properly reproduced. After this reweighting, the Bayesian unfolding procedure (Sec. 5.3) was applied as normal, producing results corrected for detector effects and additional pAu underlying event, which are shown in the next section with the same set of systematic uncertainties applied as in the pp analysis of Sec. 5.4.2. See Table 6.1 for example values.

6.3 Results

The fully corrected jet mass results for low event activity[2] are shown in Fig. 6.8. We compare to the previously published pp results [108] (for inclusive EA). The other jet substructure observables for this low event activity selection are shown in Appendix F. Although

---

[1] The result was not seen to be sensitive to the particular definition of ref. mult. that was chosen.

[2] Selections correspond to the vertical lines in Fig. 6.4, starting from 100% (lowest activity) at the origin, and moving by deciles to 0% (highest activity) at the far right.

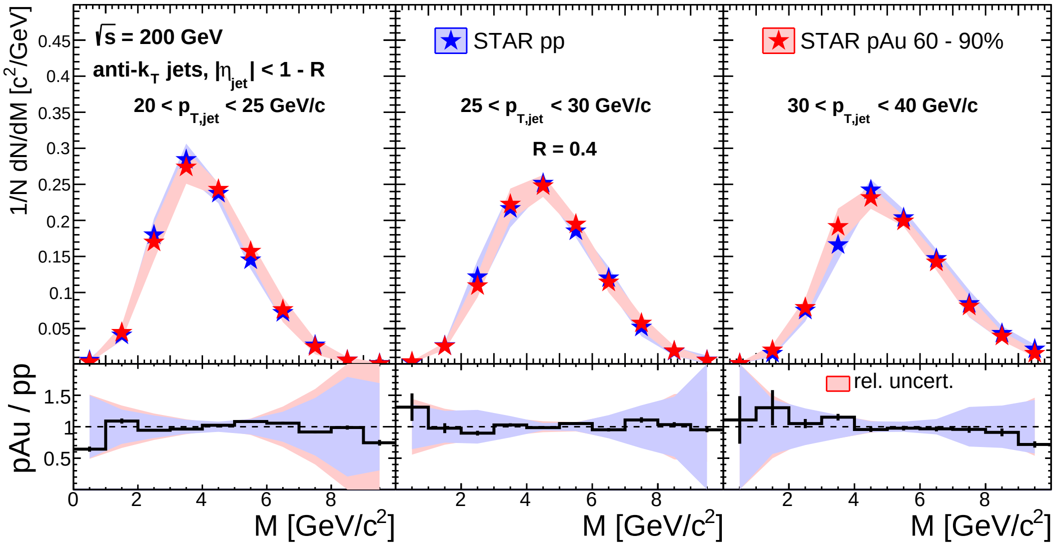


Figure 6.8: Fully corrected inclusive jet mass in pAu collisions for three selections of jet p_{T}, and fixed jet radius $R = 0.4$ and event activity selection (60–90%), shown in red star markers. The pAu data are compared to pp data [108] in blue star markers. Total systematic uncertainty for each is shown as a shaded band, while statistical uncertainties are smaller than the size of the markers. Quantitative comparison between pAu and pp jet mass is done via ratio in the bottom panels, where total relative uncertainties are shown in the shaded bands centered at unity and vertical bars denote statistical uncertainties.

Table 6.1: Systematic uncertainties for an example jet population, from 0–30% events, with $25 < p_{\mathrm{T,jet}} < 30\,\mathrm{GeV/c}$, and three example selections of jet mass corresponding to the low-mass tail, the peak region, and the high-mass tail, from top to bottom. The "Total Systematics" column is obtained by adding the four preceding columns in quadrature.

Source / Range in M	Hadronic Correction	Tower Gain	Tracking Efficiency	Unfolding Procedure	Total Systematics
$(1,2)\,\mathrm{GeV/c^2}$	6.3%	3.1%	18%	16.9%	25.7%
$(4,5)\,\mathrm{GeV/c^2}$	0.68%	∼0%	0.12%	4.5%	4.6%
$(7,8)\,\mathrm{GeV/c^2}$	5.6%	1.9%	9.3%	16.3%	19.7%

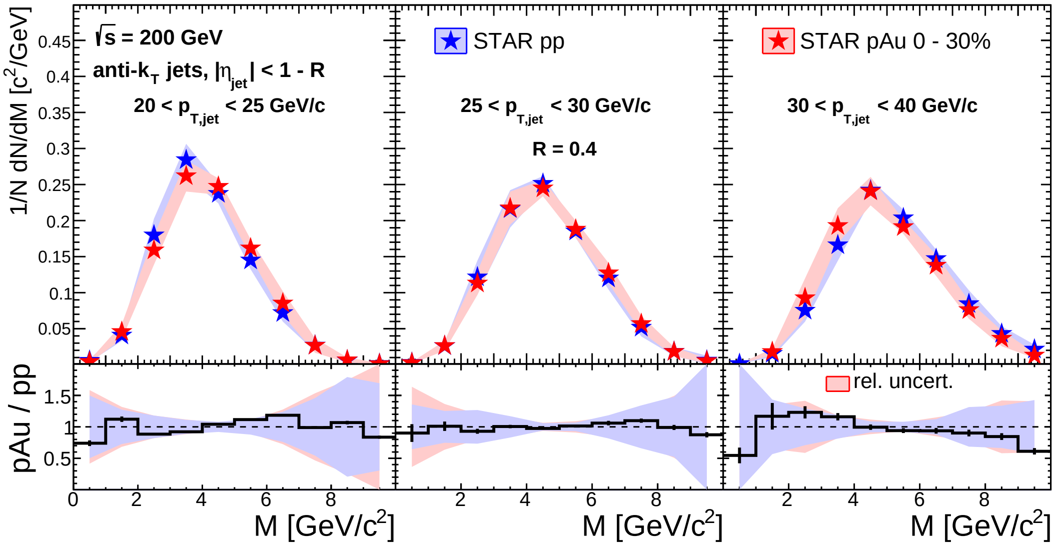


Figure 6.9: The same as Fig. 6.8, but for a high (0–30%) event activity range.

this activity selection is similar to the 60–88% centrality selection of the PHENIX R_{dAu} measurement (Fig. 6.1), we observe no modification to the jet substructure from the pp baseline that could contribute to the measured yield enhancement in this selection.

Similar results for high event activity are shown in Figs. 6.9, 6.10, 6.11, and 6.12 for M, M_{g}, z_{g}, and R_{g}, respectively. We compare both to the previously published pp results [38,108] (for inclusive EA) and, in the case of the z_{g} and R_{g}[3] to a Monte Carlo model, PYTHIA-8 Angantyr [127] (see Appendix C for exact settings). This is in order to test whether any

---

[3] Because the data has had the effect of additional UE removed, while the model has not, we only compare to z_{g} and R_{g} which are less sensitive to background than the jet mass.

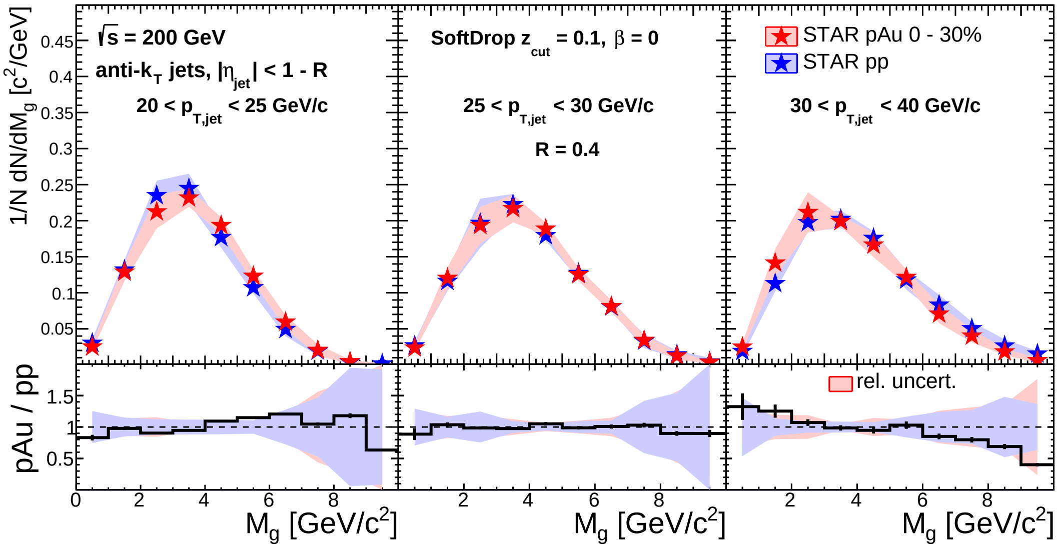


Figure 6.10: The same as Fig. 6.9 but for the groomed jet mass, M_g.

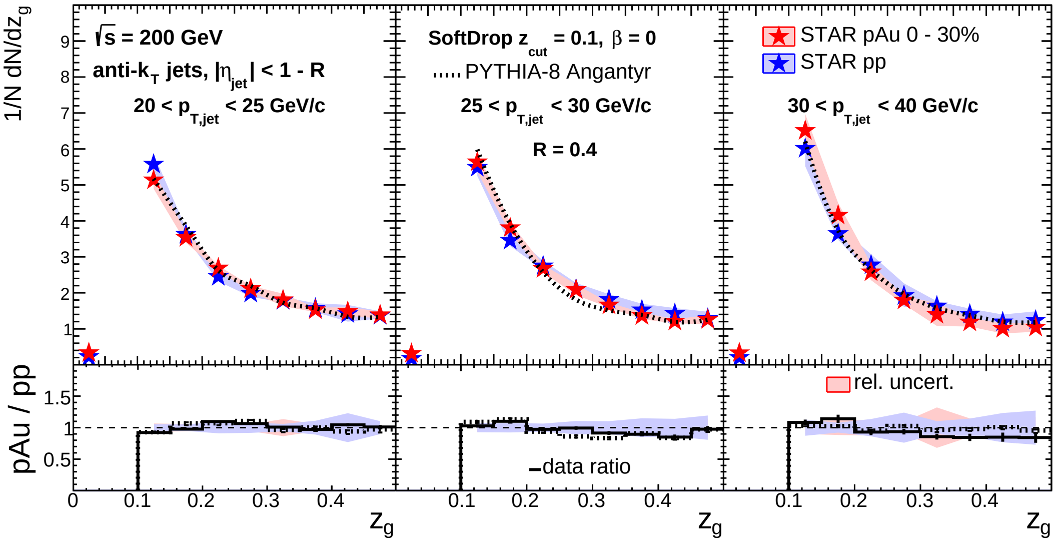


Figure 6.11: The same as Fig. 6.9 but for the subjet shared momentum fraction, z_g. The pp data are from Ref. [38]. In addition, we compare to a PYTHIA-8 Angantyr model prediction (dotted black line).

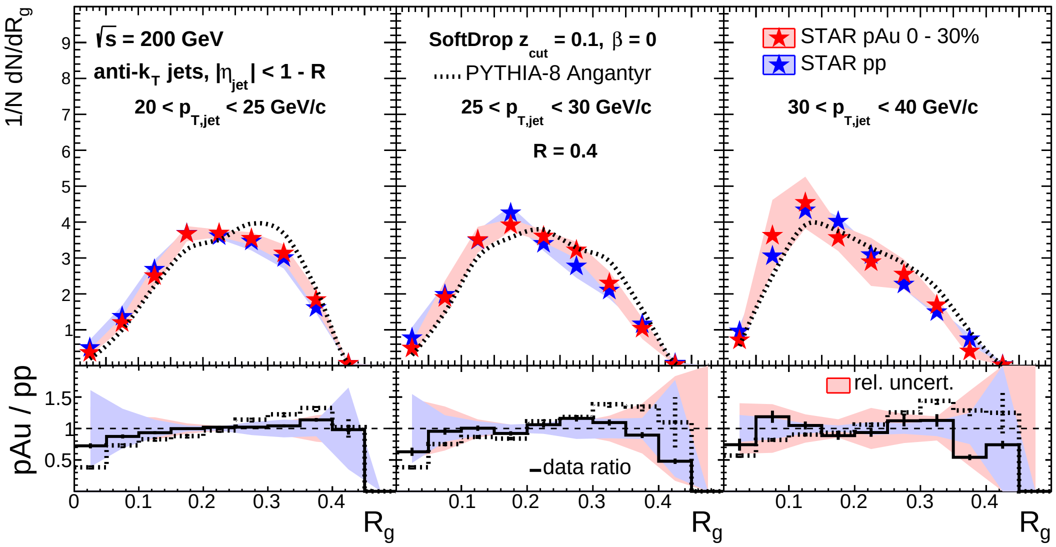


Figure 6.12: The same as Fig. 6.11 but for the groomed jet radius, R_g.

observed final state modification is due to hot nuclear matter effects, since this model of pAu collisions explicitly does not include them (or any collective effects). Angantyr is composed of a Woods-Saxon [128] nucleus geometry, a Glauber-Gribov [129, 130] initial state which includes color fluctuations in both the target and projectile nuclei, and a wounded nucleon model [131] for the final state. Otherwise, it is the same as the pp version of PYTHIA-8, including the same multi-parton interactions, parton shower and color reconnection, and string hadronization. Although it is possible to model centrality in the simulation, what is shown is also activity-inclusive. In the future, this may be updated to include an activity selection, but given the lack of jet substructure dependence on EA that we see (e.g. Fig. 6.13), it is not expected to have a large effect. There is a notable preference of Angantyr for larger angular scale (in the R_g observable), but this is consistent with what was seen for default PYTHIA-8 pp collisions (cf., e.g., Fig. 5.20), and is likely due to the more general disagreement of the default tune with STAR data (Sec. 5.7).

As for the data, the reason to study the substructure observables z_g and R_g, in addition to the groomed jet mass, is to decompose the angular and momentum scales of the parton shower which are combined as in eq. 5.2 ($M^2 \sim p_{\mathrm{T,jet}} \sum_{i \in J} p_{\mathrm{T},i} \Delta R_{iJ}^2$) to produce the (groomed) jet

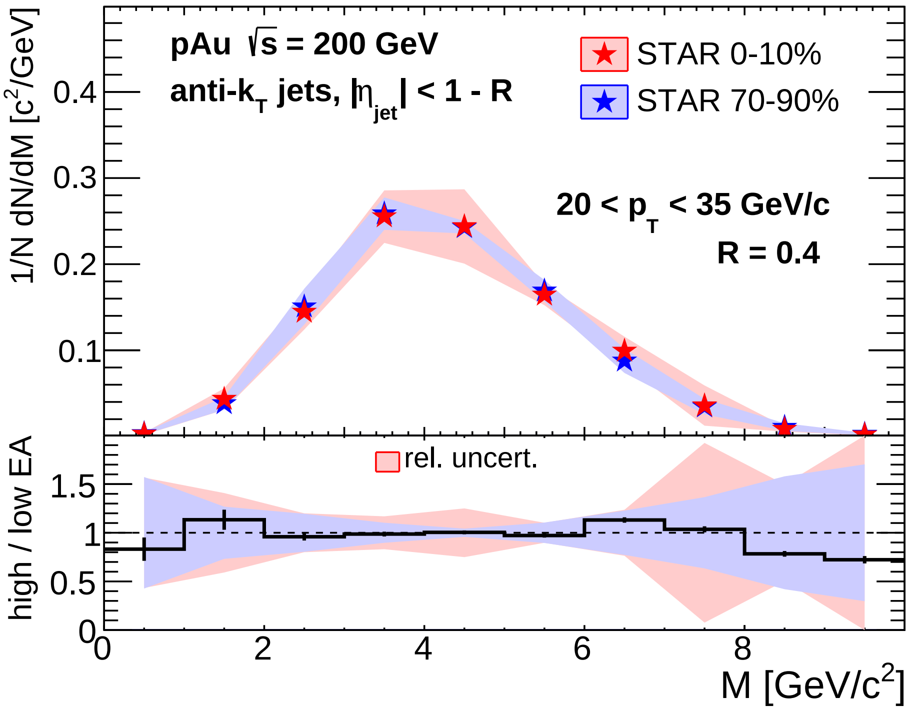


Figure 6.13: Fully corrected inclusive jet mass in pAu collisions for one broad selection of jet p_{T}, and fixed jet radius $R = 0.4$ and narrow activity selections, 0–10% and 70–90%, shown in red and blue star markers, respectively. Total systematic uncertainty for each is shown as a shaded band, while statistical uncertainties are smaller than the size of the markers. Quantitative comparison between the two activity ranges is done via ratio in the bottom panels, where total relative uncertainties are shown in the shaded bands centered at unity and vertical bars denote statistical uncertainties.

mass. Competing effects may result in an overall unmodified groomed jet mass. However, we observe consistency across all four jet substructure observables between pp and pAu collisions, even in the highest EA range, indicating that any potential hot or cold nuclear matter effects on the jets are small for both the angular and momentum scales. Although it would be possible for the SoftDrop observables to show agreement while only indicating that the hard core of the jets are unmodified, the fact that the jet mass is consistent indicates that even final state effects such as angular broadening are not significant.

Besides comparing to pp, we can also compare the jet substructure between event activity classes. Since we would expect any hot nuclear matter final state effects to be activity-dependent, this is a conceptually clean way of addressing this hypothesis. These results are

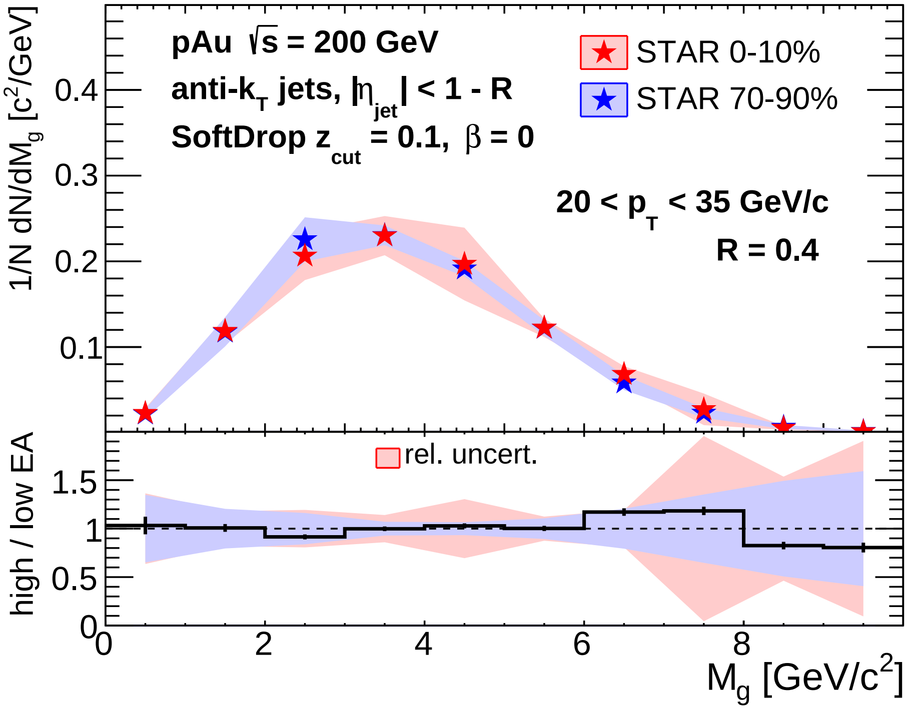


Figure 6.14: The same as Fig. 6.13 but for the groomed jet mass, $M_{\rm g}$.

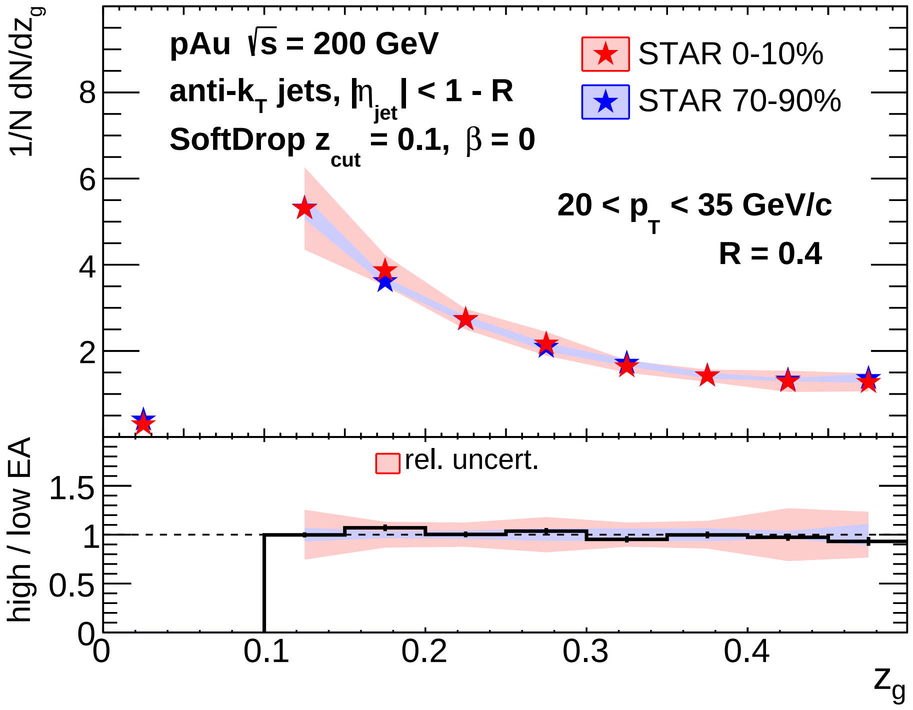


Figure 6.15: The same as Fig. 6.13 but for the subjet shared momentum fraction, $z_{\rm g}$. Here, the $z_{\rm g} = 0$ bin represents the probability of jets failing the SoftDrop criterion (Sec. 4.4).

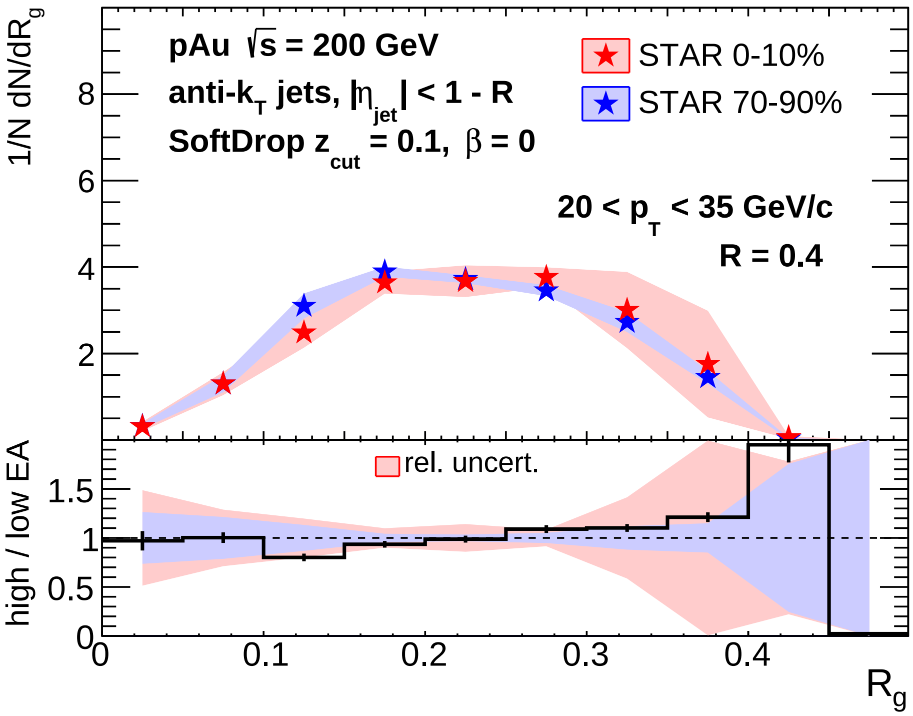


Figure 6.16: The same as Fig. 6.13 but for the groomed jet radius, R_g.

shown in Figs. 6.13, 6.14, 6.15, and 6.16. To separate the activity ranges, and therefore the collision geometry classes, as much as possible, we choose narrow EA selections 0–10% and 70–90% for this comparison. But, as mentioned in Sec. 4.4, in order to increase statistics for the comparison, we select jets in a single broad p_T range. Aside from a slight hint of an angular broadening in Fig. 6.16, the data are generally consistent between EA ranges. One point of future work will be to address the cause of the large systematic uncertainty on a single bin of the jet (groomed) mass high-EA result (4–5 GeV/c^2) and z_g (0.1–0.15) and on two bins of the R_g high-EA results (0.3–0.4), which are caused by fluctuations in the tracking efficiency uncertainty systematic that are most likely the result of unphysical fluctuations in the simulation reweighting, or just due to low statistics in general.

The next step for this analysis is to consider more carefully the degree of cancellation between correlated systematic uncertainties both between pp and pAu collisions, and between different EA ranges in pAu data. It should be the case that all detector uncertainty is canceled in comparisons between different EA selections of the same dataset, since none of them are explicitly activity-dependent. In comparisons across the two datasets, we would expect most but not all of this uncertainty to cancel – for instance, the tower gain calibration uncertainty is slightly lower for the 2015 data (3.2% [132]) than the 2012 data (3.8% [100]). As for the uncertainties related to the Bayesian unfolding, we expect a large degree of cancellation, since the shapes of the distributions are fairly similar between datasets and between activity classes, so shape variation should be independent of these classifications. For now, the systematic uncertainties on the ratios are presented separately. But even considering only statistical uncertainties, the deviation of results is within roughly 10% for almost all bins, and is similar to the level of deviation from 60–90% EA (Fig. 6.8 and Appendix F), where we would expect negligible jet substructure modification with respect to the vacuum baseline.

6.4 Discussion

Although we have shown fairly definitively in this chapter that the activity-dependent yield modification that has been seen at PHENIX and ATLAS (Sec. 2.4) is not due to, or related to, jet quenching, the question remains: What is determining these effects? Although an answer lies beyond the measurements presented in this thesis, there are some current hypotheses and ongoing work worth discussing in more detail. For ATLAS's part, they show in their paper [72] that in pPb collisions the jet modification on the Pb-going side depends on the initiating parton's Bjorken-x, denoted x_p, through its relation to the jet energy. The hypothesis is then that the initiating parton's kinematics explain this effect, without the need to resort to final state effects. This question has been studied at STAR in a paper currently in preparation. First, as confirmation of the hypothesis that the hard activity of the collision is linked to the soft dynamics at early times, it is observed that the leading jet p_{T} and the EA measured in the BBC (across a large rapidity gap) are anti-correlated (Fig. 6.17). This bias could feasibly result in a misclassification of the event activity and a resulting apparent yield modification, for instance in semi-inclusive jet spectra, seen in Ref. [101]. To test for jet quenching, it is necessary to remove the effect of the parton kinematics, for which a

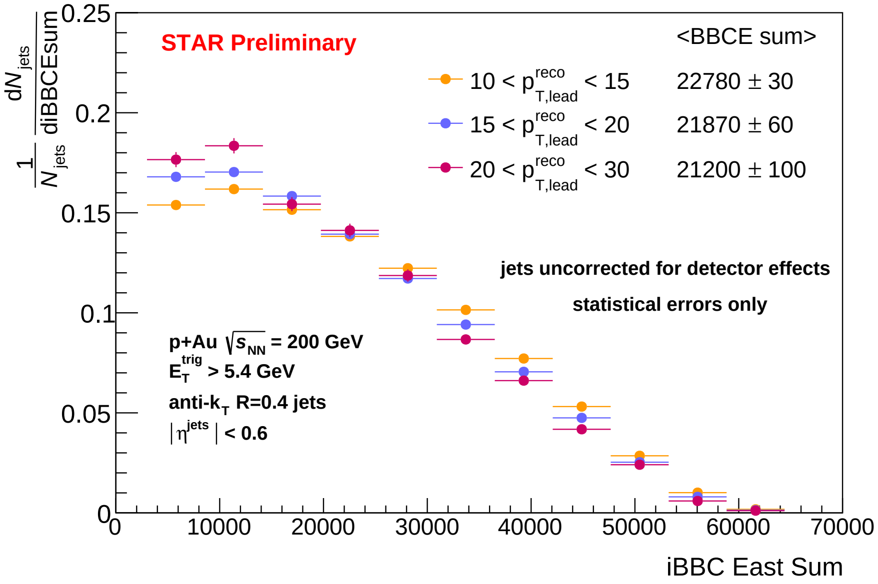


Figure 6.17: The distribution of inner BBC charge on the Au-going side, for three selections of leading jet p_{T}, from STAR pAu collisions, requiring that a high-energy tower fires in the BEMC. [126]

dijet measurement is studied (Fig. 6.18), which is normalized per-dijet, circumventing the bias on the per-trigger normalized spectra. This measurement of the dijet imbalance, A_{J}, has been used previously as a definitive measure of energy loss in the quark-gluon plasma due to path-length-dependent jet quenching [133] in a large system (AuAu) compared to an unmodified reference (pp). However, in this case, comparing the A_{J} between high- and low-activity classes of pAu collisions shows no modification, which is confirmation of the results presented in this chapter which showed no observable final state effects, either due to jet quenching or otherwise.

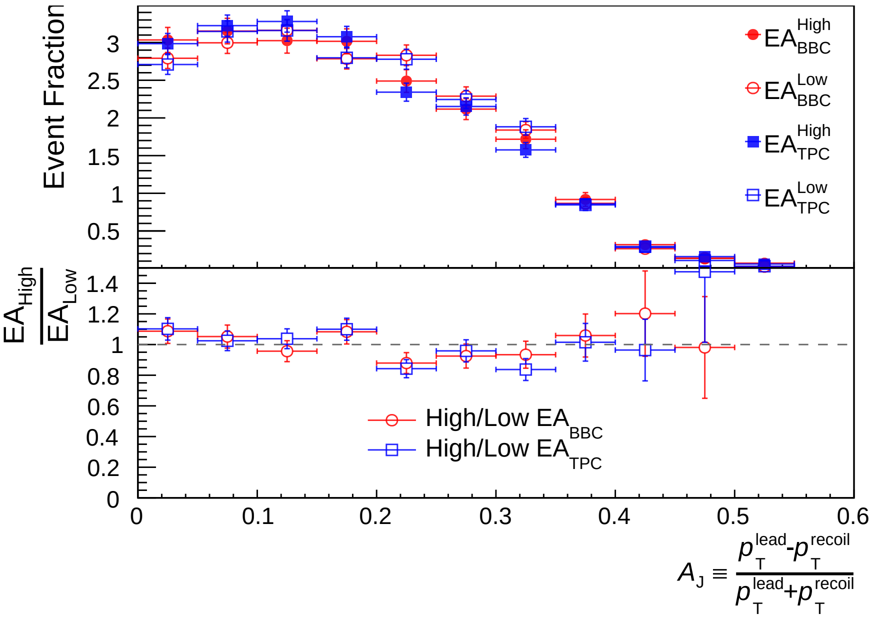


Figure 6.18: Upper: the dijet imbalance, A_J, shown for high (full markers) and low (open markers) event activity ranges, as measured either at mid-rapidity (blue squares), or large rapidity (red circles). Lower: the ratio of high and low EA distributions. [101]

CHAPTER 7 SUMMARY

In this thesis, I presented the first fully corrected measurements at RHIC of the inclusive jet mass in pp collisions, published in 2021 [108], and jet substructure in pAu collisions. Data were presented differentially in jet transverse momentum, and in the case of the pp analysis, the jet radius. This showed the dependence of the jet mass on the momentum and angular scales inherent in the radiation pattern of the jet – namely, that jet mass increases gradually with increasing jet transverse momentum, and rapidly with increasing jet radius. It was found in this analysis that the data were well-described by a STAR-tuned Monte Carlo model, PYTHIA-6, but not by LHC-tuned models HERWIG-7 and PYTHIA-8, which predict smaller and larger jet masses on average, respectively. Comparison to a theoretical calculation at the parton-level demonstrated the significant contribution of hadronization and other non-perturbative effects to the jet mass at RHIC energies, with large discrepancies between the hadron-level data and parton-level calculation in all jet kinematics. Using these measurements to derive an R-dependent NP shape function to apply to theoretical quantities would allow hadronization to be accounted for in a data-driven way. The measurements were also used as part of a Monte Carlo tuning procedure which resulted in the first RHIC tune of PYTHIA-8: the Detroit tune.

In the pAu analysis, data were selected by event activity classes to study potential modification to the jet substructure due to cold or hot nuclear matter final state effects. We note a high level of agreement between pAu and pp collisions for all jet substructure observables measured, even in the highest event activity selection of 0–10%, as measured in the large-rapidity Beam-Beam Counter. Comparison to a model, Angantyr, showed a significant disagreement with the groomed jet radius, but one consistent with the disagreement shown by the pp version of the model with the pp data presented in Ch. 5 (with larger angle being related to larger mass). For a clear view of the effect of final state interactions which increases with increasing event activity, we also compare jet substructure in high and low event activity classes of pAu collisions. Again, no effects on the jet substructure are observed, suggesting instead a bias caused by an anticorrelation between early time dynamics and the hard process (discussed in detail in Sec. 6.4). The measurements presented in this thesis are an integral component of this emerging picture, in addition to other RHIC measurements [101,126,134] and complementary substructure measurements at the LHC for a different range of initiating parton kinematics [55,135–137], of minimal final state effects on jets in small systems collisions.

The pAu analysis is reaching maturity, but there are a few more areas for improvement before publication. Jet radius dependence, similarly to the pp analysis, will be added. However, we will need to check whether $R = 0.2$ jets will be possible to analyze due to the atypical jet patch requirement on the gold-going side of the detector during data acquisition in 2015. Otherwise, the analysis should follow exactly the same steps as what was presented here. If there is a hot nuclear matter effect on the substructure of jets in pAu collisions, it may be made most evident by comparing modification as a function of jet radius [138]. Systematic uncertainties will also be studied in more detail to determine the cause of some of the large uncertainties in the highest event activity selection (0–10%), and whether they are purely statistical in nature. Before publication, we will also determine the appropriate degree of cancellation of systematic uncertainties in the high-to-low event activity and pAu-to-pp ratios, which were shown here with no assumption made on this point.

7.1 Outlook

Although the field of jet physics has existed for roughly half a century, novel developments are occurring every year. Jet substructure has become a precision topic driven by the high luminosity and broad jet cross section at the LHC, and theoretical improvements like the sophisticated handling of non-global logarithms [139], but there is more to study, especially at lower energies. The jet substructure measurements presented in this thesis, by motivating future development of non-perturbative corrections and further Monte Carlo tuning exercises, will lead to similar advances at RHIC kinematics. But they are just one piece of a larger puzzle. At STAR, measurements are being performed which can give insight into the parton shower topology (for instance, using SoftDrop iteratively), and non-perturbative effects such as hadronization (for instance, studying the jet charge in pp collisions). New techniques are also being applied, such as Omnifold [140] which unfolds many observables simultaneously, allowing for a more complete picture of the jet substructure. Hopefully, the jet mass measurement in pp collisions will also partly motivate further Monte Carlo tuning exercises for a better agreement of models with RHIC data. Finally, when the pAu jet substructure analysis is completed, the corresponding jet mass measurement in AuAu collisions at STAR should be carried out on the 2014 dataset that was prepared by Nick Elsey, using the pp and pAu jet substructure observables as vacuum and cold nuclear matter baselines, respectively, for potential modification due to the quark gluon plasma.

In small systems, jets have played a crucial role in validating claims of final state effects, and a forthcoming paper including a suite of jet observables such as the dijet imbalance, dijet acoplanarity, and mid- and large-rapidity event activity as a function of jet kinematics, will help to narrow more conclusively the list of possible causes of previously observed centrality-dependent modification of jet yields. So far, it has been shown that there is an early-time kinematic effect due to the Bjorken-x of the initiating parton in the proton which leads to a bias in the jet spectrum after classification of events into activity classes, rather than any final state effects which modify the jets.

The longer-term future looks exciting as well, with the sPHENIX experiment at RHIC coming online in the first quarter of 2023 with impressive jet detection capability expected (Fig. 7.1). The improvement in statistics will allow for a broader kinematic reach of jet measurements, which also increases complementarity with the LHC. It will be possible, for instance, to compare populations of jets with the same momentum but much different quark and gluon fractions. And subsequently, RHIC will be transformed into the Electron-Ion Collider (EIC), with operations beginning in the early 2030s. For jet physics, this offers a cleaner initial state (Fig. 7.2), which will allow for a focus on final-state interactions of parton showers with nuclear media.

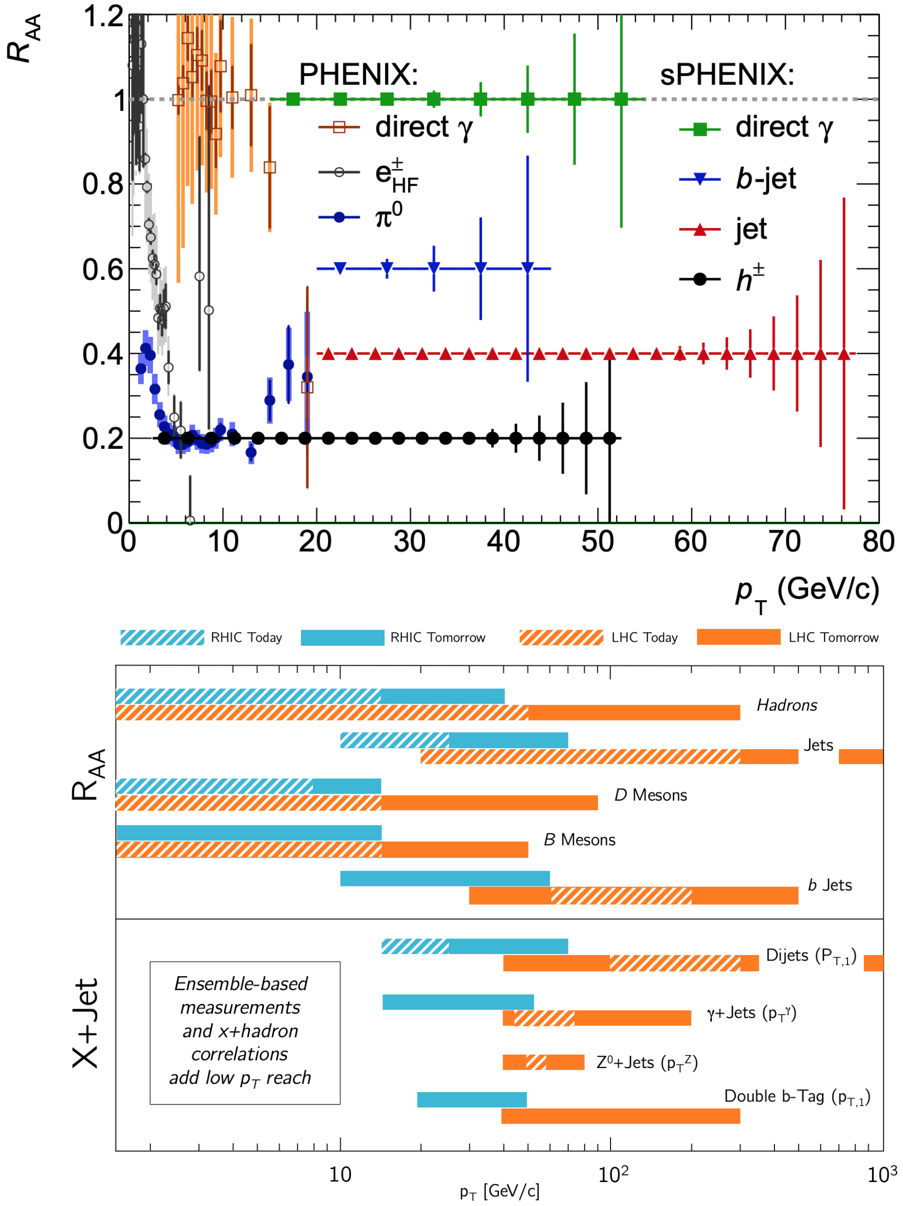


Figure 7.1: Upper: sPHENIX statistics projections with a comparison to PHENIX data. Lower: expected overlap with the LHC. [141].

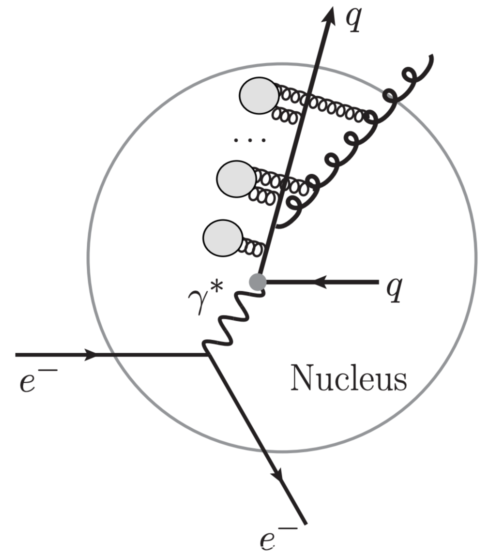


Figure 7.2: A hard parton traversing the nucleus in an electron-ion collision, and resultant medium-induced gluon bremsstrahlung [142].

APPENDIX A COLLIDER KINEMATIC VARIABLES AND CONVENTIONS

Position The Cartesian coordinate system for position within the STAR experiment (see Fig. A.1) is aligned as follows; $+z$ points along the beamline roughly west, $+y$ points perpendicular to the beamline and toward to the sky, and $+x$ points along the ground, away from the center of the RHIC ring. $(x, y, z) = (0, 0, 0)$ is set at the center of the STAR barrel. For the most part, the only important direction of these three for our purposes is the z direction, because the cylindrical symmetry of the detector allows for the use of cylindrical coordinates. Here, the $x - y$ plane can be reformulated with a radius, r, and angle, φ: $r = \sqrt{x^2 + y^2}$ and $\varphi = \text{atan2}(y, x)$[1], while z is unchanged.

Momentum Because we are considering highly relativistic collisions, we replace classical mechanical three-momentum (space) vectors with the equivalent four-momentum (space-time) vectors, with $(E/c, p_x, p_y, p_z) = (E/c, p_\mathrm{T}, p_z)$, where p_T, the transverse momentum, is the cylindrical radius in momentum space. For bodies in motion relative to some frame, Einstein's energy equation for the rest energy, $E = mc^2$, is extended to include the kinetic contribution to the total energy:

$$E^2 = \mathbf{p}^2c^2 + m^2c^4, \tag{A.1}$$

where $\mathbf{p}$ is the three-momentum. However, in quantum field theory, intermediate particles which do not appear in the final state can violate this relation. The amount to which it is violated is called the virtuality or off-shellness[2] of the particle. A higher virtuality of the initiating parton of a jet will produce a jet with a higher invariant mass, $M = |\sum_{i\in J} p_i^2|$, where a script p refers to the four-momentum.

---

[1] atan2(y,x) is basically arctan(y/x), but well-defined for all x and y.

[2] The "shell" here refers to the hyperboloid formed by solutions to eq. A.1.

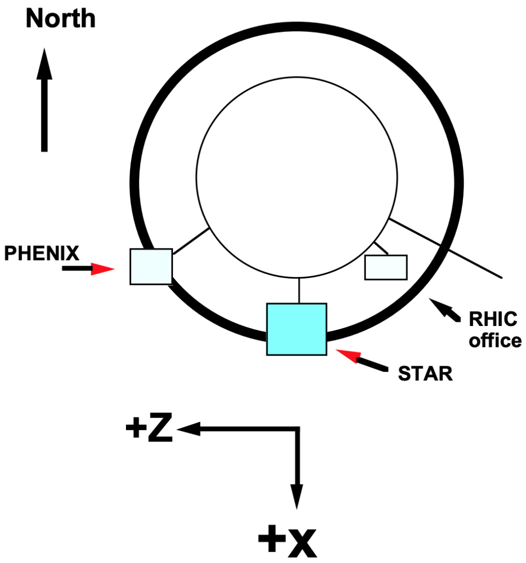


Figure A.1: A schematic of the RHIC collider as viewed from above. [143]

Energy An invariant measure of the energy of a collision of two particles 1 and 2 is $s = (p_1 + p_2)^2$. In the center-of-mass (CM) frame, where $\mathbf{p}_1 + \mathbf{p}_2 = 0$, this simplifies to $\sqrt{s} = E_1 + E_2$. This is the $\sqrt{s}$ referred to in Sec. 3.1, for instance. We can also expand the energies using eq. A.1, for a useful formula for calculating $\sqrt{s}$ if the energy of a nucleon in each beam is known, and assuming $m \ll E/c^2$: $E \approx 4E_1E_2$. When nuclei larger than protons are colliding, we use the term $\sqrt{s_{\mathrm{NN}}}$ to refer to the CM energy per nucleon (N) pair. Energies at RHIC and the LHC are typically given in GeV or TeV, which are giga- (10^9) and tera- (10^{12}) electron-volts, respectively. An electron-volt is the amount of kinetic energy gained by an electron accelerated from rest by a 1 V potential, and its relation to everyday quantities is $1\,\mathrm{eV} \approx 1.6 \times 10^{-19}$ J. Using a convention of particle physics called natural units, we set $c = 1 = \hbar$, which allows for simple conversion between energy and distance using the relation $1 = \hbar c \approx 197\,\mathrm{MeV} \cdot \mathrm{fm}$ so $1\,\mathrm{fm} = 10^{-15}\,\mathrm{m} \approx 5\,\mathrm{GeV}^{-1}$. Similarly, we have a relation between distance and time: $1\,\mathrm{fm} \approx 3 \times 10^{-24}\,\mathrm{s}$. However, throughout this thesis, natural units are avoided for the purpose of clarity.

Rapidity So far, the angle θ in the $y-z$ plane has not been discussed. This is because it is actually replaced in collider physics by a quantity called rapidity, typically denoted y, but to avoid confusion, we will call it $\tilde{y}$ in this Appendix. The definition is

$$\tilde{y} = \frac{1}{2}\ln\frac{E+p_zc}{E-p_zc}. \tag{A.2}$$

The benefit of rapidity, from an experimentalist's point of view, is that relative positions of particles in rapidity space are unaffected by the frame in which they are measured. So as an example, when FastJet calculates the distance between two measured pseudojets, d_{ij}, using $\Delta R_{ij}^2 = (\Delta\tilde{y})^2 + (\Delta\phi)^2$ (eq. 2.2), the fact that the measurement was made in the frame of the observer rather than, say, the CM (zero momentum) frame of the collision, is immaterial. These can be different even for symmetric collisions, due to e.g. varying longitudinal momentum fractions (Sec. 1.3) carried by the partons which interact. In asymmetric collisions, this is even more important, due to the large Lorentz boost in the longitudinal direction between the observation frame and the CM frame of the collision (e.g. 5.02 TeV pPb collisions at the LHC, which have $\Delta\tilde{y} = 0.465$ (eq. 2.7 of Ref. [78]) in the proton-going direction[3] compared to the observation frame). This is caused by the different energy-per-nucleon of each beam (see Sec. 3.1 for an explanation of this difference in energy).

Lastly, it should be noted that even rapidity is often supplanted by a more accessible quantity called "pseudorapidity", η, defined as

$$\eta = \frac{1}{2}\ln\frac{|\mathbf{p}|+p_z}{|\mathbf{p}|-p_z}. \tag{A.3}$$

This definition is useful because the energy of particles is not always known – for instance, in the TPC (Ref. 3.2.1), we measure momentum rather than energy. For massless particles, the two definitions are equivalent. For massive particles in the high-energy limit, using eq. A.1 and letting $E \gg mc^2$, we have $E^2 \approx \mathbf{p}^2c^2$, so $|\tilde{y}| - |\eta| \to 0^-$ as $E \to \infty$. Note that due

---

[3] In pAu collisions at RHIC, we take the proton-going direction to be positive $\tilde{y}$ (and z).

to this agreement at high energy, often the word “rapidity” is used, when what is meant is “pseudorapidity” (including multiple instances in the chapters of this thesis).

APPENDIX B LOW-JET-MASS DETECTOR RESPONSE STUDY

It was mentioned in Sec. 5.2 that the jet mass has a poor detector response at $M \sim 1\,\mathrm{GeV/c^2}$, and some explanation for this was given. However, it was important to understand this behavior because it persists at slightly above the threshold that was applied to detector-level jets. Therefore, several hypotheses were generated and tested on jets with $1.0 < M_{\mathrm{det}} < 1.5\,\mathrm{GeV/c^2}$ in an example p_{T} range of $(20, 25)\,\mathrm{GeV/c}$ (call this population of jets "L", for "low-mass, low-p_{T}"). This detailed study is presented here. The hypotheses for the cause of the secondary peak at roughly $M_{\mathrm{det}}/M_{\mathrm{part}}$ (or equivalently, for the cause of a distinct population of particle-level jets with $M \sim 4\,\mathrm{GeV/c^2}$ which lose roughly 75% of their mass) are as follows:

1. Jets with M_{det} in this selection originated as particle-level jets with typical M_{part} but lost a large percentage of their tracks due to detector efficiency.

2. These jets did not lose more tracks in the detector than average, but the tracks lost (either due to tracking inefficiency or momentum smearing reducing their momentum below the selection threshold of 200 MeV/c) were at wide angles to the jet axis, reducing the mass by a larger amount than otherwise equivalent tracks close to the jet axis would.

3. These jets may be ones which experienced a significant shift in the jet axis from clustering the particle-level and detector-level events. This shift may cause wide-angle particles at the boundary of the jet cone to be removed (or included).

4. This population of jets may be one for which our nominal hadronic correction of 100% to the towers, known to be an overcorrection in general, was much higher than the actual subtraction that should have been done, causing more towers to be removed from these jets than from the typical population.

5. As mentioned in Sec. 5.2, these jets may have a higher-than-typical number of heavy particles, which are then assigned $m = 0$, or $m = m_{\pi^{\pm}}$ at detector-level.

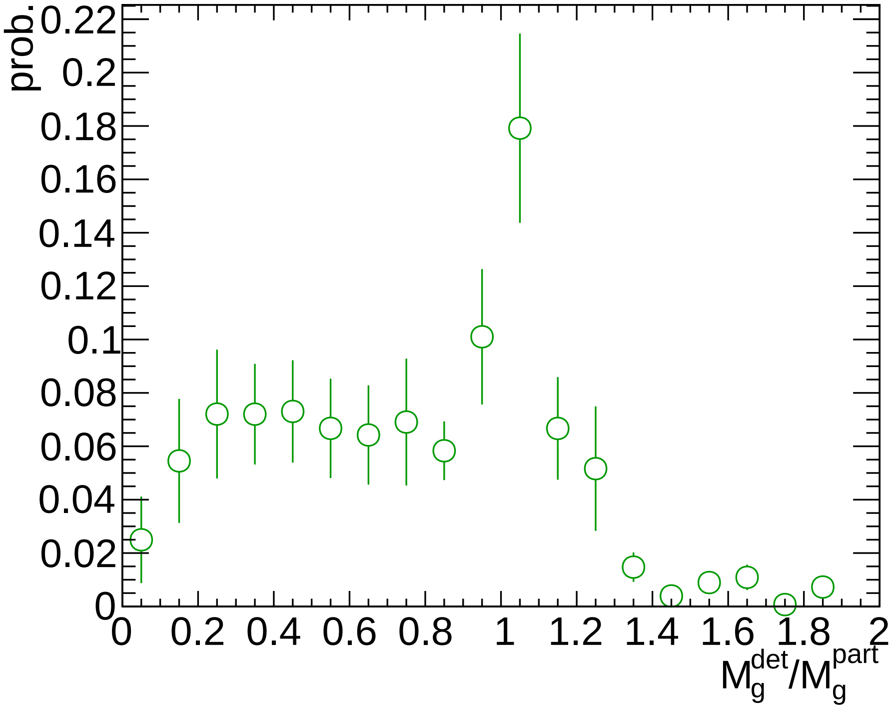
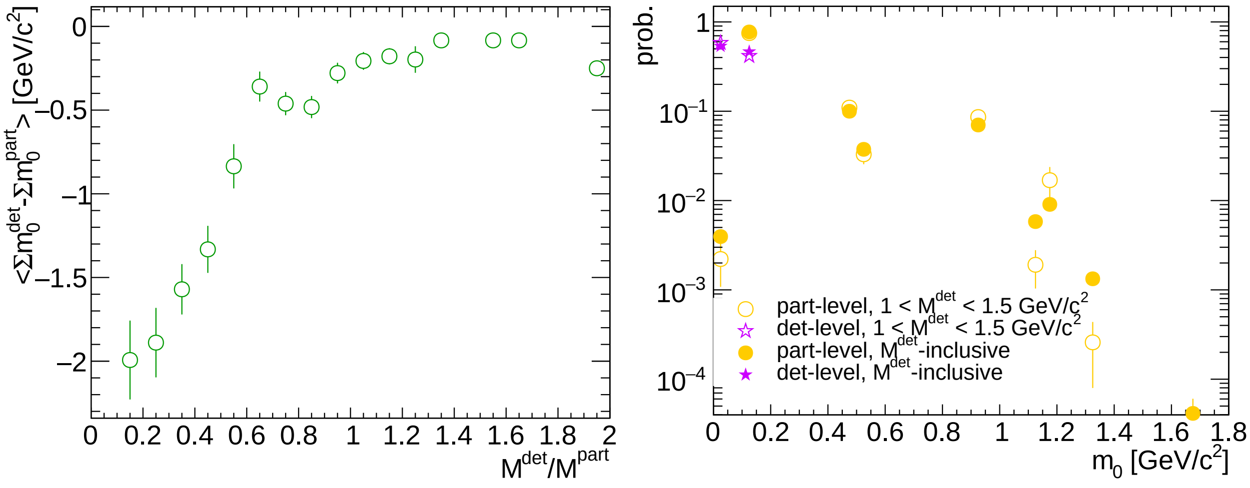


Figure B.1: The groomed jet mass response, using jets originating from ungroomed jets in population L (see text for definition).

6. For completeness, although this last hypothesis was not checked in this study due to limitations of our simulation, it should also be mentioned that the particle-level simulation, PYTHIA-6, has weak decays turned off because they are handled by GEANT at the same time as it models detector effects. So a jet may have a heavy particle which decays into two lighter particles, one of which is lost. These jets would indeed have an initially large mass, as the mother particle(s) would be wide-angle and have a large rest mass. And if would occur frequently enough, it could be a significant effect.

Hypothesis 1: These detector-level jets did indeed lose more than half of their tracks to tracking inefficiency more often than did other jets, but this was still not nearly as likely as the scenario in which the jet lost no tracks, in either case. This hypothesis is not supported. *Hypothesis 2*: Low-p_T particles are more often found at wide angles, so if they are lost due to tracking inefficiency (which is worse for low p_T) or momentum smearing reducing their p_T below the analysis threshold at 200 MeV/c, this could have a dramatic effect on the resulting



(a) (b)

Figure B.2: (a): Average rest mass lost between particle- and detector-levels as a function of the jet mass detector response, for population L (see text for details). (b): Probability of finding a particle with a given rest mass within a particle-level (orange) and detector-level (violet) jet. Both the inclusive jet population (solid markers) and population L (open markers) are shown.

jet mass. The p_{T} turns out not to be an important factor, which we tested by reducing the threshold to 0 GeV/c, and observed no significant effect. However, to test whether soft, wide-angle particles were being lost to tracking efficiency at a higher rate than for the larger jet sample, we examined the effect of grooming on these jets. Since SoftDrop grooms away soft, wide-angle particles, if there were any reduction in the secondary peak in the mass response, this could be seen as evidence in favor of this hypothesis. This is exactly what we observed (Fig. B.1), although it did not completely eliminate this effect. This hypothesis is partially supported.

Hypothesis 3: The typical shift of the jet axis from particle- to detector-level was minuscule (less than the tower size). This hypothesis is not supported.

Hypothesis 4: The nominal hadronic correction of 100% was varied to 50% to test whether it was actually an overcorrection for these jets. No significant change to the resolution was observed. This hypothesis is not supported.

Hypothesis 5: To test whether the secondary low-mass peak in the response is due to particles' rest masses being misassigned, we can sum the rest masses at particle- and detector-levels and take the average difference between the two for each bin in the response. This is shown in Fig. B.2a. We observe that around the secondary peak in the mass response at $M_{\text{det}}/M_{\text{part}} = 0.4$, jets in population L have lost on average about 50% of their mass just from rest mass loss, as opposed to the roughly 15% loss for jets for which $M_{\text{det}}/M_{\text{part}} \approx 1$. One thing to note is that this rest mass loss observable conflates tracking losses with misassignment rest mass loss, but as mentioned previously, tracking losses are small. So it is clear that this is a large contribution to the secondary peak in the detector response at low values for low-mass jets. The last question to answer is whether the particle-level jets for population L, for whatever reason, have more heavy particles than typical, or whether they are normal in that respect, but the relative effect of a few misassignments is just larger for them, due to their already low jet mass. We see from Fig. B.2b that the former is not the case: the relative abundance of e.g. kaons and protons in particle-level jets is quite similar. So it seems that the only pathological feature of these jets is that they lose more soft, wide-angle radiation than is typical, and their low overall mass means that rest mass misassignment has a larger relative effect on the measured detector-level jet mass.

APPENDIX C SIMULATION SETTINGS AND TUNES

In this section, we list the various parameters that were used to run the models for comparison to the corrected data, presented in Sec. 5.6. HERWIG-7.1.4 using MMHT2014lo68cl PDFs [144] was run by Dr. Raghav Kunnawalkam Elayavalli, with the following settings (reproduced from the STAR-internal analysis note for Ref. [38]):

- Dijet QCD

```
cd /Herwig/MatrixElements/
insert SubProcess:MatrixElements[0] MEQCD2to2
cd /Herwig/Cuts/
set JetKtCut:MinKT XY.0*GeV #minimum pthat of the event, varied from 5-60
set JetKtCut:MaxKT XY.0*GeV #maximum pthat of the event, varied from 10-80
```

- Underlying event

```
set /Herwig/UnderlyingEvent/MPIHandler:EnergyExtrapolationPower
set /Herwig/UnderlyingEvent/MPIHandler:ReferenceScale 200.*GeV
set /Herwig/UnderlyingEvent/MPIHandler:Power 0.24
set /Herwig/UnderlyingEvent/MPIHandler:pTmin0 3.91*GeV
```

- Color reconnection settings

```
set /Herwig/Hadronization/ColourReconnector:ColourReconnection Yes
set /Herwig/Hadronization/ColourReconnector:ReconnectionProbability 0.61
```

- Color disrupt settings

```
set /Herwig/Partons/RemnantDecayer:colourDisrupt 0.75
```

- Inverse hadron radius

```
set /Herwig/UnderlyingEvent/MPIHandler:InvRadius 1.35
```

- MPI model settings

```
set /Herwig/UnderlyingEvent/MPIHandler:softInt Yes
set /Herwig/UnderlyingEvent/MPIHandler:twoComp Yes
set /Herwig/UnderlyingEvent/MPIHandler:DLmode 2
```

PYTHIA-6.428 [106] with the Perugia 2012 tune (#370) [145] using CTEQ6L1 LO PDFs [146] was found to match STAR pp data well for $p_{\mathrm{T}} \gtrsim 3\,\mathrm{GeV/c}$. However, below $3\,\mathrm{GeV/c}$, one sees dramatic disagreement between the PYTHIA prediction and STAR data. For this reason, a single parameter called PARP(90), which is part of the infrared regularization scaling function from the reference energy scale of $\sqrt{s} = 7\,\mathrm{TeV}$, was reduced from the default value for Perugia 2012 by 11.25%, resulting in much better agreement at low energy [100]. Perugia 2012 with this change will be referred to throughout this thesis as the STAR tune. PYTHIA-8.303 [112] with the Monash 2013 tune [147] using NNPDF2.3 QCD+QED LO PDFs ($\alpha_S(m_Z) = 0.130$) [148] was run with default settings (with one exception – see Sec. 5.1.2) for 1 million $2 \to 2$ QCD quark/gluon production events each, specified with the switch:

```
HardQCD:all = on
```

in 11 $\hat{p}_{\mathrm{T}}$ bins. The p_{T} of the hard partonic process is specified for two reasons: first, the cross section diverges at low p_{T}, so we must cut off at some value; second, the spectrum is steeply-falling, especially at $\sqrt{s} = 200\,\mathrm{GeV}$ where the cross section goes like p_{T}^{-6} [149], which means that if we run PYTHIA events with say $\hat{p}_{\mathrm{T}} > 1\,\mathrm{GeV/c}$ and no upper bound, we might expect roughly ten $35\,\mathrm{GeV/c}$ jets for every million $5\,\mathrm{GeV/c}$ jets. So to build enough statistics throughout the entire jet spectrum would take immense computing resources. Instead, as mentioned, we ran 1 million events per bin ($5 < \hat{p}_{\mathrm{T}} < 10\,\mathrm{GeV/c}$, and up to $50\,\mathrm{GeV/c}$ in

5 GeV/c intervals, followed by larger bins, $(50, 60)$ and $(60, 80)$ GeV/c, where statistics are less important).

After these 1 million events per $\hat{p}_{\mathrm{T}}$ are generated, we must stitch them together to form a physical steeply-falling jet spectrum, e.g. like the one in Fig. 1.5. This is done by summing the values given by

```
pythia.info.sigmaGen(0)
```

for each event and dividing by

```
pythia.info.weightSum()
```

after all events have been recorded, with a similar procedure for HERWIG. This gives the average cross section for all allowed processes in this $\hat{p}_{\mathrm{T}}$-bin. Then, when jets are clustered and filled into trees and histograms, the jet observables are weighted by this cross section weight depending on which $\hat{p}_{\mathrm{T}}$-bin they originated from. For the most part, this gives a smooth, steeply-falling jet spectrum. However, since the partonic cross section does not exactly set the jet cross section, it is possible in rare cases to obtain jets with $p_{\mathrm{T,jet}} \gg \hat{p}_{\mathrm{T}}$. Since the cross section weight diminishes monotonically as p_{T} increases, these few jets will receive a large weight, compared to jets of the same p_{T} from a more typical $\hat{p}_{\mathrm{T}}$ bin. The spectrum will then have unphysical spikes with large uncertainties. To reduce this effect, in the analysis we veto events with a jet of $p_{\mathrm{T}} > 2\hat{p}_{\mathrm{T}}{-}\mathrm{max}$. This gives a reasonable cross section.

Lastly, for the pAu analysis presented in Ch. 6, a PYTHIA-8 model called Angantyr [127] was used for comparisons with the small-systems data. Angantyr is invoked simply by specifying at least one beam nucleus to be larger than a proton. In this case, PYTHIA-8.243 was run (again for one million events per $\hat{p}_{\mathrm{T}}$-bin) with the following beam parameter settings[1]:

```
Beams:frameType = 2 ! enables asymmetric back-to-back beams

Beams:idA = 2212 ! first beam, p
```

---

[1] Beam energies obtained from https://www.agsrhichome.bnl.gov/RHIC/Runs/index.html#Run-15

```
Beams:idB = 1000791970 ! second beam, Au

Beams:eA = 103.88 ! first beam energy

Beams:eB = 97.74 ! second beam energy
```

but with otherwise the exact same user-defined settings as the pp model.

APPENDIX D FIGURES FOR $R = 0.2, 0.6$

The jet mass scale and resolution are shown in Figs. D.1 and D.2. These figures are in the same style as the figure shown in the paper for $R = 0.4$ jets (Ref. [108]).

The MC closure tests for $R = 0.2, 0.6$ are shown in Fig. D.3 and Fig. D.4, respectively. Note that the non-closure in the high-mass tail of the $30 < p_{\mathrm{T}} < 40\,\mathrm{GeV/c}$ for $R = 0.6$ jets is a statistical fluke due to a small number of jets in this region. Unfolding subpopulation A with the response matrix from subpopulation B produces the result shown, but unfolding sub-population B with the response matrix from subpopulation A produces a result roughly within 10% throughout, even at high jet mass.

The systematic uncertainties for $R = 0.2, 0.6$ jets are shown in Figs. D.5 and D.7, respectively, while the maximum envelope of the unfolding uncertainty is shown in Figs. D.6 and D.8, respectively.

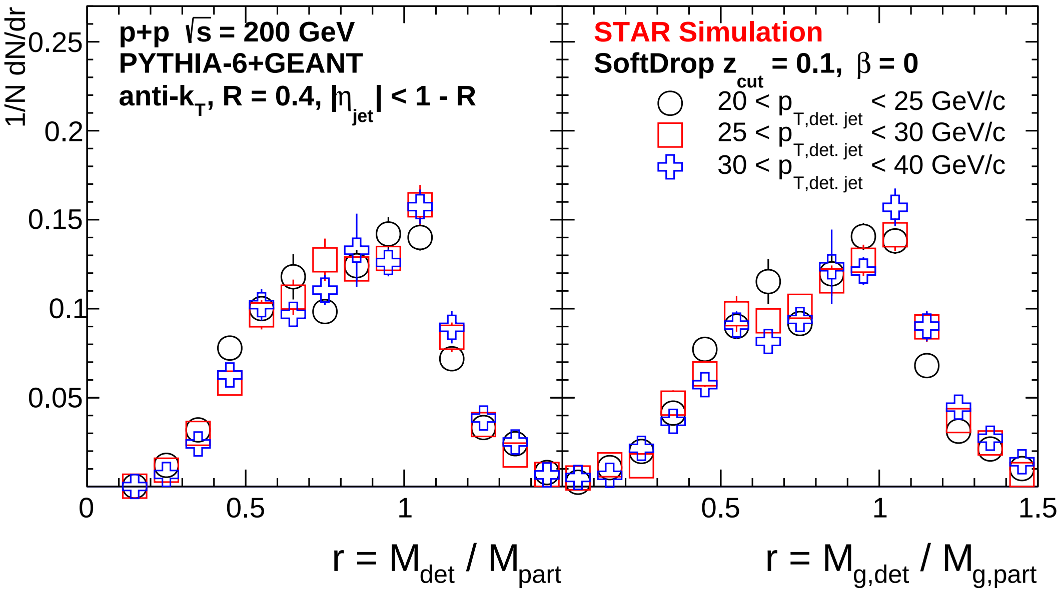


Figure D.1: Similar to Figs. 5.3 and 5.4, but for $R = 0.2$ jets, and with the p_{T} selections all shown on the same panel, and with the left panel showing the JMS and JMR, and the right panel showing the $\mathrm{JM_gS}$ and $\mathrm{JM_gR}$. Here the p_{T} selections shown are the same as for the results.

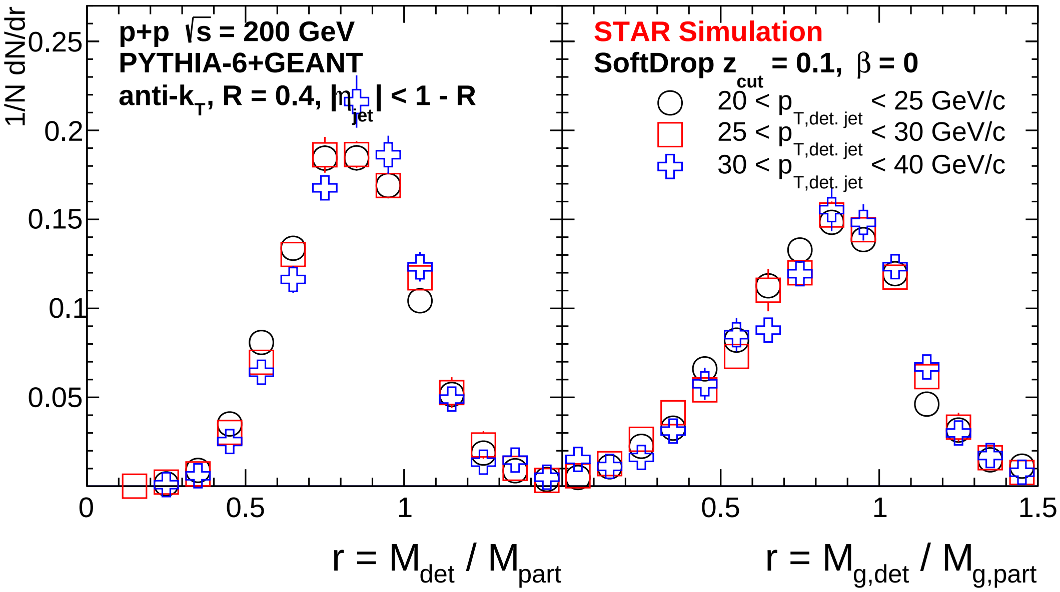


Figure D.2: The same as Fig. D.1, but for $R = 0.6$ jets.

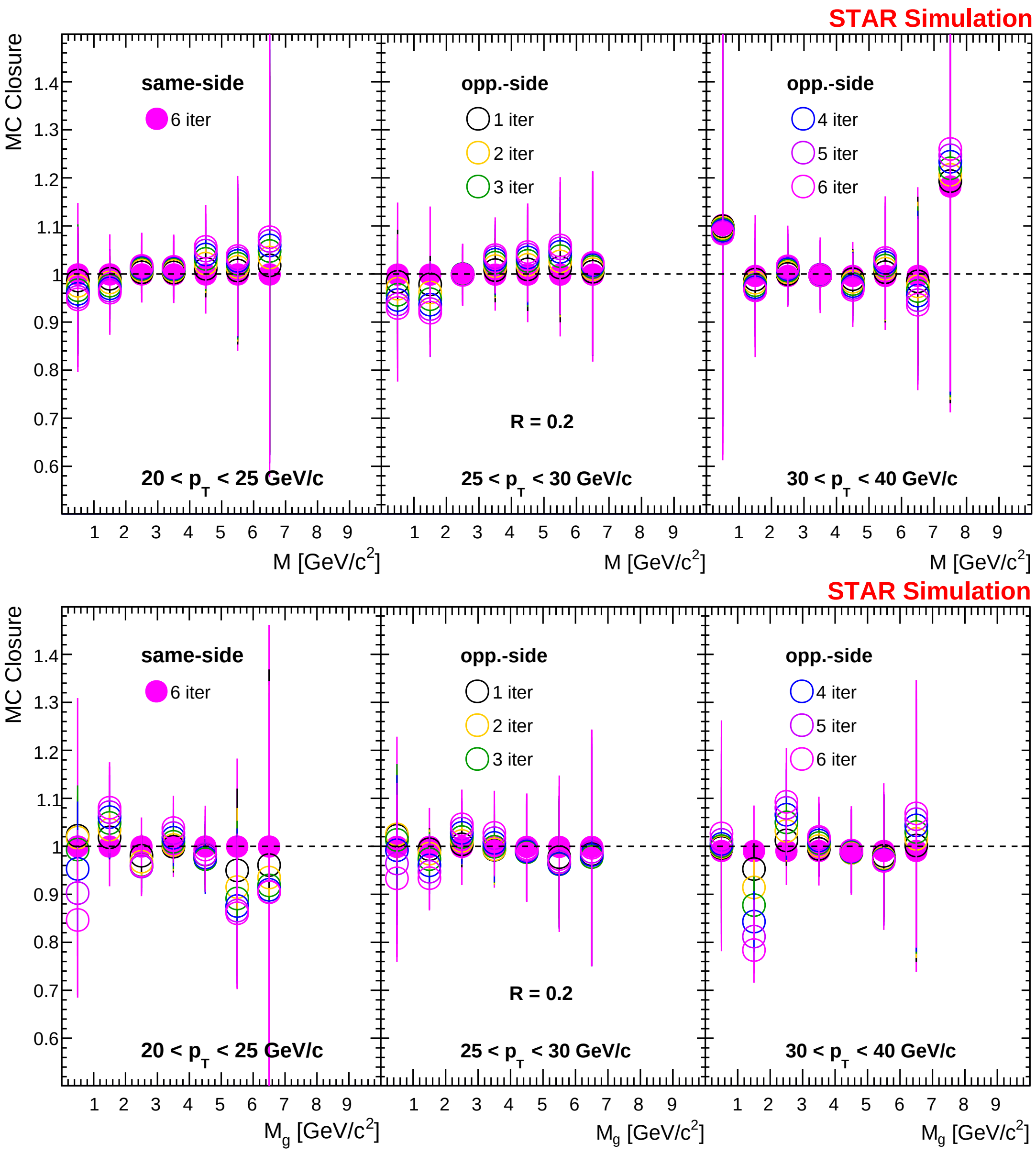


Figure D.3: The same as Fig. 5.14 but for $R = 0.2$ jets.

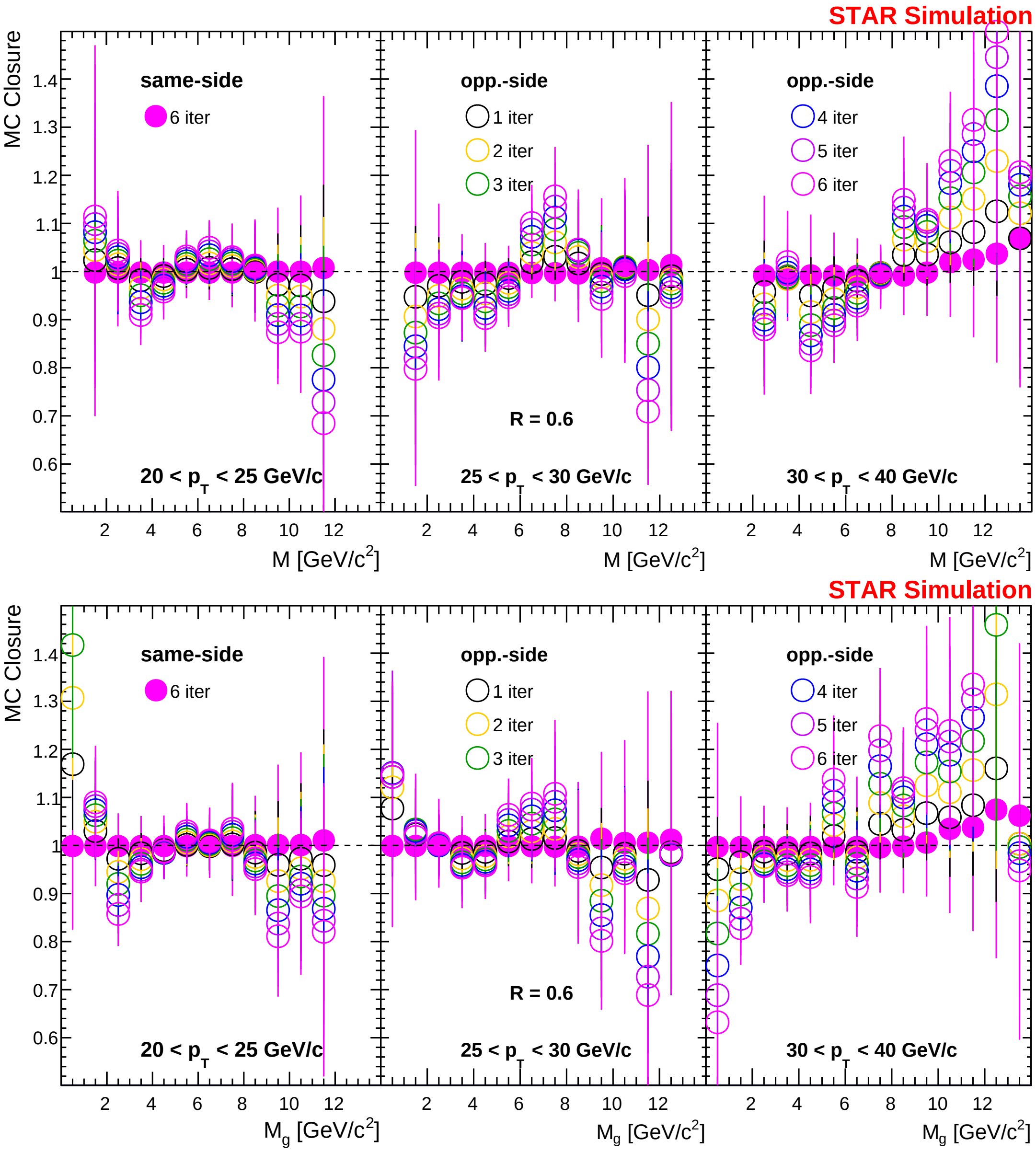


Figure D.4: The same as Fig. 5.14 but for $R = 0.6$ jets.

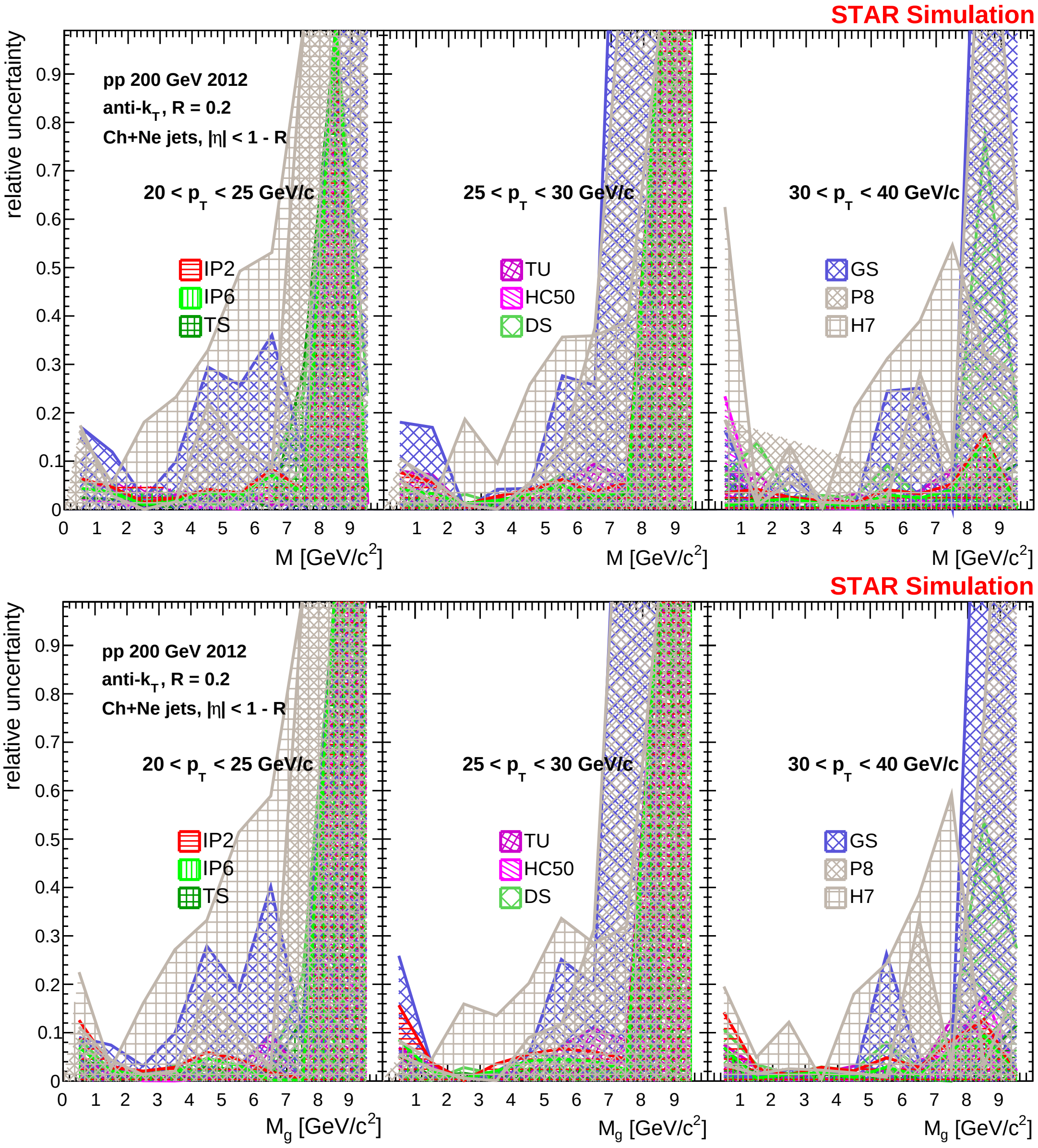


Figure D.5: The same as Fig. 5.16 but for $R = 0.2$ jets.

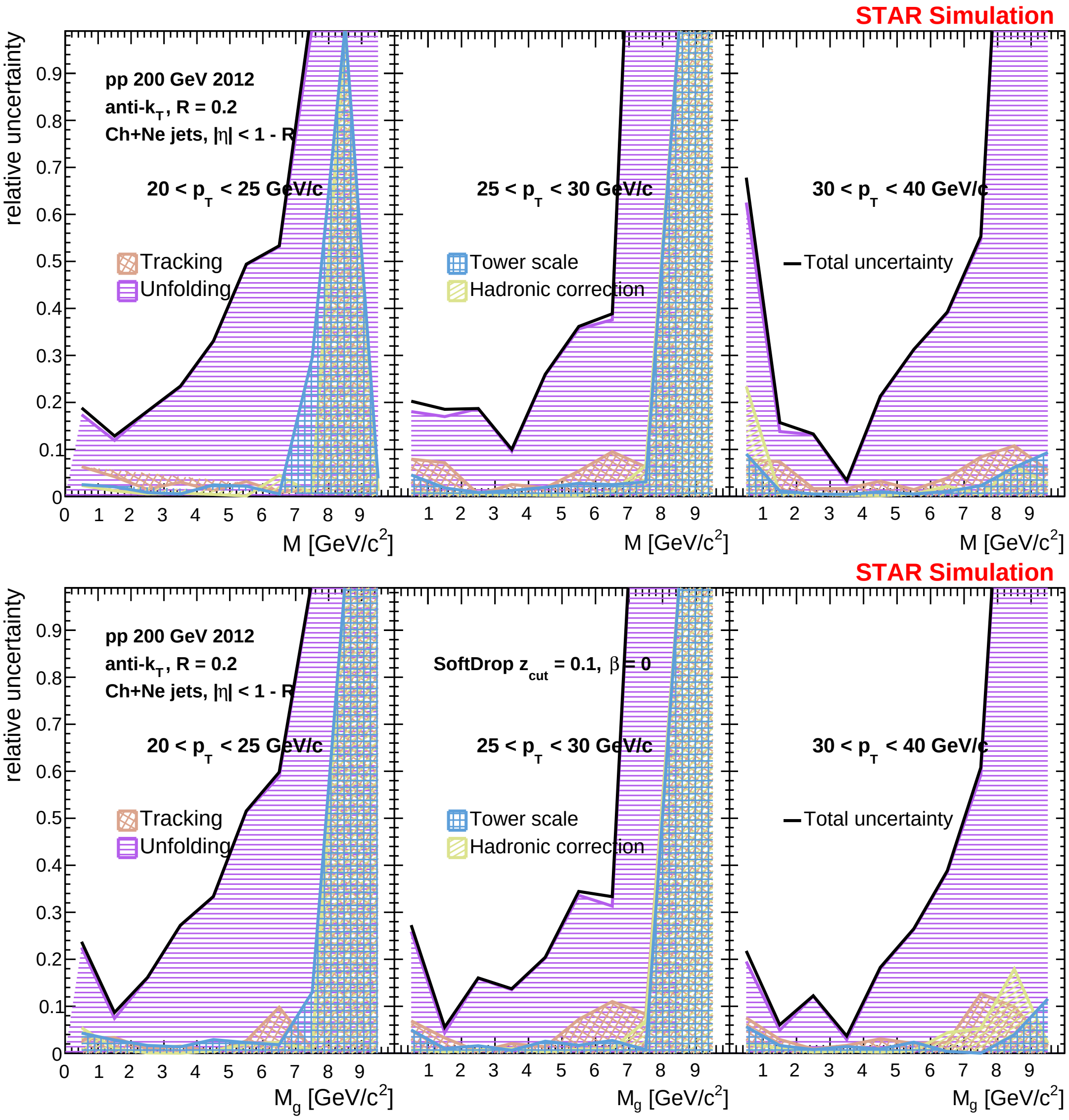


Figure D.6: The same as Fig. 5.17 but for $R = 0.2$ jets.

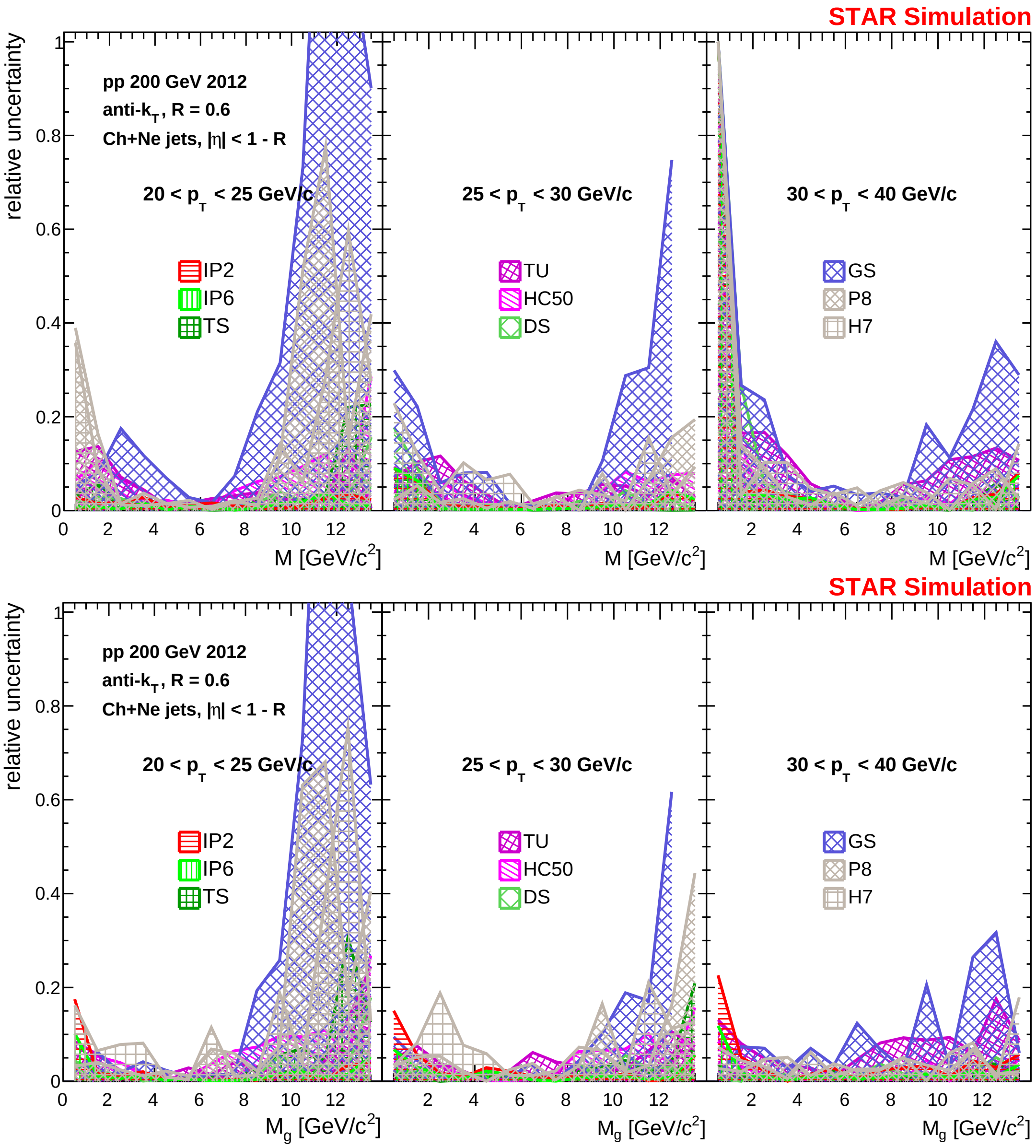


Figure D.7: The same as Fig. 5.16 but for $R = 0.6$ jets.

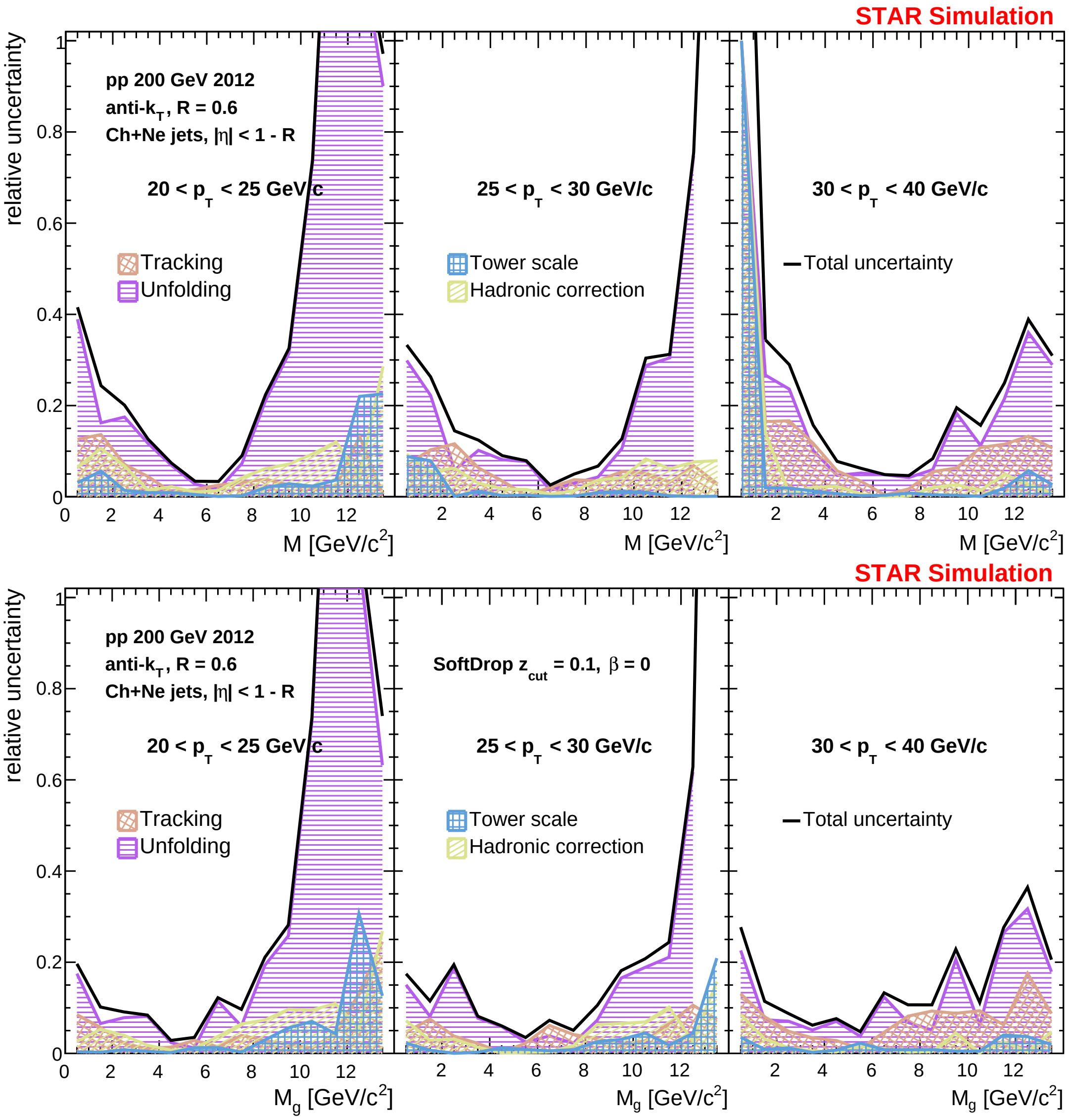


Figure D.8: The same as Fig. 5.17 but for $R = 0.6$ jets.

APPENDIX E pAu ANALYSIS DETAILS

The MC closure tests for pAu collisions are shown in Figs. E.1, E.2, E.3, and E.4 for M, M_{g}, z_{g}, and R_{g}, respectively, with jets from 0–30% high-EA (60–90% low-EA) events in the upper (lower) panel. We observe quite good closure of each observable in each activity range, with 10%-level or better agreement in nearly every bin near the peaks of the distributions.

The systematic uncertainties for 0–30% high-EA (60–90% low-EA) pAu collisions are shown in the upper (lower) panels of Figs. E.5, E.6, E.7, and E.8 for M, M_{g}, z_{g}, and R_{g}, respectively. The maximum envelopes of the uncertainties on the unfolding are shown alongside the various detector uncertainties, and the total uncertainty as a quadrature sum of these, in Figs. E.9, E.10, E.11, and E.12. We observe that in general, systematic uncertainties are similar in magnitude between EA classes, but with larger fluctuations in some bins near the high-EA distribution peaks (e.g. for $30 < p_{\mathrm{T}} < 40\,\mathrm{GeV/c}$ jets at $R_{\mathrm{g}} \approx 0.2$).

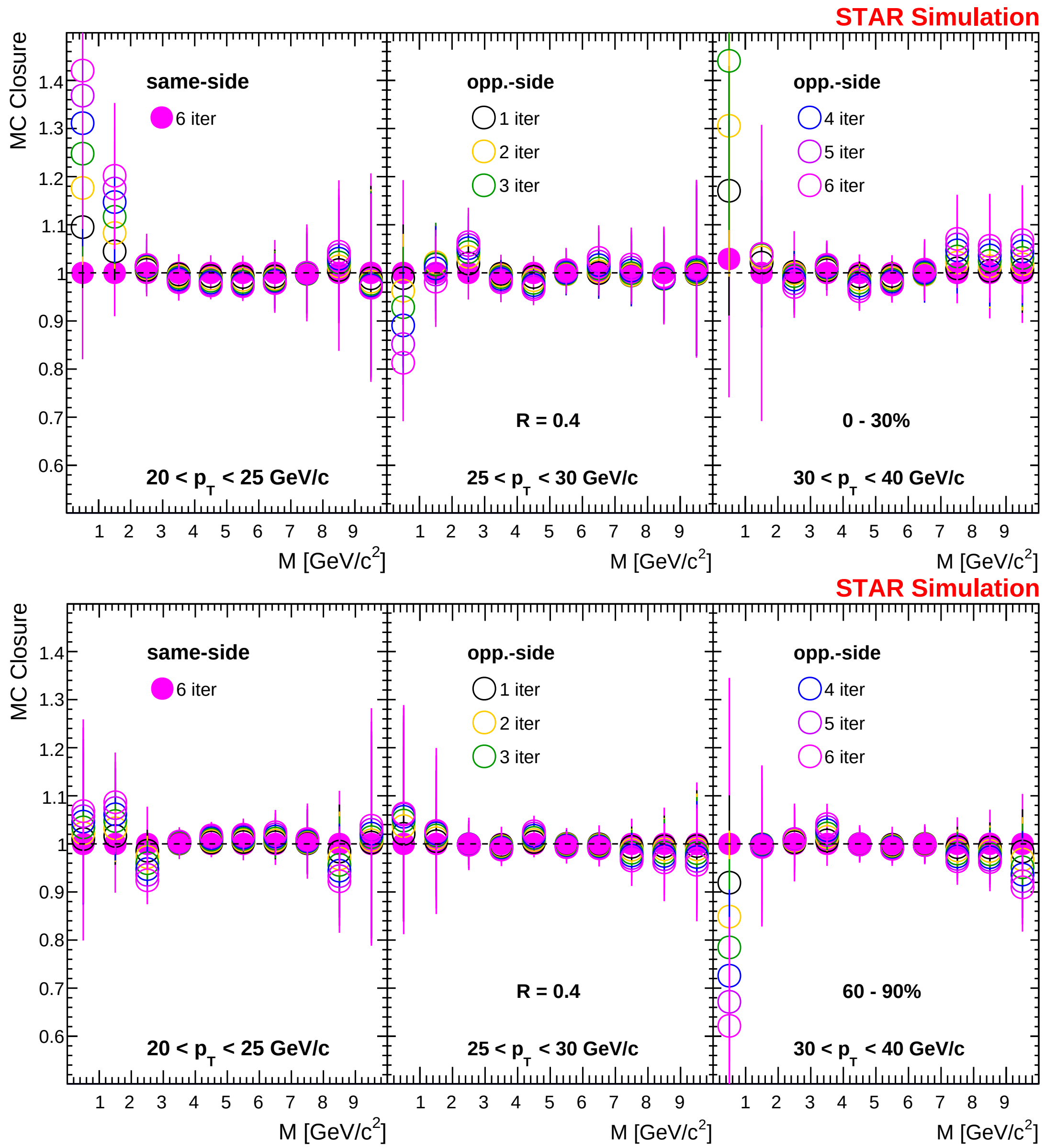


Figure E.1: The same as Fig. 5.14, but for the jet mass in high (0–30%) EA pAu collisions (upper panel) and low (60–90%) EA pAu collisions (lower panel).

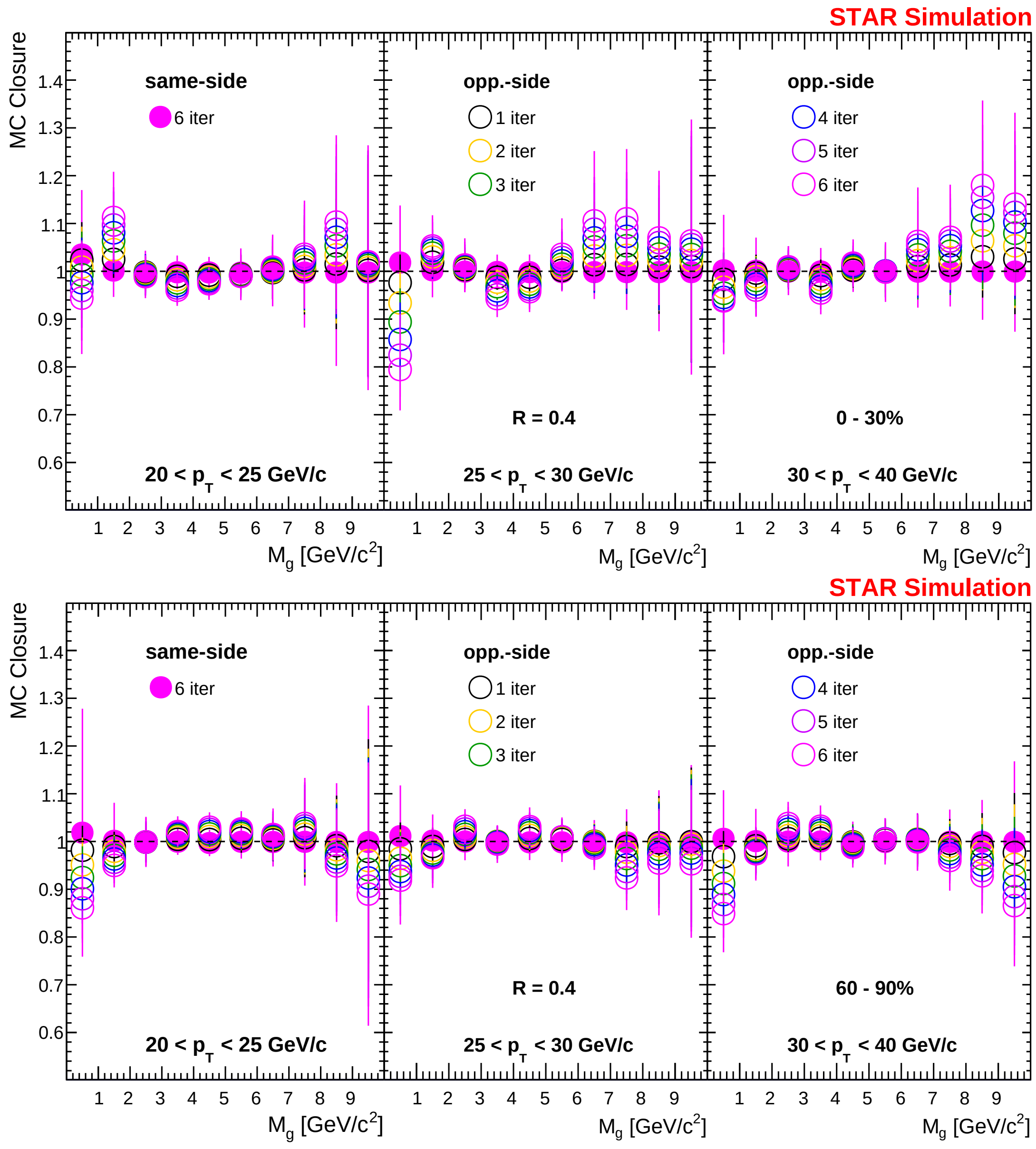


Figure E.2: The same as Fig. E.1, but for the groomed jet mass, M_g.

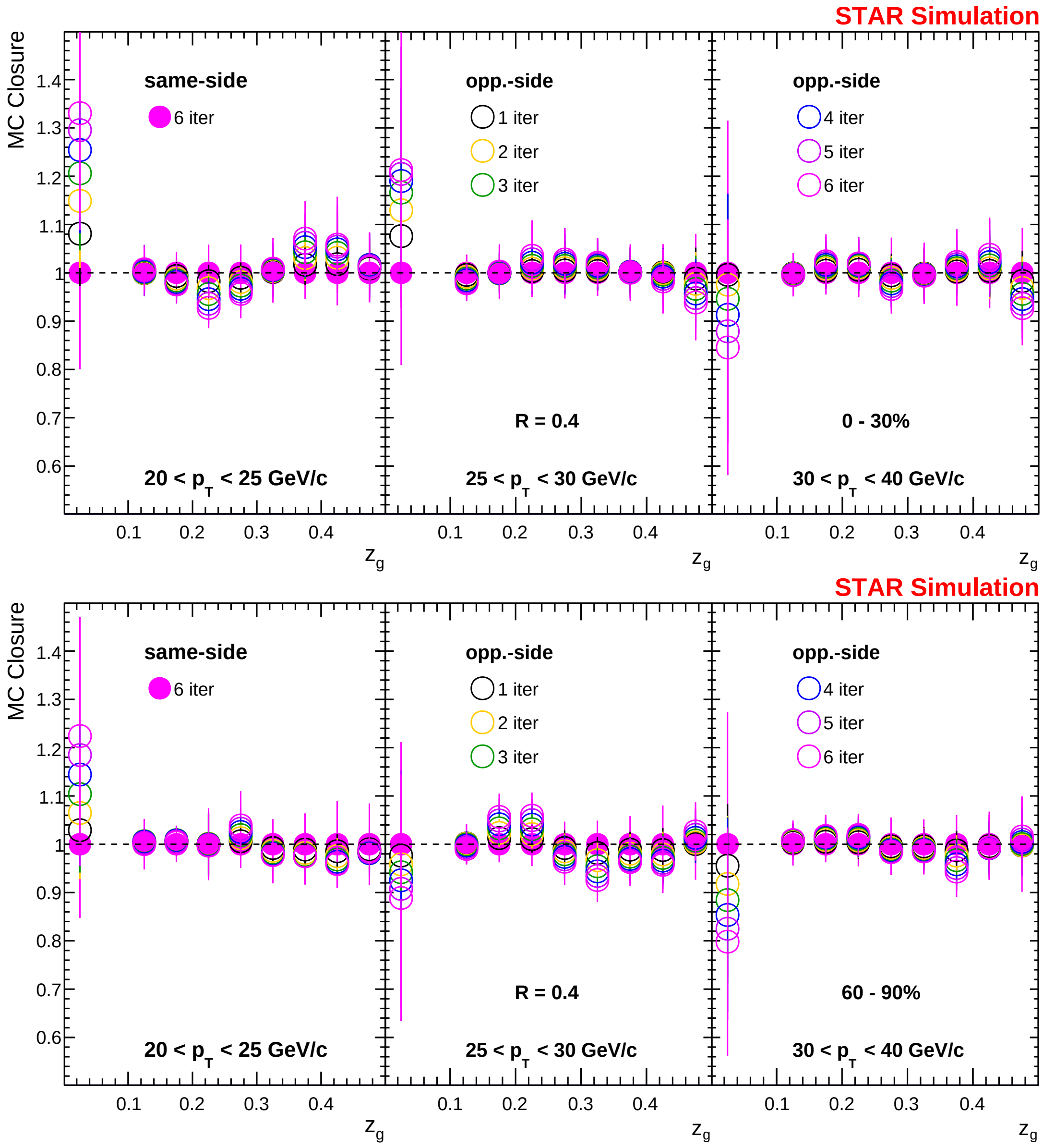


Figure E.3: The same as Fig. E.1, but for the subjet shared momentum fraction, z_g.

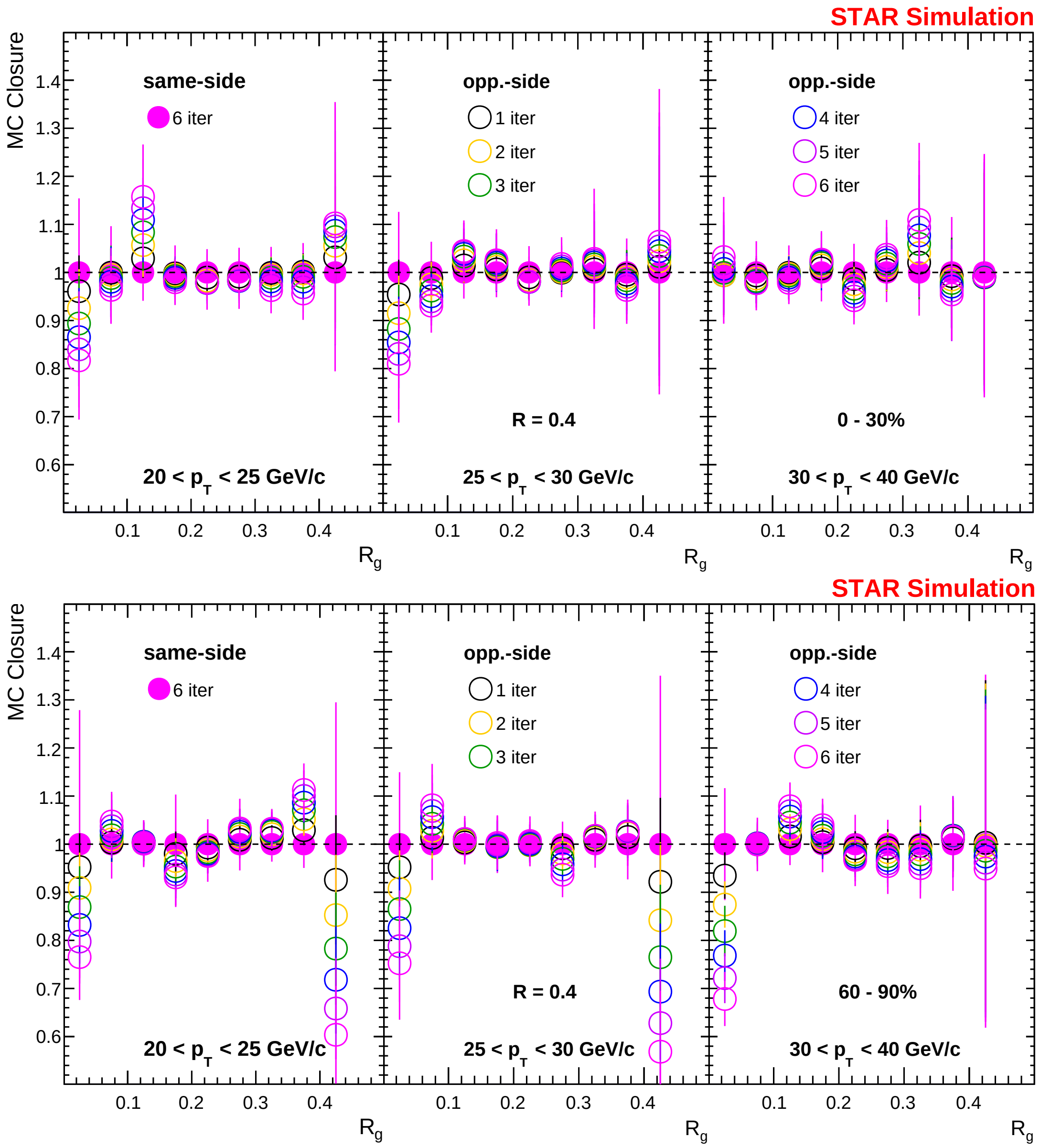


Figure E.4: The same as Fig. E.1, but for the groomed jet radius, R_g.

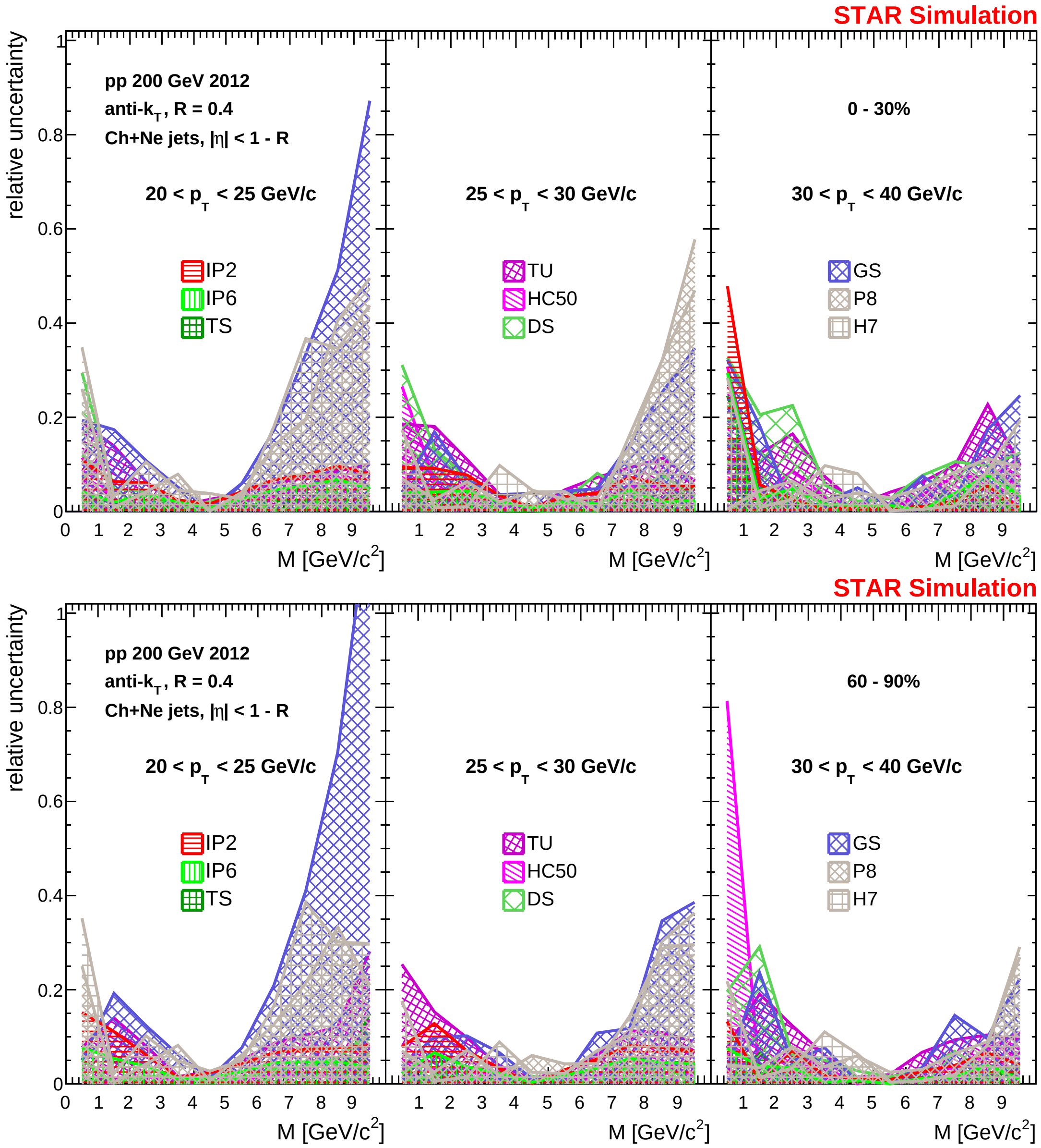


Figure E.5: The same as Fig. 5.16, but for the jet mass in high (0–30%) EA pAu collisions (upper panel) and low (60–90%) EA pAu collisions (lower panel).

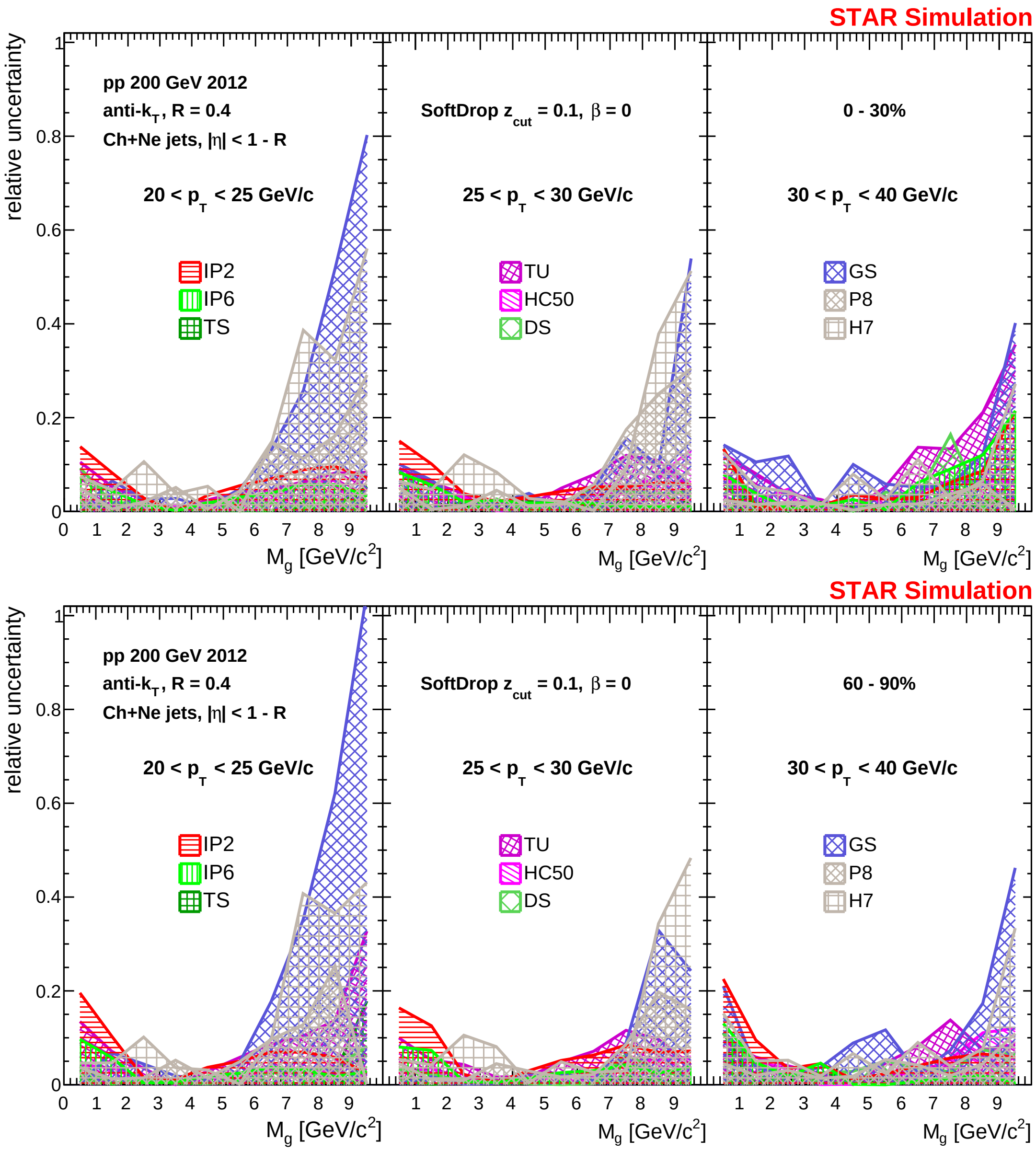


Figure E.6: The same as Fig. E.5, but for the groomed jet mass, M_g.

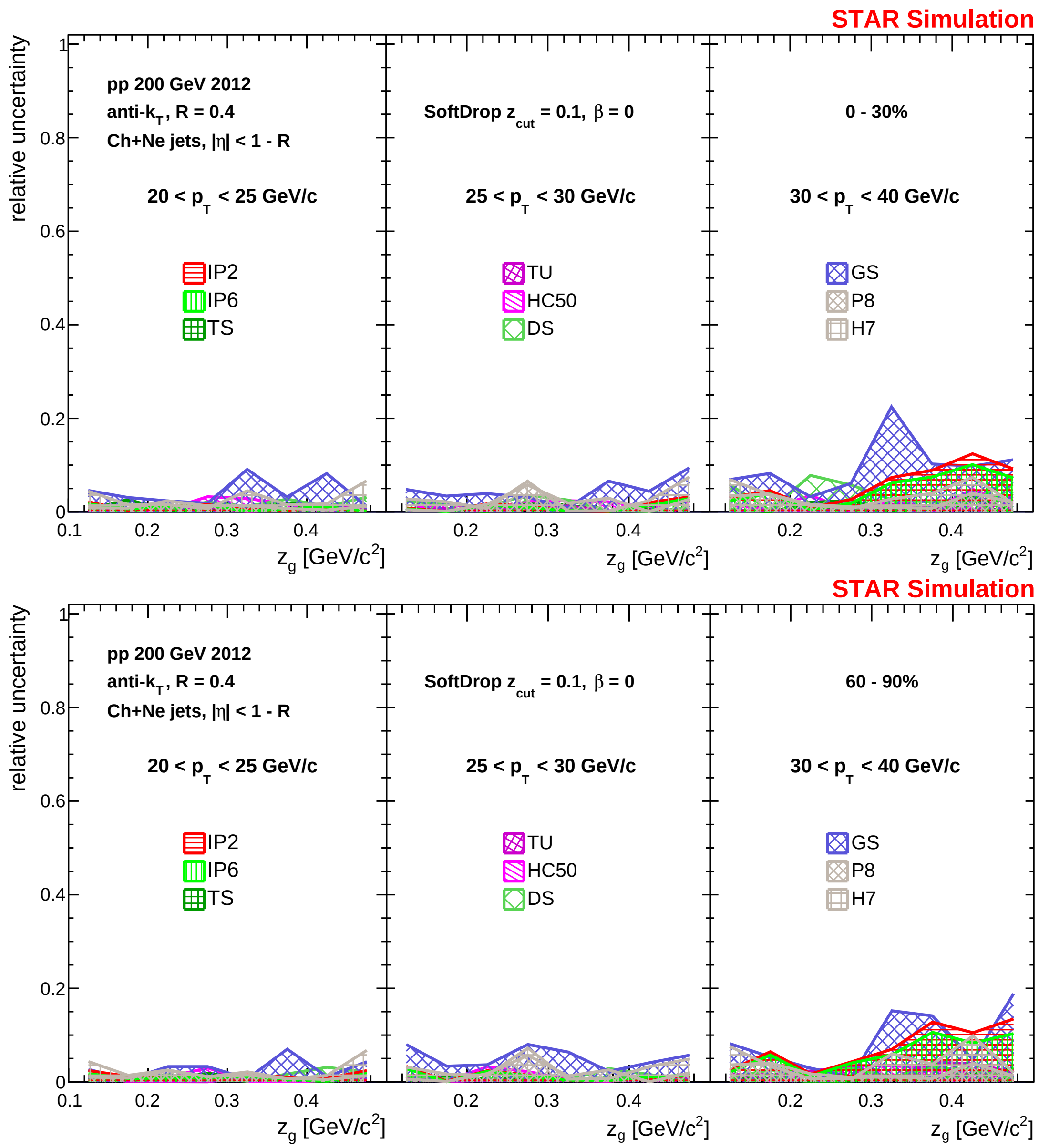


Figure E.7: The same as Fig. E.1, but for the subjet shared momentum fraction, z_g.

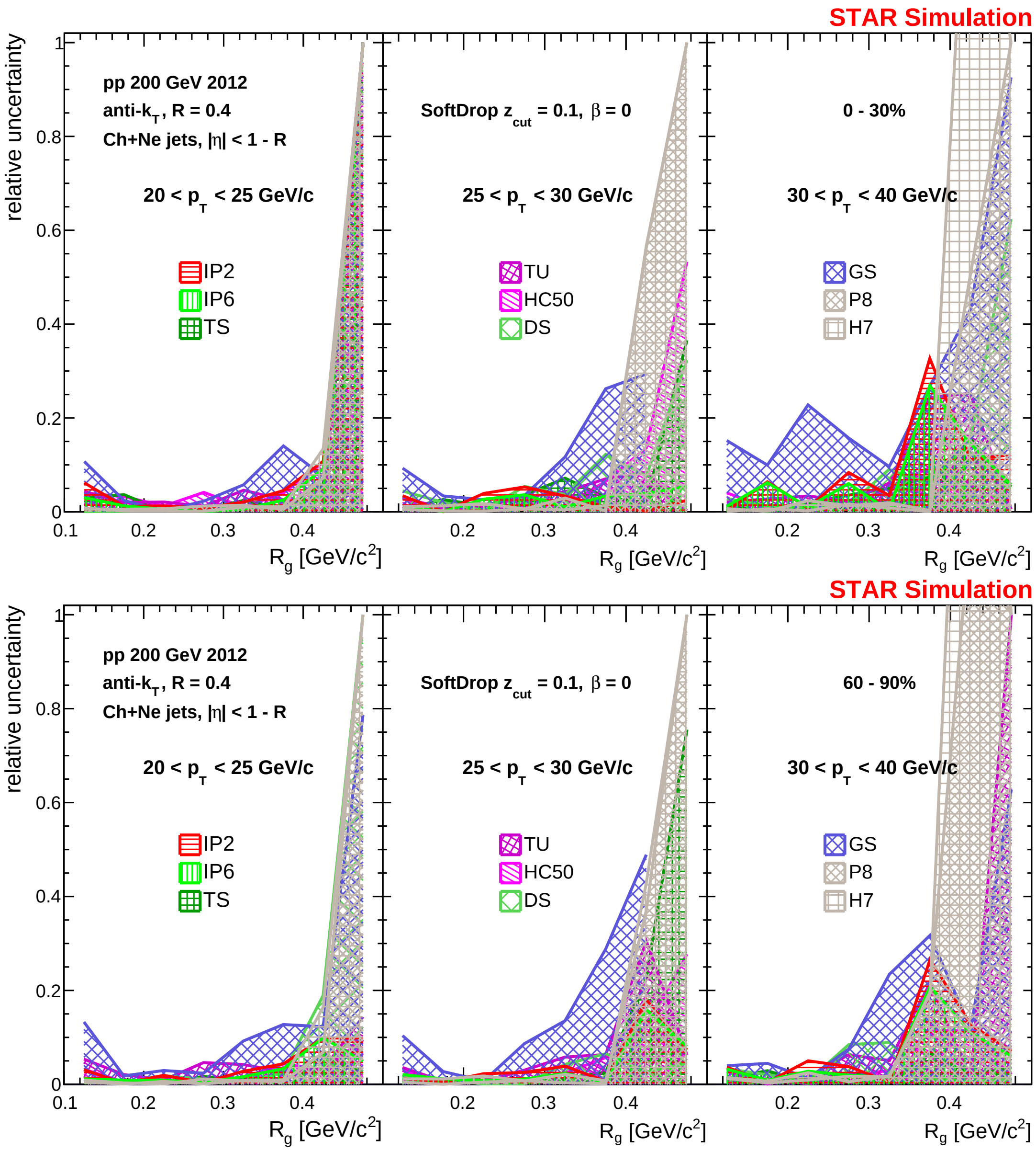


Figure E.8: The same as Fig. E.1, but for the groomed jet radius, R_g.

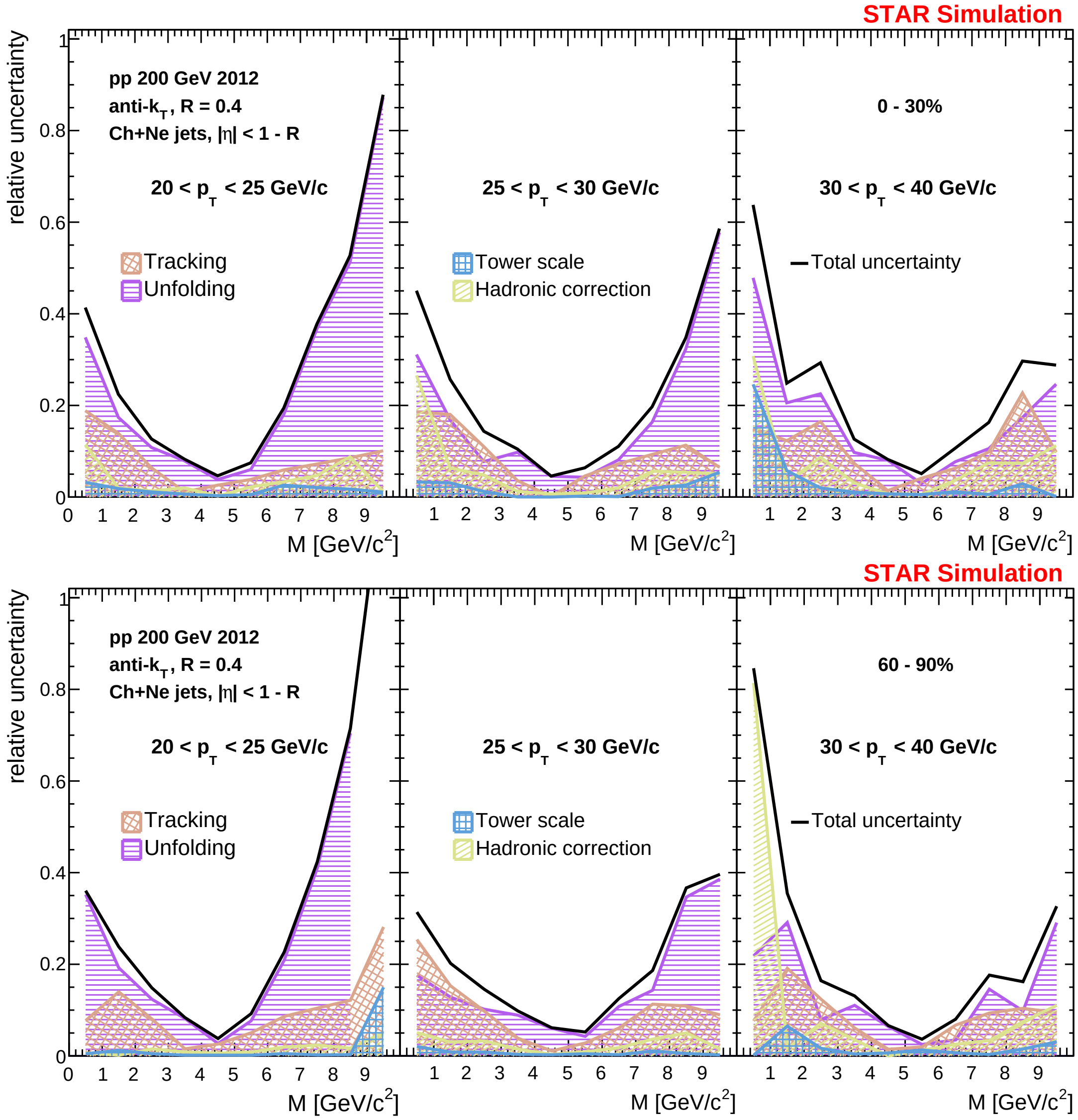


Figure E.9: The same as Fig. 5.16, but for the jet mass in high (0–30%) EA pAu collisions (upper panel) and low (60–90%) EA pAu collisions (lower panel).

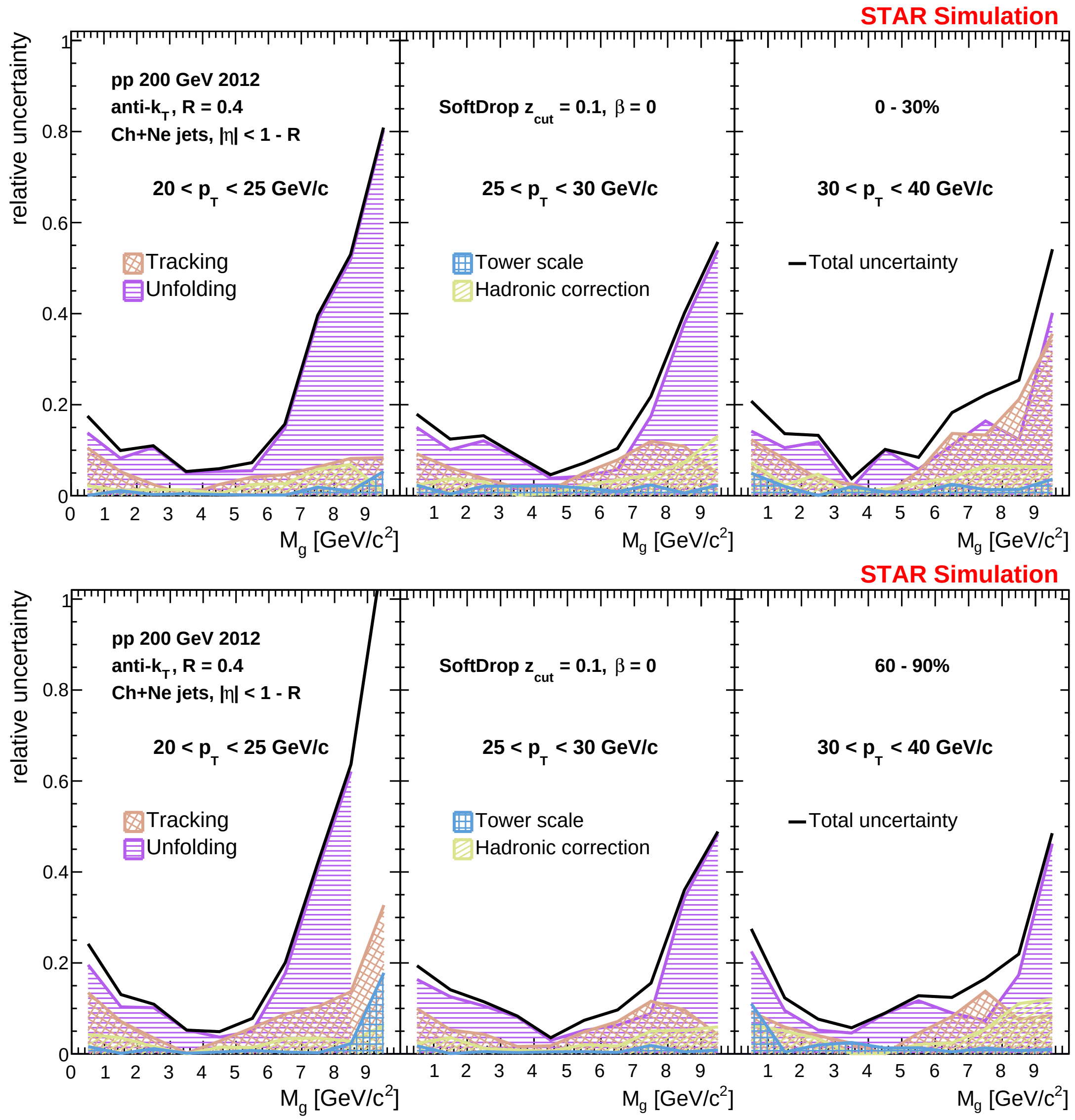


Figure E.10: The same as Fig. E.5, but for the groomed jet mass, M_g.

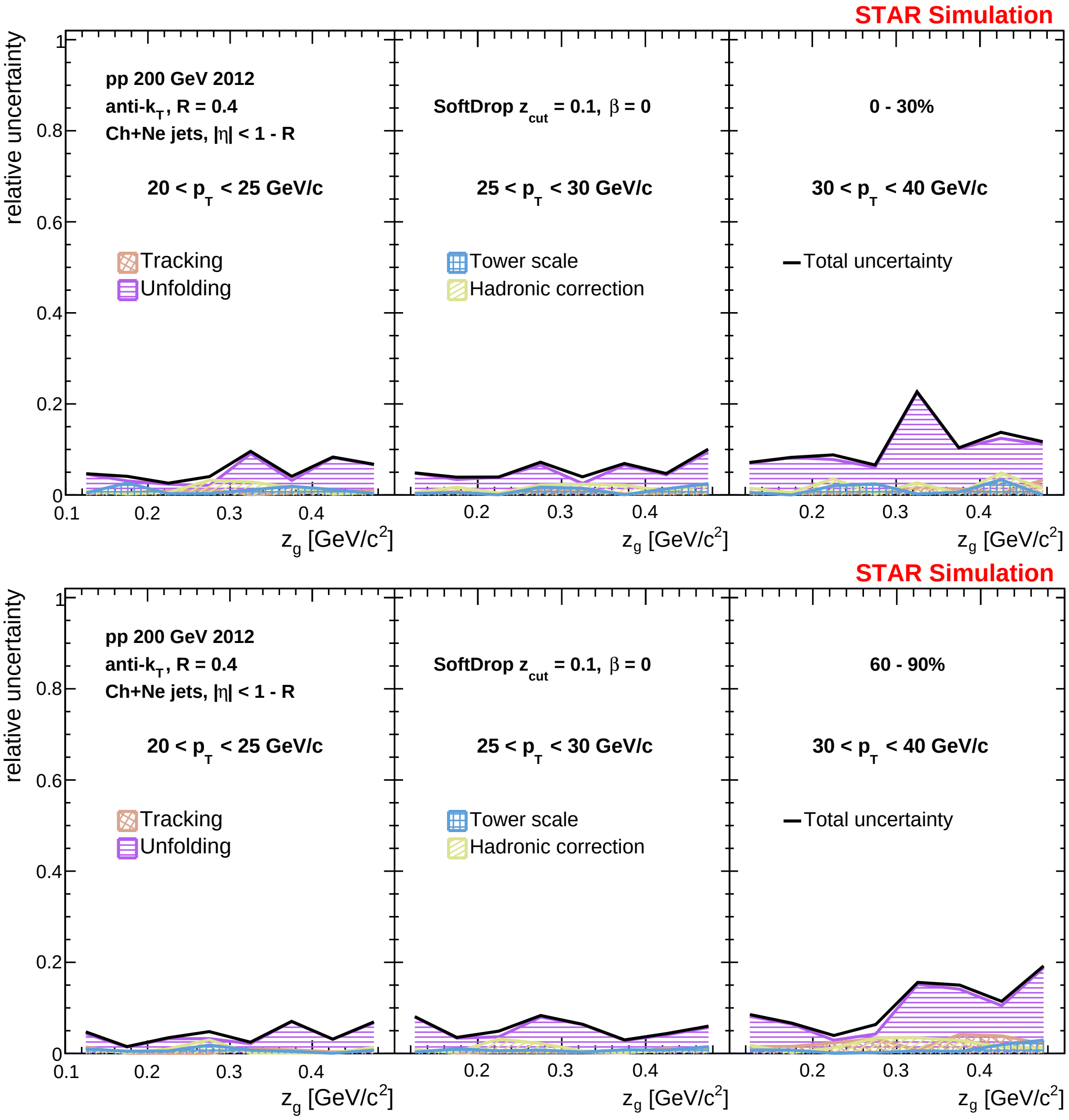


Figure E.11: The same as Fig. E.5, but for the subjet shared momentum fraction, z_g.

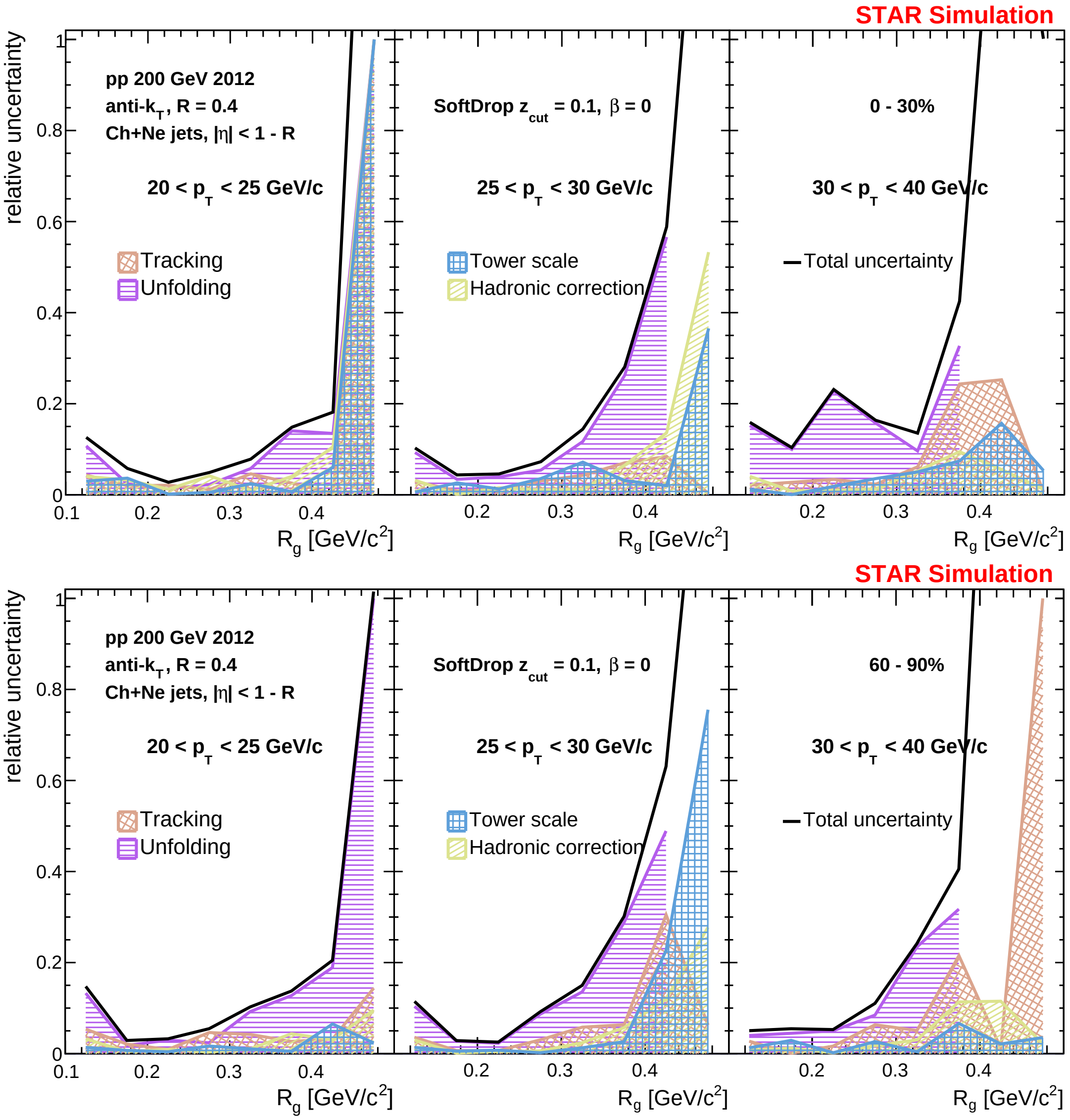


Figure E.12: The same as Fig. E.5, but for the groomed jet radius, R_g.

APPENDIX F FIGURES FOR 60–90% EVENT ACTIVITY

The remaining fully corrected jet substructure results not shown in the text, for a low event activity range (60–90%) where we would expect good agreement with the pp results, are shown in Figs. F.1, F.2, and F.3. We see in general a similar level of agreement with the pp data and a similar overprediction by the model of the angular scale (with agreement on the subjet shared momentum fraction).

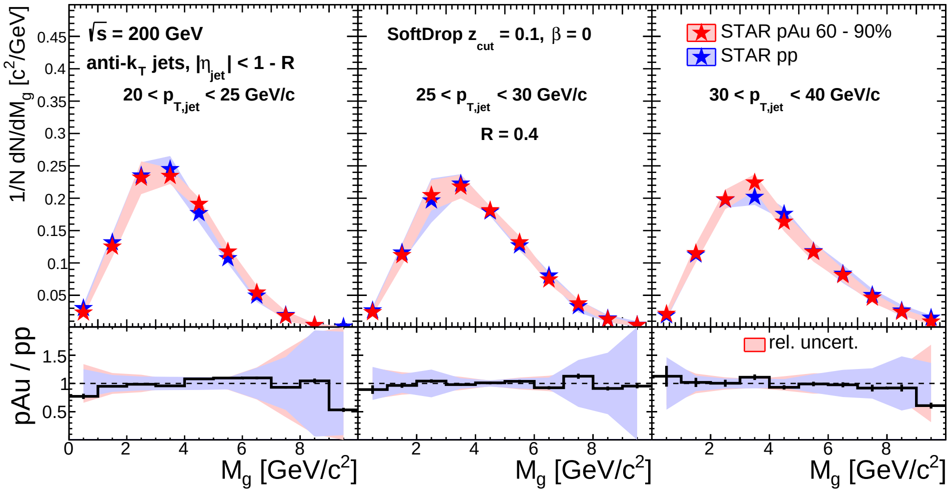


Figure F.1: The same as Fig. 6.10, but for a low (60–90%) event activity range.

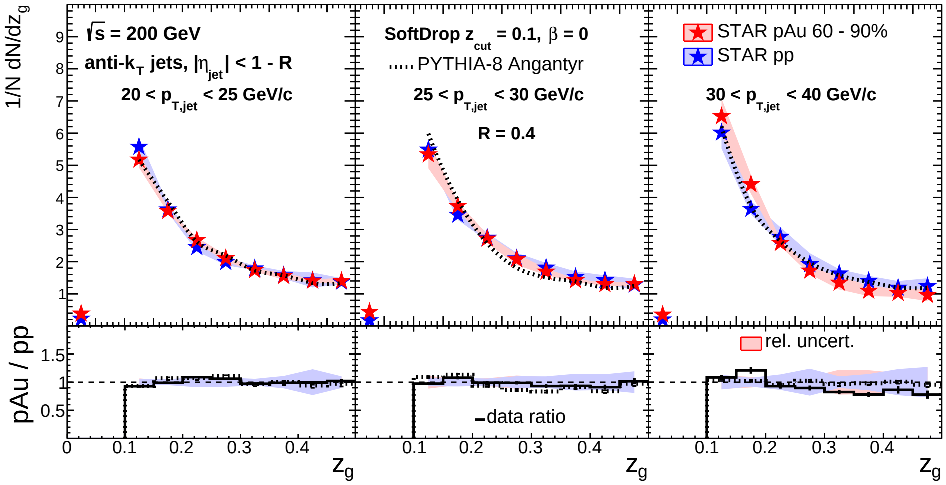


Figure F.2: The same as Fig. 6.11, but for a low (60–90%) event activity range.

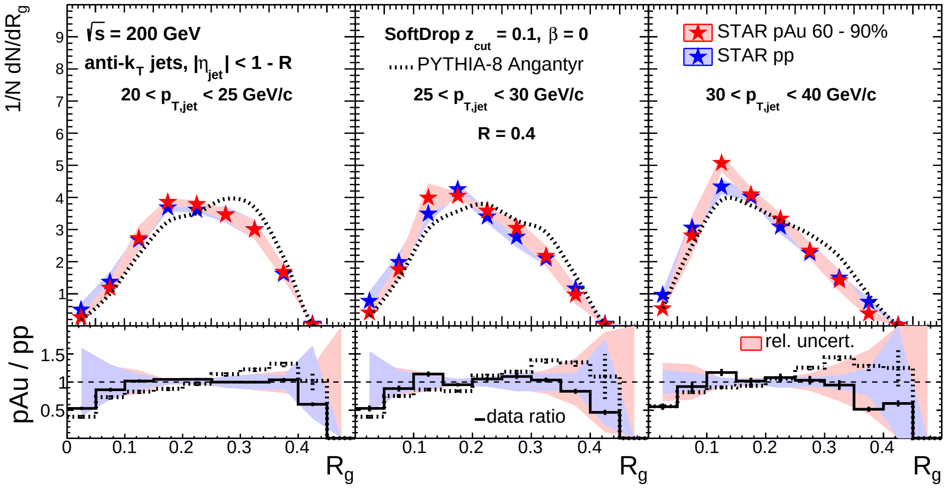


Figure F.3: The same as Fig. 6.12, but for a low (60–90%) event activity range.

BIBLIOGRAPHY

[1] R. P. Feynman. *QED: The Strange Theory of Light and Matter.* Princeton University Press, 1986.

[2] John Von Neumann. Method in the Physical Sciences. In L. Leary, editor, *The Unity of Knowledge*, pages 157–164. Doubleday, Garden City, NY, 1955.

[3] Carsten Burgard. Example: Standard Model of physics. `https://texample.net/tikz/examples/model-physics/`. Image by David Galbraith and Carsten Burgard. Programmed in TikZ by Carsten Burgard. Accessed: 4/13/2022.

[4] G.E.P. Box. Robustness in the strategy of scientific model building. In Robert L. Launer and Graham N. Wilkinson, editors, *Robustness in Statistics*, pages 201–236. Academic Press, 1979.

[5] Robert N. Cahn. The eighteen arbitrary parameters of the Standard Model in your everyday life. *Rev. Mod. Phys.*, 68:951–960, 1996.

[6] Oscar Miyamoto Gomez. Five mysteries the Standard Model can't explain. `https://www.symmetrymagazine.org/article/five-mysteries-the-standard-model-cant-explain`. Accessed: 12/24/2021.

[7] Jonathan L. Feng. Naturalness and the Status of Supersymmetry. *Ann. Rev. Nucl. Part. Sci.*, 63:351–382, 2013.

[8] Arthur Jaffe and Edward Witten. Yang–Mills and Mass Gap. `https://www.claymath.org/millennium-problems/yang%E2%80%93mills-and-mass-gap`. Accessed: 12/24/2021.

[9] CERN. The Higgs boson. `https://home.cern/science/physics/higgs-boson`. Accessed: 4/24/2022.

[10] Siegfried Bethke. Experimental tests of asymptotic freedom. *Prog. Part. Nucl. Phys.*, 58:351–386, 2007.

[11] H. David Politzer. Reliable Perturbative Results for Strong Interactions? *Phys. Rev. Lett.*, 30:1346–1349, 1973.

[12] David J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346, 1973.

[13] Kenneth G. Wilson. Confinement of quarks. *Phys. Rev. D*, 10:2445–2459, 1974.

[14] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[15] Sarah Webb. Early-universe soup. `https://deixismagazine.org/2016/06/early-universe-soup/`, June 2016. Image by Swagato Mukherjee. Accessed: 12/24/2021.

[16] R. Hagedorn. Statistical thermodynamics of strong interactions at high-energies. *Nuovo Cim. Suppl.*, 3:147–186, 1965.

[17] Raghunath Sahoo, Tapan K. Nayak, Jan-e Alam, Basanta K. Nandi, and Sonja Kabana. Probing the QGP Phase Boundary with Thermal Properties of ϕ Mesons. *Int. J. Mod. Phys. A*, 26:4145–4152, 2011.

[18] ATLAS Collaboration. Azimuthal momentum anisotropy. `https://atlas.cern/glossary/azimuthal-momentum-anisotropy`. Accessed: 5/24/2022.

[19] J. Adams et al. Azimuthal anisotropy in Au+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$. *Phys. Rev. C*, 72:014904, 2005.

[20] S. S. Adler et al. Elliptic flow of identified hadrons in Au+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$. *Phys. Rev. Lett.*, 91:182301, 2003.

[21] Wit Busza, Krishna Rajagopal, and Wilke van der Schee. Heavy Ion Collisions: The Big Picture, and the Big Questions. *Ann. Rev. Nucl. Part. Sci.*, 68:339–376, 2018.

[22] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989.

[23] Elke-Caroline Aschenauer, Kyle Lee, B. S. Page, and Felix Ringer. Jet angularities in photoproduction at the Electron-Ion Collider. *Phys. Rev. D*, 101(5):054028, 2020.

[24] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017.

[25] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[26] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.

[27] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

[28] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*, volume 8 of *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology.* Cambridge University Press, February 2011.

[29] Victor S. Fadin, E. A. Kuraev, and L. N. Lipatov. On the Pomeranchuk Singularity in Asymptotically Free Theories. *Phys. Lett. B*, 60:50–52, 1975.

[30] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. Multi - Reggeon Processes in the Yang-Mills Theory. *Sov. Phys. JETP*, 44:443–450, 1976.

[31] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. The Pomeranchuk Singularity in Nonabelian Gauge Theories. *Sov. Phys. JETP*, 45:199–204, 1977.

[32] I. I. Balitsky and L. N. Lipatov. The Pomeranchuk Singularity in Quantum Chromodynamics. *Sov. J. Nucl. Phys.*, 28:822–829, 1978.

[33] Francois Gelis, Edmond Iancu, Jamal Jalilian-Marian, and Raju Venugopalan. The Color Glass Condensate. *Ann. Rev. Nucl. Part. Sci.*, 60:463–489, 2010.

[34] Yuri V. Kovchegov and Eugene Levin. *Quantum Chromodynamics at High Energy*, volume 33 of *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology*. Cambridge University Press, August 2012.

[35] Tai Sakuma. *Inclusive Jet and Dijet Production in Polarized Proton-proton Collisions at $\sqrt{s}$ = 200 GeV at RHIC*. PhD thesis, MIT, 2009. `https://drupal.star.bnl.gov/STAR/files/tai_sakuma_thesis_0.pdf`.

[36] Zhong-Bo Kang, Kyle Lee, and Felix Ringer. Jet angularity measurements for single inclusive jet production. *JHEP*, 04:110, 2018.

[37] Stefan Höche. Introduction to parton-shower event generators. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pages 235–295, 2015.

[38] Jaroslav Adam et al. Measurement of groomed jet substructure observables in p+p collisions at $\sqrt{s}$ =200 GeV with STAR. *Phys. Lett. B*, 811:135846, 2020.

[39] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, March 2014.

[40] John E. Huth et al. Toward a standardization of jet definitions. In *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90)*, pages 0134–136, December 1990.

[41] Klaus Rabbertz. *Jet Physics at the LHC: The Strong Force beyond the TeV Scale*, volume 268 of *Springer Tracts in Modern Physics*. Springer, Berlin, 2017.

[42] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962.

[43] T. D. Lee and M. Nauenberg. Degenerate Systems and Mass Singularities. *Phys. Rev.*, 133:B1549–B1562, 1964.

[44] Gavin P. Salam and Gregory Soyez. A practical seedless infrared-safe cone jet algorithm. *JHEP*, 05:086, 2007.

[45] S. Catani, Yuri L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber. New clustering algorithm for multi - jet cross-sections in e+ e- annihilation. *Phys. Lett. B*, 269:432–438, 1991.

[46] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, 1993.

[47] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.

[48] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. In *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, pages 270–279, April 1998.

[49] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[50] Daniele Bertolini, Tucker Chan, and Jesse Thaler. Jet Observables Without Jet Algorithms. *JHEP*, 04:013, 2014.

[51] A. Banfi and M. Dasgupta. Problems in resumming interjet energy flows with k_t clustering. *Phys. Lett. B*, 628:49–56, 2005.

[52] Jaroslav Adam et al. Underlying event measurements in $p + p$ collisions at $\sqrt{s} = 200\,\mathrm{GeV}$ at RHIC. *Phys. Rev. D*, 101(5):052004, 2020.

[53] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119–126, 2008.

[54] Jaroslav Adam et al. Measurement of charged jet production cross sections and nuclear modification in p–pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. *Physics Letters B*, 749:68–81, 2015.

[55] S. Acharya et al. First measurement of jet mass in Pb–Pb and p–Pb collisions at the LHC. *Phys. Lett. B*, 776:249–264, 2018.

[56] Roman Kogler et al. Jet Substructure at the Large Hadron Collider: Experimental Review. *Rev. Mod. Phys.*, 91(4):045003, 2019.

[57] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014.

[58] André H. Hoang, Sonny Mantry, Aditya Pathak, and Iain W. Stewart. Nonperturbative Corrections to Soft Drop Jet Mass. *JHEP*, 12:002, 2019.

[59] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *JHEP*, 09:029, 2013.

[60] J. Adam et al. Longitudinal double-spin asymmetry for inclusive jet and dijet production in pp collisions at $\sqrt{s} = 510$ GeV. *Phys. Rev. D*, 100(5):052005, 2019.

[61] Vardan Khachatryan et al. Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC. *JHEP*, 09:091, 2010.

[62] Jaroslav Adam et al. Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions. *Nature Phys.*, 13:535–539, 2017.

[63] Georges Aad et al. Centrality and rapidity dependence of inclusive jet production in $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ proton-lead collisions with the ATLAS detector. *Phys. Lett. B*, 748:392–413, 2015.

[64] Jaroslav Adam et al. Measurement of charged jet production cross sections and nuclear modification in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. *Phys. Lett. B*, 749:68–81, 2015.

[65] Vardan Khachatryan et al. Measurement of inclusive jet production and nuclear modifications in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. *Eur. Phys. J. C*, 76(7):372, 2016.

[66] Jaroslav Adam et al. Measurement of dijet k_T in p–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. *Phys. Lett. B*, 746:385–395, 2015.

[67] S. Acharya et al. First measurement of jet mass in Pb–Pb and p–Pb collisions at the LHC. *Phys. Lett. B*, 776:249–264, 2018.

[68] Morad Aaboud et al. Measurement of jet fragmentation in 5.02 TeV proton-lead and proton-proton collisions with the ATLAS detector. *Nucl. Phys. A*, 978:65, 2018.

[69] U. Acharya et al. Measurement of jet-medium interactions via direct photon-hadron correlations in Au+Au and d +Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$. *Phys. Rev. C*, 102(5):054910, 2020.

[70] Shreyasi Acharya et al. Jet fragmentation transverse momentum distributions in pp and p-Pb collisions at $\sqrt{s}$, $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. *JHEP*, 09:211, 2021.

[71] L. Adamczyk et al. Dijet imbalance measurements in $Au + Au$ and pp collisions at $\sqrt{s_{NN}} = 200\,\mathrm{GeV}$ at STAR. *Phys. Rev. Lett.*, 119(6):062301, 2017.

[72] Georges Aad et al. Centrality and rapidity dependence of inclusive jet production in $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ proton-lead collisions with the ATLAS detector. *Phys. Lett. B*, 748:392–413, 2015.

[73] A. Adare et al. Centrality-dependent modification of jet-production rates in deuteron-gold collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$. *Phys. Rev. Lett.*, 116(12):122301, 2016.

[74] Nestor Armesto, Doga Can Gülhan, and Jose Guilherme Milhano. Kinematic bias on centrality selection of jet events in pPb collisions at the LHC. *Phys. Lett. B*, 747:441–445, 2015.

[75] Massimiliano Alvioli, Brian A. Cole, Leonid Frankfurt, D. V. Perepelitsa, and Mark Strikman. Evidence for x-dependent proton color fluctuations in pA collisions at the CERN Large Hadron Collider. *Phys. Rev. C*, 93(1):011902, 2016.

[76] Kari J. Eskola, Petja Paakkinen, Hannu Paukkunen, and Carlos A. Salgado. EPPS21: a global QCD analysis of nuclear PDFs. *Eur. Phys. J. C*, 82(5):413, 2022.

[77] Ivan Vitev. Non-Abelian energy loss in cold nuclear matter. *Phys. Rev. C*, 75:064906, 2007.

[78] Kari J. Eskola, Hannu Paukkunen, and Carlos A. Salgado. A perturbative QCD study of dijets in p+Pb collisions at the LHC. *JHEP*, 10:213, 2013.

[79] Zhong-Bo Kang, Jared Reiten, Ivan Vitev, and Boram Yoon. Light and heavy flavor dijet production and dijet mass modification in heavy ion collisions. *Phys. Rev. D*, 99(3):034006, 2019.

[80] J. Casalderrey-Solana, G. Milhano, D. Pablos, and K. Rajagopal. Modification of Jet Substructure in Heavy Ion Collisions as a Probe of the Resolution Length of Quark-Gluon Plasma. *JHEP*, 01:044, 2020.

[81] Brookhaven National Laboratory. 2011 site aerials. `https://www.flickr.com/photos/brookhavenlab/albums/72157627989559182`. Accessed: 2/24/2022.

[82] H. Hahn et al. The RHIC design overview. *Nucl. Instrum. Meth. A*, 499:245–263, 2003.

[83] Brookhaven National Laboratory. 200 MeV linear accelerator. `https://www.bnl.gov/rhic/linac.php`. Accessed: 2/25/2022.

[84] A. Pikin, J. G. Alessi, E. N. Beebe, A. Kponou, R. Lambiase, R. Lockey, D. Raparia, J. Ritter, L. Snydstrup, and Y. Tan. RHIC EBIS: Basics of design and status of commissioning. *JINST*, 5:C09003, 2010.

[85] D. B. Steski and P. Thieberger. Stripping foils at RHIC. *Nucl. Instrum. Meth. A*, 613:439–441, 2010.

[86] C. Liu et al. RHIC operations with asymmetric collisions in 2015, 2015. `https://www.agsrhichome.bnl.gov/RHIC/Runs/RhicRun15pA2.pdf`.

[87] CERN. H-jet measures beam polarization at RHIC. `https://cerncourier.com/a/h-jet-measures-beam-polarization-at-rhic/`. Accessed: 3/01/2022.

[88] K. Adcox et al. PHENIX detector overview. *Nucl. Instrum. Meth. A*, 499:469–479, 2003.

[89] EIC. Electron-Ion Collider achieves critical decision 1 approval. https://www.energy.gov/science/articles/electron-ion-collider-achieves-critical-decision-1-approval. Accessed: 3/01/2022.

[90] Alexander Schmah for the STAR Collaboration. Event anisotropy v_2 of identified hadrons and light nuclei in Au+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 7.7$, 11.5 and 39 GeV with STAR. https://indico.cern.ch/event/30248/contributions/1666942/attachments/572189/787999/QuarkMatter_aschmah_Final.pdf. Image by Maria Schmah and Alex Schmah. Accessed: 5/12/2022.

[91] K. H. Ackermann et al. STAR detector overview. *Nucl. Instrum. Meth. A*, 499:624–632, 2003.

[92] M. Anderson et al. The STAR Time Projection Chamber: A Unique Tool for Studying High Multiplicity Events at RHIC. *Nucl. Instrum. Meth. A*, 499:659–678, 2003.

[93] Nuclear Science and Technology. STAR Detector. https://nsww.org/projects/bnl/star/sub-systems.php. Accessed: 3/07/2022.

[94] M. Beddo et al. The STAR Barrel Electromagnetic Calorimeter. *Nucl. Instrum. Meth. A*, 499:725–739, 2003.

[95] C. W. Fabjan and T. Ludlam. Calorimetry in High-energy Physics. *Ann. Rev. Nucl. Part. Sci.*, 32:335–389, 1982.

[96] Les Bland. STAR Beam-Beam Counter (BBC) Front View. https://www.star.bnl.gov/public/bbc/geom/front_view.html. Accessed: 3/08/2022.

[97] W. J. Llope et al. The STAR Vertex Position Detector. *Nucl. Instrum. Meth. A*, 759:23–28, 2014.

[98] C. A. Whitten. The Beam-Beam Counter: A Local Polarimeter at STAR. *AIP Conf. Proc.*, 980(1):390–396, 2008.

[99] W. Herr and B. Muratori. Concept of luminosity. In *CERN Accelerator School and DESY Zeuthen: Accelerator Physics*, pages 361–377, September 2003.

[100] James Kevin Adkins. *Studying Transverse Momentum Dependent Distributions in Polarized Proton Collisions via Azimuthal Single Spin Asymmetries of Charged Pions in Jets*. PhD thesis, Kentucky U., 2017. `https://drupal.star.bnl.gov/STAR/files/KevinAdkinsThesis_Final.pdf`.

[101] David Stewart. *Jet to Event Activity Correlations in Small System Collisions at STAR*. PhD thesis, Yale U., 2021. `https://drupal.star.bnl.gov/STAR/files/Stewart_Thesis_nonumbers.pdf`.

[102] Isaac Mooney. Jet substructure in p+p and p+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV at STAR. *PoS*, HardProbes2020:144, 2021.

[103] Karen McNulty Walsh. Star Detector has a New Inner Core. `https://www.bnl.gov/newsroom/news.php?a=214492`. Accessed: 4/25/2022.

[104] T. M. Cormier, A. I. Pavlinov, M. V. Rykov, V. L. Rykov, and K. E. Shestermanov. STAR Barrel Electromagnetic Calorimeter absolute calibration using 'minimum ionizing particles' from collisions at RHIC. *Nucl. Instrum. Meth. A*, 483:734–746, 2002.

[105] L. Adamczyk et al. Precision Measurement of the Longitudinal Double-spin Asymmetry for Inclusive Jet Production in Polarized Proton Collisions at $\sqrt{s} = 200\,\mathrm{GeV}$. *Phys. Rev. Lett.*, 115(9):092002, 2015.

[106] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05(2006):026, 2006.

[107] René Brun, F. Bruyant, Federico Carminati, Simone Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban. GEANT Detector Description and Simulation Tool. October 1994.

[108] Mohamed Abdallah et al. Invariant Jet Mass Measurements in pp Collisions at $\sqrt{s} = 200\,\text{GeV}$ at RHIC. *Phys. Rev. D*, 104(5):052007, 2021.

[109] A. Majumder and J. Putschke. Mass depletion: a new parameter for quantitative jet modification. *Phys. Rev. C*, 93(5):054909, 2016.

[110] A. Moraes, C. Buttar, and I. Dawson. Prediction for minimum bias and the underlying event at LHC energies. *Eur. Phys. J. C*, 50:435–466, 2007.

[111] Daniele Anderle, Mrinal Dasgupta, Basem Kamal El-Menoufi, Jack Helliwell, and Marco Guzzi. Groomed jet mass as a direct probe of collinear parton dynamics. *Eur. Phys. J. C*, 80(9):827, 2020.

[112] Christian Bierlich et al. A comprehensive guide to the physics and usage of PYTHIA 8.3. March 2022. arXiv:2203.11601.

[113] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76:196, 2016.

[114] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J. C*, 58:639–707, 2008.

[115] The Herwig collaboration. Minimum-bias and underlying-event tunes. `https://herwig.hepforge.org/tutorials/mpi/tunes.html`, 2019. Accessed: 4/08/2022.

[116] P. K. Khandai, P. Sett, P. Shukla, and V. Singh. Hadron spectra in p+p collisions at RHIC and LHC energies. *Int. J. Mod. Phys. A*, 28:1350066, 2013.

[117] E. Norrbin and T. Sjostrand. Production mechanisms of charm hadrons in the string model. *Phys. Lett. B*, 442:407–416, 1998.

[118] G. D'Agostini. Improved iterative Bayesian unfolding. In *Alliance Workshop on Unfolding and Data Correction*, October 2010.

[119] Tim Adye. Unfolding algorithms and tests using RooUnfold. In *PHYSTAT 2011*, pages 313–318, Geneva, 2011. CERN.

[120] Armen Tumasyan et al. Study of quark and gluon jet substructure in Z+jet and dijet events from pp collisions. *JHEP*, 01:188, 2022.

[121] Zhong-Bo Kang, Kyle Lee, Xiaohui Liu, and Felix Ringer. Soft drop groomed jet angularities at the LHC. *Phys. Lett. B*, 793:41–47, 2019.

[122] Zhong-Bo Kang, Kyle Lee, and Felix Ringer. Jet angularity measurements for single inclusive jet production. *JHEP*, 04:110, 2018.

[123] Iain W. Stewart, Frank J. Tackmann, and Wouter J. Waalewijn. Dissecting Soft Radiation with Factorization. *Phys. Rev. Lett.*, 114(9):092001, 2015.

[124] Manny Rosales Aguilar, Zilong Chang, Raghav Kunnawalkam Elayavalli, Renee Fatemi, Yang He, Yuanjing Ji, Dmitry Kalinkin, Matthew Kelsey, Isaac Mooney, and Veronica Verkest. PYTHIA 8 underlying event tune for RHIC energies. *Phys. Rev. D*, 105(1):016011, 2022.

[125] John Lajoie. Recent reconstructed jet results from PHENIX. https://moriond.in2p3.fr/QCD/2021/FridayAfternoon/Lajoie.pdf. Accessed: 4/18/2022.

[126] Veronica Verkest for the STAR Collaboration. Measurements of jet and soft activity in $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ p+Au collisions at STAR. https://indico.cern.ch/event/895086/contributions/4724929/attachments/2421980/4145675/vverkest_QM2022_slides.pdf. Accessed: 4/19/2022.

[127] Christian Bierlich, Gösta Gustafson, Leif Lönnblad, and Harsh Shah. The Angantyr model for Heavy-Ion Collisions in PYTHIA8. *JHEP*, 10:134, 2018.

[128] Roger D. Woods and David S. Saxon. Diffuse Surface Optical Model for Nucleon-Nuclei Scattering. *Phys. Rev.*, 95:577–578, 1954.

[129] R. J. Glauber. Cross-sections in deuterium at high-energies. *Phys. Rev.*, 100:242–248, 1955.

[130] V. N. Gribov. Glauber corrections and the interaction between high-energy hadrons and nuclei. *Sov. Phys. JETP*, 29:483–487, 1969.

[131] A. Bialas and W. Czyz. Wounded nucleon model and Deuteron-Gold collisions at RHIC. *Acta Phys. Polon. B*, 36:905–918, 2005.

[132] Mohamed Abdallah et al. Longitudinal double-spin asymmetry for inclusive jet and dijet production in polarized proton collisions at $\sqrt{s} = 200\,\mathrm{GeV}$. *Phys. Rev. D*, 103(9):L091103, 2021.

[133] L. Adamczyk et al. Dijet imbalance measurements in $Au + Au$ and pp collisions at $\sqrt{s_{NN}} = 200\,\mathrm{GeV}$ at STAR. *Phys. Rev. Lett.*, 119(6):062301, 2017.

[134] U. Acharya et al. Measurement of jet-medium interactions via direct photon-hadron correlations in Au+Au and d +Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$. *Phys. Rev. C*, 102(5):054910, 2020.

[135] Shreyasi Acharya et al. Jet fragmentation transverse momentum distributions in pp and p-Pb collisions at $\sqrt{s}$, $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. *JHEP*, 09:211, 2021.

[136] Morad Aaboud et al. Measurement of jet fragmentation in 5.02 TeV proton-lead and proton-proton collisions with the ATLAS detector. *Nucl. Phys. A*, 978:65, 2018.

[137] Kirill Lapidus for the ALICE Collaboration. Hard substructure of jets probed in p-Pb collisions. `https://indico.cern.ch/event/433345/contributions/2358200/attachments/1407231/2150818/qm_poster_lapidus_2017_MF_final.pdf`. Accessed: 4/20/2022.

[138] Hannah Bossi for the ALICE Collaboration. R-dependence of inclusive jet suppression and groomed jet splittings in heavy-ion collisions with ALICE. `https://indico.cern.ch/event/895086/contributions/4736731/attachments/2423036/4147730/HBossi_QuarkMatter_2022.pdf`. Accessed: 4/20/2022.

[139] Andrew J. Larkoski, Ian Moult, and Duff Neill. Non-Global Logarithms, Factorization, and the Soft Substructure of Jets. *JHEP*, 09:143, 2015.

[140] Anders Andreassen, Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. OmniFold: A Method to Simultaneously Unfold All Observables. *Phys. Rev. Lett.*, 124(18):182001, 2020.

[141] A. Adare et al. An Upgrade Proposal from the PHENIX Collaboration. January 2015.

[142] Hai Tao Li, Ze Long Liu, and Ivan Vitev. Heavy meson tomography of cold nuclear matter at the electron-ion collider. *Phys. Lett. B*, 816:136261, 2021.

[143] Howard Matis, Rene Bellwied, Gary Eppley, Detlef Irmscher, Peter Jacobs, Spencer Klein, Tom LeCompte, Mike Lisa, Iwona Sakrejda, Marguerite Tonjes, Dhammika Weerasundara, Pablo Yepes, and Roy Bossingham. STAR Geometry. `https://www.star.bnl.gov/public/archives/integration/STAR/STAR_Notes/csn0229B.pdf`. Accessed: 4/14/2022.

[144] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J. C*, 75(5):204, 2015.

[145] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev. D*, 82:074018, 2010.

[146] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.

[147] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C*, 74(8):3024, 2014.

[148] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, and Juan Rojo. Parton distributions with QED corrections. *Nucl. Phys. B*, 877:290–320, 2013.

[149] Zilong Chang. Inclusive Jet Cross Section Measurements in pp Collisions at $\sqrt{s} = 200$ and 510 GeV with STAR. In *Particles and Nuclei International Conference*, November 2021. arXiv:2111.08149.

ABSTRACT

JET SUBSTRUCTURE MEASUREMENTS IN PROTON-PROTON AND PROTON-GOLD COLLISIONS

by

ISAAC MOONEY

August 2022

Advisor: Dr. Joern Putschke

Major: Physics

Degree: Doctor of Philosophy

Observing a modification of high-energy probes in the hot nuclear environment created by ultrarelativistic heavy-ion collisions is critical for understanding the microscopic structure of the QCD medium. However, before a modification can be observed, the probes must be calibrated. For this reason, measurements of jet observables in vacuum (pp collisions) and in the presence of cold nuclear matter (pAu collisions) are presented, which will allow for future comparison of jet substructure measurements in AuAu collisions to these baseline measurements.

Data from the STAR detector on the Relativistic Heavy Ion Collider at Brookhaven National Laboratory are used for both analyses. The pp measurement uses data taken in 2012 with protons colliding with center-of-mass energy of $\sqrt{s} = 200\,\mathrm{GeV}$, while the pAu measurement uses data taken in 2015 at the same energy per nucleon pair. Data are fully corrected for detector effects using a two-dimensional Bayesian unfolding procedure, including a reweighting of the detector response matrix in the pAu analysis to account for the additional underlying event in the simulated event embedding.

The observables presented in the pp analysis are the jet mass, defined as the magnitude of the four-momentum sum of jet constituents, and the SoftDrop groomed jet mass. The SoftDrop grooming algorithm is known to reduce the effect of non-perturbative radiation on jet substructure. The pp results are presented differentially in jet p_{T} and anti-k_{T} jet

resolution parameter, R, to investigate the dependence of jet mass on the momentum and angular scales of the jet shower. Results are compared to a number of leading-order Monte Carlo simulations as well as an NLL calculation by Lee et al. It is found in this first measurement of inclusive jet mass at RHIC that the effect of hadronization is large in this regime and can indeed be reduced significantly with SoftDrop. Tension between some models and the data helped to motivate a now-completed simulation tuning effort at STAR, and it is hoped that the disagreement between the data and the calculation can be used to obtain a non-perturbative shape function for convolution with parton-level calculations.

In the pAu analysis, a suite of SoftDrop observables are presented, including the SoftDrop subjet shared momentum fraction, z_{g}, groomed jet radius, R_{g}, and groomed jet mass, M_{g}, as well as the inclusive jet mass, M. Comparison to the published pp data shows that regardless of event activity, and across a range in p_{T}, the jet substructure is consistent between these two collision systems in this kinematic regime. Additional comparisons to a heavy-ion model with no collective effects are inconclusive due to the need to tune the model to RHIC kinematics. Finally, by comparing low- and high-event-activity classes in the data, it is shown that there are no significant activity-dependent final state effects, which can be seen as evidence against jet quenching in a hot nuclear environment. In general, no cold or hot nuclear matter effects on the jet substructure are observed.

AUTOBIOGRAPHICAL STATEMENT

Name: Isaac A. Mooney

Education:
B.S. Physics and Mathematics, University of Michigan, Ann Arbor, MI, 2016

Teaching:
TA, Physics for the Life Sciences Laboratory, 2017

Selected Publications:
PYTHIA8 underlying event tune for RHIC energies – analysis workforce. Published in Phys. Rev. D **105**, 016011 (2022).
Invariant Jet Mass Measurements in pp Collisions at $\sqrt{s} = 200\,\mathrm{GeV}$ *at RHIC* – primary analysis contact. Published in Phys. Rev. D **104**, 052007 (2021).
Jet substructure in p+p and p+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ *at STAR*. I. Mooney for the STAR Collaboration. Proceedings of the 10th International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions (Hard Probes 2020). Published in PoS HardProbes2020 (2021) 144.
Co-author on an additional 47 STAR publications and 2 SeaQuest publications. See my INSPIRE author profile and additional publications here.

Selected Talks:

Jet and jet substructure measurements in small systems at RHIC and the LHC. I. Mooney. **Institute for Nuclear Theory program 21-2b: Probing QCD at High Energy and Density with Jets**. Jul. 26 – Aug. 27, 2021, Seattle, Washington (virtual). (Invited talk).

Jets and heavy quarkonia production in small system collisions at RHIC. I. Mooney. **RHIC & AGS Annual Users' Meeting, Small Systems Workshop**. Jun. 8 – 11, 2021 (virtual). (Invited parallel).

Summary of recent results on jet production in pp, pA, and AA collisions. I. Mooney. **STAR Collaboration Meeting**. Sep. 14 – 25, 2020, Tirupati, India (virtual). (Invited plenary).

Jet substructure in p+p and p+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$ *at STAR*. I. Mooney for the STAR Collaboration. **10th International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions (Hard Probes 2020)**. May 31 – Jun. 5, 2020, Austin, TX, USA (virtual). (Parallel).

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346 USA